IEEE
Signal Processing
MAGAZINE

[VOLUME 32 NUMBER 6 NOVEMBER 2015]

THE SCIENCE BEHIND OUR DIGITAL LIFE

EUCLIDEAN DISTANCE MATRICES

PLAYING WITH DUALITY FOR LARGE-SCALE OPTIMIZATION

EXPRESSION CONTROL IN SINGING VOICE SYNTHESIS

SPEAKER RECOGNITION

SPARSE AND TENSOR MODELS FOR BRAIN IMAGING

IEEE Signal Processing Society

IEEE

USB & Ethernet Programmable ATTENUATORS

New Models up to 120 dB!

0–30, 60, 90, 110 & 120 dB 0.25 dB Step 1 MHz to 6 GHz from $395*

Mini-Circuits' new programmable attenuators offer precise attenuation from 0 up to 120 dB, supporting even more applications and greater sensitivity level measurements! Now available in models with maximum attenuation of 30, 60, 90, 110, and 120 dB with 0.25 dB attenuation steps, they provide the widest range of level control in the industry with accurate, repeatable performance for a variety of applications including fading simulators, handover system evaluation, automated test equipment and more! Our unique designs maintain linear attenuation change per dB over the entire range of attenuation settings, while USB, Ethernet and RS232 control options allow setup flexibility and easy remote test management. Supplied with user-friendly GUI control software, DLLs for programmers† and everything you need for immediate use right out of the box, Mini-Circuits programmable attenuators offer a wide range of solutions to meet your needs and fit your budget. Visit minicircuits.com for detailed performance specs, great prices, and off the shelf availability. Place your order today for delivery as soon as tomorrow!

RoHS compliant

	Models	Attenuation Range	Attenuation Accuracy	Step Size	USB Control	Ethernet Control	RS232 Control	Price Qty. 1-9
	RUDAT-6000-30	0-30 dB	±0.4 dB	0.25 dB	✓	-	✓	$395
	RCDAT-6000-30	0-30 dB	±0.4 dB	0.25 dB	✓	✓	-	$495
	RUDAT-6000-60	0-60 dB	±0.3 dB	0.25 dB	✓	-	✓	$625
	RCDAT-6000-60	0-60 dB	±0.3 dB	0.25 dB	✓	✓	-	$725
	RUDAT-6000-90	0-90 dB	±0.4 dB	0.25 dB	✓	-	✓	$695
	RCDAT-6000-90	0-90 dB	±0.4 dB	0.25 dB	✓	✓	-	$795
NEW	RUDAT-6000-110	0-110 dB	±0.45 dB	0.25 dB	✓	-	✓	$895
NEW	RCDAT-6000-110	0-110 dB	±0.45 dB	0.25 dB	✓	✓	-	$995
NEW	RUDAT-4000-120	0-120 dB	±0.5 dB	0.25 dB	✓	-	✓	$895
NEW	RCDAT-4000-120	0-120 dB	±0.5 dB	0.25 dB	✓	✓	-	$995

*120 dB models specified from 1-4000 MHz.
†No drivers required. DLL objects provided for 32/64-bit Windows® and Linux® environments using ActiveX® and .NET® frameworks.

Mini-Circuits®

www.minicircuits.com P.O. Box 350166, Brooklyn, NY 11235-0003 (718) 934-4500 sales@minicircuits.com

523 Rev D

[CONTENTS]

[VOLUME 32 NUMBER 6]

[FEATURES]

Digital Object Identifier 10.1109/MSP.2015.2467197

[COLUMNS]

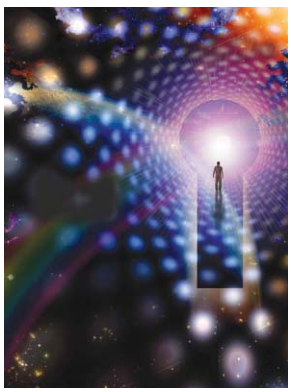
[DEPARTMENTS]

IEEE SignalProcessing Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page Qmags THE WORLD'S NEWSSTAND®

IEEE SIGNAL PROCESSING magazine

[COVER]

ISTOCKPHOTO.COM/BESTDESIGNS



IEEE SIGNAL PROCESSING MAGAZINE (ISSN 1053-5888) (ISPREG) is published bimonthly by the Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, 17th Floor, New York, NY 10016-5997 USA (+1 212 419 7900). Responsibility for the contents rests upon the authors and not the IEEE, the Society, or its members. Annual member subscriptions included in Society fee. Nonmember subscriptions available upon request. Individual copies: IEEE Members US$20.00 (first copy only), nonmembers US$213.00 per copy. Periodicals postage paid at New York, NY, and at additional mailing offices. Postmaster: Send address changes to IEEE Signal Processing Magazine, IEEE, 445 Hoes Lane, Piscataway, NJ 08854 USA. Canadian GST #125634188

Printed in the U.S.A.

Digital Object Identifier 10.1109/MSP.2015.2476995

SCOPE: *IEEE Signal Processing Magazine* publishes tutorial-style articles on signal processing research and applications, as well as columns and forums on issues of interest. Its coverage ranges from fundamental principles to practical implementation, reflecting the multidimensional facets of interests and concerns of the community. Its mission is to bring up-to-date, emerging and active technical developments, issues, and events to the research, educational, and professional communities. It is also the main Society communication platform addressing important issues concerning all members.

[from the EDITOR]

Min Wu
Editor-in-Chief
minwu@umd.edu

Engaging Undergraduate Students

Twenty years ago I joined the IEEE and the IEEE Signal Processing Society (SPS) as a Student Member. I still remember my excitement when I received my copy of *IEEE Signal Processing Magazine* (*SPM*) in the mail, which was a big deal for an undergraduate student! I probably only had the background to understand part of the content in the magazine, but still, it was valuable exposure to this exciting field.

Jack Deller was the editor-in-chief of the very first issue of *SPM* I received. It was only last year that I had the opportunity to meet him in person, but his leadership effort paved a foundation for the critical growth of *SPM* in 1991–1997. *SPM* was attractive to many young people, including me, to pursue signal processing.

My graduate study years coincided mostly with Aggelos Katsaggelos' term as editor-in-chief (1997–2002). *SPM* has served as an important reference for graduate students like me. I remember reading the wonderful series of overviews reflecting the past, present, and future of a number of technical areas in celebration of the 50th anniversary of SPS. Under the transformative leadership of K.J. Ray Liu (2003–2005), *SPM* reformed its operation with openness and diversity, expanded its content coverage, modernized its design, and topped the citation impact ranking. Since then, I was fortunate to have opportunities to become a part of the *SPM* team and work closely with three recent editors-in-chief, Shih-Fu Chang (2006–2008), Li Deng (2009–2011), and Abdelhak Zoubir (2012–2014). Through their collective efforts, these colleagues before me have brought about a high reputation for the magazine.

Digital Object Identifier 10.1109/MSP.2015.2468691
Date of publication: 13 October 2015

Given the depth and breadth of *SPM* articles, it is not surprising that this magazine contributes to the technical growth and enrichment of graduate students and researchers. Still, I can't help but recalling where I first started reading the magazine—as an undergraduate. What can *SPM* do to serve and engage undergraduate students, the future generation of our Society? Here are a few highlights.

This year, we engaged in active discussions with the magazine team and many readers on how to make articles accessible, particularly for students and practitioners. We reached a consensus to uphold *SPM*'s tradition in keeping the number of mathematical equations to the minimum amount necessary; combined with other practices on presentation styles, the goal is to make articles appealing to the majority of our readership. It's easier said than done, and this may take some time for authors to work on their articles. We appreciate their cooperation.

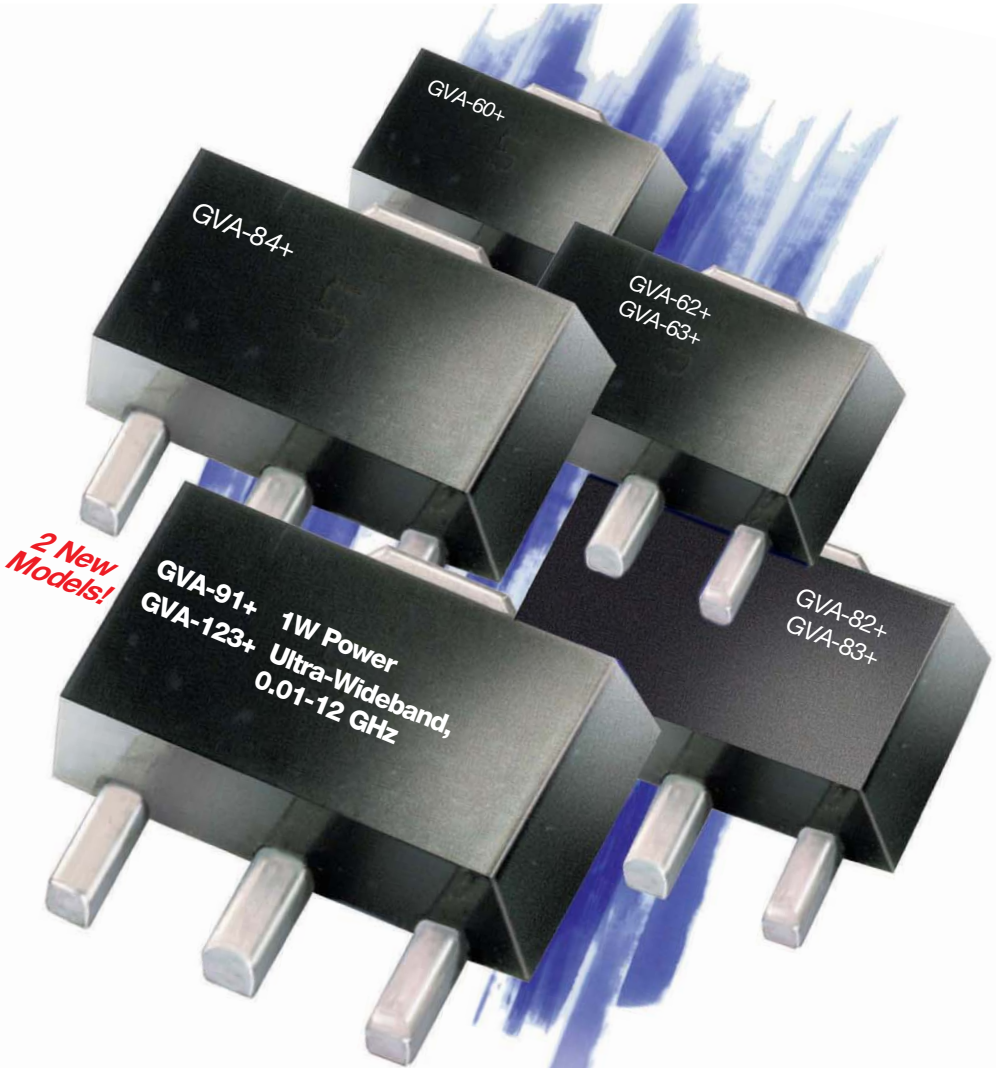
We have also been soliciting articles and special issues on timely topics that can draw readers' attention and stimulate their interests. Signal processing for computational photography and smart vehicles are two such examples that students and other readers can relate to their everyday lives. We look forward to sharing these with you in the coming year.

In parallel, we are bringing in-depth coverage of student activities. The July, September, and November 2015 issues of the magazine have featured a series of articles on the SP Cup competition, the Society's new initiative to engage undergraduate students. Special thanks to the past and current Student Service Directors Kenneth Lam and Patrizio Campisi, respectively, and the competition organizers, Carlos Sorzano and Zhilin Zhang, for their informative articles about the first two SP Cup competitions. The SP Cup is now open for the third edition. You can find more information in the "SP Education" column on page 113 in this issue.

We have also opened up the prestigious platform of the magazine to the students' voices and thoughts so that the magazine is not just a passive one-way communication to these burgeoning minds. For the first time, articles in the magazine included reflections in the students' own words as they participated in (and won) the SP Cup competition. Invitations have also been extended to the broad community to share their thoughts about career perspectives and signal processing in everyday life. In addition, we have been working with a group of volunteers to gather and compile contributions from undergraduate students and educators on exciting undergraduate design projects related to signal and information processing. Stay tuned for this content, and please encourage undergraduate students to contribute by answering the call for contributions that are open.

Beyond pursuing cutting-edge research, many undergraduate and graduate students with signal processing training usually join industry workforces. Students need to stay current, track the technical trends, gather practical tips and know-how, and build and extend their professional network. We are working on shaping timely, accessible, and informative content to meet their needs. It is a privilege for *SPM* to welcome undergraduates at the beginning of their careers and stay by their sides to offer them a helping hand. Please do not hesitate to give us feedback on how we are doing and suggestions on what we can do to serve you better.

[SP]



GVA AMPLIFIERS

NOW DC* to 12 GHz up to 1W Output Power from 94¢ ea. (qty. 1000)

GVA amplifiers now offer more options and more capabilities to support your needs. The new **GVA-123+** provides ultra-wideband performance with flat gain from 0.01 to 12 GHz, and new model **GVA-91+** delivers output power up to 1W with power added efficiency up to 47%! These new MMIC amplifiers are perfect solutions for many applications from cellular to satellite and more! The GVA series now covers bands from DC to 12 GHz with various combinations of gain, P1dB, IP3, and noise figure to fit your application. Based on high-performance InGaP HBT technology, these amplifiers are unconditionally stable and designed for a single 5V supply in tiny SOT-89 packages. All models are in stock for immediate delivery! Visit minicircuits.com for detailed specs, performance data, export info, **free X-parameters**, and everything you need to choose your GVA today! *US patent 6,943,629*

*Low frequency cut-off determined by coupling cap.
For GVA-60+, GVA-62+, GVA-63+, and GVA-123+ low cut off at 10 MHz.
For GVA-91+, low cut off at 869 MHz.

NOTE: GVA-62+ may be used as a replacement for RFMD SBB-4089Z
GVA-63+ may be used as a replacement for RFMD SBB-5089Z
See model datasheets for details

FREE X-Parameters-Based
Non-Linear Simulation Models for ADS Modelithics®
http://www.modelithics.com/mvp/Mini-Circuits.asp

Mini-Circuits®

www.minicircuits.com P.O. Box 350166, Brooklyn, NY 11235-0003 (718) 934-4500 sales@minicircuits.com

458 rev P

[president's MESSAGE]

Alex Acero
2014–2015 SPS President
a.acero@ieee.org

Signal Processing: The Science Behind Our Digital Life

Signal processing is found in almost every modern gadget. Smartphones allow a user to input text with his/her voice, take high-quality photos, and authenticate him/herself through fingerprint analysis. Wearable devices reveal heart rate and calories burned during a run. Consumers' TV experiences include access to content in 3D and 4K, with conveniences such as pause and rewind. Game consoles let users interact with the game by tracking their arm motions. Hearing aids improve millions of lives. Ultrasound machines and medical scans are life-saving advances in health care. These are just a few of the benefits we gain from signal processing.

While signal processing is a key technology in most consumer devices, the term remains invisible—or irrelevant—to most people. If we want our gadgets to continue to expand the range of features powered by signal processing, we need a higher number of engineers trained in this field coming out of engineering schools. But many college freshmen don't know what signal processing is and thus may decide to pursue other fields. If we want to advance the field of signal processing, we also need continued research funding—yet decision makers in funding government agencies don't necessarily understand that, to build such new capabilities, you need more than just computer scientists.

Digital Object Identifier 10.1109/MSP.2015.2472615
Date of publication: 13 October 2015

To tackle our visibility challenge, the Board of Governors of the IEEE Signal Processing Society (SPS) set up a committee to investigate this issue with the help of a public relations firm. The strategic awareness plan focused on creating excitement for signal processing and spurring a desire for students to pursue the field as a viable career path. The first outcome of this plan was the tagline "Signal Processing: The Science Behind Our Digital Life." This message emphasizes the criticality of signal processing in daily life and is flexible, allowing us to create a variety of stories that relate to students, professionals, and grant writers while maintaining a consistent brand.

In addition, SPS President-Elect Rabab Ward has been leading the development of short videos that explain what signal processing is and, at the same time, motivate viewers to learn more. A two-minute video, "What Is Signal Processing?," with an overview of the field [1] was uploaded in September 2014. The next step was videos focused on areas within signal processing, with a six-minute video [2] "Signal Processing and Machine Learning," which was uploaded in July 2015. I encourage you to watch both of them. And stay tuned for more videos!

We are also working on other initiatives to address the needs of the four constituencies we've identified: students, practicing engineers, academics, and the general public. Initiatives include a new website, to go live soon, and social media efforts to build on the Society's existing Facebook, LinkedIn, and Twitter presence. We have set up an IT committee led by Rony Ferzli to oversee the computing infrastructure needed to support our website, as well as mobile apps for Society conferences.

This process starts by creating and driving a focused dialogue among industry influencers, publishers, other companies, and customers. We have identified lead spokespeople, academic and industry contributors, and social engagement drivers to help us in this quest. Some of the things we'll be doing include: proactive media outreach, launch and promote monthly Twitter chats, promote SPS Technical Committee demonstration videos, upload a series of whiteboard videos on signal processing applications, and create a toolkit for SPS Chapters to engage students and industry members locally.

If you have any suggestions on how to improve the visibility of signal processing, please contact me at a.acero@ieee.org or SPS Membership and Content Administrator Jessica Perry at jessica.perry@ieee.org.

REFERENCES

[1] YouTube.com. "What is signal processing?" [Online]. Available: https://youtu.be/EErkgr1MWw0

[2] YouTube.com. "Signal processing and machine learning." [Online]. Available: https://youtu.be/EmexN6d8QF9o

[SP]

[special REPORTS]

John Edwards

Opening the Door to Innovative Consumer Technologies

For decades, signal processing has played a key role in the development of sophisticated consumer products. From personal audio and video systems to cameras to smartphones to satellite navigation systems and beyond, signal processing has helped manufacturers worldwide develop a wide range of innovative and affordable consumer devices.

Now, consumer electronics manufacturers are entering fresh areas, including new types of personal entertainment, diet management, and health-monitoring devices. As products in these and other emerging fields are designed, tested, and brought to market, signal processing continues to play essential roles in advancing device innovation, performance, and efficiency.

MUSIC THAT KNOWS YOU

The Internet is awash with millions upon millions of songs, more melodies than anyone could ever listen to in several lifetimes. As the sea of Internet music expands and deepens, finding new songs that matches one's personal tastes and preferences is becoming increasingly more difficult and time-consuming.

Back in 2006, Gert Lanckriet (Figure 1), a professor of electrical and computer engineering at the University of California, San Diego (UCSD), recognized this problem and began designing algorithms and related coding that would enable computers to automatically analyze and annotate musical content. Working with researchers at the UCSD Computer Audition Laboratory, Lanckriet started assembling the foundation for a new breed of Internet music search and recommendation engines that would automate discovery and play-listing for online music platforms.

"We decided to use signal processing and machine learning to build a system and a technology that ingested audio waveforms, basically songs, and then automatically associated those songs with tags," Lanckriet says. "We built this technology where a computer listens to millions of songs—with signal processing and machine learning—to associate them with the appropriate tags, and one could then search for music by entering certain semantic descriptions." Such descriptions could be almost anything related to a musical style, genre, instrument, or mood, such as "cool jazz" or "romantic soprano opera aria."

The research led to the development of a prototype music search engine that Lanckriet describes as a "Google for Music." To give the pattern recognition technology enough examples to work with, the researchers placed a crowdsourcing game called *Herd It* on Facebook. "The person who was playing the game would listen to a song, and then would provide feedback as to which tags are applicable to that song," Lanckriet says.

"You take all these examples and our algorithms will process them using signal processing, and then pattern recognition, to figure out what audio patterns associated with some tags have in common," Lanckriet says. Once enough sample tagged songs are collected, the algorithms are able to work at maximum effectiveness and efficiency to apply relevant tags to any music their host computer encounters. "We need this type of automation, since even people manually tagging songs all day long could never keep pace with the number of songs now being uploaded to the Internet."

The researchers then moved to improve the technology by giving users the ability to find songs that are similar to tunes that they already know and listen to regularly. "We were able to also create algorithms where users would create a query not with a semantic question, but with five or ten songs they like, and we would generate playlists of similar songs," Lanckriet explains. "Again, it uses signal processing and machine learning to do that."

Now, Lanckriet and his team are planning to break new ground by giving people the ability to listen to music that matches their current mood, environment, and activity via the various sensors built into mobile phones, smart watches, and other wearable devices. "These devices are literally jam-packed with things like accelerometers, gyroscopes, microphones, light sensors, temperature sensors, heart rate monitors, and so on," Lanckriet says. "We can use all of these sensors to detect the wearer's current activity and mood and then deliver the most appropriate music for that situation."

Lanckriet admits that, as the project moves forward, he and his coresearchers are facing a complex challenge. "What features or descriptors will allow us to extract the most meaningful information from each of these sensor signals?" he asks. "You have all of these different types of signals, and you can't process every signal in the same way; an audio signal is not the same as an accelerometer signal." With many mobile devices now incorporating as many as a dozen sensors, the task is daunting. "That is one very specific challenge on the signal processing level that we will have to address," Lanckriet says.

Although the project is still in its early stages, Lanckriet is already looking toward potential consumer applications. "Eventually, we hope that there will be personalized radio stations that select songs for you that are adapted to your mood or activity at various times during the day," he says.

Digital Object Identifier 10.1109/MSP.2015.2457471
Date of publication: 13 October 2015

1053-5888/15©2015IEEE

[FIG1] Gert Lanckriet, a professor of electrical and computer engineering at UCSD, records a sample melody. (Photo courtesy of Gert Lanckriet, UCSD.)

[FIG2] Edward Sazonov, an associate professor of electrical and computer engineering at the University of Alabama, working at his test bench. (Photo courtesy of Edward Sazonov, University of Alabama.)

"The system, for example, would know you like to listen to a certain type of music while you are waking up or working out and play it automatically at those times."

HONESTY IN EATING

Obesity is one of the world's major health concerns. Compounding the problem and its treatment is the fact that people are not always honest with themselves when it comes to tracking what they eat and then reporting that information to physicians, nutritionists, and other health professionals.

Edward Sazonov, an associate professor of electrical and computer engineering at the University of Alabama (Figure 2), wants to increase diet data accuracy with a sensor device that is worn unobtrusively around its user's ear. The unit would silently and automatically track meals, giving both consumers and health professionals honest, accurate information. "The sensor could provide objective data, helping us better understand patterns of food intake associated with obesity and eating disorders," he says. Sazonov is the lead on a US$1.8 million, five-year grant from the National Institutes of Health to test the practical accuracy of the wearable sensor in tracking diet.

According to Sazonov, the Automatic Ingestion Monitor (AIM) records its user's food intake by automatically capturing food images and then estimating the mass and the energy content of ingested food. The sensor feels vibrations from movement in the jaw during food intake, and the device is designed to filter out jaw motions, such as speaking, that are not related to either drinking or eating. "Through signal processing and pattern recognition, we are able to recognize this jaw motion on top of other activities such as speaking, or physical activity such as walking, and differentiate these," Sazonov explains.

The current prototype (Figure 3) is based on an earpiece with a camera that is mounted on the top of the unit and a piezoelectric jaw motion sensor located on the bottom. Collected data is relayed wirelessly via Bluetooth technology to an application running on a smartphone or tablet.

An earlier lanyard-worn prototype used a custom-built electronic circuit powered by a Li-polymer 3.7 V 680 mA h battery. The circuit incorporated an MSP430F2417 processor with an eight-channel, 12-bit analog-to-digital converter that was used to sample analog sensor signals. Also included were an RN-42 Bluetooth module with a serial port profile, a preamplifier for the jaw motion sensor (sampled at 1 kHz), a radio-frequency receiver for the hand-to-mouth gesture sensor (sampled at 10 Hz), a low-power three-axis accelerometer for capturing body acceleration (sampled at 100 Hz), and a self-report push button (sampled at 10 Hz) that was used in tests for pattern recognition algorithm development and validation and is not required in current models.

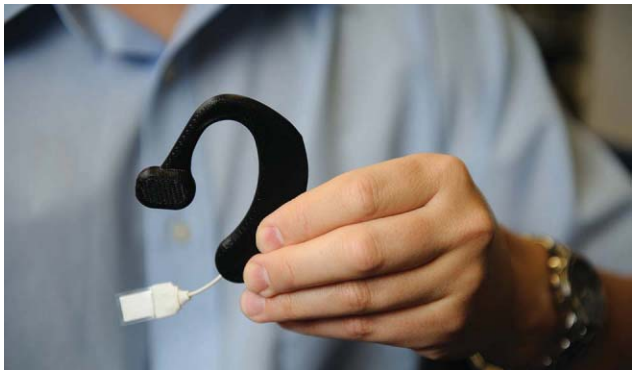
Signal processing plays a critical role in the device's operation. "We use a number of different techniques, particularly filtering and noise cancellation techniques," Sazonov says. Because AIM's sensors register information from multiple sources, independent component analysis is used to differentiate the sources. "We want to hear the chewing but eliminate the other physical activity in the signal," Sazonov says.

As with any wearable device, it is important to find ways of speeding performance while minimizing power consumption and overall device size and weight. "We are actually looking at signal processing methods that have lower computational intensity," Sazonov says. "For example, if you can substitute one feature that requires a lot of processing power with a similar feature that can do the job in recognizing food intake events but requires maybe hundreds or thousands time less computing power, that is what we are doing."

The information AIM generates could be used to improve behavioral weight-loss strategies or to develop new kinds of weight-loss interventions. In addition, the AIM could also provide an objective method of assessing the effectiveness of pharmacological and behavioral interventions for eating disorders.

BREATHING EASIER

Helping people determine the quality of the air they are breathing was the goal of a team of Carnegie Mellon University robotics researchers four years ago as they began developing the technology that would ultimately be known as Speck. The researchers, led by Illah Nourbakhsh, a Carnegie Mellon professor of robotics, envisioned Speck as a personal air pollution monitor that would enable users to monitor the



[FIG3] Edward Sazonov, an associate professor of electrical and computer engineering at the University of Alabama, holds a prototype AIM. (Photo courtesy of Edward Sazonov, University of Alabama.)

[FIG4] The Speck air-quality monitor allows users to view the level of fine particulate matter suspended in the air inside their homes. (Photo courtesy of Airviz, Inc.)

level of fine particulate matter suspended in the air inside their homes, helping them assess if their health is at risk.

With Speck now completed and available for sale (Figure 4), the researchers feel that they have reached their planned goal. Speck offers consumers insights into exposure to particulates known as PM2.5. PM2.5 particles are air pollutants with a diameter of 2.5 μm or lower, small enough to invade even the smallest airways. These particles generally come from activities that burn fossil fuels, such as traffic, as well as industrial activities such as smelting and metal processing. Knowledge of current particulate levels can help people to reduce exposure by opening or closing windows, altering activities, or taking action such as using high-efficiency particulate air filters. "It is designed to be used by your average, everyday citizen as well as scientists," says Mike Taylor, a Carnegie Mellon Ph.D. student who worked on Speck's signal processing.

"The device is more than just a sensor; it is a complete data system," observes Taylor. With a display screen that shows whether unhealthy levels of particulates are present, Speck users can view the current estimate of 2 μm particle concentration as well as a scaled estimate of PM2.5 in $\mu g/m^3$. The interface can also graph the past hour or past 12 hours of data on screen, allowing for quick access to historical data. Speck contains onboard signal processing and storage in addition to a color LCD touchscreen for the user interface. Power is supplied via a USB cable, and data can be downloaded directly to any Macintosh or Windows computer.

Speck is also Wi-Fi-enabled, which allows monitoring data to be uploaded to a user-controlled database for future reference. A companion website stores the data and provides analytical tools, including links to readings from federal air monitoring stations. Users are given free access to the site and can decide whether to share their monitoring data. Speck data can also integrate vital signs and other personal data recorded by personal fitness monitoring devices, such as Fitbit and Jawbone.

Inexpensive particulate sensor products tend to be relatively inaccurate, producing readings that sometimes do not even match the readings of identical models. The researchers claim they were able to achieve substantially higher accuracy by employing machine-learning algorithms that learn to recognize and compensate for the spurious noise in the sensor signals while maintaining affordability.

To reduce costs, Speck uses a commonly available, inexpensive DSM501a dust sensor instead of custom optics. The dust sensor's output is a digital pin that is pulled low when particles are detected in the optical chamber. The duty cycle is approximately proportional to the number of detected particles. The period of the sensor varies greatly, however, especially at low particle concentrations. While the duration of a low pulse (indicating detected particles) rarely exceeds 100 ms, the duration between pulses can last from under one second to more than one minute. "We observe that single-cycle readings are too noisy to be used directly," Taylor says. "Instead, our algorithm samples the sensor 10,000 times per second and uses the number of low samples each second as an input to an asymmetric filtering function."

Ultimately, the researchers plan to have the Speck measure particle counts as well as mass concentration in $\mu g/m^3$, the most common reporting method for PM2.5 among state and federal monitoring stations. This would allow users to compare their indoor air quality with the outdoor air quality of their local region, Taylor says.

This goal raises some new challenges, however, since the current inexpensive sensor is optical rather than mass based, which would be needed for accurate mass concentration reporting. In the current model, mass readings are estimated using a linear scale factor generated by applying Speck particle count data to that of a colocated tapered element oscillating microbalance monitor owned by the Allegheny County (Pennsylvania) Health Department. Other, less challenging, planned improvements include onboard humidity and temperature measurements that will refine both the particle count accuracy as well as the mass estimate.

Speck was developed over a period of four years in Carnegie Mellon's Community Robotics, Education, and Technology Empowerment Lab, and Nourbakhsh has established a spinoff company, Airviz Inc., to market the device.

AUTHOR

John Edwards (jedwards@johnedwardsmedia.com) is a technology writer based in the Phoenix, Arizona, area. [SP]

2016 IEEE Technical Field Award Recipients Announced

Each year, the IEEE recognizes individuals who have made outstanding contributions or exercised leadership within IEEE-designated technical areas. The IEEE Signal Processing Society is honored to announce five of its members as recipients of the 2016 IEEE Technical Field Awards:

IEEE James L. Flanagan Speech and Audio Processing Award: presented to Takehiro Moriya "for contributions to speech and audio coding algorithms and standardization"

IEEE Fourier Award for Signal Processing: presented to Bede Liu "for foundational contributions to the analysis, design, and implementation of digital signal processing systems"

IEEE Gustav Robert Kirchhoff Award: presented to P.P. Vaidyanathan "for fundamental contributions to digital signal processing"

CONGRATULATIONS TO ALL OF THE RECIPIENTS! THE FULL LIST OF 2016 IEEE TECHNICAL FIELD AWARDS RECIPIENTS CAN BE FOUND IN [1].

IEEE Leon K. Kirchmayer Graduate Teaching Award: presented to K.J. Ray Liu "for exemplary teaching and curriculum development, inspirational mentoring of graduate students, and broad educational impact in signal processing and communications"

IEEE Kiyo Tomiyasu Award: presented to Yonina Eldar "for development of the theory and implementation of sub-Nyquist sampling with applications to radar, communications, and ultrasound."

Congratulations to all of the recipients! The full list of 2016 IEEE Technical Field Awards recipients can be found in [1].

REFERENCE

[1] [Online]. Available: http://www.ieee.org/about/awards/2016_ieee_tfa_recipients_and_citations_list.pdf

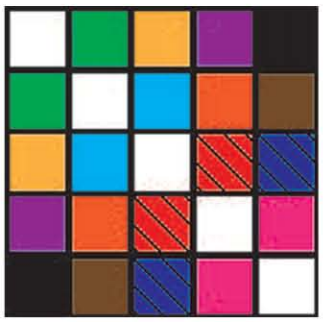
[SP]

Digital Object Identifier 10.1109/MSP.2015.2452851
Date of publication: 13 October 2015

Digital Object Identifier 10.1109/MSP.2015.2487698

Ivan Dokmanić, Reza Parhizkar, Juri Ranieri, and Martin Vetterli

Euclidean Distance Matrices



Essential theory, algorithms, and applications

Euclidean distance matrices (EDMs) are matrices of the squared distances between points. The definition is deceivingly simple; thanks to their many useful properties, they have found applications in psychometrics, crystallography, machine learning, wireless sensor networks, acoustics, and more. Despite the usefulness of EDMs, they seem to be insufficiently known in the signal processing community. Our goal is to rectify this mishap in a concise tutorial. We review the fundamental properties of EDMs, such as rank or (non)definiteness, and show how the various EDM properties can be used to design algorithms for completing and denoising distance data. Along the way, we demonstrate applications to microphone position calibration, ultrasound tomography, room reconstruction from echoes, and phase retrieval. By spelling out the essential algorithms, we hope to fast-track the readers in applying EDMs to their own problems. The code for all of the described algorithms and to generate the figures in the article is available online at http://lcav.epfl.ch/ivan.dokmanic. Finally, we suggest directions for further research.

INTRODUCTION

Imagine that you land at Geneva International Airport with the Swiss train schedule but no map. Perhaps surprisingly, this may be sufficient to reconstruct a rough (or not so rough) map of the Alpine country, even if the train times poorly translate to distances or if some of the times are unknown. The way to do it is by using EDMs; for an example, see "Swiss Trains (Swiss Map Reconstruction)."

We often work with distances because they are convenient to measure or estimate. In wireless sensor networks, for example, the sensor nodes measure the received signal strengths of the packets sent by other nodes or the time of arrival (TOA) of pulses emitted by their neighbors [1]. Both of these proxies allow for distance estimation between pairs of nodes; thus, we can attempt to reconstruct the network topology. This is often termed *self-localization* [2]–[4]. The molecular conformation problem is another instance of a distance problem [5], and so is reconstructing a room's geometry from echoes [6]. Less obviously, sparse phase retrieval [7] can be converted to a distance problem and addressed using EDMs.

Sometimes the data are not metric, but we seek a metric representation, as it happens commonly in psychometrics [8]. As a matter of fact, the psychometrics community is at the root of the development of a number of tools related to EDMs, including multidimensional scaling (MDS)—the problem of finding the best point set representation of a given set of distances. More abstractly, we can study EDMs for objects such as images, which live in high-dimensional vector spaces [9].

EDMs are a useful description of the point sets and a starting point for algorithm design. A typical task is to retrieve the original point configuration: it may initially come as a surprise that this requires no more than an eigenvalue decomposition (EVD) of a symmetric matrix. In fact, the majority of Euclidean distance problems require the reconstruction of the point set but always with one or more of the following twists:

1) The distances are noisy.
2) Some distances are missing.

Digital Object Identifier 10.1109/MSP.2015.2398954
Date of publication: 13 October 2015

1053-5888/15©2015IEEE

3) The distances are unlabeled.

For examples of applications requiring solutions of EDM problems with different complications, see Figure 1.

There are two fundamental problems associated with distance geometry [10]: 1) given a matrix, determine whether it is an EDM and 2) given a possibly incomplete set of distances, determine whether there exists a configuration of points in a given embedding dimension—the dimension of the smallest affine space comprising the points—that generates the distances.

LITERATURE REVIEW

The study of point sets through pairwise distances, and that of EDMs, can be traced back to the works of Menger [11], Schoenberg [12], Blumenthal [13], and Young and Householder [14]. An important class of EDM tools was initially developed for the purpose of data visualization. In 1952, Torgerson introduced the notion of MDS [8]. He used distances to quantify the dissimilarities between pairs of objects that are not necessarily vectors in a metric space. Later, in 1964, Kruskal suggested the notion of stress as a measure of goodness of fit for nonmetric data [15], again representing experimental dissimilarities between objects.

A number of analytical results on EDMs were developed by Gower [16], [17]. In 1985 [17], he gave a complete characterization of the EDM rank. Optimization with EDMs requires adequate geometric intuitions about matrix spaces. In 1990, Glunt et al. [18] and Hayden et al. [19] provided insights into the structure of the convex cone of EDMs. An extensive treatise on EDMs with many original results and an elegant characterization of the EDM cone is provided by Dattorro [20].

In the early 1980s, Williamson, Havel, and Wüthrich developed the idea of extracting the distances between pairs of hydrogen atoms in a protein using nuclear magnetic resonance (NMR). The extracted distances were then used to reconstruct three-dimensional (3-D) shapes of molecules [5]. (Wüthrich received the Nobel Prize for chemistry in 2002.) The NMR spectrometer (together with some postprocessing) outputs the distances between the pairs of atoms in a large molecule. The distances are not specified for all atom pairs, and they are uncertain—i.e., given only up to an interval. This setup lends itself naturally to EDM treatment; for example, it can be directly addressed using MDS [21]. Indeed, the crystallography community also contributed a large number of important results on distance geometry. In a different biochemical application, comparing distance matrices yields efficient algorithms for comparing proteins from their 3-D structure [22].

In machine learning, one can learn manifolds by finding an EDM with a low embedding dimension that preserves the local geometry. Weinberger and Saul use it to learn image manifolds [9]. Other examples of using Euclidean distance geometry in machine learning are the results by Tenenbaum, De Silva, and Langford [23] on image understanding and handwriting recognition; Jain and Saul [24] on speech and music; and Demaine et al. [25] on music and musical rhythms.

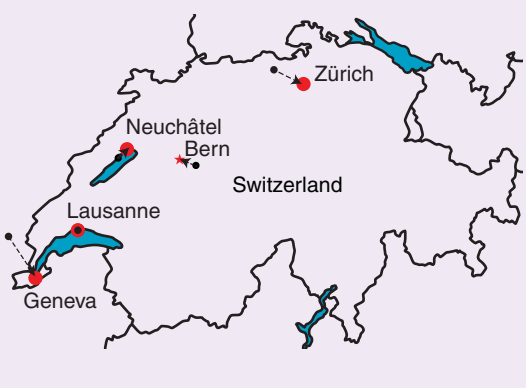
With the increased interest in sensor networks, several EDM-based approaches were proposed for sensor localization [2]–[4], [20]. The connections between EDMs, multilateration, and semidefinite programming are expounded in depth in [26], especially in the context of sensor network localization (SNL).

SWISS TRAINS (SWISS MAP RECONSTRUCTION)

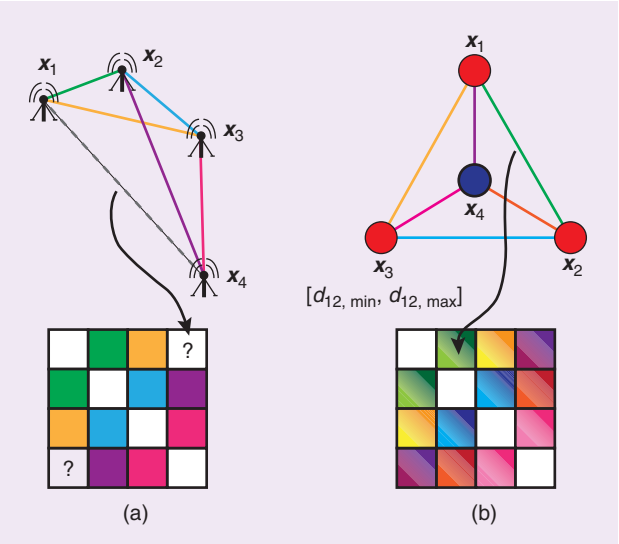
Consider the following matrix of the time in minutes it takes to travel by train between some Swiss cities (see Figure S1):

$$\begin{array}{c} \\ \text{Lausanne} \\ \text{Geneva} \\ \text{Zürich} \\ \text{Neuchâtel} \\ \text{Bern} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{L} & \text{G} & \text{Z} & \text{N} & \text{B} \end{array} \\ \begin{pmatrix} 0 & 33 & 128 & 40 & 66 \\ 33 & 0 & 158 & 64 & 101 \\ 128 & 158 & 0 & 88 & 56 \\ 40 & 64 & 88 & 0 & 34 \\ 66 & 101 & 56 & 34 & 0 \end{pmatrix} \end{array}.$$

The numbers were taken from the Swiss railways timetable. The matrix was then processed using the classical MDS algorithm (Algorithm 1), which is basically an EVD. The obtained city configuration was rotated and scaled to align with the actual map. Given all of the uncertainties involved, the fit is remarkably good. Not all trains drive with the same speed, they have varying numbers of stops, and railroads are not straight lines (i.e., because of lakes and mountains). This result may be regarded as anecdotal, but, in a fun way, it illustrates the power of the EDM toolbox. Classical MDS could be considered the simplest of the available tools, yet it yields usable results with erroneous data. On the other hand, it might be that Swiss trains are just *that* good.



[FIGS1] A map of Switzerland with the true locations of five cities (red) and their locations estimated by using classical MDS on the train schedule (black).



[FIG1] Two real-world applications of EDMs. (a) SNL from estimated pairwise distances is illustrated with one distance missing because the corresponding sensor nodes are too far apart to communicate. (b) In the molecular conformation problem, we aim to estimate the locations of the atoms in a molecule from their pairwise distances. Here, because of the inherent measurement uncertainty, we know the distances only up to an interval.

Position calibration in ad hoc microphone arrays is often done with sources at unknown locations, such as hand claps, finger snaps, or randomly placed loudspeakers [27]–[29]. This gives us the distances (possibly up to an offset time) between the microphones and the sources and leads to the problem of multidimensional unfolding (MDU) [30].

All of the mentioned applications work with labeled distance data. In certain TOA-based applications, one loses the labels, i.e., the correct permutation of the distances. This issue arises when reconstructing the geometry of a room from echoes [6]. Another example of unlabeled distances is in sparse phase retrieval, where the distances between the unknown nonzero lags in a signal are revealed in its autocorrelation function (ACF) [7]. Recently, motivated by problems in crystallography, Gujarahati et al. published an algorithm for the reconstruction of Euclidean networks from unlabeled distance data [31].

OUR MISSION

We were motivated to write this tutorial after realizing that EDMs are not common knowledge in the signal processing community, perhaps for the lack of a compact introductory text. This is effectively illustrated by the anecdote that, not long before writing this article, one of the authors of this article had to add the (rather fundamental) rank property to the Wikipedia page on EDMs (search for "Euclidean distance matrix"). (We are working on improving that page substantially.) In a compact tutorial, we do not attempt to be exhaustive; much more thorough literature reviews are available in longer exposés on EDMs and distance geometry [10], [32], [33]. Unlike these works, which take the most general approach through graph realizations, we opt to show simple cases through examples and explain and spell out a set of basic algorithms that anyone can use immediately. Two big topics that we discuss are not commonly treated in the EDM literature: localization from unlabeled distances and MDU (applied to microphone localization). On the other hand, we choose to not explicitly discuss the SNL problem as the relevant literature is abundant.

Implementations of all of the algorithms in this article are available online at http://lcav.epfl.ch/ivan.dokmanic. Our hope is that this will provide a solid starting point for those who wish to learn much more while inspiring new approaches to old problems.

FROM POINTS TO EDMs AND BACK

The principal EDM-related task is to reconstruct the original point set. This task is an inverse problem to the simpler forward problem of finding the EDM given the points. Thus, it is desirable to have an analytic expression for the EDM in terms of the point matrix. Beyond convenience, we can expect such an expression to provide interesting structural insights. We will define the notation as it becomes necessary—a summary is provided in Table 1.

Consider a collection of n points in a d-dimensional Euclidean space, ascribed to the columns of matrix $X \in \mathbb{R}^{d\times n}$, $X = [x_1, x_2, \cdots, x_n]$, $x_i \in \mathbb{R}^d$. Then the squared distance between x_i and x_j is given as

$$d_{ij} = \| x_i - x_j \|^2, \tag{1}$$

where $\| \cdot \|$ denotes the Euclidean norm. Expanding the norm yields

$$d_{ij} = (x_i - x_j)^\top (x_i - x_j) = x_i^\top x_i - 2x_i^\top x_j + x_j^\top x_j. \tag{2}$$

From here, we can read out the matrix equation for $D = [d_{ij}]$

$$\operatorname{edm}(X) \stackrel{\text{def}}{=} \mathbf{1}\operatorname{diag}(X^\top X)^\top - 2X^\top X + \operatorname{diag}(X^\top X)\mathbf{1}^\top, \tag{3}$$

where $\mathbf{1}$ denotes the column vector of all ones and $\operatorname{diag}(A)$ is the column vector of the diagonal entries of A. We see that $\operatorname{edm}(X)$ is in fact a function of $X^\top X$. For later reference, it is convenient to define an operator $\mathcal{K}(G)$ similar to $\operatorname{edm}(X)$, which operates directly on the Gram matrix $G = X^\top X$

$$\mathcal{K}(G) \stackrel{\text{def}}{=} \operatorname{diag}(G)\mathbf{1}^\top - 2G + \mathbf{1}\operatorname{diag}(G)^\top. \tag{4}$$

The EDM assembly formula (3) or (4) reveals an important property: because the rank of X is at most d (i.e., it has d rows), then the rank of $X^\top X$ is also at most d. The remaining two summands in (3) have rank one. By rank inequalities, the rank of a sum of matrices cannot exceed the sum of the ranks of the summands. With this observation, we proved one of the most notable facts about EDMs:

Theorem 1 (Rank of EDMs): *The rank of an EDM corresponding to points in* $\mathbb{R}^d$ *is at most* $d+2$.

This is a powerful theorem; it states that the rank of an EDM is independent of the number of points that generate it. In many

applications, d is three or less while n can be in the thousands. According to Theorem 1, the rank of such practical matrices is at most five. The proof of this theorem is simple, but, to appreciate that the property is not obvious, you may try to compute the rank of the matrix of nonsquared distances.

What really matters in Theorem 1 is the affine dimension of the point set, i.e., the dimension of the smallest affine subspace that contains the points, which is denoted by $\text{aff dim}(X)$. For example, if the points lie on a plane (but not on a line or a circle) in $\mathbb{R}^3$, the rank of the corresponding EDM is four, not five. This will be made clear from a different perspective in the section "Essential Uniqueness," as any affine subspace is just a translation of a linear subspace. An illustration for a one-dimensional (1-D) subspace of $\mathbb{R}^2$ is provided in Figure 2. Subtracting any point in the affine subspace from all of its points translates it to the parallel linear subspace that contains the zero vector.

ESSENTIAL UNIQUENESS

When solving an inverse problem, we need to understand what is recoverable and what is forever lost in the forward problem. Representing sets of points by distances usually increases the size of the representation. For most interesting n and d, the number of pairwise distances is larger than the size of the coordinate description, $(1/2)n(n-1) > nd$, so an EDM holds more scalars than the list of point coordinates. Nevertheless, some information is lost in this encoding such as the information about the absolute position and orientation of the point set. Intuitively, it is clear that rigid transformations (including reflections) do not change the distances between the fixed points in a point set. This intuitive fact is easily deduced from the EDM assembly formula (3). We have seen in (3) and (4) that $\text{edm}(X)$ is in fact a function of the Gram matrix $X^\top X$.

This makes it easy to show algebraically that rotations and reflections do not alter the distances. Any rotation/reflection can be represented by an orthogonal matrix $Q \in \mathbb{R}^{d\times d}$ acting on the points x_i. Thus, for the rotated point set $X_r = QX$, we can write

$$X_r^\top X_r = (QX)^\top (QX) = X^\top Q^\top QX = X^\top X, \tag{5}$$

where we invoked the orthogonality of the rotation/reflection matrix $Q^\top Q = I$.

Translation by a vector $b \in \mathbb{R}^d$ can be expressed as

$$X_t = X + b1^\top. \tag{6}$$

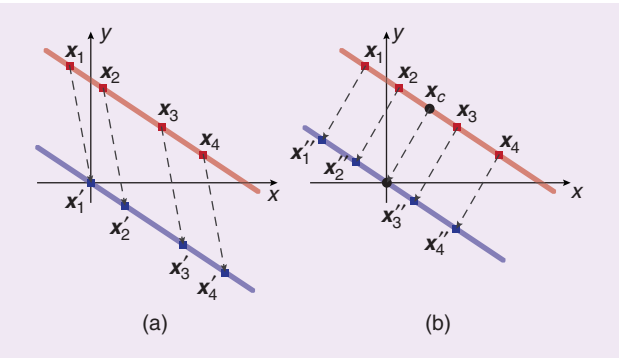
Using $\text{diag}(X_t^\top X_t) = \text{diag}(X^\top X) + 2X^\top b + \|b\|^2 1$, one can directly verify that this transformation leaves (3) intact. In summary,

$$\text{edm}(QX) = \text{edm}(X + b1^\top) = \text{edm}(X). \tag{7}$$

The consequence of this invariance is that we will never be able to reconstruct the absolute orientation of the point set using only the distances, and the corresponding degrees of freedom will be chosen freely. Different reconstruction procedures will lead to different realizations of the point set, all of them being rigid transformations of each other. Figure 3 illustrates a point set under a rigid transformation; it is clear that the distances between the points are the same for all three shapes.

[TABLE 1] A SUMMARY OF THE NOTATIONS.

SYMBOL	MEANING
n	NUMBER OF POINTS (COLUMNS) IN $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$
d	DIMENSIONALITY OF THE EUCLIDEAN SPACE
a_{ij}	ELEMENT OF A MATRIX $\boldsymbol{A}$ ON THE iTH ROW AND THE jTH COLUMN
$\boldsymbol{D}$	AN EDM
$\text{edm}(\boldsymbol{X})$	AN EDM CREATED FROM THE COLUMNS IN $\boldsymbol{X}$
$\text{edm}(\boldsymbol{X}, \boldsymbol{Y})$	A MATRIX CONTAINING THE SQUARED DISTANCES BETWEEN THE COLUMNS OF $\boldsymbol{X}$ AND $\boldsymbol{Y}$
$\mathcal{K}(\boldsymbol{G})$	AN EDM CREATED FROM THE GRAM MATRIX $\boldsymbol{G}$
$\boldsymbol{J}$	A GEOMETRIC CENTERING MATRIX
$\boldsymbol{A}_{\boldsymbol{W}}$	RESTRICTION OF $\boldsymbol{A}$ TO NONZERO ENTRIES IN $\boldsymbol{W}$
$\boldsymbol{W}$	MASK MATRIX, WITH ONES FOR OBSERVED ENTRIES
$\mathbb{S}_+^n$	A SET OF REAL SYMMETRIC POSITIVE-SEMIDEFINITE (PSD) MATRICES IN $\mathbb{R}^{n\times n}$
$\text{aff dim}(\boldsymbol{X})$	AFFINE DIMENSION OF THE POINTS LISTED IN $\boldsymbol{X}$
$\boldsymbol{A} \circ \boldsymbol{B}$	HADAMARD (ENTRYWISE) PRODUCT OF $\boldsymbol{A}$ AND $\boldsymbol{B}$
ε_{ij}	NOISE CORRUPTING THE (i, j) DISTANCE
$\boldsymbol{e}_i$	iTH VECTOR OF THE CANONICAL BASIS
$\|\boldsymbol{A}\|_F$	FROBENIUS NORM OF $\boldsymbol{A}$, $\left(\sum_{ij} a_{ij}^2\right)^{1/2}$

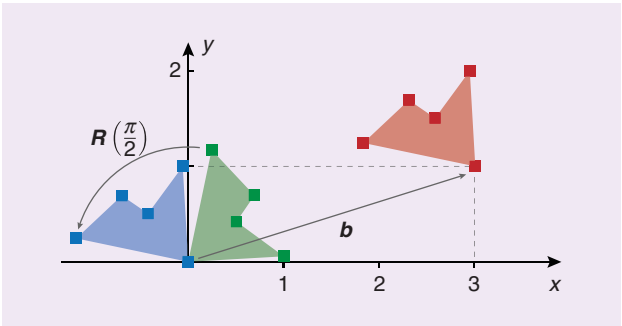


[FIG2] An illustration of the relationship between an affine subspace and its parallel linear subspace. The points $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4]$ live in an affine subspace—a line in $\mathbb{R}^2$ that does not contain the origin. In (a), the vector $\boldsymbol{x}_1$ is subtracted from all the points, and the new point list is $\boldsymbol{X}' = [\boldsymbol{0}, \boldsymbol{x}_2 - \boldsymbol{x}_1, \boldsymbol{x}_3 - \boldsymbol{x}_1, \boldsymbol{x}_4 - \boldsymbol{x}_1]$. While the columns of $\boldsymbol{X}$ span $\mathbb{R}^2$, the columns of $\boldsymbol{X}'$ only span a 1-D subspace of $\mathbb{R}^2$—the line through the origin. In (b), we subtract a different vector from all points: the centroid $(1/4)\boldsymbol{X}\boldsymbol{1}$. The translated vectors $\boldsymbol{X}'' = [\boldsymbol{x}_1'', \ldots, \boldsymbol{x}_4'']$ again span the same 1-D subspace.

RECONSTRUCTING THE POINT SET FROM DISTANCES

The EDM equation (3) hints at a procedure to compute the point set starting from the distance matrix. Consider the following choice: let the first point x_1 be at the origin. Then, the first column of D contains the squared norms of the point vectors

$$d_{i1} = \|x_i - x_1\|^2 = \|x_i - 0\|^2 = \|x_i\|^2. \tag{8}$$



[FIG3] An illustration of a rigid transformation in 2-D. Here, the point set is transformed as $RX + b\mathbf{1}^\top$. The rotation matrix $R = [0\ 1; -1\ 0]$ (MATLAB notation) corresponds to a counterclockwise rotation of 90°. The translation vector is $b = [3, 1]^\top$. The shape is drawn for visual reference.

Consequently, we can construct the term $\mathbf{1}\,\mathrm{diag}(X^\top X)$ and its transpose in (3), as the diagonal of $X^\top X$ contains exactly the norms squared $\|x_i\|^2$. Concretely,

$$\mathbf{1}\,\mathrm{diag}(X^\top X) = \mathbf{1}\,d_1^\top, \tag{9}$$

where $d_1 = De_1$ is the first column of D. We thus obtain the Gram matrix from (3) as

$$G = X^\top X = -\frac{1}{2}(D - \mathbf{1}\,d_1^\top - d_1\mathbf{1}^\top). \tag{10}$$

The point set can then be found by an EVD, $G = U\Lambda U^\top$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ with all eigenvalues λ_i nonnegative and U orthonormal, as G is a symmetric positive-semidefinite (PSD) matrix. Throughout this article, we assume that the eigenvalues are sorted in the order of decreasing magnitude, $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. We can now set $\widehat{X} \overset{\text{def}}{=} [\mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_d}), 0_{d\times(n-d)}]U^\top$. Note that we could have simply taken $\Lambda^{1/2}U^\top$ as the reconstructed point set, but if the Gram matrix really describes a d-dimensional point set, the trailing eigenvalues will be zeroes, so we choose to truncate the corresponding rows.

It is straightforward to verify that the reconstructed point set $\widehat{X}$ generates the original EDM, $D = \mathrm{edm}(X)$; as we have learned, $\widehat{X}$ and X are related by a rigid transformation. The described procedure is called the *classical MDS*, with a particular choice of the coordinate system: x_1 is fixed at the origin.

In (10), we subtract a structured rank-2 matrix $(\mathbf{1}\,d_1^\top + d_1\mathbf{1}^\top)$ from D. A more systematic approach to the classical MDS is to use a generalization of (10) by Gower [16]. Any such subtraction that makes the right-hand side of (10) PSD, i.e., that makes G a Gram matrix, can also be modeled by multiplying D from both sides by a particular matrix. This is substantiated in the following result.

Theorem 2 (Gower [16]): *D is an EDM if and only if*

$$-\frac{1}{2}(I - \mathbf{1}\,s^\top)D(I - s\,\mathbf{1}^\top) \tag{11}$$

is PSD for any s such that $s^\top\mathbf{1} = 1$ and $s^\top D \neq 0$.

Algorithm 1: The classical MDS.

```
function ClassicalMDS (D, d)
    J ← I − (1/n)11^⊤        ▷ Geometric centering matrix
    G ← −(1/2) JDJ           ▷ Compute the Gram matrix
    U, [λ_i]_{i=1}^n ← EVD(G)
    return [diag(√λ_1, ..., √λ_d), 0_{d×(n−d)}] U^⊤
end function
```

In fact, if (11) is PSD for one such s, then it is PSD for all of them. In particular, define the geometric centering matrix as

$$J \overset{\text{def}}{=} I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top. \tag{12}$$

Then, $-(1/2)\,JDJ$ being PSD is equivalent to D being an EDM. Different choices of s correspond to different translations of the point set.

The classical MDS algorithm with the geometric centering matrix is spelled out in Algorithm 1. Whereas so far we have assumed that the distance measurements are noiseless, Algorithm 1 can handle noisy distances too as it discards all but the d largest eigenvalues.

It is straightforward to verify that (10) corresponds to $s = e_1$. Think about what this means in terms of the point set: Xe_1 selects the first point in the list, x_1. Then, $X_0 = X(I - e_1\mathbf{1}^\top)$ translates the points so that x_1 is translated to the origin. Multiplying the definition (3) from the right by $(I - e_1\mathbf{1}^\top)$ and from the left by $(I - \mathbf{1}\,e_1^\top)$ will annihilate the two rank-1 matrices, $\mathrm{diag}(G)\mathbf{1}^\top$ and $\mathbf{1}\,\mathrm{diag}(G)^\top$. We see that the remaining term has the form $-2X_0^\top X_0$, and the reconstructed point set will have the first point at the origin.

On the other hand, setting $s = (1/n)\mathbf{1}$ places the centroid of the point set at the origin of the coordinate system. For this reason, the matrix $J = I - (1/n)\mathbf{1}\mathbf{1}^\top$ is called the *geometric centering matrix*. To better understand why, consider how we normally center a set of points given in X: first, we compute the centroid as the mean of all the points,

$$x_c = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}X\,\mathbf{1}. \tag{13}$$

Second, we subtract this vector from all the points in the set

$$X_c = X - x_c\,\mathbf{1}^\top = X - \frac{1}{n}X\,\mathbf{1}\mathbf{1}^\top = X(I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top). \tag{14}$$

In complete analogy with the reasoning for $s = e_1$, we can see that the reconstructed point set will be centered at the origin.

ORTHOGONAL PROCRUSTES PROBLEM

Since the absolute position and orientation of the points are lost when going over to distances, we need a method to align the reconstructed point set with a set of anchors, i.e., points whose coordinates are fixed and known.

This can be achieved in two steps, sometimes called *Procrustes analysis*. Ascribe the anchors to the columns of Y, and suppose that we want to align the point set X with the columns of Y. Let X_a denote the submatrix (a selection of columns) of X that should be aligned with the anchors. We note that the number of anchors (the columns in X_a) is typically small compared with the total number of points (the columns in X).

In the first step, we remove the means $\boldsymbol{y}_c$ and $\boldsymbol{x}_{a,c}$ from matrices Y and X_a, obtaining the matrices $\overline{Y}$, and $\overline{X}_a$. In the second step, termed *orthogonal Procrustes analysis*, we are searching for the rotation and reflection that best maps $\overline{X}_a$ onto $\overline{Y}$

$$R = \underset{Q:QQ^\top = I}{\arg\min} \| Q\overline{X}_a - \overline{Y} \|_F^2. \tag{15}$$

The Frobenius norm $\| \cdot \|_F$ is simply the ℓ^2-norm of the matrix entries, $\| A \|_F^2 \stackrel{\text{def}}{=} \sum a_{ij}^2 = \text{trace}\,(A^\top A)$.

The solution to (15), found by Schönemann in his Ph.D. thesis [34], is given by the singular value decomposition (SVD). Let $\overline{X}_a\,\overline{Y}^\top = U\Sigma V^\top$; then, we can continue computing (15) as follows:

$$\begin{aligned} R &= \underset{Q:QQ^\top = I}{\arg\min} \| Q\overline{X}_a \|_F^2 + \| \overline{Y} \|_F^2 - \text{trace}\,(Y^\top Q \overline{X}_a) \\ &= \underset{\widetilde{Q}:\widetilde{Q}\widetilde{Q}^\top = I}{\arg\min} \text{trace}\,(\widetilde{Q}\,\Sigma), \end{aligned} \tag{16}$$

where $\widetilde{Q} \stackrel{\text{def}}{=} V^\top Q U$ and we used the orthogonal invariance of the Frobenius norm and the cyclic invariance of the trace. The last trace expression in (16) is equal to $\sum_{i=1}^n \sigma_i \widetilde{q}_{ii}$. Noting that $\widetilde{Q}$ is also an orthogonal matrix, its diagonal entries cannot exceed 1. Therefore, the maximum is achieved when $\widetilde{q}_{ii} = 1$ for all i, meaning that the optimal $\widetilde{Q}$ is an identity matrix. It follows that $R = VU^\top$.

Once the optimal rigid transformation has been found, the alignment can be applied to the entire point set as

$$R(X - \boldsymbol{x}_{a,c}\mathbf{1}^\top) + \boldsymbol{y}_c\mathbf{1}^\top. \tag{17}$$

COUNTING THE DEGREES OF FREEDOM

It is interesting to count how many degrees of freedom there are in different EDM-related objects. Clearly, for n points in $\mathbb{R}^d$, we have

$$\#_X = n \times d \tag{18}$$

degrees of freedom: if we describe the point set by the list of coordinates, the size of the description matches the number of degrees of freedom. Going from the points to the EDM (usually) increases the description size to $(1/2)n(n-1)$, as the EDM lists the distances between all the pairs of points. By Theorem 1, we know that the EDM has rank at most $d+2$.

Let us imagine for a moment that we do not know any other EDM-specific properties of our matrix except that it is symmetric, positive, zero-diagonal (or hollow), and that it has rank $d+2$. The purpose of this exercise is to count the degrees of freedom associated with such a matrix and to see if their number matches the intrinsic number of the degrees of freedom of the point set, $\#_X$. If it did, then these properties would completely characterize an EDM. We can already anticipate from Theorem 2 that we need more properties: a certain matrix related to the EDM—as given in (11)—must be PSD. Still, we want to see how many degrees of freedom we miss.

We can do the counting by looking at the EVD of a symmetric matrix, $D = U\Lambda U^\top$. The diagonal matrix Λ is specified by $d+2$ degrees of freedom because D has rank $d+2$. The first eigenvector of length n takes up $n-1$ degrees of freedom due to the normalization; the second one takes up $n-2$, as it is in addition orthogonal to the first one; for the last eigenvector, number $(d+2)$, we need $n-(d+2)$ degrees of freedom. We do not need to count the other eigenvectors because they correspond to zero eigenvalues. The total number is then

$$\begin{aligned} \#_{\text{DOF}} &= \underbrace{(d+2)}_{\text{Eigenvalues}} + \underbrace{(n-1) + \cdots + [n-(d+2)]}_{\text{Eigenvectors}} - \underbrace{n}_{\text{Hollowness}} \\ &= n \times (d+1) - \frac{(d+1) \times (d+2)}{2}. \end{aligned}$$

For large n and fixed d, it follows that

$$\frac{\#_{\text{DOF}}}{\#_X} \sim \frac{d+1}{d}. \tag{19}$$

Therefore, even though the rank property is useful and we will show efficient algorithms that exploit it, it is still not a *tight* property (with symmetry and hollowness included). For $d = 3$, the ratio (19) is $(4/3)$, so loosely speaking, the rank property has 30% too many determining scalars, which we need to set consistently. In other words, we need 30% more data to exploit the rank property than we need to exploit the full EDM structure. We can assert that, for the same amount of data, the algorithms perform at least $\approx$30% worse if we only exploit the rank property without EDMness.

MISSING ENTRIES ARISE BECAUSE OF THE LIMITED RADIO RANGE OR BECAUSE OF THE NATURE OF THE SPECTROMETER.

The one-third gap accounts for various geometrical constraints that must be satisfied. The redundancy in the EDM representation is what makes denoising and completion algorithms possible, and thinking in terms of degrees of freedom gives us a fundamental understanding of what is achievable. Interestingly, the previous discussion suggests that for large n and large $d = o(n)$, little is lost by only considering rank.

Finally, in the previous discussion, for the sake of simplicity we ignored the degrees of freedom related to absolute orientation. These degrees of freedom, which are not present in the EDM, do not affect the large n behavior.

SUMMARY

Let us summarize what we have achieved in this section:

- We explained how to algebraically construct an EDM given the list of point coordinates.
- We discussed the essential uniqueness of the point set; information about the absolute orientation of the points is irretrievably lost when transitioning from points to an EDM.

- We explained classical MDS—a simple EVD-based algorithm (Algorithm 1) for reconstructing the original points—along with discussing parameter choices that lead to different centroids in reconstruction.
- Degrees of freedom provide insight into scaling behavior. We showed that the rank property is satisfactory, but there is more to it than just rank.

EDMs AS A PRACTICAL TOOL

We rarely have a perfect EDM. Not only are the entries of the measured matrix plagued by errors, but often we can measure just a subset. There are various sources of error in distance measurements: we already know that in NMR spectroscopy, we get intervals instead of exact distances. Measuring the distance using received powers or TOAs is subject to noise, sampling errors, and model mismatch.

Missing entries arise because of the limited radio range or because of the nature of the spectrometer. Sometimes the nodes in the problem at hand are asymmetric by definition; in microphone calibration, we have two types: microphones and calibration sources. This results in a particular block structure of the missing entries (see Figure 4 for an illustration).

It is convenient to have a single statement for both EDM approximation and EDM completion as the algorithms described in this section handle them at once.

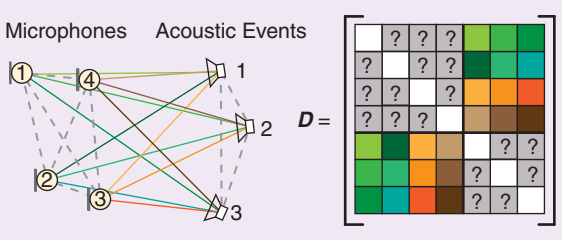
Problem 1: *Let* $D = \operatorname{edm}(X)$. *We are given a noisy observation of the distances between* $p \leq (1/2)n(n-1)$ *pairs of points from* X. *That is, we have a noisy measurement of* $2p$ *entries in* D

$$\widetilde{d}_{ij} = d_{ij} + \varepsilon_{ij}, \tag{20}$$

for $(i, j) \in E$, *where* E *is some index set and* ε_{ij} *absorbs all errors. The goal is to reconstruct the point set* $\widehat{X}$ *in the given embedding dimension, so that the entries of* $\operatorname{edm}(\widehat{X})$ *are close in some metric to the observed entries* $\widetilde{d}_{ij}$.

To concisely write down completion problems, we define the mask matrix W as follows:

$$w_{ij} \stackrel{\text{def}}{=} \begin{cases} 1, (i, j) \in E \\ 0, \text{otherwise.} \end{cases} \tag{21}$$



[FIG4] **The microphone calibration as an example of MDU. We can measure only the propagation times from acoustic sources at unknown locations to microphones at unknown locations. The corresponding revealed part of the EDM has a particular off-diagonal structure, leading to a special case of EDM completion.**

This matrix then selects elements of an EDM through a Hadamard (entrywise) product. For example, to compute the norm of the difference between the observed entries in A and B, we write $\|W \circ (A - B)\|$. Furthermore, we define the indexing A_W to mean the restriction of A to those entries where W is nonzero. The meaning of $B_W \leftarrow A_W$ is that we assign the observed part of A to the observed part of B.

EXPLOITING THE RANK PROPERTY

Perhaps the most notable fact about EDMs is the rank property established in Theorem 1: the rank of an EDM for points living in $\mathbb{R}^d$ is at most $d + 2$. This leads to conceptually simple algorithms for EDM completion and denoising. Interestingly, these algorithms exploit only the rank of the EDM. There is no explicit Euclidean geometry involved, at least not before reconstructing the point set.

We have two pieces of information: a subset of potentially noisy distances and the desired embedding dimension of the point configuration. The latter implies the rank property of the EDM that we aim to exploit. We may try to alternate between enforcing these two properties and hope that the algorithm produces a sequence of matrices that converges to an EDM. If it does, we have a solution. Alternatively, it may happen that we converge to a matrix with the correct rank that is not an EDM or that the algorithm never converges. The pseudocode is listed in Algorithm 2.

Algorithm 2: The alternating rank-based EDM completion.

```
function RankCompleteEDM (W, D̃, d)
    D_W ← D̃_W                          ▷ Initialize observed entries
    D_{11^T−W} ← μ                     ▷ Initialize unobserved entries
    repeat
        D ← EVThreshold(D, d + 2)
        D_W ← D̃_W                      ▷ Enforce known entries
        D_I ← 0                        ▷ Set the diagonal to zero
        D ← (D)_+                      ▷ Zero the negative entries
    until Convergence or MaxIter
    return D
end function

function EVThreshold (D, r)
    U, [λ_i]_{i=1}^n ← EVD(D)
    Σ ← diag(λ_1, ..., λ_r, 0, ..., 0)   (n − r times)
    D ← UΣU^T
    return D
end function
```

A different, more powerful approach is to leverage algorithms for low-rank matrix completion developed by the compressed sensing community. For example, OptSpace [35] is an algorithm for recovering a low-rank matrix from noisy, incomplete data. Let us take a look at how OptSpace works. Denote by $M \in \mathbb{R}^{m \times n}$ the rank-r matrix that we seek to recover, by $Z \in \mathbb{R}^{m \times n}$ the measurement noise, and by $W \in \mathbb{R}^{m \times n}$ the mask corresponding to the

measured entries; for simplicity, we chose $m \le n$. The measured noisy and incomplete matrix is then given as

$$\widetilde{M} = W \circ (M + Z). \tag{22}$$

Effectively, this sets the missing (nonobserved) entries of the matrix to zero. OptSpace aims to minimize the following cost function:

$$F(A, S, B) \stackrel{\text{def}}{=} \frac{1}{2} \| W \circ (\widetilde{M} - ASB^{\top}) \|_F^2, \tag{23}$$

where $S \in \mathbb{R}^{r \times r}$, $A \in \mathbb{R}^{m \times r}$ and $B \in \mathbb{R}^{n \times r}$, such that $A^{\top}A = B^{\top}B = I$. Note that S need not be diagonal.

The cost function (23) is not convex, and minimizing it is a priori difficult [36] because of many local minima. Nevertheless, Keshavan, Montanari, and Oh [35] show that using the gradient descent method to solve (23) yields the global optimum with high probability, provided that the descent is correctly initialized.

Let $\widetilde{M} = \sum_{i=1}^{m} \sigma_i \boldsymbol{a}_i \boldsymbol{b}_i^{\top}$ be the SVD of $\widetilde{M}$. Then, we define the scaled rank-r projection of $\widetilde{M}$ as $\widetilde{M}_r \stackrel{\text{def}}{=} \alpha^{-1} \sum_{i=1}^{r} \sigma_i \boldsymbol{a}_i \boldsymbol{b}_i^{\top}$. The fraction of observed entries is denoted by α so that the scaling factor compensates the smaller average magnitude of the entries in $\widetilde{M}$ in comparison with M. The SVD of $\widetilde{M}_r$ is then used to initialize the gradient descent, as detailed in Algorithm 3.

Algorithm 3: OptSpace [35].

1: **function** `OptSpace` $(\widetilde{M}, r)$
2: $\quad \widetilde{M} \leftarrow \text{Trim}(\widetilde{M})$
3: $\quad \widetilde{A}, \widetilde{\Sigma}, \widetilde{B} \leftarrow \text{SVD}(\alpha^{-1}\widetilde{M})$
4: $\quad A_0 \leftarrow$ First r columns of $\widetilde{A}$
5: $\quad B_0 \leftarrow$ First r columns of $\widetilde{B}$
6: $\quad S_0 \leftarrow \underset{S \in \mathbb{R}^{r \times r}}{\text{argmin}} F(A_0, S, B_0)$ ▷ Eq. (23)
7: $\quad A, B \leftarrow \underset{A^{\top}A = B^{\top}B = I}{\text{argmin}} F(A, S_0, B)$ ▷ See the note below
8: $\quad$ **return** $AS_0B^{\top}$
9: **end function**
▷ Line 7: gradient descent starting at A_0, B_0

Two additional remarks are due in the description of OptSpace. First, it can be shown that the performance is improved by zeroing the overrepresented rows and columns. A row (respectively, column) is overrepresented if it contains more than twice the average number of observed entries per row (respectively, column). These heavy rows and columns bias the corresponding singular vectors and values, so (perhaps surprisingly) it is better to throw them away. We call this step "Trim" in Algorithm 3.

Second, the minimization of (23) does not have to be performed for all variables at once. In [35], the authors first solve the easier, convex minimization for S, and then with the optimizer S fixed, they find the matrices A and B using the gradient descent. These steps correspond to lines 6 and 7 of Algorithm 3. For an application of OptSpace in the calibration of ultrasound measurement rigs, see "Calibration in Ultrasound Tomography."

CALIBRATION IN ULTRASOUND TOMOGRAPHY

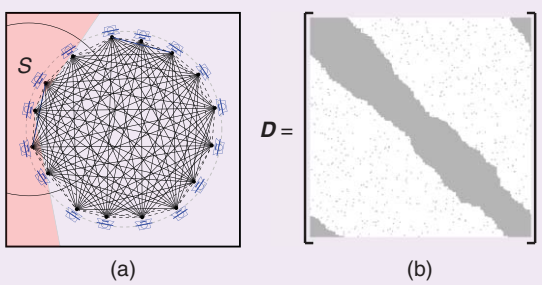
The rank property of EDMs, introduced in Theorem 1, can be leveraged in the calibration of ultrasound tomography devices. An example device for diagnosing breast cancer is a circular ring with thousands of ultrasound transducers placed around the breast [37]. The setup is shown in Figure S2(a).

Because of manufacturing errors, the sensors are not located on a perfect circle. This uncertainty in the positions of the sensors negatively affects the algorithms for imaging the breast. Fortunately, we can use the measured distances between the sensors to calibrate their relative positions. We can estimate the distances by measuring the times of flight (TOF) between pairs of transducers in a homogeneous environment, e.g., in water.

We cannot estimate the distances between all pairs of sensors because the sensors have limited beamwidths. (It is hard to manufacture omnidirectional ultrasonic sensors.) Therefore, the distances between the neighboring sensors are unknown, contrary to typical SNL scenarios where only the distances between nearby nodes can be measured. Moreover, the distances are noisy and some of them are unreliably estimated. This yields a noisy and incomplete EDM whose structure is illustrated in Figure S2(b).

Assuming that the sensors lie in the same plane, the original EDM produced by them would have a rank less than five. We can use the rank property and a low-rank matrix completion method, such as OptSpace (Algorithm 3), to complete and denoise the measured matrix [38]. Then, we can use the classical MDS in Algorithm 1 to estimate the relative locations of the ultrasound sensors.

For the reasons mentioned previously, SNL-specific algorithms are suboptimal when applied to ultrasound calibration. An algorithm based on the rank property effectively solves the problem and enables one to derive upper bounds on the performance error calibration mechanism, with respect to the number of sensors and the measurement noise. The authors in [38] show that the error vanishes as the number of sensors increases.



[FIGS2] (a) Ultrasound transducers lie on an approximately circular ring. The ring surrounds the breast and after each transducer fires an ultrasonic signal, the sound speed distribution of the breast is estimated. A precise knowledge of the sensor locations is needed to have an accurate reconstruction of the enclosed medium. (b) Because of the limited beamwidth of the transducers, noise, and imperfect TOF estimation methods, the measured EDM is incomplete and noisy. The gray areas show the missing entries of the matrix.

MULTIDIMENSIONAL SCALING

MDS refers to a group of techniques that, given a set of noisy distances, find the best fitting point conformation. It was originally proposed in psychometrics [8], [15] to visualize the (dis) similarities between objects. Initially, MDS was defined as the problem of representing distance data, but now the term is commonly used to refer to methods for solving the problem [39].

Various cost functions were proposed for solving MDS. In the section "Reconstructing the Point Set from Distances," we already encountered one method: the classical MDS. This method minimizes the Frobenius norm of the difference between the input Gram matrix and the Gram matrix of the points in the target embedding dimension.

THE S-STRESS OBJECTIVE IS EVERYWHERE DIFFERENTIABLE, BUT A DISADVANTAGE IS THAT IT FAVORS LARGE OVER SMALL DISTANCES.

The Gram matrix contains inner products, but it is better to work directly with the distances. A typical cost function represents the dissimilarity of the observed distances and the distances between the estimated point locations. An essential observation is that the feasible set for these optimizations is not convex (i.e., EDMs with embedding dimensions smaller than $n-1$ lie on the boundary of a cone [20], which is a nonconvex set).

A popular dissimilarity measure is *raw stress* [40], defined as the value of

$$\underset{X\in\mathbb{R}^{d\times n}}{\text{minimize}} \sum_{(i,j)\in E} \left(\sqrt{\text{edm}(X)_{ij}} - \sqrt{\widetilde{d}_{ij}}\right)^2, \tag{24}$$

where E defines the set of revealed elements of the distance matrix D. The objective function can be concisely written as $\|W\circ(\sqrt{\text{edm}(X)}-\sqrt{\widetilde{D}})\|_F^2$; a drawback of this cost function is that it is not globally differentiable. The approaches described in the literature comprise iterative majorization [41], various methods using convex analysis [42], and steepest descent methods [43].

Another well-known cost function, first studied by Takane, Young, and De Leeuw [44], is termed *s-stress*,

$$\underset{X\in\mathbb{R}^{d\times n}}{\text{minimize}} \sum_{(i,j)\in E} \left(\text{edm}(X)_{ij} - \widetilde{d}_{ij}\right)^2. \tag{25}$$

Again, we write the objective concisely as $\|W\circ(\text{edm}(X)-\widetilde{D})\|_F^2$. Conveniently, the s-stress objective is globally differentiable, but a disadvantage is that it puts more weight on errors in larger distances than on errors in smaller ones. Gaffke and Mathar [45] propose an algorithm to find the global minimum of the s-stress function for embedding dimension $d=n-1$. EDMs with this embedding dimension exceptionally constitute a convex set [20], but we are typically interested in embedding dimensions much smaller than n. The s-stress minimization in (25) is not convex for $d<n-1$. It was analytically shown to have saddle points [46], but interestingly, no analytical nonglobal minimizer has been found [46].

Browne proposed a method for computing s-stress based on Newton–Raphson root finding [47]. Glunt reports that the method by Browne converges to the global minimum of (25) in 90% of the test cases in his data set [48]. (While the experimental setup of Glunt [48] is not detailed, it was mentioned that the EDMs were produced randomly.)

The cost function in (25) is separable across points i and across coordinates k, which is convenient for distributed implementations. Parhizkar [46] proposed an alternating coordinate descent method that leverages this separability by updating a single coordinate of a particular point at a time. The s-stress function restricted to the kth coordinate of the ith point is a fourth-order polynomial

$$f(x;\boldsymbol{\alpha}^{(i,k)}) = \sum_{\ell=0}^{4} \alpha_\ell^{(i,k)} x^\ell, \tag{26}$$

where $\boldsymbol{\alpha}^{(i,k)}$ lists the polynomial coefficients for the ith point and the kth coordinate. For example, $\alpha_0^{(i,k)} = 4\sum_j w_{ij}$, that is, four times the number of points connected to point i. Expressions for the remaining coefficients are given in [46]; in the pseudocode (Algorithm 4), we assume that these coefficients are returned by the function "GetQuadricCoeffs," given the noisy incomplete matrix $\widetilde{D}$, the observation mask W, and the dimensionality d. The global minimizer of (26) can be found analytically by calculating the roots of its derivative (a cubic). The process is then repeated over all coordinates k and points i until convergence. The resulting algorithm is remarkably simple yet empirically converges fast. It naturally lends itself to a distributed implementation. We spell it out in Algorithm 4.

Algorithm 4: Alternating descent [46].

```
function AlternatingDescent(D̃, W, d)
  X ∈ R^{d×n} ← X_0 = 0                ▷ Initialize the point set
  repeat
    for i ∈ {1, ..., n} do             ▷ Points
      for k ∈ {1, ..., d} do           ▷ Coordinates
        α^{(i,k)} ← GetQuadricCoeffs(W, D̃, d)
        x_{i,k} ← argmin_x f(x; α^{(i,k)})   ▷ Eq. (26)
      end for
    end for
  until Convergence or MaxIter
  return X
end function
```

When applied to a large data set of random, noiseless, and complete distance matrices, Algorithm 4 converges to the global minimum of (25) in more than 99% of the cases [46].

SEMIDEFINITE PROGRAMMING

Recall the characterization of EDMs (11) in Theorem 2. It states that D is an EDM if and only if the corresponding geometrically centered Gram matrix $-(1/2)JDJ$ is PSD. Thus, it establishes a

one-to-one correspondence between the cone of EDMs, denoted by $\mathbb{EDM}^n$ and the intersection of the symmetric positive-semidefinite cone $\mathbb{S}_+^n$ with the geometrically centered cone $\mathbb{S}_c^n$. The latter is defined as the set of all symmetric matrices whose column sum vanishes,

$$\mathbb{S}_c^n = \{\boldsymbol{G} \in \mathbb{R}^{n\times n} \,|\, \boldsymbol{G} = \boldsymbol{G}^\top, \boldsymbol{G}\,\mathbf{1} = \mathbf{0}\}. \tag{27}$$

We can use this correspondence to cast EDM completion and approximation as semidefinite programs. While (11) describes an EDM of an n-point configuration in any dimension, we are often interested in situations where $d \ll n$. It is easy to adjust for this case by requiring that the rank of the centered Gram matrix be bounded. One can verify that

$$\left.\begin{array}{l}\boldsymbol{D} = \operatorname{edm}(\boldsymbol{X}) \\ \operatorname{affdim}(\boldsymbol{X}) \le d\end{array}\right\} \iff \begin{cases} -\dfrac{1}{2}\boldsymbol{J}\boldsymbol{D}\boldsymbol{J} \succeq 0 \\ \operatorname{rank}(\boldsymbol{J}\boldsymbol{D}\boldsymbol{J}) \le d, \end{cases} \tag{28}$$

when $n \ge d$. That is, EDMs with a particular embedding dimension d are completely characterized by the rank and definiteness of $\boldsymbol{J}\boldsymbol{D}\boldsymbol{J}$.

Now we can write the following rank-constrained semidefinite program for solving Problem 1:

$$\begin{array}{ll}\underset{\boldsymbol{G}}{\text{minimize}} & \|\boldsymbol{W} \circ (\widetilde{\boldsymbol{D}} - \mathcal{K}(\boldsymbol{G}))\|_F^2 \\ \text{subject to} & \operatorname{rank}(\boldsymbol{G}) \le d \\ & \boldsymbol{G} \in \mathbb{S}_+^n \cap \mathbb{S}_c^n. \end{array} \tag{29}$$

The second constraint is just shorthand for writing $\boldsymbol{G} \succeq 0$, $\boldsymbol{G}\mathbf{1} = \mathbf{0}$. We note that this is equivalent to MDS with the s-stress cost function thanks to the rank characterization (28).

Unfortunately, the rank property makes the feasible set in (29) nonconvex, and solving it exactly becomes difficult. This makes sense, as we know that s-stress is not convex. Nevertheless, we may relax the hard problem by simply omitting the rank constraint and hope to obtain a solution with the correct dimensionality:

$$\begin{array}{ll}\underset{\boldsymbol{G}}{\text{minimize}} & \|\boldsymbol{W} \circ (\widetilde{\boldsymbol{D}} - \mathcal{K}(\boldsymbol{G}))\|_F^2 \\ \text{subject to} & \boldsymbol{G} \in \mathbb{S}_+^n \cap \mathbb{S}_c^n. \end{array} \tag{30}$$

We call (30) a semidefinite relaxation (SDR) of the rank-constrained program (29).

The constraint $\boldsymbol{G} \in \mathbb{S}_c^n$, or equivalently, $\boldsymbol{G}\mathbf{1} = \mathbf{0}$, means that there are no strictly positive definite solutions. ($\boldsymbol{G}$ has a nullspace, so at least one eigenvalue must be zero.) In other words, there exist no strictly feasible points [32]. This may pose a numerical problem, especially for various interior point methods. The idea is then to reduce the size of the Gram matrix through an invertible transformation, somehow removing the part of it responsible for the nullspace. In what follows, we describe how to construct this smaller Gram matrix.

A different, equivalent way to phrase the multiplicative characterization (11) is the following statement: a symmetric hollow matrix $\boldsymbol{D}$ is an EDM if and only if it is negative semidefinite on $\{\mathbf{1}\}^\perp$ (on all vectors $\boldsymbol{t}$ such that $\boldsymbol{t}^\top\mathbf{1} = 0$). Let us construct an orthonormal basis for this orthogonal complement—a subspace of dimension $(n-1)$—and arrange it in the columns of matrix $\boldsymbol{V} \in \mathbb{R}^{n\times(n-1)}$. We demand

$$\begin{aligned} \boldsymbol{V}^\top\mathbf{1} &= 0 \\ \boldsymbol{V}^\top\boldsymbol{V} &= \boldsymbol{I}. \end{aligned} \tag{31}$$

There are many possible choices for $\boldsymbol{V}$, but all of them obey that $\boldsymbol{V}\boldsymbol{V}^\top = \boldsymbol{I} - (1/n)\mathbf{1}\mathbf{1}^\top = \boldsymbol{J}$. The following choice is given in [2]:

$$\boldsymbol{V} = \begin{bmatrix} p & p & \cdots & p \\ 1+q & q & \cdots & q \\ q & 1+q & \cdots & q \\ \vdots & \cdots & \ddots & \vdots \\ q & q & \cdots & 1+q \end{bmatrix}, \tag{32}$$

where $p = -1/(n+\sqrt{n})$ and $q = -1/\sqrt{n}$.

With the help of the matrix $\boldsymbol{V}$, we can now construct the sought Gramian with reduced dimensions. For an EDM $\boldsymbol{D} \in \mathbb{R}^{n\times n}$,

$$\mathcal{G}(\boldsymbol{D}) \overset{\text{def}}{=} -\frac{1}{2}\boldsymbol{V}^\top\boldsymbol{D}\boldsymbol{V} \tag{33}$$

is an $(n-1)\times(n-1)$ PSD matrix. This can be verified by substituting (33) in (4). Additionally, we have that

$$\mathcal{K}(\boldsymbol{V}\mathcal{G}(\boldsymbol{D})\boldsymbol{V}^\top) = \boldsymbol{D}. \tag{34}$$

Indeed, $\boldsymbol{H} \mapsto \mathcal{K}(\boldsymbol{V}\boldsymbol{H}\boldsymbol{V}^\top)$ is an invertible mapping from $\mathbb{S}_+^{n-1}$ to $\mathbb{EDM}^n$ whose inverse is exactly $\mathcal{G}$. Using these notations, we can write down an equivalent optimization program that is numerically more stable than (30) [2],

$$\begin{array}{ll}\underset{\boldsymbol{H}}{\text{minimize}} & \|\boldsymbol{W} \circ (\tilde{\boldsymbol{D}} - \mathcal{K}(\boldsymbol{V}\boldsymbol{H}\boldsymbol{V}^\top))\|_F^2 \\ \text{subject to} & \boldsymbol{H} \in \mathbb{S}_+^{n-1}. \end{array} \tag{35}$$

On the one hand, with the previous transformation, the constraint $\boldsymbol{G}\mathbf{1} = \mathbf{0}$ became implicit in the objective, as $\boldsymbol{V}\boldsymbol{H}\boldsymbol{V}^\top\mathbf{1} \equiv \mathbf{0}$ by (31); on the other hand, the feasible set is now the full semidefinite cone $\mathbb{S}_+^{n-1}$.

Still, as Krislock and Wolkowicz mention [32], by omitting the rank constraint, we allow the points to move about in a larger space, so we may end up with a higher-dimensional solution even if there is a completion in dimension d.

There exist various heuristics for promoting lower rank. One such heuristic involves the trace norm—the convex envelope of rank. The trace or nuclear norm is studied extensively by the compressed sensing community. In contrast to the common wisdom in compressed sensing, the trick here is to maximize the trace norm, not to minimize it. The mechanics are as follows: maximizing the sum of squared distances between the points will stretch the configuration as much as possible, subject to available constraints. But stretching favors smaller affine dimensions (e.g., imagine pulling out a roll of paper or stretching a bent string). Maximizing the sum of squared distances can be rewritten as maximizing the sum of norms in a centered point configuration—but that is exactly the trace of the Gram matrix $\boldsymbol{G} = -(1/2)\boldsymbol{J}\boldsymbol{D}\boldsymbol{J}$ [9]. This idea has been

successfully put to work by Weinberger and Saul [9] in manifold learning and by Biswas et al. in SNL [49].

Noting that $\operatorname{trace}(H) = \operatorname{trace}(G)$ because $\operatorname{trace}(JDJ) = \operatorname{trace}(V^\top DV)$, we write the following SDR:

$$\begin{aligned} \underset{H}{\text{maximize}} \quad & \operatorname{trace}(H) - \lambda \| W \circ (\widetilde{D} - \mathcal{K}(VHV^\top)) \|_F \\ \text{subject to} \quad & H \in \mathbb{S}_+^{n-1}. \end{aligned} \tag{36}$$

Here, we opted to include the data fidelity term in the Lagrangian form, as proposed by Biswas et al. [49], but it could also be moved to constraints. Finally, in all of the above relaxations, it is straightforward to include upper and lower bounds on the distances. Because the bounds are linear constraints, the resulting programs remain convex; this is particularly useful in the molecular conformation problem. A MATLAB/CVX [50], [51] implementation of the SDR (36) is given in Algorithm 5.

Algorithm 5: SDR (MATLAB/CVX).

```
function EDM = sdr_complete_edm(D, W, lambda)

n = size(D, 1);
x = -1/(n + sqrt(n));
y = -1/sqrt(n);
V = [y*ones(1, n - 1); x*ones(n - 1) + eye(n - 1)];
e = ones(n, 1);

cvx_begin sdp
    variable G(n - 1, n - 1) symmetric;
    B = V*G*V';
    E = diag(B)*e' + e*diag(B)' - 2*B;
    maximize trace(G) - lambda * norm(W .* (E - D), 'fro');
    subject to
        G >= 0;
cvx_end

EDM = diag(B)*e' + e*diag(B)' - 2*B;
```

MULTIDIMENSIONAL UNFOLDING: A SPECIAL CASE OF COMPLETION

Imagine that we partition the point set into two subsets and that we can measure the distances between the points belonging to different subsets but not between the points in the same subset. MDU [30] refers to this special case of EDM completion.

MDU is relevant for the position calibration of ad hoc sensor networks, particularly of microphones. Consider an ad hoc array of m microphones at unknown locations. We can measure the distances to k point sources, also at unknown locations, for example, by emitting a pulse. (We assume that the sources and the microphones are synchronized.) We can always permute the points so that the matrix assumes the structure shown in Figure 4, with the unknown entries in two diagonal blocks. This is a standard scenario described, for example, in [27].

One of the early approaches to metric MDU is that of Schönemann [30]. We go through the steps of the algorithm and then explain how to solve the problem using the EDM toolbox. The goal is to make a comparison and emphasize the universality and simplicity of the introduced tools.

Denote by $R = [r_1, \ldots, r_m]$ the unknown microphone locations and by $S = [s_1, \ldots, s_k]$ the unknown source locations. The distance between the ith microphone and the jth source is

$$\delta_{ij} = \| r_i - s_j \|^2, \tag{37}$$

so that, in analogy with (3), we have

$$\Delta = \operatorname{edm}(R, S) = \operatorname{diag}(R^\top R)\mathbf{1}^\top - 2R^\top S + \mathbf{1}\operatorname{diag}(S^\top S), \tag{38}$$

where we overloaded the edm operator in a natural way. We use Δ to avoid confusion with the standard Euclidean D. Consider now two geometric centering matrices of sizes m and k, denoted as J_m and J_k. Similar to (14), we have

$$RJ_m = R - r_c\mathbf{1}^\top, \qquad SJ_k = S - s_c\mathbf{1}^\top. \tag{39}$$

This means that

$$J_m \Delta J_k = \widetilde{R}^\top \widetilde{S} \overset{\text{def}}{=} \widetilde{G} \tag{40}$$

is a matrix of the inner products between vectors $\widetilde{r}_i$ and $\widetilde{s}_j$. We used tildes to differentiate this from the real inner products between r_i and s_j because in (40), the points in $\widetilde{R}$ and $\widetilde{S}$ are referenced to different coordinate systems. The centroids r_c and s_c generally do not coincide. There are different ways to decompose $\widetilde{G}$ into a product of two full rank matrices, call them A and B

$$\widetilde{G} = A^\top B. \tag{41}$$

We could, for example, use the SVD, $\widetilde{G} = U\Sigma V^\top$ and set $A^\top = U$ and $B = \Sigma V^\top$. Any two such decompositions are linked by some invertible transformation $T \in \mathbb{R}^{d\times d}$

$$\widetilde{G} = A^\top B = \widetilde{R}^\top T^{-1} T\widetilde{S}. \tag{42}$$

We can now write down the conversion rule from what we can measure to what we can compute

$$\begin{aligned} R &= T^\top A + r_c\mathbf{1}^\top \\ S &= (T^{-1})^\top B + s_c\mathbf{1}^\top, \end{aligned} \tag{43}$$

where A and B can be computed according to (41). Because we cannot reconstruct the absolute position of the point set, we can arbitrarily set $r_c = 0$, and $s_c = \alpha e_1$. Recapitulating, we have that

$$\Delta = \operatorname{edm}(T^\top A, (T^{-1})^\top B + \alpha e_1\mathbf{1}^\top), \tag{44}$$

and the problem is reduced to computing $\boldsymbol{T}$ and α so that (44) holds, or in other words, so that the right-hand side is consistent with the data Δ. We reduced MDU to a relatively small problem: in 3-D, we need to compute only ten scalars. Schönemann [30] provides an algebraic method to find these parameters and mentions the possibility of least squares, while Crocco, Del Bue, and Murino [27] propose a different approach using nonlinear least squares.

This procedure seems quite convoluted. Rather, we see MDU as a special case of matrix completion, with the structure illustrated in Figure 4.

More concretely, represent the microphones and the sources by a set of $n = k + m$ points ascribed to the columns of matrix $X = [R\ S]$. Then, $\operatorname{edm}(X)$ has a special structure as seen in Figure 4,

$$\operatorname{edm}(X) = \begin{bmatrix} \operatorname{edm}(R) & \operatorname{edm}(R, S) \\ \operatorname{edm}(S, R) & \operatorname{edm}(S) \end{bmatrix}. \tag{45}$$

We define the mask matrix for MDU as

$$W_{\text{MDU}} \overset{\text{def}}{=} \begin{bmatrix} 0_{m\times m} & 1_{m\times k} \\ 1_{k\times m} & 0_{k\times k} \end{bmatrix}. \tag{46}$$

With this matrix, we can simply invoke the SDR in Algorithm 5. We could also use Algorithm 2 or Algorithm 4. The performance of different algorithms is compared in the next section.

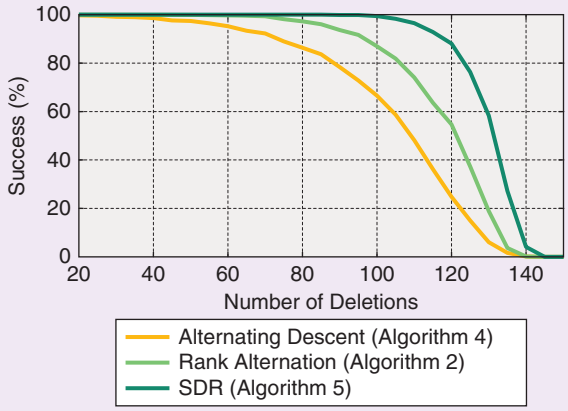
It is worth mentioning that SNL-specific algorithms that exploit the particular graph induced by limited range communication do not perform well on MDU. This is because the structure of the missing entries in MDU is in a certain sense opposite to the one of SNL.

PERFORMANCE COMPARISON OF ALGORITHMS

We compare the described algorithms in two different EDM completion settings. In the first experiment (Figures 5 and 6), the entries to delete are chosen uniformly at random. The second experiment (Figures 7 and 8) tests the performance in MDU, where the nonobserved entries are highly structured. In Figures 5 and 6, we assume that the observed entries are known exactly, and we plot the success rate (percentage of accurate EDM reconstructions) against the number of deletions in the first case and the number of calibration events in the second case. Accurate reconstruction is defined in terms of the relative error. Let $\boldsymbol{D}$ be the true and $\widehat{\boldsymbol{D}}$ the estimated EDM. The relative error is then $\|\widehat{\boldsymbol{D}} - \boldsymbol{D}\|_F / \|\boldsymbol{D}\|_F$, and we declare success if this error is below 1%.

To generate Figures 6 and 8, we varied the amount of random, uniformly distributed jitter added to the distances, and for each jitter level, we plotted the relative error. The exact values of intermediate curves are less important than the curves for the smallest and largest jitter and the overall shape of the ensemble.

A number of observations can be made about the performance of algorithms. Notably, OptSpace (Algorithm 3) does not perform well for randomly deleted entries when $n = 20$; it was designed for larger matrices. For this matrix size, the mean relative reconstruction error achieved by OptSpace is the worst of all algorithms (Figure 6). In fact, the relative error in the noiseless case was rarely below the success threshold (set to 1%), so we omitted the corresponding near-zero curve from Figure 5. Furthermore, OptSpace assumes that the pattern of missing entries is random; in the case of a blocked deterministic structure associated with MDU, it never yields a satisfactory completion.

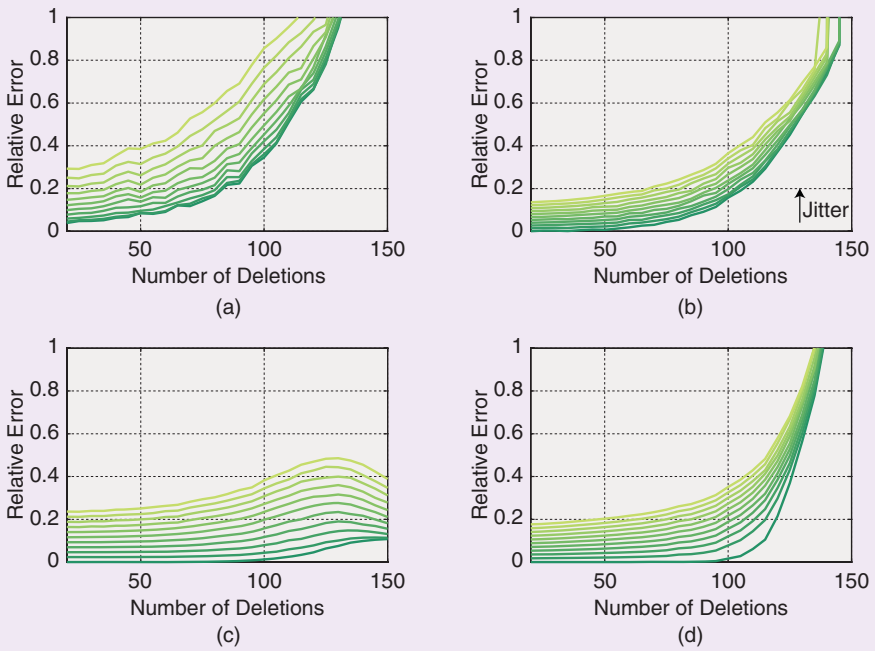


[FIG5] A comparison of different algorithms applied to completing an EDM with random deletions. For every number of deletions, we generated 2,000 realizations of 20 points uniformly at random in a unit square. The distances to delete were chosen uniformly at random among the resulting (1/2) * 20 * (20 − 1) = 190 pairs; 20 deletions correspond to ≈ 10% of the number of distance pairs and to 5% of the number of matrix entries; 150 deletions correspond to ≈ 80% of the distance pairs and to ≈ 38% of the number of matrix entries. Success was declared if the Frobenius norm of the error between the estimated matrix and the true EDM was less than 1% of the Frobenius norm of the true EDM.

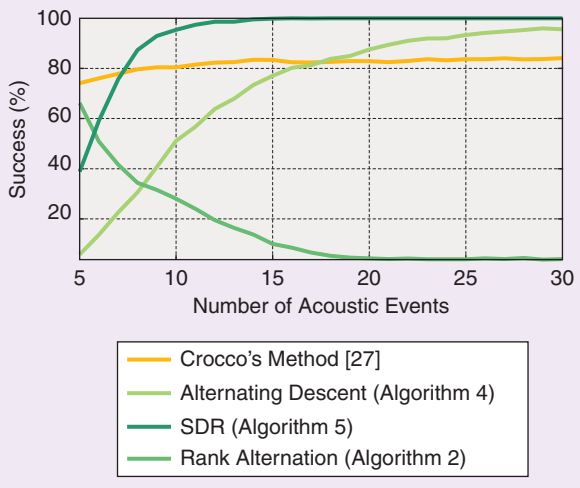
On the other hand, when the unobserved entries are randomly scattered in the matrix, and the matrix is large—in the ultrasonic calibration example, the number of sensors n was 200 or more—OptSpace is a very fast and attractive algorithm. To fully exploit OptSpace, n should be even larger, in the thousands or tens of thousands.

SDR (Algorithm 5) performs well in all scenarios. For both the random deletions and the MDU, it has the highest success rate and it behaves well with respect to noise. Alternating coordinate descent (Algorithm 4) performs slightly better in noise for a small number of deletions and a large number of calibration events, but Figures 5 and 7 indicate that, for certain realizations of the point set, it gives large errors. If the worst-case performance is critical, SDR is a better choice. We note that, in the experiments involving the SDR, we have set the multiplier λ in (36) to the square root of the number of missing entries. This simple choice was empirically found to perform well.

The main drawback of SDR is the speed; it is the slowest among the tested algorithms. To solve the semidefinite program, we used CVX [50], [51], a MATLAB interface to various interior point methods. For larger matrices (e.g., $n = 1,000$), CVX runs out of memory on a desktop computer, and essentially never finishes. MATLAB implementations of alternating coordinate descent, rank alternation (Algorithm 2), and OptSpace are all much faster.



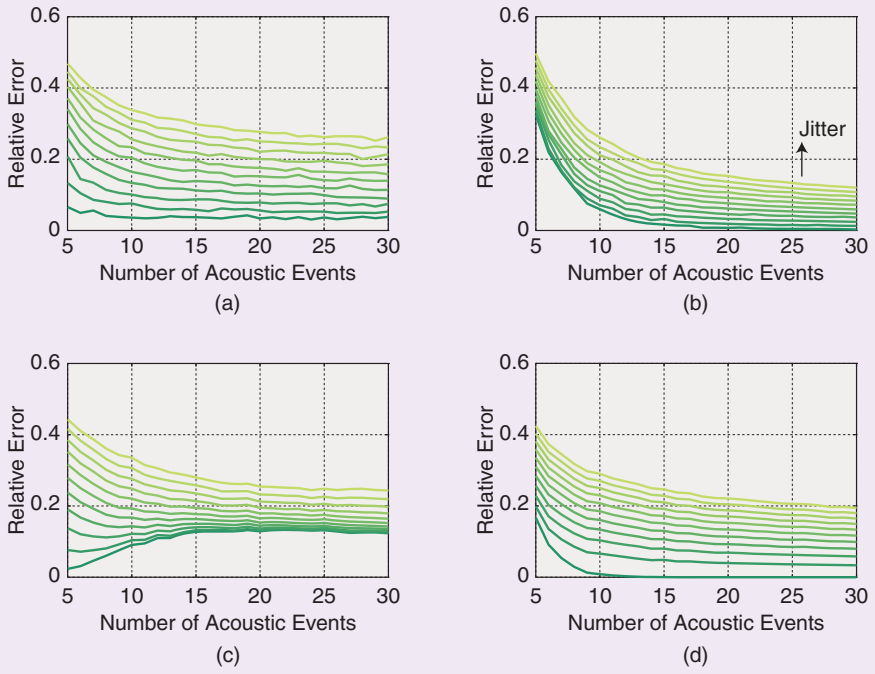
[FIG6] A comparison of different algorithms applied to completing an EDM with random deletions and noisy distances. For every number of deletions, we generated 1,000 realizations of 20 points uniformly at random in a unit square. In addition to the number of deletions, we varied the amount of jitter added to the distances. Jitter was drawn from a centered uniform distribution, with the level increasing in the direction of the arrow, from $\mathcal{U}[0, 0]$ (no jitter) for the darkest curve at the bottom, to $\mathcal{U}[-0.15, 0.15]$ for the lightest curve at the top, in 11 increments. For every jitter level, we plotted the mean relative error $\|\hat{D} - D\|_F / \|D\|_F$ for all algorithms. (a) OptSpace (Algorithm 3). (b) Alternating descent (Algorithm 4). (c) The rank alternation (Algorithm 2). (d) SDR (Algorithm 5).



[FIG7] A comparison of different algorithms applied to MDU with a varying number of acoustic events k. For every number of acoustic events, we generated 3,000 realizations of $m = 20$ microphone locations uniformly at random in a unit cube. The percentage of the missing matrix entries is given as $(k^2 + m^2)/(k + m)^2$ so that the ticks on the abscissa correspond to [68, 56, 51, 50, 51, 52] % (nonmonotonic in k with the minimum for $k = m = 20$). Success was declared if the Frobenius norm of the error between the estimated matrix and the true EDM was less than 1% of the Frobenius norm of the true EDM.

The microphone calibration algorithm by Crocco [27] performs equally well for any number of acoustic events. This may be explained by the fact that it always reduces the problem to ten unknowns. It is an attractive choice for practical calibration problems with a smaller number of calibration events. The algorithm's success rate can be further improved if one is prepared to run it for many random initializations of the nonlinear optimization step.

Interesting behavior can be observed for the rank alternation in MDU. Figures 7 and 8 show that, at low noise levels, the performance of the rank alternation becomes worse with the number of acoustic events. At first glance, this may seem counterintuitive, as more acoustic events means more information; one could simply ignore some of them and perform at least equally well as with fewer events. But this reasoning presumes that the method is aware of the geometrical meaning of the matrix entries; on the contrary, rank alternation is using only rank. Therefore, even if the percentage of the observed matrix entries grows until a certain point, the size of the structured blocks of unknown entries grows as well (and the percentage of known entries in columns/rows corresponding to acoustic events decreases). This makes it harder for a method that does not use geometric relationships to complete the matrix. A loose comparison can be made to image inpainting: If the pixels are missing randomly, many methods will do a good job, but if a large patch is missing, we cannot do much without additional structure (in our case geometry) no matter how large the rest of the image is.



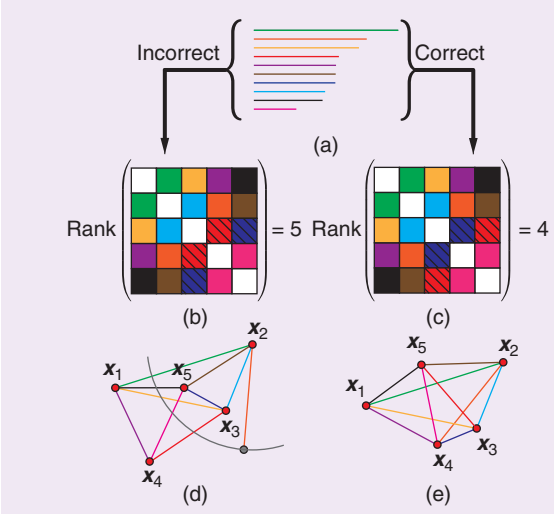
[FIG8] A comparison of different algorithms applied to MDU with a varying number of acoustic events k and noisy distances. For every number of acoustic events, we generated 1,000 realizations of $m = 20$ microphone locations uniformly at random in a unit cube. In addition to the number of acoustic events, we varied the amount of random jitter added to the distances, with the same parameters as in Figure 6. For every jitter level, we plotted the mean relative error $\|\widehat{D} - D\|_F / \|D\|_F$ for all algorithms. (a) Crocco's method [27]. (b) Alternating descent (Algorithm 4). (c) Rank alternation (Algorithm 2) and SDR (Algorithm 5).

To summarize, for smaller and moderately sized matrices, the SDR seems to be the best overall choice. For large matrices, the SDR becomes too slow and one should turn to alternating coordinate descent, rank alternation, or OptSpace. Rank alternation is the simplest algorithm, but alternating coordinate descent performs better. For very large matrices (n on the order of thousands or tens of thousands), OptSpace becomes the most attractive solution. We note that we deliberately refrained from making detailed running time comparisons due to the diverse implementations of the algorithms.

SUMMARY

In this section, we discussed:

- the problem statement for EDM completion and denoising and how to easily exploit the rank property (Algorithm 2)
- standard objective functions in MDS, raw stress and s-stress, and a simple algorithm to minimize s-stress (Algorithm 4)
- different SDRs that exploit the connection between EDMs and PSD matrices
- MDU and how to solve it efficiently using EDM completion
- performance of the introduced algorithms in two very different scenarios: EDM completion with randomly unobserved entries and EDM completion with a deterministic block structure of unobserved entries (MDU).



[FIG9] An illustration of the uniqueness of EDMs for unlabeled distances. A set of unlabeled distance (a) is distributed in two different ways in a tentative EDM with embedding dimension two (b) and (c). The correct assignment yields the matrix with the expected rank (c), and the point set is easily realized in the plane (e). On the contrary, swapping just two distances [the hatched squares in (b) and (c)] makes it impossible to realize the point set in the plane (d). Triangles that do not coincide with the swapped edges can still be placed, but in the end, we are left with a hanging orange stick that cannot attach itself to any of the five nodes.

EDM PERSPECTIVE ON SPARSE PHASE RETRIEVAL (THE UNEXPECTED DISTANCE STRUCTURE)

In many cases, it is easier to measure a signal in the Fourier domain. Unfortunately, it is common in these scenarios that we can only reliably measure the magnitude of the Fourier transform (FT). We would like to recover the signal of interest from just the magnitude of its FT, hence the name *phase retrieval*. X-ray crystallography [54] and speckle imaging in astronomy [55] are classic examples of phase retrieval problems. In both of these applications, the signal is spatially sparse. We can model it as

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} c_i \delta(\boldsymbol{x} - \boldsymbol{x}_i), \tag{S1}$$

where c_i are the amplitudes and $\boldsymbol{x}_i$ are the locations of the n Dirac deltas in the signal. In what follows, we discuss the problem on 1-D domains, that is, for $\boldsymbol{x} \in \mathbb{R}$, knowing that a multidimensional phase retrieval problem can be solved by solving multiple 1-D problems [7].

Note that measuring the magnitude of the FT of $f(\boldsymbol{x})$ is equivalent to measuring its ACF. For a sparse $f(\boldsymbol{x})$, the ACF is also sparse and is given as

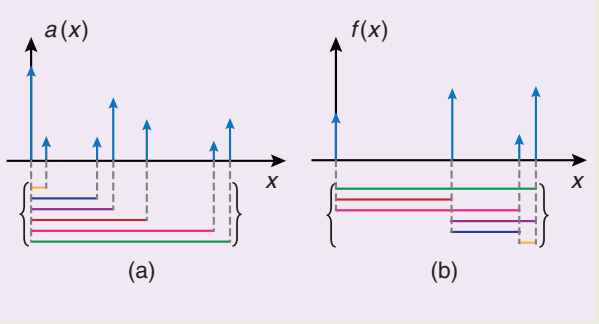
$$a(\boldsymbol{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \delta(\boldsymbol{x} - (\boldsymbol{x}_i - \boldsymbol{x}_j)), \tag{S2}$$

where we note the presence of differences between the locations $\boldsymbol{x}_i$ in the support of the ACF. As $a(\boldsymbol{x})$ is symmetric, we do not know the order of $\boldsymbol{x}_i$ and so we can only know these differences up to a sign, which is equivalent to knowing the distances $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ (Figure S3).

For the following reasons, we focus on the recovery of the support of the signal $f(\boldsymbol{x})$ from the support of the ACF $a(\boldsymbol{x})$: 1) in certain applications, the amplitudes c_i may be all equal, thus limiting their role in the reconstruction and 2) knowing the support of $f(\boldsymbol{x})$ and its ACF is sufficient to exactly recover the signal $f(\boldsymbol{x})$ [7].

The recovery of the support of $f(\boldsymbol{x})$ from the one of $a(\boldsymbol{x})$ corresponds to the localization of a set of n points from their unlabeled distances: we have access to all the pairwise distances but we do not know which pair of points corresponds to any given distance. This can be recognized as an instance of the turnpike problem, whose computational complexity is believed not to be NP-hard but for which no polynomial time algorithm is known [56].

From an EDM perspective, we can design a reconstruction algorithm recovering the support of the signal $f(\boldsymbol{x})$ by labeling the distances obtained from the ACF such that the resulting EDM has a rank that is less than or equal to three. This can be regarded as unidimensional scaling with unlabeled distances, and the algorithm to solve it is similar to echo sorting (Algorithm 6).



[FIGS3] **A graphical representation of the phase retrieval problem for 1-D sparse signals. (a) We measure the ACF of the signal and we recover a set of distances (sticks in Figure 9) from its support. (b) These are the unlabeled distances between all the pairs of Dirac deltas in the signal $f(\boldsymbol{x})$. We exactly recover the support of the signal if we correctly label the distances.**

UNLABELED DISTANCES

In certain applications, we can measure the distances between the points, but we do not know the correct labeling. That is, we know all the entries of an EDM, but we do not know how to arrange them in the matrix. As illustrated in Figure 9(a), we can imagine having a set of sticks of various lengths. The task is to work out the correct way to connect the ends of different sticks so that no stick is left hanging open-ended.

In this section, we exploit the fact that, in many cases, distance labeling is not essential. For most point configurations, there is no other set of points that can generate the corresponding set of distances up to a rigid transformation.

Localization from unlabeled distances is relevant in various calibration scenarios where we cannot tell apart distance measurements belonging to different points in space. This can occur when we measure the TOAs of echoes, which correspond to the distances between the microphones and the image sources (ISs) (see Figure 10) [6], [29]. Somewhat surprisingly, the same problem of unlabeled distances appears in sparse phase retrieval; see "EDM Perspective on Sparse Phase Retrieval (the Unexpected Distance Structure)."

No efficient algorithm currently exists for localization from unlabeled distances in the general case of noisy distances. We should mention, however, a recent polynomial-time algorithm (albeit of a high degree) by Gujarathi et al. [31] that can reconstruct relatively large point sets from unordered, noiseless distance data.

At any rate, the number of assignments to test is sometimes small enough that an exhaustive search does not present a problem. We can then use EDMs to find the best labeling. The key to the unknown permutation problem is the following fact.

Theorem 3: *Draw $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n \in \mathbb{R}^d$ independently from some absolutely continuous probability distribution (e.g., uniformly at random) on $\Omega \subseteq \mathbb{R}^d$. Then, with probability 1, the obtained point configuration is the unique (up to a rigid*

transformation) point configuration in Ω *that generates the set of distances* $\{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|, 1 \le i < j \le n\}$.

This fact is a simple consequence of a result by Boutin and Kemper [52] who provide a characterization of point sets reconstructable from unlabeled distances. Figure 9(b) and (c) shows two possible arrangements of the set of distances in a tentative EDM; the only difference is that the two hatched entries are swapped. But this simple swap is not harmless: there is no way to attach the last stick in Figure 9(d) while keeping the remaining triangles consistent. We could do it in a higher embedding dimension, but we insist on realizing it in the plane.

What Theorem 3 does not tell us is how to identify the correct labeling. But we know that for most sets of distances, only one (correct) permutation can be realized in the given embedding dimension. Of course, if all the labelings are unknown and we have no good heuristics to trim the solution space, finding the correct labeling is difficult, as noted in [31]. Yet there are interesting situations where this search is feasible because we can augment the EDM point by point. We describe one such situation next.

AN IMPORTANT APPLICATION OF EDMs WITH UNLABELED DISTANCES IS THE RECONSTRUCTION OF THE ROOM SHAPE FROM ECHOES.

HEARING THE SHAPE OF A ROOM

An important application of EDMs with unlabeled distances is the reconstruction of the room shape from echoes [6]. An acoustic setup is shown in Figure 10(a), but one could also use radio signals. Microphones pick up the convolution of the sound emitted by the loudspeaker with the room impulse response (RIR), which can be estimated by knowing the emitted sound. An example RIR recorded by one of the microphones is illustrated in Figure 10(b), with peaks highlighted in green. Some of these peaks are first-order echoes coming from different walls, and some are higher-order echoes or just noise.

Echoes are linked to the room geometry by the image source (IS) model [53]. According to this model, we can replace echoes by ISs—mirror images of the true sources across the corresponding walls. The position of the IS of $\boldsymbol{s}$ corresponding to wall i is computed as

$$\widetilde{\boldsymbol{s}}_i = \boldsymbol{s} + 2\langle \boldsymbol{p}_i - \boldsymbol{s}, \boldsymbol{n}_i \rangle \boldsymbol{n}_i, \tag{47}$$

where $\boldsymbol{p}_i$ is any point on the ith wall and $\boldsymbol{n}_i$ is the unit normal vector associated with the ith wall [see Figure 10(a)].

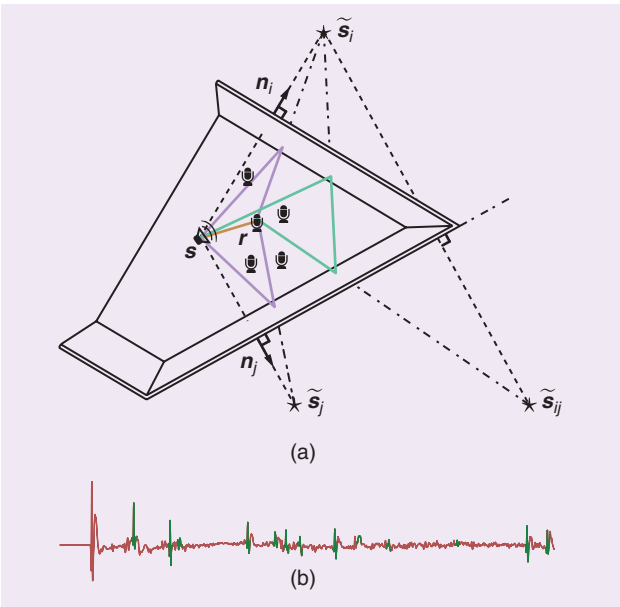
A convex room with planar walls is completely determined by the locations of first-order ISs [6], so by reconstructing their locations, we actually reconstruct the room's geometry.

We assume that the loudspeaker and the microphones are synchronized so that the times at which the echoes arrive directly correspond to distances. The challenge is that the distances—the green peaks in Figure 10(b)—are unlabeled: it might happen that the kth peak in the RIR from microphone 1 and the kth peak in the RIR from microphone 2 come from different walls, especially for larger microphone arrays. Thus, we have to address the problem of echo sorting to group peaks corresponding to the same IS in RIRs from different microphones.

Assuming that we know the pairwise distances between the microphones $\boldsymbol{R} = [\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m]$, we can create an EDM corresponding to the microphone array. Because echoes correspond to ISs, and ISs are just points in space, we attempt to grow that EDM by adding one point—an IS—at a time. To do that, we pick one echo from every microphone's impulse response, augment the EDM based on echo arrival times, and check how far the augmented matrix is from an EDM with embedding dimension three, as we work in 3-D space. The distance from an EDM is measured with the s-stress cost function. It was shown in [6] that a variant of Theorem 3 applies to ISs when microphones are positioned at random. Therefore, if the augmented matrix satisfies the EDM properties, almost surely we have found a good IS. With probability 1, no other combination of points could have generated the used distances.

The main reason for using EDMs and s-stress instead of, for instance, the rank property is that we get robust algorithms. The echo arrival times are corrupted with various errors, and relying on the rank is too brittle. It was verified experimentally [6] that EDMs and s-stress yield a very robust filter for the correct combinations of echoes.

Thus, we may try all feasible combinations of echoes and expect to get exactly one "good" combination for every IS that is



[FIG10] (a) An illustration of the IS model for first- and second-order echoes. Vector $\boldsymbol{n}_i$ is the outward-pointing unit normal associated with the ith wall. The stars denote the IS, and $\widetilde{\boldsymbol{s}}_{ij}$ is the IS corresponding to the second-order echo. The sound rays corresponding to first reflections are shown in purple, and the ray corresponding to the second-order reflection is shown in green. (b) The early part of a typical recorded RIR.

"visible" in the impulse responses. In this case, as we are only adding a single point, the search space is small enough to be rapidly traversed exhaustively. Geometric considerations allow for a further trimming of the search space: because we know the diameter of the microphone array, we know that an echo from a particular wall must arrive at all the microphones within a temporal window corresponding to the array's diameter.

The procedure is as follows: collect all echo arrival times received by the ith microphone in the set T_i and fix $t_1 \in T_1$ corresponding to a particular IS. Then, Algorithm 6 finds echoes in other microphones' RIRs that correspond to this same IS. Once we group all the peaks corresponding to one IS, we can determine its location by multilateration (e.g., by running the classical MDS) and then repeat the process for other echoes in T_1.

To get a ballpark idea of the number of combinations to test, suppose that we detect 20 echoes per microphone and that the diameter of the five-microphone array is 1 m. (We do not need to look beyond early echoes corresponding to at most three bounces; this is convenient as echoes of higher orders are challenging or impossible to isolate.) Thus, for every peak time $t_1 \in T_1$, we have to look for peaks in the remaining four microphones that arrived within a window around t_1 of length $2 \times (1\,\text{m}/343\,\text{m/s})$, where 343 m/s is the speed of sound. This is approximately 6 ms, and in a typical room, we can expect about five early echoes within a window of that duration. Thus, we have to compute the s-stress for $20 \times 5^4 = 12{,}500$ matrices of size 6×6, which can be done in a matter of seconds (or less) on a desktop computer. In fact, once we assign an echo to an IS, we can exclude it from further testing, so the number of combinations can be further reduced.

Algorithm 6: Echo sorting [6].

1: function `EchoSort` $(R, t_1, \ldots, T_m)$
2: $D \leftarrow \text{edm}(R)$
3: $s_{\text{best}} \leftarrow +\text{Inf}$
4: for all $t = [t_2, \ldots, t_m]$, such that $t_i \in T_i$ do
5: $d \leftarrow c \cdot [t_1, t^\top]^\top$ ▷ c is the sound speed
6: $D_{\text{aug}} \leftarrow \begin{bmatrix} D & d \\ d^\top & 0 \end{bmatrix}$
7: if s-stress$(D_{\text{aug}}) < s_{\text{best}}$ then
8: $s_{\text{best}} \leftarrow$ s-stress(D_{aug})
9: $d_{\text{best}} \leftarrow d$
10: end if
11: end for
12: return d_{best}
13: end function

Algorithm 6 was used to reconstruct rooms with centimeter precision [6] with one loudspeaker and an array of five microphones. The same algorithm also enables a dual application: indoor localization of an acoustic source using only one microphone—a feat not possible if we are not in a room [57].

[TABLE 2] APPLICATIONS OF EDMs WITH DIFFERENT TWISTS.

APPLICATION	MISSING DISTANCES	NOISY DISTANCES	UNLABELED DISTANCES
WIRELESS SENSOR NETWORKS	✔	✔	×
MOLECULAR CONFORMATION	✔	✔	×
HEARING THE SHAPE OF A ROOM	×	✔	✔
INDOOR LOCALIZATION	×	✔	✔
CALIBRATION	✔	✔	×
SPARSE PHASE RETRIEVAL	×	✔	✔

SUMMARY

To summarize this section:

- We explained that for most point sets, the distances they generate are unique; there are no other point sets generating the same distances.
- In room reconstruction from echoes, we need to identify the correct assignment of the distances to ISs. EDMs act as a robust filter for echoes coming from the same IS.
- Sparse phase retrieval can be cast as a distance problem, too. The support of the ACF gives us the distances between the deltas in the original signal. Echo sorting can be adapted to solve the problem from the EDM perspective.

IDEAS FOR FUTURE RESEARCH

Even problems that at first glance seem to have little to do with EDMs sometimes reveal a distance structure when you look closely. A good example is sparse phase retrieval.

The purpose of this article is to convince the reader that EDMs are powerful objects with a multitude of applications (Table 2 lists various flavors) and that they should belong to any practitioner's toolbox. We have an impression that the power of EDMs and the associated algorithms has not been sufficiently recognized in the signal processing community, and our goal is to provide a good starting reference. To this end, and perhaps to inspire new research directions, we list several EDM-related problems that we are curious about and believe are important.

DISTANCE MATRICES ON MANIFOLDS

If the points lie on a particular manifold, what can be said about their distance matrix? We know that if the points are on a circle, the EDM has a rank of three instead of four, and this generalizes to hyperspheres [17]. But what about more general manifolds? Are there invertible transforms of the data or of the Gram matrix that yield EDMs with a lower rank than the embedding dimension suggests? What about different distances, e.g., the geodesic distance on the manifold? The answers to these questions have immediate applications in machine learning, where the data can be approximately assumed to be on a smooth surface [23].

PROJECTIONS OF EDMs ON LOWER-DIMENSIONAL SUBSPACES

What happens to an EDM when we project its generating points to a lower-dimensional space? What is the minimum number of

projections that we need to be able to reconstruct the original point set? The answers to these questions have a significant impact on imaging applications such as X-ray crystallography and seismic imaging. What happens when we only have partial distance observations in various subspaces? What are the other useful low-dimensional structures on which we can observe the high-dimensional distance data?

EFFICIENT ALGORITHMS FOR DISTANCE LABELING

Without application-specific heuristics to trim down the search space, identifying correct labeling of the distances quickly becomes an arduous task. Can we identify scenarios for which there are efficient labeling algorithms? What happens when we do not have the labeling, but we also do not have the complete collection of sticks? What can we say about the uniqueness of incomplete unlabeled distance sets? Some of the questions have been answered by Gujarathi et al. [31], but many remain. The quest is on for faster algorithms as well as algorithms that can handle noisy distances.

In particular, if the noise distribution on the unlabeled distances is known, what can we say about the distribution of the reconstructed point set (taking in some sense the best reconstruction over all labelings)? Is it compact, or can we jump to totally wrong assignments with positive probability?

ANALYTICAL LOCAL MINIMUM OF S-STRESS

Everyone agrees that there are many, but, to the best of our knowledge, no analytical minimum of s-stress has yet been found.

CONCLUSIONS

We hope that we have succeeded in showing how universally useful EDMs are and that readers will be inspired to dig deeper after coming across this material. Distance measurements are so common that a simple, yet sophisticated tool like EDMs deserves attention. A good example is the SDR: even though it is generic, it is the best-performing algorithm for the specific problem of ad hoc microphone array localization. Continuing research on this topic will bring new revolutions like it did in the 1980s in crystallography. Perhaps the next one will be fueled by solving the labeling problem.

ACKNOWLEDGMENTS

We would like to thank Dr. Farid M. Naini for his help in expediting the numerical simulations. We would also like to thank the anonymous reviewers for their numerous insightful suggestions that have improved the article. Ivan Dokmanić and Juri Ranieri were supported by the ERC Advanced Grant—Support for Frontier Research—SPARSAM, number 247006. Ivan Dokmanić was also supported by the Google Ph.D. Fellowship.

AUTHORS

Ivan Dokmanić (ivan.dokmanic@epfl.ch) is a Ph.D. candidate in the Audiovisual Communications Laboratory (LCAV) at the École Polytechnique Fédérale de Lausanne (expected graduation in May 2015). His interests include inverse problems, audio and acoustics, signal processing for sensor arrays/networks, and fundamental signal processing. He was previously a teaching assistant at the University of Zagreb, a codec developer for MainConcept AG, and a digital audio effects designer for Little Endian Ltd. During the summer of 2013, he was a research intern with Microsoft Research in Redmond, Washington, where he worked on ultrasonic sensing. For his work on room shape reconstruction using sound, he received the Best Student Paper Award at the 2011 IEEE International Conference on Acoustics, Speech, and Signal Processing. In 2014, he received a Google Ph.D. fellowship.

Reza Parhizkar (reza.parhizkar@gmail.com) received his B.Sc. degree in electrical engineering from Sharif University, Tehran, Iran, in 2003 and his M.Sc. and Ph.D. degrees in communication systems from the École Polytechnique Fédérale de Lausanne in 2009 and 2013. He was an intern at the Nokia research center, Lausanne, and Qualcomm, Inc., San Diego, California. He is currently the head of research in macx red AG-Zug, Switzerland, where he works on optimization frameworks in finance. His work on sensor calibration for ultrasound tomography devices won the Best Student Paper Award at the 2011 IEEE International Conference on Acoustics, Speech, and Signal Processing. His Ph.D. thesis "Euclidean Distance Matrices: Properties, Algorithms, and Applications" was nominated by the thesis committee for the ABB Best Ph.D. Thesis Award in 2014. His research interests include mathematical signal processing, Euclidean distance matrices, and finance.

Juri Ranieri (juri.ranieri@epfl.ch) received his M.S. and B.S. degrees in electronic engineering in 2009 and 2007, respectively, from the Universitá di Bologna, Italy. From July to December 2009, he was a visiting student at the Audiovisual Communications Laboratory (LCAV) at the École Polytechnique Fédérale de Lausanne (EPFL). From January to August 2010, he was with IBM Zurich to investigate the lithographic process as a signal processing problem. From September 2010 to September 2014, he was with the doctoral school at EPFL, where he obtained his Ph.D. degree at LCAV under the supervision of Prof. Martin Vetterli and Prof. Amina Chebira. From April to July 2013, he was an intern at Lyric Labs of Analog Devices, Cambridge, Massachusetts. His main research interests are inverse problems, sensor placement, and sparse phase retrieval.

Martin Vetterli (martin.vetterli@epfl.ch) received his engineering degree from Eidgenöessische Technische Hochschule, Zürich, Switzerland, his M.S. degree from Stanford University, California, and his doctorate degree from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. He was on the faculty of Columbia University and the University of California, Berkeley, before joining EPFL. He is currently the president of the National Research Council of the Swiss National Science Foundation. His research interests are in signal processing and communications, e.g., wavelet theory and applications, sparse sampling, joint source–channel coding, and sensor networks. He received the Best Paper Awards of the IEEE Signal Processing Society (1991, 1996, and 2006) and the IEEE Signal Processing Technical and Society Awards (2002 and 2010). He is a Fellow of the IEEE, the Association for Computing Machinery, and the European Association for Signal Processing. He is the coauthor of *Wavelets and Subband Coding* (1995), *Signal Processing for Communications* (2008), and *Foundations of Signal Processing* (2014). He is a highly cited researcher in engineering (Thomson ISI Web of Science, 2007 and 2014).

REFERENCES

[1] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Processing Mag*., vol. 22, no. 4, pp. 54–69, July 2005.

[2] A. Y. Alfakih, A. Khandani, and H. Wolkowicz, "Solving Euclidean distance matrix completion problems via semidefinite programming," *Comput. Optim. Appl*., vol. 12, nos. 1–3, pp. 13–30, Jan. 1999.

[3] L. Doherty, K. Pister, and L. El Ghaoui, "Convex position estimation in wireless sensor networks," in *Proc. IEEE Conf. Computer Communications (INFOCOM)*, 2001, vol. 3, pp. 1655–1663.

[4] P. Biswas and Y. Ye, "Semidefinite programming for ad hoc wireless sensor network localization," in *Proc. ACM/IEEE Int. Conf. Information Processing in Sensor Networks,* 2004, pp. 46–54.

[5] T. F. Havel and K. Wüthrich, "An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution," *J. Mol. Biol*., vol. 182, no. 2, pp. 281–294, 1985.

[6] I. Dokmanic´, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli, "Acoustic echoes reveal room shape," *Proc. Natl. Acad. Sci*., vol. 110, no. 30, pp. 12186–12191, June 2013.

[7] J. Ranieri, A. Chebira, Y. M. Lu, and M. Vetterli, "Phase retrieval for sparse signals: Uniqueness conditions," *IEEE Trans. Inform. Theory*, arXiv:1308.3058v2.

[8] W. S. Torgerson, "Multidimensional scaling: I. Theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, 1952.

[9] K. Q. Weinberger and L. K. Saul, "Unsupervised learning of image manifolds by semidefinite programming," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. II-988–II-995, 2004.

[10] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino, "Euclidean distance geometry and applications," *SIAM Rev*., vol. 56, no. 1, pp. 3–69, 2014.

[11] K. Menger, "Untersuchungen über allgemeine metrik," *Math. Ann*., vol. 100, no. 1, pp. 75–163, Dec. 1928.

[12] I. J. Schoenberg, "Remarks to Maurice Frechet's article 'Sur la définition axiomatique d'une classe d'espace distancés vectoriellement applicable sur l'espace de Hilbert,'" *Ann. Math*., vol. 36, no. 3, p. 724, July 1935.

[13] L. M. Blumenthal, *Theory and Applications of Distance Geometry*. Oxford, U.K.: Clarendon Press, 1953.

[14] G. Young and A. Householder, "Discussion of a set of points in terms of their mutual distances," *Psychometrika*, vol. 3, no. 1, pp. 19–22, 1938.

[15] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.

[16] J. C. Gower, "Euclidean distance geometry," *Math. Sci*., vol. 7, pp. 1–14, 1982.

[17] J. C. Gower, "Properties of Euclidean and non-Euclidean distance matrices," *Linear Algebra Appl*., vol. 67, pp. 81–97, June 1985.

[18] W. Glunt, T. L. Hayden, S. Hong, and J. Wells, "An alternating projection algorithm for computing the nearest Euclidean distance matrix," *SIAM J. Matrix Anal. Appl*., vol. 11, no. 4, pp. 589–600, 1990.

[19] T. L. Hayden, J. Wells, W.-M. Liu, and P. Tarazaga, "The cone of distance matrices," *Linear Algebra Appl*., vol. 144, pp. 153–169, Jan. 1990.

[20] J. Dattorro, *Convex Optimization & Euclidean Distance Geometry*. Palo Alto, California: Meboo, 2011.

[21] M. W. Trosset, "Applications of multidimensional scaling to molecular conformation," *Comp. Sci. Stat*., vol. 29, pp. 148–152, 1998.

[22] L. Holm and C. Sander, "Protein structure comparison by alignment of distance matrices," *J. Mol. Biol*., vol. 233, no. 1, pp. 123–138, Sept. 1993.

[23] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[24] V. Jain and L. Saul, "Exploratory analysis and visualization of speech and music by locally linear embedding," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP),* Philadelphia, PA, 2004, vol. 2, pp. II-988–II-995.

[25] E. D. Demaine, F. Gomez-Martin, H. Meijer, D. Rappaport, P. Taslakian, G. T. Toussaint, T. Winograd, and D. R. Wood, "The distance geometry of music," *Comput. Geom*., vol. 42, no. 5, pp. 429–454, July 2009.

[26] A. M.-C. So and Y. Ye, "Theory of semidefinite programming for sensor network localization," *Math. Program*., vol. 109, nos. 2–3, pp. 367–384, Mar. 2007.

[27] M. Crocco, A. D. Bue, and V. Murino, "A bilinear approach to the position self-calibration of multiple sensors," *IEEE Trans. Signal Processing*, vol. 60, no. 2, pp. 660–673, 2012.

[28] M. Pollefeys and D. Nister, "Direct computation of sound and microphone locations from time-difference-of-arrival data," in *Proc. Int. Workshop HSC*, Las Vegas, NV, 2008, pp. 2445–2448.

[29] I. Dokmanić, L. Daudet, and M. Vetterli, "How to localize ten microphones in one fingersnap," in *Proc. European Signal Processing Conf.* (*EUSIPCO*), pp. 2275–2279, 2014.

[30] P. H. Schönemann, "On metric multidimensional unfolding," *Psychometrika*, vol. 35, no. 3, pp. 349–366, 1970.

[31] S. R. Gujarathi, C. L. Farrow, C. Glosser, L. Granlund, and P. M. Duxbury, "Ab-initio reconstruction of complex Euclidean networks in two dimensions," *Phys. Rev. E*, vol. 89, no. 5, p. 053311, 2014.

[32] N. Krislock and H. Wolkowicz, "Euclidean distance matrices and applications," in *Handbook on Semidefinite, Conic and Polynomial Optimization*. Boston, MA: Springer, Jan. 2012, pp. 879–914.

[33] A. Mucherino, C. Lavor, L. Liberti, and N. Maculan, *Distance Geometry: Theory, Methods, and Applications*. New York, NY: Springer Science & Business Media, Dec. 2012.

[34] P. H. Schönemann, "A solution of the orthogonal procrustes problem with applications to orthogonal and oblique rotation," Ph.D. dissertation, Univ. of Illinois at Urbana-Champaign, 1964.

[35] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Trans. Inform. Theory*, vol. 56, no. 6, pp. 2980–2998, June 2010.

[36] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from noisy entries," in *Advances in Neural Information Processing Systems* 22, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, Eds. Curran Associates, Inc., 2009, pp. 952–960.

[37] N. Duric, P. Littrup, L. Poulo, A. Babkin, R. Pevzner, E. Holsapple, O. Rama, and C. Glide, "Detection of breast cancer with ultrasound tomography: First results with the computed ultrasound risk evaluation (CURE) prototype," *J. Med. Phys*., vol. 34, no. 2, pp. 773–785, 2007.

[38] R. Parhizkar, A. Karbasi, S. Oh, and M. Vetterli, "Calibration using matrix completion with application to ultrasound tomography," *IEEE Trans. Signal Processing*, vol. 61, no. 20, pp. 4923–4933, July 2013.

[39] I. Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and Applications*. New York: Springer, 2005.

[40] J. B. Kruskal, "Nonmetric multidimensional scaling: A numerical method," *Psychometrika*, vol. 29, no. 2, pp. 115–129, 1964.

[41] J. De Leeuw, "Applications of convex analysis to multidimensional scaling," in *Recent Developments in Statistics*, J. Barra, F. Brodeau, G. Romier, and B. V. Cutsem, Eds. Amsterdam: North Holland Publishing Company, 1977, pp. 133–146.

[42] R. Mathar and P. J. F. Groenen, "Algorithms in convex analysis applied to multidimensional scaling," in *Symbolic-Numeric Data Analysis and Learning*, E. Diday and Y. Lechevallier, Eds. Hauppauge, New York: Nova Science Publishers, 1991, pp. 45–56.

[43] L. Guttman, "A general nonmetric technique for finding the smallest coordinate space for a configuration of points," *Psychometrika*, vol. 33, no. 4, pp. 469–506, 1968.

[44] Y. Takane, F. Young, and J. De Leeuw, "Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features," *Psychometrika*, vol. 42, no. 1, pp. 7–67, 1977.

[45] N. Gaffke and R. Mathar, "A cyclic projection algorithm via duality," *Metrika*, vol. 36, no. 1, pp. 29–54, 1989.

[46] R. Parhizkar, "Euclidean distance matrices: Properties, algorithms and applications," Ph.D. dissertation, School of Computer and Communication Sciences, Ecole Polytechnique Federale de Lausanne (EPFL), 2013.

[47] M. Browne, "The young-householder algorithm and the least squares multidimensional scaling of squared distances," *J. Classif*., vol. 4, no. 2, pp. 175–190, 1987.

[48] W. Glunt, T. L. Hayden, and W.-M. Liu, "The embedding problem for predistance matrices," *Bull. Math. Biol*., vol. 53, no. 5, pp. 769–796, 1991.

[49] P. Biswas, T. C. Liang, K. C. Toh, Y. Ye, and T. C. Wang, "Semidefinite programming approaches for sensor network localization with noisy distance measurements," *IEEE Trans. Autom. Sci. Eng*., vol. 3, no. 4, pp. 360–371, 2006.

[50] M. Grant and S. Boyd. (2014, Mar.). CVX: MATLAB software for disciplined convex programming, version 2.1. [Online]. Available: http://cvxr.com/cvx

[51] M. Grant and S. Boyd. (2008). Graph implementations for nonsmooth convex programs. In *Recent Advances in Learning and Control* (ser. Lecture Notes in Control and Information Sciences). V. Blondel, S. Boyd, and H. Kimura, Eds. New York: Springer-Verlag, pp. 95–110. Available: http://stanford.edu/~boyd/graph_dcp.html

[52] M. Boutin and G. Kemper, "On reconstructing N-point configurations from the distribution of distances or areas," *Adv. Appl. Math*., vol. 32, no. 4, pp. 709–735, May 2004.

[53] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer*., vol. 65, no. 4, pp. 943–950, 1979.

[54] R. P. Millane, "Phase retrieval in crystallography and optics," *J. Opt. Soc. Am. A*, vol. 7, no. 3, pp. 394–411, Mar. 1990.

[55] W. Beavers, D. E. Dudgeon, J. W. Beletic, and M. T. Lane, "Speckle imaging through the atmosphere," *Lincoln Lab. J*., vol. 2, no. 2, pp. 207–228, 1989.

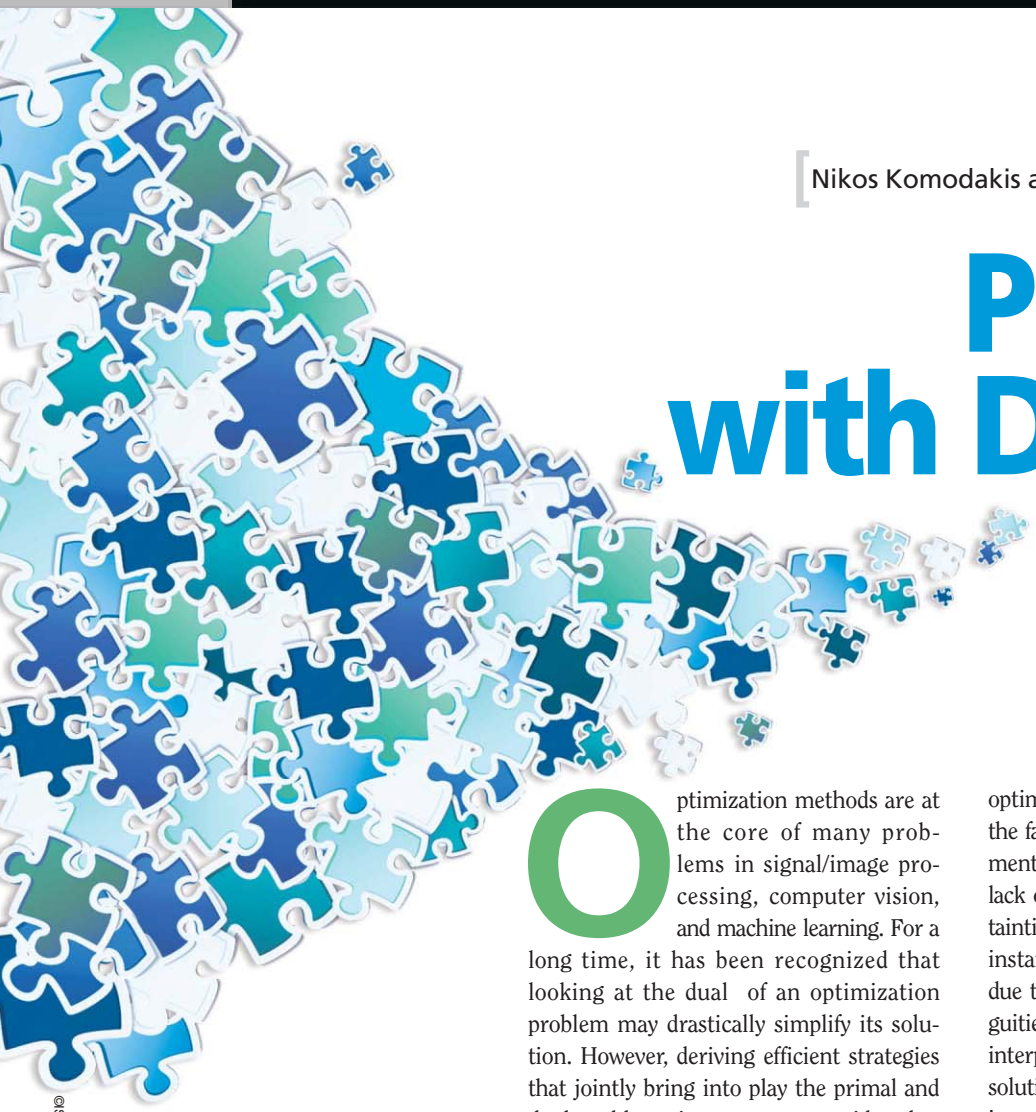
[56] S. S. Skiena, W. D. Smith, and P. Lemke, "Reconstructing sets from interpoint distances," in *Proc. ACM Symposium on Computational Geometry* (*SCG*), 1990, pp. 332–339.

[57] R. Parhizkar, I. Dokmanić, and M. Vetterli, "Single-channel indoor microphone localization," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing* (*ICASSP*), Florence, 2014.

[SP]

Nikos Komodakis and Jean-Christophe Pesquet

Playing with Duality



An overview of recent primal–dual approaches for solving large-scale optimization problems

Optimization methods are at the core of many problems in signal/image processing, computer vision, and machine learning. For a long time, it has been recognized that looking at the dual of an optimization problem may drastically simplify its solution. However, deriving efficient strategies that jointly bring into play the primal and dual problems is a more recent idea that has generated many important new contributions in recent years. These novel developments are grounded in the recent advances in convex analysis, discrete optimization, parallel processing, and nonsmooth optimization with an emphasis on sparsity issues. In this article, we aim to present the principles of primal–dual approaches while providing an overview of the numerical methods that have been proposed in different contexts. Last but not least, primal–dual methods lead to algorithms that are easily parallelizable. Today, such parallel algorithms are becoming increasingly important for efficiently handling high-dimensional problems.

INTRODUCTION

Optimization is an extremely popular paradigm that constitutes the backbone of many branches of applied mathematics and engineering, such as signal processing, computer vision, machine learning, inverse problems, and network communications, to mention just a few [1]. The popularity of optimization approaches often stems from the fact that many problems from the fields mentioned are typically characterized by a lack of closed-form solutions and by uncertainties. In signal and image processing, for instance, uncertainties can be introduced due to noise, sensor imperfection, or ambiguities that are often inherent in the visual interpretation. As a result, perfect or exact solutions hardly exist, whereas one aims at inexact but optimal (in a statistical or application-specific sense) solutions and their efficient computation. At the same time, one important characteristic that today is shared by an increasingly large number of optimization problems encountered in the mentioned areas is the fact that these problems are often very large scale. A good example is the field of computer vision, where one often needs to solve low-level problems that require associating at least one (and typically more than one) variable to each pixel of an image (or even worse of an image sequence as in the case of a video) [2]. This leads to problems that can easily contain millions of variables, which are the norm rather than the exception in this context.

Similarly, in fields such as machine learning [3], [4], because of the great ease with which data can now be collected and stored, quite often one has to cope with truly massive data sets and train very large models, which naturally leads to optimization problems of very high dimensionality [5]. Of course, a similar situation arises in many other scientific domains, including

Digital Object Identifier 10.1109/MSP.2014.2377273
Date of publication: 13 October 2015

application areas such as inverse problems (e.g., medical image reconstruction or satellite image restoration) or telecommunications (e.g., network design and provisioning) and industrial engineering. Because of this fact, computational efficiency constitutes a major issue that needs to be thoroughly addressed. This, therefore, makes mandatory the use of tractable optimization techniques that are able to properly exploit the problem structures and remain applicable to as many problems as possible.

There have been many important advances in recent years concerning a particular class of optimization approaches known as *primal–dual methods*. As their name implies, these approaches proceed by concurrently solving a primal problem (corresponding to the original optimization task) as well as a dual formulation of this problem. As it turns out, in doing so, they are able to more efficiently exploit the problem-specific properties, thus offering in many cases important computational advantages, some of which are briefly mentioned in the following sections for two very broad classes of problems: convex and discrete optimization problems.

CONVEX OPTIMIZATION

Primal–dual methods have primarily been employed in convex optimization problems [6]–[8] where strong duality holds. They have been successfully applied to various types of nonlinear and nonsmooth cost functions that are prevalent in the previously mentioned application fields.

Many such applied problems can essentially be expressed as a minimization of a sum of terms, where each term is given by the composition of a convex function with a linear operator. One first advantage of primal–dual methods pertains to the fact that they can yield very efficient splitting optimization schemes, according to which a solution to the original problem is iteratively computed through solving a sequence of easier subproblems, each one involving only one of the terms appearing in the objective function.

The resulting primal–dual splitting schemes can also handle both differentiable and nondifferentiable terms, the former by the use of gradient operators (i.e., through explicit steps) and the latter by the use of proximity operators (i.e., through implicit steps) [9], [10]. Depending on the target functions, either explicit or implicit steps may be easier to implement. Therefore, the derived optimization schemes exploit the properties of the input problem in a flexible manner, thus leading to very efficient first-order algorithms.

Even more importantly, primal–dual techniques are able to achieve what is known as *full splitting* in the optimization literature, meaning that each of the operators involved in the problem (i.e., not only the gradient or proximity operators but also the involved linear operators) is used separately [11]. As a result, no call to the inversion of a linear operator, which is an expensive operation for large-scale problems, is required during the optimization process. This is an important feature that gives these methods a significant computational advantage over all other splitting-based approaches.

Last but not least, primal–dual methods lead to algorithms that are easily parallelizable. Today, such parallel algorithms are becoming increasingly important for efficiently handling high-dimensional problems.

DISCRETE OPTIMIZATION

In addition to convex optimization, another important area where primal–dual methods play a prominent role is discrete optimization. This is of particular significance given that a large variety of tasks, such as signal processing, computer vision, and pattern recognition, are formulated as discrete labeling problems, where one seeks to optimize some measure related to the quality of the labeling [12]. This includes image analysis tasks such as image segmentation, optical flow estimation, image denoising, and stereo matching, to mention a few. The resulting discrete optimization problems are not only of very large size, but they also typically exhibit highly nonconvex objective functions, which are generally intricate to optimize.

Similarly to the case of convex optimization, primal–dual methods again offer many computational advantages, often leading to very fast graph-cut or message-passing-based algorithms, which are also easily parallelizable, thus providing in many cases a very efficient way to handle discrete optimization problems that are encountered in practice [13]–[16]. Besides being efficient, they are also successful in making small compromises regarding the quality of the estimated solutions. Techniques such as the so-called primal–dual schema are known to provide a principled way for deriving powerful approximation algorithms to solve difficult combinatorial problems, thus allowing primal–dual methods to often exhibit theoretical (i.e., worst-case) approximation properties. Furthermore, apart from the aforementioned worst-case guaranties, primal–dual algorithms also provide for free per-instance approximation guaranties. This is essentially made possible by the fact that these methods are estimating not only primal but also dual solutions.

Convex optimization and discrete optimization rely on different background theories. Convex optimization may appear to be the most tractable topic in optimization for which many efficient algorithms have been developed, allowing a broad class of problems to be solved. By contrast, combinatorial optimization problems are generally NP-hard. However, many convex relaxations of certain discrete problems can provide good approximate solutions to the original ones [17], [18]. The problems encountered in discrete optimization, therefore, constitute a source of inspiration for developing novel convex optimization techniques.

OUR OBJECTIVES

Based on the previous observations, our objectives are

1) provide a thorough introduction that intuitively explains the basic principles and ideas behind primal–dual approaches
2) describe how these methods can be employed both in the context of continuous and discrete optimization
3) explain some of the recent advances that have taken place concerning primal–dual algorithms for solving large-scale optimization problems
4) detail useful connections between primal–dual methods and some widely used optimization techniques such as the alternating direction method of multipliers (ADMM) [19], [20]
5) provide examples of useful applications in the context of image analysis and signal processing.

OPTIMIZATION BACKGROUND

In this section, we introduce the necessary mathematical definitions and concepts used for introducing primal–dual algorithms in later sections. Although the following framework holds for general Hilbert spaces, for simplicity, we will focus on the finite-dimensional case.

NOTATION

In this article, we will consider functions from $\mathbb{R}^N$ to $]-\infty, +\infty]$. The fact that we allow functions to take $+\infty$ value is useful in modern optimization to discard some "forbidden part" of the space when searching for an optimal solution. (For example, in image processing problems, the components of the solution are often intensity values, which must be nonnegative.) The domain of a function $f: \mathbb{R}^N \to]-\infty, +\infty]$ is the subset of $\mathbb{R}^N$ where this function takes finite values, i.e., $\operatorname{dom} f = \{x \in \mathbb{R}^N \mid f(x) < +\infty\}$. A function with a nonempty domain is said to be *proper*. A function f is said to be *convex* if

$$(\forall (x, y) \in (\mathbb{R}^N)^2)(\forall \lambda \in [0,1]) \quad f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y). \tag{1}$$

The class of functions for which most of the main results in convex analysis have been established is $\Gamma_0(\mathbb{R}^N)$, the class of proper, convex, lower-semicontinuous functions from $\mathbb{R}^N$ to $]-\infty, +\infty]$. Recall that a function $f: \mathbb{R}^N \to]-\infty, +\infty]$ is lower semicontinuous if its epigraph $\operatorname{epi} f = \{(x, \zeta) \in \operatorname{dom} f \times \mathbb{R} \mid f(x) \leq \zeta\}$ is a closed set (see Figure 1).

If C is a nonempty subset of $\mathbb{R}^N$, the indicator function of C is defined as

$$(\forall x \in \mathbb{R}^N) \quad \iota_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases} \tag{2}$$

This function belongs to $\Gamma_0(\mathbb{R}^N)$ if and only if C is a nonempty closed convex set.

The Moreau subdifferential of a function $f: \mathbb{R}^N \to]-\infty, +\infty]$ at $x \in \mathbb{R}^N$ is defined as

$$\partial f(x) = \{u \in \mathbb{R}^N \mid (\forall y \in \mathbb{R}^N)\ f(y) \geq f(x) + u^\top (y - x)\}. \tag{3}$$

Any vector u in $\partial f(x)$ is called a *subgradient* of f at x (see Figure 2).

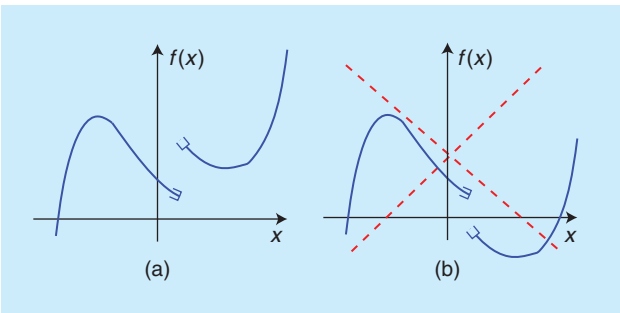
Fermat's rule states that zero is a subgradient of f at x if and only if x belongs to the set of global minimizers of f. If f is a proper convex function that is differentiable at x, then its subdifferential at x reduces to the singleton consisting of its gradient, i.e., $\partial f(x) = \{\nabla f(x)\}$. Note that, in the nonconvex case, extended definitions of the subdifferential, such as the limiting subdifferential [21], may be useful, but this one reduces to the Moreau subdifferential when the function is convex.

PROXIMITY OPERATOR

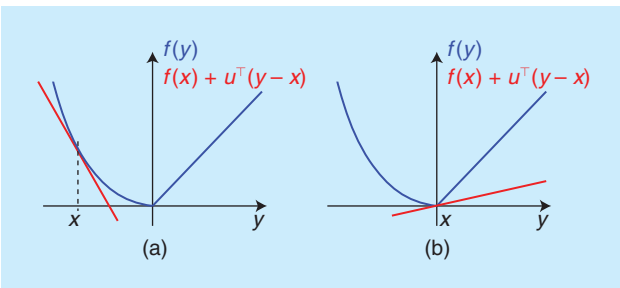
A concept that has been of growing importance in recent developments in optimization is the concept of proximity operator. It must be pointed out that the proximity operator was introduced in the early work by Moreau [9]. The proximity operator of a function $f \in \Gamma_0(\mathbb{R}^N)$ is defined as

$$\operatorname{prox}_f: \mathbb{R}^N \to \mathbb{R}^N: x \mapsto \operatorname*{argmin}_{y \in \mathbb{R}^N} f(y) + \frac{1}{2}\|y - x\|^2, \tag{4}$$

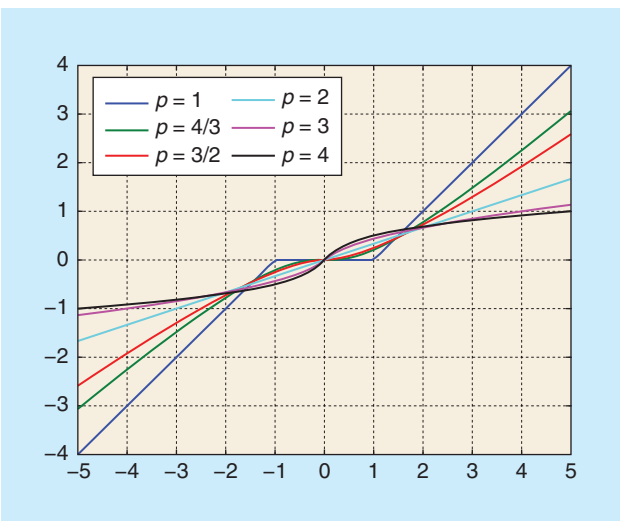
where $\|\cdot\|$ denotes the Euclidean norm. For every $x \in \mathbb{R}^N$, $\operatorname{prox}_f x$ can thus be interpreted as the result of a regularized minimization of f in the neighborhood of x. Note that the minimization to be performed to calculate $\operatorname{prox}_f x$ always has a unique solution. Figure 3 shows the variations of the prox_f function when $f: \mathbb{R} \to \mathbb{R}: x \mapsto |x|^p$ with $p \geq 1$. In the case when $p = 1$, the classical soft-thesholding operation is obtained.



[FIG1] An illustration of the lower-semicontinuity property.



[FIG2] Examples of subgradients u of a function f at x.



[FIG3] Graph of $\operatorname{prox}_{|\cdot|^p}$. This power p function is often used to regularize inverse problems.

In the case when f is equal to the indicator function of a nonempty closed convex set $C \subset \mathbb{R}^N$, the proximity operator of f reduces to the projection P_C onto this set, i.e., $(\forall x \in \mathbb{R}^N)$ $P_C x = \operatorname{argmin}_{y \in C} \|y - x\|$. This shows that proximity operators can be viewed as extensions of projections onto convex sets. The proximity operator enjoys many properties of the projection, particularly its nonexpansiveness. The firm nonexpansiveness can be viewed as a generalization of the strict contraction property, which is the engine behind the Banach–Picard fixed-point theorem. This property makes the proximity operator successful in ensuring the convergence of fixed-point algorithms grounded on its use. For more details about proximity operators and their rich properties, see the tutorial papers in [5], [10], and [22]. The definition of the proximity operator can be extended to nonconvex lower-semicontinuous functions, which are lower bounded by an affine function, but $\operatorname{prox}_f x$ is no longer guaranteed to be uniquely defined at any given point x.

CONJUGATE FUNCTION

A fundamental notion when dealing with duality issues is the notion of conjugate function. The conjugate of a function $f: \mathbb{R}^N \to]-\infty, +\infty]$ is the function f^* defined as

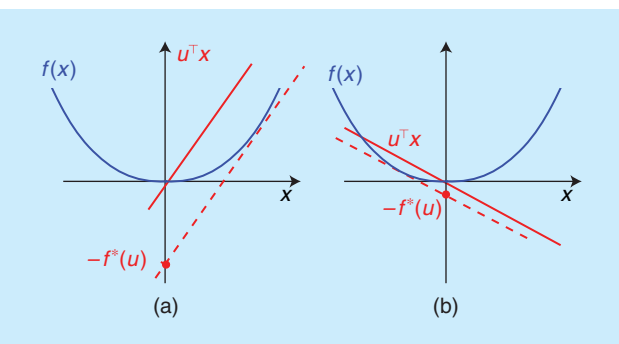
$$f^*: \mathbb{R}^N \to]-\infty, +\infty]: u \mapsto \sup_{x \in \mathbb{R}^N} (x^\top u - f(x)). \tag{5}$$

This concept was introduced by Legendre in the one-variable case, and it was generalized by Fenchel. A graphical illustration of the conjugate function is provided in Figure 4. In particular, for every vector $x \in \mathbb{R}^N$ such that the supremum in (5) is attained, u is a subgradient of f at x.

It must be emphasized that, even if f is nonconvex, f^* is a (not necessarily proper) lower-semicontinuous convex function. In addition, when $f \in \Gamma_0(\mathbb{R}^N)$, then $f^* \in \Gamma_0(\mathbb{R}^N)$, and also the biconjugate of f (that is, the conjugate of its conjugate) is equal to f. This means that we can express any function f in $\Gamma_0(\mathbb{R}^N)$ as

$$(\forall x \in \mathbb{R}^N) \qquad f(x) = \sup_{u \in \mathbb{R}^N} (u^\top x - f^*(u)). \tag{6}$$

A geometrical interpretation of this result is that the epigraph of any proper lower-semicontinuous convex function is always an intersection of closed half spaces.



[FIG4] A graphical interpretation of the conjugate function.

As we have seen, the subdifferential plays an important role in the characterization of the minimizers of a function. A natural question, therefore, regards the relations existing between the subdifferential of a function $f: \mathbb{R}^N \to]-\infty, +\infty]$ and the subdifferential of its conjugate function. An answer is provided by the following important properties:

$$\begin{aligned} u \in \partial f(x) &\Rightarrow x \in \partial f^*(u) \quad \text{if } f \text{ is proper} \\ u \in \partial f(x) &\Leftrightarrow x \in \partial f^*(u) \quad \text{if } f \in \Gamma_0(\mathbb{R}^N). \end{aligned} \tag{7}$$

Another important property is Moreau's decomposition formula, which links the proximity operator of a function $f \in \Gamma_0(\mathbb{R}^N)$ to the proximity operator of its conjugate

$$\begin{aligned} &(\forall x \in \mathbb{R}^N)(\forall \gamma \in]0, +\infty[) \\ &\quad x = \operatorname{prox}_{\gamma f} x + \gamma \operatorname{prox}_{\gamma^{-1} f^*}(\gamma^{-1} x). \end{aligned} \tag{8}$$

Other useful properties of the conjugation operation are listed in Table 1 (throughout the article, int S denotes the interior of a set S), where a parallel is drawn with the multidimensional Fourier transform, which is a more familiar tool in signal and image processing. Conjugation also makes it possible to build an insightful bridge between the main two kinds of nonsmooth convex functions encountered in signal and image processing problems, which are indicator functions of feasibility constraints and sparsity measures (see "Conjugates of Support Functions").

DUALITY RESULTS

A wide array of problems in signal and image processing can be expressed under the following variational form:

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \; f(x) + g(Lx), \tag{9}$$

where $f: \mathbb{R}^N \to]-\infty, +\infty]$, $g: \mathbb{R}^K \to]-\infty, +\infty]$, and $L \in \mathbb{R}^{K \times N}$. Equation (9) is usually referred to as the *primal problem*, which is associated with the following *dual problem* [6], [8], [26]:

$$\underset{v \in \mathbb{R}^K}{\text{minimize}} \; f^*(-L^\top v) + g^*(v). \tag{10}$$

(see "Consensus and Sharing Are Dual Problems").

The latter problem may be easier to solve than the former one, especially when K is much smaller than N.

A question, however, is whether solving the dual problem may provide information for the solution of the primal one. The Fenchel–Rockafellar duality theorem first answers this question by basically stating that solving the dual problem provides a lower bound on the minimum value that can be obtained in the primal one. More precisely, if f and g are proper functions and if μ and μ^* denote the infima of the functions minimized in the primal and dual problems, respectively, then *weak duality* holds, which means that $\mu \geq -\mu^*$. If μ is finite, $\mu + \mu^*$ is called the *duality gap*. In addition, if $f \in \Gamma_0(\mathbb{R}^N)$ and $g \in \Gamma_0(\mathbb{R}^K)$, then, under appropriate qualification conditions (for example, this property is satisfied if the intersection of the interior of the domain of g and the image of the domain of f by L is nonempty), there always exists a solution to the dual

[TABLE 1] THE PARALLELISM BETWEEN THE PROPERTIES OF THE LEGENDRE–FENCHEL CONJUGATION [10] AND OF THE FOURIER TRANSFORM.

	CONJUGATION		FOURIER TRANSFORM	
PROPERTY	$h(x)$	$h^*(u)$	$h(x)$	$\hat{h}(\nu)$
I INVARIANT FUNCTION	$(1/2)\|\|x\|\|^2$	$(1/2)\|\|u\|\|^2$	$\exp(-\pi\|\|x\|\|^2)$	$\exp(-\pi\|\|\nu\|\|^2)$
II TRANSLATION $c \in \mathbb{R}^N$	$f(x-c)$	$f^*(u)+c^\top u$	$f(x-c)$	$\exp(-\jmath 2\pi c^\top x)\hat{f}(\nu)$
III DUAL TRANSLATION $c \in \mathbb{R}^N$	$f(x)+c^\top x$	$f^*(u-c)$	$\exp(\jmath 2\pi c^\top x)f(x-c)$	$\hat{f}(\nu-c)$
IV SCALAR MULTIPLICATION $\alpha \in]0,+\infty[$	$\alpha f(x)$	$\alpha f^*\left(\frac{u}{\alpha}\right)$	$\alpha f(x)$	$\alpha \hat{f}(\nu)$
V INVERTIBLE LINEAR TRANSFORM $L \in \mathbb{R}^{N\times N}$	$f(Lx)$	$f^*((L^{-1})^\top u)$	$f(Lx)$	$\frac{1}{\lvert\det(L)\rvert}\hat{f}((L^{-1})^\top \nu)$
VI SCALING $\alpha \in \mathbb{R}^*$	$f\left(\frac{x}{\alpha}\right)$	$f^*(\alpha u)$	$f\left(\frac{x}{\alpha}\right)$	$\lvert\alpha\rvert\hat{f}(\alpha\nu)$
VII REFLECTION	$f(-x)$	$f^*(-u)$	$f(-x)$	$\hat{f}(-\nu)$
VIII SEPARABILITY	$\sum_{j=1}^N \varphi_j(x^{(j)}),\ x=(x^{(j)})_{1\le j\le N}$	$\sum_{j=1}^N \varphi_j^*(u^{(j)}),\ u=(u^{(j)})_{1\le j\le N}$	$\prod_{j=1}^N \varphi_j(x^{(j)}),\ x=(x^{(j)})_{1\le j\le N}$	$\prod_{j=1}^N \hat{\varphi}_j(\nu^{(j)}),\ \nu=(\nu^{(j)})_{1\le j\le N}$
IX ISOTROPY	$\psi(\|\|x\|\|)$	$\psi^*(\|\|u\|\|)$	$\psi(\|\|x\|\|)$	$\tilde{\psi}(\|\|\nu\|\|)$
X INF-CONVOLUTION/ CONVOLUTION	$(f \,\square\, g)(x)=\inf_{y\in\mathbb{R}^N} f(y)+g(x-y)$	$f^*(u)+g^*(u)$	$(f\star g)(x)=\int_{\mathbb{R}^N} f(y)g(x-y)dy$	$\hat{f}(\nu)\hat{g}(\nu)$
XI SUM/PRODUCT	$f(x)+g(x),\ f\in\Gamma_0(\mathbb{R}^N),\ g\in\Gamma_0(\mathbb{R}^N)$ $\operatorname{dom} f\cap\operatorname{int}(\operatorname{dom} g)\neq\varnothing$	$(f^* \,\square\, g^*)(u)$	$f(x)g(x)$	$(\hat{f}\star\hat{g})(\nu)$
XIII IDENTITY ELEMENT OF CONVOLUTION	$\iota_{\{0\}}(x)$	0	$\delta(x)$	1
XIV IDENTITY ELEMENT OF ADDITION/PRODUCT	0	$\iota_{\{0\}}(u)$	1	$\delta(\nu)$
XIV OFFSET $\alpha \in \mathbb{R}$	$f(x)+\alpha$	$f^*(u)-\alpha$	$f(x)+\alpha$	$\hat{f}(\nu)+\alpha\delta(\nu)$
XV INFINUM/SUM	$\inf_{1\le m\le M} f_m(x)$	$\sup_{1\le m\le M} f_m^*(u)$	$\sum_{m=1}^M f_m(x)$	$\sum_{m=1}^M \hat{f}_m(\nu)$
XVI VALUE AT ZERO	$f^*(0)=-\inf f$		$\hat{f}(0)=\int_{\mathbb{R}^N} f(x)dx$	

f is a function defined on $\mathbb{R}^N$, f^* denotes its conjugate, and $\hat{f}$ is its Fourier transform such that $\hat{f}(\nu)=\int_{\mathbb{R}^N} f(x)\exp(-\jmath 2\pi x^\top \nu)dx$ where $\nu\in\mathbb{R}^N$ and $\jmath$ is the imaginary unit (a similar notation is used for other functions); h, g, and $(f_m)_{1\le m\le M}$ are functions defined on $\mathbb{R}^N$; $(\varphi_j)_{1\le j\le N}$ are functions defined on $\mathbb{R}$; ψ is an even function defined on $\mathbb{R}$; $\tilde{\psi}$ is defined as $\tilde{\psi}(\rho)=2\pi\rho^{(2-N)/2}\int_0^{+\infty} r^{N/2}J_{(N-2)/2}(2\pi r\rho)\psi(r)dr$, where $\rho\in\mathbb{R}$ and $J_{(N-2)/2}$ is the Bessel function of order $(N-2)/2$; and δ denotes the Dirac distribution. (Some properties of the Fourier transform may require some technical assumptions.)

CONJUGATES OF SUPPORT FUNCTIONS

The support function of a set $C\subset\mathbb{R}^N$ is defined as

$$(\forall u\in\mathbb{R}^N)\qquad \sigma_C(u)=\sup_{x\in C} x^\top u. \tag{S1}$$

In fact, a function f is the support function of a nonempty closed convex set C if and only if it belongs to $\Gamma_0(\mathbb{R}^N)$ and it is positively homogeneous [8], i.e.,

$$(\forall x\in\mathbb{R}^N)(\forall \alpha\in]0,+\infty[)\qquad f(\alpha x)=\alpha f(x).$$

Examples of such functions are norms, e.g., the ℓ_1-norm

$$(\forall x=(x^{(j)})_{1\le j\le N}\in\mathbb{R}^N)\qquad f(x)=\|x\|_1=\sum_{j=1}^N |x^{(j)}|,$$

which is a useful convex sparsity-promoting measure in Lasso estimation [23] and compressive sensing [24]. Another famous example is the total variation seminorm [25], which is popular in image processing for retrieving constant areas with sharp contours. An important property is that, if C is a nonempty closed convex set, the conjugate of its support function is the indicator function of C. For example, the conjugate function of the ℓ_1-norm is the indicator function of the hypercube $[-1,1]^N$. This shows that using sparsity measures is equivalent in the dual domain to imposing some constraints.

problem and the duality gap vanishes. When the duality gap is equal to zero, it is said that strong duality holds.

Another useful result follows from the fact that, by using the definition of the conjugate function of g, (9) can be re-expressed as the following saddle-point problem:

$$\text{Find}\quad \inf_{x\in\mathbb{R}^N}\sup_{v\in\mathbb{R}^K}\left(f(x)+v^\top Lx-g^*(v)\right). \tag{11}$$

To find a saddle point $(\hat{x},\hat{v})\in\mathbb{R}^N\times\mathbb{R}^K$, it thus appears natural to impose the inclusion relations

$$-L^\top\hat{v}\in\partial f(\hat{x}),\qquad L\hat{x}\in\partial g^*(\hat{v}). \tag{12}$$

A pair $(\hat{x},\hat{v})$ satisfying these conditions is called a *Kuhn–Tucker point*. Actually, under a technical assumption, by using Fermat's rule and (7), it can be proved that, if $(\hat{x},\hat{v})$ is a Kuhn–Tucker point, $\hat{x}$ is a solution to the primal problem and $\hat{v}$ is a

CONSENSUS AND SHARING ARE DUAL PROBLEMS

Suppose that our objective is to minimize a composite function $\sum_{m=1}^{M} g_m$, where the potential $g_m: \mathbb{R}^N \to]-\infty, +\infty]$ is computed at the vertex of index $m \in \{1, \ldots, M\}$ of a graph. A classical technique to perform this task in a distributed or parallel manner [20] consists of reformulating this problem as a consensus problem, where a variable is assigned to each vertex and the defined variables $x_1, \ldots, x_M$ are updated so as to reach a consensus: $x_1 = \ldots = x_M$. This means that, in the product space $(\mathbb{R}^N)^M$, the original optimization problem can be rewritten as

$$\underset{\mathbf{x}=(x_1,\ldots,x_m)\in(\mathbb{R}^N)^M}{\text{minimize}} \ \iota_D(\mathbf{x}) + \underbrace{\sum_{m=1}^{M} g_m(x_m)}_{g(\mathbf{x})},$$

where D is the vector space defined as $D = \{\boldsymbol{x} = (x_1, \ldots, x_M) \in (\mathbb{R}^N)^M \mid x_1 = \ldots = x_M\}$.

By noticing that the conjugate of the indicator function of a vector space is the indicator function of its orthogonal complement, it is easy to see that the dual of this consensus problem has the following form:

$$\underset{\boldsymbol{\upsilon}=(\upsilon_1,\ldots,\upsilon_M)\in(\mathbb{R}^N)^M}{\text{minimize}} \ \iota_{D^\perp}(\boldsymbol{\upsilon}) + \underbrace{\sum_{m=1}^{M} g_m^*(\upsilon_m)}_{g^*(\boldsymbol{\upsilon})},$$

where $D^\perp = \{\boldsymbol{\upsilon} = (\upsilon_1, \ldots, \upsilon_M) \in (\mathbb{R}^N)^M \mid \upsilon_1 + \cdots + \upsilon_M = 0\}$ is the orthogonal complement of D. By making the variable change $(\forall m \in \{1, \ldots, M\})\ \upsilon_m = u_m - u/M$, where u is some given vector in $\mathbb{R}^N$, and by setting $h_m(u_m) = -g_m^*(u_m - u/M)$, the latter minimization can be re-expressed as

$$\underset{\substack{u_1\in\mathbb{R}^N,\ldots,u_M\in\mathbb{R}^N \\ u_1+\cdots+u_M=u}}{\text{maximize}} \ \sum_{m=1}^{M} h_m(u_m).$$

This problem is known as a *sharing problem,* where one wants to allocate a given resource u between M agents while maximizing the sum of their welfares evaluated through their individual utility functions $(h_m)_{1\le m\le M}$.

solution to the dual one. This property especially holds when $f \in \Gamma_0(\mathbb{R}^N)$ and $g \in \Gamma_0(\mathbb{R}^K)$.

DUALITY IN LINEAR PROGRAMMING

In linear programming (LP) [27], we are interested in convex optimization problems of the form

$$\text{Primal-LP:} \quad \underset{x\in[0,+\infty[^N}{\text{minimize}} \ c^\top x \ \text{ s.t. } \ Lx \ge b, \tag{13}$$

where $L = (L^{(i,j)})_{1\le i\le K, 1\le j\le N} \in \mathbb{R}^{K\times N}$, $b = (b^{(i)})_{1\le i\le K} \in \mathbb{R}^K$, and $c = (c^{(j)})_{1\le j\le N} \in \mathbb{R}^N$. The vector inequality in (13) means that $Lx - b \in [0, +\infty[^K$. This formulation can be viewed as a special case of (9) where

$$\begin{aligned} &(\forall x \in \mathbb{R}^N) \quad f(x) = c^\top x + \iota_{[0,+\infty[^N}(x), \\ &(\forall z \in \mathbb{R}^K) \quad g(z) = \iota_{[0,+\infty[^K}(z - b). \end{aligned} \tag{14}$$

By using the properties of the conjugate function and by setting $y = -\upsilon$, it is readily shown that the dual equation (10) can be re-expressed as

$$\text{Dual-LP:} \quad \underset{y\in[0,+\infty[^K}{\text{maximize}} \ b^\top y \ \text{ s.t. } \ L^\top y \le c. \tag{15}$$

Since f is a convex function, strong duality holds in LP. If $\hat{x} = (\hat{x}^{(j)})_{1\le j\le N}$ is a solution to Primal-LP, a solution $\hat{y} = (\hat{y}^{(i)})_{1\le i\le K}$ to Dual-LP can be obtained by the primal complementary slackness condition

$$(\forall j \in \{1, \ldots, N\}) \ \text{ such that } \ \hat{x}^{(j)} > 0, \quad \sum_{i=1}^{K} L^{(i,j)} \hat{y}^{(i)} = c^{(j)}, \tag{16}$$

whereas, if $\hat{y}$ is a solution to Dual-LP, a solution $\hat{x}$ to Primal-LP can be obtained by the dual complementary slackness condition

$$(\forall i \in \{1, \ldots, K\}) \ \text{ such that } \ \hat{y}^{(i)} > 0, \quad \sum_{j=1}^{N} L^{(i,j)} \hat{x}^{(j)} = b^{(i)}. \tag{17}$$

CONVEX OPTIMIZATION ALGORITHMS

In this section, we present several primal–dual splitting methods for solving convex optimization problems, starting from the basic to the more sophisticated highly parallelized forms.

PROBLEM

A wide range of convex optimization problems can be formulated as follows:

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \ f(x) + g(Lx) + h(x), \tag{18}$$

where $f \in \Gamma_0(\mathbb{R}^N)$, $g \in \Gamma_0(\mathbb{R}^K)$, $L \in \mathbb{R}^{K\times N}$, and $h \in \Gamma_0(\mathbb{R}^N)$ is a differentiable function having a Lipschitzian gradient with a Lipschitz constant $\beta \in]0, +\infty[$. The latter assumption means that the gradient ∇h of h is such that

$$(\forall (x, y) \in (\mathbb{R}^N)^2) \quad \| \nabla h(x) - \nabla h(y) \| \le \beta \| x - y \|. \tag{19}$$

For example, the functions $f, g \circ L$, and h may model various data fidelity terms and regularization functions encountered in the solution of inverse problems. In particular, the Lipschitz differentiability property is satisfied for the least-squares criteria.

With respect to (9), we have introduced an additional smooth term h. This may be useful in offering more flexibility for taking into account the structure of the problem of interest and the properties of the involved objective function. We will, however, see that not all algorithms are able to take advantage of the fact that h is a smooth term.

Based on the results in the "Duality Results" section and property (XI) in Table 1, the dual optimization problem reads

$$\underset{\upsilon\in\mathbb{R}^K}{\text{minimize}} \ (f^* \,\square\, h^*)(-L^\top \upsilon) + g^*(\upsilon). \tag{20}$$

Note that, in the particular case when $h = 0$, the infconvolution $f^* \,\square\, h^*$ [see the definition of property (X) in

Table 1] of the conjugate functions of f and h reduces to f^* and we recover the basic form (10) of the dual problem.

The common trick used in the algorithms, which will be presented in this section, is to jointly solve (18) and (20) instead of focusing exclusively on either (18) or (20). More precisely, these algorithms aim at finding a Kuhn–Tucker point $(\hat{x}, \hat{v}) \in \mathbb{R}^N \times \mathbb{R}^K$ such that

$$-L^\top \hat{v} - \nabla h(\hat{x}) \in \partial f(\hat{x}) \quad \text{and} \quad L\hat{x} \in \partial g^*(\hat{v}). \tag{21}$$

It has to be mentioned that some specific forms of (18) (e.g., when $g = 0$) can be solved in a quite efficient manner by simpler proximal algorithms (see [10]) than those described in the following.

ADMM

The celebrated ADMM can be viewed as a primal–dual algorithm. This algorithm belongs to the class of augmented Lagrangian methods since a possible way of deriving this algorithm consists of looking for a saddle point of an augmented version of the classical Lagrange function [20]. This augmented Lagrangian is defined as

$$(\forall (x, y, z) \in \mathbb{R}^N \times (\mathbb{R}^K)^2)$$
$$\tilde{\mathcal{L}}(x, y, z) = f(x) + h(x) + g(y) + \gamma z^\top (Lx - y) + \frac{\gamma}{2}\| Lx - y \|^2, \tag{22}$$

where $\gamma \in]0, +\infty[$ and γz corresponds to a Lagrange multiplier. ADMM simply splits the step of minimizing the augmented Lagrangian with respect to (x, y) by alternating between the two variables, while a gradient ascent is performed with respect to the variable z. The resulting iterations are given in Algorithm 1.

Algorithm 1: ADMM.

Set $y_0 \in \mathbb{R}^K$ and $z_0 \in \mathbb{R}^K$
Set $\gamma \in]0, +\infty[$
For $n = 0, 1, \ldots$
$$x_n = \operatorname*{argmin}_{x \in \mathbb{R}^N} \frac{1}{2}\| Lx - y_n + z_n \|^2 + \frac{1}{\gamma}(f(x) + h(x))$$
$$s_n = Lx_n$$
$$y_{n+1} = \operatorname{prox}_{\frac{g}{\gamma}}(z_n + s_n)$$
$$z_{n+1} = z_n + s_n - y_{n+1}.$$

This algorithm has been known for a long time [19], [28], although it has recently attracted much interest in the signal and image processing communities (see, e.g., [29]–[34]). A condition for the convergence of ADMM is shown in "Convergence of ADMM."

A convergence rate analysis is conducted in [35]. It must be emphasized that ADMM is equivalent to the application of the Douglas–Rachford algorithm [36], [37], another famous algorithm in convex optimization, to the dual problem. Other primal–dual algorithms can be deduced from the Douglas–Rachford iteration [38] or an augmented Lagrangian approach [39].

Although ADMM was observed to have a good numerical performance in many problems, its applicability may be limited by the computation of x_n at each iteration $n \in \mathbb{N}$, which may be intricate due to the presence of matrix L, especially when this matrix is high dimensional and has no simple structure. In addition, functions f and h are not dealt with separately, and so the smoothness of h is not exploited here in an explicit manner.

CONVERGENCE OF ADMM

Under the assumptions that

- rank $(L) = N$
- equation (18) admits a solution
- $\operatorname{int}(\operatorname{dom} g) \cap L(\operatorname{dom} f) \neq \varnothing$ or $\operatorname{dom} g \cap \operatorname{int}(L(\operatorname{dom} f)) \neq \varnothing$, (more general qualification conditions involving the relative interiors of the domain of g and $L(\operatorname{dom} f)$ can be obtained [10]),

$(x_n)_{n \in \mathbb{N}}$ converges to a solution to the primal problem (18) and $(\gamma z_n)_{n \in \mathbb{N}}$ converges to a solution to the dual problem (20).

METHODS BASED ON A FORWARD–BACKWARD APPROACH

The methods presented here are based on a forward–backward approach [40]: they combine a gradient descent step (forward step) with a computation step involving a proximity operator. The latter computation corresponds to a kind of subgradient step performed in an implicit (or backward) manner [10]. A deeper justification of this terminology is provided by the theory of monotone operators [8], which allows us to highlight the fact that a pair $(\hat{x}, \hat{v}) \in \mathbb{R}^N \times \mathbb{R}^K$ satisfying (21) is a zero of a sum of two maximally monotone operators. We will not go into detail, which can become rather technical, but we can mention that the algorithms presented in this section can then be viewed as offspring of the forward–backward algorithm for finding such a zero [8]. Like ADMM, this algorithm is an instantiation of a recursion converging to a fixed point of a nonexpansive mapping.

One of the most popular primal–dual methods within this class is provided in Algorithm 2. In the case when $h = 0$, this algorithm can be viewed as an extension of the Arrow–Hurwitz method, which performs alternating subgradient steps with respect to the primal and dual variables to solve the saddle-point problem (11) [41]. Two step sizes τ and σ and relaxation factors $(\lambda_n)_{n \in \mathbb{N}}$ are involved in Algorithm 2, which can be adjusted by the user so as to get the best convergence profile for a given application.

Note that when $L = 0$ and $g = 0$, the basic form of the forward–backward algorithm (also called the *proximal gradient algorithm*) is recovered, a popular example of which is the iterative soft-thresholdi ng algorithm [42].

Algorithm 2: FB-based primal–dual algorithm.

Set $x_0 \in \mathbb{R}^N$ and $v_0 \in \mathbb{R}^K$
Set $(\tau, \sigma) \in]0, +\infty[^2$
For $n = 0, 1, \ldots$
$$p_n = \operatorname{prox}_{\tau f}(x_n - \tau(\nabla h(x_n) + L^\top v_n))$$
$$q_n = \operatorname{prox}_{\sigma g^*}(v_n + \sigma L(2p_n - x_n))$$
Set $\lambda_n \in]0, +\infty[$
$$(x_{n+1}, v_{n+1}) = (x_n, v_n) + \lambda_n((p_n, q_n) - (x_n, v_n)).$$

A rescaled variant of the primal–dual method (see Algorithm 3) is sometimes preferred, which can be deduced from the previous algorithm by using Moreau's decomposition (8) and by making the variable changes $q'_n \equiv q_n/\sigma$ and $v'_n \equiv v_n/\sigma$. Under this form, it can be seen that, when $N = K$, $L = \mathrm{Id}$, $h = 0$, and $\tau\sigma = 1$, the algorithm reduces to the Douglas–Rachford algorithm (see [43] for the link existing with extensions of the Douglas–Rachford algorithm).

Algorithm 3: The rescaled variant of Algorithm 2.

Set $x_0 \in \mathbb{R}^N$ and $v'_0 \in \mathbb{R}^K$
Set $(\tau, \sigma) \in]0, +\infty[^2$
For $n = 0, 1, \ldots$
$p_n = \mathrm{prox}_{\tau f}(x_n - \tau(\nabla h(x_n) + \sigma L^{\mathrm{T}} v'_n))$
$q_n = (\mathrm{Id} - \mathrm{prox}_{g/\sigma})(v'_n + L(2p_n - x_n))$
Set $\lambda_n \in]0, +\infty[$
$(x_{n+1}, v'_{n+1}) = (x_n, v'_n) + \lambda_n((p_n, q'_n) - (x_n, v'_n))$.

Also, by using the symmetry existing between the primal and dual problems, another variant of Algorithm 2 can be obtained (see Algorithm 4), which is often encountered in the literature. When $L^{\mathrm{T}}L = \mu\mathrm{Id}$ with $\mu \in]0, +\infty[$, $h = 0$, $\tau\sigma\mu = 1$, and $\lambda_n \equiv 1$, Algorithm 4 reduces to ADMM by setting $\gamma = \sigma$, and $z_n \equiv v_n/\sigma$ in Algorithm 1.

Algorithm 4: The symmetric form of Algorithm 2.

Set $x_0 \in \mathbb{R}^N$ and $v_0 \in \mathbb{R}^K$
Set $(\tau, \sigma) \in]0, +\infty[^2$
For $n = 0, 1, \ldots$
$q_n = \mathrm{prox}_{\sigma g^*}(v_n + \sigma L x_n))$
$p_n = \mathrm{prox}_{\tau f}(x_n - \tau(\nabla h(x_n) + L^{\top}(2q_n - v_n)))$
Set $\lambda_n \in]0, +\infty[$
$(x_{n+1}, v_{n+1}) = (x_n, v_n) + \lambda_n((p_n, q_n) - (x_n, v_n))$.

Convergence guarantees were established in [44] and for a more general version of this algorithm in [45] (see "Convergence of Algorithms 2 and 4").

CONVERGENCE OF ALGORITHMS 2 AND 4
Under the following sufficient conditions:
- $\tau^{-1} - \sigma\|L\|_{\mathrm{S}}^2 \geq \beta/2$, where $\|L\|_{\mathrm{S}}$ is the spectral norm of L
- $(\lambda_n)_{n\in\mathbb{N}}$ is a sequence in $]0, \delta[$ such that $\sum_{n\in\mathbb{N}} \lambda_n \times (\delta - \lambda_n) = +\infty$ where $\delta = 2 - \beta(\tau^{-1} - \sigma\|L\|_{\mathrm{S}}^2)^{-1}/2 \in [1, 2[$
- equation (18) admits a solution
- $\mathrm{int}(\mathrm{dom} g) \cap L(\mathrm{dom} f) \neq \varnothing$ or $\mathrm{dom} g \cap \mathrm{int}(L(\mathrm{dom} f)) \neq \varnothing$,

the sequences $(x_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ are such that the former converges to a solution to the primal problem (18) and the latter converges to a solution to the dual problem (20).

Algorithm 2 also constitutes a generalization of [46]–[48] [designated by some authors as the *primal–dual hybrid gradient* (*PDHG*) algorithm]. Preconditioned or adaptive versions of this algorithm were proposed in [49]–[52], which may accelerate its convergence. Convergence rate results were also recently derived in [53].

Another primal–dual method (see Algorithm 5) was proposed in [54] and [55], which also results from a forward–backward approach [52]. This algorithm is restricted to the case when $f = 0$ in (18) (see "Convergence of Algorithm 5").

Algorithm 5: FBF-based primal–dual algorithm.

Set $x_0 \in \mathbb{R}^N$ and $v_0 \in \mathbb{R}^K$
Set $(\tau, \sigma) \in]0, +\infty[^2$
For $n = 0, 1, \ldots$
$s_n = x_n - \tau\nabla h(x_n)$
$y_n = s_n - \tau L^{\mathrm{T}} v_n$
$q_n = \mathrm{prox}_{\sigma g^*}(v_n + \sigma L y_n)$
$p_n = s_n - \tau L^{\top} q_n$
set $\lambda_n \in]0, +\infty[$
$(x_{n+1}, v_{n+1}) = (x_n, v_n) + \lambda_n((p_n, q_n) - (x_n, v_n))$.

As shown by the next convergence result, the conditions on the step sizes τ and σ are less restrictive than for Algorithm 2.

CONVERGENCE OF ALGORITHM 5
Under the assumptions that
- $\tau\sigma\|L\|_{\mathrm{S}}^2 < 1$ and $\tau < 2/\beta$
- $(\lambda_n)_{n\in\mathbb{N}}$ is a sequence in $]0, 1]$ such that $\inf_{n\in\mathbb{N}} \lambda_n > 0$
- equation (18) admits a solution
- $\mathrm{int}(\mathrm{dom} g) \cap \mathrm{ran}(L) \neq \varnothing$,

the sequence $(x_n)_{n\in\mathbb{N}}$ converges to a solution to the primal problem (18) (where $f = 0$) and $(v_n)_{n\in\mathbb{N}}$ converges to a solution to the dual problem (20).

Note also that the dual forward–backward approach that was proposed in [56] for solving (18) in the specific case when $h = \|\cdot - r\|^2/2$ with $r \in \mathbb{R}^N$ belongs to the class of primal–dual forward–backward approaches.

It must be emphasized that Algorithms 2–5 present two interesting features that are very useful in practice. At first, they allow us to deal with the functions involved in the optimization problem at hand either through their proximity operator or through their gradient. Indeed, for some functions, especially nondifferentiable or nonfinite ones, the proximity operator can be a very powerful tool [57], but, for some smooth functions (e.g., the Poisson–Gauss neg-log-likelihood [58]), the gradient may be easier to handle. Second, these algorithms do not require us to invert any matrix but only to apply L and its adjoint. This advantage is of main interest when large-size problems have to be solved for which the inverse of L (or $L^{\top}L$) does not exist or has a no tractable expression.

METHODS BASED ON A FORWARD–BACKWARD–FORWARD APPROACH

Primal–dual methods based on a forward–backward–forward (FBF) approach were among the first primal–dual proximal methods proposed in the optimization literature, inspired from

the seminal work in [59]. They were first developed in the case when $h = 0$ [60], then extended to more general scenarios in [11] (see also [61] and [62]).

The convergence of the algorithm is guaranteed by the result shown in "Convergence of Algorithm 6."

Algorithm 6: FBF-based primal–dual algorithm.

Set $x_0 \in \mathbb{R}^N$ and $v_0 \in \mathbb{R}^K$
For $n = 0, 1, \ldots$
  Set $\gamma_n \in]0, +\infty[$
  $y_{1,n} = x_n - \gamma_n(\nabla h(x_n) + L^\top v_n)$
  $y_{2,n} = v_n + \gamma_n L x_n$
  $p_{1,n} = \operatorname{prox}_{\gamma_n f} y_{1,n}$
  $p_{2,n} = \operatorname{prox}_{\gamma_n g^*} y_{2,n}$
  $q_{1,n} = p_{1,n} - \gamma_n(\nabla h(p_{1,n}) + L^\top p_{2,n})$
  $q_{2,n} = p_{2,n} + \gamma_n L p_{1,n}$
  $(x_{n+1}, v_{n+1}) = (x_n - y_{1,n} + q_{1,n}, v_n - y_{2,n} + q_{2,n})$.

Algorithm 6 is often referred to as the *Monotone + Lipschitz FBF* (*M+LFBF*) algorithm. It enjoys the same advantages as the FB-based primal–dual algorithms we have seen before. It, however, makes it possible to compute the proximity operators of scaled versions of functions f and g^* in parallel. In addition, the choice of its parameters to satisfy convergence conditions may appear more intuitive than for Algorithms 2–4. With respect to FB-based algorithms, an extra forward step needs to be performed. This may lead to a slower convergence if, for example, the computational cost of the gradient is high and an iteration of a FB-based algorithm is at least as efficient as an iteration of Algorithm 6.

CONVERGENCE OF ALGORITHM 6
Under the following assumptions:
- $(\gamma_n)_{n\in\mathbb{N}}$ is a sequence in $[\varepsilon, (1-\varepsilon)/\mu]$ where $\varepsilon \in]0, 1/(1+\mu)[$ and $\mu = \beta + \|L\|_S$
- equation (18) admits a solution
- $\operatorname{int}(\operatorname{dom} g) \cap L(\operatorname{dom} f) \neq \varnothing$ or $\operatorname{dom} g \cap \operatorname{int}(L(\operatorname{dom} f)) \neq \varnothing$,

the sequence $(x_n, v_n)_{n\in\mathbb{N}}$ converges to to a pair of primal–dual solutions.

A PROJECTION-BASED PRIMAL–DUAL ALGORITHM

Another primal–dual algorithm was recently proposed in [63], which relies on iterative projections onto half spaces including the set of Kuhn–Tucker points (see Algorithm 7).

We then have the convergence result shown in "Convergence of Algorithm 7." Although few numerical experiments have been performed with this algorithm, one of its potential advantages is that it introduces few constraints on the choice of the parameters γ_n, μ_n, and λ_n at iteration n and that it does not require any knowledge on the norm of the matrix L. Nonetheless, the use of this algorithm does not allow us to exploit the fact that h is a differentiable function.

Algorithm 7: Projection based primal–dual algorithm.

Set $x_0 \in \mathbb{R}^N$ and $v_0 \in \mathbb{R}^K$
For $n = 0, 1, \ldots$
  Set $(\gamma_n, \mu_n) \in]0, +\infty[$
  $a_n = \operatorname{prox}_{\gamma_n(f+h)}(x_n - \gamma_n L^\top v_n)$
  $l_n = L x_n$
  $b_n = \operatorname{prox}_{\mu_n g}(l_n + \mu_n v_n)$
  $s_n = \gamma_n^{-1}(x_n - a_n) + \mu_n^{-1} L^\top(l_n - b_n)$
  $t_n = b_n - L a_n$
  $\tau_n = \|s_n\|^2 + \|t_n\|^2$
  if $\tau_n = 0$
    $\hat{x} = a_n$
    $\hat{v} = v_n + \mu_n^{-1}(l_n - b_n)$
    return
  else
    Set $\lambda_n \in]0, +\infty[$
    $\theta_n = \lambda_n(\gamma_n^{-1}\|x_n - a_n\|^2 + \mu_n^{-1}\|l_n - b_n\|^2)/\tau_n$
    $x_{n+1} = x_n - \theta_n s_n$
    $v_{n+1} = v_n - \theta_n t_n$.

CONVERGENCE OF ALGORITHM 7
Assume that
- $(\gamma_n)_{n\in\mathbb{N}}$ and $(\mu_n)_{n\in\mathbb{N}}$ are sequences such that $\inf_{n\in\mathbb{N}}\gamma_n > 0$, $\sup_{n\in\mathbb{N}}\gamma_n < +\infty$, $\inf_{n\in\mathbb{N}}\mu_n > 0$, $\sup_{n\in\mathbb{N}}\mu_n < +\infty$
- $(\lambda_n)_{n\in\mathbb{N}}$ is a sequence in $\mathbb{R}$ such that $\inf_{n\in\mathbb{N}}\lambda_n > 0$ and $\sup_{n\in\mathbb{N}}\lambda_n < 2$
- equation (18) admits a solution
- $\operatorname{Int}(\operatorname{dom} g) \cap L(\operatorname{dom} f) \neq \varnothing$ or $\operatorname{dom} g \cap \operatorname{int}(L(\operatorname{dom} f)) \neq \varnothing$,

then, either the algorithm terminates in a finite number of iterations at a pair of primal–dual solutions $(\hat{x}, \hat{v})$, or it generates a sequence $(x_n, v_n)_{n\in\mathbb{N}}$ converging to such a point.

EXTENSIONS

More generally, one may be interested in more challenging convex optimization problems of the form

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\; f(x) + \sum_{m=1}^{M} (g_m \,\square\, \ell_m)(L_m x) + h(x), \tag{23}$$

where $f \in \Gamma_0(\mathbb{R}^N)$, $h \in \Gamma_0(\mathbb{R}^N)$, and, for every $m \in \{1, \ldots, M\}$, $g_m \in \Gamma_0(\mathbb{R}^{K_m})$, $\ell_m \in \Gamma_0(\mathbb{R}^{K_m})$, and $L_m \in \mathbb{R}^{K_m \times N}$. The dual problem then reads

$$\underset{v1\in\mathbb{R}^{K_1},\ldots,\upsilon m\in\mathbb{R}^{K_M}}{\text{minimize}}\; (f^* \,\square\, h^*)\left(-\sum_{m=1}^{M} L_m^\top v_m\right) + \sum_{m=1}^{M} (g_m^*(v_m) + \ell_m^*(v_m)). \tag{24}$$

Some comments can be made on this general formulation. At first, one of its benefits is to split an original objective function in a sum of a number of simpler terms. Such a splitting strategy is often the key to efficient resolution of difficult optimization problems. For example, the proximity operator of the global objective function may be quite involved, while the proximity operators of the

individual functions may have an explicit form. A second point is that we have now introduced in the formulation additional functions $(\ell_m)_{1\le m\le M}$. These functions may be useful in some models [64], but they also present the conceptual advantage to make the primal problem and its dual form quite symmetric. For instance, this fact accounts for the symmetric roles played by Algorithms 2 and 4. An assumption that is commonly adopted is to assume that, whereas h is Lipschitz differentiable, the functions $(\ell_m)_{1\le m\le M}$ are strongly convex, i.e., their conjugates are Lipschitz differentiable. A last point to be emphasized is that such split forms are amenable to efficient parallel implementations (see "How to Parallelize Primal–dual Methods"). Using parallelized versions of primal–dual algorithms on multicore architectures may render these methods even more successful for dealing with large-scale problems.

DISCRETE OPTIMIZATION ALGORITHMS

BACKGROUND ON DISCRETE OPTIMIZATION

As already mentioned in the "Introduction" section, another common class of problems in signal processing and image analysis are discrete optimization problems, for which primal–dual algorithms also play an important role. Problems of this type are often stated as *integer linear programs* (*ILPs*), which can be expressed under the following form:

$$\text{Primal-ILP:}\quad \underset{x\in\mathbb{R}^N}{\text{minimize}}\quad c^\top x \quad \text{s.t.}\quad Lx\ge b,\qquad x\in\mathcal{N}\subset\mathbb{N}^N,$$

where $L=(L^{(i,j)})_{1\le i\le K,1\le j\le N}$ represents a matrix of size $K\times N$, and $b=(b^{(i)})_{1\le i\le K}$, $c=(c^{(j)})_{1\le j\le N}$ are column vectors of size K and N, respectively. Note that the ILP provides a very general formulation suitable for modeling a very broad range of problems and will, thus, form the setting that we will consider hereafter. Among the problems encountered in practice, many of them lead to a Primal-ILP that is NP-hard to solve. In such cases, a principled approach for finding an approximate solution is through the use of convex relaxations (see "Relaxations and Discrete Optimization"), where the original NP-hard problem is approximated with a surrogate one (the so-called relaxed problem), which is convex and, thus, much easier to solve. The premise is the following: to the extent that the surrogate problem provides a reasonably good approximation to the original optimization task, one can expect to obtain an approximately optimal solution for the latter by essentially making use of or solving the former.

The type of relaxations that are typically preferred in large-scale discrete optimization are based on LP, involving the minimization of a linear function subject to linear inequality constraints. These can be naturally obtained by simply relaxing the integrality constraints of Primal-ILP, thus leading to the relaxed primal problem (13) as well as its dual (15). It should be noted that the use of LP-relaxations is often dictated by the need for maintaining a reasonable computational cost. Although more powerful convex relaxations do exist in many cases, these may become intractable as the number of variables grows larger, especially for semidefinite programming or second-order cone programming relaxations.

Based on these observations, in the following, we aim to present some very general primal–dual optimization strategies that can be used in this context, focusing a lot on their underlying principles, which are based on two powerful techniques: the so-called primal–dual schema and dual decomposition. We will see

HOW TO PARALLELIZE PRIMAL–DUAL METHODS

Two main ideas can be used to put a primal–dual method under a parallel form. Let us first consider the following simplified form of (23):

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\ \sum_{m=1}^{M} g_m(L_m x). \tag{S2}$$

A possibility consists of reformulating this equation in a higher-dimensional space as

$$\underset{y_1\in\mathbb{R}^{K_1},\dots,y_M\in\mathbb{R}^{K_M}}{\text{minimize}}\ f(\boldsymbol{y})+\sum_{m=1}^{M} g_m(y_m), \tag{S3}$$

where $\boldsymbol{y}=[y_1^\top,\dots,y_M^\top]^\top\in\mathbb{R}^K$ with $K=K_1+\cdots+K_M$, and f is the indicator function of $\operatorname{ran}(\boldsymbol{L})$, where $\boldsymbol{L}=[L_1^\top,\dots,L_M^\top]^\top\in\mathbb{R}^{K\times N}$. Function f serves to enforce the constraint: $(\forall m\in\{1,\dots,M\})$ $y_m=L_m x$. By defining the separable function $g:\boldsymbol{y}\mapsto\sum_{m=1}^{M} g_m(y_m)$, we are, thus, led to the minimization of $f+g$ in the space $\mathbb{R}^K$. This optimization can be performed by the various primal–dual methods that we have described. The proximity operator of f reduces to the linear projection onto $\operatorname{ran}(\boldsymbol{L})$, whereas the separability of g ensures that its proximity operator can be obtained by computing in parallel the proximity operators of the functions $(g_m)_{1\le m\le M}$. Note that, when $L_1=\dots=L_M=\mathrm{Id}$, we recover a consensus-based approach that we have already discussed. This technique can be used to derive parallel forms of the Douglas–Rachford algorithm: the parallel proximal algorithm (PPXA) [65] and PPXA+ [66], as well as parallel versions of ADMM (simultaneous direction method of multipliers) [67].

The second approach is even more direct since it requires no projection onto $\operatorname{ran}(\boldsymbol{L})$. For simplicity, let us consider the following instance of (23):

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\ f(x)+\sum_{m=1}^{M} g_m(L_m x)+h(x). \tag{S4}$$

By defining the function g and the matrix $\boldsymbol{L}$ as in the previous approach, the problem can be recast as

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\ f(x)+g(Lx)+h(x). \tag{S5}$$

Once again, under appropriate assumptions on the involved functions, this formulation allows us to employ the algorithms proposed in the sections "Methods Based on a Forward–Backward Approach," "Methods Based on a Forward–Backward–Forward Approach," and "A Projection-Based Primal–Dual Algorithm," and we still have the ability to compute the proximity operator of g in a parallel manner.

RELAXATIONS AND DISCRETE OPTIMIZATION

Relaxations are very useful for solving approximately discrete optimization problems. Formally, given a problem

$$(\mathcal{P}): \underset{x \in C}{\text{minimize}} \; f(x),$$

where C is a subset of $\mathbb{R}^N$, we say that

$$(\mathcal{P}'): \underset{x \in C'}{\text{minimize}} \; f'(x)$$

with $C' \subset \mathbb{R}^N$ is a relaxation of $(\mathcal{P})$ if and only if (i) $C \subset C'$, and (ii) $(\forall x \in C')\, f(x) \geq f'(x)$.

For instance, let us consider the ILP defined by $(\forall x \in \mathbb{R}^N)$ $f(x) = c^\top x$ and $C = S \cap \mathbb{Z}^N$, where $c \in \mathbb{R}^N \backslash \{0\}$ and S is a nonempty closed polyhedron defined as

$$S = \{x \in \mathbb{R}^N \mid Lx \geq b\}$$

with $L \in \mathbb{R}^{K \times N}$ and $b \in \mathbb{R}^K$. One possible LP relaxation of $(\mathcal{P})$ is obtained by setting $f' = f$ and $C' = S$, which is typically much easier than $(\mathcal{P})$ (which is generally NP-hard). The quality of $(\mathcal{P}')$ is quantified by its so-called integrality gap defined as $\inf f(C)/\inf f'(C') \geq 1$ (provided that $-\infty < \inf f'(C') \neq 0$).

Hence, for approximation purposes, LP relaxations are not all of equal value. If

$$(\mathcal{P}''): \underset{x \in C''}{\text{minimize}} \; c^\top x$$

is another relaxation of $(\mathcal{P})$ with $C'' \subset C'$, then relaxation $(\mathcal{P}'')$ is tighter. Interestingly, $(\mathcal{P})$ always has a tight LP relaxation (with integrality gap 1) given by $C'' = \text{conv}(S \cap \mathbb{Z}^N)$, where $\text{conv}(C)$ is the convex hull polyhedron of C. Note, however, that if $(\mathcal{P})$ is NP-hard, polyhedron $\text{conv}(S \cap \mathbb{Z}^N)$ will involve exponentially many inequalities.

The relaxations in all of the previous examples involve expanding the original feasible set. But, as mentioned previously, we can also derive relaxations by modifying the original objective function. For instance, in so-called submodular relaxations [68], [69], one uses as a new objective a maximum submodular function that lower bounds the original objective. More generally, convex relaxations allow us to make use of the well-developed duality theory of convex programming for dealing with discrete nonconvex problems.

that to estimate an approximate solution to Primal-ILP, both approaches make heavy use of the dual of the underlying LP relaxation, i.e., (15). However, their strategies for doing so are quite different: the second essentially aims at solving this Dual-LP (and then converting the fractional solution into an integral one, trying not to increase the cost too much in the process), whereas the first simply uses it in the design of the algorithm.

THE PRIMAL–DUAL SCHEMA FOR ILPs

The primal–dual schema is a well-known technique in the combinatorial optimization community that has its origins in LP duality theory. It is worth noting that it started as an exact method for solving linear programs. As such, it had initially been used in deriving exact polynomial-time algorithms for many cornerstone problems in combinatorial optimization that have a tight LP relaxation. Its first use probably goes back to Edmond's famous Blossom algorithm for constructing maximum matchings on graphs, but it has also been applied to many other combinatorial problems including max flow (e.g., Ford and Fulkerson's augmenting path-based techniques for max flow can essentially be understood in terms of this schema), shortest path, minimum branching, and minimum spanning tree [70]. In all of these cases, the primal–dual schema is driven by the fact that optimal LP solutions should satisfy the complementary slackness conditions [see (16) and (17)]. Starting with an initial primal–dual pair of feasible solutions, it therefore iteratively steers them toward satisfying these complementary slackness conditions (by trying at each step to minimize their total violation). Once this is achieved, both solutions (the primal and the dual) are guaranteed to be optimal. Moreover, since the primal is always chosen to be updated integrally during the iterations, it is ensured that an integral optimal solution is obtained at the end. A notable feature of the primal–dual method is that it often reduces the original LP, which is a weighted optimization problem, to a series of purely combinatorial unweighted ones (related to minimizing the violation of complementary slackness conditions at each step).

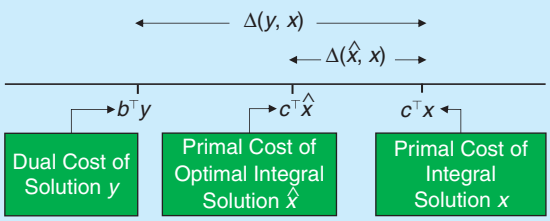
Interestingly, today the primal–dual schema is no longer used for providing exact algorithms. Instead, its main use concerns deriving approximation algorithms to NP-hard discrete problems that admit an ILP formulation, for which it has proved to be a very powerful and widely applicable tool. As such, it has been applied to many NP-hard combinatorial problems until now, including set cover, Steiner-network, scheduling, Steiner tree, feedback vertex set, and facility location, to mention only a few [17], [18]. With regard to problems from the domains of computer vision and image analysis, the primal–dual schema was recently introduced in [13] and [71] and has been used for modeling a broad class of tasks from these fields.

It should be noted that for NP-hard ILPs, an integral solution is no longer guaranteed to satisfy the complementary slackness conditions (since the LP-relaxation is not exact). How could it then be possible to apply this schema to such problems? It turns out that the answer to this question consists of using an appropriate relaxation of the previously described conditions. To understand exactly how we need to proceed in this case, let us consider the problem Primal-ILP. As already explained, the goal is to compute an optimal solution to it, but, due to the integrality constraints $x \in \mathcal{N}$, this is assumed to be NP-hard, and so we can only estimate an approximate solution. To achieve this, we will first need to relax the integrality constraints, thus giving rise to the relaxed primal problem in (13) as well as its dual (15). A primal–dual algorithm attempts to compute an approximate solution to Primal-ILP by relying on the following principle (see "Explanation of the Primal–Dual Principle in the Discrete Case").

EXPLANATION OF THE PRIMAL–DUAL PRINCIPLE IN THE DISCRETE CASE

Essentially, the proof of this principle relies on the fact that the sequence of optimal costs of problems Dual-LP, Primal-LP, and Primal-ILP is increasing.

Specifically, by weak LP duality, the optimal cost of Dual-LP is known to not exceed the optimal cost of Primal-LP. As a result of this fact, the cost $c^\top \hat{x}$ (of an unknown optimal integral solution $\hat{x}$) is guaranteed to be at least as large as the cost $b^\top y$ of any dual feasible solution y. On the other hand, by definition, $c^\top \hat{x}$ cannot exceed the cost $c^\top x$ of an integral-primal feasible solution x. Therefore, as can be seen in Figure S1, if the gap $\Delta(y, x)$ between the costs of y and x is small (e.g., it holds $c^\top x \le \nu b^\top y$), the same will be true for the gap $\Delta(\hat{x}, x)$ between the costs of $\hat{x}$ and x (i.e., $c^\top x \le \nu c^\top \hat{x}$), thus proving that x is a ν-approximation to optimal solution $\hat{x}$.



[FIGS1] A visual illustration of the primal–dual principle.

PRIMAL–DUAL PRINCIPLE IN THE DISCRETE CASE

Let $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^K$ be integral-primal and dual feasible solutions (i.e., $x \in \mathcal{N}$ and $Lx \ge b$, and $y \in [0, +\infty[^K$ and $L^\top y \le c$). Assume that there exists $\nu \in [1, +\infty[$ such that

$$c^\top x \le \nu\, b^\top y. \tag{25}$$

Then, x can be shown to be a ν-approximation to an unknown optimal integral solution $\hat{x}$, i.e.,

$$c^\top \hat{x} \le c^\top x \le \nu\, c^\top \hat{x}. \tag{26}$$

Although this principle lies at the heart of many primal–dual techniques (i.e., in one way or another, primal–dual methods often try to fulfill the assumptions imposed by this principle), it does not directly specify how to estimate a primal–dual pair of solutions (x, y) that satisfies these assumptions. This is where the so-called relaxed complementary slackness conditions come into play, as they typically provide an alternative and more convenient (from an algorithmic viewpoint) way for generating such a pair of solutions. These conditions generalize the complementary slackness conditions associated with an arbitrary pair of primal–dual linear programs (see the section "Duality in Linear Programming"). The latter conditions apply only in cases when there is no duality gap, such as between Primal-LP and Dual-LP, but they are not applicable to cases like Primal-ILP and Dual-LP when a duality gap exists as a result of the integrality constraint imposed on variable x. As in the exact case, two types of relaxed complementary slackness conditions exist, depending on whether the primal or dual variables are checked for being zero.

RELAXED PRIMAL COMPLEMENTARY SLACKNESS CONDITIONS WITH RELAXATION FACTOR $\nu_{\text{primal}} \le 1$

For given $x = (x^{(j)})_{1 \le j \le N} \in \mathbb{R}^N$, $y = (y^{(i)})_{1 \le i \le K} \in \mathbb{R}^K$, the following conditions are assumed to hold:

$$(\forall j \in J_x) \qquad \nu_{\text{primal}}\, c^{(j)} \le \sum_{i=1}^{K} L^{(i,j)} y^{(i)} \le c^{(j)}, \tag{27}$$

where $J_x = \{j \in \{1, \ldots, N\} \mid x^{(j)} > 0\}$.

RELAXED DUAL COMPLEMENTARY SLACKNESS CONDITIONS WITH RELAXATION FACTOR $\nu_{\text{dual}} \ge 1$

For given $y = (y^{(i)})_{1 \le i \le K} \in \mathbb{R}^K$, $x = (x^{(j)})_{1 \le j \le N} \in \mathbb{R}^N$, the following conditions are assumed to hold:

$$(\forall i \in I_y) \qquad b^{(i)} \le \sum_{j=1}^{N} L^{(i,j)} x^{(j)} \le \nu_{\text{dual}}\, b^{(i)}, \tag{28}$$

where $I_y = \{i \in \{1, \ldots, K\} \mid y^{(i)} > 0\}$.

When both $\nu_{\text{primal}} = 1$ and $\nu_{\text{dual}} = 1$, we recover the exact complementary slackness conditions in (16) and (17). The use of these conditions in the context of a primal–dual approximation algorithm becomes clear by the following result: If $x = (x^{(j)})_{1 \le j \le N}$ and $y = (y^{(i)})_{1 \le i \le K}$ are feasible with respect to Primal-ILP and Dual-LP, respectively, and satisfy the relaxed complementary slackness conditions (27) and (28), then the pair (x, y) satisfies the primal–dual principle in the discrete case with $\nu = \nu_{\text{dual}}/\nu_{\text{primal}}$. Therefore, x is a ν-approximate solution to Primal-ILP.

This result simply follows from the inequalities:

$$\begin{aligned} c^\top x &= \sum_{j=1}^{N} c^{(j)} x^{(j)} \overset{(27)}{\le} \sum_{j=1}^{N} \left(\frac{1}{\nu_{\text{primal}}} \sum_{i=1}^{K} L^{(i,j)} y^{(i)} \right) x^{(j)} \\ &= \frac{1}{\nu_{\text{primal}}} \sum_{i=1}^{K} \left(\sum_{j=1}^{N} L^{(i,j)} x^{(j)} \right) y^{(i)} \overset{(28)}{\le} \frac{\nu_{\text{dual}}}{\nu_{\text{primal}}} \sum_{i=1}^{K} b^{(i)} y^{(i)} = \frac{\nu_{\text{dual}}}{\nu_{\text{primal}}} b^\top y. \end{aligned} \tag{29}$$

Based on this result, iterative schemes can be devised, yielding a primal–dual ν-approximation algorithm. For example, we can employ the following algorithm:

Note that, in this scheme, primal solutions are always updated integrally. Also note that, when applying the primal–dual schema, different implementation strategies are possible. The strategy described in Algorithm 8 is to maintain feasible primal–dual solutions (x_n, y_n) at iteration n, and iteratively improve how tightly the (primal or dual) complementary slackness conditions get satisfied. This is performed through the introduction of slackness variables $(q^{(i)})_{i \in I_{y_n}}$ and $(r^{(j)})_{j \in J_{x_n}}$, the sums of which measure the degrees of violation of each relaxed

Algorithm 8: The primal–dual schema.

Generate a sequence $(x_n, y_n)_{n\in\mathbb{N}}$ of elements of $\mathbb{R}^N \times \mathbb{R}^K$ as follows:

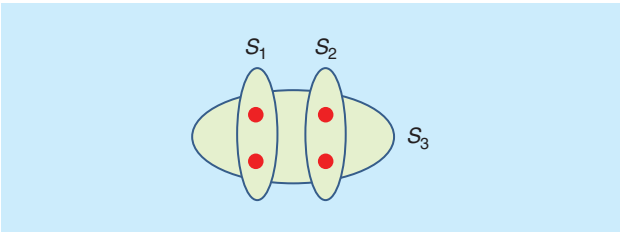
Set $\nu_{\text{primal}} \leq 1$ and $\nu_{\text{dual}} \geq 1$
Set $y_0 \in [0, +\infty[^K$ such that $L^\top y_0 \leq c$
For $n = 0, 1, \ldots$
  Find $x_n \in \{x \in \mathcal{N} \mid Lx \geq b\}$ minimizing
    $\sum_{i\in I_{y_n}} q^{(i)}$ s.t.
    $(\forall i \in I_{y_n})\ \sum_{j=1}^{N} L^{(i,j)} x^{(j)} \leq \nu_{\text{dual}}\, b^{(i)} + q^{(i)}, q^{(i)} \geq 0$
  Find $y_{n+1} \in \{y \in [0, +\infty[^K \mid L^\top y \leq c\}$ minimizing
    $\sum_{j\in J_{x_n}} r^{(j)}$ s.t.
    $(\forall j \in J_{x_n})\ \sum_{i=1}^{K} L^{(i,j)} y^{(i)} + r^{(j)} \geq \nu_{\text{primal}}\, c^{(j)}, r^{(j)} \geq 0.$ (30)

slackness condition and, thus, have to be minimized. Alternatively, for example, we can opt to maintain solutions (x_n, y_n), which satisfy the relaxed complementary slackness conditions but may be infeasible, and iteratively improve the feasibility of the generated solutions. For instance, if we start with a feasible dual solution but with an infeasible primal solution, such a scheme would result in improving the feasibility of the primal solution as well as the optimality of the dual solution at each iteration, ensuring that a feasible primal solution is obtained at the end. No matter which one of the two strategies we choose to follow, the end result will be to gradually bring the primal and dual costs $c^\top x_n$ and $b^\top y_n$ closer together so that asymptotically the primal–dual principle is satisfied with the desired approximation factor. Essentially, at each iteration, through the coupling by the complementary slackness conditions, the current primal solution is used to improve the dual, and vice versa.

Three remarks are worth making at this point: the first one relates to the fact that the two processes, i.e., the primal and the dual, make only local improvements to each other. Yet, in the end, they manage to yield a result that is almost globally optimal. The second point to emphasize is that, for computing this approximately optimal result, the algorithm requires no solution to the Primal-LP or Dual-LP to be computed, which are replaced by simpler optimization problems. This is an important advantage from a computational standpoint since, for large-scale problems, solving these relaxations can often be quite costly. In fact, in most cases where we apply the primal–dual schema, purely combinatorial algorithms can be obtained, which contain no sign of LP in the end. A last point to be noted is that these algorithms require appropriate choices of the relaxation factors ν_{primal} and ν_{dual}, which are often application guided.

APPLICATION TO THE SET COVER PROBLEM

For a simple illustration of the primal–dual schema, let us consider the problem of set cover, which is known to be NP-hard. In this problem, we are given as input a finite set $\mathcal{V}$ of K elements $(v^{(i)})_{1\leq i\leq K}$, a collection of (nondisjoint) subsets $\mathcal{S} = \{S_j\}_{1\leq j\leq N}$ where, for every $j \in \{1, \ldots, N\}$, $S_j \subset \mathcal{V}$, and



[FIG5] A toy set-cover instance with $K = 4$ and $N = 3$, where $\varphi(S_1) = (1/2)$, $\varphi(S_2) = 1$, and $\varphi(S_3) = 2$. In this case, the optimal set-cover is $\{S_1, S_2\}$ and has a cost of $(3/2)$.

$\bigcup_{j=1}^{N} S_j = \mathcal{V}$. Let $\varphi: \mathcal{S} \to \mathbb{R}$ be a function that assigns a cost $c_j = \varphi(S_j)$ for each subset S_j. The goal is to find a set cover (i.e., a subcollection of $\mathcal{S}$ that covers all elements of $\mathcal{V}$) that has minimum cost (see Figure 5).

This problem can be expressed as the following ILP:

$$\underset{x=(x^{(j)})_{1\leq j\leq N}}{\text{minimize}} \sum_{j=1}^{N} \varphi(S_j)\, x^{(j)} \tag{31}$$

$$\text{s.t. } (\forall i \in \{1, \ldots, K\}) \sum_{\substack{j\in\{1,\ldots,N\} \\ v^{(i)}\in S_j}} x^{(j)} \geq 1, \quad x \in \{0,1\}^N, \tag{32}$$

where indicator variables $(x^{(j)})_{1\leq j\leq N}$ are used for determining if a set in $\mathcal{S}$ has been included in the set cover or not, and (32) ensures that each one of the elements of $\mathcal{V}$ is contained in at least one of the sets that were chosen for participating to the set cover.

An LP-relaxation for this problem is obtained by simply replacing the Boolean constraint with the constraint $x \in [0, +\infty[^N$. The dual of this LP relaxation is given by the following linear program:

$$\underset{y=(y^{(i)})_{1\leq i\leq K}\in[0,+\infty[^K}{\text{maximize}} \sum_{i=1}^{K} y^{(i)} \tag{33}$$

$$\text{s.t. } (\forall j \in \{1, \ldots, N\}) \sum_{\substack{i\in\{1,\ldots,K\} \\ v^{(i)}\in S_j}} y^{(i)} \leq \varphi(S_j). \tag{34}$$

Let us denote by $F_{\max}$ the maximum frequency of an element in $\mathcal{V}$, where by the term *frequency* we mean the number of sets to which this element belongs. In this case, we will use the primal–dual schema to derive an $F_{\max}$-approximation algorithm by choosing $\nu_{\text{primal}} = 1$, $\nu_{\text{dual}} = F_{\max}$. This results in the following complementary slackness conditions, which we will need to satisfy:

Primal complementary slackness conditions:

$$(\forall j \in \{1, \ldots, N\}) \text{ if } x^{(j)} > 0 \text{ then } \sum_{\substack{i\in\{1,\ldots,K\} \\ v^{(i)}\in S_j}} y^{(i)} = \varphi(S_j). \tag{35}$$

Relaxed dual complementary slackness conditions (with relaxation factor $F_{\max}$):

$$(\forall i \in \{1, \ldots, K\}) \text{ if } y^{(i)} > 0 \text{ then } \sum_{\substack{j\in\{1,\ldots,N\} \\ v^{(i)}\in S_j}} x^{(j)} \leq F_{\max}. \tag{36}$$

Algorithm 9: The primal–dual schema for the set cover.

Set $x_0 \leftarrow 0$, $y_0 \leftarrow 0$
Declare all elements in $\mathcal{V}$ as uncovered
While $\mathcal{V}$ contains uncovered elements
  Select an uncovered element $v^{(i)}$ with $i \in \{1, \ldots, K\}$ and increase $y^{(i)}$ until some set becomes packed
  For every packed set S_j with $j \in \{1, \ldots, N\}$, set $x^{(j)} \leftarrow 1$ (include all the sets that are packed in the cover)
  Declare all the elements belonging to at least one set S_j with $x^{(j)} = 1$ as covered.

A set S_j with $j \in \{1, \ldots, N\}$ for which $\sum_{i \in \{1, \ldots, K\}: v^{(i)} \in S_j} y^{(i)} = \varphi(S_j)$ will be called *packed*. Based on this definition, and given that the primal variables $(x^{(j)})_{1 \le j \le N}$ are always kept integral (i.e., either 0 or 1) during the primal–dual schema, (35) basically says that only packed sets can be included in the set cover [note that overpacked sets are already forbidden by feasibility constraints (34)]. Similarly, (36) requires that an element $v^{(i)}$ with $i \in \{1, \ldots, K\}$ associated with a nonzero dual variable $y^{(i)}$ should not be covered more than $F_{\max}$ times, which is, of course, trivially satisfied given that $F_{\max}$ represents the maximum frequency of any element in $\mathcal{V}$.

Based on the previous observations, the iterative method whose pseudocode is shown in Algorithm 9 emerges naturally as a simple variant of Algorithm 8. Upon its termination, both x and y will be feasible given that there will be no uncovered element and no set that violates (34). Furthermore, given that the final pair (x, y) satisfies the relaxed complementary slackness conditions with $\nu_{\text{primal}} = 1$, $\nu_{\text{dual}} = F_{\max}$, the set cover defined by x will provide an $F_{\max}$-approximate solution.

DUAL DECOMPOSITION

We will next examine a different approach for discrete optimization based on the principle of dual decomposition [1], [14], [72]. The core idea behind this principle essentially follows a divide-and-conquer strategy: i.e., given a difficult or high-dimensional optimization problem, we decompose it into smaller, easy-to-handle subproblems and then extract an overall solution by cleverly combining the solutions from these subproblems.

To explain this technique, we will consider the general problem of minimizing the energy of a discrete Markov random field (MRF), which is an ubiquitous problem in the fields of computer vision and image analysis (applied with great success on a wide variety of tasks from these domains such as stereo matching, image segmentation, optical flow estimation, image restoration, inpainting, and object detection) [2]. This problem involves a graph G with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$ [i.e., $G = (\mathcal{V}, \mathcal{E})$] plus a finite label set $\mathcal{L}$. The goal is to find a labeling $z = (z^{(p)})_{p \in \mathcal{V}} \in \mathcal{L}^{|\mathcal{V}|}$ for the graph vertices that has minimum cost, i.e.,

$$\underset{z \in \mathcal{L}^{|\mathcal{V}|}}{\text{minimize}} \sum_{p \in \mathcal{V}} \varphi_p(z^{(p)}) + \sum_{e \in \mathcal{E}} \varphi_e(\mathrm{z}^{(e)}), \tag{37}$$

where, for every $p \in \mathcal{V}$ and $e \in \mathcal{E}$, $\varphi_p : \mathcal{L} \to]-\infty, +\infty[$ and $\varphi_e : \mathcal{L}^2 \to]-\infty, +\infty[$ represent the unary and pairwise costs (also known connectively as MRF potentials $\varphi = \{\{\varphi_p\}_{p \in \mathcal{V}}, \{\varphi_e\}_{e \in \mathcal{E}}\}$), and $\mathrm{z}^{(e)}$ denotes the pair of components of z defined by the variables corresponding to vertices connected by e (i.e., $\mathrm{z}^{(e)} = (z^{(p)}, z^{(q)})$ for $e = (p, q) \in \mathcal{E}$).

This problem is NP-hard, and much of the recent work on MRF optimization revolves around the following equivalent ILP formulation of (37) [73], which is the one that we will also use here:

$$\begin{aligned} \underset{x \in C_G}{\text{minimize}} \; f(x; \varphi) = & \sum_{p \in \mathcal{V}, z^{(p)} \in \mathcal{L}} \varphi_p(z^{(p)}) x_p(z^{(p)}) \\ & + \sum_{e \in \mathcal{E}, \mathrm{z}^{(e)} \in \mathcal{L}^2} \varphi_e(\mathrm{z}^{(e)}) \mathrm{x}_e(\mathrm{z}^{(e)}), \end{aligned} \tag{38}$$

where the set C_G is defined for any graph $G = (\mathcal{V}, \mathcal{E})$, as shown (39) in the box at the bottom of this page.

In (39), for every $p \in \mathcal{V}$ and $e \in \mathcal{E}$, the unary binary function $x_p(\cdot)$ and the pairwise binary function $\mathrm{x}_e(\cdot)$ indicate the labels assigned to vertex p and to the pair of vertices connected by edge $e = (p', q')$, respectively, i.e.,

$$\begin{aligned} (\forall z^{(p)} \in \mathcal{L}) \quad & x_p(z^{(p)}) = 1 \\ & \Leftrightarrow \; p \text{ is assigned label } z^{(p)} \end{aligned} \tag{40}$$

$$\begin{aligned} (\forall \mathrm{z}^{(e)} = (z^{(p')}, z^{(q')}) \in \mathcal{L}^2) \quad & \mathrm{x}_e(\mathrm{z}^{(e)}) = 1 \\ & \Leftrightarrow \; p', q' \text{ are assigned} \\ & \qquad \text{labels } z^{(p')}, z^{(q')}. \end{aligned} \tag{41}$$

Minimizing with respect to the vector x regrouping all these binary functions is equivalent to searching for an optimal binary vector of dimension $N = |\mathcal{V}||\mathcal{L}| + |\mathcal{E}||\mathcal{L}|^2$. The first constraints in (39) simply encode the fact that each vertex must be assigned exactly one label, whereas the rest of the constraints enforces consistency between unary functions $x_p(\cdot)$, $x_q(\cdot)$, and the pairwise function $\mathrm{x}_e(\cdot)$ for edge $e = (p, q)$, ensuring essentially that if $x_p(z^{(p)}) = x_q(z^{(q)}) = 1$, then $\mathrm{x}_e(z^{(p)}, z^{(q)}) = 1$.

$$C_G = \left\{ x = \{\{x_p\}_{p \in \mathcal{V}, z \in \mathcal{L}}, \{\mathrm{x}_e\}_{e \in \mathcal{V}, z \in \mathcal{L}^2}\} \;\middle|\; \begin{array}{ll} (\forall p \in \mathcal{V}) & \sum_{z^{(p)} \in \mathcal{L}} x_p(z^{(p)}) = 1 \\ (\forall e = (p, q) \in \mathcal{E})\,(\forall z^{(q)} \in \mathcal{L}) & \sum_{z^{(e)} \in \mathcal{L} \times \{z^{(q)}\}} \mathrm{x}_e(z^{(e)}) = x_q(z^{(q)}) \\ (\forall e = (p, q) \in \mathcal{E})\,(\forall z^{(p)} \in \mathcal{L}) & \sum_{z^{(e)} \in \{z^{(p)}\} \times \mathcal{L}} \mathrm{x}_e(z^{(e)}) = x_p(z^{(p)}) \\ (\forall p \in \mathcal{V}) & x_p(\cdot) : \mathcal{L} \mapsto \{0, 1\} \\ (\forall e \in \mathcal{E}) & \mathrm{x}_e(\cdot) : \mathcal{L}^2 \to \{0, 1\} \end{array} \right\}. \tag{39}$$

As mentioned previously, our goal will be to decompose the MRF problem (38) into easier subproblems (called *slaves*), which, in this case, involve optimizing MRFs defined on subgraphs of G. More specifically, let $\{G_m = (\mathcal{V}_m, \mathcal{E}_m)\}_{1\le m\le M}$ be a set of subgraphs that form a decomposition of $G = (\mathcal{V}, \mathcal{E})$ (i.e., $\cup_{m=1}^{M}\mathcal{V}_m = \mathcal{V}, \cup_{m=1}^{M}\mathcal{E}_m = \mathcal{E}$). On each of these subgraphs, we define a local MRF with corresponding (unary and pairwise) potentials $\varphi^m = \{\{\varphi_p^m\}_{p\in\mathcal{V}_m}, \{\varphi_e^m\}_{e\in\mathcal{E}_m}\}$, whose cost function $f^m(x;\varphi^m)$ is thus given by

$$f^m(x;\varphi^m) = \sum_{p\in\mathcal{V}_m, z^{(p)}\in\mathcal{L}} \varphi_p^m(z^{(p)})x_p(z^{(p)}) + \sum_{e\in\mathcal{E}_m, z^{(e)}\in\mathcal{L}^2} \varphi_e^m(\mathrm{z}^{(e)})\mathrm{x}_e(\mathrm{z}^{(e)}). \quad (42)$$

Moreover, the sum (over m) of the potential functions φ^m is ensured to give back the potentials φ of the original MRF on G, i.e.,

$$(\forall p\in\mathcal{V})(\forall e\in\mathcal{E}) \quad \sum_{m\in\{1,\dots,M\}:\, p\in\mathcal{V}m} \varphi_p^m = \varphi_p, \quad \sum_{m\in\{1,\dots,M\}:e\in\mathcal{E}m} \varphi_e^m = \varphi_e. \quad (43)$$

[To ensure (43), we can simply set: $(\forall m\in\{1,\dots,M\})\,\varphi_p^m = (\varphi_p)/(|\{m'|p\in\mathcal{V}_{m'}\}|)$, and $\varphi_e^m = (\varphi_e)/(|\{m'|e\in\mathcal{E}_{m'}\}|)$.] This guarantees that $f = \sum_{m=1}^{M} f^m$, thus allowing us to re-express (38) as follows:

$$\underset{x\in C_G}{\text{minimize}} \ \sum_{m=1}^{M} f^m(x;\varphi^m). \quad (44)$$

An assumption that often holds in practice is that minimizing separately each of the f^m (over x) is easy, but minimizing their sum is hard. Therefore, to leverage this fact, we introduce, for every $m\in\{1,\dots,M\}$, an auxiliary copy $x^m\in C_{G_m}$ for the variables of the local MRF defined on G_m, which are thus constrained to coincide with the corresponding variables in vector x, i.e., it holds $x^m = x_{|G_m}$, where $x_{|G_m}$ is used to denote the subvector of x containing only those variables associated with vertices and edges of subgraph G_m. In this way, (44) can be transformed into

$$\underset{x\in C_G,\{x^m\in C_{G_m}\}1\le m\le M}{\text{minimize}} \ \sum_{m=1}^{M} f^m(x^m;\varphi^m) \quad \text{s.t. } (\forall m\in\{1,\dots,M\}) \quad x^m = x_{|G_m}. \quad (45)$$

By considering the dual of (45), using a technique similar to the one described in "Consensus and Sharing Are Dual Problems," and noticing that

$$x\in C_G \ \Leftrightarrow \ (\forall m\in\{1,\dots,M\}) \ \ x^m\in C_{G_m}, \quad (46)$$

we finally end up with the following problem:

$$\underset{(\upsilon^m)1\le m\le M\in\Lambda}{\text{maximize}} \ \sum_{m=1}^{M} h^m(\upsilon^m), \quad (47)$$

where, for every $m\in\{1,\dots,M\}$, the dual variable υ^m consists of $\{\{\upsilon_p^m\}_{p\in\mathcal{V}_m}, \{\mathrm{v}_e^m\}_{e\in\mathcal{E}_m}\}$ similarly to φ^m, and function h^m is related to the following optimization of a slave MRF on G_m:

$$h^m(\upsilon^m) = \min_{x^m\in C_{Gm}} f^m(x^m;\varphi^m+\upsilon^m). \quad (48)$$

The feasible set Λ is given by (49), shown in the box at the bottom of this page. This dual problem provides a relaxation to the original problem (38)–(39). Furthermore, note that this relaxation leads to a convex optimization problem [to see this, notice that $h^m(\upsilon^m)$ is equal to a pointwise minimum of a set of linear functions of υ^m, and, thus, it is a concave function], although the original one is not. As such, it can always be solved in an optimal manner. A possible way of doing this consists of using a projected subgradient method. Exploiting the form of the projection onto the vector space Λ yields Algorithm 10, where $(\gamma_n)_{n\in\mathbb{N}}$ is a summable sequence of positive step-sizes and $\left\{\{\hat{x}_{p,n}^m\}_{p\in\mathcal{V}_m}, \{\hat{\mathrm{x}}_{e,n}^m\}_{e\in\mathcal{E}_m}\right\}$ corresponds to a subgradient of function h^m with $m\in\{1,\dots,M\}$ computed at iteration n [14]. Note that this algorithm requires only solutions to local subproblems to be computed, which is, of course, a much easier task that furthermore can be executed in a parallel manner. The solution to the master MRF is filled in from local solutions $\left\{\{\hat{x}_{p,n}^m\}_{p\in\mathcal{V}_m}, \{\hat{\mathrm{x}}_{e,n}^m\}_{e\in\mathcal{E}_m}\right\}_{1\le m\le M}$ after convergence of the algorithm.

For a better intuition for the updates of the variables $\{\{\varphi_{p,n}^m\}_{p\in\mathcal{V}_m}, \{\varphi_{e,n}^m\}_{e\in\mathcal{E}_m}\}_{1\le m\le M, n\in\mathbb{N}}$ in Algorithm 10, we should note that their aim is essentially to bring a consensus among the solutions of the local subproblems. In other words, they try to adjust the potentials of the slave MRFs so that, in the end, the corresponding local solutions are consistent with each other, i.e., all variables corresponding to a common vertex or edge are assigned the same value by the different subproblems. If this condition is satisfied (i.e., there is a full consensus), then the overall solution that results from combining the consistent local solutions is guaranteed to be optimal. In general, though, this might not always be true given that the aforementioned procedure is solving only a relaxation of the original NP-hard problem. (See also "Master-Slave Communication" for another interpretation of the updates of Algorithm 10.)

$$\Lambda = \left\{ \upsilon = \{\{\upsilon_p^m\}_{p\in\mathcal{V}_m}, \{\mathrm{v}_e^m\}_{e\in\mathcal{E}_m}\}_{1\le m\le M} \;\middle|\; \begin{array}{ll} (\forall p\in\mathcal{V})(\forall z^{(p)}\in\mathcal{L}) & \sum_{m\in\{1,\dots,M\}:\, p\in\mathcal{V}m} \upsilon_p^m(z^{(p)}) = 0, \\ (\forall e\in\mathcal{E})(\forall \mathrm{z}^{(e)}\in\mathcal{L}^2) & \sum_{m\in\{1,\dots,M\}:\, e\in\mathcal{E}m} \mathrm{v}_e^m(\mathrm{z}^{(e)}) = 0 \\ (\forall m\in\{1,\dots,M\})(\forall p\in\mathcal{V}) & \upsilon_p^m(\cdot):\mathcal{L}\mapsto\mathbb{R} \\ (\forall m\in\{1,\dots,M\})(\forall e\in\mathcal{E}) & \mathrm{v}_e^m(\cdot):\mathcal{L}^2\mapsto\mathbb{R} \end{array} \right\}. \quad (49)$$

Algorithm 10: The dual decomposition for MRF optimization.

Choose a decomposition $\{G_m = (\mathcal{V}_m, \mathcal{E}_m)\}_{1 \le m \le M}$ of G
Initialize potentials of slave MRFs:

$$(\forall m \in \{1, \ldots, M\})\,(\forall p \in \mathcal{V}_m)\ \varphi^m_{p,0} = \frac{\varphi_p}{|\{m' \,|\, p \in \mathcal{V}_{m'}\}|},$$

$$(\forall e \in \mathcal{E}_m)\ \varphi^m_{e,0} = \frac{\varphi_e}{|\{m' \,|\, e \in \mathcal{E}_{m'}\}|}$$

for $n = 0, \ldots$

Compute minimizers of slave MRF problems:

$$(\forall m \in \{1, \ldots, M\})\{\{\hat{x}^m_{p,n}\}_{p \in \mathcal{V}_m}, \{\hat{x}^m_{e,n}\}_{e \in \mathcal{E}_m}\} \in \operatorname*{Argmin}_{x^m \in C_{G_m}} f^m(x^m; \varphi^m_n)$$

Update potentials of slave MRFs:

$$(\forall m \in \{1, \ldots, M\})\,(\forall p \in \mathcal{V}_m)\ \varphi^m_{p,n+1} = \varphi^m_{p,n+1} + \gamma_n \left(\hat{x}^m_{p,n} - \frac{\sum_{m: p \in \mathcal{V}_m} \hat{x}^m_{p,n}}{|\{m' \,|\, p \in \mathcal{V}_{m'}\}|} \right)$$

$$(\forall m \in \{1, \ldots, M\})\,(\forall e \in \mathcal{E}_m)\ \varphi^m_{e,n+1} = \varphi^m_{e,n} + \gamma_n \left(\hat{x}^m_{e,n} - \frac{\sum_{m: e \in \mathcal{E}_m} \hat{x}^m_{e,n}}{|\{m' \,|\, p \in \mathcal{V}_{m'}\}|} \right).$$

Interestingly, if we choose to use a decomposition consisting only of subgraphs that are trees, then the resulting relaxation can be shown to actually coincide with the standard LP-relaxation of linear integer program (38) (generated by replacing the integrality constraints with nonnegativity constraints on the variables). This also means that when this LP-relaxation is tight, an optimal MRF solution is computed. This, for instance, leads to the result that dual decomposition approaches can estimate a globally optimal solution for binary submodular MRFs [although it should be noted that much faster graph-cut-based techniques exist for submodular problems of this type (see "Graph Cuts and MRF Optimization")]. Furthermore, when using subgraphs that are trees, a minimizer to each slave problem can be computed efficiently by applying the belief propagation algorithm [74], which is a message-passing method. Therefore, in this case, Algorithm 10 essentially reduces to a continuous exchange of messages between the nodes of graph G. Such an algorithm relates to or generalizes various other message-passing approaches [15], [75]–[79]. In general, besides tree-structured subgraphs, other types of decompositions or subproblems can be used as well (such as binary planar problems, or problems on loopy subgraphs with small tree-width, for which MRF optimization can still be solved efficiently), which can lead to even tighter relaxations (see "Decompositions and Relaxations") [80]–[85].

Furthermore, besides the projected subgradient method, one can alternatively apply an ADMM scheme for solving relaxation (47) (see the "ADMM" section). The main difference, in this case, is that the optimization of a slave MRF problem is performed by solving a (usually simple) local quadratic problem, which can again be solved efficiently for an appropriate choice of the decomposition (see the "Extensions" section). This method again penalizes disagreements among slaves, but it does so even more aggressively than the subgradient method since there is no longer a requirement for step-sizes $(\gamma_n)_{n \in \mathbb{N}}$ converging to zero. Furthermore, alternative smoothed accelerated schemes exist and can be applied as well [88]–[90].

APPLICATIONS

Although the presented primal–dual algorithms can be applied virtually to any area where optimization problems have to be solved, we now mention a few common applications of these techniques.

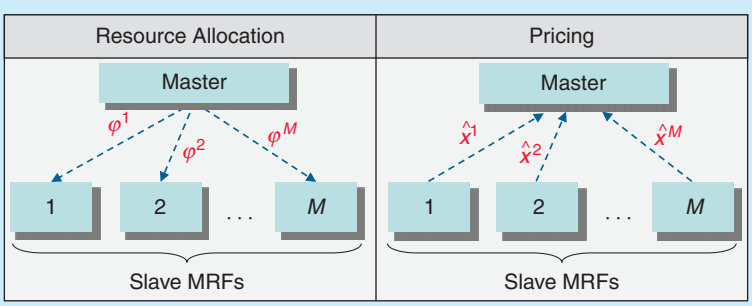
INVERSE PROBLEMS

For a long time, convex optimization approaches have been successfully used for solving inverse problems such as signal restoration, signal reconstruction, or interpolation of missing data. Most of the time, these problems are ill posed, and to recover the signal of interest in a satisfactory manner, some prior information needs to be introduced. To do this, an objective function can be minimized, which includes a data fidelity term modeling knowledge about the noise statistics and possibly involves a linear observation matrix (e.g., a convolutive blur), and a regularization (or penalization) term, which corresponds to the additional prior information. This formulation can also often be justified statistically as the determination of a maximum a posteriori (MAP) estimate. In early developed methods, in particular, in Tikhonov regularization, a quadratic penalty function is employed.

MASTER–SLAVE COMMUNICATION

As shown in Figure S2, during dual decomposition, a communication between a master process and the slaves (local subproblems) can be thought of as taking place, which can also be interpreted as a resource allocation/pricing stage.

- **Resource allocation:** At each iteration, the master assigns new MRF potentials (i.e., resources) $(\varphi^m)_{1 \le m \le M}$ to the slaves based on the current local solutions $(\hat{x}^m)_{1 \le m \le M}$.
- **Pricing:** The slaves respond by adjusting their local solutions $(\hat{x}^m)_{1 \le m \le M}$ (i.e., the prices) so as to maximize their welfares based on the newly assigned resources $(\hat{x}^m)_{1 \le m \le M}$.



[FIGS2] Dual decomposition as pricing and resource allocation.

GRAPH CUTS AND MRF OPTIMIZATION

For certain MRFs, optimizing their cost is known to be equivalent to solving a polynomial mincut problem [86], [87]. These are exactly all the binary MRFs ($|\mathcal{L}| = 2$) with submodular pairwise potentials such that, for every $e \in \mathcal{E}$

$$\varphi_e(0,0) + \varphi_e(1,1) \leq \varphi_e(0,1) + \varphi_e(1,0). \tag{S6}$$

Because of (S6), the cost $f(x)$ of a binary labeling $x = (x^{(p)})_{1 \leq p \leq |\mathcal{V}|} \in \{0,1\}^{|\mathcal{V}|}$ for such MRFs can always be written (up to an additive constant) as

$$f(x) = \sum_{p \in \mathcal{V}_P} a_p x^{(p)} + \sum_{p \in \mathcal{V}_N} a^{(p)}(1 - x^{(p)}) + \sum_{(p,q) \in \mathcal{E}} a_{p,q} x^{(p)}(1 - x^{(q)}), \tag{S7}$$

where all coefficients $(a_p)_{p \in \mathcal{V}}$ and $(a_{p,q})_{(p,q) \in \mathcal{E}}$ are nonnegative ($\mathcal{V}_P \subset \mathcal{V}$, $\mathcal{V}_N \subset \mathcal{V}$).

In this case, we can associate to f a capacitated network that has vertex set $\mathcal{V}_f = \mathcal{V} \cup \{s,t\}$. A source vertex s and a sink one t have thus been added. The new edge set $\mathcal{E}_f$ is deduced from the one used to express f

$$\mathcal{E}_f = \{(p,t) \,|\, p \in \mathcal{V}_P\} \cup \{(s,p) \,|\, p \in \mathcal{V}_N\} \cup \mathcal{E},$$

and its edge capacities are defined as $(\forall p \in \mathcal{V}_P \cup \mathcal{V}_N)\ c_{p,t} = c_{s,p} = a_p$ and $(\forall (p,q) \in \mathcal{E})\ c_{p,q} = a_{p,q}$.

A one-to-one correspondence between $s-t$ cuts and MRF labelings then exists

$$x \in \{0,1\}^{|\mathcal{V}|} \leftrightarrow \text{cut}(x) = \{s\} \cup \{p \,|\, x^{(p)} = 1\},$$

for which it is easy to see that

$$f(x) = \sum_{u \in \text{cut}(x),\, v \notin \text{cut}(x)} c_{u,v} = \text{cost of cut}(x).$$

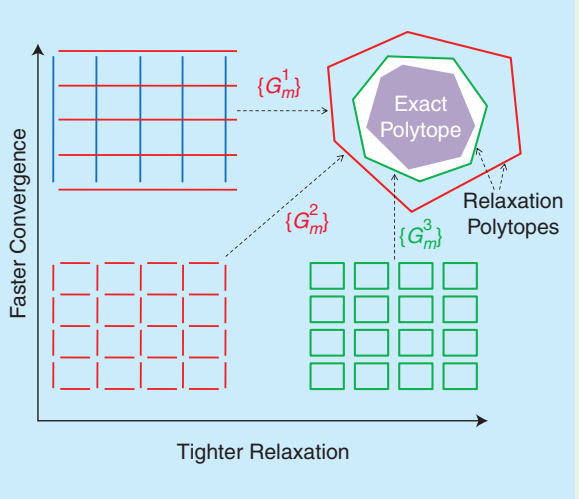
Computing a mincut, in this case, solves the LP relaxation of (38), which is tight, whereas computing a max-flow solves the dual LP.

DECOMPOSITIONS AND RELAXATIONS

Different decompositions can lead to different relaxations and/or can affect the speed of convergence. For instance, we show in Figure S3 three possible decompositions for an MRF assumed to be defined on a 5×5 image grid.

Decompositions $\{G_m^1\}$, $\{G_m^2\}$, and $\{G_m^3\}$ consist, respectively, of one suproblem per row and column, one subproblem per edge, and one subproblem per 2×2 subgrid of the original 5×5 grid. Both $\{G_m^1\}$ and $\{G_m^2\}$ (due to using solely subgraphs that are trees) lead to the same LP relaxation of (37), whereas $\{G_m^3\}$ leads to a relaxation that is tighter (due to containing loopy subgraphs).

On the other hand, decomposition $\{G_m^1\}$ leads to faster convergence compared with $\{G_m^2\}$ due to using larger subgraphs that allow a faster propagation of information during message passing.



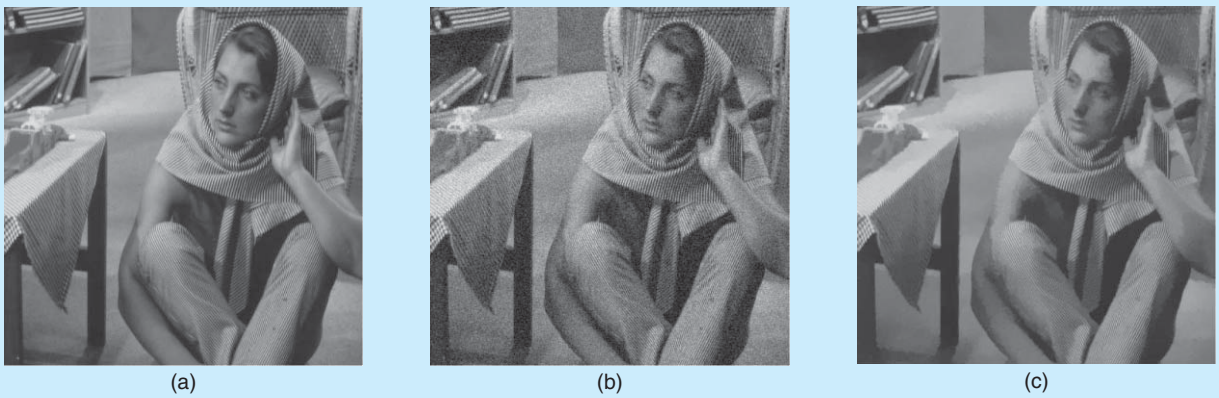
[FIGS3] Different decompositions can lead to different relaxations and also affect the speed of convergence.

Alternatively, hard constraints can be imposed on the solution (for example, bounds on the signal values), leading to signal feasibility problems. Today, a hybrid regularization [91] may be preferred so as to combine various kinds of regularity measures, possibly computed for different representations of the signal (Fourier, wavelets, etc.), some of them like total variation [25] and its nonlocal extensions [92] being tailored for preserving discontinuities such as image edges. In this context, constraint sets can be translated into penalization terms being equal to the indicator functions of these sets [see (2)]. Altogether, these lead to global cost functions which can be quite involved, often with many variables, for which the splitting techniques described in the "Extensions" section are very useful. An extensive literature exists on the use of ADMM methods for solving inverse problems (e.g., see [29]–[33]). With the advent of more recent primal–dual algorithms, many works have been mainly focused on image recovery applications [46]–[49], [51], [54], [55], [58], [62], [64], [93]–[97]. Two examples are given next.

In [98], a generalization of the total variation is defined for an arbitrary graph to address a variety of inverse problems. For denoising applications, the optimization problem to be solved is of the form (18) with

$$f = 0, \qquad g = \sigma_C, \qquad h\colon x \mapsto \frac{1}{2}\|x - y\|^2, \tag{50}$$

where x is a vector of variables associated with each vertex of a weighted graph, and $y \in \mathbb{R}^N$ is a vector of data observed at each vertex. The matrix $L \in \mathbb{R}^{K \times N}$ is equal to $\text{Diag}(\sqrt{\omega}_1, \ldots, \sqrt{\omega}_K)A$, where $(\omega_1, \ldots \omega_K) \in [0, +\infty[^K$ is the vector of edge weights and $A \in \mathbb{R}^{K \times N}$ is the graph incidence matrix playing a role similar to a gradient operator on the graph. The set C is defined as an intersection of closed semiballs in such a way that its support function σ_C [see (S1)] allows us to define a class of functions extending the total variation seminorm (see [98] for more details). Good image denoising results can be obtained by building the graph in a nonlocal

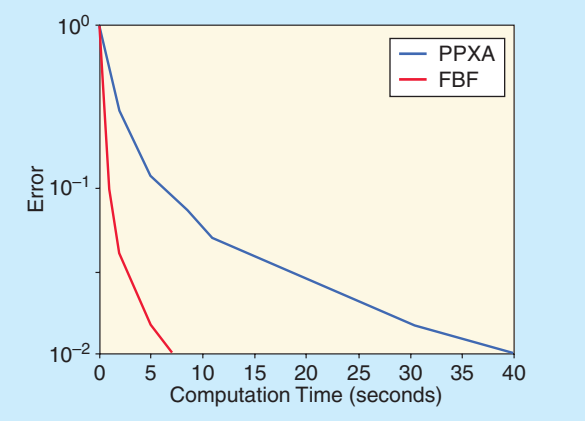


[FIG6] Nonlocal denoising (additive white zero-mean Gaussian noise with variance $\sigma^2 = 20$): (a) original image, (b) noisy signal-to-noise ratio (SNR) = 14.47 dB, and (c) nonlocal TV SNR = 20.78 dB.

manner following the strategy in [92]. Results obtained for the "Barbara" image are displayed in Figure 6. Interestingly, the ability of methods such as those presented in the section "Methods Based on a Forward–Backward–Forward Approach" to circumvent matrix inversions leads to a significant decrease of the convergence time for irregular graphs in comparison with algorithms based on the Douglas–Rachford iteration or ADMM (see Figure 7).

Another application example of primal–dual proximal algorithms is parallel magnetic resonance imaging reconstruction. A set of measurement vectors $(z_j)_{1\le j\le J}$ is acquired from J coils. These observations are related to the original full-field-of-view image $\overline{x} \in \mathbb{C}^N$ corresponding to a spin density. An estimate of $\overline{x}$ is obtained by solving the following problem:

$$\underset{x\in\mathbb{C}^N}{\text{minimize}} \quad f(x) + g(Lx) + \underbrace{\sum_{j=1}^{J} \|\Sigma F S_j x - z_j\|^2_{\Lambda_j^{-1}}}_{h(x)}, \tag{51}$$



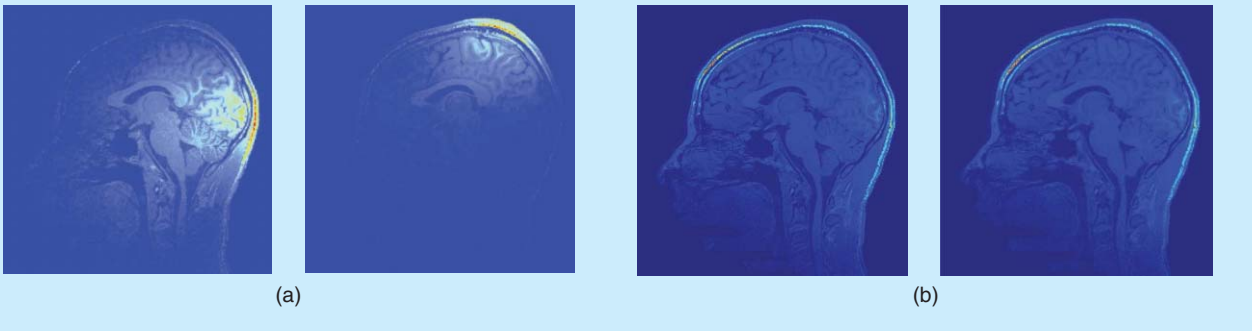
[FIG7] A comparison of the convergence speed of a Douglas–Rachford-based algorithm (PPXA [65]) (blue) and an FBF-based primal–dual algorithm (red) for image denoising using a nonregular graph. The MATLAB implementation was done on an Intel Xeon 2.5-GHz, eight-core system.

where $(\forall j \in \{1,\dots,J\})$ $\|\cdot\|^2_{\Lambda_j^{-1}} = (\cdot)^H \Lambda_j^{-1}(\cdot)$, Λ_j is the noise covariance matrix for the jth channel, $S_j \in \mathbb{C}^{N\times N}$ is a diagonal matrix modeling the sensitivity of the coil, $F \in \mathbb{C}^{N\times N}$ is a two-dimensional (2-D) discrete Fourier transform, $\Sigma \in \{0,1\}^{\lfloor N/R\rfloor \times N}$ is a subsampling matrix, $g \in \Gamma_0(\mathbb{C}^K)$ is a sparsity measure (e.g., a weighted ℓ_1-norm), $L \in \mathbb{C}^{K\times N}$ is a (possibly redundant) frame analysis operator, and f is the indicator function of a vector subspace of $\mathbb{C}^N$ serving to set to zero the image areas corresponding to the background [$(\cdot)^H$ denotes the transconjugate operation and $\lfloor\cdot\rfloor$ designates the lower rounding operation]. Combining suitable subsampling strategies in the k-space with the use of an array of coils allows us to reduce the acquisition time while maintaining a good image quality. The subsampling factor $R > 1$ thus corresponds to an acceleration factor. For a more detailed account on the considered approach, see [99] and [100] and the references therein. The reconstruction results are shown in Figure 8. Figure 9 also allows us to evaluate the convergence time for various algorithms. It can be observed that smaller differences between the implemented primal–dual strategies are apparent in this example. Because of the form of the subsampling matrix, the matrix inversion involved at each iteration of ADMM requires us to make use of a few subiterations of a linear conjugate gradient method.

Note that convex primal–dual proximal optimization algorithms have been applied to other fields besides image recovery, in particular, to machine learning [5], [101], system identification [102], audio processing [103], optimal transport [104], empirical mode decomposition [105], seismics [106], database management [107], and data streaming over networks [108].

COMPUTER VISION AND IMAGE ANALYSIS

The great majority of problems in computer vision involve image observation data that are of very high dimensionality, inherently ambiguous, noisy, incomplete, and often only provide a partial view of the desired space. Hence, any successful model that aims to explain such data usually requires a reasonable regularization, a robust data measure, and a compact structure between the variables of interest to efficiently characterize their relationships.

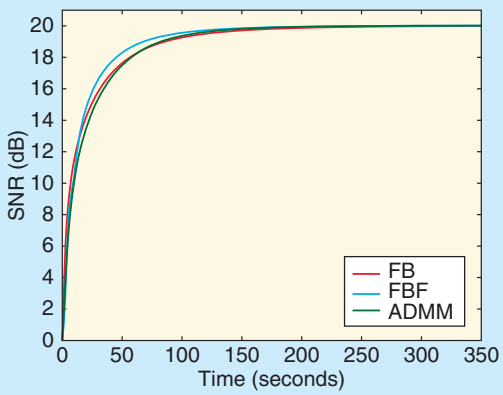


[FIG8] (a) The effects of the sensitivity matrices in the spatial domain in the absence of subsampling: the moduli of the images corresponding to $(S_j\overline{x})_{2\le j\le 3}$ are displayed for two channels out of 32. (b) The reconstruction quality: moduli of the original slice $\overline{x}$ and the reconstructed one with SNR $= 20.03$ dB (from left to right) using polynomial sampling of order 1 with $R = 5$, a wavelet frame, and an ℓ_1-regularization.

Probabilistic graphical models, and, in particular, discrete MRFs, have led to a suitable methodology for solving such visual perception problems [12], [16]. These types of models offer great representational power, and are able to take into account dependencies in the data, encode prior knowledge, and model (soft or hard) contextual constraints in a very efficient and modular manner. Furthermore, they offer the important ability to make use of very powerful data likelihood terms consisting of arbitrary nonconvex and noncontinuous functions that are often crucial for accurately representing the problem at hand. As a result, a MAP inference for these models leads to discrete optimization problems that are (in most cases) highly nonconvex (NP-hard) and also of very large scale [109], [110]. These discrete problems take the form (37), where typically the unary terms $\varphi_p(\cdot)$ encode the data likelihood and the higher-order terms $\varphi_e(\cdot)$ encode problem-specific priors.

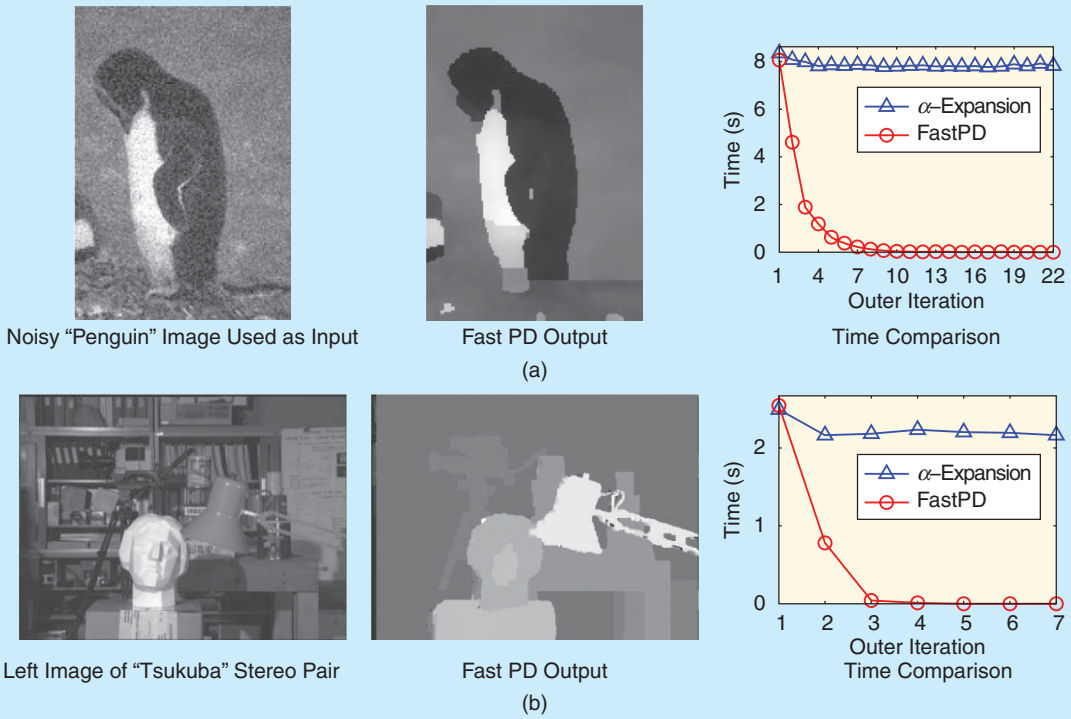
Primal–dual approaches can offer important computational advantages when dealing with such problems. One such characteristic example is the FastPD algorithm [13], which currently provides one of the most efficient methods for solving generic MRF optimization problems of this type, also guaranteeing at the same time the convergence to solutions that are approximately optimal. The theoretical derivation of this method relies on the use of the primal–dual schema described in the section "Discrete Optimization Algorithms," which results, in this case, in a very fast graph-cut-based inference scheme that generalizes previous state-of-the-art approaches such as the α-expansion algorithm [111] (see Figure 10). More generally, because of the very wide applicability of MRF models to computer vision or image analysis problems, primal–dual approaches can be and have been applied to a broad class of both low- and high-level problems from these domains, including image segmentation [112]–[115], stereo matching and three-dimensional (3-D) multiview reconstruction [116], [117], graph-matching [118], 3-D surface tracking [119], optical flow estimation [120], scene understanding [121], image deblurring [122], panoramic image stitching [123], category-level segmentation [124], and motion tracking [125]. In the following, we mention very briefly just a few examples.

A primal–dual based optimization framework has been recently proposed in [127] and [128] for the problem of deformable registration/fusion, which forms one of the most central and

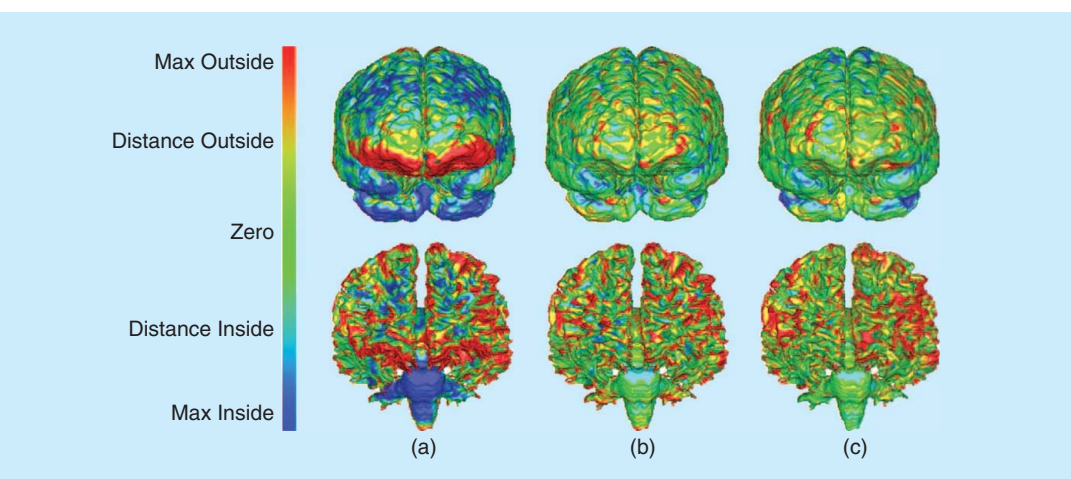


[FIG9] The SNR as a function of the computation time using ADMM and FB- or FBF-based primal–dual methods for a given slice. The MATLAB implementation was done on a 2.9-GHz Intel i7-3520M central processing unit (CPU).

challenging tasks in medical image analysis. This problem consists of recovering a nonlinear dense deformation field that aligns two signals that have, in general, an unknown relationship both in the spatial and intensity domain. In this framework, toward dimensionality reduction on the variables, the dense registration field is first expressed using a set of control points (registration grid) and an interpolation strategy. Then, the registration cost is expressed using a discrete sum over image costs projected on the control points and a smoothness term that penalizes local deviations on the deformation field according to a neighborhood system on the grid. One advantage of the resulting optimization framework is that it is able to encode even very complex similarity measures (such as normalized mutual information and Kullback–Leibler divergence) and, therefore, can be used even when seeking transformations between different modalities (interdeformable registration). Furthermore, it admits a broad range of regularization terms and can also be applied to both 2-D–2-D and 3-D–3-D registration, as an arbitrary underlying graph structure can be readily employed (see Figure 11 for the result on 3-D intersubject brain registration).



[FIG10] The FastPD [126] results for (a) an image denoising and (b) a stereomatching problem. Each plot in (a) and (b) compares the corresponding running time per iteration of the above primal–dual algorithm with the α-expansion algorithm, which is a primal-based method (experiments conducted on a 1.6-GHz CPU).



[FIG11] A color-coded visualization of the surface distance between the warped and expert segmentation after (a) affine, (b) free-form deformation (FFD)-based [129], and (c) primal–dual-based registration for the Brain 1 data set. The color range is scaled to a maximum and minimum distance of 3 mm. The average surface distance (ASD) after registration for the gray matter is 1.66, 1.14, and 1.00 mm for the affine, FFD-based, and primal–dual method, respectively. For the white matter, the resulting ASD is 1.92, 1.31, and 1.06 mm for the affine, FFD-based, and primal–dual method, respectively. Note also that the FFD-based method is more than 30 times slower than the primal–dual approach.

Another application of primal–dual methods is in stereo reconstruction [130], where given as input a pair of left and right images I_L, I_R, we seek to estimate a function $u: \Omega \to \Gamma$ representing the depth $u(s)$ at a point s in the domain $\Omega \subset \mathbb{R}^2$ of the left image (here $\Gamma = [\upsilon_{\min}, \upsilon_{\max}]$ denotes the allowed depth range). To accomplish this, the following variational problem is proposed in [130]:

$$\underset{u}{\text{minimize}} \int_{\Omega} f(u(s), s)\, ds + \int_{\Omega} |\nabla u(s)|\, ds, \tag{52}$$

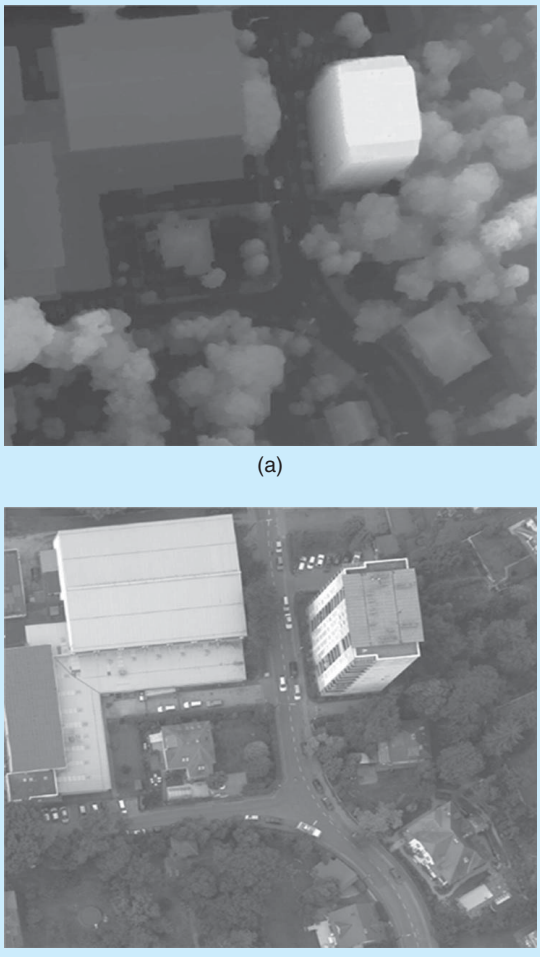
where $f(u(s), s)$ is a data term favoring different depth values by measuring the absolute intensity differences of respective patches projected in the two input images, and the second term is a total variation regularizer that promotes spatially smooth depth fields. Criterion (52) is nonconvex (due to the use of the data term f), but it turns out that there exists an equivalent convex formulation obtained by lifting the original problem to a higher-dimensional space, in which u is represented in terms of its level sets

$$\underset{\phi \in D}{\text{minimize}} \int_{\Sigma} (|\nabla \phi(s,\upsilon)| + f(s,\upsilon)|\partial_{\upsilon}\phi(s,\upsilon)|) ds d\upsilon. \quad (53)$$

In this formulation, $\Sigma = \Omega \times \Gamma$, $\phi: \Sigma \to \{0,1\}$ is a binary function such that $\phi(s,\upsilon)$ equals one if $u(s) > \upsilon$ and zero otherwise, and the feasible set is defined as $D = \{\phi: \Sigma \to \{0,1\} \,|\, (\forall s \in \Omega)\, \phi(s,\upsilon_{\min}) = 1,\ \phi(s,\upsilon_{\max}) = 0\}$. A convex relaxation of the latter problem is obtained by using $D' = \{\phi: \Sigma \to [0,1] \,|\, (\forall s \in \Omega)\, \phi(s,\upsilon_{\min}) = 1,\ \phi(s,\upsilon_{\max}) = 0\}$ instead of D. A discretized form of the resulting optimization problem can be solved with the algorithms described in the section "Methods Based on a Forward–Backward Approach." Figure 12 shows a sample result of this approach.

Recently, primal–dual approaches have also been developed for discrete optimization problems that involve higher-order terms [131]–[133]. They have been applied successfully to various tasks, for instance, in stereo matching [131]. In this case, apart from a data term that measures the similarity between the corresponding pixels in two images, a discontinuity-preserving smoothness prior of the form $\phi(s_1, s_2, s_3) = \min(|s_1 - 2s_2 + s_3|, \kappa)$ with $\kappa \in]0, +\infty[$ has been employed as a regularizer that penalizes depth surfaces of high curvature. Indicative stereo matching results from an algorithm based on the dual decomposition principle described in the section "Dual Decomposition" are shown in Figure 13.

It should be also mentioned that an advantage of all primal–dual algorithms (which is especially important for NP-hard problems) is that they also provide (for free) per-instance approximation bounds, specifying how far the cost of an estimated solution can be from the unknown optimal cost. This directly follows from the fact that these methods are computing both primal and dual solutions, which (in the case of a minimization task) provide, respectively, upper and lower limits to the true optimum. These approximation bounds are continuously updated throughout an algorithm's execution and, thus, can be directly used for assessing the performance of a primal–dual method with respect to any particular problem instance (and without essentially any extra computational cost). Moreover, often in practice, these sequences converge to a common value, which means that the corresponding estimated solutions are almost optimal (see, e.g., the charts in Figure 13).



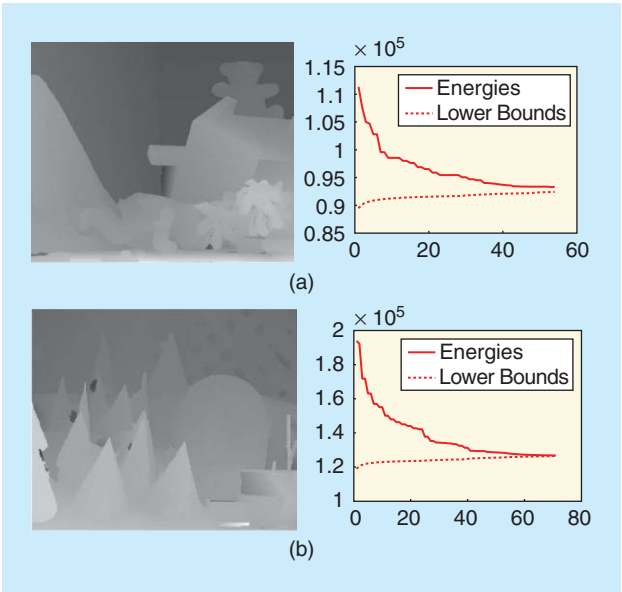
(a)

(b)

[FIG12] (a) An estimated depth map for a large aerial stereo data set of Graz using the primal–dual approach in [130]. (b) One of the images of the corresponding stereoscopic pair (of size 1,500 × 1,400).

CONCLUSIONS

In this article, we reviewed a number of primal–dual optimization methods, which can be employed for solving signal and image processing problems. The links existing between convex approaches and discrete ones were little explored in the literature, and one of the goals of this article is to put them in a unifying perspective. Although the presented algorithms have proved to be quite effective in numerous problems, there remains much room for extending their scope to other application fields and also for improving them so as to accelerate their convergence. In particular, the parameter choices in these methods may have a strong influence on the convergence speed, and it would be interesting to design automatic procedures for setting these parameters. Various techniques can also be devised for designing faster variants of these methods (e.g., preconditioning, activation of blocks of variables, combination with stochastic strategies, and distributed implementations). Another issue is the robustness to numerical errors, although it can be mentioned that most of the existing proximal algorithms are tolerant to summable errors. Concerning discrete optimization methods, we have shown that the key to success lies in tight relaxations of combinatorial NP-hard problems. Extending these methods to more challenging problems, e.g., those involving higher-order Markov fields or extremely large



[FIG13] The stereo matching results for (a) "Teddy" and (b) "Cones" when using a higher-order discontinuity preserving the smoothness prior. We show plots for the corresponding sequences of upper and lower bounds generated during the primal–dual method. Notice that these sequences converge to the same limit, meaning that the estimated solution converges to the optimal value.

label sets, appears to be of main interest in this area. More generally, developing primal–dual strategies that further bridge the gap between continuous and discrete approaches and solve other kinds of nonconvex optimization problems, such as those encountered in phase reconstruction or blind deconvolution, opens the way to appealing investigations. So, the floor is yours now to play with duality!

AUTHORS

Nikos Komodakis (nikos.komodakis@enpc.fr) received his Ph.D. degree in computer science (with highest honors) from the University of Crete, Greece, in 2006 and his Habilitation à Diriger des Recherches degree from the University Paris-Est, France, in 2013. He is an associate professor at Universite Paris-Est, École des Ponts ParisTech, and a research scientist at the Laboratoire d'Informatique Gaspard-Monge, Centre National de la Recherche Scientifique (CNRS). He is also an affiliated adjunct professor at the École Normale Supérieure de Cachan. He is currently an editorial board member of *International Journal of Computer Vision* and *Computer Vision and Image Understanding*. His research interests are in the areas of computer vision, image processing, machine learning, and medical imaging, and he has published numerous papers in the most prestigious journals and conferences from the above domains. He is a Member of the IEEE.

Jean-Christophe Pesquet (jean-christophe.pesquet@u-pem.fr) received his engineering degree from Supélec, Gif-sur-Yvette, France, in 1987; his Ph.D. degree from the University Paris-Sud (XI), Paris, France, in 1990; and the Habilitation à Diriger des Recherches from the University Paris-Sud in 1999. From 1991 to 1999, he was a maître de conférences at the University Paris-Sud and a research scientist at the Laboratoire des Signaux et Systèmes, Centre National de la Recherche Scientifique (CNRS), Gif-sur-Yvette. He is currently a professor (classe exceptionnelle) at the Université de Paris-Est Marne-la-Vallée, France, and the deputy director of the Laboratoire d'Informatique of the university (UMR–CNRS 8049). He is a Fellow of the IEEE.

REFERENCES

[1] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Nashua, NH: Athena Scientific, 2004.

[2] A. Blake, P. Kohli, and C. Rother, *Markov Random Fields for Vision and Image Processing*. Cambridge, MA: MIT Press, 2011.

[3] S. Sra, S. Nowozin, and S. J. Wright, *Optimization for Machine Learning*. Cambridge, MA: MIT Press, 2012.

[4] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*,. San Diego, CA: Academic Press, 2015.

[5] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Found. Trends Machine Learn*., vol. 4, no. 1, pp. 1–106, 2012.

[6] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge Univ. Press, 2004.

[8] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York: Springer, 2011.

[9] J. J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bull. Soc. Math. France*, vol. 93, no. 3, pp. 273–299, 1965.

[10] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. New York: Springer-Verlag, 2011, pp. 185–212.

[11] P. L. Combettes and J.-C. Pesquet, "Primal–dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued Var. Anal*., vol. 20, no. 2, pp. 307–330, June 2012.

[12] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, 3rd ed. London: Springer-Verlag, 2009.

[13] N. Komodakis, G. Tziritas, and N. Paragios, "Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal–dual strategies," *Comput. Vis. Image Understand*., vol. 112, no. 2, pp. 14–29, Oct. 2008.

[14] N. Komodakis, N. Paragios, and G. Tziritas, "MRF energy minimization and beyond via dual decomposition," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 33, no. 3, pp. 531–552, Jan. 2011.

[15] M. Wainwright, T. Jaakkola, and A. Willsky, "MAP estimation via agreement on trees: message-passing and linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 11, pp. 3697–3717, Nov. 2005.

[16] C. Wang, N. Komodakis, and N. Paragios, "Markov random field modeling, inference & learning in computer vision & image understanding: A survey," *Comput. Vis. Image Understand.*, vol. 117, no. 11, pp. 1610–1627, Nov. 2013.

[17] V. V. Vazirani, *Approximation Algorithms*. New York, NY: Springer-Verlag, 2001.

[18] D. S. Hochbaum, Ed., *Approximation Algorithms for NP-Hard Problems*. Boston, MA: PWS Publishing, 1997.

[19] M. Fortin and R. Glowinski, Eds., *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*. Amsterdam: Elsevier Science, 1983.

[20] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Machine Learn*., vol. 8, no. 1, pp. 1–122, 2011.

[21] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation, Vol. I: Basic Theory* (Series of Comprehensive Studies in Mathematics, vol. 330). Berlin, Heidelberg: Springer-Verlag, 2006.

[22] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim*., vol. 1, no. 3, pp. 123–231, 2013.

[23] R Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Stat. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

[24] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Mag*., vol. 25, no. 2, pp. 21–30, Mar. 2008.

[25] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.

[26] R. I. Boţ, *Conjugate Duality in Convex Optimization* (Lecture Notes in Economics and Mathematical Systems, vol. 637). Berlin, Heidelberg: Springer-Verlag, 2010.

[27] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*. Nashua, NH: Athena Scientific, 1997.

[28] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite elements approximations," *Comput. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.

[29] J.-F. Giovannelli and A. Coulais, "Positive deconvolution for superimposed extended source and point sources," *Astron. Astrophys.*, vol. 439, no. 1, pp. 401–412, Aug. 2005.

[30] T. Goldstein and S. Osher, "The split Bregman method for ℓ_1-regularized problems," *SIAM J. Imaging Sci.*, vol. 2, no. 2, pp. 323–343, 2009.

[31] M. A. T. Figueiredo and R. D. Nowak, "Deconvolution of Poissonian images using variable splitting and augmented Lagrangian optimization," in *Proc. IEEE Workshop Statistical Signal Processing,* Cardiff, United Kingdom, Aug. 31–Sept. 3, 2009, pp. 733–736.

[32] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Trans. Image Processing*, vol. 19, no. 12, pp. 3133–3145, Dec. 2010.

[33] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Processing*, vol. 20, no. 3, pp. 681–695, Mar. 2011.

[34] Q. Tran-Dinh and V. Cevher. (2014). A primal–dual algorithmic framework for constrained convex minimization. [Online]. Available: http://arxiv.org/pdf/1406.5403.pdf

[35] M. Hong and Z.-Q. Luo. (2013). On the linear convergence of the alternating direction method of multipliers. [Online]. Available: http://arxiv.org/abs/1208.3922

[36] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1–3, pp. 293–318, 1992.

[37] P. L. Combettes and J.-C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE J. Select. Topics Signal Processing*, vol. 1, no. 4, pp. 564–574, Dec. 2007.

[38] R. I. Boţ and C. Hendrich, "A Douglas-Rachford type primal–dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators," *SIAM J. Optim.*, vol. 23, no. 4, pp. 2541–2565, Dec. 2013.

[39] G. Chen and M. Teboulle, "A proximal-based decomposition method for convex minimization problems," *Math. Program.*, vol. 64, no. 1–3, pp. 81–101, 1994.

[40] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.

[41] A. Nedić and A. Ozdaglar, "Subgradient methods for saddle-point problems," *J. Optim. Theory Appl.*, vol. 142, no. 1, pp. 205–228, 2009.

[42] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.

[43] D. Davis. (2014). Convergence rate analysis of the Forward-Douglas-Rachford splitting scheme. [Online]. Available: http://arxiv.org/abs/1410.2654

[44] L. Condat, "A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, Aug. 2013.

[45] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comput. Math.*, vol. 38, no. 3, pp. 667–681, Apr. 2013.

[46] A. Chambolle and T. Pock, "A first-order primal–dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[47] E. Esser, X. Zhang, and T. Chan, "A general framework for a class of first order primal–dual algorithms for convex optimization in imaging science," *SIAM J. Imaging Sci.*, vol. 3, no. 4, pp. 1015–1046, 2010.

[48] B. He and X. Yuan, "Convergence analysis of primal–dual algorithms for a saddle-point problem: from contraction perspective," *SIAM J. Imaging Sci.*, vol. 5, no. 1, pp. 119–149, 2012.

[49] T. Pock and A. Chambolle, "Diagonal preconditioning for first order primal–dual algorithms in convex optimization," in *Proc. IEEE Int. Conf. Computer Vision*, Barcelona, Spain, Nov. 6–13, 2011, pp. 1762–1769.

[50] P. L. Combettes and B. C. Vũ, "Variable metric forward-backward splitting with applications to monotone inclusions in duality," *Optimization*, vol. 63, no. 9, pp. 1289–1318, Sept. 2014.

[51] T. Goldstein, E. Esser, and R. Baraniuk. (2013). Adaptive primal–dual hybrid gradient methods for saddle-point problems. [Online]. Available: http://arxiv.org/abs/1305.0546

[52] P. L. Combettes, L. Condat, J.-C. Pesquet, and B. C. Vũ, "A forward-backward view of some primal–dual optimization methods in image recovery," in *Proc. Int. Conf. Image Processing*, Paris, France, 27–30 Oct. 2014, pp. 4141–4145.

[53] J. Liang, J. Fadili, and G. Peyré. (2014). Convergence rates with inexact nonexpansive operators. [Online]. Available: http://arxiv.org/abs/1404.4837

[54] I. Loris and C. Verhoeven, "On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty," *Inverse Problems*, vol. 27, no. 12, pp. 125007, 2011.

[55] P. Chen, J. Huang, and X. Zhang, "A primal–dual fixed point algorithm for convex separable minimization with applications to image restoration," *Inverse Problems*, vol. 29, no. 2, pp. 025011, 2013.

[56] P. L. Combettes, D. Dũng, and B. C. Vũ, "Dualization of signal recovery problems," *Set-Valued Var. Anal.*, vol. 18, pp. 373–404, Dec. 2010.

[57] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame-based inverse problems," *Inverse Problems*, vol. 23, no. 4, pp. 1495–1518, June 2007.

[58] A. Jezierska, E. Chouzenoux, J.-C. Pesquet, and H. Talbot, "A primal–dual proximal splitting approach for restoring data corrupted with Poisson-Gaussian noise," in *Proc. Int. Conf. Acoustics, Speech Signal Processing*, Kyoto, Japan, 25–30 Mar. 2012, pp. 1085–1088.

[59] P. Tseng, "A modified forward-backward splitting method for maximal monotone mappings," *SIAM J. Control Optim.*, vol. 38, pp. 431–446, 2000.

[60] L. M. Briceño-Arias and P. L. Combettes, "A monotone + skew splitting model for composite monotone inclusions in duality," *SIAM J. Optim.*, vol. 21, no. 4, pp. 1230–1250, Oct. 2011.

[61] P. L. Combettes, "Systems of structured monotone inclusions: duality, algorithms, and applications," *SIAM J. Optim.*, vol. 23, no. 4, pp. 2420–2447, Dec. 2013.

[62] R. I. Boţ and C. Hendrich, "Convergence analysis for a primal–dual monotone + skew splitting algorithm with applications to total variation minimization," *J. Math. Imaging Vision*, vol. 49, no. 3, pp. 551–568, 2014.

[63] A. Alotaibi, P. L. Combettes, and N. Shahzad, "Solving coupled composite monotone inclusions by successive Fejér approximations of their Kuhn-Tucker set," *SIAM J. Optim.*, to be published, vol. 24, no. 4, pp. 2076–2095, Dec. 2014.

[64] S. R. Becker and P. L. Combettes, "An algorithm for splitting parallel sums of linearly composed monotone operators, with applications to signal recovery," *Nonlinear Convex Anal.*, vol. 15, no. 1, pp. 137–159, Jan. 2014.

[65] P. L. Combettes and J.-C. Pesquet, "A proximal decomposition method for solving convex variational inverse problems," *Inverse Problems*, vol. 24, no. 6, p. 065014, Dec. 2008.

[66] J.-C. Pesquet and N. Pustelnik, "A parallel inertial proximal optimization method," *Pac. J. Optim.*, vol. 8, no. 2, pp. 273–305, Apr. 2012.

[67] S. Setzer, G. Steidl, and T. Teuber, "Deblurring Poissonian images by split Bregman techniques," *J. Visual Commun. Image Represent.*, vol. 21, no. 3, pp. 193–199, Apr. 2010.

[68] V. Kolmogorov, "Generalized roof duality and bisubmodular functions," in *Proc. Annu. Conf. Neural Information Processing Systems,* Vancouver, Canada, 6–9 Dec. 2010, pp. 1144–1152.

[69] F. Kahl and P. Strandmark, "Generalized roof duality," *Discrete Appl. Math.*, vol. 160, no. 16–17, pp. 2419–2434, 2012.

[70] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. Englewoods Cliffs, NJ: Prentice Hall, 1982.

[71] N. Komodakis and G. Tziritas, "Approximate labeling via graph-cuts based on linear programming," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1436–1453, Aug. 2007.

[72] N. Komodakis, N. Paragios, and G. Tziritas, "MRF optimization via dual decomposition: Message-passing revisited," in *Proc. IEEE Int. Conf. Computer Vision*, Rio de Janeiro, Brazil, 14–21 Oct. 2007, pp. 1–8.

[73] C. Chekuri, S. Khanna, J. Naor, and L. Zosin, "Approximation algorithms for the metric labeling problem via a new linear programming formulation," in *Proc. 12th Annu. ACM-SIAM Symp. Discrete Algorithms*, Washington, DC, USA, 7–9 Jan. 2001, pp. 109–118.

[74] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann, 1988.

[75] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1568–1583, Aug. 2006.

[76] T. Werner, "A linear programming approach to max-sum problem: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 7, pp. 1165–1179, July 2007.

[77] A. Globerson and T. Jaakkola, "Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations," in *Proc. Annu. Conf. Neural Information Processing Systems*, Vancouver and Whistler, Canada, 3–6 Dec. 2007, pp. 553–560.

[78] C. Yanover, T. Talya Meltzer, and Y. Weiss, "Linear programming relaxations and belief propagation—an empirical study," *J. Mach. Learn. Res.*, vol. 7, pp. 1887–1907, Sept. 2006.

[79] T. Hazan and A. Shashua, "Norm-product belief propagation: Primal–dual message-passing for approximate inference," *IEEE Trans. Inform. Theory*, vol. 56, no. 12, pp. 6294–6316, Dec. 2010.

[80] S. Jegelka, F. Bach, and S. Sra, "Reflection methods for user-friendly submodular optimization," in *Proc. Ann. Conf. Neural Information Processing Systems*, Lake Tahoe, NV, USA, 5–10 Dec. 2013, pp. 1313–1321.

[81] N. N. Schraudolph, "Polynomial-time exact inference in NP-hard binary MRFs via reweighted perfect matching," in *Proc. 13th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010, pp. 717–724.

[82] A. Osokin, D. Vetrov, and V. Kolmogorov, "Submodular decomposition framework for inference in associative Markov networks with global constraints," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Colorado Springs, USA, 21–23 June 2011, pp. 1889–1896.

[83] J. Yarkony, R. Morshed, A. T. Ihler, and C. Fowlkes, "Tightening MRF relaxations with planar subproblems," in *Proc. Conf. Uncertainty in Artificial Intelligence*, Barcelona, Spain, 14–17 July 2011, pp. 770–777.

[84] D. Sontag, T. Meltzer, A. Globerson, Y. Weiss, and T. Jaakkola, "Tightening LP relaxations for MAP using message passing," in *Proc. Conf. Uncertainty in Artificial Intelligence*, Helsinki, Finland, 9–12 July 2008, pp. 656–664.

[85] N. Komodakis and N. Paragios, "Beyond loose LP-relaxations: Optimizing MRFs by repairing cycles," in *Proc. European Conf. Computer Vision*, Marseille, France, 12–18 Oct. 2008, pp. 806–820.

[86] E. Boros and P. L. Hammer, "Pseudo-Boolean optimization," *Discrete Appl. Math*., vol. 123, no. 1–3, pp. 155–225, 2002.

[87] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 26, no. 2, pp. 147–159, Feb. 2004.

[88] V. Jojic, S. Gould, and D. Koller, "Fast and smooth: Accelerated dual decomposition for MAP inference," in *Proc. Int. Conf. Machine Learning*, Haifa, Israel, 21–24 June 2010, pp. 503–510.

[89] B. Savchynskyy, J. H. Kappes, S. Schmidt, and C. Schnörr, "A study of Nesterov's scheme for Lagrangian decomposition and MAP labeling," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Colorado Springs, USA, 21–23 June 2011, pp. 1817–1823.

[90] B. Savchynskyy, S. Schmidt, J. H. Kappes, and C. Schnörr, "Efficient MRF energy minimization via adaptive diminishing smoothing," in *Proc. Conf. Uncertainty in Artificial Intelligence*, Catalina Island, USA, 15–17 Aug. 2012, pp. 746–755.

[91] N. Pustelnik, C. Chaux, and J.-C. Pesquet, "Parallel ProXimal Algorithm for image restoration using hybrid regularization," *IEEE Trans. Image Processing*, vol. 20, no. 9, pp. 2450–2462, Sept. 2011.

[92] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," *SIAM J. Imaging Sci*., vol. 3, no. 3, pp. 253–276, 2010.

[93] S. Bonettini and V. Ruggiero, "On the convergence of primal–dual hybrid gradient algorithms for total variation image restoration," *J. Math. Imaging Vision*, vol. 44, no. 3, pp. 236–253, 2012.

[94] A. Repetti, E. Chouzenoux, and J.-C. Pesquet, "A penalized weighted least squares approach for restoring data corrupted with signal-dependent noise," in *Proc. European Signal and Image Processing Conf*., Bucharest, Romania, 27–31 Aug. 2012, pp. 1553–1557.

[95] S. Harizanov, J.-C. Pesquet, and G. Steidl, "Epigraphical projection for solving least squares Anscombe transformed constrained optimization problems," in *Proc. 4th Int. Conf. Scale-Space and Variational Methods in Computer Vision*, A. Kuijper, K. Bredies, T. Pock, and H. Bischof, Eds., Schloss Seggau, Leibnitz, Austria, 2–6 June 2013 (Lecture Notes in Computer Science, vol. 7893). Berlin: Springer-Verlag, pp. 125–136.

[96] T. Teuber, G. Steidl, and R.-H. Chan, "Minimization and parameter estimation for seminorm regularization models with I-divergence constraints," *Inverse Problems*, vol. 29, pp. 035007, Mar. 2013.

[97] M. Burger, A. Sawatzky, and G. Steidl. (2014). First order algorithms in variational image processing. [Online]. Available: http://arxiv.org/abs/1412.4237

[98] C. Couprie, L. Grady, L. Najman, J.-C. Pesquet, and H. Talbot, "Dual constrained TV-based regularization on graphs," *SIAM J. Imaging Sci*., vol. 6, no. 3, pp. 1246–1273, 2013.

[99] L. Chaari, J.-C. Pesquet, A. Benazza-Benyahia, and Ph. Ciuciu, "A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging," *Med. Image Anal*., vol. 15, no. 2, pp. 185–201, Apr. 2011.

[100] A. Florescu, E. Chouzenoux, J.-C. Pesquet, Ph. Ciuciu, and S. Ciochina, "A Majorize-Minimize Memory Gradient method for complex-valued inverse problems," *Signal Process*. (Special issue on Image Restoration and Enhancement: Recent Advances and Applications), vol. 103, pp. 285–295, Oct. 2014.

[101] S. Mahadevan, B. Liu, P. Thomas, W. Dabney, S. Giguere, N. Jacek, I. Gemp, and J. Liu, "Proximal reinforcement learning: A new theory of sequential decision making in primal–dual spaces," [Online]. Available: http://http://arxiv.org/abs/1405.6757.

[102] S. Ono, M. Yamagishi, and I. Yamada, "A sparse system identification by using adaptively-weighted total variation via a primal–dual splitting approach," in *Proc. Int. Conf. Acoustics, Speech Signal Processing*, Vancouver, Canada, 26–31 May 2013, pp. 6029–6033.

[103] I. Bayram and O. D. Akyildiz, "Primal–dual algorithms for audio decomposition using mixed norms," *Signal Image Video Process*., vol. 8, no. 1, pp. 95–110, Jan. 2014.

[104] N. Papadakis, G. Peyré, and E. Oudet, "Optimal transport with proximal splitting," *SIAM J. Imaging Sci*., vol. 7, no. 1, pp. 212–238, 2014.

[105] N. Pustelnik, P. Borgnat, and P. Flandrin, "Empirical Mode Decomposition revisited by multicomponent nonsmooth convex optimization," *Signal Process*., vol. 102, pp. 313–331, Sept. 2014.

[106] M.-Q. Pham, C. Chaux, L. Duval, and J.-C. Pesquet, "Sparse template-based adaptive filtering with a primal–dual proximal algorithm: Application to seismic multiple removal," *IEEE Trans. Signal Processing*, vol. 62, no. 16, pp. 4256–4269, Aug. 2014.

[107] G. Moerkotte, M. Montag, A. Repetti, and G. Steidl. (2015). Proximal operator of quotient functions with application to a feasibility problem in query optimization. *J. Comput. Appl. Math*. [Online]. Available: http://hal.archives-ouvertes.fr/docs/00/94/24/53/PDF/Quotient Functions.pdf

[108] Z. J. Towfic and A. H. Sayed. (2015). Stability and performance limits of adaptive primal–dual networks. *IEEE Trans. Signal Processing*, [Online]. Available: http://arxiv.org/pdf/1408.3693.pdf

[109] J. H. Kappes, B. Andres, F. A. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B. X. Kausler, et al., "A comparative study of modern inference techniques for discrete energy minimization problems," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Portland, OR, USA, 25–27 June 2013, pp. 1328–1335.

[110] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 30, no. 6, pp. 1068–1080, June 2008.

[111] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[112] P. Strandmark, F. Kahl, and T. Schoenemann, "Parallel and distributed vision algorithms using dual decomposition," *Computer Vision and Image Understanding*, vol. 115, no. 12, pp. 1721–1732, 2011.

[113] S. Vicente, V. Kolmogorov, and C. Rother, "Joint optimization of segmentation and appearance models," in *Proc. IEEE Int. Conf. Computer Vision*, Kyoto, Japan, 29 Sept.–2 Oct. 2009, pp. 755–762.

[114] T. Pock, A. Chambolle, D. Cremers, and H. Bischof, "A convex relaxation approach for computing minimal partitions," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Miami, FL, USA, 20–25 June 2009, pp. 810–817.

[115] O. J. Woodford, C. Rother, and V. Kolmogorov, "A global perspective on MAP inference for low-level vision," in *Proc. IEEE Int. Conf. Computer Vision*, Kyoto, Japan, 27 Sept.–4 Oct. 2009, pp. 2319–2326.

[116] D. Cremers, P. Thomas, K. Kolev, and A. Chambolle, "Convex relaxation techniques for segmentation, stereo and multiview reconstruction," in *Markov Random Fields for Vision and Image Processing*, A. Blake, P. Kohli, and C. Rother, Eds. Boston: MIT, 2011, pp. 185–200.

[117] C. Hane, C. Zach, A. Cohen, R. Angst, and M. Pollefeys, "Joint 3D scene reconstruction and class segmentation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Portland, OR, USA, 25–27 June 2013, pp. 97–104.

[118] L. Torresani, V. Kolmogorov, and C. Rother, "A dual decomposition approach to feature correspondence," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 35, no. 2, pp. 259–271, Feb. 2013.

[119] Y Zeng, C. Wang, Y. Wang, X. Gu, D. Samaras, and N. Paragios, "Intrinsic dense 3D surface tracking," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Colorado Springs, CO, 21–23 June 2011, pp. 1225–1232.

[120] B. Glocker, N. Paragios, N. Komodakis, G. Tziritas, and N. Navab, "Optical flow estimation with uncertainties through dynamic MRFs," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Anchorage, AK, USA, 23–28 June 2008, pp. 1–8.

[121] M. P. Kumar and D. Koller, "Efficiently selecting regions for scene understanding," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco, CA, USA, 13–18 June 2010, pp. 3217–3224.

[122] N. Komodakis and N. Paragios, "MRF-based blind image deconvolution," in *Proc. Asian Conf. Computer Vision*, Daejeon, Korea, 5–9 Nov. 2012, pp. 361–374.

[123] V. Kolmogorov and A. Shioura, "New algorithms for convex cost tension problem with application to computer vision," *Discrete Optim*., vol. 6, no. 4, pp. 378–393, 2009.

[124] D. Batra, P. Yadollahpour, A. Guzman-Rivera, and G. Shakhnarovich, "Diverse m-best solutions in Markov random fields," in *Proc. European Conf. Computer Vision*, Florence, Italy, 7–13 Oct. 2012, pp. 1–16.

[125] D. Tsai, M. Flagg, A. Nakazawa, and J. M. Rehg, "Motion coherent tracking using multi-label MRF optimization," *Int. J. Comp. Vis*., vol. 100, no. 2, pp. 190–202, Nov. 2012.

[126] N. Komodakis, G. Tziritas, and N. Paragios, "Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal–dual strategies," *Comput. Vis. Image Understand*., vol. 112, no. 1, pp. 14–29, Oct. 2008.

[127] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios, "Dense image registration through MRFs and efficient linear programming," *Med. Image Anal*., vol. 12, no. 6, pp. 731–741, Dec. 2008.

[128] B. Glocker, A. Sotiras, N. Paragios, and N. Komodakis, "Deformable medical image registration: setting the state of the art with discrete methods," *Annu. Rev. Biomed. Eng*., vol. 13, pp. 219–244, Aug. 2011.

[129] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imag*., vol. 18, no. 8, pp. 712–721, Aug. 1999.

[130] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers, "A convex formulation of continuous multi-label problems," in *Proc. European Conf. Computer Vision*, Marseille, France, 12–18 Oct. 2008, pp. 792–805.

[131] N. Komodakis and N. Paragios, "Beyond pairwise energies: Efficient optimization for higher-order MRFs," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Miami, FL, USA, 20–25 June 2009, pp. 2985–2992.

[132] A. Fix, C. Wang, and R. Zabih, "A primal–dual method for higher-order multilabel Markov random fields," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Columbus, OH, USA, 23–28 June 2014, pp. 1138–1145.

[133] C. Arora, S. Banerjee, P. Kalra, and S. N. Maheshwari, "Generic cuts: An efficient algorithm for optimal inference in higher order MRF-MAP," in *Proc. European Conf. Computer Vision*, Florence, Italy, 7–13 Oct. 2012, pp. 17–30.

[SP]

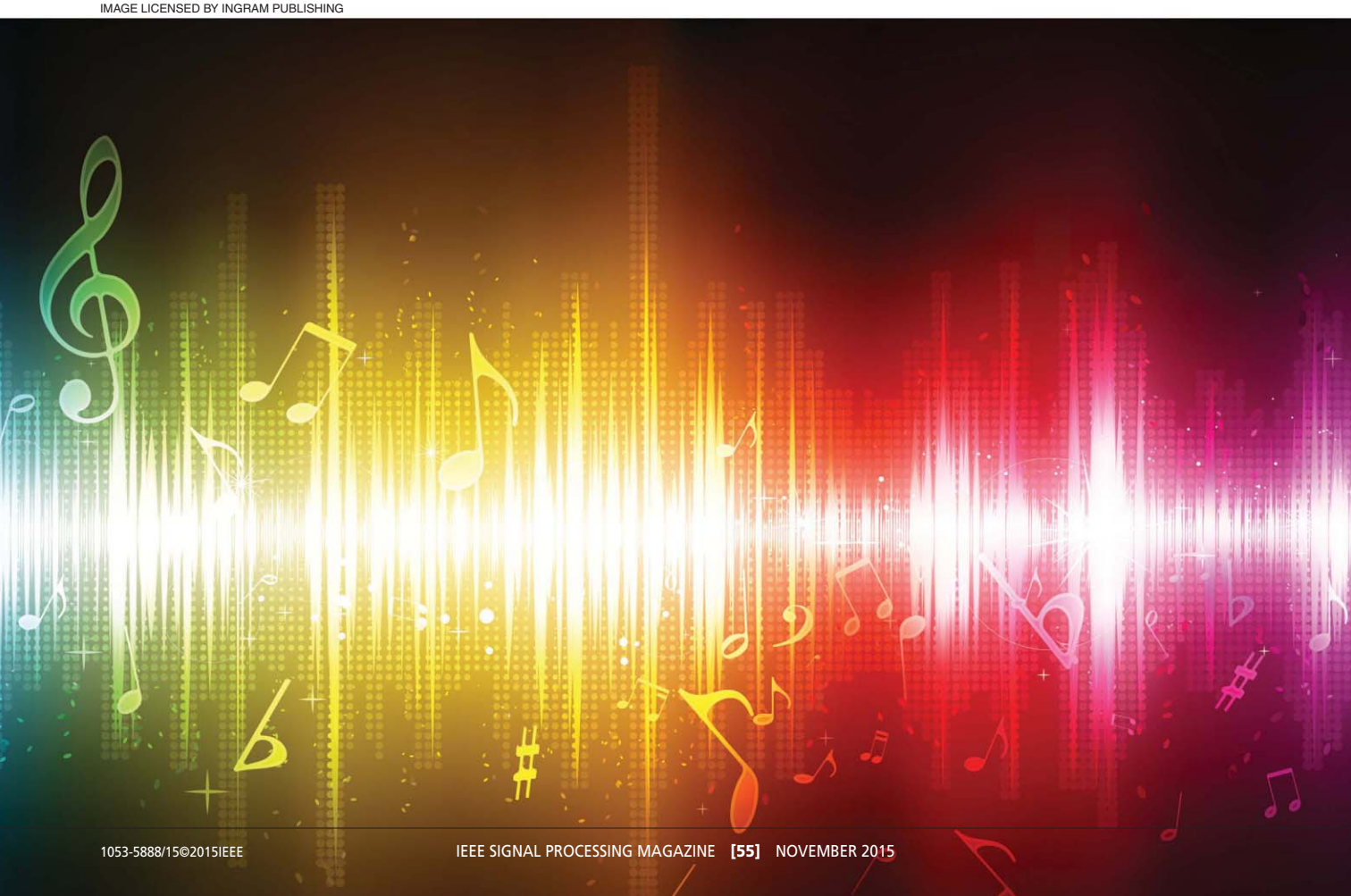
Expression Control in Singing Voice Synthesis

Features, approaches, evaluation, and challenges

Martí Umbert, Jordi Bonada, Masataka Goto, Tomoyasu Nakano, and Johan Sundberg

In the context of singing voice synthesis, expression control manipulates a set of voice features related to a particular emotion, style, or singer. Also known as *performance modeling*, it has been approached from different perspectives and for different purposes, and different projects have shown a wide extent of applicability. The aim of this article is to provide an overview of approaches to expression control in singing voice synthesis. We introduce some musical applications that use singing voice synthesis techniques to justify the need for an accurate control of expression. Then, expression is defined and related to speech and instrument performance modeling. Next, we present the commonly studied set of voice parameters that can change

Digital Object Identifier 10.1109/MSP.2015.2424572
Date of publication: 13 October 2015



1053-5888/15©2015IEEE

[TABLE 1] RESEARCH PROJECTS USING SINGING VOICE SYNTHESIS TECHNOLOGIES.

PROJECT	WEBSITE
CANTOR	HTTP://WWW.VIRSYN.DE
CANTOR DIGITALIS	HTTPS://CANTORDIGITALIS.LIMSI.FR/
CHANTER	HTTPS://CHANTER.LIMSI.FR
FLINGER	HTTP://WWW.CSLU.OGI.EDU/TTS/FLINGER
LYRICOS	HTTP://WWW.CSLU.OGI.EDU/TTS/DEMOS
ORPHEUS	HTTP://WWW.ORPHEUS-MUSIC.ORG/V3
SINSY	HTTP://WWW.SINSY.JP
SYMPHONIC CHOIRS VIRTUAL INSTRUMENT	HTTP://WWW.SOUNDSONLINE.COM/SYMPHONIC-CHOIRS
VOCALISTENER	HTTPS://STAFF.AIST.GO.JP/T.NAKANO/VOCALISTENER
VOCALISTENER (PRODUCT VERSION)	HTTP://WWW.VOCALOID.COM/LINEUP/VOCALIS
VOCALISTENER2	HTTPS://STAFF.AIST.GO.JP/T.NAKANO/VOCALISTENER2
VOCALOID	HTTP://WWW.VOCALOID.COM
VOCAREFINER	HTTPS://STAFF.AIST.GO.JP/T.NAKANO/VOCAREFINER
VOCAWATCHER	HTTPS://STAFF.AIST.GO.JP/T.NAKANO/VOCAWATCHER

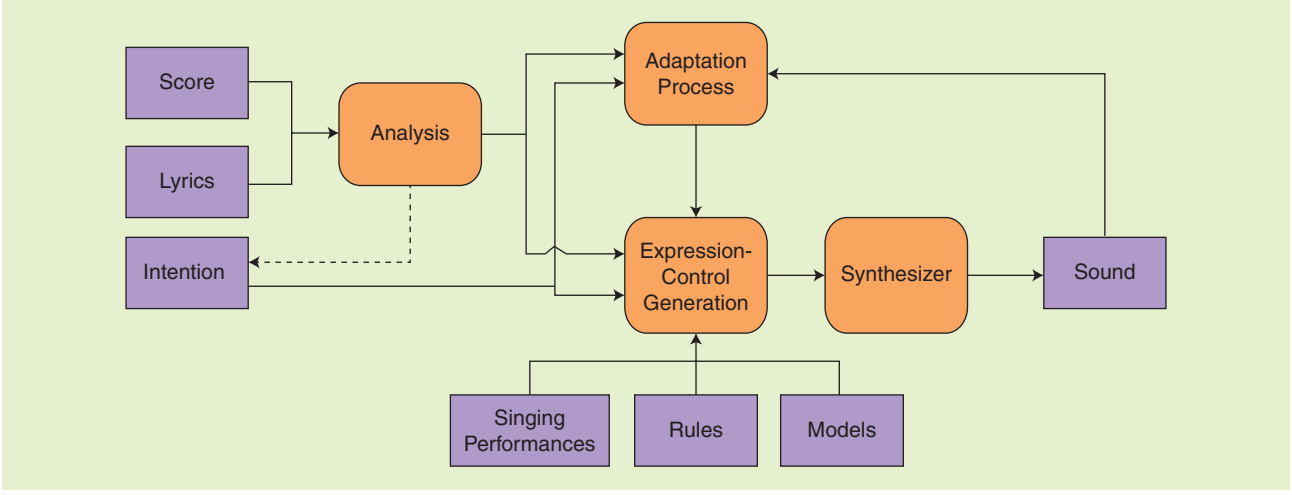
perceptual aspects of synthesized voices. After that, we provide an up-to-date classification, comparison, and description of a selection of approaches to expression control. Then, we describe how these approaches are currently evaluated and discuss the benefits of building a common evaluation framework and adopting perceptually-motivated objective measures. Finally, we discuss the challenges that we currently foresee.

SINGING VOICE SYNTHESIS SYSTEMS

In recent decades, several applications have shown how singing voice synthesis technologies can be of interest for composers [1], [2]. Technologies for the manipulation of voice features have been increasingly used to enhance tools for music creation and postprocessing, singing a live performance, to imitate a singer, and even to generate voices that are difficult to produce naturally (e.g., castrati). More examples can be found with pedagogical purposes or as tools to identify perceptually relevant voice properties [3]. These applications of the so-called music information research field may have a great impact on the way we interact with music [4]. Examples of research projects using singing voice synthesis technologies are listed in Table 1.

The generic framework of these systems is represented in Figure 1, based on [5]. The input may consist of the score (e.g., the note sequence, contextual marks related to loudness, or note transitions), lyrics, and the intention (e.g., the style or emotion). The intention may be derived from the lyrics and score content (shown by the dashed line). The input may be analyzed to get the phonetic transcription, the alignment with a reference performance, or contextual data. The expression control generation block represents the implicit or explicit knowledge of the system as a set of reference singing performances, a set of rules, or statistical models. Its output is used by the synthesizer to generate the sound, which may be used iteratively to improve the expression controls.

A key element of such technologies is the singer voice model [1], [2], [6], although due to space constraints, it is not described here in detail. For the purpose of this article, it is more interesting to classify singing synthesis systems with respect to the control parameters. As shown in Table 2, those systems are classified into model-based and concatenative synthesizers. While, in signal models, the control parameters are mostly related to a perception perspective, in physical models, these are related to physical aspects of the vocal organs. In concatenative synthesis, a cost criterion is used to retrieve sound segments (called *units*) from a corpus that are then transformed and concatenated to generate the output utterance. Units may cover a fixed number of linguistic units, e.g., diphones that cover the transition between two phonemes or a more flexible and wider scope. In this case, control parameters are also related to perceptual aspects.



[FIG1] Generic framework blocks for expression control.

Within the scope of this review, we focus on the perceptual aspects of the control parameters, which are used to synthesize expressive performances by taking a musical score, lyrics, or an optional human performance as the input. However, this article does not discuss voice conversion and morphing in which input voice recordings are analyzed and transformed [7], [8].

EXPRESSION IN MUSICAL PERFORMANCE AND SINGING

Expression is an intuitive aspect of a music performance, but it is complex to define. In [5, p. 2], it is viewed as "the strategies and changes which are not marked in a score but which performers apply to the music." In [9, p. 1], expression is "the added value of a performance and is part of the reason that music is interesting to listen to and sounds alive." A complete definition is given in [10, p. 150], relating the liveliness of a score to "the artist's understanding of the structure and 'meaning' of a piece of music, and his/her (conscious or unconscious) expression of this understanding via expressive performance." From a psychological perspective, Juslin [11, p. 276] defines it as "a set of perceptual qualities that reflect psychophysical relationships between 'objective' properties of the music, and 'subjective' impressions of the listener."

Expression has a key impact on the perceived quality and naturalness. As pointed out by Ternström [13], "even a single sine wave can be expressive to some degree if it is expertly controlled in amplitude and frequency." Ternström says that musicians care more about instruments being adequately expressive than sounding natural. For instance, in Clara Rockmore's performance of *Vocalise* by Sergei Vasilyevich Rachmaninoff, a skillfully controlled Theremin expresses her intentions to a high degree (all cited sounds have been collected and shown online; see [51]), despite the limited degrees of freedom.

In the case of the singing voice, achieving a realistic sound synthesis implies controlling a wider set of parameters than just the amplitude and frequency. These parameters can be used by a singing voice synthesizer or to transform a recording. From a psychological perspective, pitch contour, vibrato features, intensity contour, tremolo, phonetic timing, and others related to timbre are the main control parameters that are typically used to transmit a message with a certain mood or emotion [12] and shaped by a musical style [14]. These are described in detail in the section "Singing Voice Performance Features."

Nominal values for certain parameters can be inferred from the musical score through the note's pitch, dynamics, and duration as well as its articulation, such as staccato or legato marks. However, these values are not intrinsically expressive per se. In other words, expression contributes to the differences between these values and a real performance. Different strategies for generating expression controls are explained in the section "Expression-Control Approaches."

It is important to note that there is more than one acceptable expressive performance for a given song [1], [3], [15]. Such variability complicates the evaluation and comparison of different expression-control approaches. This issue is tackled in the "Evaluation" section. Besides singing, expression has been studied in speech and instrumental music performance.

CONNECTION TO SPEECH AND INSTRUMENTAL MUSICAL PERFORMANCE

There are several common aspects of performing expressively through singing voice, speech, and musical instruments. In speech, the five acoustic attributes of prosody have been widely studied [16], for instance, to convey emotions [17]. The most studied attribute is the fundamental frequency (F0) of the voice source signal. Timing is the acoustic cue of rhythm, and it is a rather complex attribute given the number of acoustic features to which it is related [16, p. 43]. Other attributes are intensity, voice quality (related to the glottal excitation), and articulation (largely determined by the phonetic context and speech rate).

Expressive music performance with instruments has also been widely studied. Several computational models are reviewed in [18, p. 205], such as the KTH model, which is based "on performance rules that predict the timing, dynamics, and articulation from local musical context." The Todd model links the musical structure to a performance with simple rules like measurements of human performances. The Mazzola model analyzes musical structure features such as tempo and melody and iteratively modifies the expressive parameters of a synthesized performance. Finally, a machine-learning model discovers patterns within a large amount of data; it focuses, for instance, on timing, dynamics, and more abstract structures like phrases and manipulates them via tempo, dynamics, and articulation. In [5], 30 more systems are classified into nonlearning methods, linear regression, artificial neural networks, and rule-/case-based learning models, among others.

In this review, we adopt a signal processing perspective to focus on the acoustic cues that convey a certain emotion or evoke a singing style in singing performances. As mentioned in [12, p. 799], "vocal expression is the model on which musical

[TABLE 2] SINGING VOICE SYNTHESIS SYSTEMS AND CONTROL PARAMETERS.

	SINGING SYNTHESIS SYSTEMS			
	MODEL-BASED SYNTHESIS		CONCATENATIVE SYNTHESIS	
	SIGNAL MODELS	PHYSICAL MODELS	FIXED LENGTH UNITS	NONUNIFORM LENGTH UNITS
PARAMETERS	F0, RESONANCES (CENTER FREQUENCY AND BANDWIDTH), SINUSOID FREQUENCY, PHASE, AMPLITUDE, GLOTTAL PULSE SPECTRAL SHAPE, AND PHONETIC TIMING	VOCAL APPARATUS-RELATED PARAMETERS (TONGUE, JAW, VOCAL TRACT LENGTH AND TENSION, SUBGLOTTAL AIR PRESSURE, AND PHONETIC TIMING)	F0, AMPLITUDE, TIMBRE, AND PHONETIC TIMING	

expression is based," which highlights the topic relevance for both the speech and the music performance community. Since there is room for improvement, the challenges that we foresee are described in the "Challenges" section.

SINGING VOICE PERFORMANCE FEATURES

In the section "Expression in Musical Performance and Singing," we introduced a wide set of low-level parameters for singing voice expression. In this section, we relate them to other musical elements. Then, the control parameters are described, and finally, we illustrate them by analyzing a singing voice excerpt.

FEATURE CLASSIFICATION

As in speech prosody, music can also be decomposed into various musical elements. The main musical elements, such as melody, dynamics, rhythm, and timbre, are built on low-level acoustic features. The relationships between these elements and the acoustic features can be represented in several ways [19, p. 44]. Based on this, Table 3 relates the commonly modeled acoustic features of the singing voice to the elements to which they belong. Some acoustic features spread transversally over several elements. Some features are instantaneous, such as F0 and intensity frame values, some span over a local time window, such as articulation and attack, and others have a more global temporal scope, such as F0 and intensity contours or vibrato and tremolo features.

Next, for each of these four musical elements, we provide introductory definitions to their acoustic features. Finally, these are related to the analysis of a real singing voice performance.

MELODY-RELATED FEATURES

The F0 contour, or the singer's rendition of the melody (note sequence in a score), is the sequence of F0 frame-based values [20]. F0 represents the "rate at which the vocal folds open and close across the glottis," and acoustically it is defined as "the lowest periodic cycle component of the acoustic waveform" [12, p. 790]. Perceptually, it relates to the pitch, defined as "the aspect of auditory sensation whose variation is associated with musical melodies" [21, p. 2]. In the literature, however, the pitch and F0 terms are often used indistinctly to refer to F0.

The F0 contour is affected by microprosody [22], i.e., fluctuations in pitch and dynamics due to phonetics (not attributable to expression). While certain phonemes such as vowels may have stable contours, other phonemes such as velar consonants may fluctuate because of articulatory effects.

A skilled singer can show expressive ability through the melody rendition and modify it more expressively than unskilled singers. Pitch deviations from the theoretical note can be intentional as an expressive resource [3]. Moreover, different articulations, i.e., the F0 contour in a transition between consecutive notes, can be used expressively. For example, in staccato, short pauses are introduced between notes. In the section "Transverse Features," the use of vibrato is detailed.

DYNAMICS-RELATED FEATURES

As summarized in [12, p. 790], the intensity (related to the perceived loudness of the voice) is a "measure of energy in the acoustic signal" usually from the waveform amplitude. It "reflects the effort required to produce the speech" or singing voice and is measured by energy at a frame level. A sequence of intensity values provides the intensity contour, which is correlated to the waveform envelope and the F0 since the energy increases with the F0 so as to produce a similar auditory loudness [23]. Acoustically, vocal effort is primarily related to the spectrum slope of the glottal sound source rather than the overall sound level. Tremolo may also be used, as detailed in the section "Transverse Features."

Microprosody also has an influence on intensity. The phonetic content of speech may produce intensity increases as in plosives or reductions like some unvoiced sounds.

RHYTHM-RELATED FEATURES

The perception of rhythm involves cognitive processes such as "movement, regularity, grouping, and yet accentuation and differentiation" [24, p. 588], where it is defined as "the grouping and strong/weak relationships" among the beats or "the sequence of equally spaced phenomenal impulses which define a tempo for the music." The tempo corresponds to the number of beats per minute. In real-life performances, there are timing deviations from the nominal score [12].

Similar to the role of speech rate in prosody, phoneme onsets are also affected by singing voice rhythm. Notes and lyrics are aligned so that the first vowel onset in a syllable is synchronized with the note onset and any preceding phoneme in the syllable is advanced [3], [25].

TIMBRE-RELATED FEATURES

The timbre mainly depends on the vocal tract dimensions and on the mechanical characteristics of the vocal folds, which affect the voice source signal [23]. Timbre is typically characterized by an amplitude spectrum representation and is often decomposed into source and vocal tract components.

[TABLE 3] CLASSIFICATION OF SINGING VOICE EXPRESSION FEATURES.

<table>
<tr><th>MELODY</th><th colspan="2">DYNAMICS</th><th>RHYTHM</th><th>TIMBRE</th></tr>
<tr><td colspan="2">VIBRATO AND TREMOLO (DEPTH AND RATE)</td><td colspan="2">PAUSES</td><td>VOICE SOURCE</td></tr>
<tr><td colspan="2">ATTACK AND RELEASE</td><td colspan="2">PHONEME TIME LAG</td><td>SINGER'S FORMANT</td></tr>
<tr><td>ARTICULATION</td><td colspan="3">PHRASING</td><td>SUBHARMONICS</td></tr>
<tr><td>F0 CONTOUR</td><td>INTENSITY CONTOUR</td><td>NOTE/PHONEME ONSET/DURATION</td><td>FORMANT TUNING</td></tr>
<tr><td>F0 FRAME VALUE</td><td rowspan="2">INTENSITY FRAME VALUE</td><td>TIMING DEVIATION</td><td rowspan="2">APERIODICITY SPECTRUM</td></tr>
<tr><td>DETUNING</td><td>TEMPO</td></tr>
</table>

The voice source can be described in terms of its F0, amplitude, and spectrum (i.e., vocal loudness and mode of phonation). In the frequency domain, the spectrum of the voice source is generally approximated by an average slope of –12 dB/octave, but it typically varies with vocal loudness [23]. The voice source is relevant for expression and is used differently among singing styles [14].

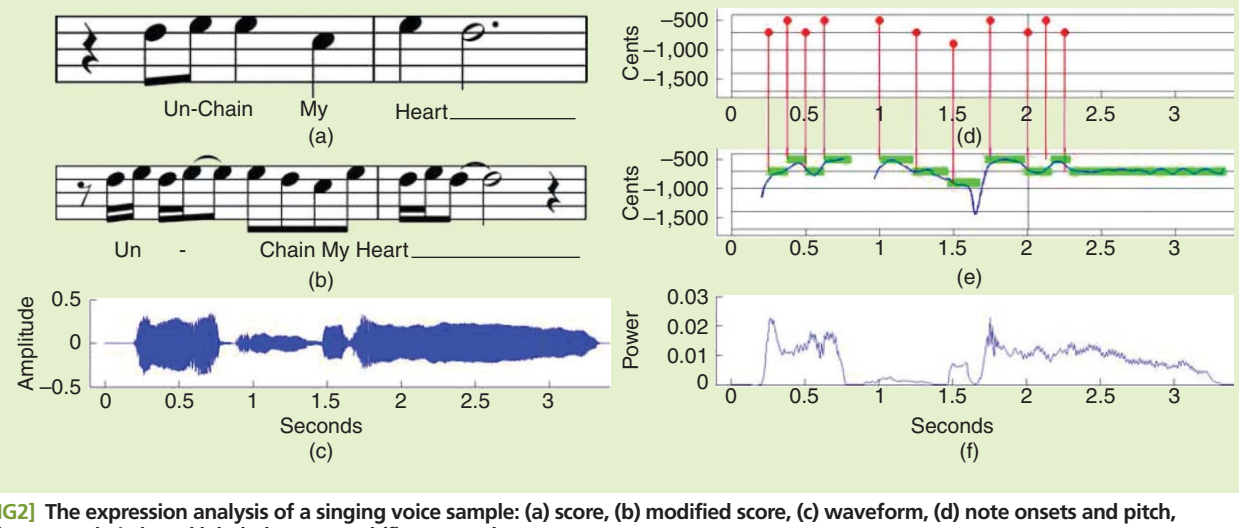
The vocal tract filters the voice source, emphasizing certain frequency regions or formants. Although formants are affected by all vocal tract elements, some have a higher effect on certain formants. For instance, the first two formants are related to the produced vowel, with the first formant being primarily related to the jaw opening and the second formant to the tongue body shape. The next three formants are related to timbre and voice identity, with the third formant being particularly influenced by the region under the tip of the tongue and the fourth to the vocal tract length and dimensions of the larynx [23]. In western male operatic voices, the third, fourth, and fifth formants typically cluster, producing a marked spectrum envelope peak around 3 kHz, the so-called singer's formant cluster [23]. This makes it easier to hear the singing voice over a loud orchestra. The affected harmonic frequencies (multiples of F0) are radiated most efficiently toward the direction where the singer is facing, normally the audience.

Changing modal voice into other voice qualities can be used expressively [26]. A rough voice results from a random modulation of the F0 of the source signal (jitter) or of its amplitude (shimmer). In a growl voice, subharmonics emerge because of half-periodic vibrations of the vocal folds, and, in breathy voices, the glottis does not completely close, increasing the presence of aperiodic energy.

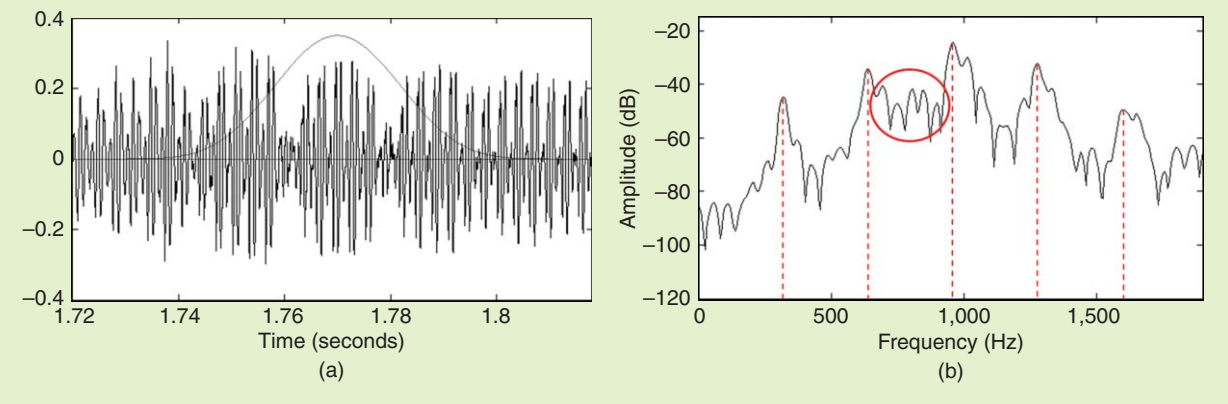
TRANSVERSE FEATURES

Several features from Table 3 can be considered transversal given that they are spread over several elements. In this section, we highlight the most relevant ones.

Vibrato is defined [23] as a nearly sinusoidal fluctuation of F0. In operatic singing, it is characterized by a rate that tends to range from 5.5 to 7.5 Hz and a depth around ±0.5 or 1 semitones.



[FIG2] The expression analysis of a singing voice sample: (a) score, (b) modified score, (c) waveform, (d) note onsets and pitch, (e) extracted pitch and labeled notes, and (f) extracted energy.



[FIG3] The growl analysis of a singing voice sample: (a) waveform and (b) spectrum with harmonics (dashed lines) and subharmonics (circle).

Tremolo [23] is the vibrato counterpart observed in intensity. It is caused by the vibrato oscillation when the harmonic with the greatest amplitude moves in frequency, increasing and decreasing the distance to a formant, thus making the signal amplitude vary. Vibrato may be used for two reasons [23, p. 172]. Acoustically, it prevents harmonics from different voices from falling into close regions and producing beatings. Also, vibratos are difficult to produce under phonatory difficulties such as pressed phonation. Aesthetically, vibrato shows that the singer is not running into such problems when performing a difficult note or phrase such as a high-pitched note.

Attack is the musical term to describe the pitch and intensity contour shapes and duration at the beginning of a musical note or phrase. *Release* is the counterpart of attack, referring to the pitch and intensity contour shapes at the end of a note or phrase.

As summarized in [27], grouping is one of the mental structures that are built while listening to a piece that describes the hierarchical relationships between different units. Notes, the lowest-level units, are grouped into motifs, motifs are grouped into phrases, and phrases are grouped into sections. The piece is the highest-level unit. Phrasing is a transversal aspect that can be represented as an "arch-like shape" applied to both the tempo and intensity during a phrase [15, p. 149]. For example, a singer may increase the tempo at the beginning of a phrase or decrease it at the end for classical music.

SINGING VOICE PERFORMANCE ANALYSIS

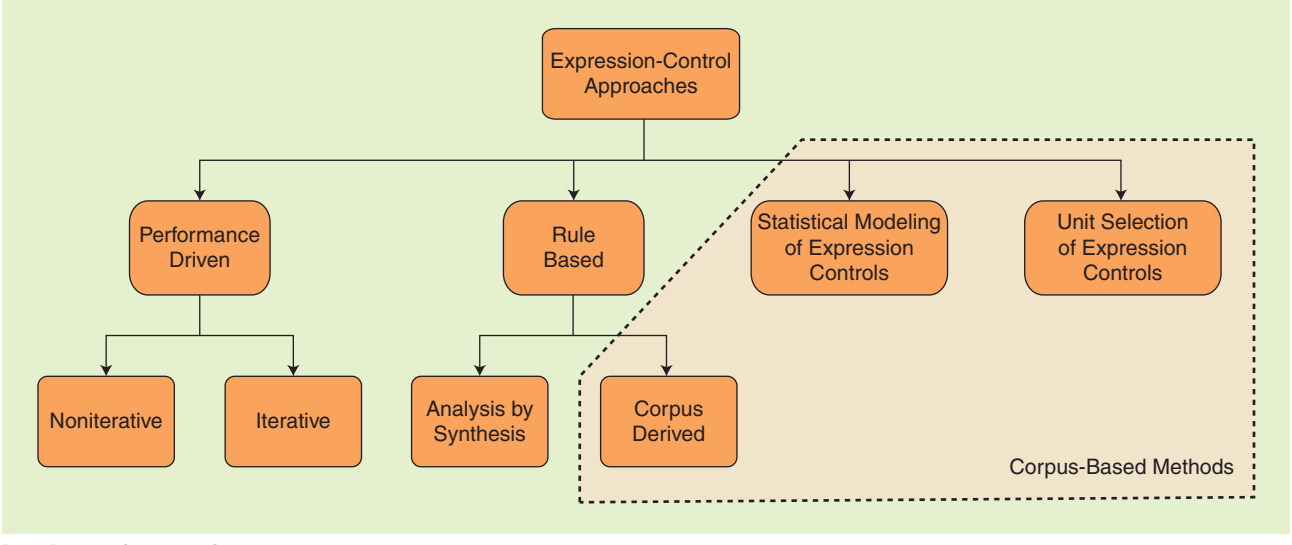
To illustrate the contribution of the acoustic features to expression, we analyze a short excerpt from a real singing performance (an excerpt from the song "Unchain My Heart;" see [51]). It contains clear expressive features such as vibrato in pitch, dynamics, timing deviations in rhythm, and growl in timbre. The result of the analysis is shown in Figures 2 and 3. (The dashed lines indicate harmonic frequencies, and the circle is placed at the subharmonics.) The original score and lyrics are shown in Figure 2(a), where each syllable corresponds to one note except for the first and last ones, which correspond to two notes. The singer introduces some changes, such as ornamentation and syncopation, which are represented in Figure 2(b). In (c), the note pitch is specified by the expected frequency in cents, and the note onsets are placed at the expected time using the note figures and a 120-beats/minute tempo. Figure 2(d) shows the extracted F0 contour in blue and the notes in green. The microprosody effects can be observed, for example, in a pitch valley during the attack to the word "heart." At the end, vibrato is observed. The pitch stays at the target pitch for a short period of time, especially in the ornamentation notes.

In a real performance, the tempo is not generally constant throughout a score interpretation. In general, beats are not equally spaced through time, leading to tempo fluctuation. Consequently, note onsets and rests are not placed where expected with respect to the score. In Figure 2(d), time deviations can be observed between the labeled notes and the projection colored in red from the score. Also, the note durations differ from the score.

The recording's waveform and energy, which are aligned to the estimated F0 contour, are shown in Figure 2(e) and (f), respectively. The intensity contour increases/decays at the beginning/end of each segment or note sequence. Energy peaks are especially prominent at the beginning of each segment since a growl voice is used, and increased intensity is needed to initiate this effect.

We can take a closer look at the waveform and spectrum of a windowed frame, as shown in Figure 3. In the former, we can see the pattern of a modulation in the amplitude or macroperiod, which spans over several periods. In the latter, we can see that, for the windowed frame, apart from the frequency components related to F0 around 320 Hz, five subharmonic components appear between F0 harmonics, which give the growl voice quality. Harmonics are marked with a dashed line and subharmonics between the second and the third harmonics with a red circle.

If this set of acoustic features is synthesized appropriately, the same perceptual aspects can be decoded. Several approaches that generate these features are presented next.



[FIG4] **Classification of expression-control methods in singing voice synthesis.**

[TABLE 4] A COMPARISON OF APPROACHES FOR EXPRESSION CONTROL IN SINGING VOICE SYNTHESIS.

TYPE	REFERENCE	CONTROL FEATURES	SYNTHESIZER	STYLE OR EMOTION	INPUT	LANGUAGE
PERFORMANCE DRIVEN	[29]	TIMING, F0, INTENSITY, SINGER'S FORMANT CLUSTER	UNIT SELECTION	OPERA	SCORE, SINGING VOICE	GERMAN
	[30]	TIMING, F0, INTENSITY, VIBRATO	SAMPLE BASED	GENERIC	LYRICS, MIDI NOTES, SINGING VOICE	SPANISH
	[31]	TIMING, F0, INTENSITY	SAMPLE BASED	POPULAR MUSIC (RWC-MDB)[1]	LYRICS, SINGING VOICE	JAPANESE
	[32]	TIMING, F0, INTENSITY, TIMBRE	SAMPLE BASED	MUSIC GENRE (RWC-MDB)[2]	LYRICS, SINGING VOICE	JAPANESE
	[33]	TIMING, F0, SINGER FORMANT	RESYNTHESIS OF SPEECH	CHILDREN'S SONGS	SCORE, TEMPO, SPEECH	JAPANESE
RULE BASED	[3]	TIMING, CONSONANT DURATION, VOWEL ONSET, TIMBRE CHANGES, FORMANT TUNING, OVERTONE SINGING, ARTICULATION SILENCE TO NOTE	FORMANT SYNTHESIS	OPERA	SCORE, MIDI, OR KEYBOARD	NOT SPECIFIED
	[37]	TIMING, MICROPAUSES, TEMPO AND PHRASING, F0, INTENSITY, VIBRATO AND TREMOLO, TIMBRE QUALITY	SAMPLE BASED	ANGRY, SAD, HAPPY	SCORE, LYRICS, TEMPO, EXPRESSIVE INTENTIONS	SWEDISH, ENGLISH
	[40]	TIMBRE (MANUAL), PHONETICS, TIMING, F0, INTENSITY, MUSICAL ARTICULATION, SUSTAINS, VIBRATO AND TREMOLO (RATE AND DEPTH)	SAMPLE BASED	GENERIC	SCORE, LYRICS, TEMPO	JAPANESE, ENGLISH, SPANISH
STATISTICAL MODELING	[25]	TIMING, F0, TIMBRE	HMM BASED	CHILDREN'S SONGS	SCORE AND LYRICS	JAPANESE
	[42]	TIMING, F0, VIBRATO AND TREMOLO, TIMBRE, SOURCE	HMM BASED	CHILDREN'S SONGS	MUSICXML[2] SCORE	JAPANESE, ENGLISH
	[22]	BASELINE F0 (RELATIVE TO NOTE), VIBRATO RATE AND DEPTH (NOT TREMOLO), INTENSITY	SAMPLE BASED	CHILDREN'S SONGS	SCORE (NO LYRICS TO CREATE MODELS)	JAPANESE
UNIT SELECTION	[43]	F0, VIBRATO, TREMOLO, INTENSITY	SAMPLE BASED	JAZZ STANDARDS	SCORE	LANGUAGE INDEPENDENT

[1] Real World Computing (RWC) Music Database: https://staff.aist.go.jp/m.goto/RWC-MDB/
[2] MUSICXML: http://www.musicxml.com

EXPRESSION-CONTROL APPROACHES

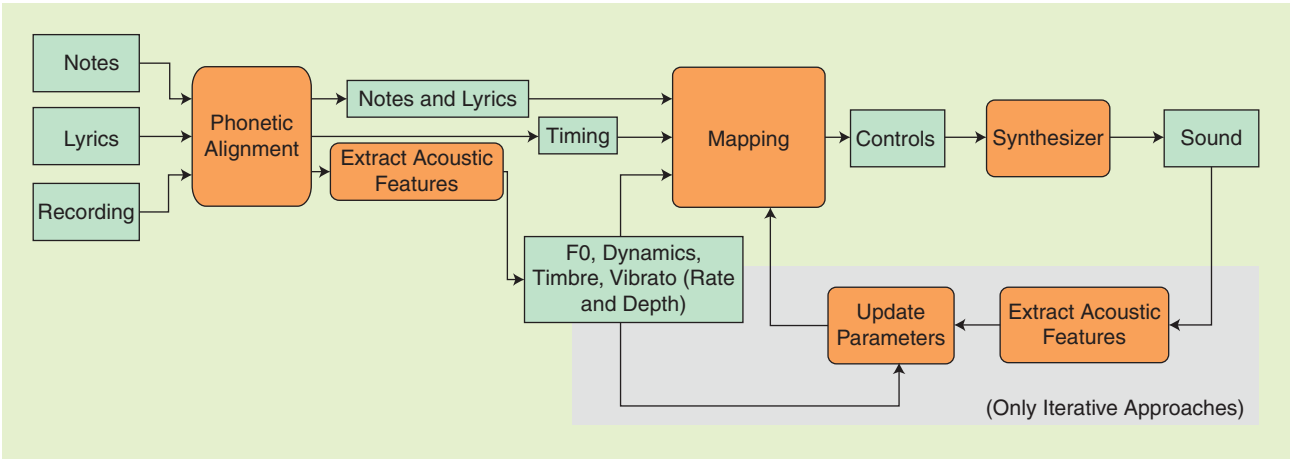
In the section "Singing Voice Performance Features," we defined the voice acoustic features and related them to aspects of music perception. In this section, we focus on how different approaches generate expression controls. First, we propose a classification of the reviewed approaches and then we compare and describe them. As will be seen, acoustic features generally map one to one to expressive controls at the different temporal scopes, and the synthesizer is finally controlled by the lowest-level acoustic features (i.e., F0, intensity, and spectral envelope representation).

CLASSIFICATION OF APPROACHES

To see the big picture of the reviewed works on expression control, we propose a classification in Figure 4. Performance-driven approaches use real performances as the control for a synthesizer, taking advantage of the implicit rules that the singer has applied to interpret a score. Expression controls are estimated and applied directly to the synthesizer. Rule-based methods derive a set of rules that reflect the singers' cognitive process. In analysis by synthesis, rules are evaluated by synthesizing singing voice performances. Corpus-derived, rule-based approaches generate expression controls from the observation of singing voice contours and imitating their behavior. Statistical approaches generate singing voice expression features using techniques such as hidden Markov models (HMMs). Finally, unit selection-based approaches select, transform, and concatenate expression contours from excerpts of a singing voice database (DB). Approaches using a training database of expressive singing have been labeled as corpus-based methods. The difficulties of the topic reviewed in this article center on how to generate control parameters that are perceived as natural. The success of conveying natural expression depends on a comprehensive control of the acoustic features introduced in the section "Singing Voice Performance Features." Currently, statistical approaches are the only type of system that jointly model all of the expression features.

COMPARISON OF APPROACHES

In this article, we review a set of works that model the features that control singing voice synthesis expression. Physical modeling perspective approaches can be found, for instance, in [28].



[FIG5] **The general framework for performance-driven approaches.**

Within each type of approach in Figure 4, there are one or more methods for expression control. In Table 4, we provide a set of items that we think can be useful for comparison. The "Type" column refers to the type of expression control from Figure 4 to which the reference belongs. In the "Control Feature" column, we list the set of features addressed by the approach. The "Synthesizer" column provides the type of synthesizer used to generate the singing voice, and the "Style or Emotion" column provides the emotion, style, or sound to which the expression is targeted. The "Input" column details the input to the system (e.g., the score, lyrics, tempo, and audio recording). The "Language" column lists the language dependency of each method, if any.

We have collected samples in [51] as examples of the results of the reviewed expression-control approaches. Listeners will observe several differences among them. First, some samples consist of a cappella singing voices, and others are presented with background music, which may mask the synthesized voice and complicate the perception of the generated expression. Second, the samples correspond to different songs, making it difficult to compare approaches. Athough the lyrics in most cases belong to a particular language, in some samples, they are made by repeating the same syllable, such as /la/. We believe that the evaluation of a synthesized song can be performed more effectively in a language spoken by the listener. Finally, the quality of the synthetic voice is also affected by the type of synthesizer used in each sample. The difficulties in comparing them and the subsequent criticisms are discussed in the "Evaluation" and "Challenges" sections.

PERFORMANCE-DRIVEN APPROACHES

Performance-driven approaches use a real performance to control the synthesizer. The knowledge applied by the singer, implicit in the extracted data, can be used in two ways. In the first one, control parameters such as F0, intensity, and timing from the reference recording are mapped to the input controls of the synthesizer so that the rendered performance follows the input signal expression. Alternatively, speech audio containing the target lyrics is transformed to match the pitch and timing of the input score. Figure 5 summarizes the commonalities of these approaches on the inputs (reference audio, lyrics, and, possibly, the note sequence) and intermediate steps (phonetic alignment, acoustic feature extraction, and mapping) that generate internal data such as timing information, acoustic features, and controls used by the synthesizer.

In Table 5, we summarize the correspondence between the extracted acoustic features and the synthesis parameters for each of

[TABLE 5] MAPPING FROM ACOUSTIC FEATURES TO SYNTHESIZER CONTROLS.

	MAPPED SYNTHESIS PARAMETERS				
ACOUSTIC FEATURES	**[29]**	**[30]**	**[31]**	**[32]**	**[33]**
F0	F0	SMOOTHED AND CONTINUOUS PITCH	MIDI NOTE NUMBER, PITCH BEND AND SENSITIVITY	MIDI NOTE NUMBER, PITCH BEND AND SENSITIVITY	F0
VIBRATO	INCLUDED IN F0 IMPLICITLY	VIBRATOS FROM INPUT OR FROM DB SINGER	INCLUDED IN F0 IMPLICITLY	INCLUDED IN F0 IMPLICITLY	INCLUDED IN F0 IMPLICITLY
ENERGY	DYNAMICS	DYNAMICS	DYNAMICS	DYNAMICS	DYNAMICS
PHONETIC ALIGNMENT	PHONEME TIMING	ONSETS OF VOWELS OR VOICED PHONEMES	NOTE ONSET AND DURATION	NOTE ONSET AND DURATION	PHONEME TIMING
TIMBRE	SINGER'S FORMANT CLUSTER AMPLITUDE	NOT USED	NOT USED	MIXING DIFFERENT VOICE QUALITY dBs	SINGER'S FORMANT CLUSTER AMPLITUDE AND AMPLITUDE MODULATION OF THE SYNTHESIZED SIGNAL

these works. The extracted F0 can be mapped directly into the F0 control parameter, processed into a smoothed and continuous version, or split into the Musical Instrument Digital Instrument (MIDI) note, pitch bend, and its sensitivity parameters. Vibrato can be implicitly modeled in the pitch contour, extracted from the input, or selected from a database. Energy is generally mapped directly into dynamics. From the phonetic alignment, note onsets and durations are derived, mapped directly to phoneme timing, or mapped either to onsets of vowels or voiced phonemes. Concerning timbre, some approaches focus on the singer's formant cluster, and, in a more complex case, the output timbre comes from a mixture of different voice quality databases.

Approaches using estimated controls achieve different levels of robustness depending on the singing voice synthesizers and voice databases. In the system presented in [29], a unit selection framework is used to create a singing voice synthesizer from a particular singer's recording in a nearly automatic procedure. In comparison to a sample-based system, where the design criterion is to minimize the size of the voice database with only one possible unit sample (e.g., diphones), the criterion in unit selection is related to redundancy to allow the selection of consecutive units in the database at the expense of having a larger database. The system automatically segments the recorded voice into phonemes by aligning it to the score and feeding the derived segmentation constraints to an HMM recognition system. Units are selected to minimize a cost function that scores the amount of time, frequency, and timbre transformations. Finally, units are concatenated. In this approach, the main effort is put on the synthesis engine. Although it uses a unit selection-based synthesizer, the expression controls for pitch, timing, dynamics, and timbre, like the singer's formant, are extracted from a reference singing performance of the target score. These parameters are directly used by the synthesizer to modify the selected units with a combination of sinusoidal modeling (SM) with time domain pitch synchronous overlap add (TD-PSOLA) called *SM-PSOLA*. Editing is allowed by letting the user participate in the unit selection process, change some decisions, and modify the unit boundaries. Unfortunately, this approach only manipulates the singer's formant feature of timbre so that other significant timbre-related features in the opera singing style are not handled.

In [30], the steps followed are: extraction of acoustic features such as energy, F0, and automatic detection of vibrato sections; mapping into synthesis parameters; and phonetic alignment. The mapped controls and the input score are used to build an internal score that matches the target timing, pitch, and dynamics, and minimizes the transformation cost of samples from a database. However, this approach is limited since timbre is not handled and also because the expression features of the synthesized performance are not compared to the input values. Since this approach lacks a direct mapping of acoustic features to control parameters, these differences are likely to happen. On the other hand, the possibility of using a singer DB to produce vibratos other than the extracted ones from the reference recording provides a new degree of freedom to the user.

Toward a more robust methodology to estimate the parameters, in [31], the authors study an iterative approach that takes the target singing performance and lyrics as input. The musical score or note sequence is automatically generated from the input. The first iteration provides an initialization of the system similar to the previous approach [30]. At this point, these controls can be manually edited by applying pitch transposition, correction, vibrato modifications, and pitch and intensity smoothing. The iterative process continues by analyzing the synthesized waveform and adjusting the control parameters so that, in the next iteration, the results are closer to the expected performance. In [32], the authors extend this approach by including timbre. Using different voice quality databases from the same singer, the corresponding versions of the target song are synthesized as in the previous approach. The system extracts the spectral envelopes of each one to build a three-dimensional (3-D) voice timbre space. Next, a temporal trajectory in this space is estimated from the reference target performance to represent its spectral timbre changes. Finally, singing voice synthesis output is generated using the estimated trajectory to imitate the target timbre change. Although expression control is more robust than the previous approach, thanks to iteratively updating the parameters and by allowing a certain degree of timbre control, these approaches also have some limitations. First, it cannot be assured that the iterative process will converge to the optimal set of parameter values. Second, the timbre control is limited to the variability within the set of available voice quality databases.

In [33], naturally spoken readings of the target lyrics are transformed into a singing voice by matching the target song properties described in the musical score. Other input data are the phonetic segmentation and the synchronization of phonemes and notes. This approach first extracts acoustic features such as F0, spectral envelope, and the aperiodicity index from the input speech. Then, a continuous F0 contour is generated from discrete notes, phoneme durations are lengthened, and the singer's formant cluster is generated. The fundamental frequency contour takes into account four types of fluctuations: 1) overshoot (F0 exceeds the target note after a note change), 2) vibrato, 3) preparation (similar to overshoot before the note change), and 4) fine fluctuations. The first three types of F0 fluctuations are modeled by a single second-order transfer function that depends mainly on a damping coefficient, a gain factor, and a natural frequency. A rule-based approach is followed for controlling phoneme durations by splitting consonant-to-vowel transitions into three parts. First, the transition duration is not

[TABLE 6] SINGING VOICE-RELATED KTH RULES' DEPENDENCIES.

ACOUSTIC FEATURE	DEPENDENCIES
CONSONANT DURATION	PREVIOUS VOWEL LENGTH
VOWEL ONSET	SYNCHRONIZED WITH TIMING
FORMANT FREQUENCIES	VOICE CLASSIFICATION
FORMANT FREQUENCIES	PITCH, IF OTHERWISE F0 WOULD EXCEED THE FIRST FORMANT
SPECTRUM SLOPE	DECREASE WITH INCREASING INTENSITY
VIBRATO	INCREASE DEPTH WITH INCREASING INTENSITY
PITCH IN COLORATURA PASSAGES	EACH NOTE REPRESENTED AS A VIBRATO CYCLE
PITCH PHRASE ATTACK (AND RELEASE)	AT PITCH START (END) FROM (AT) 11 SEMITONES BELOW TARGET F0

modified for singing. Then, the consonant part is transformed based on a comparative study of speech and singing voices. Finally, the vowel section is modified so that the duration of the three parts matches the note duration. With respect to timbre, the singer's formant cluster is handled by an emphasis function in the spectral domain centered at 3 kHz. Amplitude modulation is also applied to the synthesized singing voice according to the generated vibrato's parameters. Although we have classified this approach as performance-driven since the core data are found in the input speech recording, some aspects are modeled, such as the transfer function for F0, rules for phonetic duration, and a filter for the singer's formant cluster. Similarly to [29], in this approach, timbre control is limited to the singer formant, so the system cannot change other timbre features. However, if the reference speech recording contains voice quality variations that fit the target song, this can add some naturalness to the synthesized singing performance.

Performance-driven approaches achieve a highly expressive control since performances implicitly contain knowledge naturally applied by the singer. These approaches are especially convenient for creating parallel database recordings, which are used in voice conversion approaches [8]. On the other hand, the phonetic segmentation may cause timing errors if not manually corrected. The noniterative approach lacks robustness because the differences between the input controls and the controls extracted from the synthesized sound are not corrected. In [32], timbre control is limited by the number of available voice qualities. We note that a human voice input for natural singing control is required for these approaches, which can be considered a limitation since it may not be available in most cases. When such a reference is not given, other approaches are necessary to derive singing control parameters from the input musical score.

RULE-BASED APPROACHES

Rules can be derived from work with synthesizing and analyzing sung performances. Applying an analysis-by-synthesis method, an ambitious rule-based system for western music was developed at KTH in the 1970s and improved over the last three decades [3]. By synthesizing sung performances, this method aims at identifying acoustic features that are perceptually important, either individually or jointly [15]. The process of formulating a rule is iterative. First, a tentative rule is formulated and implemented and the resulting synthesis is assessed. If its effect on the performance needs to be changed or improved, the rule is modified and the effect of the resulting performance is again assessed. On the basis of parameters such as phrasing, timing, metrics, note articulation, and intonation, the rules modify pitch, dynamics, and timing. Rules can be combined to model emotional expressions as well as different musical styles. Table 6 lists some of the acoustic features and their dependencies.

The rules reflect both physical and musical phenomena. Some rules are compulsory and others optional. The consonant duration rule, which lengthens consonants following short vowels, also applies to speech in some languages. The vowel onset rule corresponds to the general principle that the vowel onset is synchronized with the onset of the accompaniment, even though lag and lead of onset are often used for expressive purposes [34]. The spectrum slope rule is compulsory as it reflects the fact that vocal loudness is controlled by subglottal pressure and an increase of this pressure leads to a less steeply sloping spectrum envelope. The pitch in coloratura passages rule implies that the fundamental frequency makes a rising–falling gesture around the target frequency in legato sequences of short notes [35]. The pitch phrase attack, in lab jargon referred to as *bull's roaring onset*, is an ornament used in excited moods and would be completely out of place in a tender context. Interestingly, results close to the KTH rules have been confirmed by machine-learning approaches [36].

A selection of the KTH rules [15] has been applied to the Vocaloid synthesizer [37]. Features are considered at the note level (start and end times), intra- and internote (within and between note changes), and to timbre variations (not related to KTH rules). The system implementation is detailed in [38] along with the acoustic cues, which are relevant for conveying basic emotions such as anger, fear, happiness, sadness, and love/tenderness [12]. The rules are combined in expressive palettes indicating to what degree the rules need to be applied to convey a target

[TABLE 7] THE SELECTION OF RULES FOR SINGING VOICE: LEVEL OF APPLICATION AND AFFECTED ACOUSTIC FEATURES.

LEVEL	RULES	AFFECTED ACOUSTIC FEATURES
NOTE	DURATION CONTRAST	DECREASE DURATION AND INTENSITY OF SHORT NOTES PLACED NEXT TO LONG NOTES
	PUNCTUATION	INSERT MICROPAUSES IN CERTAIN PITCH INTERVAL AND DURATION COMBINATIONS
	TEMPO	CONSTANT VALUE FOR THE NOTE SEQUENCE (MEASURED IN BEATS/min)
	INTENSITY	SMOOTH/STRONG ENERGY LEVELS, HIGH PITCH NOTES INTENSITY INCREASES 3 dB/OCTAVE
	TRANSITIONS	LEGATO, STACCATO (PAUSE IS SET TO MORE THAN 30% OF INTERONSET INTERVAL)
	PHRASING ARCH	INCREASE/DECREASE TEMPO AT PHRASE BEGINNING/END, SAME FOR ENERGY
	FINAL RITARDANDO	DECREASE TEMPO AT THE END OF A PIECE
INTRA-/INTERNOTE	ATTACK	PITCH SHAPE FROM STARTING PITCH UNTIL TARGET NOTE, ENERGY INCREASES SMOOTHLY
	NOTE ARTICULATION	PITCH SHAPE FROM THE STARTING TO THE ENDING NOTE, SMOOTH ENERGY
	RELEASE	ENERGY DECREASES SMOOTHLY TO ZERO, DURATION IS MANUALLY EDITED
	VIBRATO AND TREMOLO	MANUAL CONTROL OF POSITION, DEPTH, AND RATE (COSINE FUNCTION AND RANDOM FLUCTUATIONS)
TIMBRE	BRIGHTNESS	INCREASE HIGH FREQUENCIES DEPENDING ON ENERGY
	ROUGHNESS	SPECTRAL IRREGULARITIES
	BREATHINESS	MANUAL CONTROL OF NOISE LEVEL (NOT INCLUDED IN EMOTION PALETTES)

[TABLE 8] CONTEXTUAL FACTORS IN HMM-BASED SYSTEMS.

HMM-BASED APPROACHES	LEVELS	CONTEXTUAL FACTORS
[25]	PHONEME	P/C/N PHONEMES
	NOTE	P/C/N NOTE F0, DURATIONS, AND POSITIONS WITHIN THE MEASURE
[42]	PHONEME	FIVE PHONEMES (CENTRAL AND TWO PRECEDING AND SUCCEEDING)
	MORA	NUMBER OF PHONEMES IN THE P/C/N MORA
		POSITION OF THE P/C/N MORA IN THE NOTE
	NOTE	MUSICAL TONE, KEY, TEMPO, LENGTH, AND DYNAMICS OF THE P/C/N NOTE
		POSITION OF THE CURRENT NOTE IN THE CURRENT MEASURE AND PHRASE
		TIES AND SLURRED ARTICULATION FLAG
		DISTANCE BETWEEN CURRENT NOTE AND NEXT/PREVIOUS ACCENT AND STACCATO
		POSITION OF THE CURRENT NOTE IN THE CURRENT CRESCENDO OR DECRESCENDO
	PHRASE	NUMBER OF PHONEMES AND MORAS IN THE P/C/N PHRASE
	SONG	NUMBER OF PHONEMES, MORAS, AND PHRASES IN THE SONG
[22]	NOTE REGION	MANUALLY SEGMENTED BEHAVIOR TYPES (BEGINNING, SUSTAINED, ENDING)
	NOTE	MIDI NOTE NUMBER AND DURATION (IN 50-MILLISECOND UNITS)
		DETUNING: MODEL F0 BY THE RELATIVE DIFFERENCE TO THE NOMINAL NOTE

P/C/N: PREVIOUS, CURRENT, AND NEXT.

emotion. The relationship between application level, rules, and acoustic features is shown in Table 7. As an example of the complexity of the rules, the punctuation rule at the note level inserts a 20-millisecond micropause if a note is three tones lower than the next one and its duration is 20% larger. Given that this work uses a sample-based synthesizer, voice quality modifications are applied to the retrieved samples. In this case, the timbre variations are limited to rules affecting brightness, roughness, and breathiness and, therefore, do not cover the expressive possibilities of a real singer.

Apart from the KTH rules, in corpus-derived rule-based systems, heuristic rules are obtained to control singing expression by observing recorded performances. In [6], expression controls are generated from high-level performance scores where the user specifies the note articulation, pitch, intensity, and vibrato data that are used to retrieve templates from recorded samples. This work, used in the Vocaloid synthesizer [39], models the singer's performance with heuristic rules [40]. The parametric model is based on anchor points for pitch and intensity, which are manually derived from the observation of a small set of recordings. At synthesis, the control contours are obtained by interpolating the anchor points generated by the model. The number of points used for each note depends on its absolute duration. The phonetics relationship with timing is handled by synchronizing the vowel onset with the note onset. Moreover, manual editing is permitted for the degree of articulation application as well as its duration, pitch and dynamics contours, phonetic transcription, timing, vibrato and tremolo depth and rate, and timbre characteristics.

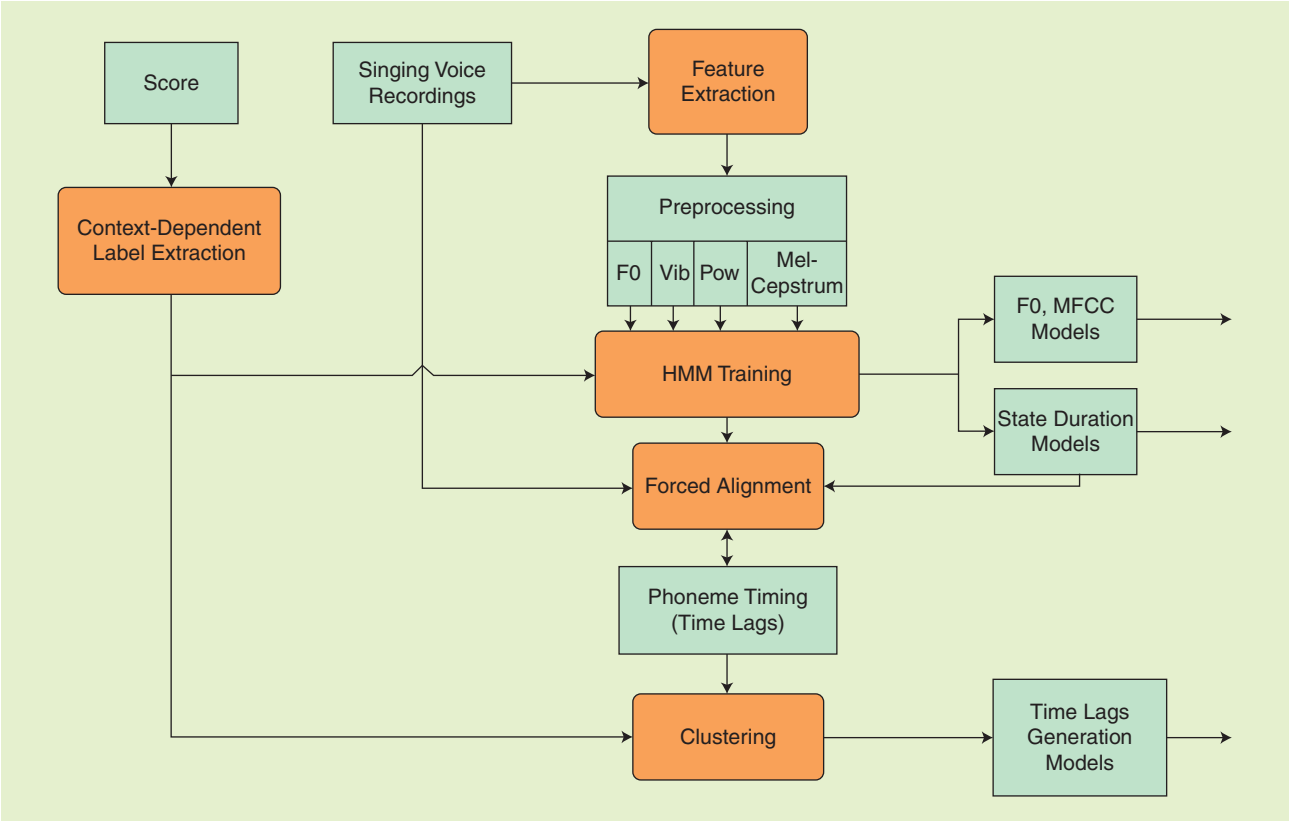
The advantage of these approaches is that they are relatively straightforward and completely deterministic. Random variations can be easily introduced so that the generated contours are different for each new synthesis of the same score, resulting in distinct interpretations. The main drawbacks are that either the models are based on few observations that do not fully represent a given style or they are more elaborate but become unwieldy due to the complexity of the rules.

STATISTICAL MODELING APPROACHES

Several approaches have been used to statistically model and characterize expression-control parameters using HMMs. They have a common precedent in speech synthesis [41], where the parameters such as spectrum, F0, and state duration are jointly modeled. Compared to unit selection, HMM-based approaches tend to produce lower speech quality, but they need a smaller data set to train the system without needing to cover all combinations of contextual factors. Modeling a singing voice with HMMs amounts to using similar contextual data as those used for speech synthesis, adapted to singing voice specificities. Moreover, new voice characteristics can be easily generated by changing the HMM parameters.

These systems operate in two phases: training and synthesis. In the training part, acoustic features are first extracted from the training recordings, such as F0, intensity, vibrato parameters, and Mel-cepstrum coefficients. Contextual labels, i.e., the relationships of each note, phoneme, or phrase with the preceding and succeeding values, are derived from the corresponding score and lyrics. Contextual labels vary in their scope at different levels, such as phoneme, note, or phrase, according to the approach, as summarized in Table 8. This contextual data are used to build the HMMs that relate how these acoustic features behave according to the clustered contexts. The phoneme timing is also modeled in some approaches. These generic steps for the training part in HMM-based synthesis are summarized in Figure 6. The figure shows several blocks found in the literature, which might not be present simultaneously in each approach. We refer to [41] for the detailed computations that HMM training involves.

In the synthesis part, given a target score, contextual labels are derived as in the training part from the note sequence and lyrics. Models can be used in two ways. All necessary parameters for singing voice synthesis can be generated from them; therefore, state durations, F0, vibrato, and Mel-cepstrum observations are generated to synthesize the singing voice. On the other hand, if another synthesizer is used, only control parameters, such as F0, vibrato depth and rate, and dynamics need to be generated, which are then used as input of the synthesizer.



[FIG6] **Generic blocks for the training part of HMM-based approaches.**

As introduced in the section "Classification of Approaches," statistical methods jointly model the largest set of expression features among the reviewed approaches. This gives them a better generalization ability. As long as singing recordings for training involve different voice qualities, singing styles or emotions, and the target language phonemes, these will be reproducible at synthesis given the appropriate context labeling. Model interpolation allows new models to be created as a combination of existing ones. New voice qualities can be created by modifying the timbre parameters. However, this flexibility is possible at the expense of having enough training recordings to cover the combinations of the target singing styles and voice qualities. In the simplest case, a training database of a set of songs representing a single singer and style in a particular language would be enough to synthesize it. As a drawback, training HMMs with large databases tends to produce smoother time series than the original training data, which may be perceived as unnatural.

In [25], a corpus-based singing voice synthesis system based on HMMs is presented. The contexts are related to phonemes, note F0 values, and note durations and positions, as we show in Table 8 (dynamics are not included). Also, synchronization between notes and phonemes needs to be handled adequately, mainly because phoneme timing does not strictly follow the score timing, and phonemes might be advanced with respect to the nominal note onsets (negative time lag).

In this approach, the training part generates three models: 1) for the spectrum and excitation (F0) parts extracted from the training database, 2) for the duration of context-dependent states, and 3) to model the time lag. The second and third model note timing and phoneme durations of real performances, which are different than what can be inferred from the musical score and its tempo. Time lags are obtained by forced alignment of the training data with context-dependent HMMs. Then, the computed time lags are related to their contextual factors and clustered by a decision tree.

The singing voice is synthesized in five steps: 1) the input score (note sequence and lyrics) is analyzed to determine note duration and contextual factors, 2) a context-dependent label sequence of contextual factors as shown in Table 8 is generated, 3) the song HMM is generated, 4) its state durations are jointly determined with the note time lags, and 5) spectral and F0 parameters are generated, which are used to synthesize the singing voice. The authors claim that the synthesis performance achieves a natural singing voice, which simulates expression elements of the target singer such as voice quality and singing style (i.e., F0 and time lag).

In [25], the training database consists of 72 minutes of a male voice singing 60 Japanese children's songs in a single voice quality. These are the characteristics that the system can reproduce in a target song. The main limitation of this approach is that the contextual factors scope is designed to only cover phoneme and note descriptors. Longer scopes than just the previous and next note are necessary to model higher-level expressive features such as phrasing. Although we could not get samples from this work, an evolved system is presented next.

The system presented in [25] has been improved and is publicly available as Sinsy, an online singing voice synthesizer [42]. The new

characteristics of the system include reading input files in MusicXML format with F0, lyrics, tempo, key, beat, and dynamics as well as extended contextual factors used in the training part, vibrato rate and depth modeling, and a reduction of the computational cost. Vibrato is jointly modeled with the spectrum and F0 by including the depth and rate in the observation vector in the training step.

The new set of contexts automatically extracted from the musical score and lyrics used by the Sinsy approach are also shown in Table 8. These factors describe the context such as previous, current, and next data at different hierarchical levels: phoneme, mora (the sound unit containing one or two phonemes in Japanese), note, phrase, and the entire song. Some of them are strictly related to musical expression aspects, such as musical tone, key, tempo, length and dynamics of notes, articulation flags, or distance to accents and staccatos.

Similar to [25], in this case, the training database consists of 70 minutes of a female voice singing 70 Japanese children's songs in a single voice quality. However, it is able to reproduce more realistic expression control since vibrato parameters are also extracted and modeled. Notes are described with a much richer set of factors than the previous work. Another major improvement is the scope of the contextual factors shown in Table 8, which spans from the phoneme level to the whole song and is, therefore, able to model phrasing.

In [22], a statistical method is able to model singing styles. This approach focuses on baseline F0; vibrato features such as its extent, rate, and evolution over time, not tremolo; and dynamics. These parameters control the Vocaloid synthesizer, and so the timbre is not controlled by the singing style modeling system but is dependent on the database.

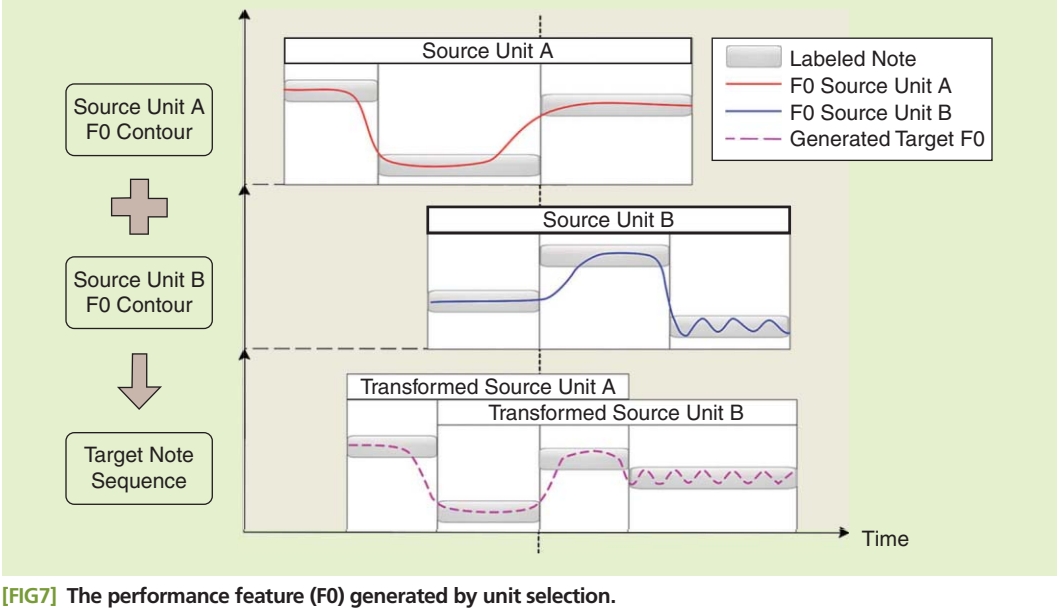
A preprocessing step is introduced after extracting the acoustic features such as F0 and dynamics to get rid of the microprosody effects on such parameters by interpolating F0 in unvoiced sections and flattening F0 valleys of certain consonants. The main assumption here is that expression is not affected by phonetics, which is reflected in erasing such dependencies in the initial preprocessing step and also in training note HMMs instead of phoneme HMMs. Also, manual checking is done to avoid errors in F0 estimation and MIDI events such as note on and note off estimated from the phonetic segmentation alignment. A novel approach estimates the vibrato shape and rate, which at synthesis is added to the generated baseline melody parameter. The shape is represented with the low-frequency bins of the Fourier transform of single vibrato cycles. In this approach, context-dependent HMMs model the expression parameters summarized in Table 8. Feature vectors contain melody, vibrato shape and rate, and dynamics components.

This last HMM-based work focuses on several control features except for timbre, which is handled by the Vocaloid synthesizer. This makes the training database much smaller in size. It consists of 5 minutes of five Japanese children's songs since there is no need to cover a set of phonemes. Contextual factors are rich at the note level since the notes are divided into three parts (begin, sustain, and end), and the detuning is also modeled relative to the nominal note. On the other hand, this system lacks the modeling of wider temporal aspects such as phrasing.

UNIT SELECTION APPROACHES

The main idea of unit selection [29] is to use a database of singing recordings segmented into units that consist of one or more phonemes or other units such as diphones or half phones. For a target score, a sequence of phonemes with specific features such as pitch or duration is retrieved from the database. These are generally transformed to match the exact required characteristics.

An important step in this kind of approach is the definition of the target and concatenation cost functions as the criteria on which unit selection is built. The former is a distance measure of the unit transformation in terms of a certain acoustic feature such as pitch or duration. Concatenation costs measure the perceptual consequences



[FIG7] **The performance feature (F0) generated by unit selection.**

[TABLE 9] UNIT SELECTION COST FUNCTIONS.

COST	DESCRIPTION	COMPUTATION
NOTE DURATION	COMPARE SOURCE AND TARGET UNIT NOTE DURATIONS	OCTAVE RATIO (SOURCE/TARGET UNIT NOTES)
PITCH INTERVAL	COMPARE SOURCE AND TARGET UNIT NOTE INTERVALS	OCTAVE RATIO (SOURCE/TARGET UNIT INTERVALS)
CONCATENATION	FAVOR COMPATIBLE UNITS FROM THE DB	ZERO IF CONSECUTIVE UNITS
CONTINUITY	FAVOR SELECTION OF CONSECUTIVE UNITS	PENALIZE SELECTION OF NONCONSECUTIVE UNITS

of joining nonconsecutive units. These cost functions' contributions are weighted and summed to get the overall cost of the unit sequence. The goal is then to select the sequence with the lowest cost.

Unit selection approaches present the disadvantages of requiring a large database, which needs to be labeled, and the subcost weights need to be determined. On the other hand, the voice quality and naturalness are high because of the implicit rules applied by the singer within the units.

A method to model pitch, vibrato features, and dynamics based on selecting units from a database of performance contours has recently been proposed [43]. We illustrate it in Figure 7 for the F0 contour showing two selected source units for a target note sequence where units are aligned at the transition between the second and third target notes. The target note sequence is used as input to generate the pitch and dynamics contours. A reference database is used that contains extracted pitch, vibrato features, and dynamics from expressive recordings of a single singer and style. In addition to these features, the database is labeled with the note pitches, durations, and strength as well as the start and end times of note transitions. This approach splits the task of generating the target song expression contours into first finding similar and shorter note combinations (source units A and B in Figure 7), and then transforming and concatenating the corresponding pitch and dynamics contours to match the target score (the dashed line in Figure 7). These shorter contexts are the so-called units, defined by three consecutive notes or silences, so that consecutive units overlap by two notes. The contour of dynamics is generated similarly from the selected units.

With regard to unit selection, the cost criterion consists of the combination of several subcost functions, as summarized in Table 9. In this case, there are four functions and unit selection is implemented with the Viterbi algorithm. The overall cost function considers the amount of transformation in terms of note durations (note duration cost) and pitch interval (pitch interval cost) to preserve as much as possible the contours as originally recorded. It also measures how appropriate it is to concatenate two units (concatenation cost) as a way of penalizing the concatenation of units from different contexts. Finally, the overall cost function also favors the selection of long sequences of consecutive notes (continuity cost), although the final number of consecutive selected units depends on the resulting cost value. This last characteristic is relevant to be able to reflect the recorded phrasing at synthesis.

Once a sequence is retrieved, each unit is time scaled and pitch shifted. The time scaling is not linear; instead, most of the transformation is applied in the sustain part and keeping the transition (attacks and releases) durations as close to the original as possible. Vibrato is handled with a parametric model, which allows the original rate and depth contour shapes to be kept.

The transformed unit contours are overlapped and added after applying a cross-fading mask, which mainly keeps the shape of the attack to the unit central note. This is done separately for the intensity, baseline pitch and vibrato rate, and vibrato depth contours. The generated baseline pitch is then tuned to the target note pitches to avoid strong deviations. Then, vibrato rate and depth contours are used to compute the vibrato oscillations, which are added to the baseline pitch.

The expression database contains several combinations of note durations, pitch intervals, and note strength. Such a database can be created systematically [44] to cover a relevant portion of possible units. Notes are automatically detected and then manually checked. Vibrato sections are manually segmented, and the depth and rate contours are estimated. An important characteristic of such a database is that it does not contain sung text, only sung vowels to avoid microprosody effects when extracting pitch and dynamics.

This approach controls several expression features except for timbre aspects of the singing voice. In our opinion, a positive characteristic is that it can generate expression features without suffering from smoothing as is the case in HMMs. The selected units contain the implicit rules applied by the singer to perform a vibrato, an attack, or a release. In addition, the labeling and cost functions for unit selection are designed in a way that favors the selection of long sequences of consecutive notes in the database to help the implicit reproduction of high expression features such as phrasing. Similarly to the KTH rules, this approach is independent of phonetics since this is handled separately by the controlled synthesizer, which makes it convenient for any language. The lack of an explicit timbre control could be addressed in the future by adding control features such as the degree of breathiness or brightness.

WHEN TO USE EACH APPROACH

The use of each approach has several considerations: the limitations of each one; whether singing voice recordings are available since these are needed in model training or unit selection; the reason for synthesizing a song, which could be for database creation or rule testing; or flexibility requirements such as model interpolation. In this section, we provide brief guidelines on the suitability of each type of approach.

Performance-driven approaches are suitable to be applied, by definition, when the target performance is available, since the expression of the singer is implicit in the reference audio and it can be used to control the synthesizer. Another example of applicability is the creation of parallel databases for different purposes such as voice conversion [8]. An application example for the case of speech to singing synthesis is the generation of singing

performances for untrained singers, whose timbre is taken from the speech recording and the expression for pitch and dynamics can be obtained from a professional singer.

Rule-based approaches are suitable to be applied to verify the defined rules and also to see how these are combined, for example, to convey a certain emotion. If no recordings are available, rules can still be defined with the help of an expert so that these approaches are not fully dependent on singing voice databases.

Statistical modeling approaches are also flexible, given that it is possible to interpolate models and create new voice characteristics. They have the advantage that, in some cases, these are part of complete singing voice synthesis systems, i.e., those that have the score as input and that generate both the expression parameters and output voice.

Similarly to rule-based and statistical modeling approaches, unit selection approaches do not need the target performance, although they can benefit from it. On the other hand, unit selection approaches share a common characteristic with performance-driven approaches. The implicit knowledge of the singer is contained in the recordings, although in unit selection it is extracted from shorter audio segments. Unlike statistical models, no training step is needed, so the expression databases can be improved just by adding new labeled singing voice recordings.

EVALUATION

In the beginning of this article, we explained that a score can be interpreted in several acceptable ways, making expression a subjective aspect to rate. However, "procedures for systematic and rigorous evaluation do not seem to exist today" [1, p. 105], especially if there is no ground truth to compare with. In this section, we first summarize typical evaluation strategies. Then, we propose the initial ideas to build a framework that solves some detected issues, and finally, we discuss the need for automatic measures to rate expression.

CURRENT EVALUATION STRATEGIES

Expression control can be evaluated from subjective or objective perspectives. The former typically consists of listening tests where participants perceptually evaluate some psychoacoustic characteristic such as voice quality, vibrato, and overall expressiveness of the generated audio files. A common scale is the mean opinion score (MOS), with a range from one (bad) to five (good). In pairwise comparisons, using two audio files obtained with different system configurations, preference tests rate which option achieves a better performance. Objective evaluations help to compare how well the generated expression controls match a reference real performance by computing an error.

Within the reviewed works, subjective tests outnumber the objective evaluations. The evaluations are summarized in Table 10. For each approach, several details are provided such as a description of the evaluation (style, voice quality, naturalness, expression, and singer skills), the different rated tests, and information on the subjects if available. Objective tests are done only

[TABLE 10] CONDUCTED SUBJECTIVE AND OBJECTIVE EVALUATIONS PER APPROACH.

			TESTS	
TYPE	**APPROACH**	**METHOD**	**DESCRIPTION**	**SUBJECTS**
PERFORMANCE DRIVEN	[29]	SUBJECTIVE	RATE VOICE QUALITY WITH PITCH MODIFICATION OF TEN PAIRS OF SENTENCES (SM-PSOLA VERSUS TD-PSOLA)	10 SUBJECTS
	[30]	SUBJECTIVE	INFORMAL LISTENING TEST	NOT SPECIFIED
	[31]	OBJECTIVE	TWO TESTS: LYRICS ALIGNMENT AND MEAN ERROR VALUE OF EACH ITERATION FOR F0 AND INTENSITY COMPARED TO TARGET	NO SUBJECTS
	[32]	OBJECTIVE	TWO TESTS: 3-D VOICE TIMBRE REPRESENTATION AND EUCLIDEAN DISTANCE BETWEEN REAL AND MEASURED TIMBRE	NO SUBJECTS
	[33]	SUBJECTIVE	PAIRED COMPARISONS OF DIFFERENT CONFIGURATIONS TO RATE NATURALNESS OF SYNTHESIS IN A SEVEN-STEP SCALE (–3 TO 3)	10 STUDENTS WITH NORMAL HEARING ABILITY
RULE BASED	[3]	SUBJECTIVE	LISTENING TESTS OF PARTICULAR ACOUSTIC FEATURES	15 SINGERS OR SINGING TEACHERS
	[37]	NONE	NONE	NONE
	[40]	SUBJECTIVE	LISTENING TESTS RATINGS (1–5)	50 SUBJECTS WITH DIFFERENT LEVELS OF MUSICAL TRAINING
STATISTICAL MODELING	[25]	SUBJECTIVE	LISTENING TEST (1–5 RATINGS) OF 15 MUSICAL PHRASES. TWO TESTS: WITH AND WITHOUT TIME-LAG MODEL	14 SUBJECTS
	[42]	SUBJECTIVE	NOT DETAILED (BASED ON [25])	NOT SPECIFIED
	[22]	SUBJECTIVE	RATE STYLE AND NATURALNESS LISTENING TESTS RATINGS (1–5) OF TEN RANDOM PHRASES PER SUBJECT	10 SUBJECTS
UNIT SELECTION	[43]	SUBJECTIVE	RATE EXPRESSION, NATURALNESS, AND SINGER SKILLS LISTENING TESTS RATINGS (1–5)	17 SUBJECTS WITH DIFFERENT LEVELS OF MUSICAL TRAINING

for performance-driven approaches, i.e., when a ground truth is available. In the other approaches, no reference is directly used for comparison, so only subjective tests are carried out. However, in the absence of a reference of the same target song, the generated performances could be compared to the recording of another song, as is done in the case of speech synthesis.

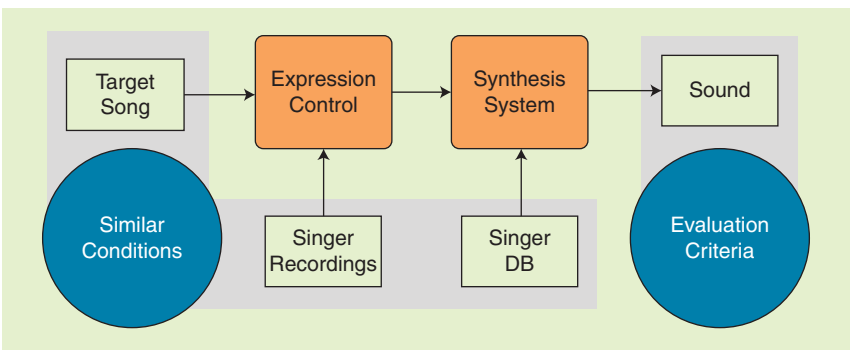
In our opinion, the described evaluation strategies are devised for evaluating a specific system and, therefore, focus on a concrete set of characteristics that are particularly relevant for that system. For instance, the evaluations summarized in Table 10 do not include comparisons to other approaches. This is due to the substantial differences between systems, which make the evaluation and comparison between them a complex task. These differences can be noted in the audio excerpts of the accompanying Web site to this article, which were introduced in the section "Comparison of Approaches." At this stage, it is difficult to decide which method more efficiently evokes a certain emotion or style, performs better vibratos, changes the voice quality in a better way, or has a better timing control. There are limitations in achieving such a comprehensive evaluation and comparing the synthesized material.

TOWARD A COMMON EVALUATION FRAMEWORK

The evaluation methodology could be improved by building the systems under similar conditions to reduce the differences among performances and by sharing the evaluation criteria. Building a common framework would help to easily evaluate and compare the singing synthesis systems.

The main blocks of the reviewed works are summarized in Figure 8. For a given target song, the expression parameters are generated to control the synthesis system. To share as many commonalities as possible among systems, these could be built under similar conditions and tested by a shared evaluation criterion. Then, the comparison would benefit from focusing on the technological differences and not on other aspects such as the target song and singer databases.

Concerning the conditions, several aspects could be shared among approaches. Currently, there are differences in the target songs synthesized by each approach, the set of controlled expression features, and the singer recordings (e.g., singer gender, style, or emotion) used to derive rules, train models, build expression databases, and build the singer voice models.



[FIG8] The proposed common evaluation framework.

A publicly available data set of songs, with scores (e.g., in MusicXML format) and reference recordings, could be helpful if used as target songs to evaluate how expression is controlled by each approach. In addition, deriving the expression controls and building the voice models from a common set of recordings would have a great impact on developing this evaluation framework. If all approaches shared such a database, it would be possible to compare how each one captures expression and generates the control parameters since the starting point would be the same for all of them. Additionally, both sample- and HMM-based synthesis systems would derive from the same voice. Thus, it would be possible to test a single expression-control method with several singing voice synthesis technologies. The main problem we envisage is that some approaches are initially conceived for a particular synthesis system. This might not be a major problem for the pitch contour control, but it would be more difficult to apply the voice timbre modeling of HMM-based systems to sample-based systems.

The subjective evaluation process is worthy of particular note. Listening tests are time-consuming tasks, and several aspects need to be considered in their design. The different backgrounds related to singing voice synthesis, speech synthesis, technical skills, and the wide range of musical skills of the selected participants can be taken into consideration by grouping the results according to such expertise, and clear instructions have to be provided on what to rate, e.g., which specific acoustic features of the singing voice to focus on, and how to rate using pairwise comparisons or MOS. Moreover, uncontrolled biases in the rating of stimuli due to the order in which these are listened to can be avoided by presenting them using pseudorandom methods such as Latin squares, and the session duration has to be short enough so as not to decrease the participant's level of attention. However, often, the reviewed evaluations are designed differently and are not directly comparable. Next, we introduce a proposal to overcome this issue.

PERCEPTUALLY MOTIVATED OBJECTIVE MEASURES

The constraints in the section "Toward a Common Evaluation Framework" make it unaffordable to extensively evaluate different configurations of systems by listening to many synthesized performances. This can be solved if objective measures that correlate with perception are established. Such perceptually motivated objective measures can be computed by learning the relationship between MOS and extracted features at a local or global scope. The measure should be ideally independent from the style and the singer, and it should provide ratings for particular features such as timing, vibrato, tuning, voice quality, or the overall performance expression. These measures, besides helping to improve the systems' performance, would represent a standard for evaluation and allow for scalability.

The development of perceptually motivated objective measures could benefit from approaches in the speech and audio processing fields. Psychoacoustic and cognitive models have been used to build objective

metrics for assessing audio quality and speech intelligibility [45], and its effectiveness has been measured by its correlation to MOS ratings. Interestingly, method-specific measures have been computed in unit selection cost functions for speech synthesis [46]. Other approaches for speech quality prediction are based on a log-likelihood measure as a distance between a synthesized utterance and an HMM model built from features based on MFCCs and F0 of natural recordings [47]. This gender-dependent measure is correlated to subjective ratings such as naturalness. For male data, it can be improved by linearly combining it with parameters typically used in narrow-band telephony applications, such as noise or robotization effects. For female data, it can be improved by linearly combining it with parameters related to signal-like duration, formants, or pitch. The research on automatic evaluation of expressive performances is considered an area to exploit, although it is still not mature enough [48]; e.g., it could be applied to develop better models and training tools for both systems and students.

Similar to the speech and instrumental music performance communities, the progress in the singing voice community could be incentivized through evaluation campaigns. These types of evaluations help to identify the aspects that need to be improved and can be used to validate perceptually motivated objective measures. Examples of past evaluation campaigns are the Synthesis Singing Challenge [52] and the Performance Rendering Contest (Rencon) http://renconmusic.org/ [48]. In the first competition, one of the target songs was compulsory and the same for each team. The performances were rated by 60 participants with a five-point scale involving the quality of the voice source, quality of the articulation, expressive quality, and the overall judgment. The organizers concluded that "the audience had a difficult task, since not all systems produced both a baritone and a soprano version, while the quality of the voices used could be quite different (weaker results for the female voice)" [52]. Rencon's methodology is also interesting. Expressive performances are generated from the same Disklavier grand piano so that the differences among approaches are only due to the performance and are subjectively evaluated by an audience and experts. In 2004, voice synthesizers were also invited. Favorable reviews were received but not included in the ranking.

CHALLENGES

While expression control has advanced in recent years, there are many open challenges. In this section, we discuss some specific challenges and consider the advantages of hybrid approaches. Next, we discuss important challenges in approaching a more human-like naturalness in the synthesis. Then, requirements for intuitive and flexible singing voice synthesizers' interfaces are analyzed, along with the importance of associating a synthetic voice with a character.

TOWARD HYBRID APPROACHES

Several challenges have been identified in the described approaches. Only one of the performance-driven approaches deals with timbre, and it depends on the available voice quality databases. This approach would benefit from techniques for the analysis of the target voice quality, its evolution over time, and techniques for voice quality transformations so to be able to synthesize any type of voice quality. The same analysis and transformation techniques would be useful for the unit selection approaches. Rule-based approaches would benefit from machine-learning techniques that learn rules from singing voice recordings to characterize a particular singer and explore how these are combined. Statistical modeling approaches currently do not utilize comprehensive databases that cover a broad range of styles, emotions, and voice qualities. If we could take databases that efficiently cover different characteristics of a singer in such a way, it would lead to interesting results such as model interpolation.

We consider the combination of existing approaches to have great potential. Rule-based techniques could be used as a preprocessing step to modify the nominal target score so that it contains variations such as ornamentations and timing changes related to the target style or emotion. The resulting score could be used as the target score for statistical and unit selection approaches where the expression parameters would be generated.

MORE HUMAN-LIKE SINGING SYNTHESIS

One of the ultimate goals of singing synthesis technologies is to synthesize human-like singing voices that cannot be distinguished from human singing voices. Although the naturalness of synthesized singing voices has been increasing, perfect human-like naturalness has not yet been achieved. Singing synthesis technologies will require more dynamic, complex, and expressive changes in the voice pitch, loudness, and timbre. For example, voice quality modifications could be related to emotions, style, or lyrics.

Moreover, automatic context-dependent control of those changes will also be another important challenge. The current technologies synthesize words in the lyrics without knowing their meanings. In the future, the meanings of the lyrics could be reflected in singing expressions as human singers do. Human-like singing synthesis and realistic expression control may be a very challenging goal, given how complex this has been proven for speech.

When human-like naturalness increases, the "Uncanny Valley" hypothesis [49] states that some people may feel a sense of creepiness. Although the Uncanny Valley is usually associated with robots and computer graphics, it is applicable even to singing voices. In fact, when a demonstration video by VocaListener [31] first appeared in 2008, the Uncanny Valley was often mentioned by listeners to evaluate its synthesized voices. An exhibition of a singer robot driven by VocaWatcher [50] in 2010 also elicited more reactions related to the Uncanny Valley. However, we believe that such a discussion of this valley should not discourage further research. What this discussion means is that the current technologies are in a transitional stage towards future technologies that will go beyond the Uncanny Valley [50] and that it is important for researchers to keep working toward such future technologies.

Note, however, that human-like naturalness is not always demanded. As sound synthesis technologies are often used to provide artificial sounds that cannot be performed by natural instruments, synthesized singing voices that cannot be performed by human singers are also important and should be pursued in parallel, sometimes even for aesthetic reasons. Some possible examples are extremely fast singing or singing with pitch or timbre quantization.

MORE FLEXIBLE INTERFACES FOR SINGING SYNTHESIS

User interfaces for singing synthesis systems will play a more important role in the future. As various score- and performance-driven interfaces are indispensable for musicians in using general sound synthesizers, singing synthesis interfaces have also had various options such as score-driven interfaces based on the piano-roll or score editor and performance-driven interfaces in which a user can just sing along with a song and a synthesis system then imitates him or her (as mentioned in the section "Performance-Driven Approaches"). More intuitive interfaces that do not require time-consuming manual adjustment will be an important goal for ultimate singing interfaces. So far, direct manipulator-style interfaces, such as the aforementioned score- or performance-driven interfaces, are used for singing synthesis systems, but indirect producer-style interfaces, such as those that enable users to verbally communicate with and ask a virtual singer to sing in different ways, will also be attractive to help users focus on how to express the user's message or intention through a song, although such advanced interfaces have yet to be developed. More flexible expression control of singing synthesis in real time is also another challenge.

MULTIMODAL ASPECTS OF SINGING SYNTHESIS

Attractive singing synthesis itself must be a necessary condition for its popularity, but it is not a sufficient condition. The most famous virtual singer, Hatsune Miku, has shown that having a character can be essential to make singing synthesis technologies popular. Hatsune Miku is the name of the most popular singing synthesis software package in the world. She is based on Vocaloid and has a synthesized voice in Japanese and English with an illustration of a cartoon girl. After Hatsune Miku originally appeared in 2007, many people started listening to a synthesized singing voice as the main vocal of music, something rare and almost impossible before Hatsune Miku. Many amateur musicians have been inspired and motivated by her character image together with her voice and have written songs for her. Many people realized that having a character facilitated writing lyrics for a synthesized singing voice and that multimodality is an important aspect in singing synthesis.

An important multimodal challenge, therefore, is to generate several attributes of a singer, such as a voice, face, and body. The face and body can be realized by computer graphics or robots. An example of simultaneous control of voice and face was shown in the combination of VocaListener [31] and VocaWatcher [50], which imitates singing expressions of the voice and face of a human singer.

In the future, speech synthesis could also be fully integrated with singing synthesis. It will be challenging to develop new voice synthesis systems that could seamlessly generate any voice produced by a human or virtual singer/speaker.

ACKNOWLEDGMENTS

We would like to thank Alastair Porter for proofreading and Merlijn Blaauw for reviewing the article. Some works presented in the article were supported in part by the Core Research for Evolutional Science and Technology funding program provided by the Japan Science and Technology Agency.

AUTHORS

Martí Umbert (marti.umbert@upf.edu) earned his M.S. degree in telecommunications at the Universitat Politècnica de Catalunya, Barcelona, Spain, in 2004. In 2010, he earned his M.Sc. degree in sound and music computing from the Universitat Pompeu Fabra (UPF) in Barcelona, Spain. He is currently a Ph.D. candidate in the Music Technology Group at UPF. He has worked on speech technologies both at the Universitat Politècnica de Catalunya and in the private sector. His research is carried out within the audio signal processing team, and he is interested in singing voice synthesis and expression modeling based on unit selection and corpus generation.

Jordi Bonada (jordi.bonada@upf.edu) earned his Ph.D. degree in computer science and digital communications from the Universitat Pompeu Fabra (UPF), Barcelona, Spain, in 2009. He is currently a senior researcher with the Music Technology Group at UPF. He is interested in singing voice modeling and synthesis. His research trajectory has an important focus on technology transfer acknowledged by more than 50 patents. He has received several awards, including the Rosina Ribalta Prize by the Epson Foundation in 2007 and the Japan Patent Attorneys Association Prize by the Japan Institute of Invention and Innovation in 2013.

Masataka Goto (m.goto@aist.go.jp) received the doctor of engineering degree from Waseda University in Tokyo, Japan, 1998. He is currently a prime senior researcher and the leader of the Media Interaction Group at the National Institute of Advanced Industrial Science and Technology, Japan. Over the past 23 years, he has published more than 220 papers in refereed journals and international conferences and has received 40 awards, including several best paper awards, best presentation awards, the Tenth Japan Academy Medal, and the Tenth Japan Society for the Promotion of Science Prize. In 2011, as the research director, he began the OngaCREST Project, a five-year Japan Science and Technology Agency-funded research project on music technologies.

Tomoyasu Nakano (t.nakano@aist.go.jp) earned his Ph.D. degree in informatics from the University of Tsukuba, Japan, in 2008. He is currently a senior researcher at the National Institute of Advanced Industrial Science and Technology, Japan. His research interests include singing information processing, human–computer interaction, and music information retrieval. He has received several awards including the Information Processing Society of Japan (IPSJ) Yamashita SIG Research Award and the Best Paper Award from the Sound and Music Computing Conference 2013. He is a member of the IPSJ and the Acoustical Society of Japan.

Johan Sundberg (jsu@csc.kth.se) earned his Ph.D. degree in musicology at Uppsala University Sweden, in 1966 and is doctor honoris causae at the University of York, United Kingdom (1996) and at the University of Athens, Greece (2014). He had a personal chair in music acoustics at KTH and was the founder and head of its music acoustics research group, which he was part of until 2001. His research concerns particularly the singing voice and music performance. He is the author of *The Science of the Singing Voice* (1987) and *The Science of Musical Sounds* (1991). He is a

member of the Royal Swedish Academy of Music, a member of the Swedish Acoustical Society (president 1976–1981), and a fellow of the Acoustical Society of America. He was awarded the Silver Medal in Musical Acoustics in 2003.

REFERENCES

[1] X. Rodet, "Synthesis and processing of the singing voice," in *Proc. 1st IEEE Benelux Workshop on Model-Based Processing and Coding of Audio (MPCA)*, 2002, pp. 99–108.

[2] P. R. Cook, "Singing voice synthesis: History, current work, and future directions," *Comput. Music J.*, vol. 20, no. 3, pp. 38–46, 1996.

[3] J. Sundberg, "The KTH synthesis of singing," *Adv. Cognit. Psychol.*, vol. 2, nos. 2–3, pp. 131–143, 2006.

[4] M. Goto, "Grand challenges in music information research," in M. Müller, M. Goto, and M. Schedl (Eds.), *Multimodal Music Processing*. Dagstuhl Publishing, Saarbrücken/Wadern, Germany, 2012, vol. 3, pp. 217–225.

[5] A. Kirke and E. R. Miranda, *Guide to Computing for Expressive Music Performance*. New York: Springer, 2013, Ch. 1.

[6] J. Bonada and X. Serra, "Synthesis of the singing voice by performance sampling and spectral models," *IEEE Signal Processing Mag.*, vol. 24, no. 2, pp. 67–79, 2007.

[7] H. Kawahara, R. Nisimura, T. Irino, M. Morise, T. Takahashi, and B. Banno, "Temporally variable multi-aspect auditory morphing enabling extrapolation without objective and perceptual breakdown," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, 2009, pp. 3905–3908.

[8] H. Doi, T. Toda, T. Nakano, M. Goto, and S. Nakamura, "Singing voice conversion method based on many-to-many eigenvoice conversion and training data generation using a singing-to-singing synthesis system," in *APSIPA/ASC*, Dec. 2012, pp. 1–6.

[9] S. Canazza, G. De Poli, C. Drioli, A. Roda, and A. Vidolin, "Modeling and control of expressiveness in music performance," *Proc. IEEE Special Issue Eng. Music*, vol. 92, no. 4, pp. 686–701, 2004.

[10] G. Widmer, "Using AI and machine learning to study expressive music performance: Project survey and first report," *AI Commun.*, vol. 14, no. 3, pp. 149–162, 2001.

[11] P. N. Juslin, "Five facets of musical expression: A psychologist's perspective on music performance," *Psychol. Music*, vol. 31, no. 3, pp. 273–302, 2003.

[12] P. N. Juslin and P. Laukka, "Communication of emotions in vocal expression and music performance: Different channels, same code," *Psychol. Bull.*, vol. 129, no. 5, pp. 770–814, 2003.

[13] S. Ternström. (2002, June). Session on naturalness in synthesized speech and music. in *Proc. 143rd Acoustical Society of America (ASA) Meeting*. [Online]. Available: http://www.pvv.ntnu.no/~farner/sonata/ternstrom02.html

[14] M. Thalén and J. Sundberg, "Describing different styles of singing: A comparison of a female singer's voice source in 'classical,' 'pop,' 'jazz,' and 'blues'," *Logoped. Phoniatrics Vocol.*, vol. 26, no. 2, pp. 82–93, 2001.

[15] A. Friberg, R. Bresin, and J. Sundberg, "Overview of the KTH rule system for musical performance," *Adv. Cognit. Psychol.*, vol. 2, nos. 2–3, pp. 145–161, 2006.

[16] N. Obin. (2011). MeLos: Analysis and modelling of speech prosody and speaking style. Ph.D. dissertation, Université Pierre et Marie Curie-Paris VI, France. [Online]. Available: http://articles.ircam.fr/textes/Obin11e/index.pdf

[17] M. Schröder, "Expressive speech synthesis: Past, present, and possible futures," in *Affective Information Processing*. London: Springer, May 2009, pp. 111–126.

[18] G. Widmer and W. Goebl, "Computational models of expressive music performance: The state of the art," *J. New Music Res.*, vol. 33, no. 3, pp. 203–216, 2004.

[19] M. Lesaffre. (2005). Music information retrieval conceptual framework, annotation and user behaviour. Ph.D. dissertation, Ghent University. [Online]. Available: https://biblio.ugent.be/publication/3258568

[20] J. Salamon, E. Gómez, D. P. W. Ellis, and G. Richard, "Melody extraction from polyphonic music signals: Approaches, applications and challenges," *IEEE Signal Processing Mag.*, vol. 31, no. 2, pp. 118–134, Mar. 2014.

[21] C. J. Plack and A. J. Oxenham, "Overview: The present and future of pitch," in *Pitch—Neural Coding and Perception*, Springer Handbook of Auditory Research. New York: Springer, 2005, vol. 24, Chap. 1, pp. 1–6.

[22] K. Saino, M. Tachibana, and H. Kenmochi, "A singing style modeling system for singing voice synthesizers," in *Proc. Interspeech*, Makuhari, Japan, Sept. 2010, pp. 2894–2897.

[23] J. Sundberg, *The Science of the Singing Voice*. Northern Illinois Univ. Press, DeKalb, Illinois, 1987.

[24] E. D. Scheirer, "Tempo and beat analysis of acoustic musical signals," *J. Acoust. Soc. Amer.*, vol. 103, no. 1, pp. 588–601, 1998.

[25] K. Saino, H. Zen, Y. Nankaku, A. Lee, and K. Tokuda, "An HMM-based singing voice synthesis system," in *Proc. Interspeech*, Pittsburgh, PA, USA, Sept. 2006, pp. 1141–1144.

[26] A. Loscos and J. Bonada, "Emulating rough and growl voice in spectral domain," in *Proc. 7th Int. Conf. Digital Audio Effects (DAFx)*, Naples, Italy, Oct. 2004, pp. 49–52.

[27] R. L. de Mantaras and J. Ll. Arcos, "AI and music: From composition to expressive performance," *AI Mag.*, vol. 23, no. 3, pp. 43–57, 2002.

[28] M. Kob, "Singing voice modelling as we know it today," *Acta Acustica*, vol. 90, no. 4, pp. 649–661, 2004.

[29] Y. Meron. (1999). High quality singing synthesis using the selection-based synthesis scheme. Ph.D. dissertation, Univ. of Tokyo. [Online]. Available: http://www.gavo.t.u-tokyo.ac.jp/~meron/sing.html

[30] J. Janer, J. Bonada, and M. Blaauw, "Performance-driven control for sample-based singing voice synthesis," in *Proc. 9th Int. Conf. Digital Audio Effects (DAFx)*, 2006, vol. 6, pp. 42–44.

[31] T. Nakano and M. Goto, "VocaListener: A singing-to-singing synthesis system based on iterative parameter estimation," in *Proc. 6th Sound and Music Computing Conf. (SMC)*, July 2009, pp. 343–348.

[32] T. Nakano and M. Goto, "VocaListener2: A singing synthesis system able to mimic a user's singing in terms of voice timbre changes as well as pitch and dynamics," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 453–456.

[33] T. Saitou, M. Goto, M. Unoki, and M. Akagi, "Speech-to-singing synthesis: Converting speaking voices to singing voices by controlling acoustic features unique to singing voices," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, Oct. 2007, pp. 215–218.

[34] J. Sundberg and J. Bauer-Huppmann, "When does a sung tone start," *J. Voice*, vol. 21, no. 3, pp. 285–293, 2007.

[35] J. Sundberg, "Synthesis of singing, in Musica e Tecnologia: Industria e Cultura per lo Sviluppo del Mezzagiorno," in *Proc. Symp. Venice*. Venedig: Unicopli, 1981, pp. 145–162.

[36] M. C. Marinescu and R. Ramirez, "A machine learning approach to expression modeling for the singing voice," in *Proc. Int. Conf. Computer and Computer Intelligence (ICCCI)*. New York: ASME Press, 2011, vol. 31, no. 12, pp. 311–316.

[37] M. Alonso. (2005). Expressive performance model for a singing voice synthesizer. Master's thesis, Universitat Pompeu Fabra. [Online]. Available: http://mtg.upf.edu/node/2223

[38] R. Bresin and A. Friberg, "Emotional coloring of computer-controlled music performances," *Comput. Music J.*, vol. 24, no. 4, pp. 44–63, Winter 2000.

[39] H. Kenmochi and H. Ohshita, "VOCALOID—Commercial singing synthesizer based on sample concatenation," in *Proc. Interspeech*, Antwerp, Belgium, Aug. 2007, pp. 4009–4010.

[40] J. Bonada. (2008). Voice processing and synthesis by performance sampling and spectral models. Ph.D. dissertation, Universitat Pompeu Fabra, Barcelona. [Online]. Available: http://www.mtg.upf.edu/node/1231

[41] T. Yoshimura, K. Tokuda, T. Masuko, T. Kobayashi, and T. Kitamura, "Simultaneous modeling of spectrum, pitch and duration in HMM-based speech synthesis," in *Proc. 6th European Conf. Speech Communication and Technology (Eurospeech)*, Budapest, Hungary, 1999, pp. 2347–2350.

[42] K. Oura, A. Mase, T. Yamada, S. Muto, Y. Nankaku, and K. Tokuda, "Recent development of the HMM-based singing voice synthesis system—Sinsy," in *Proc. Int. Speech Communication Association (ISCA), 7th Speech Synthesis Workshop (SSW7)*, Tokyo, Japan, Sept. 2010, pp. 211–216.

[43] M. Umbert, J. Bonada, and M. Blaauw, "Generating singing voice expression contours based on unit selection," in *Proc. Stockholm Music Acoustics Conference (SMAC)*, Stockholm, Sweden, Aug. 2013, pp. 315–320.

[44] M. Umbert, J. Bonada, and M. Blaauw, "Systematic database creation for expressive singing voice synthesis control," in *Proc. Int. Speech Communication Association (ISCA), 8th Speech Synthesis Workshop (SSW8)*, Barcelona, Spain, Sept. 2013, pp. 213–216.

[45] D. Campbell, E. Jones, and M. Glavin, "Audio quality assessment techniques—A review, and recent developments," *Signal Process.*, vol. 89, no. 8, pp. 1489–1500, Aug. 2009.

[46] M. Chu and H. Peng, "An objective measure for estimating MOS of synthesized speech," in *Proc. 7th European Conf. Speech Communication and Technology (Eurospeech)*, Aalborg, Denmark, Sept. 2001 pp. 2087–2090.

[47] S. Möller, F. Hinterleitner, T. H. Falk, and T. Polzehl, "Comparison of approaches for instrumentally predicting the quality of text-to-speech systems," in *Proc. Interspeech*, Makuhari, Japan, Sept. 2010, pp. 1325–1328.

[48] H. Katayose, M. Hashida, G. De Poli, and K. Hirata, "On evaluating systems for generating expressive music performance: The Rencon experience," *J. New Music Res.*, vol. 41, no. 4, pp. 299–310, 2012.

[49] M. Mori, K. F. MacDorman, and N. Kageki, "The uncanny valley [from the field]," *IEEE Robot. Automat. Mag.*, vol. 19, no. 2, pp. 98–100, June 2012.

[50] M. Goto, T. Nakano, S. Kajita, Y. Matsusaka, S. Nakaoka, and K. Yokoi, "VocaListener and VocaWatcher: imitating a human singer by using signal processing," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 5393–5396.

[51] Music Technology Group of the Universitat Pompeu Fabra website. [Online]. Available: www.mtg.upf.edu/publications/ExpressionControlinSingingVoiceSynthesis

[52] [Online]. Available: http://www.interspeech2007.org/Technical/synthesis_of_singing_challenge.php

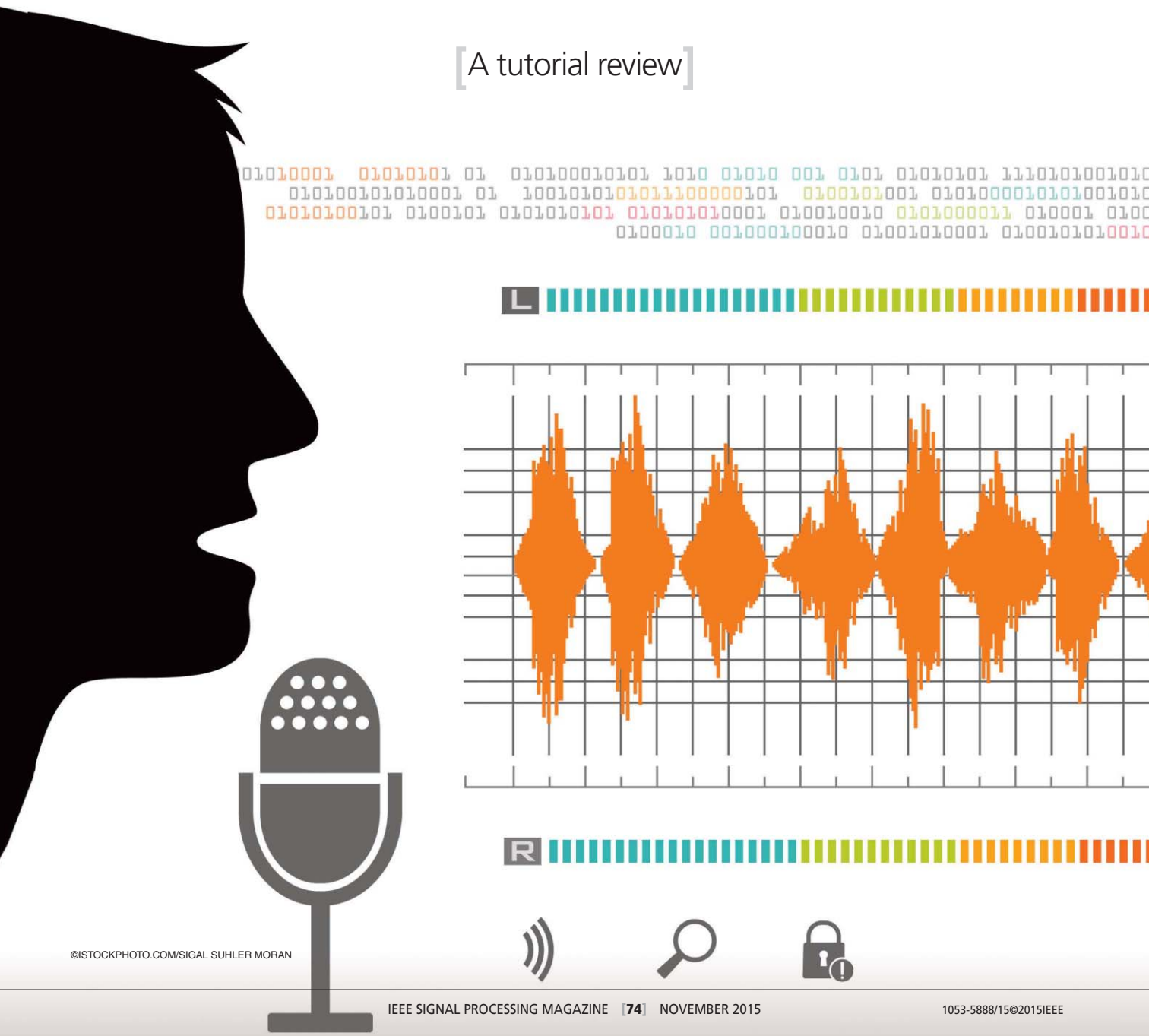
[SP]

IEEE SignalProcessing | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | Qmags THE WORLD'S NEWSSTAND®

John H.L. Hansen and Taufiq Hasan

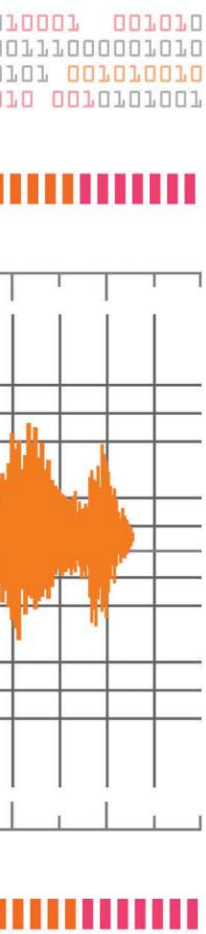
Speaker Recognition by Machines and Humans

A tutorial review



1053-5888/15©2015IEEE

Identifying a person by his or her voice is an important human trait most take for granted in natural human-to-human interaction/communication. Speaking to someone over the telephone usually begins by identifying who is speaking and, at least in cases of familiar speakers, a subjective verification by the listener that the identity is correct and the conversation can proceed. Automatic speaker-recognition systems have emerged as an important means of verifying identity in many e-commerce applications as well as in general business interactions, forensics, and law enforcement. Human experts trained in forensic speaker recognition can perform this task even better by examining a set of acoustic, prosodic, and linguistic characteristics of speech in a general approach referred to as *structured listening*. Techniques in forensic speaker recognition have been developed for many years by forensic speech scientists and linguists to help reduce any potential bias or preconceived understanding as to the validity of an unknown audio sample and a reference template from a potential suspect. Experienced researchers in signal processing and machine learning continue to develop automatic algorithms to effectively perform speaker recognition—with ever-improving performance—to the point where automatic systems start to perform on par with human listeners. In this article, we review the literature on speaker recognition by machines and humans, with an emphasis on prominent speaker-modeling techniques that have emerged in the last decade for automatic systems. We discuss different aspects of automatic systems, including voice-activity detection (VAD), features, speaker models, standard evaluation data sets, and performance metrics. Human speaker recognition is discussed in two parts—the first part involves forensic speaker-recognition methods, and the second illustrates how a naïve listener performs this task from a neuroscience perspective. We conclude this review with a comparative study of human versus machine speaker recognition and attempt to point out strengths and weaknesses of each.



INTRODUCTION

Speaker recognition and verification have gained increased visibility and significance in society as speech technology, audio content, and e-commerce continue to expand. There is an ever-increasing need to search for audio materials, and searching based on speaker identity is a growing interest. With emerging technologies such as Watson, IBM's supercomputer [1], which can compete with expert human players in the game of "Jeopardy," and Siri [2], Apple's powerful speech-recognition-based personal assistant, it is not hard to imagine a future when handheld devices will be an extension of our identity—highly intelligent, sympathetic, and fully functional personal assistants, which will not only understand the meaning of what we say but also recognize and track us by our voice or other identifiable traits.

As we increasingly realize how much sensitive information our personal handheld devices can contain, it will become critical that effective biometric authentication be an integral part of access to information and files contained on the device, with a potential range of public/private access. Because speech is the most natural means of human communication, these devices will unavoidably lean toward automatic voice-based authentication in addition to other forms of biometrics. Apple's recent iPhone models have already introduced fingerprint scanners, reflecting the industry trend. The latest Intel technology on laptops employs face recognition as the password for access. Our digital personal assistant, in theory, could also replace most forms of traditional key locks as well for our home and vehicles, again making security of such a personal device more important.

Apart from personal authentication for access control, speaker recognition is an important tool in law enforcement, national security, and forensics in general. Because of widespread availability, cell phones have become the primary means of communication for the general public, and, unfortunately, also for criminals. Unlike the domain of personal authentication for personal files/information access, these individuals usually do not want to be recognized. In such cases, many criminals may attempt to alter their voice to prevent them from being identified. This introduces additional challenges for developers of speaker-recognition technology—"Is the participant a willing individual in being assessed?" In law enforcement, any voice recorded as part of evidence may be disguised or even synthesized, to obscure recognition, adding to the difficulty of being recognized. Over a number of years, forensic speech scientists have devised different strategies to overcome these difficulties.

Interestingly, humans routinely recognize individuals by their voices with striking accuracy, especially when the degree of familiarity with the subject is high (i.e., close acquaintances or public figures). Many times, even a short nonlinguistic queue, such as a laugh, is enough for us to recognize a familiar person [3]. On the other hand, it is also common knowledge that we cannot recognize a once-heard voice very easily and sometimes have difficulty in recognizing familiar voices over the phone. With these ideas in mind, a naïve person may wonder what exactly makes speaker recognition difficult and why is it a topic of such rigorous research.

Digital Object Identifier 10.1109/MSP.2015.2462851
Date of publication: 13 October 2015

From the discussion so far, it is safe to say that speaker recognition can be accomplished in three ways.

- We can recognize familiar voices with considerable ease without any conscious training. This form of speaker recognition can be termed *naïve speaker recognition*.
- In forensic investigations, speech samples from a telephone call are often compared to recordings of potential suspects (i.e., from a phone threat, emergency 911 call, or known criminal). In these cases, trained listeners are involved in systematically comparing the speech samples to provide an informed decision concerning their similarities. We would classify this as *forensic speaker recognition*.
- Finally, we have *automatic speaker recognition,* where the complete speech analysis and decision-making process is performed using computer analysis.

In both naïve and forensic speaker recognition, humans are directly involved in the process, even though some automatic or computer-assisted means may be used to supplement knowledge extraction for the purposes of comparison in the forensic scenario. However, it should be noted that both the forensic and automatic methods are highly systematic, and the procedures from both disciplines are technical in nature.

The forensic and automatic speaker-recognition research communities have developed various methods more or less independently for several decades. Conversely, naïve recognition is a natural ability of humans—which is, at times, very accurate and effective. Recent studies on brain imaging [4], [5] have revealed many details on how we perform cognitive-based speaker recognition, which may inspire new directions for both automatic and forensic methods.

In this article, we present a tutorial review of the automatic speaker-recognition methods, especially those developed in the last decade, while providing the reader with a perspective on how humans also perform speaker recognition, especially by forensics experts and naïve listeners. The aim is to provide a discussion on the three classes of speaker recognition, highlighting the important similarities and differences among them. We emphasize how automatic techniques have evolved over time toward more current approaches. Many speech-processing techniques, such as Mel-scale filter-bank analysis and concepts in noise masking, are inspired by human auditory perception. Also, there are similarities between the methods used by forensic voice experts and automated systems—even though, in many cases, the research communities are separate. We believe that incorporating the perspective of speech perception by humans in this review, including highlights of both strengths and weaknesses in speaker recognition compared to machines, will help broaden the view of the reader and perhaps inspire new research directions in the area.

SPEAKER-RECOGNITION TASKS

First, to consider the overall research domain, it would be useful to clarify what is encompassed by the term *speaker recognition*, which consists of two alternative tasks: speaker identification and verification. In speaker identification, the task is to identify an unknown speaker from a set of known speakers. In other words, the goal is to find the speaker who sounds closest to the speech stemming from an unknown speaker within an audio sample. When all speakers within a given set are known, it is called a *closed* or *in-set scenario*. Alternatively, if the potential input test subject could also be from outside the predefined known speaker group, this becomes an open-set scenario, and, therefore, a world model or universal background model (UBM) [6] is needed. This scenario is called *open-set speaker recognition* (also *out-of-set speaker identification*).

In speaker verification, an unknown speaker claims an identity, and the task is to verify if this claim is true. This essentially comes down to comparing two speech samples/utterances and deciding if they are spoken by the same speakers. In some methods, this is done by comparing the unknown sample against two alternative models, the claimed speaker model and a world model. In the forensic scenario, the general task is to identify the unknown speaker, who is suspected of a crime, but, in many instances, verification is also necessary.

Speaker recognition can be based on an audio stream that is text dependent or text independent. This is more relevant in authentication applications—where a claimed user says something specific, such as a password or personal identification number, to gain access to some resource/information. Throughout this article, the focus will be on text-independent speaker verification, especially in the treatment of automatic systems.

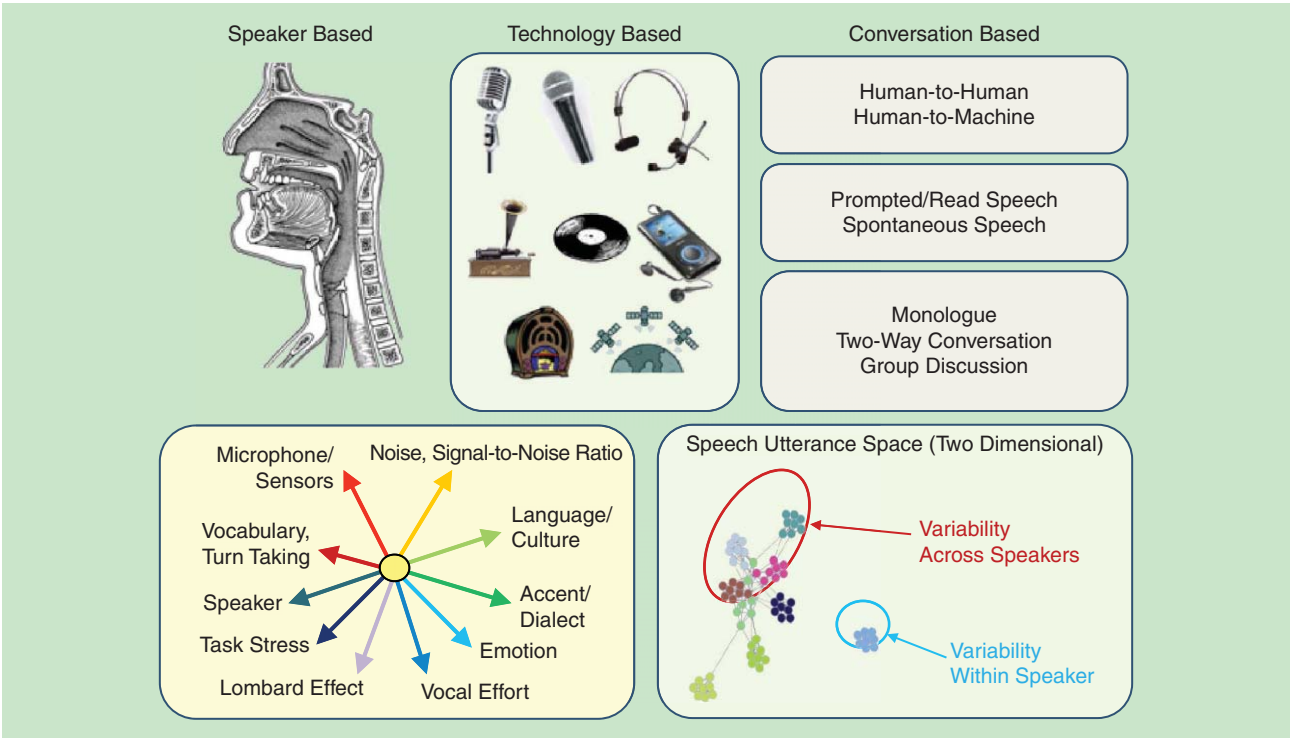
CHALLENGES IN SPEAKER RECOGNITION

Unlike other forms of biometrics (e.g., fingerprints, irises, facial features, gait, and hand geometry) [7], human speech is a performance biometric. Simply put, the identity information of the speaker is embedded (primarily) in how speech is spoken, not necessarily in what is being said (although in many voice forensic applications, it is also necessary to identify who said what within a multispeaker discussion). This makes speech signals prone to a large degree of variability. It is important to note that even the same person does not say the same words in exactly the same way every time (this is known as *style shifting* or *intraspeaker variability*) [8]. Also, various recording devices and transmission methods commonly used exacerbate the problem. For example, we may find it difficult to recognize someone's voice through a telephone or maybe when the person is not healthy (i.e., has a cold) or is performing another task or speaking with a different level of vocal effort (i.e., whispering or shouting).

SOURCES OF VARIABILITY IN SPEAKER RECOGNITION

To consider variability, Figure 1 highlights a range of factors that can contribute to mismatch for speaker recognition. These can be partitioned based on three broad classes: 1) speaker based, 2) conversation based, and 3) technology based. Also, variability for speakers can be within speakers and across speakers.

- *Speaker-based variability sources:* these reflect a range of changes in how a speaker produces speech and will affect system performance for speaker recognition. These can be thought of as *intrinsic* or *within-speaker variability* and include the following factors.



[FIG1] **Sources of variability in speaker recognition.**

- *Situational task stress*—the subject is performing some task while speaking, such as operating a vehicle (car, plane, truck, etc.), hands-free voice input (factory setting, emergency responders/fire fighters, etc.), which can include cognitive as well as physical task stress [9].
- *Vocal effort/style*—the subject alters his or her speech production from normal phonation, resulting in whispered [10], [11], soft, loud, or shouted speech; the subject alters his or her speech production mechanism to speak effectively in the presence of noise [12], known as the Lombard effect; or the subject is singing versus speaking [13].
- *Emotion*—the subject is communicating his or her emotional state while speaking (e.g., anger, sadness, happiness, etc.) [14].
- *Physiological*—the subject has some illness or is intoxicated or under the influence of medication; this can include aging as well.
- *Disguise*—the subject intentionally alters his or her voice to circumvent the system. This can be by natural means (speaking in a harsh voice to avoid detection, mimicking another person's voice, etc.) or using a voice-conversion system.

■ *Conversation-based/higher-level mode/language of speaking variability sources*: these reflect different scenarios with respect to the voice interaction with either another person or technology system, or differences with respect to the specific language or dialect spoken, and can include

- *human-to-human*: speech that includes two or more individuals interacting or one person speaking and addressing an audience
  - —language or dialect spoken
  - —if speech is read/prompted (through visual display or through headphones), spontaneous, conversational, or disguised speech
  - —monologue, two-way conversation, public speech in front of an audience or for TV or radio, group discussion
- *human-to-machine*: speech produced where the subject is directing his or her speech toward a piece of technology (e.g., cell/smart/landline telephone and computer)
  - —*prompted speech*: voice input to a computer
  - —*voice input for telephone/dialog system/computer input*: interacting with a voice-based system.

■ *Technology- or external-based variability sources*: these include how and where the audio is captured and the following issues:

- *electromechanical*—transmission channel, handset (cell, cordless, and landline) [15]–[17] microphone
- *environmental*—background noise [18] (stationary, impulsive, time-varying, etc.), room acoustics [19], reverberation [20], and distant microphone
- *data quality*—duration, sampling rate, recording quality, and audio codec/compression.

These multifaceted sources of variation pose the greatest challenge in accurately modeling and recognizing a speaker, whether automatic algorithms are used, or if human listening/assessment is performed. Given that speech will contain variability, the task of speaker verification is deciding if the variability is due to the same speaker (intra {within}-speaker) or different speakers (inter {across}-speaker).

In current automated speaker-recognition technology, various mathematical tools are used to mitigate the effects of these

variability/degradations, especially the extrinsic ones. Additive noise and transmission channel variability have received much attention recently. Intrinsic variability in speech is very difficult to quantify and account/address for in automatic assessment. Higher-level knowledge may become important in these cases. For example, even if a person's voice (spectral characteristics) may change due to his or her current health (e.g., a cold) or aging, the person's accent or style of speech remains generally the same. Forensic experts pay special attention to these details when detecting a subject's voice from potential suspects' speech recordings.

CHALLENGES IN SPEAKER RECOGNITION

Early efforts in speaker recognition involving technology focused more on the telecommunications domain, where telephone handset and communication channel variation was the primary concern. In the United States, when telephone systems were confined to handheld rotary phones in the home and public phone booths in public settings, technology- and telephony-based variability was an issue, but it was, to a large degree, significantly less important than it is today. With mobile cell phone/smartphone technology dominating the world's telecommunications market, the diversity of telephony scenarios has expanded considerably. Virtually all cell phones have a speaker option, which allows voice interaction at a distance from the microphone, and movement of the device introduces a wider range of channel variability.

Voice is also a time-varying entity. Research has shown that intersession variability, the inherent changes present within audio files captured at different times, results in changes in speaker-recognition performance. Analysis of the Multisession Audio Research Project corpus collected using the same audio equipment in the same location on a monthly basis over a 36-month period showed measurable differences in speaker-recognition performance [21], [22]. However, the changes in speaker-identification performance seem to be independent of the time difference between training and testing [21], [23]. While no aging effects were noted for the 36-month period, other research has demonstrated long-term changes in speech physiology and production due to aging [23]. More extensive research that explores the evolution of speaker structure for speaker recognition over a 20–60-year period (at least for a small subset of speakers) has shown measurable changes and suggested methods to address changes due to aging [24], [25].

These examples of variation point to the sensitivity of existing speaker-recognition technology. It is possible to employ such technology in a way that could lead to noncredible results. A recent example of how to wrongly use automatic speaker recognition was seen during the recent U.S. legal case involving George Zimmerman, who was accused of shooting Trayvon Martin during an argument [26]. In that case, a 911 emergency call captured a scream for help heard in the background. The defense team claimed that it was Zimmerman who was yelling while supposedly being attacked by Trayvon Martin, who was killed; alternatively, the prosecutors argued that it was the unarmed victim who was shouting. Parents of both parties testified that the voice heard on the 911 call belonged to their own son. Some forensic experts did attempt to use semiautomatic methods to compare the original scream and a simulated scream obtained from Zimmerman. The issue of using automatic assessment schemes for scream analysis to assess identity was controversial, as experts from the U.S. Federal Bureau of Investigation (FBI) and U.S. National Institute of Standards and Technology (NIST) testified that these methods are unreliable. A brief probe analysis of scream and speaker-recognition technology confirmed the limitations of current technology [27].

Most forensic speaker-identification scenarios, however, are not as complicated. When there is sufficient speech material available from the offender and the suspect, systematic analysis can be performed to extract speaker idiosyncratic characteristics, also known as feature parameters, from the speech data, and a comparison between the samples can be made. Also, in automatic speaker-identification systems, features designed to differentiate among speakers are first extracted and mathematically modeled to perform a meaningful comparison. Thus, in the next section, we consider what traits help identify a person from his or her speech—in other words, what are the feature parameters that we should consider in making an assessment?

SPEAKER CHARACTERIZATION: FEATURE PARAMETERS

Every speaker has some characteristic traits in his or her voice that are unique. Individual speaker characteristics may not be so easily distinguishable but are unique mainly due to speaker vocal tract physiology and learned habits of articulation. Even identical twins have differences in their voices, though studies show they have similar vocal tract shape [28] and acoustic properties [29], and it is difficult to distinguish them from a perceptual/forensics perspective [30], [31]. Researchers in voice forensics have even participated in the National Twins Day event held in Twinsburg, Ohio, [32] in an effort to capture voice and other biometrics to explore the challenges in distinguishing closely related individuals. Thus, whether recognition is performed by humans (an expert or naïve listener) or by machines, some measurable and predefined aspects of speech need to be considered to make meaningful comparisons among voices. Generally, we refer to these characterizing aspects as *feature parameters*.

One might expect that a unique voice must have unique features, but this is not always true. For example, two different speakers may have the same speaking rate (which is a valid feature parameter) but differ in average pitch. This is complicated by the variability and degradations discussed previously, which is why considering multiple feature parameters is critical.

PROPERTIES OF IDEAL FEATURES

As outlined by Nolan [33], ideally a feature parameter should

1) show high between-speaker variability and low within-speaker variability
2) be resistant to attempted disguise or mimicry
3) have a high frequency of occurrence in relevant materials
4) be robust in transmission
5) be relatively easy to extract and measure.

These properties, though mentioned in the forensic speaker-identification context, apply in general. Interestingly, Wolf [34] discussed very similar sets of properties in the context of features for

automatic speaker recognition, independently, preceding Nolan [33]. We refer to these properties as *ideal property 1–5* throughout this article. It should be reiterated that variability in features will always exist, but the important task is to determine if the origin of the variability is the same speaker or different speakers.

We now discuss various feature parameters used in forensic speaker identification, which can be and are also useful for general speech understanding. There is no fixed set of rules for what parameters should be used in forensic speaker recognition. This is largely dependent on the circumstances or availability [35]. Some forensic experts may choose parameters to compare based on the most obvious aspect of the voices under consideration. Feature parameters can be broadly classified into auditory versus acoustic, linguistic versus nonlinguistic, and short-term versus long-term features.

AUDITORY VERSUS ACOUSTIC FEATURES

Some aspects of speech are better suited for auditory analysis (i.e., through listening). Auditory features are thus defined as aspects of speech that can "be heard and objectively described" by a trained listener [36]. These can be specific ways of uttering individual speech sounds (e.g., the pronunciation of the vowel sounds in the word *hello* can be used as auditory features).

Acoustic features, on the other hand, are mathematically defined parameters derived from the speech signal using automatic algorithms. Clearly, these kinds of features are used in automatic systems, but they are also used in computer-assisted forensic speaker recognition. Fundamental frequency (F0) and formant frequency bandwidth are examples of acoustic features. Automatic systems frequently use acoustic features derived from the short-term power spectrum of speech.

Both auditory and acoustic features have their strengths and weaknesses. Two speech samples may sound very similar but have highly variant acoustic parameters [37]. Alternatively, speech samples may sound very different yet have similar acoustic features [28]. It is thus generally accepted that both auditory and acoustic features are indispensable for forensic investigations [35]. One might argue that if reverse engineering of the human auditory system [38] is fully successful, auditory features can also be extracted using automatic algorithms.

LINGUISTIC VERSUS NONLINGUISTIC FEATURES

Linguistic feature parameters can provide contrast "within the structure of a given language or across languages or dialects" [35]. They can be acoustic or auditory in nature and are classified further as phonological, morphological, and syntactic [36]. A simple example of a linguistic feature is whether the "r" sound at the end of a word, e.g., *car,* is pronounced or silent—in some dialects of English, this type of "r" sound is not pronounced (i.e., Lancashire versus Yorkshire dialects of U.K. English). This is different from an auditory analysis of how the "r" sound is pronounced, since, in this case, this speech sound will be compared across different words.

Nonlinguistic features include aspects of speech that are not related to the speech content. Typical nonlinguistic features may include: speech quality (nasalized, breathy, husky, etc.), fluency, speech pauses (frequency and type), speaking rate, average fundamental frequency, and nonspeech sounds (coughs, laughs, etc.). Again, these features can be auditory or acoustic in nature. Referring back to the Zimmerman case, the manner of screaming (i.e., loudness, pitch, and duration) could be a potential feature if it could be properly measured/parameterized.

SHORT-TERM VERSUS LONG-TERM FEATURES

Depending on the time span of the feature parameters, they can be categorized as short versus long term. Most features discussed so far are short term or segmental in nature. Popular automatic systems mostly use short-term acoustic features, especially the ones extracted from the speech spectrum. The short-term features are also effective in auditory forensic analysis, for example, direct comparison of the "r" sound and consonant–vowel transition [33].

The long-term features are usually averaged short-term parameters, (e.g., fundamental frequency, short-term spectrum). These parameters have the benefit of being insensitive to fluctuations due to individual speech sounds and provide a smoother measurement from a speech segment. The long-term features also include energy, pitch, and formant contours, which are measured/averaged over long time periods. Recent automatic systems also successfully used such features [39]–[41]. If a feature parameter is extracted from an entire speech utterance, we refer to it as an *utterance-level feature,* or *utterance feature* for short. This concept will become very useful as we proceed with the discussion to automatic systems.

FORENSIC SPEAKER RECOGNITION

While the focus in this review is on automatic machine-based speaker recognition, we also briefly consider both forensic and naïve speaker recognition. The need for forensic speaker recognition/identification arises when a criminal leaves his or her voice as evidence, be it as a telephone recording or speech heard by an earwitness. The use of technology for forensic speaker recognition has been discussed as early as 1926 [42] with speech waveforms. Later, the spectrographic representation of speech was developed at AT&T Bell Laboratories during World War II. It was popularized much later, in the 1970s, when it came to be known as the *voiceprint* [43]. As the name suggests, the voiceprint was presented as being analogous to fingerprints and with very high expectations. Later, the reliability of the voiceprint for voice identification, from its operating mechanisms to formal procedure, was thoroughly questioned and argued [44], [45], even called "an idea gone wrong" [45]. It was simply not accurate with speech being so subject to variability. Most researchers today believe it to be controversial at best. A chronological history of voiceprints can be found in [46], and an overview discussion on forensic speaker recognition can be found in [47]. Here, we present an overview with respect to current trends.

In the general domain of forensic science, the United States has recently formed the Organization of Scientific Area Committees (OSAC) (http://www.nist.gov/forensics/osac.cfm), which is overseen by NIST. The legacy structure before OSAC was Forensic Science Working Groups. The current OSAC organization was established to help formalize the process of best practices for standards as they relate to researchers, practitioners, legal and law enforcement as well as government agencies. It also allows for a

more transparent process in which experts and users of the various technologies can provide feedback and help shape best practices. Currently, OSAC is establishing a number of working documents to build consensus among the various forensic subfields. A good source of current information from OSAC is the NIST Forensic Science Publications website (http://www.nist.gov/forensics/publications.cfm).

Today, forensic speaker identification is commonly performed by expert phoneticians who generally have backgrounds in linguistics and statistics. This is a very complex procedure, and varies among practitioners. There is no standard set of procedures every practitioner agrees upon. Different aspects/features are considered when forensic experts make comparisons between utterances. The procedure is often dictated by the situation at hand—for example, if only a few seconds of screaming of the unknown speaker is available on the evidence tape, the only thing that can be done is to try to recreate a similar scream from the likely speaker (suspect) and compare, which is generally not feasible.

THE LIKELIHOOD RATIO

Regardless of the varying approaches by practitioners, forensic speaker recognition essentially entails a scientific and objective method of comparing voices (there are, apparently, people who attempt to perform this task using methods unacceptable by the general forensic community [48]). Forensic experts must testify in court concerning the similarity/dissimilarity of the speech samples in consideration in a meaningful way. However, they cannot make any categorical judgment about the voices (e.g., the two voices come from the same speaker). For this purpose, the likelihood ratio (LR) [49] measure was introduced, which forensic experts use to express the strength of their findings [50], [51]. This means that the evaluation of forensic speech samples will not yield an absolute identification or elimination of the suspect but instead provides a probabilistic confidence measure. As discussed previously, even speech samples from the same speaker will differ in realistic scenarios. The goal of the forensic voice comparison expert is thus to estimate the probability of observing the measured difference between speech samples assuming that they were spoken by 1) the same speaker and 2) different speakers [35]. The procedure for measuring the LR is given next:

- X = Speech sample recorded during a crime (evidence recording).
- Y = Speech sample obtained from suspect (exemplar).
- H_0 = The hypothesis that X and Y are spoken by the same person.
- H_1 = The hypothesis that X and Y are spoken by different persons.
- E = Observed forensic evidence (e.g., average pitch from X and Y differ by 10 Hz).

The LR formula is

$$\text{LR} = \frac{p(E \mid H_0)}{p(E \mid H_1)}.$$

As an example, if the average pitch difference between two utterances is considered the feature parameter, the forensic expert first computes the probability distribution of this feature parameter for speech data collected from many same-speaker (hypothesis H_0) and different-speaker (hypothesis H_1) pairs. In the next step, given the evidence E (average pitch from X and Y differ by 10 Hz), the conditional probabilities $p(E \mid H_0)$, and $p(E \mid H_1)$, can be computed. Note that the forensic expert does not try to estimate $p(H_0 \mid E)$ (i.e., the probability that the suspect is guilty given the observed evidence). This is because this estimation is done using Bayes' theorem, which requires the prior probabilities of the hypotheses generally not provided to the expert (and are also difficult to estimate). More discussion on this can be found in [35, Ch. 4].

APPROACHES IN FORENSIC SPEAKER IDENTIFICATION

Here, we discuss general approaches taken for forensic speaker recognition. The methods described are performed by human experts, fully or partially. While full automatic approaches are also considered for forensics, we discuss automatic speaker recognition in later sections.

AUDITORY APPROACH

This approach is practiced by auditory phoneticians and involves producing a detailed transcript of the evidence tape and exemplars. Drawing on their experience, experts listen to speech samples and attempt to detect any aspects of the voices that are unusual, distinctive, or noteworthy [51]. The experience of the expert is obviously an important factor in deciding about rarity or typicality. The auditory features discussed previously are used in this approach.

The auditory approach is fully subjective, unless it is combined with other approaches. Although the LR can be used to express the outcome of the analysis, practitioners of the auditory approach generally do not use it. Instead, based on their comparison of auditory features, they present an evidentiary statement (a formal statement describing the basis of the evidence) in court.

AUDITORY-SPECTROGRAPHIC APPROACH

As discussed previously, the spectrographic approach, previously known as voiceprint analysis, is based on visual comparison of speech spectrograms. Generally, the same word or phrase is extracted from the known and questioned voices and their spectrograms are visually analyzed. Additional foil speakers' (background speakers) spectrograms are also included to facilitate in understanding similarity versus typicality. It is believed that visual comparison using spectrograms together with listening to the audio reinforces the voice identification procedure [44], [45], which is why the approach is termed *auditory-spectrographic*.

Following the controversy on voiceprints, the spectrographic method evolved in various ways. It was not evident if forensic experts could differentiate between intraspeaker (changes of speech from the same speaker) and interspeaker (changes in speech due to different speakers) variation by a general visual comparison of spectrographs. Thus, different protocols evolved that require the forensic examiner to analyze predefined aspects of the spectrographs. According to the American Board of Recorded Evidence (ABRE) protocols, the examiner is required to visually analyze and compare

aspects such as general formant shaping and positioning, pitch striations, energy distribution, word length, and coupling (nasality). It also requires auditory comparisons of pitch, stress/emphasis, speaking rate, disguise, mode, etc. [51], [52].

The auditory-spectrographic, similar to the auditory approach, is also subjective and depends heavily on the experience of the examiner. Courts in some jurisdictions do not accept testimony based on this approach. The FBI seeks advice from auditory-spectrographic experts during investigations but does not allow them to testify in court [51].

ACOUSTIC–PHONETIC APPROACH

This approach, which is commonly taken by experts trained on acoustic–phonetics, requires quantitative acoustic measurements from speech samples, and statistical analysis of the results. Acoustic features discussed previously are ones that are considered. Generally, similar phonetic units are extracted from the known and questioned speech samples, and various acoustic parameters measured from these segments are compared. The LR can be conveniently used in this approach since it is based on numerical parameters [51].

Although the acoustic–phonetic approach is a more objective approach, it does have some subjective elements. For example, an acoustic–phonetician may identify speech sounds as being affected by stress (through listening) and then perform objective analysis. However, whether the speaker was actually under stress at that moment is a subjective quantity determined by the examiner through his or her experience. It is a matter of debate if having a human element in the forensic speaker-recognition process is advantageous [51].

Forensic speaker identification will continue to be an important research area in the coming future. As evident from the discussion, the methods are evolving toward mathematical and statistical approaches, perhaps signaling that the human element in this process may actually be a source of error. The NIST has conducted studies on human-assisted speaker recognition (HASR) comparing human experts and state-of-the-art algorithms [20]. In these experiments, a set of difficult speaker pairs (i.e., same speakers that sound different in two recordings or different speakers that sound similar) were selected. The results indicated that the state-of-the-art fully automatic systems outperformed the human-assisted systems. We discuss these studies further in the "Man Versus Machine in Speaker Recognition" section.

NAÏVE SPEAKER RECOGNITION

The ability to recognize people by their voices is an acquired human trait. Research shows that we are able to recognize our mothers' voice from as early as the fetus stage [53], [54]. We analyze many different aspects of a person's voice to identify him or her, including spectral characteristics, language, prosody, and speaking style. We learn and remember these traits constantly without even putting in a conscious effort. In this section, we discuss various aspects of how a naïve listener identifies a speaker and what is currently known about the speaker-recognition process in the human brain.

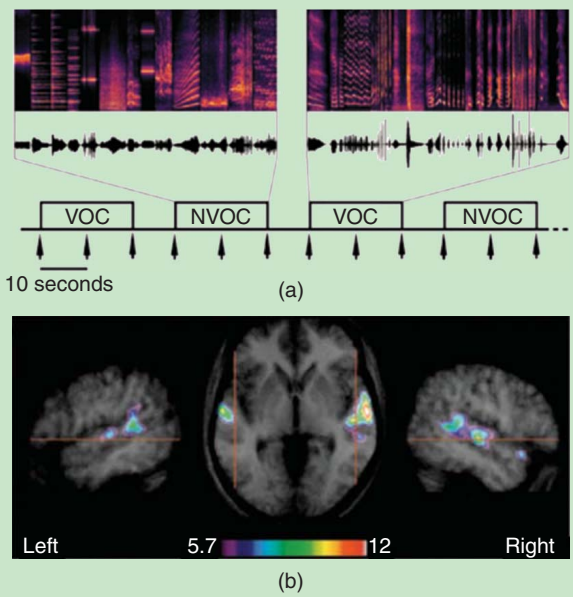
IDENTIFY SPEECH SEGMENTS

An important aspect of detecting speakers from audio samples is to first identify speech segments. Humans can efficiently distinguish between speech and nonspeech sounds from a very early age [55]. This is observed from highly voice-selective cerebral activity measured by functional magnetic resonance imaging (fMRI) in the adult human brain [4], [55], [56]. Figure 2 shows the brain regions that demonstrate higher neural activity with vocal and nonvocal stimuli. Note that in this experiment, any sound produced by a human is considered vocal (irrespective of being voiced or unvoiced), including laughs and coughs. In later sections, we discuss a very similar process required by automatic systems as a preprocessing step before performing speaker recognition.

SPEAKER RECOGNITION VERSUS DISCRIMINATION

It is obvious that we need to be familiar with a person's voice before identifying him or her. Familiarity is a subjective condition, but it is apparent that being familiar with a person depends on how much time the subject has spent in listening to that person. In other words, familiarity with a speaker depends on the amount of speech data observed by the listener. The familiar person can be a close acquaintance (e.g., a friend or relative) or someone famous (e.g., a celebrity or political leader).

Interestingly, familiar voice recognition and unfamiliar voice discrimination are known to be separate cognitive abilities [57].



[FIG2] An experiment on finding voice-selective regions of the human brain using fMRI. (a) The experimental paradigm: spectrograms (0–4 kHz) and amplitude waveforms of examples of auditory stimuli. Vocal (VOC) and nonvocal (NVOC) stimuli are presented in 20-second blocks with 10-second silence intervals. (b) Voice-sensitive activation regions in the group average: regions with significantly ($P < 0.001$) higher response to human voices than to energy-matched nonvocal stimuli are shown in color scale (t-value) on an axial slice of the group-average MRI (center) and on sagittal slices (vertical plane dividing the brain into left and right halves) of each hemisphere. (Figure adapted from [4].)

Familiar voice recognition is essentially a pattern-recognition task—humans can perform this task even if the speech signal is reversed [58]. These findings suggest that unfamiliar voice discrimination involves analysis of speech features as well as the pattern-recognition ability of the brain [57]. Forensic examiners heavily depend on the ability to discriminate since they are not usually familiar with the speakers in the speech samples involved.

The findings in [57] also imply that voice discrimination ability of the human brain is not a preprocessing step of voice recognition, since these two processes are found to be independent. For automatic systems, however, this is not usually true. The same algorithms can be used (usually with slight modification) to discriminate between speakers or identify a specific speaker. In many cases, discriminative training methods are used to learn speaker models, which can later be used to identify speakers. We discuss automatic systems further in the "Automatic Speaker Recognition" section.

FAMILIARITY WITH LANGUAGE

It is observed in [59] that humans are better at recognizing people who are familiar and speak a known language. Experiments reported in this study show that native English speakers with normal reading ability could identify voices speaking English significantly more accurately than voices speaking Chinese. Thus, the voice-recognition ability of humans depends on their familiarity with the phonology of the particular language. Humans can still recognize people speaking an unknown language, but with much lower accuracy [59].

ABSTRACT REPRESENTATIONS OF SPEECH

The human brain forms efficient abstract representations from relevant audio features that contain both phonetic and speaker identity information. These representations aid in efficient processing and high robustness due to noise and other forms of degradations. These aspects of the brain were studied in [5], where the authors have shown that it is possible to decipher both speech content and speaker identity by observing neural activity of the human listener. The brain activities were measured by fMRI and it was found that there are certain observable patterns corresponding to speech and voice stimuli elicit in the listener's auditory cortex. This is illustrated in Figure 3, where vowel (red) and speaker (blue) discriminative regions in the brain are shown.

SPEAKER RECOGNITION IN THE BRAIN: FINAL REMARKS

There is still much more to discover about the human brain and how it processes information. From what we already know, the human brain performs complex spectral and temporal audio processing [60], is sensitive to vocal stimuli [4], shows familiarity to the phonology of languages [59], and builds abstract representations of speech and speaker information that are robust to noise and other degradations [5]. Most of these abilities are highly desirable in automatic systems, especially the brain's ability to process noisy speech. It is thus natural to attempt to mimic the human brain in solving these problems. Research efforts are already underway to reverse engineer the processes performed by the human auditory pathway [38].

As discussed previously, the human brain processes familiar speakers differently than unfamiliar ones [55], [57]. This may mean that faithfully comparing human and machine performance in a speaker-recognition task can be very difficult since it is not well understood how to quantify familiarity with a person from an automatic system's perspective—what amount of data is enough for the system to be familiar with that person? Nevertheless, it will be interesting to be able to determine exactly how the human brain stores the speaker identity information of familiar speakers. These findings may lead to breakthrough algorithmic advances in the automatic speaker-recognition area.

As we conclude this section, we want to highlight the strengths and weaknesses of humans in the speaker-recognition task. Here, humans include both forensic examiners and naïve listeners.

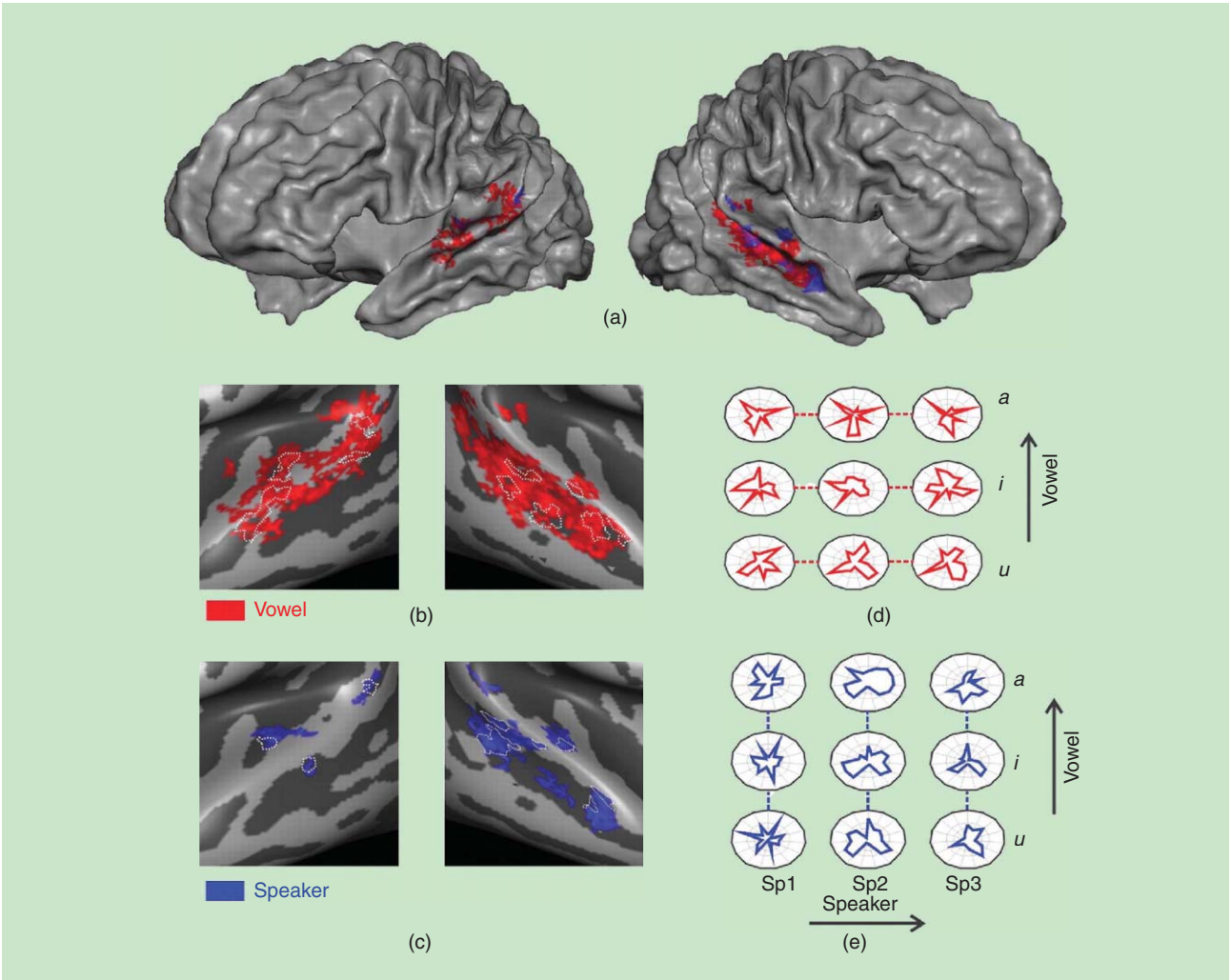
STRENGTHS OF HUMAN LISTENERS

- Humans (naïve listeners and experts alike) can identify familiar speakers with remarkable accuracy, even in challenging conditions (normal, disguised, and stressed) [61].
- Humans are good at finding the idiosyncrasies of a speaker's voice. Thus, the forensic examiner may easily identify where to look. For example, a speaker may cough in a specific manner, which a human will notice very quickly.

WEAKNESSES OF HUMAN LISTENERS

- Humans are susceptible to contextual bias [62]. For example, if the forensic examiner knows that a suspect already confessed to a crime, he is more likely to find a match between the exemplar and evidence recording.
- Humans are prone to error. The reliability of voiceprints was questioned mostly due to human errors involved in the process [46].
- Humans cannot remember a speaker's voice for a long time [63]. Memory retention ability depends on the duration of speech heard by the listener [64].
- For familiar speakers, the listener may confuse them with someone else. The subject may know that the voice is familiar but may not correctly identify exactly who the speaker is.
- Naïve listeners cannot distinguish subtle differences between voices. However, trained experts can. For example, the difference between New York and Boston accents is distinguishable by an expert but probably not by naïve listeners [35].
- Humans perform better while they are attentive. However, the attention level drops with time, and listeners tend to become fatigued after a certain time.
- The outcome of voice comparison results as LRs may not be consistent across multiple experts (or the same expert at different times).
- Human listeners (including forensic experts) may seem to identify someone from a voice recording if they are expecting to hear that person.

Concluding the discussion on speaker recognition by humans, we now move forward with the main focus of this review, which is automatic systems for speaker recognition.

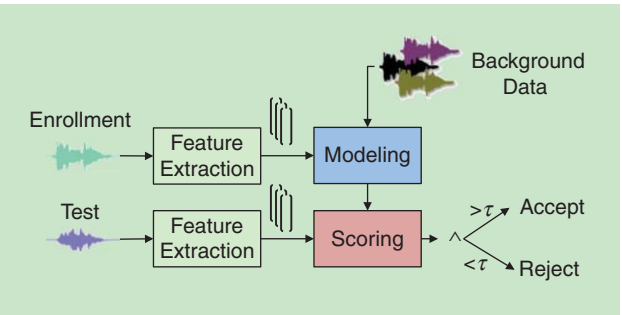


[FIG3] (a)–(c) The regions of the human brain that contribute the most in discriminating between vowels (red) and speakers (blue). (b) and (c) Enlarged representations of the auditory cortex (region of the brain sensitive to sounds). (d) and (e) Activation patterns of sounds created from the 15 most discriminative voxels (of the fMRI) for decoding (d) vowels and (e) speakers. Each axis of the polar plot forming a pattern displays the normalized activation level in a voxel. Note the similarity among the patterns of the same vowel [horizontal direction in (d)] or speaker [vertical direction in (e)]. (Figure reprinted from [5].)

AUTOMATIC SPEAKER RECOGNITION

In automatic speaker recognition, computer programs designed to operate independently with minimum human intervention identify a speaker's voice. The system user may adjust the design parameters, but to make the comparison between speech segments, all the user needs to do is provide the system with the audio recordings. In the current discussion, we focus our attention on the text-independent scenario and the speaker-verification task. Naturally, the challenges mentioned previously affect the automatic systems in the same way as they do the human listeners or forensic experts. Various speaker-verification approaches can be found in the literature that address specific challenges; see [65]–[74] for a comprehensive tutorial review on automatic speaker recognition. The research community is largely driven by standardized tasks set forth by NIST through the speaker-recognition evaluation (SRE) campaigns [75]–[78]. We discuss the NIST SRE tasks in more detail in later sections.

A simple block diagram representation of an automatic speaker-verification system is shown in Figure 4. Predefined feature parameters are first extracted from the audio recordings that are designed to capture the idiosyncratic characteristics of a



[FIG4] An overall block diagram of a basic speaker-verification system.

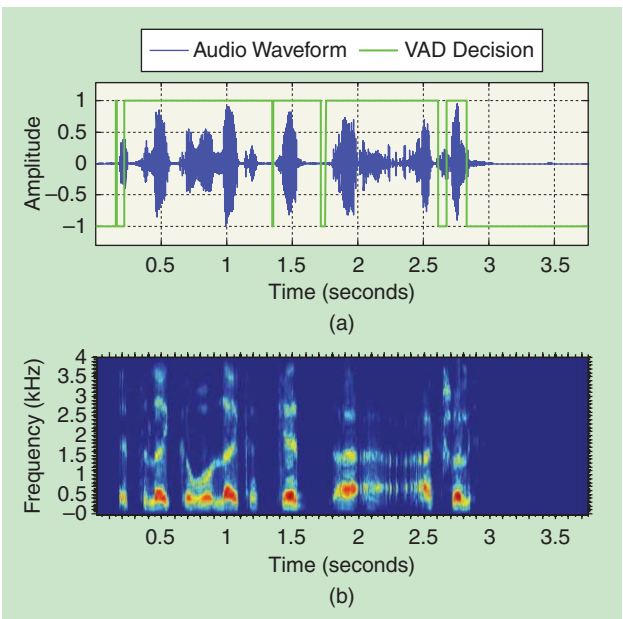
person's speech in mathematical parameters. These features obtained from an enrollment speaker are used to build/train mathematical models that summarize their speaker-dependent properties. For an unknown test segment, the same features are then extracted, and they are compared against the model of the enrollment/claimed speaker. The models are designed so that such a comparison provides a score (a scalar value) indicating whether the two utterances are from the same speaker. If this score is higher (or lower) than a predefined threshold then the system accepts (or rejects) the test speaker.

It should be noted that the block diagram in Figure 4 for speaker verification is a simplified one. As we discuss more about the standard speaker-recognition systems of today, features can be extracted from short-term segments of speech, a relatively longer duration of speech, or the entire utterance. The classification of features discussed previously also applies in this case.

In some automatic systems, the feature-extraction processes may be dependent on other speech utterances spoken by a diverse speaker population, as well as the enrollment speaker [79]. In short, the recent techniques make use of the general properties of human speech by observing many different speech recordings to make effective speaker-verification decisions. This is also intuitive, since we also learn how human speech varies across conditions over time. For example, if we only heard one language in our entire life, we would have difficulty distinguishing people speaking a different language [59].

FEATURE PARAMETERS IN AUTOMATIC SPEAKER-RECOGNITION SYSTEMS

As mentioned previously, feature parameters extracted from an entire utterance are referred to as *utterance features* in this article. This becomes more important in the automatic speaker-recognition context as many common pattern-recognition algorithms operate on fixed dimension vectors. Because of the variable length/duration property of speech, acoustic/segmental features cannot be directly used with such classifiers. However, simple methods such as averaging segmental features over time do not seem to be highly effective in this case, due to the time-varying nature and context dependency of speech [80], [81]. For example, taking speaking rate as a feature, it is obvious that two people may commonly have the same speaking rate, so this feature by itself may not be very useful. Researchers noted early on that a specific speaker's idiosyncratic features will be time varying and context/speech sound dependent [34], [66]. However, the high-level and long-term features such as dialect, accent, speaking style/rate, and prosody are also useful and can be beneficial when used together with low-level acoustic features [39], [82].



[FIG5] (a) A speech waveform with voice-activity decisions (1 versus 0 values indicate speech versus silence) and (b) a spectrogram plot of the corresponding speech waveform.

VAD

As noted previously, humans are good at distinguishing between speech and nonspeech sounds, which is also an essential part in auditory forensic speaker recognition. Clearly, in automatic systems it is also desirable that features be extracted only from speech segments of the audio waveform, which necessitates VAD [83], [84]. Detecting speech segments becomes critical when highly noisy/degraded acoustic conditions are considered. The function of VAD is illustrated in Figure 5(a), where speech presence/absence is indicated by a binary signal overlaid on the speech samples. The corresponding speech spectrogram is shown in Figure 5(b). The VAD algorithm used in this plot is presented in [83], though more advanced unsupervised solutions such as Combo-Speech Activity Detection (SAD) have recently emerged as successful in diverse audio conditions for speaker recognition [85].

SHORT-TERM FEATURES

These features refer to parameters extracted from short speech segments/frames of duration within 20–25 milliseconds. The most popular short-term acoustic features are the Mel-frequency cepstral coefficients (MFCCs) [86] and linear predictive coding (LPC)-based [87] features. For a review on different short-term acoustic features for speaker recognition, see [71] and [73]. We briefly discuss the MFCC features here. To obtain these coefficients from an audio recording, first the audio samples are divided into short overlapping segments. A typical 25-millisecond speech signal frame is shown in Figure 6(a). The signal obtained in these segments/frames is then multiplied by a window function (e.g., Hamming and Hanning), and the Fourier power spectrum is obtained. In the next step, the logarithm of the spectrum is computed and nonlinearly spaced Mel-space filter-bank analysis is performed. The logarithm operation expands the scale of the coefficients and also decomposes multiplicative components to additive [88]. The filter-bank analysis produces the spectrum energy in each channel (also known as the *filter-bank energy coefficients*), representing different frequency bands.

A typical 24-channel filter bank and its outputs are shown in Figure 6(c) and (d), respectively. As evident here, the filter bank is designed so that it is more sensitive to frequency variations in the lower end of the spectrum, similar to the human auditory system

[86]. Finally, MFCCs are obtained by performing discrete cosine transform (DCT) on the filter-bank energy parameters and retaining a number of leading coefficients. DCT has two important properties: 1) it compresses the energy of a signal to a few coefficients and 2) its coefficients are highly decorrelated. For these reasons, removing some dimensions using DCT improves modeling efficiency and reduces some nuisance components. Also, the decorrelation property of DCT helps the models that assume feature coefficients are uncorrelated. In summary, the following sequence of operations—power spectrum, logarithm, and DCT—produces the well-known cepstral representation of a signal [88]. Figure 6(e) shows the static MFCC parameters, retaining the first 12 coefficients after DCT. Generally, velocity and acceleration parameters computed across multiple frames of speech are appended to the MFCCs. These parameters (known as *deltas* and *double deltas*, respectively) represent the dynamic properties of the short-term feature coefficients.

FEATURE NORMALIZATION

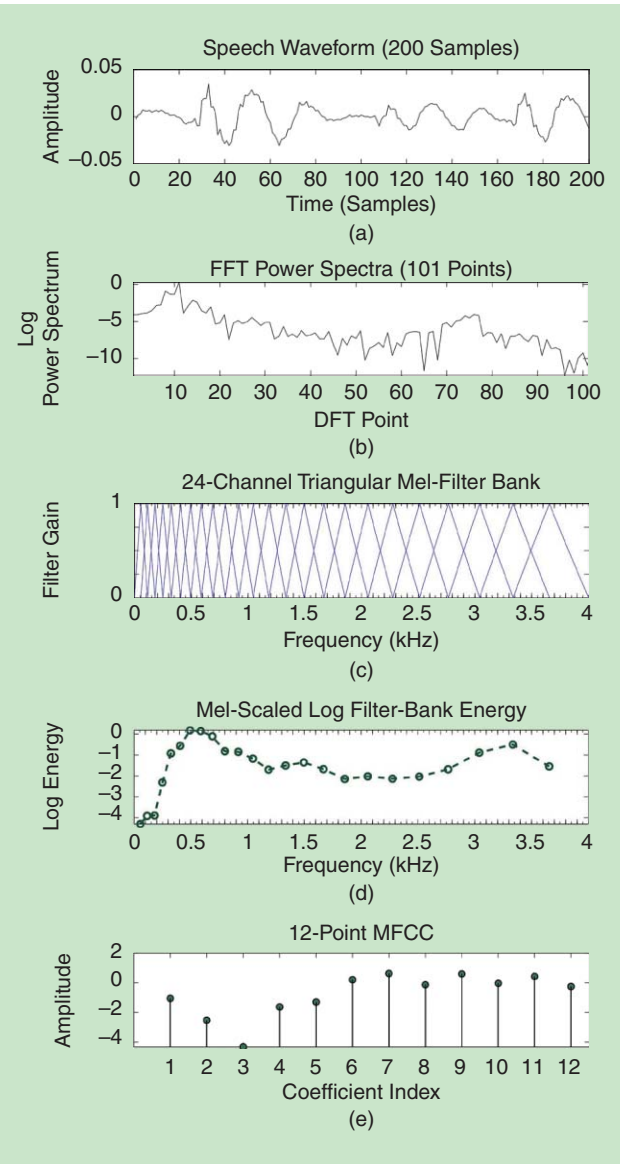
As stated previously, one of the desirable properties of acoustic features (and any feature parameter in a pattern-recognition problem) is robustness to degradation. This is one of the desirable characteristics of an ideal feature parameter [34]. In reality, it is not possible to design a feature parameter that will be absolutely unchanged in modified acoustic conditions and also provide meaningful speaker-dependent information. However, these changes can be minimized in various ways using feature-normalization techniques such as cepstral mean subtraction [89], feature warping [90], relative spectra (RASTA) processing [91], and quantile-based cepstral normalization [92]. It should be noted that normalization techniques are not designed to enhance the discriminative ability of the features (ideal property 3), rather they aim to modify the features so that they are more consistent among different speech utterances (ideal property 5). Popular normalization schemes include feature warping and cepstral mean and variance normalization.

SPEAKER MODELING

Once the audio segments are converted to feature parameters, the next task of the speaker-recognition process is modeling. In general terms, we can define modeling as a process of describing the feature properties for a given speaker. The model must also provide means of its comparison with an unknown utterance. A modeling method is robust when its characterizing process of the features is not significantly affected by unwanted distortions, even though the features are. Ideally, if features could be designed in such a way that no intraspeaker variation is present while interspeaker discrimination is maximum, the simplest methods of modeling might have sufficed. In essence, the nonideal properties of the feature extraction stage requires various compensation techniques during the modeling phase so that the effect of the nuisance variations observed in the signal are minimized during the speaker-verification process.

Most speaker-modeling techniques make various mathematical assumptions on the features (Gaussian distributed, for example). If these properties are not met by the data, we are essentially introducing imperfections during the modeling phase as well. The normalization of features can alleviate these problems to some extent, but not entirely. Consequently, mathematical models are forced to fit the features and recognition scores are derived based on these models and test data. Thus, this process introduces artifacts in the detection scores, and a family of score-normalization techniques has been proposed in the past to encounter this final-stage mismatch [17].

In summary, degradations in the acoustic signal affect features, models, and scores. Thus, improving robustness of speaker-recognition systems is important in these three domains. Recently, it has been observed that as speaker-modeling techniques are improved, score-normalization techniques become less



[FIG6] Steps in MFCC feature extraction from a speech frame: (a) 200-sample frame representing 25 milliseconds of speech sampled at a rate of 8 kHz, (b) DFT power spectrum showing first 101 points, (c) 24-channel triangular Mel-filter bank, (d) log filter-bank energy outputs from Mel-filter, and (e) 12 static MFCCs obtained by performing DCT on filter-bank energy coefficients and retaining the first 12 values.

effective [93], [94]. Similarly, we can argue that if acoustic features are improved, simple modeling techniques will be sufficient. However, from the speaker-recognition research trend in the last decade, it seems that improving feature robustness beyond a certain level (for a variety of degradations) is extremely difficult—or, in other words, data-driven modeling techniques have been more successful in improving robustness compared to new features. This is especially true if large data sets are used in training strong discriminative models. In the recent approaches for speech recognition, simple filter-bank energy features are found to be more effective than MFCCs when large neural networks are used for modeling [95]. Also, modeling techniques that aim at learning the behavior of the degradations from example speech utterances are at an advantage in improving robustness. For example, an automatic system that has observed several examples of speech recordings of different speakers in roadside noise will be better at distinguishing speakers in that environment.

In the following sections, we discuss how state-of-the-art systems have evolved during the last decade. We emphasize a few key advancements made during this time.

GAUSSIAN-MIXTURE-MODEL-BASED METHOD

A Gaussian mixture model (GMM) is a combination of Gaussian probability density functions (PDFs) generally used to model multivariate data. The GMM clusters the data in an unsupervised way (i.e., without any labeled data), but it provides a PDF of the data. Using GMMs to model a speaker's features results in a speaker-dependent PDF. Evaluating the PDF at different data points (e.g., features obtained from a test utterance) provides a probability score that can be used to compute the similarity between a speaker GMM and an unknown speaker's data. For a simple speaker-identification task, a GMM, is first obtained for each speaker. During testing, the utterance is compared against each GMM, and the most likely speaker (i.e., the highest-scoring GMM) is selected.

In text-independent speaker-recognition tasks when there is no a priori knowledge about the speech content, using GMMs to model short-term features has been found to be most effective for acoustic modeling. This is expected since the average behavior of the short-term spectral features is more speaker dependent rather than being affected by the temporal characteristics. It was first used in a speaker-recognition method in [96]. Before GMMs were introduced, the vector quantization (VQ) method [81], [97], [98] was used for speaker recognition. This technique models the speaker using a set of prototype vectors instead of PDFs. GMMs have been shown to be better speaker models compared to VQ because of their probabilistic nature for allowing greater variability. Therefore, even when the test utterance has a different acoustic condition, GMMs, being a probabilistic model, can relate to the data better than the more restrictive VQ model (see "GMM-Based Speaker Recognition: Summary").

A GMM is a mixture of Gaussian PDF parameterized by a number of mean vectors, covariance matrices, and weights of the individual mixture components. The model is represented by a weighted sum of the individual PDFs. If a random vector $\mathbf{x}_n$ can be modeled by M Gaussian components with mean vectors $\boldsymbol{\mu}_g$, covariance matrices Σ_g, where $g = 1, 2 \ldots M$ indicate the component indices, the PDF of $\mathbf{x}_n$ is given by

$$f(\mathbf{x}_n \mid \lambda) = \sum_{g=1}^{M} \pi_g \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_g, \Sigma_g), \tag{1}$$

where π_g indicates the weight of the gth mixture component. We denote the GMM model as $\lambda = \{\pi_g, \mu_g, \Sigma_g \mid g = 1 \ldots M\}$. The likelihood of a feature vector given the GMM model can be evaluated using (1). Acoustic feature vectors are generally assumed to be independent. For a sequence of feature vectors $\mathcal{X} = \{\mathbf{x}_n \mid n \in 1 \ldots T\}$, the probability of observing these features given the GMM model is computed as

$$p(\mathcal{X} \mid \lambda) = \prod_{n=1}^{T} p(\mathbf{x}_n \mid \lambda).$$

Note that the order of the features is irrelevant in calculating the likelihood, which simplifies the computation for text-dependent speaker recognition. A GMM is usually trained using the expectation-maximization (EM) algorithm [99], which iteratively increases the likelihood of the data given the model.

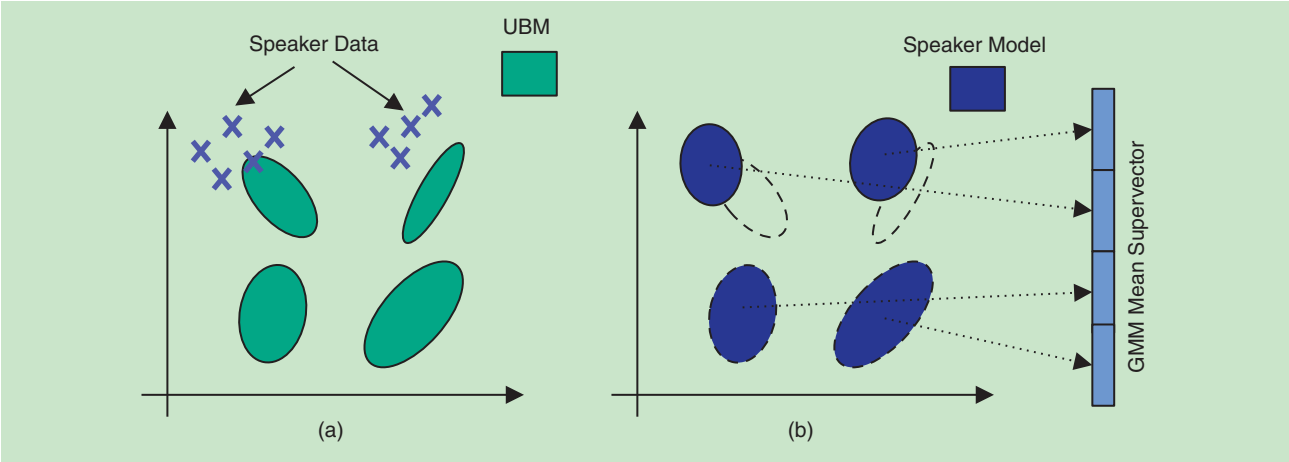
GMM-BASED SPEAKER RECOGNITION: SUMMARY

First proposed	Reynolds et al. (1995) [96]
Previous methods	Averaging of long-term features, VQ-based methods [80], [97], [98]
Proposed method	Model features using GMMs, compute similarity using feature likelihood
Why robust?	The probabilistic nature of GMM allows more variability in the data

ADAPTED GMMs: THE GMM–UBM SPEAKER-VERIFICATION SYSTEM

The GMM approach has been effective in speaker-identification tasks. For speaker verification, apart from the claimed speaker model, an alternate speaker model (representing speakers other than the target) is needed. In this way, these two models can be compared with the test data and the more likely model can be chosen, leading to an accept or reject decision. The alternate speaker model, also known as the *background* or *world model*, initiated the idea of using a UBM that represents everyone except the target speaker. It is essentially a large GMM trained to represent the speaker-independent distribution of the speech features for all speakers in general. The block diagram in Figure 4 becomes clear now since the background model is assumed to exist. Note that the UBM is assumed to be a "universal" model that serves as the alternate model for all enrolled speakers. Some methods have considered providing speaker-dependent unique background models [100], [101]. However, using a single background model has been the most effective and meaningful strategy.

The UBM was first introduced as an alternate speaker model in [102]. Later, in [6], the UBM was used as an initial model for the enrollment speaker GMMs. This concept was a significant



[FIG7] A schematic diagram of a GMM–UBM system using a four-mixture UBM. MAP adaptation procedure and supervector formation by concatenating the mean vectors are also illustrated. (a) A schematic diagram of a GMM–UBM system using a four-mixture UBM. (b)MAP adaptation procedure and supervector formation by concatenating the mean vectors are also illustrated.

advancement achieved by the so-called GMM–UBM method. In this approach, a speaker's GMM is adapted or derived from the UBM using Bayesian adaptation [103]. In contrast to performing maximum likelihood training of the GMM for an enrollment speaker, this model is obtained by updating the well-trained UBM parameters. This relation between the speaker model and the background model provides better performance than independently trained GMMs and also lays the foundation for the speaker model adaptation techniques that were developed later. We will return to these relations as we proceed. In the following subsections, we describe the formulations of this approach.

The LR Test

Given an observation O and a hypothesized speaker s, the task of speaker verification can be stated as a hypothesis test between

$$H_0 : O \text{ is from speaker } s,$$
$$H_1 : O \text{ is not from speaker } s.$$

In the GMM–UBM approach, the hypothesis H_0 and H_1 are represented by a speaker-dependent GMM λ_s and the UBM λ_0. Thus, for the set of observed feature vectors $\mathrm{X} = \{\mathbf{x}_n \,|\, n \in 1 \ldots T\}$, the LR test is performed by evaluating the following ratio:

$$\frac{p(\mathcal{X} \,|\, \lambda_s)}{p(\mathcal{X} \,|\, \lambda_0)} \begin{cases} \geq \tau \text{ accept } H_0 \\ < \tau \text{ reject } H_0 \end{cases},$$

where τ is the decision threshold. Usually, the LR test is performed in the logarithmic scale, providing the so-called log-LR

$$\Lambda(\mathcal{X}) = \log p(\mathcal{X} \,|\, \lambda_s) - \log p(\mathcal{X} \,|\, \lambda_0). \tag{2}$$

Maximum A Posteriori Adaptation of UBM

Let $\mathcal{X} = \{\mathbf{x}_n \,|\, n \in 1 \ldots T\}$ denote the set of acoustic feature vectors obtained from the enrollment speaker s. Given a UBM as in (1) and the enrollment speaker's data X, at first the probabilistic alignment of the feature vectors with respect the UBM components is calculated as

$$p(g \,|\, \mathbf{x}_n, \lambda_0) = \frac{\pi_g p(\mathbf{x}_n \,|\, g, \lambda_0)}{\sum_{g=1}^{M} \pi_g \, p(\mathbf{x}_n \,|\, g, \lambda_0)} = \gamma_n(g).$$

Next, the values of $\gamma_n(g)$ values are used to calculate the sufficient statistics for the weight, mean, and covariance parameter as

$$N_s(g) = \sum_{n=1}^{T} \gamma_n(g),$$
$$\mathrm{F}_s(g) = \sum_{n=1}^{T} \gamma_n(g)\mathbf{x}_n,$$
$$S_s(g) = \sum_{n=1}^{T} \gamma_n(g)\mathbf{x}_n\mathbf{x}_n^T.$$

These quantities are known as the zero-, first-, and second-order Baum–Welch statistics, respectively. Using these parameters, the posterior mean and covariance matrix of the features given the data vectors $\mathcal{X}$ can be found as

$$E_g[\mathbf{x}_n \,|\, \mathcal{X}] = \frac{\mathrm{F}_s(g)}{N_s(g)},$$
$$E_g[\mathbf{x}_n \, \mathbf{x}_n^T \,|\, \mathcal{X}] = \frac{\mathrm{S}_s(g)}{N_s(g)}.$$

The maximum a posteriori (MAP) adaptation update equations for weight, mean, and covariance, (3), (4), and (5), respectively, are proposed in [103] and used in [6] for speaker verification

$$\hat{\pi}_g = [\alpha_g N_s(g)/T + (1-\alpha_g)\pi_g]\beta, \tag{3}$$
$$\hat{\mu}_g = \alpha_g E_g[\mathbf{x}_n \,|\, \mathcal{X}] + (1-\alpha_g)\,\boldsymbol{\mu}_g, \tag{4}$$
$$\hat{\Sigma}_g = \alpha_g E_g[\mathbf{x}_n \, \mathbf{x}_n^T \,|\, \mathcal{X}] + (1-\alpha_g)(\Sigma_g + \mu_g\mu_g^T) - \hat{\mu}_g\hat{\mu}_g^T. \tag{5}$$

The scaling factor β in (3) is computed from all the adapted mixture weights to ensure that they sum to unity. Thus, the new GMM parameters are a weighted summation of the UBM parameters and the sufficient statistics obtained from the observed data (see "GMM–UBM System: Summary"). The variable α_g is defined as

$$\alpha_g = \frac{N_s(g)}{N_s(g) + r}. \tag{6}$$

GMM–UBM SYSTEM: SUMMARY

First proposed	Reynolds et al. (2000) [6]
Previous methods	GMM models for enrollment, cohort speakers as background
Proposed method	Adapt speaker GMMs from a UBM
Why robust?	Speaker models adapted from a well-trained UBM is more reliable than directly trained GMMs for each speaker

Here, r is known as the relevance factor. This parameter controls how the adapted GMM parameter will be affected by the observed speaker data. In the original study [6], this parameter was defined differently for the model weight, mean, and covariance. However, since only adaptation of the mean vectors turned out to be the most effective, we only use one relevance factor in our discussion here. Figure 7 shows an example of MAP adaptation for a two-dimensional feature space with a four-mixture UBM case.

THE GMM SUPERVECTORS

One of the issues with speaker recognition is that the training and test speech data can be of different durations. This requires the comparison of two utterances of different lengths. Thus, one of the efforts toward effective speaker recognition has always been to obtain a fixed-dimensional representation of a single utterance [80]. This is extremely useful since many different classifiers can be used on these utterance-level features from the machine-learning literature. One effective solution to obtaining a fixed-dimensional vector from a variable-duration utterance is the formation of a GMM supervector, which is essentially a large vector obtained by concatenating the parameters of a GMM model. Generally, a GMM supervector is obtained by concatenating the GMM mean vectors of a MAP-adapted speaker model, as illustrated in Figure 7.

The term *supervector* was first used in this context for eigenvoice speaker adaptation in speech recognition applications [104]. For speaker recognition, supervectors were first introduced in [105], motivating new model adaptation strategies involving eigenvoice and MAP adaptation. Researchers realized that these large dimensional vectors are a very good platform for designing channel compensation methods. Various effective modeling techniques were proposed to operate on the supervector space. The two dominating trends observed in these efforts were based on factor analysis (FA) and support vector machines (SVMs). They will be discussed next.

GMM SUPERVECTOR SVMs

SVMs [106] are one of the most popular supervised binary classifiers in machine learning. In [107], it was observed that GMM supervectors could be effectively used for speaker recognition/verification using SVMs. The supervectors obtained from the training utterances were used as positive examples while a set of impostor utterances were used as negative examples. Channel compensation strategies were also developed in this domain, such as nuisance attribute projection (NAP) [108] and within-class covariance normalization (WCCN) [109]. Other approaches used SVM models for speaker recognition using short- and long-term features [39], [110]. However, using GMM supervectors with SVM and NAP provided the most effective solution (see "GMM-SVM System: Summary").

GMM-SVM SYSTEM: SUMMARY

First proposed	Campbell et al. (2006) [107]
Previous methods	Adapted GMM-based methods, GMM–UBM system
Proposed method	Use GMM supervector as utterance features, classify using SVMs
Why robust?	Combines the effectiveness of adapted GMM as an utterance model and the discriminating ability of the SVM

SVMs

An SVM classifier aims at optimally separating multidimensional data points obtained from two classes using a hyperplane (a high-dimensional plane). The model can then be used to predict the class of an unknown observation depending on its location with respect to the hyperplane. Given a set of training vectors and labels $(\mathbf{x}_n, y_n)$ for $n \in \{1 \ldots T\}$, where $\mathbf{x}_n \in \Re^d$ and $y_n \in \{-1, +1\}$, the goal of SVM is to learn the function $f: \Re^d \rightarrow \Re$ so that the class label of an unknown vector $\mathbf{x}$ can be predicted as

$$I(\mathbf{x}) = \text{sign}(f(\mathbf{x})).$$

For a linearly separable data set [106], a hyperplane H given by $\mathbf{w}^T\mathbf{x} + b = 0$, can be obtained that separates the two classes, so that

$$y_n(\mathbf{w}^T\mathbf{x}_n + b) \geq 1, n = 1 \ldots T.$$

An optimal linear separator H provides the maximum margin between the classes, i.e., the distance between H and the projections of the training data from the two different classes are maximum. The maximum margin is found to be $2/\|\mathbf{w}\|$ and data points $\mathbf{x}_n$ for which $y_n(\mathbf{w}^T\mathbf{x}_n + b) = 1$, (i.e., points that lie on the margins, are known as *support vectors*). In a simple two-dimensional case, the operation of SVM is illustrated in Figure 8. When training data are not linearly separable, the features can be mapped into a higher-dimensional space using Kernel functions where the classes become linearly separable. For more details on SVM training and kernels, refer to [106] and [111]. Compensation strategies that are developed for SVM-based speaker recognition (e.g., NAP and WCCN) are discussed in later sections.

FA OF THE GMM SUPERVECTORS

FA aims at describing the variability in high-dimensional observable data vectors using a lower number of unobservable/hidden

variables. For speaker recognition, the idea of explaining the speaker- and channel-dependent variability using FA in the GMM supervector space was first discussed in [112]. Many variants of FA methods were employed since then, which finally led to the current state-of-the-art i-vector approach [79]. In this section, we discuss these methods briefly to illustrate how the techniques have evolved.

[TABLE 1] A SUMMARY OF THE LINEAR STATISTICAL MODELS IN SPEAKER RECOGNITION.

MODEL	FORMULATION	REMARKS
CLASSICAL MAP	$\mathbf{m}_s = \mathbf{m}_0 + \mathbf{D}\mathbf{z}_s$	$\mathbf{D}$ IS DIAGONAL, $\mathbf{z}_s \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$
EIGENVOICE	$\mathbf{m}_s = \mathbf{m}_0 + \mathbf{V}\mathbf{y}_s$	$\mathbf{V}$ IS LOW RANK, $\mathbf{y}_s \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$
EIGENCHANNEL	$\mathbf{m}_{s,h} = \mathbf{m}_0 + \mathbf{D}\mathbf{z}_s + \mathbf{U}\mathbf{x}_h$	$\mathbf{U}$ IS LOW RANK, $(\mathbf{z}_s, \mathbf{x}_h) \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$
JFA	$\mathbf{m}_{s,h} = \mathbf{m}_0 + \mathbf{U}\mathbf{x}_h + \mathbf{V}\mathbf{y}_s + \mathbf{D}\mathbf{z}_{s,h}$	$\mathbf{U}, \mathbf{V}$ ARE LOW RANK, $(\mathbf{x}_h, \mathbf{y}_s, \mathbf{z}_{s,h}) \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$
i-VECTOR	$\mathbf{m}_{s,h} = \mathbf{m}_0 + \mathbf{T}\mathbf{w}_{s,h}$	$\mathbf{T}$ IS LOW RANK, $\mathbf{w}_{s,h} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$

Linear Distortion Model

In the discussions to follow, a speaker-dependent GMM supervector $\mathbf{m}_s$ is generally assumed to be a linear combination of four components. These components are as follows:

1) speaker-/channel-/environment-independent component $(\mathbf{m}_0)$
2) speaker-dependent component $(\mathbf{m}_{\mathrm{spk}})$
3) channel-/environment-dependent component $(\mathbf{m}_{\mathrm{chn}})$
4) residual $(\mathbf{m}_{\mathrm{res}})$.

Component 1 is usually obtained from the UBM and is a constant. Components 2–4 are random vectors and are responsible for variability in the supervectors due to different phenomena. Using this model, a GMM supervector obtained from speaker s and session h is written as

$$\mathbf{m}_{s,h} = \mathbf{m}_0 + \mathbf{m}_{\mathrm{spk}} + \mathbf{m}_{\mathrm{chn}} + \mathbf{m}_{\mathrm{res}}. \tag{7}$$

For acoustic features of dimension d and a UBM with M mixture components, these GMM supervectors are of dimension $(Md \times 1)$. As an example, the speaker- and channel-independent supervector $\mathbf{m}_0$ is the concatenation of the UBM mean vectors. We denote the subvectors of $\mathbf{m}_0$ for the gth mixture as $\mathbf{m}_{0[g]}$, which equals μ_g. In the following sections, we discuss how well-known linear Gaussian models, including FA, can be used to develop methods based on this generic decomposition of the GMM supervectors. A summary of the various linear statistical models in speaker recognition is included in Table 1, which highlights both formulation and specifics on matrix/model traits.

Classical MAP Adaptation

We revisit the MAP adaptation technique discussed previously in the GMM–UBM system. If we examine the adaptation equation (4), which is used to update the mean vectors, it is clear that this is a linear combination of two components: one is speaker dependent and the other is independent. In a more generalized way, MAP adaptation can be represented as an operation on the GMM mean supervector as:

$$\mathbf{m}_s = \mathbf{m}_0 + \mathbf{D}\mathbf{z}_s, \tag{8}$$

where D is $(Md \times Md)$ a diagonal matrix and $\mathbf{z}_s$ is a $Md \times 1$ standard normal random vector. We dropped the subscript due to session h for simplicity. According to the linear distortion model of (7), $\mathbf{m}_{\mathrm{spk}} = \mathbf{D}\mathbf{z}_s$. As discussed in [113], in the special case when we set

$$\mathbf{D}^2 = (1/r)\,\Sigma,$$

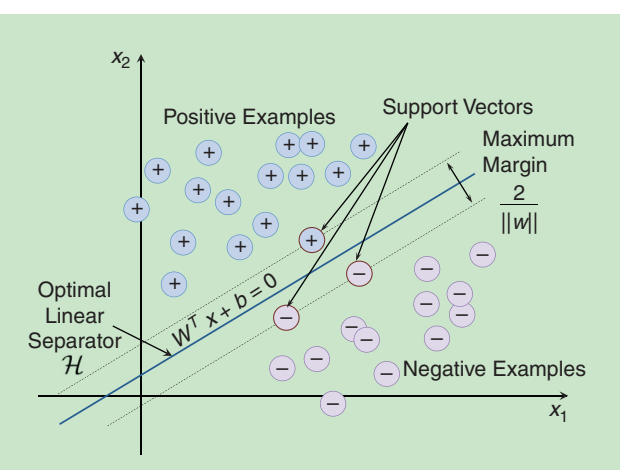
the MAP adaptation equations given in (4) [6] arises from (8), where r is the relevance factor in (6).

Eigenvoice Adaptation

Perhaps the first FA-related model used in speaker recognition was the eigenvoice method [105]. The eigenvoice method was initially proposed for speaker adaptation in speech recognition [114]. In essence, this method restricts the speaker model parameters to lie in a lower dimensional subspace, which is defined by the columns of the eigenvoice matrix. In this model, a speaker-dependent GMM mean supervector $\mathbf{m}_s$ is expressed as

$$\mathbf{m}_s = \mathbf{m}_0 + \mathbf{V}\mathbf{y}_s, \tag{9}$$

where $\mathbf{m}_0$ is the speaker-independent supervector obtained from the UBM, the columns of the matrix V spans the speaker subspace, and $\mathbf{y}_s$ are the standard normal hidden variables known as speaker factors. Here, we dropped the subscript h for simplicity. In accordance with the linear distortion model in (7), the speaker-dependent component is $\mathbf{m}_{\mathrm{spk}} = \mathbf{V}\mathbf{y}_s$. Note that this model does not have a residual noise term as in probabilistic PCA (PPCA) [115] or FA. This means that the eigenvoice model is essentially equivalent to PCA. The model covariance is $\mathbf{V}\mathbf{V}^T$. Since supervectors are usually of a large dimension, a full rank sample covariance matrix, i.e., the supercovariance matrix, is difficult to estimate with limited amount of data. Thus, EM algorithms [116], [117] are used to estimate the eigenvoices. The speaker factors need to be estimated for an enrollment speaker. Computation



[FIG8] A conceptual illustration of an SVM classifier: Positive (+) and negative (–) examples are correspondingly labeled, with the optimal linear separator and support vectors shown.

of the likelihood is carried out as provided in [16, eq. (19)], using the adapted supervector.

This model implies that the adaptation of the GMM supervector parameters is restricted by the eigenvoice matrix. The advantage with this model is that when a small amount of data is available for adaptation, the adapted model is more robust as it is restricted to live in the speaker-dependent subspace, being less affected by nuisance directions. However, the eigenvoice model does not model the channel or intraspeaker variability.

Eigenchannel Adaptation

Similar to adapting the UBM toward a speaker model, a speaker model can also be adapted to a channel model [105]. This can be useful when an unseen channel distortion is observed during testing, and the enrollment speaker model can be adapted to that channel. Similar to the eigenvoice model, the channel variability can also be assumed to lie in a subspace spanned by the principal eigenvectors of the channel covariance matrix. According to our distortion model (7), for a specific channel h, the term $\mathbf{m}_{\text{chn}} = \mathbf{U}\mathbf{x}_h$, where $\mathbf{U}$ is a low-rank matrix that spans the channel subspace, and $\mathbf{x}_h \in \mathrm{N}(0, \mathbf{I})$ are the channel factors. When eigenchannel adaptation is combined with classical MAP, we obtain the model for speaker- and session-dependent GMM supervector

$$\mathbf{m}_{s,h} = \mathbf{m}_0 + \mathbf{D}\mathbf{z}_s + \mathbf{U}\mathbf{x}_h. \tag{10}$$

More details on training the hyperparameters $\mathbf{D}$ and $\mathbf{U}$ can be found in [113]. Likelihood computation can be carried out in a similar way as the eigenvoice method.

Joint FA

The joint FA (JFA) model is formulated by combining both eigenvoice and eigenchannel together, which is accomplished by MAP adaptation for a single model (see "JFA: Summary"). This model assumes that both speaker and channel variability lie in lower dimensional subspaces of the GMM supervector space. These subspaces are spanned by the matrices V and U, as before. The model assumes, for a randomly chosen utterance obtained from speaker s and session h, that its GMM mean supervector can be represented by

$$\mathbf{m}_{s,h} = \mathbf{m}_0 + \mathbf{U}\mathbf{x}_h + \mathbf{V}\mathbf{y}_s + \mathbf{D}\mathbf{z}_{s,h}. \tag{11}$$

JFA: SUMMARY

First proposed	Kenny et al. (2004) [118]
Previous methods	MAP adapted GMM, GMM-SVM approach
Proposed method	Model speaker and channel variability in GMM supervectors
Why robust?	Exploits the behavior of speakers' features in variety of channel conditions learned using FA

Thus, this is the only model so far that considers all four components of the linear distortion model we discussed previously. Indeed, JFA was shown to outperform the other contemporary methods. More details on implementation of JFA can be found in [16] and [118].

The i-Vector Approach

As discussed previously, SVM classifiers on GMM supervectors have been a very successful approach for robust speaker recognition. FA based methods (especially the JFA technique) were also among state-of-the-art systems. In an attempt to combine the strengths of these two approaches, Dehak et al. [79], [119], [120] attempted to use JFA as a feature extractor for SVMs. In their initial attempt [119], the speaker factors estimated using JFA were used as features for SVM classifiers. Observing the fact that the channel factors also contain speaker-dependent information, the speaker and channel factors were combined into a single space termed the *total variability space* [79], [120]. In this FA model, a speaker- and session-dependent GMM supervector is represented by

$$\mathbf{m}_{s,h} = \mathbf{m}_0 + \mathbf{T}\mathbf{w}_{s,h}. \tag{12}$$

The hidden variables $\mathbf{w}_{s,h} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ in this case are called *total factors*. Similar to all of the FA methods above, the hidden variables are not observable but can be estimated by their posterior expectation. The estimates of the total factors, which can be used as features to the next stage of classifiers, came to be known as the *i-vectors*. The term *i-vector* is a short form of "identity vector," regarding the speaker-identification application, and also of "intermediate vectors," referring to its intermediate dimension between those of a supervector and an acoustic feature vector [79] (see "The i-Vector System: Summary").

THE i-VECTOR SYSTEM: SUMMARY

First proposed	Dehak et al. (2009) [79]
Previous methods	JFA and GMM–SVM-based approaches
Proposed method	Reduce supervector dimension using FA before classification
Why robust?	i-vectors effectively summarize utterances and allows using compensation methods that were not practical in large dimensional supervectors

Unlike JFA or other FA methods, the i-vector approach does not make a distinction between speaker and channel. It is simply a dimensionality reduction method of the GMM supervector. In essence, (12) is very similar to a PCA model on the GMM supervectors. The $\mathbf{T}$ matrix is trained using the same algorithms as for the eigenvoice model, except that each utterance is assumed to be obtained from a different speaker.

Mismatch Compensation In i-Vector Domain

The i-vector approach itself does not perform any compensation; on the contrary, it only provides a meaningful lower-dimensional ($400 \cong 800$) representation of a GMM supervector. Thus, it has

most of the advantages of the supervectors, but because of its lower dimension, many conventional compensation strategies can be applied to speaker recognition, which were previously not practical with the large-dimensional supervectors.

LINEAR DISCRIMINANT ANALYSIS

Linear discriminant analysis (LDA) is a commonly employed technique in statistical pattern recognition that aims at finding linear combinations of feature coefficients to facilitate discrimination of multiple classes. It finds orthogonal directions in the feature space that are more effective in discriminating the classes. Projecting the original features in these directions improve classification accuracy. Let D indicate the set of all development utterances, $\mathbf{w}_{s,i}$ indicates an utterance feature (e.g., supervector or i-vector) obtained from the ith utterance of speaker s, n_s denotes the total number of utterances belonging to speaker s, and S is the total number of speakers in D. The between-and within-class covariance matrices are given by

$$\mathbf{S}_b = \frac{1}{S}\sum_{s=1}^{S}(\bar{\mathbf{w}}_s - \bar{\mathbf{w}})(\bar{\mathbf{w}}_s - \bar{\mathbf{w}})^T \quad \text{and} \tag{13}$$

$$\mathbf{S}_w = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{n_s}\sum_{i=1}^{n_s}(\mathbf{w}_{s,i} - \bar{\mathbf{w}}_s)(\mathbf{w}_{s,i} - \bar{\mathbf{w}}_s)^T, \tag{14}$$

where the speaker-dependent and speaker-independent mean vectors are given by

$$\bar{\mathbf{w}}_s = \frac{1}{n_s}\sum_{i=1}^{n_s}\mathbf{w}_{s,i} \quad \text{and}$$

$$\bar{\mathbf{w}} = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{n_s}\sum_{i=1}^{n_s}\mathbf{w}_{s,i},$$

respectively. The LDA optimization thus aims at maximizing the between class variance while minimizing the within-class variance (due to channel variability). The projections obtained from this optimization are found by the solution of the following generalized eigenvalue problem:

$$\mathbf{S}_b\mathbf{v} = \Lambda\mathbf{S}_w\mathbf{v}. \tag{15}$$

Here, Λ is the diagonal matrix containing the eigenvalues. If the matrix $\mathbf{S}_w$ is invertible, this solution can be found by finding the eigenvalues of the matrix $\mathbf{S}_w^{-1}\mathbf{S}_b$. Generally, the first $k < R$ eigenvalues are used to prepare a matrix $\mathbf{A}_{\text{LDA}}$ of dimension $R \times k$ given by

$$\mathbf{A}_{\text{LDA}} = [\mathbf{v}_1 \ldots \mathbf{v}_k],$$

where $\mathbf{v}_1 \ldots \mathbf{v}_k$ denote the first k eigenvectors obtained by solving (15). The LDA transformation of the utterance feature $\mathbf{w}$ is thus obtained by

$$\Phi_{\text{LDA}}(\mathbf{w}) = \mathbf{A}_{\text{LDA}}^{\text{T}}\,\mathbf{w}.$$

NAP

The NAP algorithm was originally proposed in [108]. In this approach, the feature space is transformed using an orthogonal projection in the channel's complementary space, which depends only on the speaker (assuming that other variability in the data is insignificant). The projection is calculated using the within-class covariance matrix. Define a $d \times d$ projection matrix [108] of co-rank $k < d$

$$\mathbf{P} = \mathbf{I} - \mathbf{u}_{[k]}\mathbf{u}_{[k]}^T,$$

where $\mathbf{u}_{[k]}$ is a rectangular matrix of low rank whose columns are the k principal eigenvectors of the within-class covariance matrix $\mathbf{S}_w$ given in (14). NAP is performed on $\mathbf{w}$ as

$$\Phi_{\text{NAP}}(\mathbf{w}) = \mathbf{P}\mathbf{w}.$$

WCCN

This normalization was originally proposed for improving robustness in the SVM-based speaker-recognition framework [109] using a one-versus-all decision approach. The WCCN projection aims at minimizing the false-alarm and miss-error rates during SVM training.

The implementation of the strategy begins with using a data set D similar to the one that was described in the previous section. The within-class covariance matrix $\mathbf{S}_w$ is calculated using (14), and the WCCN projection is performed as

$$\Phi_{\text{WCCN}}(\mathbf{w}) = A_{\text{WCCN}}^T\,\mathbf{w},$$

where $\mathbf{A}_{\text{WCCN}}$ is computed through the Cholesky factorization of $\mathbf{S}_w^{-1}$ such that

$$\mathbf{S}_w^{-1} = \mathbf{A}_{\text{WCCN}}\,\mathbf{A}_{\text{WCCN}}^{\text{T}}.$$

In contrast to LDA and NAP, the WCCN projection conserves the directions of the feature space.

SPEAKER VERIFICATION USING i-VECTORS

After i-vectors were introduced, in essence, many previously available pattern-recognition methods were effectively applied in this domain. We discuss some of the popular methods of classification using i-vectors.

SVM Classifier

As discussed previously, the i-vector representation was discovered in an attempt to utilize JFA as a feature extractor for SVMs. Thus, initially i-vectors were used with SVMs with different kernel functions [79]. The idea is the same as SVM with GMM supervectors, except that the i-vectors are now used as utterance-dependent features. Because of the lower dimension of the i-vectors compared to supervectors, the application of LDA and WCCN projections together became more effective and were well suited.

Cosine Distance Scoring

In [79], the cosine similarity measure-based scoring was proposed for speaker verification. In this measure, the match score between a target and test i-vector $\mathbf{w}_{\text{target}}$ and $\mathbf{w}_{\text{test}}$ is computed as their normalized dot product

$$\text{CDS}(\mathbf{w}_{\text{target}}, \mathbf{w}_{\text{test}}) = \frac{\mathbf{w}_{\text{target}} \cdot \mathbf{w}_{\text{test}}}{\|\mathbf{w}_{\text{target}}\|\,\|\mathbf{w}_{\text{test}}\|}.$$

Probabilistic Linear Discriminant Analysis

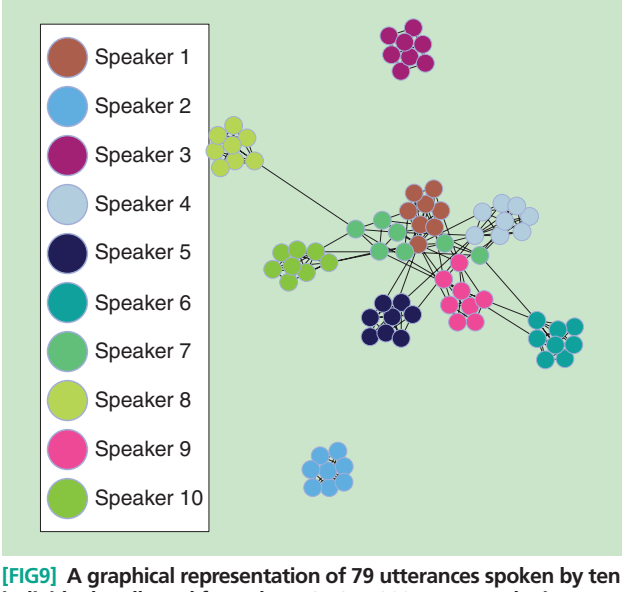
Probalistic LDA (PLDA) was first used for session variability compensation for facial recognition [121]. This essentially follows the same modeling assumptions as JFA, i.e., a pattern vector contains class-dependent and session-dependent variabilities, both lying in lower-dimensional subspaces. An i-vector extracted from utterance u is decomposed as

$$\mathbf{w}_{s,h} = \mathbf{w}_0 + \Phi\beta_s + \Gamma\alpha_h + \varepsilon_{s,h}. \tag{16}$$

Here, $\mathbf{w}_0 \in \mathbf{R}^R$ is the speaker-independent mean i-vector, Φ is the $R \times N_{ev}$ low-rank matrix representing the speaker-dependent basis functions/eigenvoices, Γ is the $R \times N_{ec}$ low-rank matrix spanning the channel subspace, $\boldsymbol{\beta}_s \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is an $N_{ev} \times 1$ hidden variable (i.e., speaker factors), $\boldsymbol{\alpha}_s \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is an $N_{ec} \times 1$ hidden variable (i.e., channel factors), and $\boldsymbol{\varepsilon}_{h,s} \in \mathrm{R}^R$ is a random vector representing the residual noise.

PLDA was first introduced in speaker verification in [94] using a heavy-tailed distribution assumption on i-vectors instead of a Gaussian assumption. Later, it was shown that when i-vectors are length normalized (i.e., they are divided by their corresponding vector length) [122], a Gaussian PLDA model performs equivalent to its heavy-tailed version. Since the latter is computationally more expensive, Gaussian PLDA models are more commonly used. Also, the use of a full-covariance noise model for $\boldsymbol{\varepsilon}_{h,s}$ is feasible in this formulation that allows one to drop the eigenchannel term ($\Gamma\boldsymbol{\alpha}_h$) from (16) without loss of performance. In this case, the PLDA model would be as follows:

$$\mathbf{w}_{s,h} = \mathbf{w}_0 + \Phi\boldsymbol{\beta}_s\boldsymbol{\varepsilon}_{s,h}.$$



[FIG9] A graphical representation of 79 utterances spoken by ten individuals collected from the NIST SRE 2004 corpus. The i-vector representation is used for each segment; the plot is generated using GUESS, an open-source graph exploration software [123] that can visualize higher-dimensional data using distance measures between samples.

We note that, though developed independently, the JFA model is very similar to PLDA. Looking at (11) and (16) and comparing the terms makes this clear. The obvious difference between these models is that JFA models the GMM supervectors, while PLDA models i-vectors. Since i-vectors are essentially dimensionality reduced versions of supervectors (incurring loss of information), JFA, in principle, should be better in modeling the within- and between-speaker variations. However, in reality, the amount of labeled training data is limited, and due to the large number of parameters in JFA, it cannot be trained as effectively as a PLDA model on lower dimensional i-vectors (using the same amount of labeled data). Besides, the total variability model (i-vector extractor) can be trained on unlabeled data sets, which are available in larger amounts.

Although the model equations are identical, there are significant differences in the training process of the two models. Since JFA was designed for GMM supervectors, the formulations involved processing the acoustic speech frames and their statistics in different mixtures of the UBM. Unlike i-vectors, the GMM supervectors are not extracted first before JFA training—instead, JFA operates directly on the acoustic features and can provide similarity scores between two utterances from their corresponding feature streams. This dependence on acoustic features (and the various order statistics) makes the scoring process more computationally expensive for JFA. For PLDA, the input features are i-vectors that are extracted beforehand, and, during the scoring process, only two i-vectors from the corresponding utterances are required—not the acoustic features. This makes PLDA much simpler in implementation.

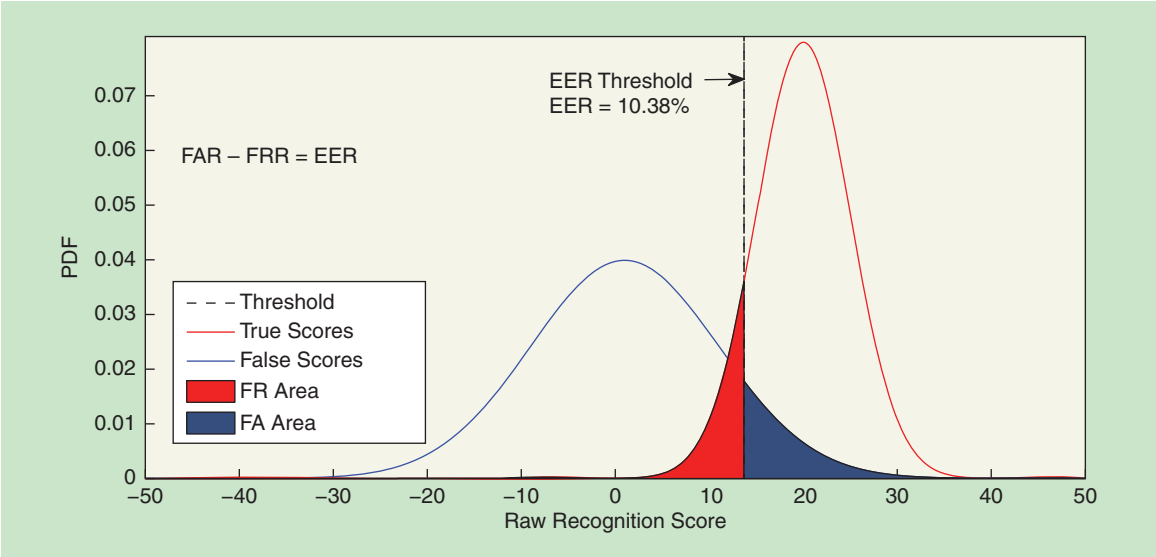
It can be argued that, with a sufficiently large labeled data set, JFA can outperform an i-vector-PLDA system. However, we are not aware of such results reported at this time.

PERFORMANCE EVALUATION IN STANDARDIZED DATA SETS

Evaluating the performance of a speaker-verification task using a standardized data set is a very important element of the research cycle. Over the years, new data sets and performance metrics have been introduced to match realistic scenarios. These, in turn, motivated researchers to discover new strategies to address the challenges, compare results among peers, and exchange ideas.

THE NIST SRE CHALLENGE

NIST has been organizing an SRE campaign for the past several years aiming at providing standard data sets, verification tasks, and performance metrics for the speaker ID community (Figure 9). Every year's evaluation introduces new challenges for the research community. These challenges include newly introduced recording conditions (e.g., microphone, handset, and room acoustics), short test utterance duration, varying vocal effort, artificial and real-life additive noise, restrictions or allowances in data-utilization strategy, new performance metrics to be optimized, etc. It is clear that the performance metric defined for a speaker-recognition task depend on the data set and train-test pairs of speech (also known as *trials*) used for the evaluation. A sufficient number of such trials needs to be provided for a statistically significant evaluation measure [78]. The performance measures can be based on hard verification decisions or soft scores, they may require log-LR as scores, and depend on the



[FIG10] An illustration of target and nontarget score distributions and the decision threshold. Areas under the curves with blue and red colors represent FAR and FRR errors, respectively.

prior probability of encountering a target speaker. For a given data set and task, systems evaluated using a specific error/cost criteria can be compared. Before discussing the common performance measures, we introduce the type of errors encountered in speaker verification.

TYPES OF ERRORS

There are mainly two types of errors in speaker verification (or any other biometric authentication) when a hard decision is made by the automatic system. From the speaker authentication point of view, we define them as

- *false accept (FA)*: granting access to an impostor speaker
- *false reject (FR)*: denying access to a legitimate speaker.

From the speaker-detection point of view (a target speaker is sought), these are called *false-alarm* and *miss errors*, respectively. According to these definitions, two error rates are defined as

$$\text{False-Acceptance Rate (FAR)} = \frac{\text{Number of FA errors}}{\text{Number of impostor attempts}}$$

$$\text{False-Rejection Rate (FRR)} = \frac{\text{Number of FR errors}}{\text{Number of legitimate attempts}}.$$

Speaker-verification systems generally output a match score between the training speaker and the test utterance. This is true for most two-class recognition/binary detection problem. This score is a scalar variable that represents the similarity between the enrolled speaker and the test speaker, with higher values indicating the speakers are more similar. To make a decision, the system needs to use a threshold (τ) as illustrated in Figure 10. If the threshold is too low, there will be a lot of FA errors, whereas if the threshold is too high, there will be too many FR/miss errors.

EQUAL ERROR RATE

The equal error rate (EER) is defined as the FAR and FRR values when they become equal. That is, by changing the threshold, we find a point where the FAR and FRR become equal. This is shown in Figure 10. The EER is a very popular performance measure for speaker-verification systems. Only the soft scores from the automatic system are required to compute the EER. No actual hard decisions are made. It should be noted that operating a speaker-verification system on the threshold corresponding to the EER might not be desirable for practical purposes. For high-security applications, one should set the threshold higher, lowering the FA errors at the cost of miss errors. However, for high convenience, the threshold may be set lower. Let us discuss some examples. In authenticating users for bank accounts, security is of utmost importance. It is thus better to deny access to the legitimate user (and ask other forms of verification) as opposed to granting access to an impostor. On the contrary, for an automated customer service, denying a legitimate speaker will cause inconvenience and frustration to the user. In this case, accepting an illegitimate speaker is not as critical as in high-security applications.

DETECTION COST FUNCTION

This is, in fact, a family of performance measures introduced by NIST over the years. As mentioned before, the EER does not differentiate between the two errors, which sometimes is not a realistic performance measure. The detection cost function (DCF), thus, introduces numerical costs/penalties for the two types of errors (FA and miss). The a priori probability of encountering a target speaker is also provided. The DCF is computed over the full range of decision threshold values as

$$\text{DCF}(\tau) = C_{\text{MISS}} P(\tau) P_{\text{Target}} + C_{\text{FA}} P_{\text{FA}}(\tau)(1 - P_{\text{Target}}).$$

Here,

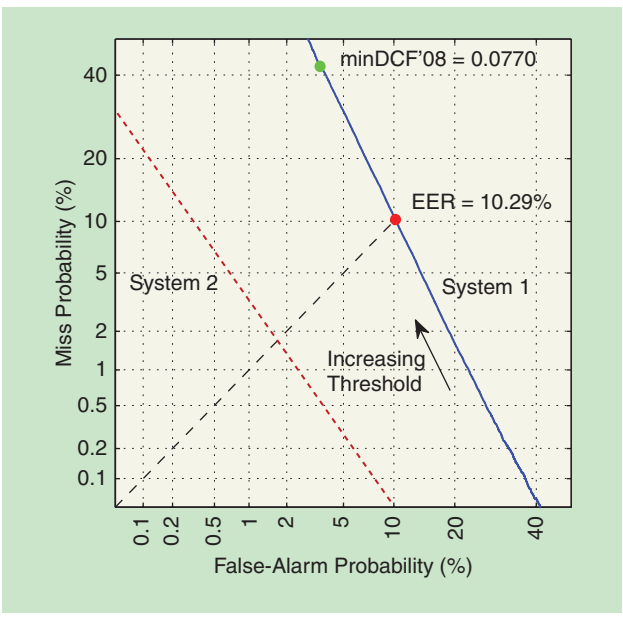
C_{Miss} = Cost of a miss/FR error
C_{FA} = Cost of an FA error
P_{Target} = Prior probability of target speaker.
$P_{\text{Miss}}(\tau)$ = Probability of (Miss | Target, Threshold = τ)
$P_{\text{FA}}(\tau)$ = Probability of (FA | Nontarget, Threshold = τ).

Usually, the DCF is normalized by dividing it by a constant [77]. The probability values here can be computed using the distribution of true and impostor scores and computing the areas under the curve as shown in Figure 10. The first three quantities above (C_{Miss}, C_{FA}, and P_{Target}) are predefined. Generally, the goal of the system designer is to find the optimal threshold value that minimizes the DCF.

In NIST SRE 2008, these DCF parameters were set as $C_{Miss} = 10$, $C_{FA} = 1$, and $P_{Target} = 0.01$. The values of the costs indicate that the system is penalized ten times more for making a miss error rather than an FA error. As a real-world example, when detecting a known criminal's voice from evidence recordings, it may be better to have false positives (e.g., to suspect and investigate an innocent speaker) than to miss the target speaker (e.g., to be unable to detect the criminal at all). If we ignore P_{Target} for the moment, setting a lower threshold (τ) would be beneficial since, in this case, the system will tolerate more FAs but will not miss too many legitimate speakers [$P_{Miss}(\tau)$ will be lower], yielding a lower DCF value for that threshold. Now, the value of the prior ($P_{Target} = 0.01$) indicates that a target speaker will be encountered by the system once in every 100 speaker-verification attempts. If this condition is considered independently, it is better to have a higher threshold since most of the attempts will be from impostors ($P_{Nontarget} = 0.99$). However, when all three parameters are considered together, finding the optimal threshold requires sweeping through all the DCF values.

By processing the DCF, two performance measures are derived: 1) the minimum DCF (MinDCF) and 2) the actual DCF (ActDCF). The MinDCF is the minimum value of DCF that can be obtained by changing the threshold, τ. The MinDCF parameter can be computed only when the soft scores are provided by the systems. When the system provides hard decisions, the actual DCF is used where the probability values involved (in the DCF equation) are simply computed by counting the errors. Both of these performance measures have been extensively used in the NIST evaluations. The most recent evaluation in 2012 introduced a DCF that is a dependent on two different operating points (two sets of error costs and target priors) instead of one.

It is important to note here that the MinDCF (or ActDCF) parameter is not an error rate in the general sense. Thus, its interpretation is not straightforward. Obviously, the lower MinDCF, the better the system performance. However, the exact value of the MinDCF can only be used to compare other systems evaluated using the same trials and performance measure. Generally, when the system EER improves, the DCF parameters also improve. An elaborate discussion on the relationship between EER and DCF can be found in [124].



[FIG11] DET curves of two speaker-verification systems (System 1 and System 2). In System 1, the points on the curve corresponding to the threshold that yields the EER and minimum DCF (as in NIST SRE 2008), and the direction of an increasing threshold are shown. Being closer to the origin, System 2 shows a better performance.

DETECTION ERROR TRADEOFF CURVE

When speaker-verification performance needs to be evaluated in a range of operating points, the detection error tradeoff (DET) curve is generally employed. The DET curve is a plot of the errors FAR versus FRR/miss. An example DET curve is shown in Figure 11. As the system performance improves, the curve moves toward the origin. As illustrated in Figure 11, the DET curve corresponding to System 2 is closer to the origin and thus represents a better system. The EER and minDCT points are shown on the DET curve of System 1.

During the preparation of the DET curve, the cumulative density functions (CDFs) of the true and impostor scores are transformed to normal deviates. This means that the true/impostor score CDF value for a given threshold is transformed by a standard normal inverse CDF (ICDF) and the resulting values are used to make the plot. This transform yields a linear DET curve when the two distributions are normal and have equal variances. Thus, even though the labels indicate the axis as error probabilities, they are actually plotted according to the corresponding normal deviate values.

RECENT ADVANCEMENTS IN AUTOMATIC SPEAKER RECOGNITION

In recent years, considerable research progress has been made in spoofing and countermeasures [125], [126], back-end classifiers [127], [128], compensation for short utterances [129]–[131], score calibration and fusion [132], [133], deep neural network (DNN) [134]–[136], and alternate acoustic modeling [137] techniques. In this section, we briefly discuss some of these topics and their possible implications in the speaker-recognition research.

NIST i-VECTOR MACHINE-LEARNING CHALLENGE AND BACK-END PROCESSING

The most recent NIST-sponsored evaluation, the i-Vector Machine-Learning Challenge, focused on back-end classifiers. In this paradigm, instead of audio data, i-vectors from speech utterances were provided to the participants [138]. In this way, the entry barrier to the evaluation was reduced as many machine-learning-focused research groups were able to participate without expertise in audio/speech processing. Significant performance

improvements were observed from top-performing systems compared to the baseline system provided by NIST [138]. Since only i-vectors were provided by NIST, the algorithmic improvements are all due to modeling and back-end processing of i-vectors. In addition, the i-vectors provided by NIST did not have any speaker labels, which also generated new ideas on utilizing unlabeled data in speaker recognition [139].

DURATION VARIABILITY COMPENSATION

Duration variability is one of the problems that has received considerable attention in recent years. Since the advent of GMM supervectors and i-vectors, variable-duration utterances could be mapped to a fixed-dimensional pattern. This has been a significant advancement since various machine-learning tools were being applied to these vectors, especially i-vectors due to their smaller dimensions. However, it is clear that an i-vector extracted from a short utterance will not be as representative of a speaker compared to the one extracted from a longer utterance. Duration mismatch between train and test is thus a major problem. One way to mitigate this problem is by including short utterances in the PLDA training [130], [140]. Alternatively, this can be addressed in the score domain [130]. In [141], Kenny et al. propose that i-vectors extracted from short utterances are less reliable and incorporates this variability by including a noise term into the PLDA model. In [142], a DNN-based method was proposed for speaker recognition in short utterances where the content of the test utterance was searched in the enrollment data to be compared.

DNN-BASED METHODS

In the last few years, DNNs have been tremendously successful at many speech-processing tasks, most prominently in speech recognition [143], [144]. Naturally, DNNs have also been used in speaker recognition. Works by Kenny et al. [136] have shown improvements in extracting Baum–Welch statistics for speaker recognition using DNNs. DNNs have also been incorporated for multisession speaker recognition [134] as well as phonetically aware DNNs for noise-robust speaker recognition [135]. DNNs have also been used to extract front-end features, also known as *bottle-neck features* [145]. Since, there are an extensive set of literature on deep learning [143], [146] and its application in speaker recognition is relatively new, we have not included a discussion on DNNs in this tutorial.

[TABLE 2] THE SPEAKER-RECOGNITION PROCESS: MAN VERSUS MACHINE.

ASPECT	HUMANS	MACHINES
TRAINING	SPEAKER RECOGNITION IS AN ACQUIRED HUMAN TRAIT AND REQUIRES TRAINING.	REQUIRES SUFFICIENT DATA TO TRAIN THE RECOGNIZERS.
VAD	DIFFERENT PARTS OF THE HUMAN BRAIN ARE ACTIVATED WHEN SPEECH AND NONSPEECH STIMULI ARE PRESENTED.	SPEECH SIGNAL PROPERTIES AND STATISTICAL MODELS ARE USED TO DETECT PRESENCE OR ABSENCE OF SPEECH.
AUDIO PROCESSING	THE HUMAN BRAIN PERFORMS BOTH SPECTRAL AND TEMPORAL PROCESSING. IT IS NOT KNOWN EXACTLY HOW THE AUDIO SIGNAL DEVELOPS THE SPEAKER- OR PHONEME-DEPENDENT ABSTRACT REPRESENTATIONS/MODELS.	ACOUSTIC FEATURE PARAMETERS DEPENDING ON SPECTRAL AND TEMPORAL PROPERTIES OF THE AUDIO SIGNAL ARE UTILIZED FOR RECOGNITION.
HIGH-LEVEL FEATURES	WE CONSIDER LEXICON, INTONATION, PROSODY, AGE, GENDER, DIALECT, SPEAKING RATE, AND MANY OTHER PARALINGUISTIC ASPECTS OF SPEECH TO REMEMBER A PERSON'S VOICE.	RECENT ALGORITHMS HAVE INCORPORATED PROSODY, PRONUNCIATION, DIALECT, AND OTHER HIGH-LEVEL FEATURES FOR SPEAKER IDENTIFICATION.
COMPACT REPRESENTATION	THE HUMAN BRAIN FORMS SPEAKER-DEPENDENT, EFFICIENT ABSTRACT REPRESENTATIONS. THESE ARE INVARIANT TO CHANGES OF THE ACOUSTIC INPUT, PROVIDING ROBUSTNESS TO NOISE AND SIGNAL DISTORTION.	HIGH-LEVEL FEATURES ARE EXTRACTED THAT SUMMARIZE THE VOICE CHARACTERISTICS OF A SUBJECT. THESE ARE EXTRACTED IN A WAY TO MINIMIZE SESSION VARIABILITY DUE TO NOISE OR DISTORTION.
LANGUAGE DEPENDENCE	SPEAKER RECOGNITION BY HUMANS IS BETTER IF THEY KNOW THE LANGUAGE BEING SPOKEN.	AUTOMATIC SYSTEM'S PERFORMANCE IS DEGRADED IF THERE IS A MISMATCH IN TRAINING AND TEST LANGUAGE.
FAMILIAR VERSUS UNFAMILIAR SPEAKERS	HUMANS ARE EXTREMELY GOOD AT IDENTIFYING FAMILIAR VOICES, BUT NOT SO FOR UNFAMILIAR ONES.	MACHINES PROVIDE CONSISTENT PERFORMANCE WHEN ADEQUATE AMOUNT OF DATA IS PROVIDED. FAMILIARITY CAN BE RELATED TO THE AMOUNT OF TRAINING DATA.
IDENTIFICATION VERSUS DISCRIMINATION	THE HUMAN BRAIN PROCESSES THESE TWO TASKS DIFFERENTLY.	IN MOST CASES, THE SAME ALGORITHM (WITH SLIGHT MODIFICATION) CAN BE USED TO IDENTIFY AND DISCRIMINATE BETWEEN SPEAKERS.
MEMORY RETENTION	HUMANS' ABILITY TO REMEMBER A PERSON'S VOICE DEGRADES WITH TIME.	A COMPUTER ALGORITHM CAN STORE THE MODELS OF A PERSON INDEFINITELY IF PROVIDED SUPPORT.
FATIGUE	HUMAN LISTENERS CANNOT PERFORM AT THE SAME LEVEL FOR A LONG DURATION.	COMPUTERS DO NOT HAVE ISSUES WITH FATIGUE. LONG RUNTIMES MAY CAUSE OVERHEATING IF NECESSARY PRECAUTIONS ARE NOT TAKEN.
IDENTIFY IDIOSYNCRASIES	HUMANS ARE VERY GOOD AT IDENTIFYING CHARACTERISTIC TRAITS OF A VOICE.	THE MACHINE ALGORITHMS HAVE TO BE SPECIFICALLY TOLD WHAT TO LOOK FOR AND COMPARE.
MISMATCHED CONDITIONS	HUMANS RELY MORE ON PARALINGUISTIC ASPECTS OF SPEECH IN SEVERE MISMATCHED CONDITIONS.	AUTOMATIC SYSTEMS ARE TRAINED ON VARIOUS ACOUSTIC CONDITIONS, AND USUALLY ARE MORE ROBUST.
SUSCEPTIBILITY TO BIAS	HUMAN JUDGMENT CAN BE BIASED BY CONTEXTUAL INFORMATION.	AUTOMATIC ALGORITHMS CAN BE BIASED TOWARD THE TRAINING DATA.

MAN VERSUS MACHINE IN SPEAKER RECOGNITION

In this section, we attempt to compare the speaker-recognition task as performed by humans and the state-of-the-art algorithms. First we must realize that it is very difficult to do such comparisons in a statistically meaningful manner. This is because getting humans to evaluate a large number of utterances reliably is quite challenging. However, attempts have been made to make such comparisons in the past [147]–[149]. In the majority of these cases, especially in the recent ones, the speaker-recognition performance of humans was found to be inferior to that of automatic systems.

In [150], the authors compared the speaker-recognition performance of human listeners to a typical algorithm (automatic system is not mentioned in the paper) using a subset of NIST SRE 1998 data. Opinions of multiple human listeners were combined to form the final speaker-recognition decision. Results showed that humans are as good as the best system and outperformed standard algorithms especially when there is a mismatch in the telephone channel (a different number was used to make the phone call).

Recently, NIST presented speaker-recognition tasks for evaluating systems that combined human and machines [20]. The task, known as the HASR, was designed in a way such that the most difficult test samples are selected for the evaluation (channel mismatch, noise, same/different speakers that sound highly dissimilar/similar, etc.). However, the total number of trials in these experiments was very low compared to evaluations designed for automatic systems. One of the motivations of this study was to evaluate if automatic systems have become good enough, in other words, is it beneficial to keep humans involved in the process? The HASR study was repeated during the 2012 NIST SRE where both noisy and channel degraded speech data were encountered.

Interestingly, machines consistently performed better than human-assisted approaches in the given NIST HASR tasks [151]–[155]. In [156], it was even claimed that by combining multiple naïve listeners' decisions, the HASR 2010 task can be performed as well as forensic experts, which somewhat undermines the role of a forensic expert. In [157], it was shown that human and machine decisions were complementary, meaning that in some cases the humans correctly identified a speaker where the automatic system failed, and vice versa. However, the HASR tasks were exceptionally difficult for human listeners because of the severe channel mismatch, unfamiliarity with the speakers, noise, and other factors. A larger and more balanced set of trials should be used for a proper evaluation of human performance. Following the HASR paradigm, further research focused on how humans can aid the decision of an automatic system, especially in the context of forensic speaker recognition [157]. An i-vector system [79] with a PLDA classifier was used in this particular study.

The performance of humans and machines was compared in a forensic context in [149], where 45 trials were used (nine target and 36 nontarget). The human system consisted of a panel of listeners whereas a GMM–UBM-based system [6] was used for the automatic system. Here again, the automatic system outperformed the human panel of listeners. However, the results should be interpreted with caution since the number of trials here was low.

In [158], human speaker-recognition performance was compared with automatic algorithms in presence of voice mimicry. A GMM–UBM system and an i-vector-based system were used in the study. The speech database consisted of five Finnish public figures and their voices were mimicked by a professional voice actor. The results show that humans are more likely to make errors when impersonation is done. On average, the automatic algorithm performed better than the human listeners.

Although most experiments so far show human performance to be inferior to automatic algorithms, these cannot be considered as definitive proof that machines are always better than humans. In many circumstances, humans will perform better, especially when paralinguistic information becomes important. As discussed previously, humans perform exceptionally well in recognizing familiar speakers. To the best of our knowledge, a comparison of familiar speaker recognition versus automatic algorithm (with sufficient training data) has not been performed yet. Thus, for familiar speakers, humans may perform much better than state-of-the-art algorithms—and this should motivate researchers to discover how the human brain stores familiar speakers' identity information. In HASR, the goal was to have humans assist the automatic system. On the other hand, automatic systems inspired by the forensic experts' methodology have already been investigated [159], where speaker nativeness, dialect, and other demographic information were considered. A generic comparison between how humans and machines perform speaker recognition is provided in Table 2.

CONCLUSIONS

A substantial amount of work still needs to be done to fully understand how the human brain makes decisions about speech content and speaker identity. However, from what we know, it can be said that automatic speaker-recognition systems should focus more on high-level features for improved performance. Humans are effective at effortlessly identifying unique traits of speakers they know very well, whereas automatic systems can only learn a specific trait if a measurable feature parameter can be properly defined. Automatic systems are better at searching over vast collections of audio and, perhaps, at being able to more effectively set aside those audio samples which are less likely to be speaker matches; whereas humans are better at comparing a smaller subset and overcoming microphone or channel mismatch more easily. It may be worthwhile to investigate what it really means to "know" a speaker from the perspective of an automatic system. The search for alternative compact representations of speakers and audio segments emphasizing the identity relevant parameters while suppressing the nuisance components will always be an ongoing challenge for system developers.

AUTHORS

John H.L. Hansen (John.Hansen@utdallas.edu) received his Ph.D. and M.S. degrees in electrical engineering from the Georgia Institute of Technology and his B.S.E.E. degree from Rutgers University, New Jersey. He is with the University of Texas at Dallas, where he is associate dean for research and professor of electrical engineering. He holds the Distinguished Chair in Telecommunications and oversees

the Center for Robust Speech Systems. He is an International Science Congress Association (ISCA) fellow; a past chair of the IEEE Speech and Language Technical Committee; a coorganizer and technical chair of the IEEE International Conference on Acoustics, Speech, and Signal Processing 2010; a coorganizer of IEEE Spoken Language Technology 2014; and the general chair/organizer of ISCA Interspeech 2002. He has supervised more than 70 Ph.D./M.S. students and coauthored more than 570 papers in the field. His research interests include digital speech processing, speech and speaker analysis, robust speech/speaker/language recognition, speech enhancement for hearing loss, and robust hands-free human-interaction in car environments. He is a Fellow of the IEEE.

Taufiq Hasan (taufiq.hasan@utdallas.edu) received his B.S. and M.S. degrees in electrical and electronic engineering (EEE) from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, in 2006 and 2008, respectively. He earned his Ph.D. degree from the Department of Electrical Engineering, Erik Jonsson School of Engineering and Computer Science, University of Texas at Dallas (UT Dallas), Richardson, in 2013. From 2006 to 2008, he was a lecturer in the Department of EEE, United International University, Dhaka. He served as the lead student from the Center for Robust Speech Systems at UT Dallas during the 2012 National Institute of Standards and Technology Speaker Recognition Evaluation submissions. Currently, he works as a research scientist at Robert Bosch Research and Technology Center in Palo Alto, California. His research interests include speaker recognition in mismatched conditions, speech recognition, enhancement and summarization, affective computing, and multimodal signal processing.

REFERENCES

[1] D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, et al., "Building Watson: An overview of the DeepQA project," *AI Mag.*, vol. 31, no. 3, pp. 59–79, 2010.

[2] J. Aron, "How innovative is Apple's new voice assistant, Siri?" *New Sci.*, vol. 212, no. 2836, pp. 24, 29 Oct. 2011.

[3] A. Eriksson, "Tutorial on forensic speech science," in *Proc. European Conf. Speech Communication and Technology*, Lisbon, Portugal, 2005, pp. 4–8.

[4] P. Belin, R. J. Zatorre, P. Lafaille, P. Ahad, and B. Pike, "Voice-selective areas in human auditory cortex," *Nature*, vol. 403, pp. 309–312, Jan. 2000.

[5] E. Formisano, F. De Martino, M. Bonte, and R. Goebel, "'Who' is saying 'what'? Brain-based decoding of human voice and speech," *Science*, vol. 322, pp. 970–973, Nov. 2008.

[6] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Process.*, vol. 10, no. 1–3, pp. 19–41, Jan. 2000.

[7] [Online]. Available: www.biometrics.gov

[8] P. Eckert and J. R. Rickford, *Style and Sociolinguistic Variation*. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[9] J. H. L. Hansen, "Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition," *Speech Commun.*, vol. 20, no. 1–2, pp. 151–173, Nov. 1996.

[10] X. Fan and J. H. L. Hansen, "Speaker identification within whispered speech audio streams," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 4, pp. 1408–1421, May 2011.

[11] C. Zhang and J. H. L. Hansen, "Whisper-island detection based on unsupervised segmentation with entropy-based speech feature processing," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 4, pp. 883–894, May 2011.

[12] J. H. L. Hansen and V. Varadarajan, "Analysis and compensation of Lombard speech across noise type and levels with application to in-set/out-of-set speaker recognition," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 2, pp. 366–378, Feb. 2009.

[13] M. Mehrabani and J. H. L. Hansen, "Singing speaker clustering based on subspace learning in the GMM mean supervector space," *Speech Commun.*, vol. 55, no. 5, pp. 653–666, June 2013.

[14] J. H. Hansen, C. Swail, A. J. South, R. K. Moore, H. Steeneken, E. J. Cupples, T. Anderson, C. R. Vloeberghs, et el., "The impact of speech under 'stress' on military speech technology," NATO Project Rep., no. 104, 2000.

[15] D. A. Reynolds, M. A. Zissman, T. F. Quatieri, G. C. O'Leary, and B. A. Carlson, "The effects of telephone transmission degradations on speaker recognition performance," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'95)*, pp. 329–332.

[16] P. Kenny, G. Boulianne, P. Ouellet, and P. Dumouchel, "Joint factor analysis versus eigenchannels in speaker recognition," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1435–1447, 2007.

[17] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas, "Score normalization for text-independent speaker verification systems," *Digital Signal Process.*, vol. 10, no. 1–3, pp. 42–54, Jan. 2000.

[18] R. C. Rose, E. M. Hofstetter, and D. A. Reynolds, "Integrated models of signal and background with application to speaker identification in noise," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 2, pp. 245–257, 1994.

[19] Q. Jin, T. Schultz, and A. Waibel, "Far-field speaker recognition," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 7, pp. 2023–2032, 2007.

[20] C. Greenberg, A. Martin, L. Brandschain, J. Campbell, C. Cieri, G. Doddington, and J. Godfrey, "Human assisted speaker recognition in NIST SRE10," in *Proc. Odyssey: The Speaker and Language Recognition Workshop*, Brno, Czech Republic, 2010, pp. 180–185.

[21] A. D. Lawson, A. Stauffer, E. J. Cupples, S. J. Wenndt, W. Bray, and J. J. Grieco, "The multi-session audio research project (MARP) corpus: Goals, design and initial findings," in *Proc. Interspeech*, Brighton, U.K., pp. 1811–1814, 2009.

[22] K. W. Godin and J. H. Hansen, "Session variability contrasts in the MARP corpus." in *Proc. Interspeech*, 2010, pp. 298–301.

[23] L. A. Ramig and R. L. Ringel, "Effects of physiological aging on selected acoustic characteristics of voice," *J. Speech Lang. Hearing Res.*, vol. 26, pp. 22–30, Mar. 1983.

[24] F. Kelly, A. Drygajlo, and N. Harte, "Speaker verification with long-term ageing data," in *Proc. Int. Conf. Biometrics,* 2012, pp. 478–483.

[25] F. Kelly, A. Drygajlo, and N. Harte, "Speaker verification in score-ageing-quality classification space," *Comput. Speech Lang.*, vol. 27, no. 5, pp. 1068–1084, Aug. 2013.

[26] Wikipedia Contributors. George Zimmerman. Entry on Wikipedia (2014, Nov. 29). [Online]. Available: http://en.wikipedia.org/wiki/George_Zimmerman

[27] J. H. L. Hansen and N. Shokouhi. (2013, Nov. 29). Speaker identification: Screaming, stress and non-neutral speech, is there speaker content?. [Online]. *IEEE SLTC Newsletter.* Available: http://www.signalprocessingsociety.org/technical-committees/list/sl-tc/spl-nl/2013-11/SpeakerIdentification/

[28] F. Nolan and T. Oh, "Identical twins, different voices," *Int. J. Speech Lang. Law*, vol. 3, no. 1, pp. 39–49, 1996.

[29] W. D. Van Gysel, J. Vercammen, and F. Debruyne, "Voice similarity in identical twins," *Acta Otorhinolaryngol. Belg.*, vol. 55, no. 1, pp. 49–55, 2001.

[30] K. M. Van Lierde, B. Vinck, S. De Ley, G. Clement, and P. Van Cauwenberge, "Genetics of vocal quality characteristics in monozygotic twins: a multiparameter approach," *J. Voice*, vol. 19, no. 4, pp. 511–518, Dec. 2005.

[31] D. Loakes, "A forensic phonetic investigation into the speech patterns of identical and non-identical twins," *Int. J. Speech Lang. Law*, vol. 15, no. 1, pp. 97–100, 2008.

[32] *Twins Day. Twinsburg, Ohio* (2015). [Online]. Available: http://www.twinsdays.org

[33] F. Nolan, *The Phonetic Bases of Speaker Recognition*. Cambridge, U.K.: Cambridge Univ. Press, 1983.

[34] J. J. Wolf, "Efficient acoustic parameters for speaker recognition," *J. Acoust. Soc. Amer*, vol. 51, no. 68, p. 2044, 1972.

[35] P. Rose, *Forensic Speaker Identification*. Boca Raton, FL: CRC Press, 2004.

[36] P. Rose, "The technical comparison of forensic voice samples," in *Expert Evidence*, 1 ed. Sydney, Australia: Thompson Lawbook Co., 2003, Chap. 99, pp. 1051–6102.

[37] F. Nolan, "The limitations of auditory-phonetic speaker identification," in *Texte Zur Theorie Und Praxis Forensischer Linguistik*, H. Kniffka, Ed. Berlin, Germany: De Gruyter, 1990, pp. 457–479.

[38] L. Watts, "Reverse-engineering the human auditory pathway," in *Advances in Computational Intelligence*. New York: Springer, 2012, pp. 47–59.

[39] E. Shriberg, L. Ferrer, S. Kajarekar, A. Venkataraman, and A. Stolcke, "Modeling prosodic feature sequences for speaker recognition," *Speech Commun.*, vol. 46, no. 3–4, pp. 455–472, July 2005.

[40] A. G. Adami, R. Mihaescu, D. A. Reynolds, and J. J. Godfrey, "Modeling prosodic dynamics for speaker recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'03)*, pp. 788–791.

[41] N. Dehak, P. Dumouchel, and P. Kenny, "Modeling prosodic features with joint factor analysis for speaker verification," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 15, no. 7, pp. 2095–2103, 2007.

[42] J. H. Wigmore, "A new mode of identifying criminals," *17 Amer Inst. Crim. L. Criminology 165*, vol. 17, no. 2, pp. 165–166, Aug. 1926.

[43] L. G. Kersta, "Voiceprint identification," *Police L. Q.*, vol. 3, no. 5, 1973–1974.

[44] B. E. Koenig, "Spectrographic voice identification: A forensic survey," *J. Acoust. Soc. Amer*, vol. 79, no. 6, pp. 2088–2090, 1986.

[45] H. F. Hollien, *Forensic Voice Identification*. New York: Academic Press, 2002.

[46] L. Yount, *Forensic Science: From Fibers to Fingerprints*. New York: Chelsea House, 2007.

[47] J. P. Campbell, W. Shen, W. M. Campbell, R. Schwartz, J. Bonastre, and D. Matrouf, "Forensic speaker recognition," *IEEE Signal Process. Mag.*, vol. 26, no. 2, pp. 95–103, 2009.

[48] G. S. Morrison, "Distinguishing between science and pseudoscience in forensic acoustics," in *Proc. Meetings on Acoustics*, 2013, pp. 060001.

[49] J. Neyman and E. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philos. Trans. R. Soc.*, vol. 231, pp. 289–337, Jan. 1933.

[50] G. S. Morrison, "Forensic voice comparison and the paradigm shift," *Sci. Justice*, vol. 49, no. 4, pp. 298–308, 2009.

[51] G. S. Morrison, "Forensic voice comparison," in *Expert Evidence 99*, 1 ed. London: Thompson Reuters, 2010, Chap. 99, p. 1051.

[52] R. H. Bolt, *On the Theory and Practice of Voice Identification*. Washington, DC: National Academy of Sciences, 1979.

[53] B. S. Kisilevsky, S. M. Hains, K. Lee, X. Xie, H. Huang, H. H. Ye, K. Zhang, and Z. Wang, "Effects of experience on fetal voice recognition," *Psychol. Sci.*, vol. 14, no. 3, pp. 220–224, May 2003.

[54] M. Latinus and P. Belin, "Human voice perception," *Curr. Biol.*, vol. 21, no. 4, pp. R143–R145, Feb. 2011.

[55] P. Belin, P. E. Bestelmeyer, M. Latinus, and R. Watson, "Understanding voice perception," *Br. J. Psychol.*, vol. 102, no. 4, pp. 711–725, Nov. 2011.

[56] J. Cacioppo, *Foundations in Social Neuroscience*. Cambridge, MA: MIT Press, 2002.

[57] D. Van Lancker and J. Kreiman, "Voice discrimination and recognition are separate abilities," *Neuropsychologia*, vol. 25, no. 5, pp. 829–834, 1987.

[58] D. Van Lancker, J. Kreiman, and K. Emmorey, "Familiar voice recognition: Patterns and parameters. Part I: Recognition of backward voices," *J. Phonetics*, vol. 13, no. 1, pp. 19–38, 1985.

[59] T. K. Perrachione, S. N. Del Tufo, and J. D. Gabrieli, "Human voice recognition depends on language ability," *Science*, vol. 333, no. 6042, pp. 595–595, July 2011.

[60] R. J. Zatorre and P. Belin, "Spectral and temporal processing in human auditory cortex," *Cereb. Cortex*, vol. 11, pp. 946–953, Oct. 2001.

[61] H. Hollien, W. Majewski, and P. Hollien, "Perceptual identification of voices under normal, stress, and disguised speaking conditions," *J. Acoust. Soc. Amer*, vol. 56, no. S53, 1974.

[62] S. M. Kassin, I. E. Dror, and J. Kukucka, "The forensic confirmation bias: Problems, perspectives, and proposed solutions," *J. Appl. Res. Memory Cognit.*, vol. 2, no. 1, pp. 42–52, Mar. 2013.

[63] G. Papcun, J. Kreiman, and A. Davis, "Long-term memory for unfamiliar voices," *J. Acoust. Soc. Amer*, vol. 85, no. 2, pp. 913, Feb. 1989.

[64] S. Cook and J. Wilding, "Earwitness testimony: Never mind the variety, hear the length," *Appl. Cognit. Psychol.*, vol. 11, no. 2, pp. 95–111, Apr. 1997.

[65] A. E. Rosenberg, "Automatic speaker verification: A review," *Proc. IEEE*, vol. 64, no. 4, pp. 475–487, 1976.

[66] B. S. Atal, "Automatic recognition of speakers from their voices," *Proc. IEEE*, vol. 64, no. 4, pp. 460–475, 1976.

[67] G. R. Doddington, "Speaker recognition—Identifying people by their voices," *Proc. IEEE*, vol. 73, no. 11, pp. 1651–1664, 1985.

[68] J. M. Naik, "Speaker verification: A tutorial," *IEEE Commun. Mag.*, vol. 28, no. 1, pp. 42–48, 1990.

[69] S. Furui, "Speaker-dependent-feature extraction, recognition and processing techniques," *Speech Commun.*, vol. 10, no. 5–6, pp. 505–520, Dec. 1991.

[70] H. Gish and M. Schmidt, "Text-independent speaker identification," *IEEE Signal Process. Mag.*, vol. 11, no. 4, pp. 18–32, 1994.

[71] R. J. Mammone, X. Zhang, and R. P. Ramachandran, "Robust speaker recognition: A feature-based approach," *IEEE Signal Process. Mag.*, vol. 13, no. 5, p. 58, 1996.

[72] S. Furui, "Recent advances in speaker recognition," *Pattern Recog. Lett.*, vol. 18, no. 9, pp. 859–872, Sept. 1997.

[73] J. P. Campbell Jr., "Speaker recognition: A tutorial," *Proc. IEEE*, vol. 85, no. 9, pp. 1437–1462, 1997.

[74] F. Bimbot, J. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-García, et al., "A tutorial on text-independent speaker verification," *EURASIP J. Appl. Signal Process.*, vol. 2004, pp. 430–451, Apr. 2004.

[75] A. F. Martin and M. A. Przybocki, "The NIST speaker recognition evaluations: 1996-2001," in *Proc. Odyssey: The Speaker and Language Recognition Workshop*, Crete, Greece, pp. 1–5, 2001.

[76] C. S. Greenberg and A. F. Martin, "NIST speaker recognition evaluations 1996-2008," in *Proc. SPIE Defense, Security, and Sensing*, 2009, pp. 732411–732411-12.

[77] A. F. Martin and C. S. Greenberg, "The NIST 2010 speaker recognition evaluation." in *Proc. Interspeech*, 2010, pp. 2726–2729.

[78] G. R. Doddington, M. A. Przybocki, A. F. Martin, and D. A. Reynolds, "The NIST speaker recognition evaluation–overview, methodology, systems, results, perspective," *Speech Commun.*, vol. 31, no. 2–3, pp. 225–254, June 2000.

[79] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 4, pp. 788–798, 2011.

[80] J. Markel, B. Oshika, and A. Gray, Jr., "Long-term feature averaging for speaker recognition," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 25, no. 4, pp. 330–337, 1977.

[81] K. Li and E. Wrench Jr., "An approach to text-independent speaker recognition with short utterances," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'83)*, 1983, pp. 555–558.

[82] D. Reynolds, W. Andrews, J. Campbell, J. Navratil, B. Peskin, A. Adami, Q. Jin, D. Klusacek, et el., "The SuperSID project: Exploiting high-level information for high-accuracy speaker recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'03)*, pp. 784–787.

[83] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," *IEEE Signal Process. Lett.*, vol. 6, no. 1, pp. 1–3, 1999.

[84] F. Beritelli and A. Spadaccini, "The role of voice activity detection in forensic speaker verification," in *Proc. Digital Signal Processing*, 2011, pp. 1–6.

[85] S. O. Sadjadi and J. H. L. Hansen, "Unsupervised speech activity detection using voicing measures and perceptual spectral flux," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 197–200, 2013.

[86] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980.

[87] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," *J. Acoust. Soc. Amer*, vol. 87, no. 4, pp. 1738, Apr. 1990.

[88] A. V. Oppenheim and R. W. Schafer, "From frequency to quefrency: A history of the cepstrum," *IEEE Signal Process. Mag.*, vol. 21, no. 5, pp. 95–106, 2004.

[89] S. Furui, "Cepstral analysis technique for automatic speaker verification," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 29, no. 2, pp. 254–272, 1981.

[90] J. Pelecanos and S. Sridharan, "Feature warping for robust speaker verification," in *Proc. Odyssey: The Speaker and Language Recognition Workshop*, Crete, Greece, pp. 1–5, 2001.

[91] H. Hermansky and N. Morgan, "RASTA processing of speech," *IEEE Trans. Speech Audio Processing*, vol. 2, no. 4, pp. 578–589, 1994.

[92] H. Boril and J. H. L. Hansen, "Unsupervised equalization of Lombard effect for speech recognition in noisy adverse environments," *IEEE Trans. Audio, Speech, Lang. Processing*, vol. 18, no. 6, pp. 1379–1393, 2010.

[93] P. Matejka, O. Glembek, F. Castaldo, M. J. Alam, O. Plchot, P. Kenny, L. Burget, and J. Cernocky, "Full-covariance UBM and heavy-tailed PLDA in i-vector speaker verification," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'11)*, pp. 4828–4831.

[94] P. Kenny, "Bayesian speaker verification with heavy tailed priors," in *Proc. Odyssey: The Speaker and Language Recognition Workshop*, Brno, Czech Republic, 2010.

[95] A. Mohamed, G. E. Dahl, and G. Hinton, "Acoustic modeling using deep belief networks," *IEEE Trans. Audio, Speech Lang. Processing*, vol. 20, no. 1, pp. 14–22, 2012.

[96] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. Speech Audio Processing*, vol. 3, no. 1, pp. 72–83, 1995.

[97] F. Soong, A. Rosenberg, L. Rabiner, and B. Juang, "A vector quantization approach to speaker recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'85)*, pp. 387–390.

[98] D. Burton, "Text-dependent speaker verification using vector quantization source coding," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 35, no. 2, pp. 133-143, 1987.

[99] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[100] A. E. Rosenberg and S. Parthasarathy, "Speaker background models for connected digit password speaker verification," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'96)*, pp. 81–84.

[101] A. E. Rosenberg, J. DeLong, C. Lee, B. Juang, and F. K. Soong, "The use of cohort normalized scores for speaker verification." in *Proc. Int. Conf. Spoken Language Processing*, 1992, pp. 599–602.

[102] D. A. Reynolds, "Comparison of background normalization methods for text-independent speaker verification," in *Proc. Eurospeech*, pp. 963–966, 1997.

[103] J. Gauvain and C. Lee, "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains," *IEEE Trans. Speech Audio Processing*, vol. 2, no. 2, pp. 291–298, 1994.

[104] R. Kuhn, P. Nguyen, J. Junqua, L. Goldwasser, N. Niedzielski, S. Fincke, K. Field, and M. Contolini, "eigenvoices for speaker adaptation." in *Proc. Int. Conf. Spoken Language Processing*, 1998, pp. 1774–1777.

[105] P. Kenny, M. Mihoubi, and P. Dumouchel, "New MAP estimators for speaker recognition," in *Proc. Interspeech*, Geneva, Switzerland, pp. 2964–2967, 2003.

[106] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learning*, vol. 20, no. 3, pp. 273–297, Sept. 1995.

[107] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 308–311, 2006.

[108] A. Solomonoff, W. M. Campbell, and I. Boardman, "Advances in channel compensation for SVM speaker recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'05)*, pp. 629–632.

[109] A. O. Hatch, S. S. Kajarekar, and A. Stolcke, "Within-class covariance normalization for SVM-based speaker recognition." in *Proc. Interspeech*, Pittsburgh, PA, pp. 1471–1474, 2006.

[110] W. M. Campbell, "Generalized linear discriminant sequence kernels for speaker recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'02)*, pp. 161–164.

[111] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

[112] P. Kenny and P. Dumouchel, "Disentangling speaker and channel effects in speaker verification," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'04)*, pp. 37–40.

[113] P. Kenny, P. Ouellet, N. Dehak, V. Gupta, and P. Dumouchel, "A study of inter-speaker variability in speaker verification," *IEEE Trans. Audio, Speech, Lang. Processing*, vol. 16, no. 5, pp. 980–988, 2008.

[114] R. Kuhn, J. Junqua, P. Nguyen, and N. Niedzielski, "Rapid speaker adaptation in eigenvoice space," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 6, pp. 695–707, 2000.

[115] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Comput.*, vol. 11, no. 2, pp. 443–482, Feb. 1999.

[116] P. Kenny, G. Boulianne, and P. Dumouchel, "Eigenvoice modeling with sparse training data," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 3, pp. 345–354, 2005.

[117] S. Roweis, "EM algorithms for PCA and SPCA," *Adv. Neural Inform. Process. Syst.*, vol. 10, no. 1, pp. 626–632, 1998.

[118] P. Kenny, "Joint factor analysis of speaker and session variability: Theory and algorithms," Tech. Rep. CRIM-06/08-13, CRIM, Montreal, Quebec, Canada, 2005.

[119] N. Dehak, P. Kenny, R. Dehak, O. Glembek, P. Dumouchel, L. Burget, V. Hubeika, and F. Castaldo, "Support vector machines and joint factor analysis for speaker verification," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'09)*, pp. 4237–4240.

[120] N. Dehak, R. Dehak, P. Kenny, N. Brümmer, P. Ouellet, and P. Dumouchel, "Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification." in *Proc. Interspeech*, 2009, pp. 1559–1562.

[121] S. J. Prince and J. H. Elder, "Probabilistic linear discriminant analysis for inferences about identity," in *Proc. IEEE Int. Conf. Computer Vision*, 2007, pp. 1–8.

[122] D. Garcia-Romero and C. Y. Espy-Wilson, "Analysis of i-vector length normalization in speaker recognition systems," in *Proc. Interspeech*, 2011, pp. 249–252.

[123] E. Adar, "GUESS: A language and interface for graph exploration," in *Proc. ACM's Special Interest Group on Computer-Human Interaction*, 2006, pp. 791–800.

[124] N. Brummer, "Measuring, refining and calibrating speaker and language information extracted from speech," Ph.D. diss. Stellenbosch, Univ. Stellenbosch, 2010.

[125] Z. Wu, C. E. Siong, and H. Li, "Detecting converted speech and natural speech for anti-spoofing attack in speaker recognition." in *Proc. Interspeech*, Portland, OR, pp. 1700–1703, 2012.

[126] Z. Wu, N. Evans, T. Kinnunen, J. Yamagishi, F. Alegre, and H. Li, "Spoofing and countermeasures for speaker verification: a survey," *Speech Commun.*, vol. 66, pp. 130–153, Feb. 2015.

[127] G. Liu and J. H. L. Hansen, "An investigation into back-end advancements for speaker recognition in multi-session and noisy enrollment scenarios," *IEEE/ACM Trans. Audio, Speech Lang. Processing*, vol. 22, no. 12, pp. 1978–1992, 2014.

[128] S. Novoselov, T. Pekhovsky, K. Simonchik, and A. Shulipa, "RBM-PLDA subsystem for the NIST i-Vector Challenge," in *Proc. Interspeech*, Singapore, Sept. 14–18, 2014, pp. 378–382.

[129] A. K. Sarkar, D. Matrouf, P. Bousquet, and J. Bonastre, "Study of the effect of I-vector modeling on short and mismatch utterance duration for speaker verification," in *Proc. Interspeech*, Portland, OR, pp. 2662–2665, 2012.

[130] T. Hasan, R. Saeidi, J. H. Hansen, and D. A. van Leeuwen, "Duration mismatch compensation for i-vector based speaker recognition systems," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'13)*, pp. 7663–7667.

[131] R. Travadi, M. Van Segbroeck, and S. Narayanan, "Modified-prior i-vector estimation for language identification of short duration utterances," in *Proc. Interspeech*, pp. 3037–3041, 2014.

[132] G. S. Morrison, "Tutorial on logistic-regression calibration and fusion: Converting a score to a likelihood ratio," *Aust. J. Foren. Sci.*, vol. 45, no. 2, pp. 173–197, Dec. 2013.

[133] V. Hautamäki, K. Lee, D. A. van Leeuwen, R. Saeidi, A. Larcher, T. Kinnunen, T. Hasan, S. O. Sadjadi, et al., "Automatic regularization of cross-entropy cost for speaker recognition fusion." in *Proc. Interspeech*, 2013, pp. 1609–1613.

[134] O. Ghahabi and J. Hernando, "i-Vector modeling with deep belief networks for multi-session speaker recognition," in *Proc. Odyssey:The Speaker and Language Recognition Workshop*, Joensuu, Finland, June 16–19, 2014, pp. 305–310.

[135] Y. Lei, N. Scheffer, L. Ferrer, and M. McLaren, "A novel scheme for speaker recognition using a phonetically-aware deep neural network," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'14)*, pp. 1695–1699.

[136] P. Kenny, V. Gupta, T. Stafylakis, P. Ouellet, and J. Alam, "Deep neural networks for extracting Baum-Welch statistics for speaker recognition," in *Odyssey: The Speaker and Language Recognition Workshop*, Joensuu, Finland.

[137] T. Hasan and J. H. L. Hansen, "Maximum likelihood acoustic factor analysis models for robust speaker verification in noise," *IEEE/ACM Trans. Audio, Speech Lang. Processing*, vol. 22, no. 2, pp. 381–391, 2014.

[138] C. S. Greenberg, D. Bansé, G. R. Doddington, D. Garcia-Romero, J. J. Godfrey, T. Kinnunen, A. F. Martin, A. McCree, et al., "The NIST 2014 speaker recognition i-vector machine learning challenge," in *Proc. Odyssey: The Speaker and Language Recognition Workshop*, pp. 224–230, 2014.

[139] G. Liu, C. Yu, A. Misra, N. Shokouhi, and J. H. Hansen, "Investigating state-of-the-art speaker verification in the case of unlabeled development data," in *Proc. Odyssey: The Speaker and Language Recognition Workshop*, Joensuu, Finland, pp. 118–122, 2014.

[140] T. Hasan, S. O. Sadjadi, G. Liu, N. Shokouhi, H. Boril, and J. H. Hansen, "CRSS systems for 2012 NIST speaker recognition evaluation," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'13)*, pp. 6783–6787.

[141] P. Kenny, T. Stafylakis, P. Ouellet, M. Alam, and P. Dumouchel, "PLDA for speaker verification with utterances of arbitrary duration," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'13)*, pp. 7649–7653.

[142] N. Scheffer and Y. Lei, "Content matching for short duration speaker recognition," in *Proc. Interspeech*, Joensuu, Finland, 2014, pp. 1317–1321.

[143] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, 2012.

[144] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *IEEE Trans. Audio, Speech Lang. Processing*, vol. 20, no. 1, pp. 30–42, 2012.

[145] T. Yamada, L. Wang, and A. Kai, "Improvement of distant-talking speaker identification using bottleneck features of DNN," in *Proc. Interspeech*, Lyon, France, 2013, pp. 3661–3664.

[146] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, 2013.

[147] A. Schmidt-Nielsen and T. H. Crystal, "Human vs. machine speaker identification with telephone speech." in *Proc. ICSLP*, 1998.

[148] D. O'Shaughnessy, *Speech Communication: Human and Machine*. India: Universities Press, 1987.

[149] J. Lindh and G. S. Morrison, "Humans versus machine: Forensic voice comparison on a small database of swedish voice recordings," in *Proc. Int. Congress of Phonetic Sciences (ICPhS)*, 2011, p. 4.

[150] A. Schmidt-Nielsen and T. H. Crystal, "Speaker verification by human listeners: Experiments comparing human and machine performance using the NIST 1998 speaker evaluation data," *Digital Signal Process.*, vol. 10, no. 1–3, pp. 249–266, Jan. 2000.

[151] V. Hautamäki, T. Kinnunen, M. Nosratighods, K. Lee, B. Ma, and H. Li, "Approaching human listener accuracy with modern speaker verification," in *Proc. Interspeech* 2010, Makuhari, Chiba, Japan, Sept. 26–30, 2010, pp. 1473–1476.

[152] R. Schwartz, J. P. Campbell, W. Shen, D. E. Sturim, W. M. Campbell, F. S. Richardson, R. B. Dunn, and R. Granville, "USSS-MITLL 2010 human assisted speaker recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'11)*, pp. 5904–5907.

[153] D. Ramos, J. Franco-Pedroso, and J. Gonzalez-Rodriguez, "Calibration and weight of the evidence by human listeners. The ATVS-UAM submission to NIST human-aided speaker recognition 2010," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'11)*, pp. 5908–5911.

[154] J. Kahn, N. Audibert, S. Rossato, and J. Bonastre, "Speaker verification by inexperienced and experienced listeners vs. speaker verification system," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'11)*, pp. 5912–5915.

[155] C. S. Greenberg, A. F. Martin, G. R. Doddington, and J. J. Godfrey, "Including human expertise in speaker recognition systems: Report on a pilot evaluation," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'11)*, pp. 5896–5899.

[156] W. Shen, J. Campbell, D. Straub, and R. Schwartz, "Assessing the speaker recognition performance of naïve listeners using mechanical turk," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'11)*, pp. 5916–5919.

[157] R. G. Hautamäki, V. Hautamäki, P. Rajan, and T. Kinnunen, "Merging human and automatic system decisions to improve speaker recognition performance," in *Proc. Interspeech*, pp. 2519–2523, 2013.

[158] R. G. Hautamäki, T. Kinnunen, V. Hautamäki, and A. Laukkanen, "Comparison of human listeners and speaker verification systems using voice mimicry data," in *Proc. Odyssey: The Speaker and Language Recognition Workshop*, Joensuu, Finland, 2014, pp. 137–144.

[159] K. J. Han, M. K. Omar, J. Pelecanos, C. Pendus, S. Yaman, and W. Zhu, "Forensically inspired approaches to automatic speaker recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'11)*, pp. 5160–5163.

[160] NIST OSAC—The Organization of Scientific Area Committees. (2015). [Online]. Available: http://www.nist.gov/forensics/osac.cfm

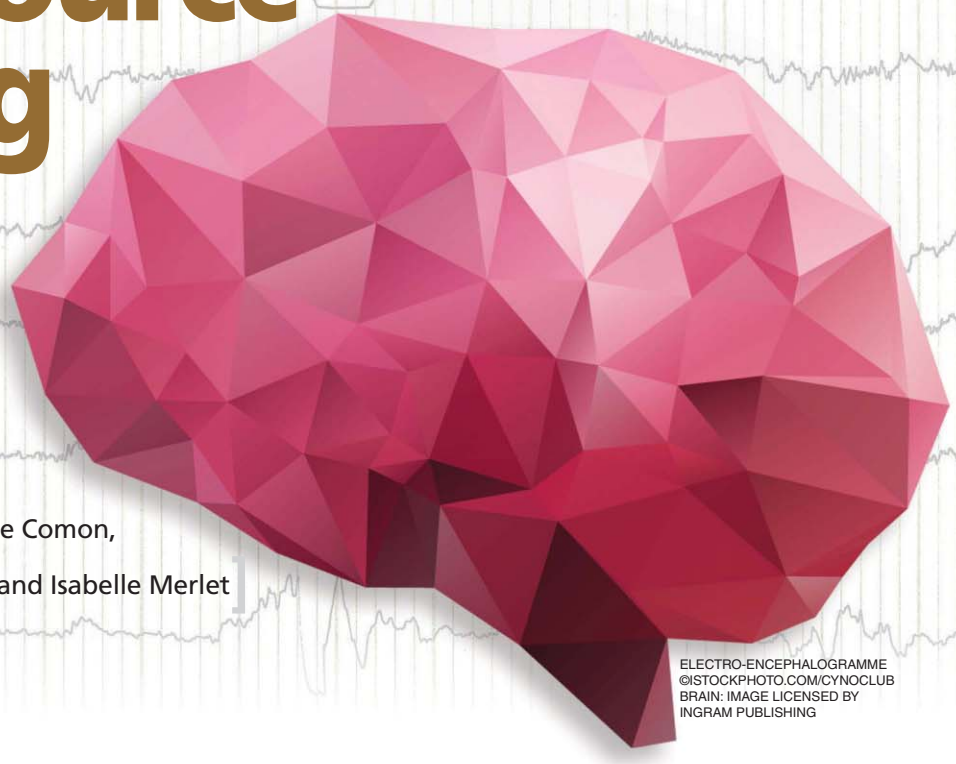
[161] NIST OSAC. (2015). NIST forensic science publications. [Online]. Available: http://www.nist.gov/forensics/publications.cfm

[SP]

Brain-Source Imaging

From sparse to tensor models

Hanna Becker, Laurent Albera, Pierre Comon, Rémi Gribonval, Fabrice Wendling, and Isabelle Merlet



ELECTRO-ENCEPHALOGRAMME ©ISTOCKPHOTO.COM/CYNOCLUB BRAIN: IMAGE LICENSED BY INGRAM PUBLISHING

A number of application areas such as biomedical engineering require solving an underdetermined linear inverse problem. In such a case, it is necessary to make assumptions on the sources to restore identifiability. This problem is encountered in brain-source imaging when identifying the source signals from noisy electroencephalographic or magnetoencephalographic measurements. This inverse problem has been widely studied during recent decades, giving rise to an impressive number of methods using different priors. Nevertheless, a thorough study of the latter, including especially sparse and tensor-based approaches, is still missing. In this article, we propose 1) a taxonomy of the algorithms based on methodological considerations; 2) a discussion of the identifiability and convergence properties, advantages, drawbacks, and application domains of various techniques; and 3) an illustration of the performance of seven selected methods on identical data sets. Directions for future research in the area of biomedical imaging are eventually provided.

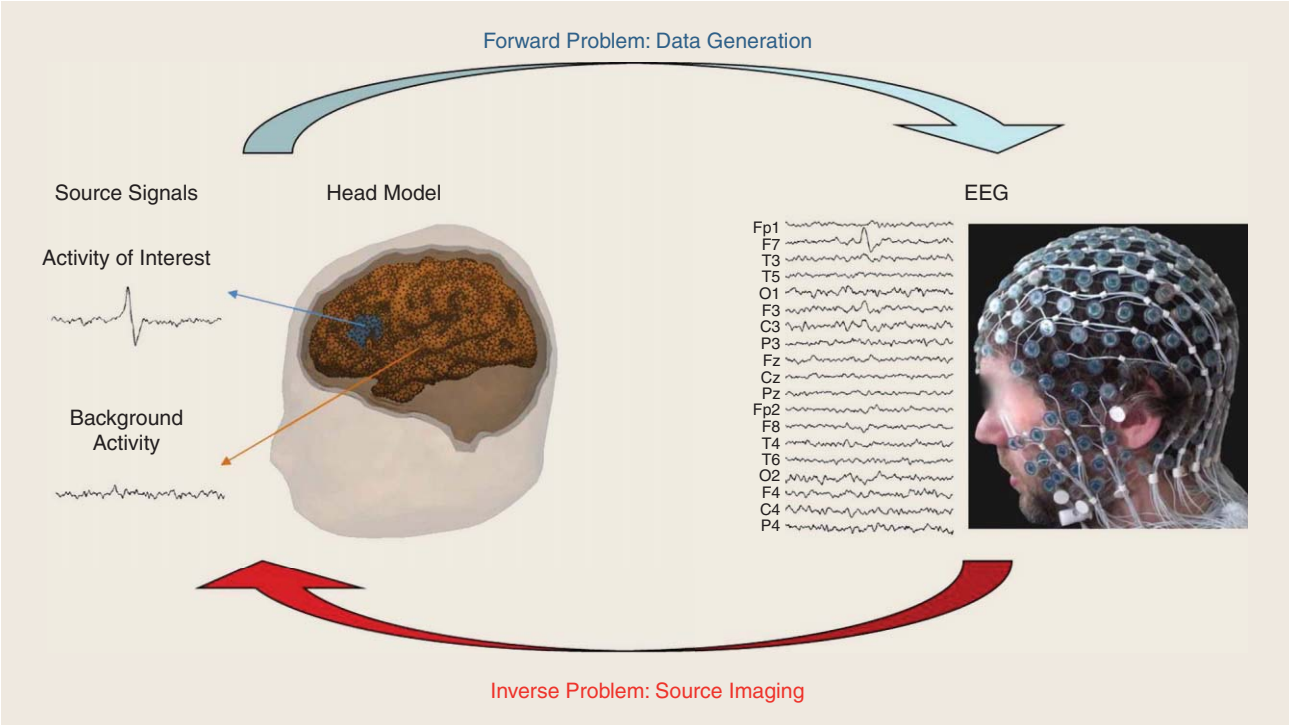
INTRODUCTION

In brain-source imaging, one is confronted with the analysis of a linear static system—the head volume conductor—that relates the electromagnetic activity originating from a number of sources located inside the brain to the surface of the head, where it can be measured with an array of electric or magnetic sensors using electroencephalography (EEG) or magnetoencephalography (MEG). The source signals and locations contain valuable information about the activity of the brain, which is crucial for the diagnosis and management of diseases such as epilepsy or for the understanding of the brain functions in neuroscience research. However, without surgical intervention, the source signals cannot be directly observed and have to be identified from the noisy mixture of signals originating from all over the brain, which is recorded by the EEG/MEG sensors at the surface of the head. This is known as the *inverse problem*. On the other hand, deriving the EEG/MEG signals for a known source configuration is referred to as the *forward problem* (see Figure 1). Thanks to refined models of head geometry and advanced mathematical tools that allow for the computation of the so-called lead-field matrix (referred to as the *mixing matrix* in other domains), solving the forward problem has become straightforward, whereas finding a solution to the inverse problem is still a challenging task.

The methods that are currently available for solving the inverse problem of the brain can be broadly classified into two types of approaches that are based on different source models: the equivalent current dipole and the distributed source [26]. Each equivalent current dipole describes the activity within a spatially extended brain region, leading to a small number of active sources with free orientations and positions anywhere within the brain. The lead-field matrix is, hence, not known but parameterized by the source positions and orientations. Equivalent current dipole methods also include the well-known multiple signal classification (MUSIC) algorithm [1], [42] and beamforming techniques (see [48] and the references therein). These methods are based on a fixed source space with a large number of dipoles, from which a small number of equivalent

Digital Object Identifier 10.1109/MSP.2015.2413711
Date of publication: 13 October 2015

1053-5888/15©2015IEEE



[FIG1] **An illustration of the forward and inverse problems in the context of EEG.**

current dipoles are identified. On the other hand, the distributed source approaches aim at identifying spatially extended source regions, which are characterized by a high number of dipoles (largely exceeding the number of sensors) with fixed locations. As the positions of the source dipoles are fixed, the lead-field matrix can be computed and, thus, is known.

We concentrate on the solution of the inverse problem for the case where the lead-field matrix is known and focus on the distributed source model. This inverse problem is one of the main topics in biomedical engineering [2], [26], [39], [54] and has been widely studied in the signal processing community, giving rise to an impressive number of methods. Our objective is to provide an overview of the currently available source-imaging methods that takes into account the recent advances in the field.

DATA MODEL AND HYPOTHESES

EEG and MEG are multichannel systems that record brain activity over a certain time interval with a number of sensors covering a large part of the head. The two-dimensional measurements are stored in a data matrix $\mathbf{X} \in \mathbb{R}^{N \times T}$, where N denotes the number of EEG/MEG sensors and T the number of recorded time samples. The brain electric and magnetic fields are known to be generated by a number of current sources within the brain, which can be modeled by current dipoles [43]. In this article, we assume that the latter correspond to the dipoles of a predefined source space, which can be derived from structural magnetic resonance imaging. Furthermore, different hypotheses on the location and orientation of the sources can be incorporated by considering either a volume grid of source dipoles with free orientations or a surface grid of source dipoles with fixed orientations. Indeed, most of the activity recorded at the surface of the head is known to originate from pyramidal cells located in the gray matter and oriented perpendicular to the cortical surface [16].

Assuming a source space with free orientation dipoles and denoting $\mathbf{S} \in \mathbb{R}^{3D \times T}$ the signal matrix that contains the temporal activity with which each of the $3D$ dipole components of the D sources contributes to the signals of interest, the measurements at the surface constitute a linear combination of the source signals

$$\mathbf{X} = \mathbf{GS} + \mathbf{N} = \mathbf{GS} + \mathbf{X}_i + \mathbf{X}_b \tag{1}$$

in the presence of noise $\mathbf{N}$. The noise is composed of two parts: instrumentation noise $\mathbf{X}_i$ introduced by the measurement system and background activity $\mathbf{X}_b = \mathbf{GS}_b$, which originates from all dipoles of the source space that do not contribute to the signals of interest but emit perturbing signals $\mathbf{S}_b \in \mathbb{R}^{3D \times T}$. The matrix $\mathbf{G} \in \mathbb{R}^{N \times 3D}$ is generally referred to as the *lead-field matrix* in the EEG/MEG context. For each dipole component of the source space, it characterizes the propagation of the source signal to the sensors at the surface.

In the case of dipoles with fixed orientations, the signal matrices $\mathbf{S}$ and $\mathbf{S}_b$ are replaced by the matrices $\tilde{\mathbf{S}} \in \mathbb{R}^{D \times T}$ and $\tilde{\mathbf{S}}_b \in \mathbb{R}^{D \times T}$, which characterize the brain activity of the D dipoles. Furthermore, the lead-field matrix $\mathbf{G}$ is replaced by the matrix $\tilde{\mathbf{G}} \in \mathbb{R}^{N \times D}$, which is given by $\tilde{\mathbf{G}} = \mathbf{G\Theta}$, where $\mathbf{\Theta} \in \mathbb{R}^{3D \times D}$ contains the fixed orientations of the dipoles. The lead-field matrix $\mathbf{G}$ can be computed numerically based on Maxwell's equations. Several methods have been developed to accomplish this, and various software packages are available [23].

We assume that the lead-field matrix is known and consider the EEG/MEG inverse problem that consists in estimating the unknown

sources $\mathbf{S}$ or $\tilde{\mathbf{S}}$ (depending on the source model) from the measurements $\mathbf{X}$. As the number of source dipoles D (several thousand) is much higher than the number of sensors (several hundred), the lead-field matrix is severely underdetermined, making the inverse problem ill posed. To restore the identifiability of the underdetermined source reconstruction problem, it is necessary to make assumptions on the sources. We discuss a large number of hypotheses that have been introduced in the context of the EEG/MEG inverse problem. In the following sections, we distinguish between three categories of assumptions depending on whether the hypotheses apply to the spatial, temporal, or spatiotemporal (deterministic or statistical) distribution of the sources, represented by "Sp," "Te," and "SpTe" respectively. Subsequently, we provide a short description of the possible hypotheses.

HYPOTHESES ON THE SPATIAL DISTRIBUTION OF THE SOURCES

Sp1 MINIMUM ENERGY

The power of the sources is physiologically limited. A popular approach thus consists in identifying the spatial distribution of minimum energy.

Sp2 MINIMUM ENERGY IN A TRANSFORMED DOMAIN

Because of a certain synchronization of adjacent neuronal populations, the spatial distribution of the sources is unlikely to contain abrupt changes and can, therefore, be assumed to be smooth. This hypothesis is generally enforced by constraining the Laplacian of the source spatial distribution to be of minimum energy.

Sp3 SPARSITY

In practice, it is often reasonable to assume that only a small fraction of the source dipoles contributes to the measured signals of interest in a significant way. For example, audio or visual stimuli lead to characteristic brain signals in certain functional areas of the brain only. The signals of the other source dipoles are, thus, expected to be zero. This leads to the concept of sparsity.

Sp4 SPARSITY IN A TRANSFORMED DOMAIN

If the number of active dipoles exceeds the number of sensors, which is generally the case for spatially extended sources, the source distribution is not sufficiently sparse for standard methods based on sparsity in the spatial domain to yield accurate results, leading to too-focused source estimates. In this context, another idea consists in transforming the sources into a domain where their distribution is sparser than in the original source space and imposing sparsity in the transformed domain. The applied transform may be redundant, including a large number of basis functions or atoms, and is not necessarily invertible.

Sp5 SEPARABILITY IN SPACE AND WAVE-VECTOR DOMAINS

For each distributed source, one can assume that the space-wave-vector matrix at each time point, which is obtained by computing a local spatial Fourier transform of the measurements, can be factorized into a function that depends on the space variable only and a function that depends on the wave-vector variable only. The space and wave-vector variables are, thus, said to be separable. In the context of brain-source imaging, this is approximately the case for superficial sources.

Sp6 GAUSSIAN JOINT PROBABILITY DENSITY FUNCTION WITH PARAMETERIZED SPATIAL COVARIANCE

For this prior, the source signals are assumed to be random variables that follow a Gaussian distribution with a spatial covariance matrix that can be described by a linear combination of a certain number of basis covariance functions. This combination is characterized by so-called hyperparameters, which have to be identified in the source-imaging process.

HYPOTHESES ON THE TEMPORAL DISTRIBUTION OF THE SOURCES

Te1 SMOOTHNESS

Since the autocorrelation function of the sources of interest usually has a full width at half maximum of several samples, the source time distribution should be smooth. For example, this is the case for interictal epileptic signals or event-related potentials.

Te2 SPARSITY IN A TRANSFORMED DOMAIN

Similar to hypothesis Sp4, this assumption implies that the source signals admit a sparse representation in a domain that is different from the original time domain. This can, for instance, be achieved by applying a wavelet transform or a redundant transformation such as the Gabor transform to the time dimension of the data. The transformed signals can then be modeled using a small number of basis functions or atoms, which are determined by the source-imaging algorithm.

Te3 PSEUDOPERIODICITY WITH VARIATIONS IN AMPLITUDE

If the recorded data comprise recurrent events such as a repeated time pattern that can be associated with the sources of interest, one can exploit the repetitions as an additional diversity. This does not necessarily require periodic or quasiperiodic signals. Indeed, the intervals between the characteristic time patterns may differ, as may the amplitudes of different repetitions. Examples of signals with repeated time patterns include interictal epileptic spikes and event-related potentials (ERPs).

Te4 SEPARABILITY IN TIME AND FREQUENCY DOMAINS

This hypothesis is the equivalent of hypothesis Sp5 and assumes that the time and frequency variables of data transformed into the time-frequency domain [e.g., by applying a short-time Fourier transform (STFT) or a wavelet transform to the measurements] separate. This is approximately the case for oscillatory signals as encountered, for example, in epileptic brain activity.

Te5 NONZERO HIGHER-ORDER MARGINAL CUMULANTS

Regarding the measurements as realizations of an N-dimensional vector of random variables, this assumption is required when

resorting to statistics of an order higher than two, which offer a better performance and identifiability than approaches based on second-order statistics. It is generally verified in practice, as the signals of interest usually do not follow a Gaussian distribution.

HYPOTHESES ON THE SPATIOTEMPORAL DISTRIBUTION OF THE SOURCES

SpTe SYNCHRONOUS DIPOLES

Contrary to point sources, which can be modeled by a single dipole, in practice, one is often confronted with so-called distributed sources. A distributed source is composed of a certain number of grid dipoles, which can be assumed to transmit synchronous signals. This hypothesis concerns both the spatial and the temporal distributions of the sources and is generally made in the context of dipoles with fixed orientations. In this case, it allows for the separation of the matrix $\tilde{\mathbf{S}}_{\mathcal{I}_r}$, which contains the signals of all synchronous dipoles of the rth distributed source, indicated by the set $\mathcal{I}_r$, into the coefficient vector ϕ_r that characterizes the amplitudes of the synchronous dipoles and thereby the spatial distribution of the rth distributed source and the signal vector $\bar{\mathbf{s}}$ that contains the temporal distribution of the distributed source. This gives rise to a new data model

$$\mathbf{X} = \mathbf{H}\bar{\mathbf{S}} + \mathbf{N}, \tag{2}$$

where the matrix $\mathbf{H} = [\mathbf{h}_1, \cdots, \mathbf{h}_R]$ contains the lead-field vectors for R distributed sources and the matrix $\bar{\mathbf{S}} \in \mathbb{R}^{R\times T}$ characterizes the associated distributed source signals. Each distributed source lead-field vector $\mathbf{h}_r$ corresponds to a linear combination of the lead-field vectors of all grid dipoles belonging to the distributed source: $\mathbf{h}_r = \tilde{\mathbf{G}}\phi_r$. The distributed source lead-field vectors can be used as inputs for source-imaging algorithms, simplifying the inverse problem by allowing for a separate localization of each source.

HYPOTHESES ON THE NOISE

While both the instrumentation noise and the background activity are often assumed to be Gaussian, the instrumentation noise can be further assumed to be spatially white, whereas the background activity is spatially correlated because signals are mixed. To meet the assumption of spatially white Gaussian noise made by many algorithms, the data can be prewhitened based on an estimate of the noise covariance matrix $\mathbf{C}_n$. More precisely, the prewhitening matrix is computed as the inverse of the square root of the estimated noise covariance matrix. To achieve prewhitening, the data and the lead-field matrices are multiplied from the left by the prewhitening matrix.

ALGORITHMS

In this section, we provide an overview of the various source-imaging methods that have been developed in the context of the EEG/MEG inverse problem. Based on methodological considerations, we distinguish four main families of techniques: regularized least-squares approaches, tensor-based approaches, Bayesian approaches, and extended source scanning approaches. Each class of methods is associated with a certain number of hypotheses that are exploited by the algorithms. The knowledge of these hypotheses leads to a better understanding of the functioning of the source-imaging techniques.

REGULARIZED LEAST-SQUARES METHODS

A natural approach to solve the ill-posed EEG/MEG inverse problem consists of finding the solution that best describes the measurements in a least-squares sense. In the presence of noise, this is generally achieved by solving an optimization problem with a cost function of the form

$$L(\mathbf{S}) = ||\mathbf{X} - \mathbf{GS}||_{\mathrm{F}}^2 + \lambda f(\mathbf{S}). \tag{3}$$

For methods that do not consider the temporal structure of the data, but work on a time-sample-by-sample basis, the data matrix $\mathbf{X}$ and the source matrix $\mathbf{S}$ are replaced by the column vectors $\mathbf{x}$ and $\mathbf{s}$, respectively.

The first term on the right-hand side of (3) is generally referred to as the *data fit term* and characterizes the difference between the measurements and the surface data reconstructed from given sources. The second is a regularization term and incorporates additional constraints on the sources according to the a priori information. The regularization parameter λ is used to manage a tradeoff between data fit and a priori knowledge and depends on the noise level since the gap between the measured and reconstructed data is expected to become larger as the signal-to-noise ratio decreases. Figure 2 provides an overview of the regularized least-squares algorithms with different regularization terms that are discussed in the following sections.

MINIMUM NORM ESTIMATES—ASSUMPTION Sp1 OR Sp2

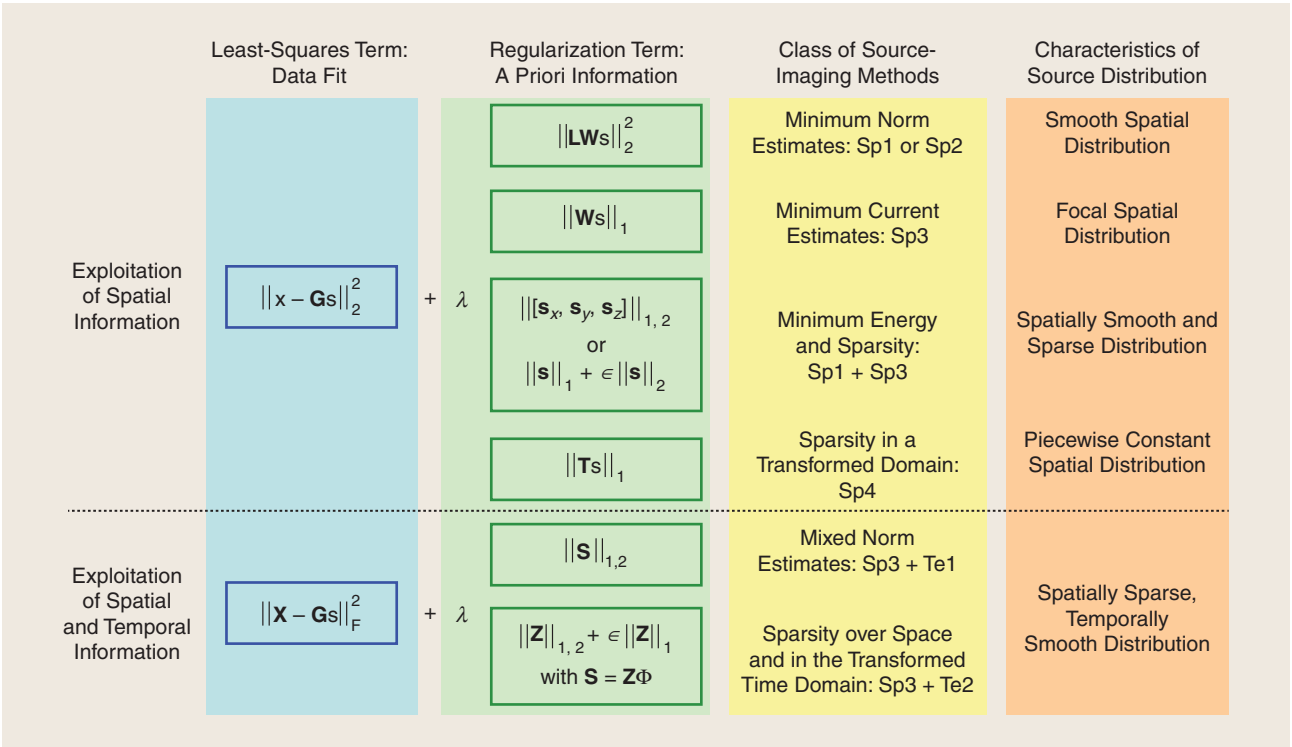
The minimum norm solution is obtained by employing a prior, which imposes a minimal signal power according to hypothesis Sp1, leading to a regularization term that is based on the L_2-norm of the signal vector: $f(\mathbf{s}) = ||\mathbf{Ws}||_2^2$. To compensate for the depth bias, the diagonal matrix $\mathbf{W} \in \mathbb{R}_+^{3D\times 3D}$ containing fixed weights was introduced in the weighted minimum norm estimates (MNE) methods. Furthermore, one can consider the variance of the noise or the sources, leading to normalized estimates. This approach is pursued by the dynamic statistical parametric mapping (dSPM) [15] algorithm, which takes into account the noise level, and standardized low-resolution brain electromagnetic tomography (sLORETA) [45], which standardizes the source estimates with respect to the variance of the sources.

The MNEs generally yield smooth source distributions. Nevertheless, spatial smoothness can also be more explicitly promoted by applying a Laplacian operator $\mathbf{L}$ to the source vector in the regularization term, leading to the popular LORETA method [46], which is based on assumption Sp2. In this case, the L_2-norm constraint is imposed on the transformed signals, yielding a regularization term of the form $f(\mathbf{s}) = \|\mathbf{LWs}\|_2^2$. More generally, the matrix $\mathbf{L}$ can be used to implement a linear operator that is applied to the sources.

The original MNEs have been developed for sources with free orientations. Modifications of the algorithms to account for orientation constraints can, e.g., be found in [34] and [53].

METHODS BASED ON SPARSITY—ASSUMPTION Sp3 OR Sp4

As the MNEs generally lead to blurred source localization results, as widely described in the literature (see, e.g., [56]), source-imaging methods based on hypothesis Sp3, which promote sparsity, were



[FIG2] An overview of regularized least-squares algorithms. (For an explanation of the employed notations for the different algorithms, see the text in the associated sections.)

developed to obtain more focused source estimates. One of the first algorithms proposed in this field was focal underdetermined system solver (FOCUSS) [22], which iteratively updates the minimum norm solution using an L_0 "norm." This gradually shrinks the source spatial distribution, resulting in a sparse solution. Around the same time, source-imaging techniques based on an L_p-norm $(0 \leq p \leq 1)$ regularization term of the form $f(\mathbf{s}) = \|\mathbf{Ws}\|_p$, where $\mathbf{W}$ is a diagonal matrix of weights, were put forward [36]. The parameter p is generally chosen to be equal to 1, leading to a convex optimization problem. (Note that the minimization of this cost function is closely related to the optimization problem $\min ||\mathbf{Ws}||_p \text{s.t.} \|\mathbf{x} - \mathbf{Gs}\|_2^2 \leq \delta$ with regularization parameter δ, on which the algorithm proposed in [36] is based.) However, by treating the dipole components independently in the regularization term, the estimated source orientations are biased. To overcome this problem, Uutela et al. [50] proposed using fixed orientations determined either from the surface normals or estimated using a preliminary minimum norm solution. This gave rise to the minimum current estimate (MCE) algorithm. Extensions of this approach, which require only the knowledge of the signs of the dipole components or which permit the incorporation of loose orientation constraints, have been treated in [29] and [34]. Another solution to the problem of orientation bias of the sparse source estimates consists in imposing sparsity dipolewise instead of componentwise [20]. In [56], a combination of the ideas of FOCUSS and L_p-norm $(p \leq 1)$ regularization was implemented in an iterative scheme.

To find a compromise between the smoothness and sparsity of the spatial distribution, the use of a prior that is composed of both an L_1-norm and an L_2-norm regularization term was proposed in [52]. Another idea consists in imposing sparsity in a transformed domain. This is generally achieved by employing a regularization term of the form $\|\mathbf{T\tilde{s}}\|_1$, where $\mathbf{T}$ is a transformation matrix. In the literature, different transformations have been considered. The authors of [10] used a surface Laplacian, thus imposing sparsity on the second-order spatial derivatives of the source distribution, in combination with classical L_1-norm regularization. Another way to promote a piecewise constant spatial distribution was proposed by Ding, giving rise to the variation-based sparse cortical current density (VB-SCCD) method [19], which is closely related to the total variation approach. A third approach that makes use of sparsity in a transformed domain considers a spatial wavelet transform that allows the signals to be compressed through a sparse representation of the sources in the wavelet domain [10], [31].

MIXED NORM ESTIMATES—ASSUMPTION Sp3 OR Sp4 AND ASSUMPTION Te1 OR Te2

To impose hypotheses simultaneously in several domains, e.g., the space-time plane, one can resort to mixed norms. Efficient algorithms that have been developed to deal with the resulting optimization problem are presented in [24]. In [44], a source-imaging method, called *mixed-norm estimate (MxNE)*, which imposes sparsity over space (hypothesis Sp3) and smoothness over time (assumption Te1) using a mixed $L_{1,2}$-norm regularization, has been proposed.

An approach that imposes sparsity over space (hypothesis Sp3) as well as in the transformed time domain (assumption Te2) is taken in the time-frequency MxNE (TF-MxNE) method.

This technique makes use of a dictionary, $\mathbf{\Phi}$, from which a small number of temporal basis functions are selected to characterize the source signals. In [25], Gabor basis functions were considered, whereas the authors of [49] employed a data-dependent temporal basis obtained using a singular value decomposition (SVD) of the measurements and a data-independent temporal basis that is given by natural cubic splines. The method is based on mixed norms and uses a composite prior of two regularization terms similar to [52].

Furthermore, in [28], one can find an approach that imposes sparsity in a spatial transform domain similar to [10], but which is based on a mixed $L_{1,2}$-norm to take into account the temporal smoothness of the source distribution. Finally, let us point out that it is also possible to consider both temporal and spatial basis functions (assumptions Sp4 and Te2) as suggested in [7] for the event sparse penalty (ESP) algorithm.

TENSOR-BASED SOURCE LOCALIZATION—ASSUMPTION SpTe; ASSUMPTION Sp5, Te3, OR Te4: AND ASSUMPTIONS Sp4 AND Sp3

The objective of tensor-based methods consists of identifying the lead-field vectors and the signals of distributed sources, i.e., matrices $\mathbf{H}$ and $\bar{\mathbf{S}}$ in data model (2), from measurements. To separate R simultaneously active distributed sources, tensor-based methods exploit multidimensional data (at least one dimension in addition to space and time) and assume a certain structure underlying the measurements. The multidimensional data are then approximated by a model that reflects the assumed structure and comprises a number of components that can be associated with the sources. A popular tensor model is the rank-R canonical polyadic (CP) decomposition [14], which imposes a multilinear structure on the data. This means that each element of a third-order tensor $\boldsymbol{X}$ can be written as a sum of R components, each being a product of three univariate functions, a_r, b_r, and d_r

$$\boldsymbol{X}(\alpha_k, \beta_\ell, \gamma_m) = \sum_{r=1}^{R} a_r(\alpha_k) b_r(\beta_\ell) d_r(\gamma_m). \tag{4}$$

The samples of functions a_r, b_r, and d_r can be stored into three loading matrices $\mathbf{A} \in \mathbb{C}^{K\times R} = [a_1, \cdots, a_R]$, $\mathbf{B} \in \mathbb{C}^{L\times R} = [\mathbf{b}_1, \cdots, \mathbf{b}_R]$, and $\mathbf{D} \in \mathbb{C}^{M\times R} = [\mathbf{d}_1, \cdots, \mathbf{d}_R]$ that characterize the tensor $\boldsymbol{X} \in \mathbb{C}^{K\times L\times M}$.

In the literature, a certain number of tensor methods based on the CP decomposition have been proposed in the context of EEG/MEG data analysis. These methods differ in the dimension(s), which is (are) exploited in addition to space and time. In this work, we focus on third-order tensors. Here, first, a distinction can be made between approaches that collect an additional diversity directly from the measurements, for instance, by taking different realizations of a repetitive event (see [40]), or methods that create a third dimension by applying a transform which preserves the two original dimensions, such as the STFT or wavelet transform. This transform can be applied either over time or over space, leading to space–time–frequency (STF) data (see, e.g., [17] and the references therein) or space–time–wave–vector (STWV) data [5]. Depending on the dimensions of the tensor, the CP decomposition involves different multilinearity assumptions: for space–time–realization (STR) data, hypothesis Te3 is required; for STF data, hypothesis Te4 is involved; and for STWV data, we resort to hypothesis Sp5.

Once several simultaneously active distributed sources have been separated, using the tensor decomposition, and estimates for the distributed source lead-field vectors have been derived, the latter can be used for source localization. The source localization is then performed separately for each distributed source. For this purpose, a dictionary of potential elementary distributed sources is defined by a number of circular-shaped cortical areas of different centers and sizes, subsequently called *disks*. Each disk describes a source region with constant amplitudes, leading to a sparse, piecewise constant source distribution, which can be attributed to hypotheses Sp3 and Sp4. For each source, a small number of disks that correspond best to the estimated distributed source lead-field vector are then identified based on a metric and are merged to reconstruct the distributed source. The steps of the algorithm based on STWV data and referred to as *STWV-DA* (disk algorithm) [5] are schematically summarized in Figure 3.

BAYESIAN APPROACHES—ASSUMPTION Sp6

Bayesian approaches are based on a probabilistic model of the data and treat the measurements, the sources, and the noise as realizations of random variables. In this context, the reconstruction of the sources corresponds to obtaining an estimate of their posterior distribution, which is given by

$$p(\mathbf{s} \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \mathbf{s}) p(\mathbf{s})}{p(\mathbf{x})}, \tag{5}$$

where $p(\mathbf{x} \mid \mathbf{s})$ is the likelihood of the data, $p(\mathbf{s})$ is the source distribution, and $p(\mathbf{x})$ is the model evidence. The crucial point consists in finding an appropriate prior distribution $p(\mathbf{s})$ for the sources, which, in the Bayesian framework, incorporates the hypotheses that regularize the ill-posed inverse problem. We can distinguish three classes of Bayesian approaches [54]: maximum a posteriori estimation for the sources, variational Bayes, and empirical Bayes. The first approach employs a fixed prior $p(\mathbf{s})$ leading to MNE, MCE, and MxNE solutions, which were addressed earlier. In this section, we focus on variational and empirical Bayesian approaches, which use a flexible, parameterized prior $p(\mathbf{s}|\gamma)$, which is modulated by the hyperparameter vector $\gamma \in \mathbb{R}^L$. More particularly, in the EEG/MEG context, the source distribution is generally assumed to be zero-mean Gaussian with a covariance matrix $\mathbf{C}_s$ that depends on hyperparameters, such that

$$p(\mathbf{s}|\gamma) \propto \exp\left(-\frac{1}{2}\mathbf{S}^{\mathrm{T}}\mathbf{C}_s^{-1}(\gamma)\mathbf{S}\right). \tag{6}$$

The hyperparameters can either directly correspond to the elements of $\mathbf{C}_s$ (as in the Champagne algorithm [55]) or parameterize the covariance matrix such that $\mathbf{C}_s = \sum_{i=1}^{I} \gamma_i \mathbf{C}_i$. Here, $\mathbf{C}_i$, $i = 1, \cdots, I$ are predefined covariance components. The hyperparameters are then learned from the data to perform some kind of model selection by choosing the appropriate components.

VARIATIONAL BAYESIAN APPROACHES

The variational Bayesian methods (see [21] and the references therein) try to obtain estimates of the posterior distributions of the

hyperparameters $\hat{p}(\gamma\,|\,\mathbf{x})$. To do this, additional assumptions are required, such as 1) statistical independence of the hyperparameters (also known as *mean-field approximation*) or 2) a Gaussian posterior distribution of the hyperparameters (also known as *Laplace approximation*). This allows us to not only approximate the distribution $p(\mathbf{s}|\,\mathbf{x})$ and thereby estimate the sources but also to provide an estimate of the model evidence $p(\mathbf{x})$, which can be used to compare different models (e.g., for different sets of covariance components).

EMPIRICAL BAYESIAN APPROACHES

The empirical Bayesian approaches (see, e.g., [37] and [55] and the references therein), on the other hand, are concerned with finding a point estimate of the hyperparameters, which is obtained by marginalization over the unknown sources $\mathbf{s}$

$$\hat{\gamma} = \operatorname*{argmax}_{\gamma} \int p(\mathbf{x}|\,\mathbf{s})p(\mathbf{s}|\,\gamma)p(\gamma)d\mathbf{s}. \tag{7}$$

For known hyperparameters, the conditional distribution $p(\mathbf{s}|\,\mathbf{x},\gamma)$ can be determined. To obtain a suitable estimate of the sources, one can, for instance, apply the expectation maximization (EM) algorithm [18], which alternates between two steps: 1) the M-step in which the maximum likelihood estimates of the hyperparameters are updated for fixed $\mathbf{s}$ and 2) the E-step in which the conditional expectation of the sources is determined based on the hyperparameters obtained in the M-step. An example of an empirical Bayesian algorithm is the Champagne algorithm introduced in [55].

EXTENDED SOURCE SCANNING METHODS

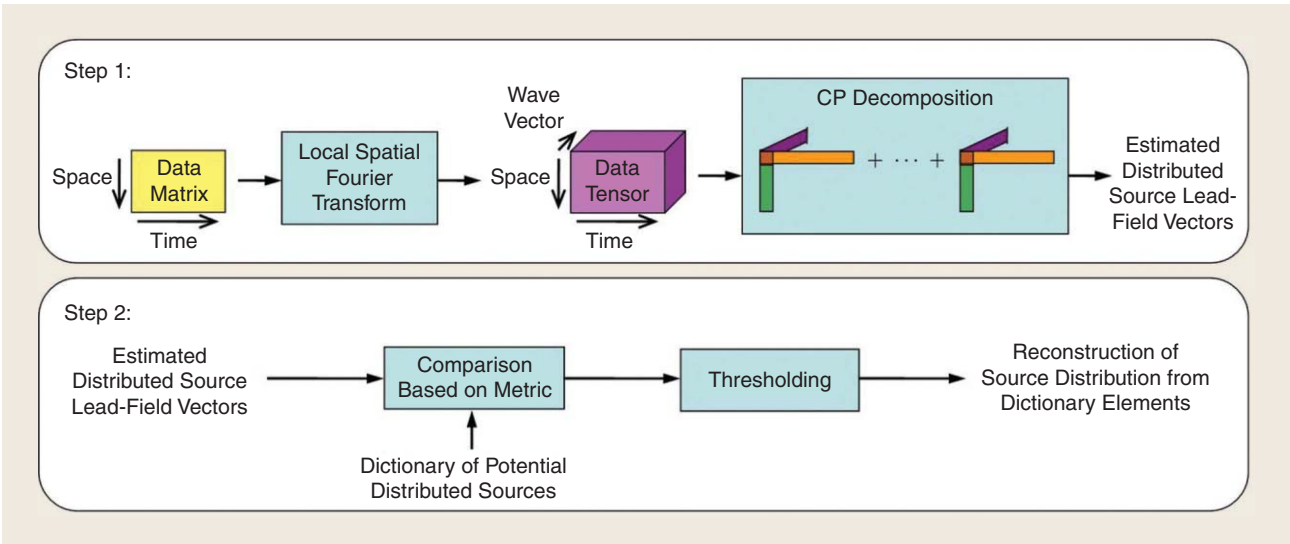
Here, the idea is to identify active sources from a dictionary of potential distributed sources. To this end, a metric is computed for each element of the dictionary. The source estimates are then obtained from the elementary source distributions that are associated with the maxima of the metric. Based on the employed metric, we subsequently distinguish two types of scanning methods that correspond to spatial filtering, also known as *beamforming*, and subspace-based approaches.

BEAMFORMING APPROACHES—ASSUMPTIONS Sp3 AND Sp4

Beamforming techniques were originally proposed in the context of equivalent current dipole localization from MEG measurements [51]. The basic approach employs the linearly constrained minimum variance (LCMV) filter, which is based on the data covariance matrix and is derived for each dipole of the source space to reconstruct its temporal activity while suppressing contributions from other sources. The filter output is then used to compute a metric that serves to identify the active dipole sources. The LCMV beamformer was shown to yield unbiased solutions in the case of a single dipole source [48], but leads to source localization errors in the presence of correlated sources. To overcome this problem, extensions of the beamforming approach to multiple, potentially correlated (dipole) sources have been considered (see [41] and the references therein). Furthermore, in [33], the beamforming approach has been extended to the localization of distributed sources. This is achieved by deriving spatial filters for all elements of a dictionary of potential source regions, also called *patches*. The source-imaging solution is then obtained from the dictionary elements associated with the maxima of the metric, which is derived from the filter outputs, resulting in a spatially sparse source distribution with a small number of active source regions according to hypotheses Sp3 and Sp4.

SUBSPACE-BASED APPROACHES—ASSUMPTIONS SpTe, Te5, Sp3, AND Sp4

Similar to Bayesian approaches, subspace-based methods also treat the measurements made by several sensors as realizations of a random vector. They then exploit the symmetric $2q$th $(q \geq 1)$-order cumulant matrix $\mathbf{C}_{2q,x}$ of this random vector from which the signal



[FIG3] A schematic representation of the STWV-DA algorithm.

and noise subspaces are identified by means of an eigenvalue decomposition. For source-imaging purposes, one then exploits the fact that the higher-order lead-field vector $\tilde{\mathbf{g}}_r^{\otimes q}$, $r = 1, \cdots, R$, where $\tilde{\mathbf{g}}_r^{\otimes q}$ is a shorthand notation for $\tilde{\mathbf{g}}_r \otimes \tilde{\mathbf{g}}_r \otimes \cdots \otimes \tilde{\mathbf{g}}_r$ with $q-1$ Kronecker products (denoted by $\otimes$), must lie in the $2q$th-order signal subspace and be orthogonal to the noise subspace. Therefore, MUSIC-like algorithms can be employed, which were first used in the context of equivalent current dipole localization [1], [42]. Recently, the $2q$-MUSIC algorithm [12] has been adapted to the identification of distributed sources [6], then referred to as *2q-ExSo-MUSIC*. In analogy to the classical MUSIC algorithm, the $2q$-ExSo-MUSIC spectrum is computed for a number of predefined parameter vectors ϕ. To this end, one defines a dictionary of disks as described in the section "Tensor-Based Source Localization," assuming a sparse, piecewise constant source distribution (corresponding to hypotheses Sp3 and Sp4) similar to VB-SCCD and STWV-DA. The spectrum is then thresholded, and all coefficient vectors ϕ for which the spectrum exceeds a fixed threshold are retained and united to model distributed sources. An advantage of subspace-based techniques exploiting the $2q$th order statistics with $q > 1$ over other source-imaging algorithms lies in their asymptotic robustness to Gaussian noise because cumulants of an order higher than two of a Gaussian random variable are null.

DISCUSSION

Here we discuss several aspects of the brain-source-imaging methods described in the previous section, including identifiability and convergence issues, advantages and drawbacks of representative algorithms, and application domains. Table 1 lists several source-imaging methods mentioned in the previous section and summarizes the exploited hypotheses.

IDENTIFIABILITY

For methods that solve the inverse problem by exploiting sparsity, the uniqueness of the solution depends on the conditioning of the lead-field matrix. More particularly, sufficient conditions that are based on the mutual or cumulative coherence of the lead-field matrix are available in the literature [11] and can easily be verified for a given lead-field matrix. However, in brain-source imaging, these conditions are generally not fulfilled because the lead-field vectors of adjacent grid dipoles are often highly correlated, making the lead-field matrix ill conditioned.

A strong motivation for the use of tensor-based methods is the fact that the CP decomposition is essentially unique under mild conditions on the tensor rank [32]. These conditions are generally verified in brain-source imaging because the rank R of the noiseless tensor corresponds to the number of distributed sources, which is usually small (fewer than ten) compared to the tensor dimensions. The limitations of the tensor-based approach thus arise from the approximations that are made when imposing a certain structure on the data and not from the identifiability conditions. Note, however, that these identifiability conditions only concern the CP decomposition, which separates the distributed sources. Additional conditions are indeed required for the uniqueness of the results of the subsequent source localization step that is applied for each distributed source separately. Nevertheless, the separation of the distributed sources facilitates their identification and may alleviate the identifiability conditions for the source localization step.

[TABLE 1] THE CLASSIFICATION OF THE DIFFERENT ALGORITHMS MENTIONED IN THE "ALGORITHMS" SECTION ACCORDING TO THE EXPLOITED HYPOTHESES.

BRAIN-SOURCE IMAGING	Sp1	Sp2	Sp3	Sp4	Sp5	Sp6	Te1	Te2	Te3	Te4	Te5	SpTe
REGULARIZED LEAST-SQUARES ALGORITHMS												
sLORETA [45]	X											
LORETA [46]		X										
MCE [50]			X									
VB-SCCD [19]				X								
MxNE [44]			X				X					
TF-MxNE [25]			X					X				
BAYESIAN APPROACHES												
CHAMPAGNE [55]						X						
EXTENDED SOURCE SCANNING METHODS												
$2q$-ExSo-MUSIC [6]				X							X	X
TENSOR-BASED METHODS												
STR-DA [5]				X					X			X
STF-DA [5]				X						X		X
STWV-DA [5]				X	X							X

HYPOTHESES ON THE SPATIAL DISTRIBUTION

Sp1: MINIMUM ENERGY
Sp2: MINIMUM ENERGY IN A TRANSFORMED DOMAIN
Sp3: SPARSITY
Sp4: SPARSITY IN A TRANSFORMED DOMAIN
Sp5: SEPARABILITY IN THE SPACE-WAVE-VECTOR DOMAIN
Sp6: PARAMETERIZED SPATIAL COVARIANCE

HYPOTHESES ON THE TEMPORAL DISTRIBUTION

Te1: SMOOTHNESS
Te2: SPARSITY IN A TRANSFORMED DOMAIN
Te3: PSEUDOPERIODICITY
Te4: SEPARABILITY IN THE TIME-FREQUENCY DOMAIN
Te5: NONZERO HIGHER-ORDER MARGINAL CUMULANTS

HYPOTHESES ON THE SPATIOTEMPORAL DISTRIBUTION

SpTe: SYNCHRONOUS DIPOLES

IEEE SignalProcessing | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | Qmags THE WORLD'S NEWSSTAND®

Finally, for subspace-based approaches, the number of sources that can be identified depends on the dimensions of the signal and noise subspaces of the cumulant matrix. In the best case, one can identify at most $N_{2q}-1$ statistically independent distributed sources, where $N_{2q} \leq N^q$ denotes the maximal rank that can be attained by the $2q$th-order distributed source lead-field matrix and N is the number of sensors, while, in the worst case, when all distributed sources are correlated, one can identify up to $N-1$ sources. In the context of brain-source imaging, these identifiability conditions are usually not very restrictive.

CONVERGENCE

The source-imaging methods exploiting sparsity may be implemented using two types of convex optimization algorithms: interior point methods such as second-order cone programming (SOCP) [9] and proximal splitting methods such as the fast iterative shrinkage-thresholding algorithm (FISTA) [3] or the alternating direction method of multipliers (ADMM) [8]. Both types of solvers are known to converge to the global solution of a convex optimization problem. However, the interior point methods are computationally too expensive to solve large-scale problems as encountered in brain-source imaging, and the simpler and more efficient proximal splitting methods are to be preferred in this case.

To solve the optimization problem associated with the CP decomposition, a wide panel of algorithms, including alternating methods such as alternating least squares, derivative-based techniques such as gradient descent (GD) or Levenberg–Marquardt [14], and direct techniques (see, e.g., [35] and [47] and the references therein) have been used. Even if the local convergence properties hold for most of these methods, there is no guarantee that they will converge to the global minimum because the cost function generally features a large number of local minima. However, in practical situations, it has been observed [30] that good results can be achieved, e.g., by combining a direct method such as the direct algorithm for canonical polyadic decomposition (DIAG) algorithm described in [35] with a derivative-based technique such as GD.

Similar to the tensor decomposition algorithm, there is no guarantee of global convergence for the EM algorithm, which is popular in empirical Bayesian approaches, or for the alternating optimization method employed by the Champagne algorithm.

ADVANTAGES AND DRAWBACKS

Since strengths and weaknesses are often specific to a given source-imaging method and cannot be generalized to other techniques of the same family of approaches, we subsequently focus on seven representative algorithms. Table 2 lists the advantages and drawbacks of each of these methods. On the one hand, the regularized least-squares techniques sLORETA, MCE, and MxNE are simple and computationally efficient, but the source estimates obtained by these algorithms tend to be very focal (for MCE and MxNE) or blurred (for sLORETA). On the other hand, VB-SCCD, STWV-DA, and 4-ExSo-MUSIC, which allow for the identification of spatially extended sources, feature a higher computational complexity. Furthermore, STWV-DA and 4-ExSo-MUSIC have additional requirements such as knowledge of the number of sources or the signal subspace dimension, a certain structure of the data (for STWV-DA), or a sufficiently high number of time samples (for 4-ExSo-MUSIC). While all of these methods require adjusting certain parameters, which are tedious to tune in practice, the main advantage of the Champagne algorithm consists in the fact that there is no parameter to adjust. However, this method also has a high computational complexity and leads to very sparse source estimates.

APPLICATION DOMAINS

Brain-source imaging finds application both in the clinical domain and in cognitive neuroscience. The most frequent clinical application is in epilepsy, where the objective consists in delineating the regions from where interictal spikes or ictal discharges arise [38]. For this purpose, brain-source-imaging methods such as VB-SCCD, STWV-DA, or 4-ExSo-MUSIC, which can identify both the spatial extent and the shape of a small number of distributed sources, are well suited. In cognitive neuroscience, multiple brain structures are often simultaneously activated, particularly when the subjects are asked to perform complex cognitive tasks during the experimental sessions [2]. The source-imaging methods employed for the analysis of these data should thus be able to deal with multiple correlated sources. This is, e.g., the case for VB-SCCD and other regularized least-squares techniques, but not for STWV-DA or 4-ExSo-MUSIC. On the other hand, during simple tasks such as those related to perceptual processes, the analysis of EEG signals of ERPs can also aim at identifying focal sources, in which case methods such as MCE, MxNE, or Champagne are preferred. Finally, there is a rising interest in the analysis of source connectivity [27]. While sLORETA, MCE, MxNE, or Champagne can be employed for this purpose, VB-SCCD, STWV-DA, and 4-ExSo-MUSIC, which enforce identical signals for dipoles belonging to the same patch, would theoretically be less suited, especially for the analysis of very local cortical networks. Nevertheless, at a macroscopic level, these algorithms may be employed to identify cortical networks that characterize the connectivity between distinct brain regions.

RESULTS

In this section, we give the reader an idea of the kind of source-imaging results that can be obtained with different types of algorithms by illustrating and comparing the performance of seven representative algorithms on simulated data for an example of epileptic EEG activity. To do this, we consider two or three quasi-simultaneous active patches and model epileptiform spike-like signals that spread from one brain region to another. The sources are localized using the sLORETA, MCE, MxNE, VB-SCCD, STWV-DA, Champagne, and 4-ExSo-MUSIC algorithms. To quantitatively evaluate the performance of the different methods, we use a measure called the *distance of localization error* (*DLE*) [13], which characterizes the difference between the original and the estimated source configuration. The DLE is averaged over 50 realizations of EEG data with different epileptiform signals and background activity. For detailed descriptions of the data generation process, the implementation of the source-imaging methods, and the evaluation criterion, see [4].

[TABLE 2] THE ADVANTAGES AND DRAWBACKS OF SOURCE-IMAGING ALGORITHMS.

ALGORITHM	ADVANTAGES	DISADVANTAGES
sLORETA [45]	■ SIMPLE TO IMPLEMENT ■ COMPUTATIONALLY EFFICIENT ■ NO LOCALIZATION ERROR FOR A SINGLE DIPOLE SOURCE IN THE ABSENCE OF NOISE ■ WORKS ON A SINGLE TIME SAMPLE	■ BLURRED RESULTS ■ ASSUMES INDEPENDENT DIPOLE SOURCES
MCE [50]	■ SIMPLE ■ CAN LOCALIZE CORRELATED SOURCES ■ WORKS ON A SINGLE TIME SAMPLE ■ LOW COMPUTATIONAL COST FOR SMALL NUMBERS OF TIME SAMPLES	■ VERY FOCAL SOURCE ESTIMATES
VB-SCCD [19]	■ IDENTIFIES SPATIALLY EXTENDED SOURCES ■ FLEXIBLE WITH RESPECT TO THE PATCH SHAPE ■ PERMITS TO LOCALIZE MULTIPLE SIMULTANEOUSLY ACTIVE (AND CORRELATED) PATCHES ■ WORKS ON A SINGLE TIME SAMPLE	■ OVERESTIMATES SIZE OF SMALL PATCHES ■ COMPUTATIONALLY EXPENSIVE ■ SYSTEMATIC ERROR ON ESTIMATED AMPLITUDES
MxNE [44]	■ EXPLOITS THE TEMPORAL STRUCTURE OF THE DATA ■ EXTRACTS SMOOTH TIME SIGNALS ■ SMALL COMPUTATIONAL COST	■ VERY FOCAL SOURCE ESTIMATES
CHAMPAGNE [55]	■ NO PARAMETER TO ADJUST MANUALLY ■ EASY TO IMPLEMENT ■ PERMITS PERFECT SOURCE RECONSTRUCTION UNDER CERTAIN CONDITIONS ■ WORKS ON A SINGLE TIME SAMPLE	■ VERY SPARSE SOURCE ESTIMATES ■ ASSUMES INDEPENDENT DIPOLE SIGNALS ■ HIGH COMPUTATIONAL COMPLEXITY
STWV-DA [5]	■ SEPARATES (CORRELATED) SOURCES ■ IDENTIFIES EXTENDED SOURCES ■ DOES NOT REQUIRE SPATIAL PREWHITENING TO YIELD ACCURATE RESULTS	■ MAKES STRONG ASSUMPTIONS ON DATA STRUCTURE THAT ARE DIFFICULT TO VERIFY IN PRACTICE ■ REQUIRES KNOWLEDGE OF THE NUMBER OF SOURCES TO SEPARATE ■ COMPUTATIONALLY EXPENSIVE FOR LONG DATA LENGTHS
4-ExSo-MUSIC [6]	■ IDENTIFIES EXTENDED SOURCES ■ ROBUST TO GAUSSIAN NOISE	■ HIGH COMPUTATIONAL COMPLEXITY ■ REQUIRES KNOWLEDGE OF THE SIGNAL SUBSPACE DIMENSION ■ REQUIRES A SUFFICIENTLY LARGE NUMBER OF TIME SAMPLES ($>$ 500) TO ESTIMATE THE DATA STATISTICS ■ DIFFICULTIES IN LOCALIZING HIGHLY CORRELATED SOURCES

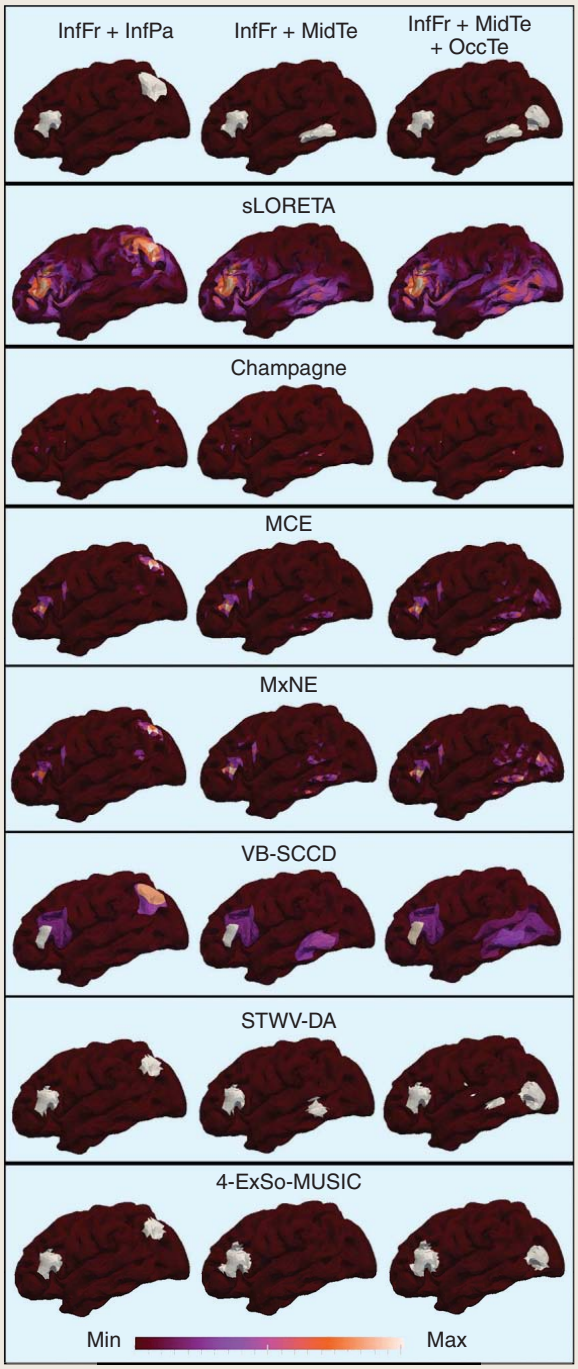
The CPU runtimes that are required for the application of the different source-imaging methods, implemented in MATLAB and run on a machine with a 2.7-GHz processor and 8 GB of random access memory, are listed in Table 3. Note that the runtime of 4-ExSo-MUSIC cannot be compared to that of the other algorithms because this method is partly implemented in C.

We first consider two scenarios with two patches of medium distance composed of a patch in the inferior frontal region (InfFr) combined once with a patch in the inferior parietal region (InfPa) and once with a patch in the middle posterior temporal gyrus (MidTe). The patches are all located on the lateral aspect of the left hemisphere, but the patch MidTe is partly located in a sulcus, leading to weaker surface signals than the patches InfFr and InfPa, which are mostly on a gyral convexity. This has an immediate influence on the performance of all source-imaging algorithms except for Champagne. For the first scenario, the algorithms exhibit high dipole amplitudes for dipoles belonging to each of the true patches. For the second scenario, on the other hand, the weak patch is difficult to make out on the estimated source distribution of sLORETA, slightly more visible on the MCE and MxNE solutions, and completely missing for 4-ExSo-MUSIC. VB-SCCD and STWV-DA both recover the second patch, but with a smaller amplitude in the case of VB-SCCD and a smaller size for STWV-DA. According to the DLE, MCE leads to the best results among the focal source-imaging algorithms while STWV-DA outperforms the other distributed source localization methods.

In the third scenario, we add a patch at the temporo-occipital function (OccTe) to the InfFr and MidTe patches, which further complicates the correct recovery of the active grid dipoles. The best result in terms of the DLE (see Figure 4 and the lower part of Table 4) is achieved by VB-SCCD. Even though this method mostly identifies the brain regions that correspond to the active patches, it does not allow the patches MidTe and OccTe to be distinguished into two separate active sources. STWV-DA, on the other hand, identifies all three patches, even though the extent of the estimated active source region that can be associated to the patch MidTe is too small. However, this method also identifies several spurious source regions of small size located between the patches MidTe and

[TABLE 3] THE AVERAGE CPU RUNTIME OF THE DIFFERENT SOURCE-IMAGING ALGORITHMS FOR THE CONSIDERED THREE-PATCH SCENARIOS.

	sLORETA	VB-SCCD	MxNE	MCE	CHAMPAGNE	STWV-DA	4-ExSo-MUSIC
CPU RUNTIME IN SECONDS	0.18	120	5.9	2.2	233	156	58



[FIG4] The original patches and source reconstructions of different source-imaging algorithms for the scenarios InfFr+InfPa, InfFr+MidTe, and InfFr+MidTe+OccTe.

InfFr. 4-ExSo-MUSIC and Champagne recover only one of the two patches located in the temporal lobe. Similar to VB-SCCD, sLORETA does not allow the patches MidTe and OccTe to be distinguished. This distinction is better performed by MCE and especially by MxNE, which displays three foci of brain activity.

CONCLUSIONS AND PERSPECTIVES

We classified existing source-imaging algorithms based on methodological considerations. Furthermore, we discussed the different techniques, both under theoretical and practical considerations, by addressing questions of identifiability and convergence, advantages and drawbacks of certain algorithms as well as application domains, and by illustrating the performance of representative source-imaging algorithms through a simulation study.

While uniqueness conditions are available for both tensor- and sparsity-based techniques, in the context of brain-source imaging, these conditions are generally only fulfilled for tensor-based approaches, which exploit the concept of distributed sources, whereas the bad conditioning of the lead-field matrix practically prohibits the unique identification of a sparse source distribution. On the other hand, while convex optimization algorithms used for sparse approaches usually converge to the global minimum, such algorithms are not available for tensor decompositions, which suffer from multiple local minima, making it almost impossible to find the global optimum. In practice, despite the limitations concerning identifiability and convergence, both tensor-based and sparse approaches often yield good source reconstruction.

Since the various source localization algorithms have different advantages, drawbacks, and requirements, source-imaging solutions may vary depending on the application. As discussed previously, for each problem, an appropriate source-imaging technique has to be chosen depending on the desired properties of the solution, the characteristics of the algorithm, and the validity of the hypotheses employed by the method. Furthermore, it is advisable to compare the results of different methods for confirmation of the identified source region(s).

To summarize the findings of the simulation study, we can say that sLORETA, Champagne, MCE, and MxNE recover well the source positions, though not their spatial extent as they are conceived for focal sources, while ExSo-MUSIC, STWV-DA, and VB-SCCD also allow for an accurate estimate of the source size. We noticed that most of the methods, except for ExSo-MUSIC and STWV-DA, require prewhitening of the data or a good estimate of the noise covariance matrix (in the case of Champagne) to yield accurate results. On the one hand, this can be explained by the hypothesis of spatially white Gaussian noise made by some approaches, while on the other hand, the prewhitening also leads to a decorrelation of the lead-field vectors and, therefore, to a better conditioning

[TABLE 4] THE DLE (IN CENTIMETERS) OF SOURCE-IMAGING ALGORITHMS FOR DIFFERENT SCENARIOS.

SCENARIO	sLORETA	CHAMPAGNE	MCE	MxNE	VB-SCCD	STWV-DA	ExSo-MUSIC
InfFr+InfPa	2.97	4.03	3.51	3.52	1.23	0.59	0.61
InfFr+MidTe	6.13	4.34	4.40	4.50	1.51	1.17	14.90
InfFr+MidTe+OccTe	5.88	4.83	4.59	4.51	2.54	5.99	4.30

of the lead-field matrix, which consequently facilitates the correct identification of active grid dipoles. Furthermore, the source-imaging algorithms generally have some difficulties in identifying mesial sources located close to the midline as well as multiple quasi-simultaneously active sources. On the whole, for the situations addressed in our simulation study, STWV-DA seems to be the most promising algorithm for distributed source localization, both in terms of robustness and source reconstruction quality. However, more detailed studies are required to confirm the observed performances of the tested algorithms before drawing further conclusions.

Based on these results, we can identify several promising directions for future research. As the VB-SCCD algorithm demonstrates, imposing sparsity in a suitable spatial transform domain may work better than applying sparsity constraints directly to the signal matrix. This type of approach should, thus, be further developed. Another track for future research consists in further exploring different combinations of a priori information, e.g., by merging the successful strategies of different recently established source-imaging approaches, such as tensor- or subspace-based approaches and sparsity. In a similar way, one could integrate the steps of two-step procedures such as STWV-DA into one single step to process all of the available information and constraints at the same time.

ACKNOWLEDGMENTS

Hanna Becker was supported by the Conseil Régional Provence Alpes Côte d'Azur (PACA) and by CNRS France. The work of Pierre Comon was funded by the FP7 European Research Council Programme, DECODA project, under grant ERC-AdG-2013-320594. The work of Rémi Gribonval was funded by the FP7 European Research Council Programme, PLEASE project, under grant ERC-StG-2011-277906. Furthermore, we acknowledge the support of Programme ANR 2010 BLAN 0309 01 (project MULTIMODEL).

AUTHORS

Hanna Becker (hanna.becker@technicolor.com) received her B.S. and M.S. degrees in electrical engineering and information technology from the Ilmenau University of Technology, Germany, in 2010 and 2011, respectively, and her Ph.D. degree from the University of Nice-Sophia Antipolis in 2014. In 2010, she received the European Signal Processing Conference Best Student Paper Award. For her Ph.D. thesis, she received the Research Award of the Société Francaise de Génie Biologique et Médical and the French IEEE Engineering in Medicine and Biology Society section. She is currently working as a postdoctoral researcher at Technicolor R&D France. Her research interests include blind source separation, tensor decompositions, and source localization.

Laurent Albera (laurent.albera@univ-rennes1.fr) received his Ph.D. degree in sciences from the University of Nice Sophia-Antipolis, France, in 2003. He is now an assistant professor at the University of Rennes 1 and is affiliated with the Institut National de la Santé et de la Recherche Médicale (INSERM) research group Laboratoire Traitement du Signal et de l'Image. He was a member of the scientific committee of the University of Rennes 1 from 2008 to 2010. In 2010 he received the Habilitation to Lead Researches degree in sciences from the University of Rennes 1, France. His research interests include human electrophysiological inverse problems based on high-order statistics, sparsity, and multidimensional algebra.

Pierre Comon (pierre.comon@gipsa-lab.fr) has been the research director at CNRS since 1998, now at Gipsa-Lab, Grenoble, France. He was previously employed by various private companies, including the Thales group. His research interests include high-order statistics, blind source separation, statistical signal and array processing, tensor decompositions, multiway factor analysis, and data mining, with applications to health and environment. He was on the editorial boards of several international journals including *IEEE Transactions on Signal Processing*, the EURASIP journal *Signal Processing*, and *IEEE Transactions on Circuits and Systems I*. He is currently an associate editor of *SIAM Journal on Matrix Analysis and Applications*. He is a Senior Member of the IEEE.

Rémi Gribonval (remi.gribonval@irisa.fr) received his Ph.D. degree in applied mathematics from the Université Paris-IX Dauphine in 1999. He is a senior researcher with Inria. His research focuses on mathematical signal processing and machine learning, with an emphasis on sparse approximation, inverse problems, and dictionary learning. He founded the series of international workshops on signal processing with adaptive/sparse representations. In 2011, he was awarded the Blaise Pascal Award in Applied Mathematics and Scientific Engineering from the Société de Mathématiques Appliquées et Industrielles (SMAI) by the French National Academy of Sciences and a starting investigator grant from the European Research Council. He is the leader of the Parsimony and New Algorithms for Audio and Signal Modeling research group on sparse audio processing. He is a Fellow of the IEEE.

Fabrice Wendling (fabrice.wendling@univ-rennes1.fr) received his biomedical engineering diploma in 1989 from the University of Technology of Compiègne, France, his M.S. degree in 1991 from the Georgia Institute of Technology, Atlanta, and his Ph.D. degree in 1996 from the University of Rennes, France. He is the director of research at the Institut National de la Santé et de la Recherche Médicale. He heads the team SESAME: "Epileptogenic Systems: Signals and Models" at the Laboratoire Traitement du Signal et de l'Image, Rennes, France. He has been working on brain signal processing and modeling for more than 20 years in close collaboration with clinicians. In 2012, he received the award Prix Michel Montpetit from the French Academy of Science. He has coauthored approximately 110 peer-reviewed articles.

Isabelle Merlet (isabelle.merlet@univ-rennes1.fr) received her Ph.D. degree in neurosciences from Lyon 1 University France, in 1997. She is a full-time research scientist at the Institut National de la Santé et de la Recherche Médicale. She was with the Epilepsy Department of the Montreal Neurological Institute from 1997 to 2000, the Neurological Hospital of Lyon from 2000 to 2005, and has been with the team SESAME: "Epileptogenic Systems: Signals and Models" of the Signal and Image Processing Laboratory of Rennes since 2005. Since 1993, her work has been devoted to the validation of source localization methods and their application to electroencephalography or magnetoencephalography signals. She is involved in the transfer of these methods to the clinical ground, particularly in the field of epilepsy research.

REFERENCES

[1] L. Albera, A. Ferréol, D. Cosandier-Rimélé, I. Merlet, and F. Wendling, "Brain source localization using a fourth-order deflation scheme," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 2, pp. 490–501, 2008.

[2] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *IEEE Signal Processing Mag.*, vol. 18, no. 6, pp. 14–30, Nov. 2001.

[3] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[4] H. Becker, "Denoising, separation and localization of EEG sources in the context of epilepsy," Ph.D. dissertation, Univ. of Nice-Sophia Antipolis, 2014.

[5] H. Becker, L. Albera, P. Comon, M. Haardt, G. Birot, F. Wendling, M. Gavaret, C. G. Bénar et al., "EEG extended source localization: Tensor-based vs. conventional methods," *NeuroImage*, vol. 96, pp. 143–157, Aug. 2014.

[6] G. Birot, L. Albera, F. Wendling, and I. Merlet, "Localisation of extended brain sources from EEG/MEG: The ExSo-MUSIC approach," *NeuroImage*, vol. 56, no. 1, pp. 102–113, May 2011.

[7] A. Bolstad, B. Van Veen, and R. Nowak, "Space-time event sparse penalization for magneto-/electroencephalography," *NeuroImage*, vol. 46, no. 4, pp. 1066–1081, July 2009.

[8] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[9] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[10] W. Chang, A. Nummenmaa, J. Hsieh, and F. Lin, "Spatially sparse source cluster modeling by compressive neuromagnetic tomography," *NeuroImage*, vol. 53, no. 1, pp. 146–160, Oct. 2010.

[11] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Processing*, vol. 54, no. 12, pp. 4634–4643, 2006.

[12] P. Chevalier, A. Ferréol, and L. Albera, "High-resolution direction finding from higher order statistics: the 2*q*-MUSIC algorithm," *IEEE Trans. Signal Processing*, vol. 54, no. 8, pp. 2986–2997, 2006.

[13] J. Yao and J. P. A. Dewald, "Evaluation of different cortical source localization methods using simulated and experimental EEG data," *NeuroImage*, vol. 25, no. 2, pp. 369–382, Apr. 2005.

[14] P. Comon, L. Luciani, and A. L. F. D. Almeida, "Tensor decompositions, alternating least squares and other tales," *J. Chemometrics*, vol. 23, no. 7–8, pp. 393–405, July–Aug. 2009.

[15] A. M. Dale, A. K. Liu, B. R. Fischl, R. L. Buckner, J. W. Belliveau, J. D. Lewine, and E. Halgren, "Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity," *Neuron*, vol. 26, no. 1, pp. 55–67, 2000.

[16] A. M. Dale and M. I. Sereno, "Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach," *J. Cognit. Neurosci.*, vol. 5, no. 2, pp. 162–176, 1993.

[17] W. Deburchgraeve, P. J. Cherian, M. De Vos, R. M. Swarte, J. H. Blok, G. H. Visser, and P. Govaert, "Neonatal seizure localization using parafac decomposition," *Clin. Neurophysiol.*, vol. 120, no. 10, pp. 1787–1796, Oct. 2009.

[18] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[19] L. Ding, "Reconstructing cortical current density by exploring sparseness in the transform domain," *Phys. Med. Biol.*, vol. 54, no. 9, pp. 2683–2697, May 2009.

[20] L. Ding and B. He, "Sparse source imaging in EEG with accurate field modeling," *Hum. Brain Map.*, vol. 19, no. 9, pp. 1053–1067, Sept. 2008.

[21] K. J. Friston, L. Harrison, J. Daunizeau, S. Kiebel, C. Phillips, N. Trujillo-Barreto, R. Henson, G. Flandin et al., "Multiple sparse priors for the M/EEG inverse problem," *NeuroImage*, vol. 39, no. 1, pp. 1104–1120, 2008.

[22] I. F. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic source imaging with FOCUSS: A recursive weighted minimum norm algorithm," *Electroencephalogr. Clin. Neurophysiol.*, vol. 95, no. 4, pp. 231–251, Oct. 1995.

[23] A. Gramfort, "Mapping, timing and tracking cortical activations with MEG and EEG: Methods and application to human vision," Ph.D. dissertation, Telecom ParisTech, 2009.

[24] A. Gramfort, M. Kowalski, and M. Hämäläinen, "Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods," *Phys. Med. Biol.*, vol. 57, no. 7, pp. 1937–1961, Apr. 2012.

[25] A. Gramfort, D. Strohmeier, J. Haueisen, M. Hämäläinen, and M. Kowalski, "Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations," *NeuroImage*, vol. 70, pp. 410–422, Apr. 2013.

[26] R. Grech, T. Cassar, J. Muscat, K. P. Camilleri, S. G. Fabri, M. Zervakis, P. Xanthopoulos, V. Sakkalis et al., "Review on solving the inverse problem in EEG source analysis," *J. NeuroEng. Rehabil.*, vol. 5, no. 25, Nov. 2008.

[27] J. Gross, J. Kujala, M. Hämäläinen, L. Timmermann, A. Schnitzler, and R. Salmelin, "Dynamic imaging of coherent sources: Studying neural interactions in the human brain," *PNAS*, vol. 98, no. 2, pp. 694–699, 2001.

[28] S. Haufe, V. Nikulin, A. Ziehe, K.-R. Mueller, and G. Nolte, "Combining sparsity and rotational invariance in EEG/MEG source reconstruction," *NeuroImage*, vol. 42, no. 2, pp. 726–738, Aug. 2008.

[29] M.-X. Huang, A. M. Dale, T. Song, E. Halgren, D. L. Harrington, I. Podgorny, J. M. Canive, S. Lewis et al., "Vector-based spatial-temporal minimum l1-norm solution for MEG," *NeuroImage*, vol. 31, no. 3, pp. 1025–1037, July 2006.

[30] A. Karfoul, L. Albera, and P. Comon, "Canonical decomposition of even order Hermitian positive semi-definite arrays," in *Proc. Proceedings of European Signal Processing Conference (EUSIPCO)*, Glasgow, Scotland, 2009, pp. 515–519.

[31] K. Liao, M. Zhu, L. Ding, S. Valette, W. Zhang, and D. Dickens, "Sparse imaging of cortical electrical current densities using wavelet transforms," *Phys. Med. Biol.*, vol. 57, no. 21, pp. 6881–6901, Nov. 2012.

[32] L.-H. Lim and P. Comon, "Blind multilinear identification," *IEEE Trans. Inform. Theory*, vol. 60, no. 2, pp. 1260–1280, 2014.

[33] T. Limpiti, B. D. Van Veen, and R. T. Wakai, "Cortical patch basis model for spatially extended neural activity," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 9, pp. 1740–1754, 2006.

[34] F. Lin, J. W. Belliveau, A. M. Dale, and M. S. Hämäläinen, "Distributed current estimates using cortical orientation constraints," *Hum. Brain Map.*, vol. 27, no. 1, pp. 1–13, Jan. 2006.

[35] X. Luciani and L. Albera, "Canonical polyadic decomposition based on joint eigenvalue decomposition," *Chemometrics Intell. Lab. Syst.*, vol. 132, pp. 152–167, Mar. 2014.

[36] K. Matsuura and Y. Okabe, "A robust reconstruction of sparse biomagnetic sources," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 8, pp. 720–726, Aug. 1997.

[37] J. Mattout, C. Phillips, W. Penny, M. Rugg, and K. Friston, "MEG source localization under multiple constraints: An extended Bayesian framework," *NeuroImage*, vol. 30, no. 3, pp. 753–767, Apr. 2006.

[38] I. Merlet, "Dipole modeling of interictal and ictal EEG and MEG," *Epileptic Disord Special Issue*, vol. 3, pp. 11–36, July 2001.

[39] C. M. Michel, M. M. Murray, G. Lantz, S. Gonzalez, L. Spinelli, and R. G. D. Peralta, "EEG source imaging," *Clin. Neurophysiol.*, vol. 115, no. 10, pp. 2195–2222, Oct. 2004.

[40] J. Möcks, "Decomposing event-related potentials: A new topographic components model," *Biol. Psychol.*, vol. 26, no. 1–3, pp. 199–215, June 1988.

[41] A. Moiseev, J. M. Gaspar, J. A. Schneider, and A. T. Herdman, "Application of multi-source minimum variance beamformers for reconstruction of correlated neural activity," *NeuroImage*, vol. 58, no. 2, pp. 481–496, Sept. 2011.

[42] J. C. Mosher, P. S. Lewis, and R. M. Leahy, "Multiple dipole modeling and localization from spatio-temporal MEG data," *IEEE Trans. Biomed. Eng.*, vol. 39, pp. 541–557, June 1992.

[43] P. L. Nunez and R. Srinivasan, *Electric Fields of the Brain*, 2nd ed. New York: Oxford Univ. Press, 2006.

[44] E. Ou, M. Hämäläinen, and P. Golland, "A distributed spatio-temporal EEG/MEG inverse solver," *NeuroImage*, vol. 44, no. 3, pp. 932–946, Feb. 2009.

[45] R. D. Pascual-Marqui, "Standardized low resolution brain electromagnetic tomography (sLORETA): Technical details," *Methods Findings Exp. Clin. Pharmacol.*, vol. 24D, pp. 5–12, 2002.

[46] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann, "Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain," *Int. J. Psychophysiol.*, vol. 18, no. 1, pp. 49–65, Oct. 1994.

[47] F. Römer and M. Haardt, "A semi-algebraic framework for approximate CP decompositions via simultaneous matrix diagonalization (SECSI)," *Signal Process.*, vol. 93, no. 9, pp. 2462–2473, Sept. 2013.

[48] K. Sekihara, M. Sahani, and S. S. Nagarajan, "Localization bias and spatial resolution of adaptive and non-adaptive spatial filters for MEG source reconstruction," *NeuroImage*, vol. 25, no. 4, pp. 1056–1067, May 2005.

[49] T. S. Tian and Z. Li, "A spatio-temporal solution for the EEG/MEG inverse problem using group penalization methods," *Stat. Interface*, vol. 4, no. 4, pp. 521–533, 2011.

[50] K. Uutela, M. Hämäläinen, and E. Somersalo, "Visualization of magnetoencephalographic data using minimum current estimates," *NeuroImage*, vol. 10, no. 2, pp. 173–180, Aug. 1999.

[51] B. C. Van Veen, W. Van Drongelen, M. Yuchtman, and A. Suzuki, "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 867–880, Sept. 1997.

[52] M. Vega-Hernández, E. Martínes-Montes, J. M. Sánchez-Bornot, A. Lage-Castellanos, and P. A. Valdés-Sosa, "Penalized least squares methods for solving the EEG inverse problem," *Stat. Sinica*, vol. 18, pp. 1535–1551, 2008.

[53] M. Wagner, M. Fuchs, H. A. Wischmann, and R. Drenckhahn, "Smooth reconstruction of cortical sources from EEG and MEG recordings," *NeuroImage*, vol. 3, no. 3, p. S168, 1996.

[54] D. Wipf and S. Nagarajan, "A unified Bayesian framework for MEG/EEG source imaging," *NeuroImage*, vol. 44, no. 3, pp. 947–966, Feb. 2009.

[55] D. Wipf, J. Owen, H. Attias, K. Sekihara, and S. Nagarajan, "Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG," *NeuroImage*, vol. 49, no. 1, pp. 641–655, Jan. 2010.

[56] P. Xu, Y. Tian, H. Chen, and D. Yao, "Lp norm iterative sparse solution for EEG source localization," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 3, pp. 400–409, Mar. 2007.

[SP]

[sp EDUCATION]

Zhilin Zhang

Undergraduate Students Compete in the IEEE Signal Processing Cup: Part 3

The IEEE Signal Processing (SP) Cup is a competition that provides undergraduate students with the opportunity to form teams and work together to solve a challenging and interesting real-world problem using signal processing techniques and methods.

The second IEEE SP Cup, held during the fall and winter of 2014, had the following topic: "Heart Rate Monitoring During Physical Exercise Using Wrist-Type Photoplethysmographic (PPG) Signals" [1]. Participating students were grouped in teams and provided with PPG signals recorded from subjects' wrists during physical exercise. Students were then asked to design an algorithm to estimate the subjects' corresponding heart rates.

BACKGROUND AND MOTIVATION

Wearable health monitoring is a popular, fast-growing area in both industry and academia. Numerous wearable devices for monitoring vital signs have been developed and sold or are selling on the consumer market, such as smart watches and smart wristbands. A key function of these wearable devices is heart rate monitoring using PPG signals recorded from users' wrists. This function can help users control the intensity of their workout according to their heart rate or, alternately, help remote health-care providers monitor the health status of the users.

However, estimating heart rate using wrist-type PPG signals during exercise is a difficult problem. The movements of the users, especially their wrist motion, can result in extremely strong motion artifacts (MAs) in recorded PPG signals, thereby seriously degrading heart rate estimation accuracy (Figure 1). Such interference calls for effective MA removal and heart rate estimation methods.

As a researcher with years of experience in PPG-based heart rate monitoring, I realized that this problem was suitable for the SP Cup for the following reasons:

- The problem can be formulated into a typical signal processing problem from different perspectives. For example, with the available simultaneous acceleration signals, it can be formulated into an adaptive noise cancellation problem. Alternatively, it can be formulated into a single-channel (or multichannel) signal decomposition problem. Therefore, students have the freedom to choose different signal processing techniques to solve this problem, based on their preferences and academic backgrounds.
- Solving this problem requires jointly using multiple signal processing algorithms, fostering collaboration between team members. A successful heart rate monitoring solution consists of many components, such as digital filtering, interference cancellation, power spectrum estimation, signal decomposition, or other advanced algorithms, depending on the kind of signal processing problems formulated by the students. Therefore, team members can divide the problem into a number of subproblems, working on them separately. But they also need to closely collaborate with each other to achieve the optimal performance of their whole solution.

As a result, I submitted a proposal to run an SP Cup on this topic last year, and I was delighted to learn that it had been accepted.

DESCRIPTION OF THE COMPETITION

Approximately 270 students, consisting of 66 teams, registered for this edition of the SP Cup. They came from 21 countries/areas. Ultimately, 49 teams submitted their results by the deadline with qualified submission materials.

The competition had two rounds. In the first round, performance evaluation was mainly based on an average absolute estimation error, defined as the difference between true heart rates and estimated heart rates averaged over the whole test database. Three teams with the best estimation performance were selected to enter the final round. The finalist teams presented their work at ICASSP 2015. A panel of judges attended their presentations and ranked them. The evaluation criteria included 1) the average absolute estimation error (mean and variance), 2) the algorithm novelty, 3) the quality of the report writing, and 4) the oral presentation. Details of the competition procedure can be found in [1] and [3].

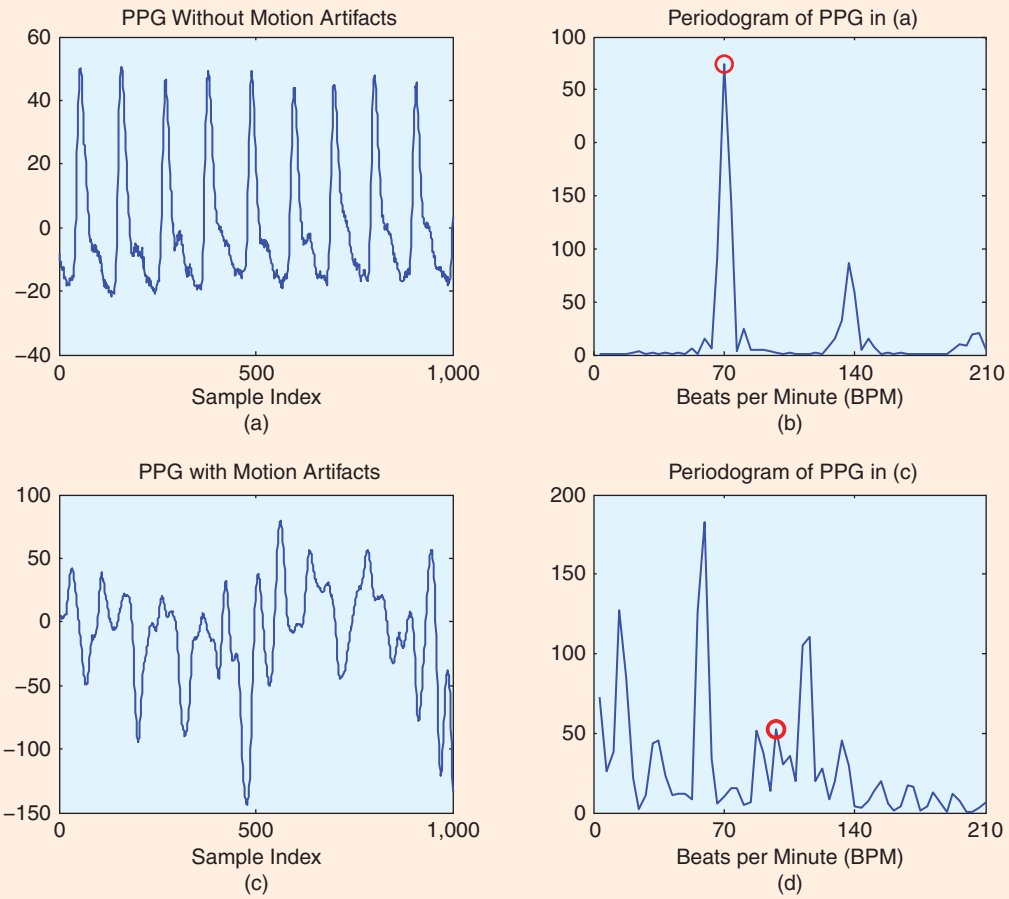
COMPETITION RESULTS

FIRST PLACE: SIGNAL PROCESSING CREW DARMSTADT

The team Signal Processing Crew Darmstadt (Alaa Alameer, Bastian Alt, Christian Sledz, Hauke Radtki, Maximilian Hüttenrauch, Patrick Wenzel, and Tim Schäck) from Technische Universität Darmstadt,

The IEEE SP Cup 2016 will be held at ICASSP 2016 with the competition topic: "Exploring Power Signatures for Location Forensics of Media Recordings." Visit http://www.signalprocessingsociety.org/community/sp-cup/ for more details.

Digital Object Identifier 10.1109/MSP.2015.2462991
Date of publication: 13 October 2015



[FIG1] A comparison between an MA-free PPG signal and an MA-contaminated PPG signal. (a)–(d) shows the MA-free PPG signal and its spectrum, and the MA-contaminated PPG signal and its spectrum, respectively. The spectra are calculated using the periodogram algorithm. The red circles in (b) and (d) indicate the spectral peaks corresponding to the heartbeat. The x-coordinates in (b) and (d) are expressed by BPM for convenience, instead of hertz. The comparison shows the difficulty of identifying heart rate from MA-contaminated PPG signals. (Figure adapted from [2].)

Germany, supervised by Dr.-Ing. Michael Muma, won first place. They adaptively estimated the time-varying transfer functions of each one of the tri-axis acceleration signals that produced artifacts in raw PPG signals. A quality-weighted combination of the outputs of the adaptive filters was then used to form a cleansed signal from which the heart rate was estimated. The method achieved the average absolute estimation error of 3.44 beats/minute (BPM) on the test database.

SECOND PLACE: SUPERSIGNAL

The team Supersignal (Sayeed Shafayet Chowdhury, Rakib Hyder, Anik Khan, Md. Samzid Bin Hafiz, and Zahid Hasan) from Bangladesh University of Engineering and Technology, Bangladesh, supervised by Prof. Mohammad Ariful Haque, took second place. This team proposed a solution, mainly based on adaptive filtering, with carefully designed reference signals from tri-axis accelerometer data and PPG signals.The team obtained the average absolute estimation error of 2.27 BPM on the test database.

THIRD PLACE: SSU

The team SSU (Gyehyun Baek, Minkyu Jung, Hyunil Kang, Jungsub Lee, Baeksan On, and Sunho Kim) from Soongsil University, South Korea, supervised by Prof. Sungbin Im, placed third. To remove motion artifacts in the raw PPG signals, the team proposed a solution based on a multiple-input, single-output (MISO) filter with tri-axis accelerometer data as inputs, where the MISO filter coefficients are estimated using the Wiener filter approach. The solution obtained the average absolute estimation error of 3.26 BPM on the test database.

Figure 2 shows the estimation results of the three teams on set 2 of the test database.

FEEDBACK FROM PARTICIPATING STUDENTS AND SUPERVISORS

I received extensive feedback from the participating students and their supervisors. Due to space constraints in this article, selected samples from the three winning teams are given next.

STUDENTS' FEEDBACK ON THE EXPERIENCE

I think the SP Cup is very helpful in a sense that one can work on close-to-real-world problems.The problem was not as designed as university tasks, and the data were collected from real experiments. Also, it showed that often not the most complex and sophisticated concepts lead to good results, but rather, one starts out with a basic idea and adds bits and pieces to this initial idea.

—Signal Processing Crew Darmstadt

We had to learn a lot of new topics like adaptive filtering, wavelet decomposition, empirical mode decomposition, singular spectrum analysis, etc. in an attempt to solve the problem. We think solving, or even attempting to solve, a real-life signal processing problem helps a lot to build up our interest, as well as understanding, in signal processing.

—Supersignal

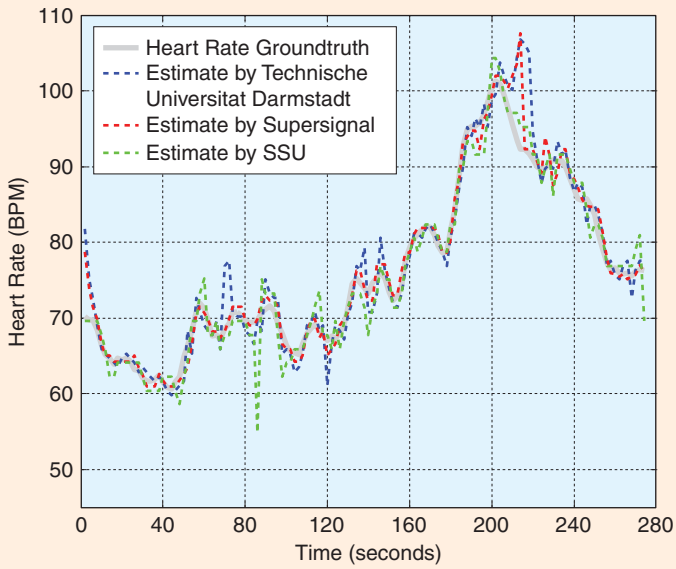
Our team members and I regularly met to study algorithms, including independent component analysis, singular spectrum analysis, sparse signal recovery, Kalman filter, and so on, during meetings. These techniques helped us deal with our goal. We tried to implement some of the techniques into MATLAB codes.

—SSU

STUDENTS' FEEDBACK ON TEAMWORK

Our SP Cup team size was bigger, and we learned how to cooperate together in smaller groups, each group being responsible for a certain problem […] This, of course, required communication skills to understand the work of each group and combine everything together. We met once/twice per week and worked together. The discussions showed us that there were always other possible ways to think of a specific problem, and these discussions helped us identify the best possible way to solve a specific problem.

—Signal Processing Crew Darmstadt



[FIG2] **The heart rate groundtruth of test set 2 and the estimates by the three winning teams. During the data recording, the subject performed various activities, including forearm and upper arm exercise, running, jumping, and push-ups. Thus, the subject's heart rate largely fluctuated. Most of the time, the estimates of the three teams closely followed the true heart rate changes.**

SUPERVISORS' FEEDBACK

The most important ingredient in the recipe (to win the SP Cup) was the motivation of the students. We did our best to keep up a good team spirit in which their creativity could be channeled into new approaches. We had regular meetings, beginning with a kick-off meeting where the students got to know each other. We also provided the team with the communication infrastructure to ensure that the information flow between the students was efficient and transparent.

The SP Cup was extremely useful for students who could apply the concepts they had learned, e.g., in our lectures on adaptive filters and digital signal processing. Also, via the SP Cup, the students were able to implement and fully understand different adaptive filters, such as the least mean squares and Kalman filters. They understood, using real data, the tradeoffs between performance and computational costs.

—Dr.-Ing. M. Muma,
Signal Processing Crew Darmstadt

The students benefitted so much from the SP Cup that it cannot be fully explained in words. It gave them the opportunity to work on a real-world problem. It taught them how to study the literature and link fundamental and advanced digital signal processing algorithms to solve a complex problem. Moreover, it gave them hands-on experience to write technical reports. Although the SP Cup is primarily intended for undergraduate students, we had to explore many advanced algorithms to find an applicable solution. As a result, the insight that I have gained, on the related signal processing algorithms, are helpful to my teaching undergraduate, as well as graduate, courses. Specifically, it helps me to put forward appropriate examples and develop suitable assignments for my students. The benefit is not just limited to teaching courses but also extends to supervising research projects and theses.

—Prof. M.A. Haque, Supersignal,
who also led students
to win the first SP Cup

The SP Cup provides students with insight about how signal processing works in practice. Undergraduate students usually learn many signal processing theories and techniques, but only from textbooks and lectures. Through this competition, they collected a lot of information from various materials such as papers, videos, and discussions, and the combination of the collected information gave good results. It is important for students to learn how to approach problems. Several students even seriously studied sparse signal processing, Kalman filtering, and independent component analysis, which are beyond the scope of the undergraduate signal processing level, to try understanding the state of the art. Introducing the SP Cup in class interests students in signal processing. The explanation of the topic of the SP Cup, related to health monitoring, helps students find applications of signal processing in our daily lives. It is most useful to reduce the distance of the students to the signal processing area.

—Prof. S. Im, SSU

CONCLUSIONS

I am glad to see that many teams proposed effective algorithms to solve this challenge. Nevertheless, it should be noted that there is still much work to do so that the algorithms can work in various scenarios (e.g., different physical activities, different skin color, and different collection devices and PPG sensors). I hope that this edition of the SP Cup managed to raise students' interest in applying their signal processing skills to solve practical problems in wearable health care.

AUTHOR

Zhilin Zhang (zhilinzhang@ieee.org) is a staff research engineer and manager with Samsung Research America, Dallas, Texas. He was a main organizer of the 2015 IEEE Signal Processing Cup and a member of the Bioimaging and Signal Processing Technical Committee of the IEEE Signal Processing Society.

REFERENCE

[1] [Online]. Available: http://www.signalprocessingsociety.org/spcup2015/index.html

[2] Z. Zhang, "Heart rate monitoring from wrist-type photoplethysmographic (PPG) signals during intensive physical exercise," in *Proc. 2014 IEEE Global Conf. Signal and Information Processing (GlobalSIP)*, IEEE, 2014, pp. 698–702.

[3] K.-M. Lam, C. O. S. Sorzano, Z. Zhang, and P. Campisi, "Undergraduate students compete in the IEEE Signal Processing Cup: Part 1," *IEEE Signal Processing Mag.*, vol. 32, no. 4, pp. 123–125, July 2015

[SP]

Digital Object Identifier 10.1109/MSP.2015.2484399

lecture NOTES

Danilo P. Mandic, Sithan Kanna
and Anthony G. Constantinides

On the Intrinsic Relationship Between the Least Mean Square and Kalman Filters

The Kalman filter and the least mean square (LMS) adaptive filter are two of the most popular adaptive estimation algorithms that are often used interchangeably in a number of statistical signal processing applications. They are typically treated as separate entities, with the former as a realization of the optimal Bayesian estimator and the latter as a recursive solution to the optimal Wiener filtering problem. In this lecture note, we consider a system identification framework within which we develop a joint perspective on Kalman filtering and LMS-type algorithms, achieved through analyzing the degrees of freedom necessary for optimal stochastic gradient descent adaptation. This approach permits the introduction of Kalman filters without any notion of Bayesian statistics, which may be beneficial for many communities that do not rely on Bayesian methods [1], [2].

There are several and not immediately patent aspects of common thinking between gradient descent and recursive state-space estimators. Because of their nonobvious or awkward nature, these are often overlooked. Hopefully the framework presented in this article, with the seamless transition between LMS and Kalman filters, will provide a straightforward and unifying platform for understanding the geometry of learning and optimal parameter selection in these approaches. In addition, the material may be useful in lecture courses in statistical signal processing, or indeed, as interesting reading for the intellectually curious and generally knowledgeable reader.

Digital Object Identifier 10.1109/MSP.2015.2461733
Date of publication: 13 October 2015

NOTATION

Lowercase letters are used to denote scalars, e.g., a; boldface letters for vectors, $\boldsymbol{a}$; and boldface uppercase letters for matrices, $\mathbf{A}$. Vectors and matrices are respectively of dimensions $M \times 1$ and $M \times M$. The symbol $(\cdot)^T$ is used for vector and matrix transposition and the subscript k for discrete time index. Symbol $E\{\cdot\}$ represents the statistical expectation operator, $\mathrm{tr}\{\cdot\}$ is the matrix trace operator, and $\|\cdot\|^2$ the l_2 norm.

PROBLEM FORMULATION

Consider a generic system identification setting

$$d_k = \boldsymbol{x}_k^T \boldsymbol{w}_k^o + n_k, \tag{1}$$

where the aim is to estimate the unknown true system parameter vector, $\boldsymbol{w}_k^o$ (optimal weight vector), which characterizes the system in (1) from observations, d_k, corrupted by observation noise, n_k. This parameter vector can be fixed, i.e., $\boldsymbol{w}_k^o = \boldsymbol{w}^o$, or time varying as in (1), while $\boldsymbol{x}_k$ designates a zero-mean input vector and n_k is a zero-mean white Gaussian process with variance $\sigma_n^2 = E\{n_k^2\}$. For simplicity, we assume that all signals are real valued.

To assist a joint discussion of state-space and regression-type models Table 1 lists the terms commonly used across different communities for the variables in the system identification paradigm in (1).

We first start the discussion with a deterministic and time-invariant optimal weight vector, $\boldsymbol{w}_k^o = \boldsymbol{w}^o$, and build up to the general case of a stochastic and time-varying system to give the general Kalman filter.

PERFORMANCE EVALUATION CRITERIA

Consider observations from an unknown deterministic system

$$d_k = \boldsymbol{x}_k^{\mathrm{T}} \boldsymbol{w}^{\mathrm{o}} + n_k. \tag{2}$$

We desire to estimate the true parameter vector $\boldsymbol{w}^{\mathrm{o}}$ recursively, based on the existing weight vector estimate $\boldsymbol{w}_{k-1}$ and the observed and input signals, i.e., $\hat{\boldsymbol{w}}^{\mathrm{o}} = \boldsymbol{w}_k = f(\boldsymbol{w}_{k-1}, d_k, \boldsymbol{x}_k)$. Notice that $\boldsymbol{w}_{k-1}, d_k, \boldsymbol{x}_k$ are related through the output error

$$e_k = d_k - \boldsymbol{x}_k^{\mathrm{T}} \boldsymbol{w}_{k-1}. \tag{3}$$

Performance of statistical learning algorithms is typically evaluated based on the mean square error (MSE) criterion, which is defined as the output error power and is given by

$$\mathrm{MSE} = \xi_k \stackrel{\mathrm{def}}{=} E\{e_k^2\}. \tag{4}$$

Since our goal is to estimate the true system parameters, it is natural to also consider the weight error vector

$$\tilde{\boldsymbol{w}}_k \stackrel{\mathrm{def}}{=} \boldsymbol{w}^{\mathrm{o}} - \boldsymbol{w}_k, \tag{5}$$

and its contribution to the output error, given by

$$e_k = \boldsymbol{x}_k^{\mathrm{T}} \tilde{\boldsymbol{w}}_{k-1} + n_k. \tag{6}$$

[TABLE 1] THE TERMINOLOGY USED IN DIFFERENT COMMUNITIES.

AREA	d_k	$\boldsymbol{x}_k$	$\boldsymbol{w}_k^{\mathrm{O}}$
ADAPTIVE FILTERING	DESIRED SIGNAL	INPUT REGRESSOR	TRUE/OPTIMAL WEIGHTS
KALMAN FILTERING	OBSERVATION	MEASUREMENT	STATE VECTOR
MACHINE LEARNING	TARGET	FEATURES	HYPOTHESIS PARAMETERS

Without loss of generality, here we treat x_k as a deterministic process, although in adaptive filtering convention it is assumed to be a zero-mean stochastic process with covariance matrix $\mathbf{R} = E\{x_k x_k^{\mathrm{T}}\}$. Our assumption conforms with the Kalman filtering literature, where the vector x_k is often deterministic (and sometimes even time invariant). Replacing the output error from (6) into (4) gives

$$\begin{aligned}\xi_k &= E\{(x_k^{\mathrm{T}}\tilde{w}_{k-1} + n_k)^2\} \\ &= x_k^{\mathrm{T}}\mathbf{P}_{k-1}x_k + \sigma_n^2 \qquad (7a)\\ &\stackrel{\mathrm{def}}{=} \xi_{\mathrm{ex},k} + \xi_{\min}, \qquad (7b)\end{aligned}$$

where $\mathbf{P}_{k-1} \stackrel{\mathrm{def}}{=} E\{\tilde{w}_{k-1}\tilde{w}_{k-1}^{\mathrm{T}}\}$ is the symmetric and positive semidefinite weight error covariance matrix, and the noise process n_k is assumed to be statistically independent from all other variables. Therefore, for every recursion step, k, the corresponding MSE denoted by ξ_k comprises two terms: 1) the time-varying excess MSE (EMSE), $\xi_{\mathrm{ex},k}$, which reflects the misalignment between the true and estimated weights (function of the performance of the estimator), and 2) the observation noise power, $\xi_{\min} = \sigma_n^2$, which represents the minimum achievable MSE (for $w_k = w^{\circ}$) and is independent of the performance of the estimator.

Our goal is to evaluate the performance of a learning algorithm in identifying the true system parameters, w°, and a more insightful measure of how closely the estimated weights, w_k, have approached the true weights, w°, is the mean square deviation (MSD), which represents the power of the weight error vector and is given by

$$\begin{aligned}\mathrm{MSD} = J_k &\stackrel{\mathrm{def}}{=} E\{\|\tilde{w}_k\|^2\} = E\{\tilde{w}_k^{\mathrm{T}}\tilde{w}_k\} \\ &= \mathrm{tr}\{\mathbf{P}_k\}. \qquad (8)\end{aligned}$$

Observe that the MSD is related to the MSE in (7a) through the weight error covariance matrix, $\mathbf{P}_k = E\{\tilde{w}_k\tilde{w}_k^{\mathrm{T}}\}$, and thus minimizing MSD also corresponds to minimizing MSE.

OPTIMAL LEARNING GAIN FOR STOCHASTIC GRADIENT ALGORITHMS

The LMS algorithm employs stochastic gradient descent to approximately minimize the MSE in (4) through a recursive estimation of the optimal weight vector, w° in (2), in the form $w_k = w_{k-1} - \mu_k \nabla_w E\{e_k^2\}$. Based on the instantaneous estimate $E\{e_k^2\} \approx e_k^2$, the LMS solution is then given by [3]

$$\begin{aligned}\text{LMS:}\quad w_k &= w_{k-1} + \Delta w_k \\ &= w_{k-1} + \mu_k x_k e_k. \qquad (9)\end{aligned}$$

The parameter μ_k is a possibly time-varying positive step-size that controls the magnitude of the adaptation steps the algorithm takes; for fixed system parameters this can be visualized as a trajectory along the error surface—the MSE plot evaluated against the weight vector, $\xi_k(w)$. Notice that the weight update $\Delta w_k = \mu_k x_k e_k$ has the same direction as the input signal vector, x_k, which makes the LMS sensitive to outliers and noise in data. Figure 1 illustrates the geometry of learning of gradient descent approaches for correlated data (elliptical contours of the error surface)—gradient descent performs locally optimal steps but has no means to follow the globally optimal shortest path to the solution, w°. It is therefore necessary to control both the direction and magnitude of adaptation steps for an algorithm to follow the shortest, optimal path to the global minimum of error surface, $\xi(w^{\circ})$.

The first step toward Kalman filters is to introduce more degrees of freedom by replacing the scalar step-size, μ_k, with a positive definite learning gain matrix, $\mathbf{G}_k$, so as to control both the magnitude and direction of the gradient descent adaptation, and follow the optimal path in Figure 1. In this way, the weight update recursion in (9) now generalizes to

$$w_k = w_{k-1} + \mathbf{G}_k x_k e_k. \qquad (10)$$

Unlike standard gradient-adaptive step-size approaches that minimize the MSE via $\partial\xi_k/\partial\mu_k$ [4], [5], our aim is to introduce an optimal step-size (and learning gain) into the LMS based on the direct minimization of the MSD in (8). For convenience, we consider a general recursive weight estimator

$$w_k = w_{k-1} + \mathbf{g}_k e_k, \qquad (11)$$

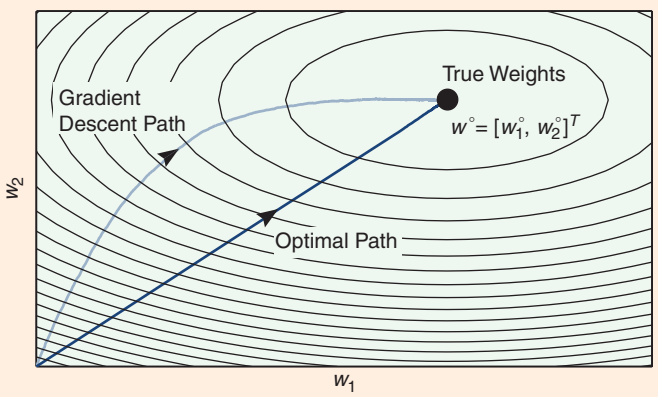
which represents both (9) and (10), where the gain vector

$$\mathbf{g}_k \stackrel{\mathrm{def}}{=} \begin{cases}\mu_k x_k, & \text{for the conventional LMS in (9),}\\ \mathbf{G}_k x_k, & \text{for a general LMS in (10).}\end{cases} \qquad (12)$$

To minimize the MSD, given by $J_k = E\{\|\tilde{w}_k\|^2\} = \mathrm{tr}\{\mathbf{P}_k\}$, we first establish the weight error vector recursion for the general LMS by subtracting w° from both sides of (11) and replacing the output error with $e_k = x_k^{\mathrm{T}}\tilde{w}_{k-1} + n_k$, to give

$$\tilde{w}_k = \tilde{w}_{k-1} - \mathbf{g}_k x_k^{\mathrm{T}}\tilde{w}_{k-1} - \mathbf{g}_k n_k. \qquad (13)$$

The recursion for the weight error covariance matrix, $\mathbf{P}_k$, is then established upon postmultiplying both sides of (13) by their



[FIG1] Mean trajectories of an ensemble of noisy single-realization gradient descent paths for correlated data. The LMS path, produced based on (9), is locally optimal but globally slower converging than the optimal path.

respective transposes and applying the statistical expectation operator $E\{\cdot\}$ to both sides, to yield

$$\begin{aligned}\mathbf{P}_k &= E\{\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathrm{T}}\}\\ &= \mathbf{P}_{k-1} - (\mathbf{P}_{k-1}\boldsymbol{x}_k\mathbf{g}_k^{\mathrm{T}} + \mathbf{g}_k\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1})\\ &\quad + \mathbf{g}_k\mathbf{g}_k^{\mathrm{T}}(\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k + \sigma_n^2). \qquad (14)\end{aligned}$$

Using the well-known matrix trace identities, $\mathrm{tr}\{\mathbf{P}_{k-1}\boldsymbol{x}_k\mathbf{g}_k^{\mathrm{T}}\} = \mathrm{tr}\{\mathbf{g}_k\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}\} = \mathbf{g}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k$ and $\mathrm{tr}\{\mathbf{g}_k\mathbf{g}_k^{\mathrm{T}}\} = \mathbf{g}_k^{\mathrm{T}}\mathbf{g}_k = \|\mathbf{g}_k\|^2$, the MSD evolution, $J_k = \mathrm{tr}\{\mathbf{P}_k\}$, is obtained as

$$\begin{aligned}J_k &= J_{k-1} - 2\mathbf{g}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k\\ &\quad + \|\mathbf{g}_k\|^2(\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k + \sigma_n^2). \qquad (15)\end{aligned}$$

OPTIMAL SCALAR STEP-SIZE FOR LMS

The standard optimal step-size approach to the LMS aims at achieving $e_{k+1|k} = d_k - \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{w}_k = 0$, where the a posteriori error, $e_{k+1|k}$, is obtained using the updated weight vector, $\boldsymbol{w}_k$, and the current input, $\boldsymbol{x}_k$. The solution is known as the normalized LMS (NLMS), given by (for more details, see [6])

$$\text{NLMS:}\quad \boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \frac{1}{\|\boldsymbol{x}_k\|^2}\boldsymbol{x}_k e_k. \qquad (16)$$

The effective LMS-type step-size, $\mu_k = 1/\|\boldsymbol{x}_k\|^2$, is now time varying and data adaptive. In practice, to stabilize the algorithm a small positive step-size ρ_k can be employed, to give $\mu_k = \rho_k/\|\boldsymbol{x}_k\|^2$. The NLMS is therefore conformal with the LMS, whereby the input vector, $\boldsymbol{x}_k$, is normalized by its norm, $\|\boldsymbol{x}_k\|^2$ (input signal power).

To find the optimal scalar step-size for the LMS in (9), which minimizes the MSD, we shall first substitute the gain $\mathbf{g}_k = \mu_k\boldsymbol{x}_k$ into (15), to give the MSD recursion

$$\begin{aligned}J_k &= J_{k-1} - 2\mu_k\underbrace{\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k}_{\xi_{ex,k}}\\ &\quad + \mu_k^2\|\boldsymbol{x}_k\|^2\underbrace{(\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k + \sigma_n^2)}_{\xi_k}. \qquad (17)\end{aligned}$$

The optimal step-size, which minimizes MSD, is then obtained by solving for μ_k in (17) via $\partial J_k/\partial\mu_k = 0$, to yield [7]

$$\begin{aligned}\mu_k &= \frac{1}{\|\boldsymbol{x}_k\|^2}\frac{\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k}{(\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k + \sigma_n^2)}\\ &= \underbrace{\frac{1}{\|\boldsymbol{x}_k\|^2}}_{\text{normalization}}\ \underbrace{\frac{\xi_{\mathrm{ex},k}}{\xi_k}}_{\text{correction}}. \qquad (18)\end{aligned}$$

REMARK 1

In addition to the NLMS-type normalization factor, $1/\|\boldsymbol{x}_k\|^2$, the optimal LMS step-size in (18) includes the correction term, $\xi_{\mathrm{ex},k}/\xi_k < 1$, a ratio of the EMSE, $\xi_{\mathrm{ex,k,}}$ to the overall MSE, ξ_k. A large deviation from the true system weights causes a large $\xi_{\mathrm{ex},k}/\xi_k$ and fast weight adaptation (cf. slow adaptation for a small $\xi_{\mathrm{ex},k}/\xi_k$). This also justifies the use of a small step-size, ρ_k, in practical NLMS algorithms, such as that in "Variants of the LMS."

FROM LMS TO KALMAN FILTER

The optimal LMS step-size in (18) aims to minimize the MSD at every time instant, however, it only controls the magnitude of gradient descent steps (see Figure 1). To find the optimal learning gain that controls simultaneously both the magnitude and direction of the gradient descent in (10), we start again from the MSD recursion [restated from (15)]

$$\begin{aligned}J_k &= J_{k-1} - 2\mathbf{g}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k\\ &\quad + \|\mathbf{g}_k\|^2(\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k + \sigma_n^2).\end{aligned}$$

The optimal learning gain vector, $\mathbf{g}_k$, is then obtained by solving the above MSD for $\mathbf{g}_k$, via $\partial J_k/\partial\mathbf{g}_k = 0$, to give

$$\begin{aligned}\mathbf{g}_k &= \frac{\mathbf{P}_{k-1}}{\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k + \sigma_n^2}\boldsymbol{x}_k = \frac{\mathbf{P}_{k-1}}{\xi_k}\boldsymbol{x}_k\\ &= \mathbf{G}_k\boldsymbol{x}_k. \qquad (19)\end{aligned}$$

This optimal gain vector is precisely the Kalman gain [8], while the gain matrix, $\mathbf{G}_k$, represents a ratio between the weight error covariance, $\mathbf{P}_{k-1}$, and the MSE, ξ_k. A substitution into the update for $\mathbf{P}_k$ in (14) yields a Kalman filter that estimates the time-invariant and deterministic weights, $\boldsymbol{w}^{\mathrm{o}}$, as outlined in Algorithm 1.

REMARK 2

For $\sigma_n^2 = 1$, the Kalman filtering equations in Algorithm 1 are identical to the recursive least squares (RLS) algorithm. In this way, this lecture note complements the classic article by Sayed and Kailath [9] that establishes a relationship between the RLS and the Kalman filter.

SCALAR COVARIANCE UPDATE

An additional insight into our joint perspective on Kalman and LMS algorithms is provided for independent and identically distributed system weight error vectors, whereby the diagonal weight error

VARIANTS OF THE LMS

To illustrate the generality of our results, consider the NLMS and the regularized NLMS (also known as $\varepsilon-$NLMS), given by

$$\text{NLMS:}\quad \boldsymbol{w}_k = \boldsymbol{w}_{k-1} - \rho_k\frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2}e_k, \qquad (S1)$$

$$\varepsilon-\text{NLMS:}\quad \boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2 + \varepsilon_k}e_k, \qquad (S2)$$

where ρ_k is a step-size and ε_k a regularization factor. Based on (17) and (18), the optimal values for ρ_k and ε_k can be found as

$$\rho_k = \frac{\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k}{\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k + \sigma_n^2}, \qquad \varepsilon_k\frac{\|\boldsymbol{x}_k\|^2\sigma_n^2}{\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k}. \qquad (S3)$$

Upon substituting ρ_k and ε_k from (S3) into their respective weight update recursions in (S1) and (S2), we arrive at

$$\boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \frac{\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k}{(\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k + \sigma_n^2)}\frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2}e_k, \qquad (S4)$$

for both the NLMS and $\varepsilon-$NLMS, which is identical to the LMS with the optimal step-size in (18). Therefore, the minimization of the mean square deviation with respect to the parameter: 1) μ_k in the LMS, 2) ρ_k in the NLMS, and 3) ε_k in the ε-NLMS, yields exactly the same algorithm, which is intimately related to the Kalman filter, as shown in Table 2 and indicated by the expression for the Kalman gain, $\mathbf{g}_k$.

Algorithm 1: The Kalman filter for deterministic states.

At each time instant $k > 0$, based on measurements $\{d_k, \boldsymbol{x}_k\}$

1) Compute the optimal learning gain (Kalman gain):

$$\mathbf{g}_k = \mathbf{P}_{k-1}\boldsymbol{x}_k/(\mathbf{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k + \sigma_n^2)$$

2) Update the weight vector estimate:

$$\boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \mathbf{g}_k(d_k - \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{w}_{k-1})$$

3) Update the weight error covariance matrix:

$$\mathbf{P}_k = \mathbf{P}_{k-1} - \mathbf{g}_k\boldsymbol{x}_k^{\mathrm{T}}\mathbf{P}_{k-1}$$

covariance matrix is given by $\mathbf{P}_{k-1} = \sigma_{P,k-1}^2\mathbf{I}$, while the Kalman gain, $\mathbf{g}_k$, in (19) now becomes

$$\mathbf{g}_k = \frac{\sigma_{P,k-1}^2}{\sigma_{P,k-1}^2\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{x}_k + \sigma_n^2}\boldsymbol{x}_k = \frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2 + \varepsilon_k}, \tag{20}$$

where $\varepsilon_k \overset{\text{def}}{=} \sigma_n^2/\sigma_{P,k-1}^2$ denotes the regularization parameter and $\sigma_{P,k-1}^2$ is the estimated weight error vector variance.

REMARK 3

A physical interpretation of the regularization parameter, ε_k, is that it models our confidence level in the current weight estimate, $\boldsymbol{w}_k$, via a ratio of the algorithm-independent minimum MSE, $\xi_{\min} = \sigma_n^2$, and the algorithm-specific weight error variance, $\sigma_{P,k-1}^2$. The more confident we are in current weight estimates, the greater the value of ε_k and the smaller the magnitude of the weight update, $\Delta\boldsymbol{w}_k = \mathbf{g}_k e_k$.

Algorithm 2: A hybrid Kalman-LMS algorithm.

At each time instant $k > 0$, based on measurements $\{d_k, \boldsymbol{x}_k\}$

1) Compute the confidence level (regularisation parameter):

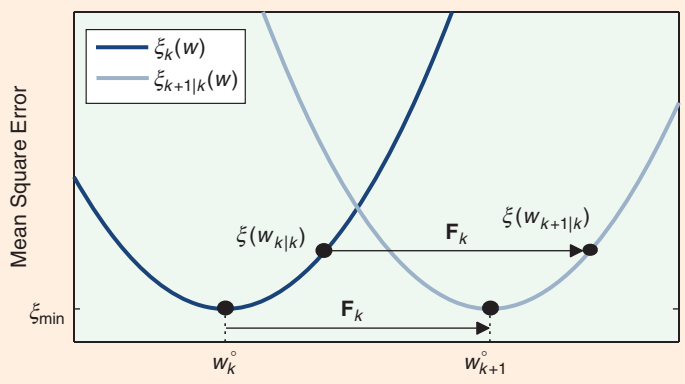
$$\varepsilon_k = \sigma_n^2/\sigma_{P,k-1}^2$$

2) Update the weight vector estimate:

$$\boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2 + \varepsilon_k}(d_k - \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{w}_{k-1})$$

3) Update the weight error variance:

$$\sigma_{P,k}^2 = \sigma_{P,k-1}^2 - \frac{\|\boldsymbol{x}_k\|^2}{M(\|\boldsymbol{x}_k\|^2 + \varepsilon_k)}\sigma_{P,k-1}^2$$



[FIG2] The time-varying state transition in (22a) results in a time-varying MSE surface. For clarity, the figure considers a scalar case without state noise. Within the Kalman filter, the prediction step in (23b) preserves the relative position of $\boldsymbol{w}_{k+1|k}$ with respect to the evolved true state, $\boldsymbol{w}_{k+1}^\circ$.

To complete the derivation, since $\mathbf{P}_k = \sigma_{P,k}^2\mathbf{I}$ and $\mathrm{tr}\{\mathbf{P}_k\} = M\sigma_{P,k}^2$, the MSD recursion in (15) now becomes

$$\sigma_{P,k}^2 = \sigma_{P,k-1}^2 - \frac{\|\boldsymbol{x}_k\|^2}{M(\|\boldsymbol{x}_k\|^2 + \varepsilon_k)}\sigma_{P,k-1}^2. \tag{21}$$

The resulting hybrid "Kalman-LMS" algorithm is given in Algorithm 2.

REMARK 4

The form of the LMS algorithm outlined in Algorithm 2 is identical to the class of generalized normalized gradient descent (GNGD) algorithms in [5] and [10], which update the regularization parameter, ε_k, using stochastic gradient descent. More recently, Algorithm 2 was derived independently in [11] as an approximate probabilistic filter for linear Gaussian data and is referred to as the *probabilistic LMS*.

FROM OPTIMAL LMS TO GENERAL KALMAN FILTER

To complete the joint perspective on the LMS and Kalman filters, we now consider a general case of a time-varying and stochastic weight vector $\boldsymbol{w}_k^\circ$ in (1), to give

$$\boldsymbol{w}_{k+1}^\circ = \mathbf{F}_k\boldsymbol{w}_k^\circ + \boldsymbol{q}_k, \quad \boldsymbol{q}_k \sim \mathcal{N}(0, Q_s), \tag{22a}$$

$$d_k = \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{w}_k^\circ + n_k, \quad n_k \sim \mathcal{N}(0, \sigma_n^2). \tag{22b}$$

The evolution of the true weight vector $\boldsymbol{w}_k^\circ$ is governed by a known state transition matrix, $\mathbf{F}_k$, while the uncertainty in the state transition model is represented by a temporally white state noise vector, $\mathbf{q}_k$, with covariance $\mathbf{Q}_s = E\{\mathbf{q}_k\mathbf{q}_k^{\mathrm{T}}\}$, which is uncorrelated with observation noise n_k. The optimal weight vector evolution in (22a) requires both the update of the current state estimate, $\boldsymbol{w}_{k|k}$, in an LMS-like fashion and the prediction of the next state, $\boldsymbol{w}_{k+1|k}$, as below

$$\boldsymbol{w}_{k|k} = \boldsymbol{w}_{k|k-1} + \mathbf{g}_k(d_k - \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{w}_{k|k-1}), \tag{23a}$$

$$\boldsymbol{w}_{k+1|k} = \mathbf{F}_k\boldsymbol{w}_{k|k}, \tag{23b}$$

where $\mathbf{g}_k$ in (23a) is the Kalman gain. Figure 2 illustrates that, unlike the standard LMS or deterministic Kalman filter in Algorithm 1,

the general Kalman filter in (23a) and (23b) employs its prediction step in (23b) to track the time-varying error surface, a "frame of reference" for optimal adaptation.

The update steps (indicated by the index $k|k$) and the prediction steps (index $k+1|k$) for all the quantities involved are defined below as

$$\begin{aligned}\tilde{\boldsymbol{w}}_{k|k} &\stackrel{\text{def}}{=} \boldsymbol{w}_k^{\text{o}} - \boldsymbol{w}_{k|k},\\ \mathbf{P}_{k|k} &\stackrel{\text{def}}{=} E\{\tilde{\boldsymbol{w}}_{k|k}\tilde{\boldsymbol{w}}_{k|k}^{\text{T}}\},\\ \tilde{\boldsymbol{w}}_{k+1|k} &\stackrel{\text{def}}{=} \boldsymbol{w}_{k+1}^{\text{o}} - \boldsymbol{w}_{k+1|k} = \mathbf{F}_k\tilde{\boldsymbol{w}}_{k|k} + \mathbf{q}_k,\\ \mathbf{P}_{k+1|k} &\stackrel{\text{def}}{=} E\{\tilde{\boldsymbol{w}}_{k+1|k}\tilde{\boldsymbol{w}}_{k+1|k}^{\text{T}}\}\\ &= \mathbf{F}_k\mathbf{P}_{k|k}\mathbf{F}_k^{\text{T}} + \mathbf{Q}_s. \end{aligned} \tag{24}$$

Much like (13)–(17), the Kalman gain is derived based on the weight error vector recursion, obtained by subtracting the optimal time-varying $\boldsymbol{w}_k^{\text{o}}$ from the state update in (23a), to yield

$$\tilde{\boldsymbol{w}}_{k|k} = \tilde{\boldsymbol{w}}_{k|k-1} - \mathbf{g}_k\boldsymbol{x}_k^{\text{T}}\tilde{\boldsymbol{w}}_{k|k-1} - \mathbf{g}_k n_k, \tag{25}$$

so that the evolution of the weight error covariance becomes

$$\begin{aligned}\mathbf{P}_{k|k} &\stackrel{\text{def}}{=} E\{\tilde{\boldsymbol{w}}_{k|k}\tilde{\boldsymbol{w}}_{k|k}^{\text{T}}\}\\ &= \mathbf{P}_{k|k-1} - (\mathbf{P}_{k|k-1}\boldsymbol{x}_k\mathbf{g}_k^{\text{T}} + \mathbf{g}_k\boldsymbol{x}_k^{\text{T}}\mathbf{P}_{k|k-1})\\ &\quad + \mathbf{g}_k\mathbf{g}_k^{\text{T}}(\boldsymbol{x}_k^{\text{T}}\mathbf{P}_{k|k-1}\boldsymbol{x}_k + \sigma_n^2). \end{aligned} \tag{26}$$

Finally, the Kalman gain, $\mathbf{g}_k$, which minimizes the MSD, $J_{k|k} = \text{tr}\{\mathbf{P}_{k|k}\}$, is obtained as [1]

$$\mathbf{g}_k = \frac{\mathbf{P}_{k|k-1}}{\boldsymbol{x}_k^{\text{T}}\mathbf{P}_{k|k-1}\boldsymbol{x}_k + \sigma_n^2}\boldsymbol{x}_k = \mathbf{G}_k\boldsymbol{x}_k. \tag{27}$$

which is conformal with the optimal LMS gain in (19). The general Kalman filter steps are summarized in Algorithm 3.

REMARK 5

Steps 1–3 in Algorithm 3 are identical to the deterministic Kalman filter that was derived starting from the LMS and is described in Algorithm 1. The essential difference is in steps 4 and 5, which cater for the time-varying and stochastic general system weights. Therefore, the fundamental principles of the Kalman filter can be considered through optimal adaptive step-size LMS algorithms.

Algorithm 3: The general Kalman filter.

At each time instant $k > 0$, based on measurements $\{d_k, \boldsymbol{x}_k\}$

1) Compute the optimal learning gain (Kalman gain):

$$\mathbf{g}_k = \mathbf{P}_{k|k-1}\boldsymbol{x}_k/(\boldsymbol{x}_k^{\text{T}}\mathbf{P}_{k|k-1}\boldsymbol{x}_k + \sigma_n^2)$$

2) Update the weight vector estimate:

$$\boldsymbol{w}_{k|k} = \boldsymbol{w}_{k|k-1} + \mathbf{g}_k(d_k - \boldsymbol{x}_k^{\text{T}}\boldsymbol{w}_{k|k-1})$$

3) Update the weight error covariance matrix:

$$\mathbf{P}_k = \mathbf{P}_{k-1} - \mathbf{g}_k\boldsymbol{x}_k^{\text{T}}\mathbf{P}_{k-1}$$

4) Predict the next (posterior) weight vector (state):

$$\boldsymbol{w}_{k+1|k} = \mathbf{F}_k\boldsymbol{w}_{k|k}$$

5) Predict the weight error covariance matrix:

$$\mathbf{P}_{k+1|k} = \mathbf{F}_k\mathbf{P}_{k|k}\mathbf{F}_k^{\text{T}} + \mathbf{Q}_s$$

[TABLE 2] A SUMMARY OF OPTIMAL GAIN VECTORS. THE OPTIMAL STEP-SIZES FOR THE LMS-TYPE ALGORITHMS ARE LINKED TO THE A PRIORI VARIANT OF THE KALMAN GAIN VECTOR, $\mathbf{g}_k$, SINCE $\mathbf{P}_{k|k-1} = \mathbf{P}_{k-1}$ FOR DETERMINISTIC AND TIME-INVARIANT SYSTEM WEIGHT VECTORS.

ALGORITHM	GAIN VECTOR	OPTIMAL GAIN VECTOR
KALMAN FILTER	$\mathbf{g}_k$	$\frac{\mathbf{P}_{k\|k-1}\boldsymbol{x}_k}{\boldsymbol{x}_k^{\text{T}}\mathbf{P}_{k\|k-1}\boldsymbol{x}_k + \sigma_n^2}$
LMS	$\mu_k\boldsymbol{x}_k$	
NLMS	$\rho_k\frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2}$	$\frac{\boldsymbol{x}_k^{\text{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k}{\boldsymbol{x}_k^{\text{T}}\mathbf{P}_{k-1}\boldsymbol{x}_k + \sigma_n^2}\frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2}$
ε − NLMS	$\frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2 + \varepsilon_k}$	which equals $\boldsymbol{x}_k^{\text{T}}\boldsymbol{g}_k\frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2}$

CONCLUSIONS

We have employed "optimal gain" as a mathematical lens to examine conjointly variants of the LMS algorithms and Kalman filters. This perspective enabled us to create a framework for unification of these two main classes of adaptive recursive online estimators. A close examination of the relationship between the two standard performance evaluation measures, the MSE and MSD, allowed us to intuitively link up the geometry of learning of Kalman filters and LMS, within both deterministic and stochastic system identification settings. The Kalman filtering algorithm is then derived in an LMS-type fashion via the optimal learning gain matrix, without resorting to probabilistic approaches [12].

Such a conceptual insight permits seamless migration of ideas from the state-space-based Kalman filters to the LMS adaptive linear filters and vice versa and provides a platform for further developments, practical applications, and nonlinear extensions [13]. It is our hope that this framework of examination of these normally disparate areas will both demystify recursive estimation for educational purposes [14], [15] and further empower practitioners with enhanced intuition and freedom in algorithmic design for the manifold applications.

ACKNOWLEDGMENTS

We thank Prof. Jonathon Chambers, Prof. Ljubisa Stankovic, and Dr. Vladimir Lucic for their insightful comments. This material was finely tuned through our lecture courses on statistical signal processing at Imperial College, notwithstanding the pivotal feedback from our students.

AUTHORS

Danilo P. Mandic (d.mandic@imperial.ac.uk) is a professor of signal processing at Imperial College London, United Kingdom. He has been working in statistical signal processing and specializes in multivariate

state-space estimation and multidimensional adaptive filters. He received the President Award for excellence in postgraduate supervision at Imperial College in 2014. He is a Fellow of the IEEE.

Sithan Kanna (shri.kanagasabapathy08@imperial.ac.uk) is a Ph.D. candidate in statistical signal processing at Imperial College London, United Kingdom, and has been working in state-space estimation and adaptive filtering. He was awarded the Rector's Scholarship at Imperial College.

Anthony G. Constantinides (a.constantinides@imperial.ac.uk) is Emeritus Professor at Imperial College London, United Kingdom. He is a pioneer of signal processing and has been actively involved in research on various aspects of digital signal processing and digital communications for more than 50 years. He is a Fellow of the Royal Academy of Engineering and the 2012 recipient of the IEEE Leon K. Kirchmayer Graduate Teaching Award. He is a Fellow of the IEEE.

REFERENCES

[1] D. J. Simon, *Optimal State Estimation: Kalman, H-Infinity and Non-Linear Approaches*. Hoboken, NJ: Wiley, 2006.

[2] D. Williams, *Probability with Martingales*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[3] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1985.

[4] V. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Trans. Signal Processing*, vol. 41, pp. 2075–2087, June 1993.

[5] S. Douglas, "Generalized gradient adaptive step sizes for stochastic gradient adaptive filters," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, May 1995, vol. 2, pp. 1396–1399.

[6] S. Douglas, "A family of normalized LMS algorithms," *IEEE Signal Processing Lett.*, vol. 1, no. 3, pp. 49–51, 1994.

[7] C. G. Lopes and J. Bermudez, "Evaluation and design of variable step size adaptive algorithms," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2001, vol. 6, pp. 3845–3848.

[8] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ: Prentice Hall, 2000.

[9] A. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Mag.*, vol. 11, pp. 18–60, July 1994.

[10] D. Mandic, "A generalized normalized gradient descent algorithm," *IEEE Signal Processing Lett.*, vol. 11, pp. 115–118, Feb. 2004.

[11] J. Fernandez-Bes, V. Elvira, and S. Van Vaerenbergh, "A probabilistic least-mean-squares filter," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2015, pp. 2199–2203.

[12] R. Faragher, "Understanding the basis of the Kalman filter via a simple and intuitive derivation," *IEEE Signal Processing Mag.*, vol. 29, no. 5, pp. 128–132, 2012.

[13] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures, and Stability*. New York: Wiley, 2001.

[14] J. Humpherys and J. West, "Kalman filtering with Newton's method," *IEEE Control Syst. Mag.*, vol. 30, no. 6, pp. 49–51, 2010.

[15] A. Nehorai and M. Morf, "A mapping result between Wiener theory and Kalman filtering for nonstationary," *IEEE Trans. Automat. Contr.*, vol. 30, no, 2, pp. 175–177, Feb. 1985.

[SP]

Digital Object Identifier 10.1109/MSP.2015.2484400

Dong Yu, Kaisheng Yao, and Yu Zhang

The Computational Network Toolkit

The computational network toolkit (CNTK) is a general-purpose machine-learning tool that supports training and evaluation of arbitrary computational networks (CNs), i.e., machine-learning models that can be described as a series of computational steps. It runs under both Windows and Linux and on both central processing unit (CPU) and Compute Unified Device Architecture (CUDA)-enabled graphics processing unit (GPU) devices. The source code, periodic release builds, documents, and example setups can all be found at http://cntk.codeplex.com.

MOTIVATION

In the past several years, powered by the significant improvements in computing facilities and the great increase of data, deep learning techniques became the new state of the art in many fields such as speech recognition and image classification.

The deep neural network (DNN) is the first successful deep learning model [1]. In DNNs, the combined hidden layers conduct complex nonlinear feature transformation, and the top layer classifies the samples. DNNs jointly optimize the feature transformation and the classification. Though powerful, DNNs do not explicitly exploit structures such as translational variability in images, nor do they explicitly apply operations such as pooling and aggregation to reduce feature variability.

The convolutional neural network (CNN) improves upon the DNN with the explicit modeling of the translational variability by tiling shared local filters across observations to detect the same pattern at different locations [2]. The pattern-detection results are then aggregated through either maximum or average pooling. However, CNNs only deal with translational variability and cannot handle other variations such as horizontal reflections or color intensity differences. Furthermore, CNNs, like DNNs, cannot take advantage of dependencies and correlations between adjacent samples in a sequence.

To address this deficiency, recurrent neural networks (RNNs) were introduced [3]. RNNs can exploit information fed back from hidden and/or output layers in the previous time steps and are often trained with the backpropagation through time algorithm. Unfortunately, simple RNNs are difficult to train and have difficulty modeling long-range dependencies.

The long short-term memory (LSTM)-RNN [3] addresses this difficulty by employing input, output, and forget gates. It significantly improves upon the simple RNN and has been successfully applied in many pattern recognition tasks. However, it may not be optimal for a specific problem at hand since LSTM is a generic model that does not take into account special structures in particular tasks.

To exploit the structure and information inside a particular task, we need to design customized models. Unfortunately, testing customized models is time consuming without proper tools. Typically, we need to design the model, derive the training algorithm, implement them, and run the tests. The majority of the time is spent in the algorithm development and model implementation, which are often error prone and time-consuming. To make things worse, the right model is rarely found on the first trial. We often need to design and evaluate many models with different architectures before settling down with the right one for a specific task. CNTK intends to provide means to reduce the effort required by these two steps and therefore increase the speed of innovation by focusing on problem analysis and model design.

COMPUTATIONAL NETWORKS

If we examine DNNs, CNNs, RNNs, and LSTM-RNNs, we notice that all of these models can be reduced as a series of computational steps. If we know how to compute each step as well as the order in which they are computed, we have an implementation of these models. This observation suggests that we can generalize and treat all these models as special cases of CNs [10].

A CN can be described as a directed graph where each vertex, called a *computation node*, represents a computation, and each edge represents the operator-operant relationship. Note that the order of operands matters for some operations such as matrix multiplication. Leaf nodes in the graph do not have children and are used to represent input values or model parameters that are not result of some computation.

Figure 1 illustrates the correspondence between the NN and the CN representations for a single-hidden-layer neural network with a recurrent loop from the hidden layer to itself. The operations performed by the neural network at time t can be captured by the following three equations:

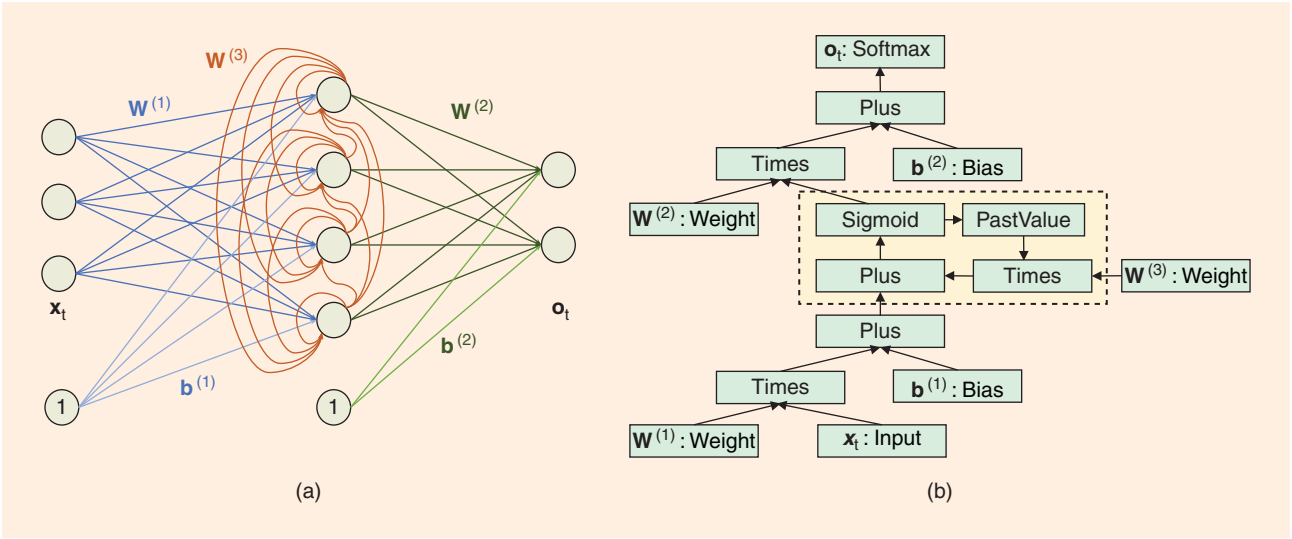
$$\mathbf{p}_t^{(1)} = \mathbf{W}^{(1)}\mathbf{x}_t + \mathbf{b}^{(1)}, \quad (1)$$

$$\mathbf{s}_t = \sigma\left(\mathbf{W}^{(3)}\mathbf{s}_{t-1}^{(1)} + \mathbf{p}_t^{(1)}\right), \quad (2)$$

$$\mathbf{o}_t = f\left(\mathbf{W}^{(2)}\mathbf{s}_t + \mathbf{b}^{(2)}\right), \quad (3)$$

where $\mathbf{W}^{(.)}$ and $\mathbf{b}^{(.)}$ are weights and bias defining the behavior of the NN and that will be learnt during the training phase. Equations (1) and (3) capture the

Digital Object Identifier 10.1109/MSP.2015.2462371
Date of publication: 13 October 2015



[FIG1] NN-CN representation correspondence for a single-hidden-layer neural network with a recurrent loop from the hidden layer to itself. Nodes of the CN are represented using the [node name]:[node type] or [node type] format. (a) Neural network representation. (b) CN representation.

behavior of a single-hidden layer NN. The first equation gives the hidden layer preactivation $\mathbf{p}_t$ for the input $\mathbf{x}_t$. All elements should then go through the sigmoid function $\sigma(p) = 1/[1+\exp(-p)]$ to obtain the hidden layer activity $\mathbf{s}_t$ that is then fed to a softmax function $f(\,)$ in (3) to obtain the output layer activity $\mathbf{o}_t$. Equation (2) accounts for the recurrent loop from the hidden layer to itself and makes the influence of the previous hidden layer activity $\mathbf{s}_{t-1}$ apparent. This flow of computations can be readily represented as a CN. The recurrent loop is depicted in the dash-line box and relies on a past-value node to indicate the dependence of the hidden layer activity on its past values.

Note that CNs can cover a significantly larger variety of and more complicated models than the standard models such as DNNs and RNNs. To support this architectural flexibility, it is necessary to employ special algorithms to evaluate and train CNs.

NODE EVALUATION

Given a CN, we need to evaluate the value of any node in it. However, for general CNs, different network structures may require a different computation order. In the cases similar to DNNs, where there is no recurrent loop in the CN, the value of a node can be computed by following a depth-first search on a directed acyclic graph (DAG), starting from that node. Node evaluation under this condition, also called *forward computation*, is very efficient because many samples can be computed concurrently.

When there are recurrent connections, efficient computation becomes harder. We cannot compute the value of several samples in a sequence as a batch since the value of the next data sample depends on the previous data sample in the same sequence. Two strategies can be exploited to speed up the forward computation in a CN with directed loops.

The first strategy identifies the loops [or strongly connected components (SCCs)] in the CN. If we treat each SCC as a composite node, the CN with loops becomes a DAG and we can use the depth-first-search-based forward computation strategy to evaluate the nodes. For regular nodes, the value of all samples in the sequence can be computed in parallel as a single matrix operation. For the composite nodes corresponding to a loop, the nodes are evaluated sample-by-sample following the time index in the sequence. The computation order of the nodes inside the loop can be easily determined once the values computed in the previous time index are assumed to be known at the current time index: the loop then becomes a DAG if we consider each time step.

The second strategy to speed up forward computation in the recurrent CNs is to process multiple sequences in parallel. To do so, we can organize sequences in a way that the frames with the same time index from different sequences are grouped together so that we can compute them in parallel.

In practice, both strategies can be exploited to speed up the evaluation and training of recurrent neural networks.

MODEL TRAINING

To train a CN, we need to define a scalar training criterion, represent it as a computation node, and insert it into the CN to result in another CN. The model parameters can then be optimized using the stochastic gradient descent (SGD) algorithm. The key here is to efficiently and automatically compute the gradient of the criterion with regard to each model parameter, no matter what CN structure is specified. Here, the well-known reverse automatic gradient computation algorithm [4] can be extended to CNs with recurrent connections similar to that in the forward computation. This algorithm assumes that each node (operator) knows how to compute the gradient of the training criterion with regard to its child nodes (operands) and is independent on other nodes.

EXISTING TOOLKITS

There were already open-source deep learning toolkits available before CNTK. Each toolkit has its own set of features and targeted users.

Theano [5] might be the first general-purpose training tool for deep learning models. It is implemented in Python and supports automatic symbolic differentiation. However, Theano lacks of mechanisms to efficiently train arbitrary recurrent neural networks. In addition, Theano requires a compilation step in the process that can cast difficulties in debugging.

Caffe [6] is probably the most popular open-source deep learning toolkit with a focus on image classification tasks. It is written in C++, with Python and MATLAB interfaces. It provides a flexible language for specifying models. Most recently, LSTM RNN has also been implemented inside Caffe. However, its support to RNNs is still very limited and lacks of functionalities needed in tasks such as speech recognition and natural language processing.

Torch [7] is a popular deep learning toolkit that supports automatic differentiation. The users need to write code in Lua scripting language to use Torch although its back end is written in C/C++ and CUDA.

CNTK shares the same goal as the aforementioned toolkits, i.e., making deep learning model development easier. In particular, CNTK has been designed to efficiently train arbitrary (including bidirectional) recurrent neural networks and sequence-level criteria, which is very important to achieve the state-of-the-art results on tasks such as speech recognition and language modeling and to explore new recurrent architectures. CNTK is written in C++ and is integrated with popular speech recognition toolkits such as Hidden Markov Model Toolkit (HTK) and Kaldi. It also comes with a real-time speech decoder and can be easily plugged into other existing speech-recognition decoders.

COMPUTATIONAL NETWORK TOOLKIT

CNTK is a C++ implementation of CN. It supports both CPU and GPU (CUDA). The toolkit is architected in modules with the core CN functionalities separated from the data readers (for different data format), training algorithms, and network definition.

To run CNTK, we need to prepare a configuration file that specifies a command. If the configuration file is named myconfig.cfg, we run CNTK as

```
cntk.exe
configFile=myconfig.cfg.
```

The top-level commands supported in CNTK are listed in Table 1. Different commands require different information. For example, the train command requires information on the model definition, data reader, and training algorithm.

In CNTK, the model structure can be specified using the network definition language (NDL), which is very similar to the math formula. For example, the line `h = Times(W, x)` indicates that the node h is the product of the weight W and the input x, where `Times` is an NDL function. See "Example in Figure 1 as Expressed in NDL" for reference, where `inputDim` is the input feature dimension, `hiddenDim` is the hidden-layer dimension, and `outputDim` is the output-layer dimension.

Table 2 lists the key functions currently supported in CNTK. Each function is associated with a type of computation node. New computation node types can be added independently following the interface defined in CNTK. NDL supports macros to simplify repeated operations.

Advantageously, CNTK provides several data readers, designed for different file formats and purposes. The UCIFastReader is designed to support space-delimitated text files often used in the UCI data sets. It can use the BinaryReader to cache and speed up. The HTKMLFReader and KaldiReader are used to read speech features and labels in HTK and Kaldi format, respectively. The LMSequenceReader and LUSequenceReader are text file sequence readers for language modeling and language understanding, respectively. If a file format is supported by the above-mentioned readers, we can either convert it to

[TABLE 1] TOP-LEVEL COMMANDS SUPPORTED IN CNTK.

COMMAND	DESCRIPTION
`Train`	TRAIN A MODEL
`Eval`	EVALUATE A MODEL
`CV`	EVALUATE MODELS AT DIFFERENT EPOCHS ON A CROSS-VALIDATION SET
`Adapt`	ADAPT AN ALREADY TRAINED MODEL
`Write`	WRITE THE VALUE OF A NODE TO A FILE
`Edit`	MODIFY AN EXISTING MODEL
`Dumpnode`	DISPLAY THE INFORMATION OF NODE(S)

EXAMPLE IN FIGURE 1 AS EXPRESSED IN NDL

```
ndlCreateNetwork(inputDim, hiddenDim, outputDim) = [
  W1 = Parameter(hiddenDim, inputDim)
  W2 = Parameter(outputDim, hiddenDim)
  W3 = Parameter(hiddenDim, hiddenDim)
  b1 = Parameter(hiddenDim, init=fixedvalue, value=0)
  b2 = Parameter(outputDim, init=fixedvalue, value=0)

  xt = Input(inputDim, tag=feature)
  p1 = Plus(Times(W1, xt), b1)
  pastS = PastValue (outputDim, s1)
  s1 = Sigmoid (Plus (Times (W3, pastS), p1))
  ot = Softmax (Plus (Times (W2, s1), b2))
]
```

[TABLE 2] FUNCTIONS SUPPORTED IN NDL.

CATEGORY	FUNCTIONS
INPUTS AND PARAMETERS	`Input, ImageInput, LookupTable, Parameter, Constant`
VECTOR, MATRIX AND TENSOR OPERATIONS	`ReLU, Sigmoid, Tanh, Log, Cos, Dropout, Negate, Softmax, LogSoftmax, SumElements, RowSlice, RowStack, Scale, Times, DiagTimes, Plus, Minus, ElementTimes, KhatriRaoProduct, Reshape`
TRAINING CRITERIA	`SquareError, CrossEntropyWithSoftmax, ClassificationError, ClassBasedCrossEntropyWithSoftMax, GMMLogLikelihood, CRF`
NORMALIZATION	`Mean, InvStdDev, PerDimMVNorm`
CNN	`Convolution, MaxPooling, AveragePooling`
RNN	`TimeReverse, PastValue, FutureValue`

one of the above file formats or write a new reader to support the new file format.

CNTK trains models using the stochastic gradient descent (SGD) algorithm. In addition, CNTK supports AdaGrad, RMSProp, model averaging, and 1-bit quantization-based data parallelization. CNTK supports manual or automatic learning rate annealing.

SUMMARY AND ADDITIONAL INFORMATION

CNTK allows us to define complex CNs and to train and evaluate the model. It can significantly reduce the effort needed to develop new models and therefore speed up the innovation. Equipped with CNTK, a machine-learning practitioner may define a model with CNs using NDL, evaluate the model, expand, reduce, or transport the model using a model editing language, and reevaluate the modified model. For reference, CNTK has already been successfully used to develop and validate novel models such as the prediction-adaptation-correction-RNN for speech recognition [8] and the sequence encoder-decoder model for text translation [9]

If you are interested in knowing more about CNTK, you can find detailed information on the toolkit in [10] and at http://cntk.codeplex.com.

AUTHORS

Dong Yu (dongyu@microsoft.com) is a principal researcher at Microsoft Research.

Kaisheng Yao (kaisheng.yao@microsoft.com) is a researcher at Microsoft Research.

Yu Zhang (sjtuzy@gmail.com) is a Ph.D. candidate at the Massachusetts Institute of Technology.

REFERENCES

[1] G. E. Hinton, L. Deng, D. Yu et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Mag.*, vol. 29, no. 6, pp. 82–97, 2012.

[2] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*. Boston: MIT Press, 1995, vol. 3361, no. 10, pp. 255–258.

[3] A. Graves, M. Liwicki, S. Fermandez, R. Bertolami, H. Bunke, and J. Schmidhuber, "A novel connectionist system for improved unconstrained handwriting recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 855–868, 2009.

[4] B. Guenter, "Efficient symbolic differentiation for graphics applications," in *Proc. ACM SIGGRAPH 2007, SIGGRAPH'07*, New York, ACM, p. 108.

[5] J. Bergstra et al., "Theano: A CPU and GPU math expression compiler," in *Proc. Python for Scientific Computing Conf. (SciPy)*, June 2010, pp. 3–10.

[6] Y. Jia et al. "Caffe: Convolutional architecture for fast feature embedding," in *in Proc. ACM Int. Conf. Multimedia (ACM MM'14)*, 2014, pp. 675–678.

[7] R. Collobert, K. Kavukcuoglu, and C. Farabet, "Torch7: A MATLAB-like environment for machine learning," in *Proc. Neural Information Processing Systems (NIPS)*, 2011.

[8] Y. Zhang, D. Yu, M. Seltzer, and J. Droppo, "Speech recognition with prediction-adaptation-correction recurrent neural networks," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'15)*.

[9] K. Yao and G. Zweig, "Sequence-to-sequence neural net for grapheme-to-phoneme conversion," in *Proc. INTERSPEECH*, 2015, [Online]. Available: http://arxiv.org/abs/1506.00196

[10] D. Yu, A. Eversole, M. Seltzer, K. Yao, Z. Huang, B. Guenter, O. Kuchaiev, Y. Zhang et al., "An introduction to computational networks and the computational network toolkit", Microsoft Technical Report MSR-TR-2014–112, 2014.

[SP]

Digital Object Identifier 10.1109/MSP.2015.2484398

advertisers INDEX

The Advertisers Index contained in this issue is compiled as a service to our readers and advertisers: the publisher is not liable for errors or omissions although every effort is made to ensure its accuracy. Be sure to let our advertisers know you found them through *IEEE Signal Processing Magazine*.

ADVERTISER	PAGE	URL	PHONE
IEEE Marketing Dept.	3	innovate.ieee.org	
IEEE MDL/Marketing	7	www.ieee.org/go/trymdl	
Mathworks	CVR 4	www.mathworks.com/ltc	+1 508 647 7040
Mini-Circuits	CVR 2, 5, CVR 3	www.minicircuits.com	+1 718 934 4500

advertisers SALES OFFICES

James A. Vick
Sr. Director, Advertising
Phone: +1 212 419 7767;
Fax: +1 212 419 7589
jv.ieeemedia@ieee.org

Marion Delaney
Advertising Sales Director
Phone: +1 415 863 4717;
Fax: +1 415 863 4717
md.ieeemedia@ieee.org

Mark David
Sr. Manager Advertising and Business Development
Phone: +1 732 465 6473
Fax: +1 732 981 1855
m.david@ieee.org

Mindy Belfer
Advertising Sales Coordinator
Phone: +1 732 562 3937
Fax: +1 732 981 1855
m.belfer@ieee.org

Product Advertising

MIDATLANTIC
Lisa Rinaldo
Phone: +1 732 772 0160;
Fax: +1 732 772 0164
lr.ieeemedia@ieee.org
NY, NJ, PA, DE, MD, DC, KY, WV

NEW ENGLAND/SOUTH CENTRAL/ EASTERN CANADA
Jody Estabrook
Phone: +1 774 283 4528;
Fax: +1 774 283 4527
je.ieeemedia@ieee.org
ME, VT, NH, MA, RI, CT, AR, LA, OK, TX
Canada: Quebec, Nova Scotia, Newfoundland, Prince Edward Island, New Brunswick

SOUTHEAST
Cathy Flynn
Phone: +1 770 645 2944;
Fax: +1 770 993 4423
cf.ieeemedia@ieee.org
VA, NC, SC, GA, FL, AL, MS, TN

MIDWEST/CENTRAL CANADA
Dave Jones
Phone: +1 708 442 5633;
Fax: +1 708 442 7620
dj.ieeemedia@ieee.org
IL, IA, KS, MN, MO, NE, ND, SD, WI, OH
Canada: Manitoba, Saskatchewan, Alberta

MIDWEST/ ONTARIO, CANADA
Will Hamilton
Phone: +1 269 381 2156;
Fax: +1 269 381 2556
wh.ieeemedia@ieee.org
IN, MI. Canada: Ontario

WEST COAST/MOUNTAIN STATES/ WESTERN CANADA
Marshall Rubin
Phone: +1 818 888 2407;
Fax: +1 818 888 4907
mr.ieeemedia@ieee.org
AZ, CO, HI, NM, NV, UT, AK, ID, MT, WY, OR, WA, CA. Canada: British Columbia

EUROPE/AFRICA/MIDDLE EAST ASIA/FAR EAST/PACIFIC RIM
Louise Smith
Phone: +44 1875 825 700;
Fax: +44 1875 825 701
les.ieeemedia@ieee.org
Europe, Africa, Middle East
Asia, Far East, Pacific Rim, Australia, New Zealand

Recruitment Advertising

MIDATLANTIC
Lisa Rinaldo
Phone: +1 732 772 0160;
Fax: +1 732 772 0164
lr.ieeemedia@ieee.org
NY, NJ, CT, PA, DE, MD, DC, KY, WV

NEW ENGLAND/EASTERN CANADA
Liza Reich
Phone: +1 212 419 7578;
Fax: +1 212 419 7589
e.reich@ieee.org
ME, VT, NH, MA, RI. Canada: Quebec, Nova Scotia, Prince Edward Island, Newfoundland, New Brunswick

SOUTHEAST
Cathy Flynn
Phone: +1 770 645 2944;
Fax: +1 770 993 4423
cf.ieeemedia@ieee.org
VA, NC, SC, GA, FL, AL, MS, TN

MIDWEST/SOUTH CENTRAL/ CENTRAL CANADA
Darcy Giovingo
Phone: +224 616 3034;
Fax: +1 847 729 4269
dg.ieeemedia@ieee.org;
AR, IL, IN, IA, KS, LA, MI, MN, MO, NE, ND, SD, OH, OK, TX, WI. Canada: Ontario, Manitoba, Saskatchewan, Alberta

WEST COAST/SOUTHWEST/ MOUNTAIN STATES/ASIA
Tim Matteson
Phone: +1 310 836 4064;
Fax: +1 310 836 4067
tm.ieeemedia@ieee.org
AZ, CO, HI, NV, NM, UT, CA, AK, ID, MT, WY, OR, WA. Canada: British Columbia

EUROPE/AFRICA/MIDDLE EAST
Louise Smith
Phone: +44 1875 825 700;
Fax: +44 1875 825 701
les.ieeemedia@ieee.org
Europe, Africa, Middle East

Digital Object Identifier 10.1109/MSP.2015.2388981

dates AHEAD

Please send calendar submissions to:
Dates Ahead, c/o Jessica Barragué
IEEE Signal Processing Magazine
445 Hoes Lane
Piscataway, NJ 08855 USA
e-mail: j.barrague@ieee.org

2015

OCTOBER

IEEE International Workshop on Multimedia Signal Processing (MMSP)
19–21 October, Xiamen, China.
General Chairs: Xiao-Ping Zhang, Oscar C. Au, and Jonathan Li
URL: http://www.mmsp2015.org/

IEEE International Conference on Signal and Image Processing Applications (ICSIPA)
19–21 October, Kuala Lumpur, Malaysia.
General Chair: Syed Khaleel
URL: http://spsocmalaysia.org/icsipa2015/

NOVEMBER

49th Asilomar Conference on Signals, Systems, and Computers (ACSSC)
8–11 November, Pacific Grove, California, United States.
General Chair: Erik G. Larsson
URL: http://www.asilomarsscconf.org/

Seventh IEEE International Workshop on Information Forensics and Security (WIFS)
16–19 November, Rome, Italy.
General Chairs: Patrizio Campisi and Nasir Memon
URL: http://www.wifs2015.org/

DECEMBER

IEEE 6th International Workshop on Computational Advances in Multisensor Adaptive Processing (CAMSAP)
13–16 December, Cancun, Mexico.
URL: http://inspire.rutgers.edu/camsap2015/

Digital Object Identifier 10.1109/MSP.2015.2467193
Date of publication: 13 October 2015

IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)
13–17 December, Scottsdale, Arizona, United States.
URL: http://www.asru2015.org/

International Conference on 3-D Imaging (IC3D)
14–15 December, Liege, Belgium.
Contact: alain@3dstereomedia.eu
URL: http://www.3dstereomedia.eu/ic3d

IEEE Global Conference on Signal and Information Processing (GlobalSIP)
14–16 December, Orlando, Florida, United States.
General Chairs: José M.F. Moura and Dapeng Oliver Wu
URL: http://2015.ieeeglobalsip.org/

IEEE Second World Forum on Internet of Things (WF-IoT)
14–16 December, Milan, Italy.
Conference Chair: Latif Ladid
URL: http://sites.ieee.org/wf-iot/

Asia-Pacific Signal and Information Processing Association Annial Summit and Conference (APSIPA)
16–19 December, Hong Kong.
Honorary General Chair: Wan-Chi Siu
General Cochairs: Kenneth Lam, Helen Meng, and Oscar Au
URL: http://www.apsipa2015.org/

2016

MARCH

41st IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)
21–25 March, Shanghai, China.
General Chairs: Zhi Ding, Zhi-Quan Luo, and Wenjun Zhang
URL: http://icassp2016.org

Data Compression Conference (DCC)
29 March–1 April, Snowbird, Utah, United States.
URL: http://www.cs.brandeis.edu/~dcc/Dates.html

APRIL

15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)
11–14 April, Vienna, Austria.
General Chair: Guoliang Xing
URL: http://ipsn.acm.org/2016/

IEEE International Symposium on Biomedical Imaging (ISBI)
13–16 April, Prague, Czech Republic.
General Chairs: Jan Kybic and Milan Sonka
URL: http://biomedicalimaging.org/2016/

JUNE

IEEE Workshop on Statistial Signal Processing (SSP)
26–29 June, Palma de Mallorca, Spain.
General Chairs: Antonio Artés-Rodríguez and Joaquín Miguez
URL: http://ssp2016.tsc.uc3m.es/

JULY

IEEE Ninth IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)
10–13 July, Rio de Janeiro, Brazil.
General Chairs: Rodrigo C. de Lamare and Martin Haardt
URL: http://delamare.cetuc.puc-rio.br/sam2016/index.html

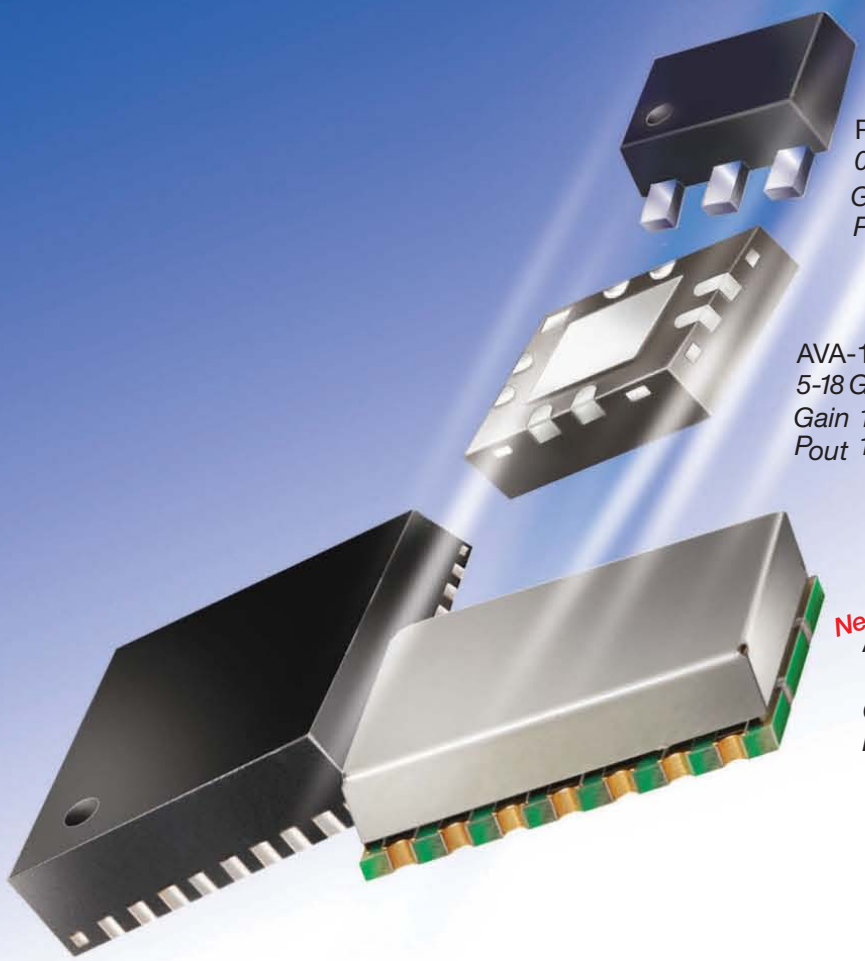
IEEE International Conference on Multimedia and Expo (ICME)
11–15 July, Seattle, Washington, United States.
General Chairs: Tsuhan Chen, Ming-Ting Sun, and Cha Zhang
URL: http://www.icme2016.org/

SP

The 2015 Annual Index for
IEEE Signal Processing Magazine (SPM)
is available in IEEE *Xplore*.
Visit http://ieeexplore.ieee.org, select the "Browse" tab, and then navigate to "Journals and Magazines" in the drop-down menu to find *SPM's* current issue.

50 MHz to 26.5 GHz

MICROWAVE MMIC AMPLIFIERS



PHA-1+
0.05-6 GHz
$1.99 ea. (qty. 20)
Gain 13.5 dB
P_{out} 22 dBm

AVA-183A+
5-18 GHz
$6.95 ea. (qty. 10)
Gain 14.0 dB
P_{out} 19 dBm

New
AVM-273HPK+
13-26.5 GHz
$36.90 ea. (qty. 10)
Gain 13.0 dB
P_{out} 27 dBm

Mini-Circuits' ***New*** ***AVM-273HPK+*** wideband microwave MMIC amplifier supports applications from 13 to 26.5 GHz with up to 0.5W output power, 13 dB gain, ±1 dB gain flatness and 58 dB isolation. The amplifier comes supplied with a voltage sequencing and DC control module providing reverse voltage protection in one tiny package to simplify your circuit design. This model is an ideal buffer amplifier for P2P radios, military EW and radar, DBS, VSAT and more!

The AVA-183A+ delivers 14 dB Gain with excellent gain flatness (±1.0 dB) from 5 to 18 GHz, 38 dB isolation, and 19 dBm power handling. It is unconditionally stable and an ideal LO driver amplifier. Internal DC blocks, bias tee, and microwave coupling capacitor simplify external circuits, minimizing your design time.

The PHA-1+ + uses E-PHEMT technology to offer ultra-high dynamic range, low noise, and excellent IP3 performance, making it ideal for LTE and TD-SCDMA. Good input and output return loss across almost 7 octaves extend its use to CATV, wireless LANs, and base station infrastructure.

We've got you covered! Visit minicircuits.com for full specs, performance curves, and free data! These models are in stock and ready to ship today!

RoHS compliant

FREE X-Parameters-Based
Non-Linear Simulation Models for ADS
Modelithics®
http://www.modelithics.com/mvp/Mini-Circuits.asp

Mini-Circuits®

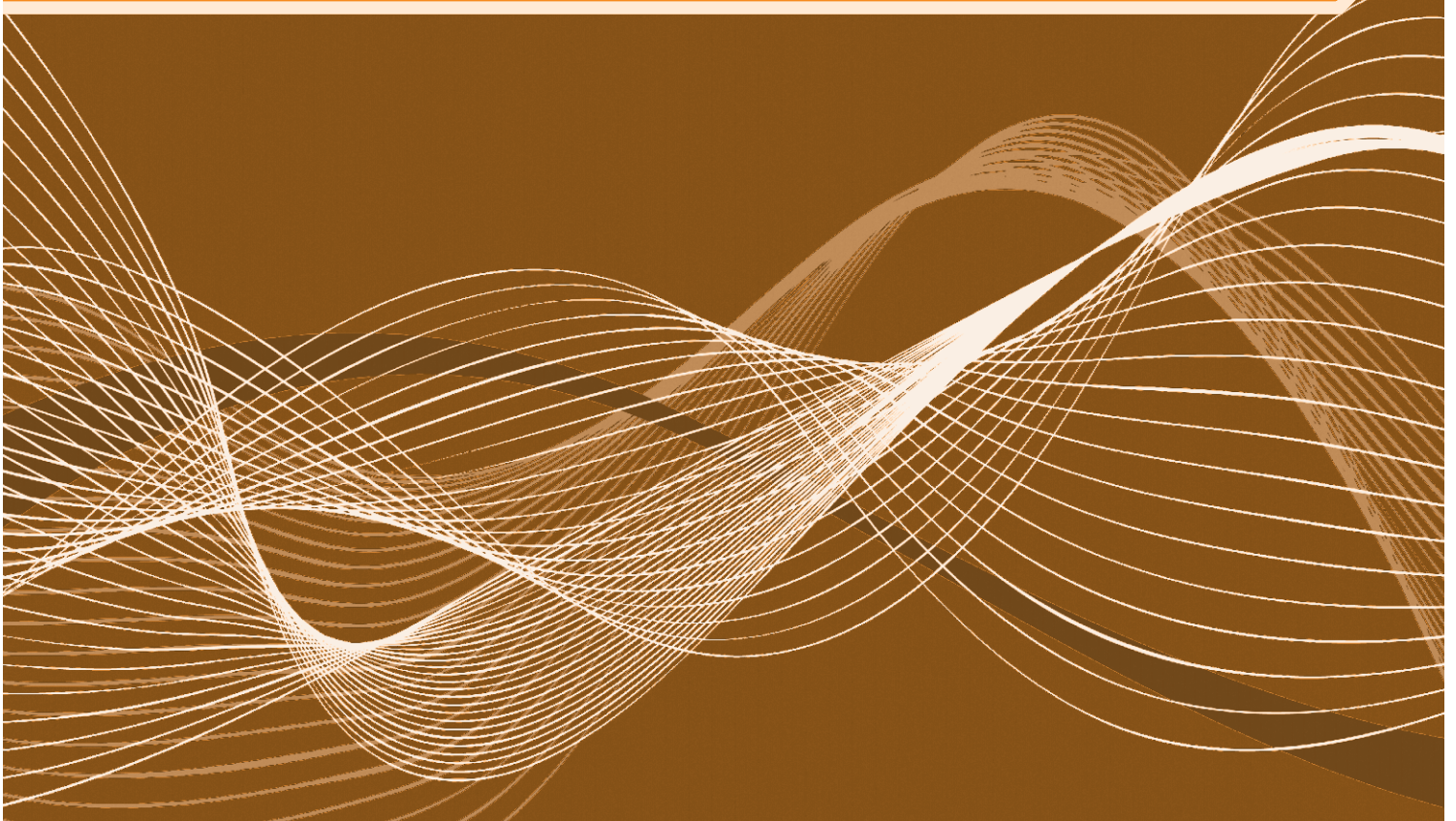
www.minicircuits.com P.O. Box 350166, Brooklyn, NY 11235-0003 (718) 934-4500 sales@minicircuits.com

478 rev Q

IEEE SIGNAL PROCESSING SOCIETY

CONTENT GAZETTE

[ISSN 2167-5023]



NOVEMBER 2015

IEEE Signal Processing Society

IEEE

GlobalSIP 2016 – Washington DC

Call for Symposium Proposals

The fourth IEEE Global Conference on Signal and Information Processing (GlobalSIP) will be held in Greater Washington, DC, USA on December 7–9, 2016. GlobalSIP has rapidly assumed flagship status within the IEEE Signal Processing Society. It focuses on signal and information processing and up-and-coming signal processing themes. The conference aims to feature world-class speakers, tutorials, exhibits, and oral and poster sessions. GlobalSIP is comprised of co-located symposia selected based on responses to the Call for Symposium Proposals. Topics include but are not limited to:

* Signal processing in communications and networks, including green communications, and signal processing in massive MIMO and millimeter-wave technology
* Big data signal processing
* Signal processing in information secrecy, privacy and security
* Information forensics
* Image and video processing
* Selected topics in speech and language processing
* Signal processing in finance
* Signal processing in energy and power systems
* Signal processing in genomics and bioengineering
* Selected topics in statistical signal processing
* Graph-theoretic signal processing
* Machine learning
* Compressed sensing, sparsity analysis and applications
* Music and multimedia transmission, indexing and retrieval, and playback challenges
* Real-time DSP implementations
* Other novel and significant applications of selected areas of signal processing

Symposium proposals should have the following information: title of the symposium; length of the symposium; paper length requirement; names, addresses, and a short CV (up to 250 words) of the organizers including the general organizers and the technical chairs of the symposium; a two-page description of the technical issues that the symposium will address, including timeliness and relevance to the signal processing community; names of (potential) participants on the technical program committee; names of the invited speakers; and a draft Call for Papers. Please pack everything together in pdf format, and email the proposals to conference TPC Chairs at tpc-chairs@2016.ieeeglobalsip.org. More detailed information can be found in the GlobalSIP Symposium Proposal Preparation Guide, available at http://2016.ieeeglobalsip.org/.

Conference Timeline:

* **February 5, 2016: Symposium proposals due**
* February 19, 2016: Symposium selection decision made
* February 29, 2016: Call for Papers for accepted symposia will be publicized
* June 5, 2016: Paper submission due
* August 5, 2016: Final Acceptance decisions notifications sent to all authors
* September 5, 2016: Camera-ready papers due.

IEEE TRANSACTIONS ON
SIGNAL PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

PubMed | MEDLINE U.S. National Library of Medicine

SEPTEMBER 15, 2015 VOLUME 63 NUMBER 18 ITPRED (ISSN 1053-587X)

REGULAR PAPERS

IEEE

IEEE TRANSACTIONS ON
SIGNAL PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

PubMed MEDLINE U.S. National Library of Medicine

OCTOBER 1, 2015 VOLUME 63 NUMBER 19 ITPRED (ISSN 1053-587X)

REGULAR PAPERS

IEEE

IEEE TRANSACTIONS ON

SIGNAL PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

PubMed MEDLINE U.S. National Library of Medicine

OCTOBER 15, 2015 VOLUME 63 NUMBER 20 ITPRED (ISSN 1053-587X)

REGULAR PAPERS

IEEE

IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS

Now accepting paper submissions

The new *IEEE Transactions on Signal and Information Processing over Networks* publishes high-quality papers that extend the classical notions of processing of signals defined over vector spaces (e.g. time and space) to processing of signals and information (data) defined over networks, potentially dynamically varying. In signal processing over networks, the topology of the network may define structural relationships in the data, or may constrain processing of the data. Topics of interest include, but are not limited to the following:

Adaptation, Detection, Estimation, and Learning
- Distributed detection and estimation
- Distributed adaptation over networks
- Distributed learning over networks
- Distributed target tracking
- Bayesian learning; Bayesian signal processing
- Sequential learning over networks
- Decision making over networks
- Distributed dictionary learning
- Distributed game theoretic strategies
- Distributed information processing
- Graphical and kernel methods
- Consensus over network systems
- Optimization over network systems

Communications, Networking, and Sensing
- Distributed monitoring and sensing
- Signal processing for distributed communications and networking
- Signal processing for cooperative networking
- Signal processing for network security
- Optimal network signal processing and resource allocation

Modeling and Analysis
- Performance and bounds of methods
- Robustness and vulnerability
- Network modeling and identification

Modeling and Analysis (cont.)
- Simulations of networked information processing systems
- Social learning
- Bio-inspired network signal processing
- Epidemics and diffusion in populations

Imaging and Media Applications
- Image and video processing over networks
- Media cloud computing and communication
- Multimedia streaming and transport
- Social media computing and networking
- Signal processing for cyber-physical systems
- Wireless/mobile multimedia

Data Analysis
- Processing, analysis, and visualization of big data
- Signal and information processing for crowd computing
- Signal and information processing for the Internet of Things
- Emergence of behavior

Emerging topics and applications
- Emerging topics
- Applications in life sciences, ecology, energy, social networks, economic networks, finance, social sciences, smart grids, wireless health, robotics, transportation, and other areas of science and engineering

Editor-in-Chief: Petar M. Djurić, Stony Brook University (USA)

To submit a paper, go to: https://mc.manuscriptcentral.com/tsipn-ieee

IEEE COMMUNICATIONS SOCIETY

IEEE Signal Processing Society

IEEE COMPUTER SOCIETY

http://www.signalprocessingsociety.org/

NEW PUBLICATION:

Transactions on Signal and Information Processing over Networks (T-SIPN)*

http://www.signalprocessingsociety.org/publications/periodicals/tsipn/

>>We are accepting paper submissions: please submit a manuscript here<<

There has been an explosion of research in network systems of various types, including physical, engineered, biological and social systems. Its aim is to find answers to fundamental questions about the systems and with them be able to understand, predict, and control them better. To that end, a core area of work is signal and information processing over networks.

Network systems represent a growing research field encompassing numerous disciplines in science and engineering. Their complexity is reflected in the diversity and the interconnectivity of their elements, which have the capacity to adapt and learn from experience. Applications of network systems are wide and include communications (wireless sensor networks, peer-to-peer networks, pervasive mobile networks, the Internet of Things), the electric power grid, biology, the Internet, the stock market, ecology, and in animal and human societies.

The Transactions on Signal and Information Processing over Networks (T-SIPN) publishes timely peer-reviewed technical articles on advances in the theory, methods, and algorithms for signal and information processing, inference, and learning in network systems. The following core topics define the scope of the Transaction:

Adaptation, Detection, Estimation, and Learning (ADEL)

- Distributed detection and estimation (ADEL-DDE)
- Distributed adaptation over networks (ADEL-DAN)
- Distributed learning over networks (ADEL-DLN)
- Distributed target tracking (ADEL-DTT)
- Bayesian learning; Bayesian signal processing (ADEL-BLSP)
- Sequential learning over networks (ADEL-SLN)
- Decision making over networks (ADEL-DMN)
- Distributed dictionary learning (ADEL-DDL)
- Distributed game theoretic strategies (ADEL-DGTS)
- Distributed information processing (ADEL-DIP)
- Graphical and kernel methods (ADEL-GKM)
- Consensus over network systems (ADEL-CNS)
- Optimization over network systems (ADEL-ONS)

Communications, Networking, and Sensing (CNS)

- Distributed monitoring and sensing (CNS-DMS)
- Signal processing for distributed communications and networking (CNS-SPDCN)
- Signal processing for cooperative networking (CNS-SPCN)
- Signal processing for network security (CNS-SPNS)
- Optimal network signal processing and resource allocation (CNS-NSPRA)

(continued on next page)

Modeling and Analysis (MA)

- Performance and bounds of methods (MA-PBM)
- Robustness and vulnerability (MA-RV)
- Network modeling and identification (MA-NMI)
- Simulations of networked information processing systems (MA-SNIPS)
- Social learning (MA-SL)
- Bio-inspired network signal processing (MA-BNSP)
- Epidemics and diffusion in populations (MA-EDP)

Imaging and Media Applications (IMA)

- Image and video processing over networks (IMA-IVPN)
- Media cloud computing and communication (IMA-MCCC)
- Multimedia streaming and transport (IMA-MST)
- Social media computing and networking (IMA-SMCN)
- Signal processing for cyber-physical systems (IMA-SPCPS)
- Wireless/mobile multimedia (IMA-WMM)

Data Analysis (DA)

- Processing, analysis, and visualization of big data (DA-BD)
- Signal and information processing for crowd computing (DA-CC)
- Signal and information processing for the Internet of Things (DA-IOT)
- Emergence of behavior (DA-EB)

Emerging topics and applications (ETA)

- Emerging topics (ETA-ET)
- Applications in life sciences, ecology, energy, social networks, economic networks, finance, social sciences etc. smart grids, wireless health, robotics, transportation, and other areas of science and engineering (ETA-APP)

>>We are accepting paper submissions: please submit a manuscript here<<

***T-SIPN is co-sponsored by the Signal Processing, Communications and Computer societies**

IEEE/ACM TRANSACTIONS ON

AUDIO, SPEECH, AND LANGUAGE PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society | acm Association for Computing Machinery | www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

PubMed | MEDLINE U.S. National Library of Medicine

SEPTEMBER 2015 VOLUME 23 NUMBER 9 ITASFA (ISSN 2329-9290)

IEEE

IEEE/ACM TRANSACTIONS ON

AUDIO, SPEECH, AND LANGUAGE PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society | acm Association for Computing Machinery | www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

PubMed | MEDLINE U.S. National Library of Medicine

OCTOBER 2015 | VOLUME 23 | NUMBER 10 | ITASFA | (ISSN 2329-9290)

IEEE

The Ninth IEEE Sensor Array and Multichannel Signal Processing Workshop

10th-13th July 2016,Rio de Janeiro, Brazil

IEEE Signal Processing Society

Call for Papers

General Chairs
Rodrigo C. de Lamare,
PUC-Rio, Brazil and University of York, United Kingdom

Martin Haardt,
TU Ilmenau, Germany

Technical Chairs
Aleksandar Dogandzic,
Iowa State University, USA

Vítor Nascimento,
University of São Paulo, Brazil

Special Sessions Chair
Cédric Richard,
University of Nice, France

Publicity Chair
Maria Sabrina Greco,
Universityof Pisa, Italy

Important Dates

Special Session Proposals
29th January , 2016

Submission of Papers
26th February, 2016

Notification of Acceptance
29th April , 2016

Final Manuscript Submission
16th May, 2016

Advance Registration
16th May, 2016

Technical Program
The SAM Workshop is an important IEEE Signal Processing Society event dedicated to sensor array and multichannel signal processing. The organizing committee invites the international community to contribute with state-of-the-art developments in the field. SAM 2016 will feature plenary talks by leading researchers in the field as well as poster and oral sessions with presentations by the participants.

Welcome to Rio de Janeiro! – The workshop will be held at the Pontifical Catholic University of Rio de Janeiro, located in Gávea, in a superb area surrounded by beaches, mountains and the Tijuca National Forest, the world's largest urban forest. Rio de Janeiro is a world renowned city for its culture, beautiful landscapes, numerous tourist attractions and international cuisine. The workshop will take place during the first half of July about a month before the 2016 Summer Olympic Games when Rio will offer plenty of cultural activities and festivities, which will make SAM 2016 a memorable experience.

Research Areas
Authors are invited to submit contributions in the following areas:
- Adaptive beamforming
- Array processing for biomedical applications
- Array processing for communications
- Blind source separation and channel identification
- Computational and optimization techniques
- Compressive sensing and sparsity-based signal processing
- Detection and estimation
- Direction-of-arrival estimation
- Distributed and adaptive signal processing
- Intelligentsystems and knowledge-based signal processing
- Microphone and loudspeaker array applications
- MIMO radar
- Multi-antenna systems: multiuser MMO, massive MIMO and space-time coding
- Multi-channel imaging and hyperspectral processing
- Multi-sensor processing for smart grid and energy
- Non-Gaussian, nonlinear, and non-stationary models
- Performance evaluations with experimental data
- Radar and sonar array processing
- Sensor networks
- Source Localization, Classification and Tracking
- Synthetic aperture techniques
- Space-time adaptive processing
- Statistical modelling for sensor arrays
- Waveform diverse sensors and systems

Submission of papers – Full-length four-page papers will be accepted only electronically.

Special session proposals – They should be submitted by e-mail to the Technical Program Chairs and the Special Sessions Chair and include a topical title, rationale, session outline, contact information, and list of invited speakers.

IEEE TRANSACTIONS ON

IMAGE PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

PubMed MEDLINE U.S. National Library of Medicine

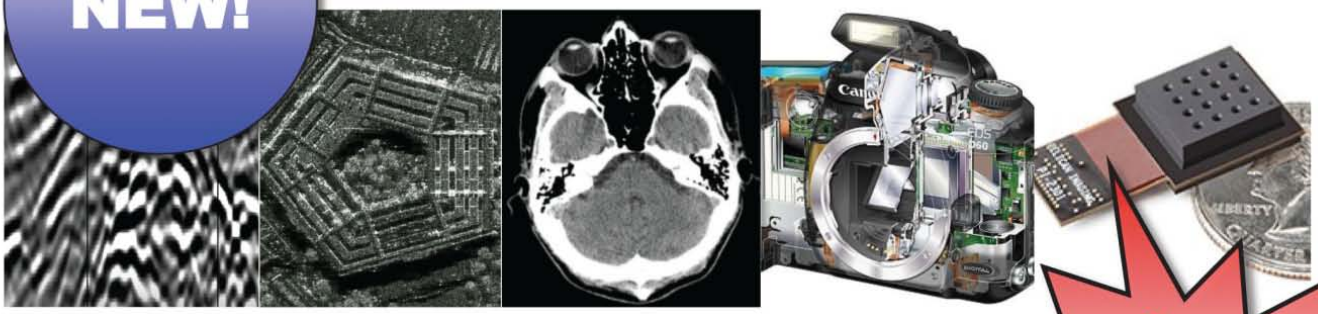
OCTOBER 2015 VOLUME 24 NUMBER 10 IIPRE4 (ISSN 1057-7149)

PAPERS

IEEE

NEW!

IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING



30% FASTER TURN AROUND THAN TIP!

Editor-in-Chief

W. Clem Karl
Boston University

Technical Committee

Charles Bouman
Eric Miller
Peter Corcoran
Jong Chul Ye
Dave Brady
William Freeman

The IEEE Transactions on Computational Imaging publishes research results where computation plays an integral role in the image formation process. All areas of computational imaging are appropriate, ranging from the principles and theory of computational imaging, to modeling paradigms for computational imaging, to image formation methods, to the latest innovative computational imaging system designs. Topics of interest include, but are not limited to the following:

Computational Imaging Methods and Models

- Coded image sensing
- Compressed sensing
- Sparse and low-rank models
- Learning-based models, dictionary methods
- Graphical image models
- Perceptual models

Computational Image Formation

- Sparsity-based reconstruction
- Statistically-based inversion methods
- Multi-image and sensor fusion
- Optimization-based methods; proximal iterative methods, ADMM

Computational Photography

- Non-classical image capture
- Generalized illumination
- Time-of-flight imaging
- High dynamic range imaging
- Plenoptic imaging

Computational Consumer Imaging

- Mobile imaging, cell phone imaging
- Camera-array systems
- Depth cameras, multi-focus imaging
- Pervasive imaging, camera networks

Computational Acoustic Imaging

- Multi-static ultrasound imaging
- Photo-acoustic imaging
- Acoustic tomography

Computational Microscopy

- Holographic microscopy
- Quantitative phase imaging
- Multi-illumination microscopy
- Lensless microscopy
- Light field microscopy

Imaging Hardware and Software

- Embedded computing systems
- Big data computational imaging
- Integrated hardware/digital design

Tomographic Imaging

- X-ray CT
- PET
- SPECT

Magnetic Resonance Imaging

- Diffusion tensor imaging
- Fast acquisition

Radar Imaging

- Synthetic aperture imaging
- Inverse synthetic aperture imaging

Geophysical Imaging

- Multi-spectral imaging
- Ground penetrating radar
- Seismic tomography

Multi-spectral Imaging

- Multi-spectral imaging
- Hyper-spectral imaging
- Spectroscopic imaging

For more information on the **IEEE Transactions on Computational Imaging** see *http://www.signalprocessingsociety.org/publications/periodicals/tci/*

NOW WITH NO PAGE CHARGES!

IEEE Signal Processing Society | EMB | GRSS

General Chair
Lina Karam
Arizona State University

General Co-Chair
Aggelos Katsaggelos
Northwestern University

Technical Program Chairs
Fernando Pereira
Instituto Superior Técnico
Gaurav Sharma
University of Rochester

Innovation Program Chairs
Haohong Wang
TCL Research America
Jeff Bier
BDTI & Embedded Vision Alliance

Finance Chair
Sohail Dianat
Rochester Institute of Technology

Plenary Chairs
Michael Marcellin
University of Arizona
Sethuraman Panchanathan
Arizona State University

Special Sessions Chairs
Dinei Florencio
Microsoft Research
Chaker Larabi
Poitiers University
Zhou Wang
University of Waterloo

Tutorials Chairs
Ghassan AlRegib
Georgia Tech
Rony Ferzli
Intel

Publicity Chair
Michel Sarkis
Qualcomm Technologies Inc.

Awards Chairs
Vivek Goyal
Boston University
Ivana Tosic
Ricoh Innovations

Exhibits Chair
David Frakes
Arizona State University & Google

Publication Chairs
Patrick Le Callet
Nantes University
Baoxin Li
Arizona State University

Local Arrangement Chairs
Jorge Caviedes
Intel
Pavan Turaga
Arizona State University

Registration Chair
Ricardo De Queiroz
Universidade de Brasilia

Conference Management
Conference Management Services

The 23rd IEEE International Conference on Image Processing (ICIP) will be held in the Phoenix Convention Centre, Phoenix, Arizona, USA, on September 25 - 28, 2016. ICIP is the world's largest and most comprehensive technical conference focused on image and video processing and computer vision. In addition to the Technical Program, ICIP 2016 will feature **an Innovation Program focused on innovative vision technologies and fostering innovation, entrepreneurship, and networking**. The conference will feature world-class speakers, tutorials, exhibits, and a vision technology showcase.

Topics in the ICIP 2016 Technical Program include but are not limited to the following:

Filtering, Transforms, Multi-Resolution Processing
Restoration, Enhancement, Super-Resolution
Computer Vision Algorithms and Technologies
Compression, Transmission, Storage, Retrieval
Computational Imaging
Color and Multispectral Processing
Multi-View and Stereoscopic Processing
Multi-Temporal and Spatio-Temporal Processing
Video Processing and Analytics
Authentication and Biometrics
Biological and Perceptual-based Processing
Visual Quality Assessment
Scanning, Display, and Printing
Document and Synthetic Visual Processing
Applications to various fields (e.g., biomedical, Advanced Driving Assist Systems, assistive living, security, learning, health and environmental monitoring, manufacturing, consumer electronics)

The ICIP 2016 innovation program will feature a **vision technology showcase** of state-of-the-art vision technologies, innovation challenges, talks by innovation leaders and entrepreneurs, tutorials, and networking.

Paper Submission: Prospective authors are invited to submit full-length papers at the conference website, with up to four pages for technical content including figures and references, and with one additional optional 5th page for references only. Submission instructions, templates for the required paper format, and information on "no show" policy are available at www.icip2016.com.

Tutorials and Special Sessions Proposals: Tutorials will be held on September 25, 2016. Tutorial proposals should be submitted to *tutorials@icip2016.com* and must include title, outline, contact information, biography and selected publications for the presenter(s), and a description of the tutorial and material to be distributed to participants. Special Sessions proposals should be submitted to *specialsessions@icip2016.com* and must include a topical title, rationale, session outline, contact information, and a list of invited papers. For detailed submission guidelines, please refer the ICIP 2016 website at www.icip2016.com.

Important Deadlines:
Special Session and Tutorial Proposals: November 16, 2015
Notification of Special Session and Tutorial Acceptance: December 18, 2015

Paper Submissions: January 25, 2016
Notification of Paper Acceptance: April 30, 2016

Visual Technology Innovator Award Nomination: March 30, 2016

Revised Paper Upload Deadline: May 30, 2016
Authors' Registration Deadline: May 30, 2016

http://www.facebook.com/icip2016
https://twitter.com/icip2016/
https://www.linkedin.com/groups/ICIP-2016-6940658

IEEE TRANSACTIONS ON

INFORMATION FORENSICS AND SECURITY

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.signalprocessingsociety.org

OCTOBER 2015 VOLUME 10 NUMBER 10 ITIFA6 (ISSN 1556-6013)

PAPERS

IEEE TRANSACTIONS ON
INFORMATION FORENSICS AND SECURITY

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.signalprocessingsociety.org

NOVEMBER 2015 VOLUME 10 NUMBER 11 (ISSN 1556-6013)

PAPERS

IEEE TRANSACTIONS ON

MULTIMEDIA

A PUBLICATION OF
THE IEEE SIGNAL PROCESSING SOCIETY
THE IEEE CIRCUITS AND SYSTEMS SOCIETY
THE IEEE COMMUNICATIONS SOCIETY

IEEE Signal Processing Society | CAS | IEEE COMMUNICATIONS SOCIETY

http://www.signalprocessingsociety.org/tmm/

TECHNICALLY COSPONSORED BY THE IEEE COMPUTER SOCIETY

IEEE computer society

OCTOBER 2015 VOLUME 17 NUMBER 10 ITMUF8 (ISSN 1520-9210)

PAPERS

IEEE

IEEE JOURNAL OF
SELECTED TOPICS IN SIGNAL PROCESSING

IEEE Signal Processing Society

www.ieee.org/sp/index.html

OCTOBER 2015 VOLUME 9 NUMBER 7 IJSTGY (ISSN 1932-4553)

ISSUE ON SIGNAL AND INFORMATION PROCESSING FOR PRIVACY

IEEE

CALL FOR PAPERS

IEEE Signal Processing Society

IEEE Journal on Selected Topics in Signal Processing

Special Issue on **Advanced Signal Processing in Brain Networks**

Aims and Scope

Network models of the brain have become an important tool of modern neurosciences to study fundamental organizational principles of brain structure & function. Their connectivity is captured by the so-called *connectome*, the complete set of structural and functional links of the network. There is still an important need for advancing current methodology; e.g., going towards increasing large-scale models; incorporating multimodal information in multiplex graph models; dealing with dynamical aspects of network models; and matching data-driven and theoretical models.

These challenges form multiple opportunities to develop and adapt emerging signal processing theories and methods at the interface of graph theory, machine learning, applied statistics, simulation, and so on, to play a key role in the analysis and modeling and to bring our understanding of brain networks to the next level for key applications in cognitive and clinical neurosciences, including brain-computer interfaces.

Topics of Interest include (but are not limited to):

- Multi-layer/multiplex networks
- Various types of brain data including (f)MRI, M/EEG, NIRS, ECoG/multi-electrode arrays, genomics, ...
- Novel subspace decompositions (e.g., tensor models, sparsity-driven regularization, low-rank properties)
- Multiscale decompositions (e.g., graph wavelets)
- Advanced statistical inference (e.g., two-step procedures, Riemannian statistics)
- Machine learning (e.g., graph kernels, structured penalties, deep neural networks)
- Dynamical systems and simulation approaches
- Time delay techniques for brain networks
- Big data methods for brain networks (e.g., approximate inference, distributed computing on graphs)
- Dynamical graphical models (e.g., Bayesian non-parametrics, structure learning)
- Clustering (e.g., overlapping/fuzzy communities)

Important Dates:

Manuscript submission due: November 1, 2015
First review completed: January 15, 2016
Revised manuscript due: February 28, 2016
Second review completed: April 15, 2016
Final manuscript due: June 1, 2016

Dimitri Van De Ville Ecole Polytechnique Fédérale de Lausanne and University of Geneva	**Viktor Jirsa** Aix-Marseille University	**Stephen Strother** Rotman Research Institute, Baycrest and University of Toronto	**Jonas Richiardi** University of Geneva	**Andrew Zalesky** The University of Melbourne

Prospective authors should visit http://www.signalprocessingsociety.org/publications/periodicals/jstsp/ for information on paper submission. Manuscripts should be submitted using Manuscript Central at http://mc.manuscriptcentral.com/jstsp-ieee.

CALL FOR PAPERS

IEEE Journal of Selected Topics in Signal Processing

Special Issue on Exploiting Interference towards Energy Efficient and Secure Wireless Communications

Interference has long been the central focus for meeting the ever increasing requirements on quality of service (QoS) in modern and future wireless communication systems. Traditional approaches aim to minimise, cancel or avoid interference. Contrary to this traditional view, which treats interference as a detrimental phenomenon, recent interest has emerged on innovative approaches that consider interference as a useful resource for developing energy efficient and secure 5G communication systems. These include exploiting constructive interference as a source of useful signal power at the modulation level by use of practical multiuser downlink precoding, and also the use of radio frequency radiation for energy harvesting that handles interference and unintended signals as a source of green energy. These techniques open new exciting opportunities in wireless communications by enabling energy self-sustainable and environmentally friendly networks with extended lifetimes, untethered mobility and independence from the power grid, and joint distribution of information and energy within networks. Interference is also being used for physical (PHY) layer secrecy, as an efficient means to jam potential eavesdroppers. This is particularly useful in networks without infrastructure to secure wireless links without the computational overhead imposed by standard cryptographic techniques. These research streams introduce a new vision about interference in wireless networks and motivate a plethora of potential new applications and services. The purpose of this special issue is to re-examine the notion of interference in communications networks and introduce a new paradigm that considers interference as a useful resource in the context of 5G communications.

This special issue seeks to bring together contributions from researchers and practitioners in the area of signal processing for wireless communications with an emphasis on new methods for exploiting interference including symbol level precoding, physical layer security, radiated energy harvesting and wireless power transfer. We solicit high-quality original research papers on topics including, but not limited to:

- Fundamental limits of communication by interference exploitation,
- Modulation level precoding for interference exploitation, interference exploitation in 5G techniques,
- Interference exploitation in the presence of channel state information errors, limited feedback and hardware imperfections,
- Energy harvesting, cooperation and relaying in wireless networks, and in conjunction with 5G methods,
- Time switching, power splitting and antenna switching for simultaneous energy and information transfer,
- Interference exploitation and management in coexisting wireless communications and power transfer systems,
- Joint optimisation of the baseband processing and RF circuit design for energy harvesting,
- Joint interference exploitation and wireless power transfer techniques at the transmitter,
- Security concerns in energy harvesting networks,
- Signal processing for information-theoretic privacy,
- PHY layer secrecy and jamming, PHY secrecy in 5G technologies,
- Introducing artificial and controlled interference for enhancing wireless security,
- Beamforming for PHY-layer secrecy and energy harvesting,
- Interference exploitation and management in coexisting wireless communications and power transfer systems

In addition to technical research results, we invite very high quality submissions of a tutorial or overview nature; we also welcome creative papers outside of the areas listed here but related to the overall scope of the special issue. Prospective authors can contact the Guest Editors to ascertain interest on topics that are not listed.

Prospective authors should visit http://www.signalprocessingsociety.org/publications/periodicals/jstsp/ for information on paper submission. Manuscripts should be submitted using the Manuscript Central system at http://mc.manuscriptcentral.com/jstsp-ieee. Manuscripts will be peer reviewed according to the standard IEEE process.

Manuscript Submission:	January 30, 2016
First review completed:	March 31, 2016
Revised manuscript due:	May 15, 2016
Second review completed:	July 1, 2016
Final manuscript due:	August 15, 2016
Publication date:	December 2016

Guest Editors

Dr. Ioannis Krikidis, University of Cyprus, Cyprus, email: krikidis.ioannis@ucy.ac.cy
Dr. Christos Masouros, University College London, UK, email: c.masouros@ucl.ac.uk
Dr. Gan Zheng, University of Essex, UK, email: ganzheng@essex.ac.uk
Prof. Rui Zhang, National University of Singapore, Singapore, email: elezhang@nus.edu.sg
Prof. Robert Schober, Universität Erlangen-Nürnberg, Germany, email: robert.schober@fau.de

IEEE

SIGNAL PROCESSING LETTERS

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.ieee.org/sp/index.html

NOVEMBER 2015 VOLUME 22 NUMBER 11 ISPLEM (ISSN 1070-9908)

LETTERS


IEEE
SignalProcessing
MAGAZINE
[VOLUME 32 NUMBER 6 NOVEMBER 2015]
THE SCIENCE BEHIND OUR DIGITAL LIFE
EUCLIDEAN DISTANCE MATRICES
PLAYING WITH DUALITY FOR LARGE-SCALE OPTIMIZATION
EXPRESSION CONTROL IN SINGING VOICE SYNTHESIS
SPEAKER RECOGNITION
SPARSE AND TENSOR MODELS FOR BRAIN IMAGING
IEEE Signal Processing Society
IEEE

CONTENTS

VOLUME 32 NUMBER 6

FEATURES

THEORIES AND METHODS

AUDIO AND SPEECH PROCESSING

BIOMEDICAL SIGNAL PROCESSING

Digital Object Identifier 10.1109/MSP.2015.2467197

COLUMNS

DEPARTMENTS

Call for Papers
IEEE Signal Processing Society
IEEE Transactions on Signal and Information Processing over Networks

SPECIAL ISSUE ON INFERENCE AND LEARNING OVER NETWORKS

Networks are everywhere. They surround us at different levels and scales, whether we are dealing with communications networks, power grids, biological colonies, social networks, sensor networks, or distributed Big Data depositories. Therefore, it is not hard to appreciate the ongoing and steady progression of network science, a prolific research field spreading across many theoretical as well as applicative domains. Regardless of the particular context, the very essence of a network resides in the interaction among its individual constituents, and Nature itself offers beautiful paradigms thereof. Many biological networks and animal groups owe their sophistication to fairly structured patterns of cooperation, which are vital to their successful operation. While each individual agent is not capable of sophisticated behavior on its own, the *combined interplay* among simpler units and the *distributed processing* of dispersed pieces of information, enable the agents to solve complex tasks and enhance dramatically their performance. Self-organization, cooperation and adaptation emerge as the essential, combined attributes of a network tasked with distributed information processing, optimization, and inference. Such a network is conveniently described as an ensemble of spatially dispersed (possibly moving) agents, linked together through a (possibly time-varying) connection topology. The agents are allowed to interact locally and to perform in-network processing, in order to accomplish the assigned inferential task. Correspondingly, several problems such as, e.g., network intrusion, community detection, and disease outbreak inference, can be conveniently described by signals on graphs, where the graph typically accounts for the topology of the underlying space and we obtain multivariate observations associated with nodes/edges of the graph. The goal in these problems is to identify/infer/learn patterns of interest, including anomalies, outliers, and existence of latent communities. Unveiling the fundamental principles that govern distributed inference and learning over networks has been the common scope across a variety of disciplines, such as signal processing, machine learning, optimization, control, statistics, physics, economics, biology, computer, and social sciences. In the realm of signal processing, many new challenges have emerged, which stimulate research efforts toward delivering the theories and algorithms necessary to (a) designing networks with sophisticated inferential and learning abilities; (b) promoting truly distributed implementations, endowed with real-time adaptation abilities, needed to face the dynamical scenarios wherein real-world networks operate; and (c) discovering and disclosing significant relationships possibly hidden in the data collected from across networked systems and entities. This call for papers therefore encourages submissions from a broad range of experts that study such fundamental questions, including but not limited to:

- Adaptation and learning over networks.
- Consensus strategies; diffusion strategies.
- Distributed detection, estimation and filtering over networks.
- Distributed dictionary learning.
- Distributed game-theoretic learning.
- Distributed machine learning; online learning.
- Distributed optimization; stochastic approximation.
- Distributed proximal techniques, sub-gradient techniques.
- Learning over graphs; network tomography.
- Multi-agent coordination and processing over networks.
- Signal processing for biological, economic, and social networks.
- Signal processing over graphs.

Prospective authors should visit http://www.signalprocessingsociety.org/publications/periodicals/tsipn/ for information on paper submission. Manuscripts should be submitted via Manuscript Central at http://mc.manuscriptcentral.com/tsipn-ieee.

Important Dates:

- Manuscript submission: February 1, 2016
- First review completed: April 1, 2016
- Revised manuscript due: May 15, 2016
- Second review completed: July 15, 2016
- Final manuscript due: September 15, 2016
- Publication: December 1, 2016

Guest Editors:

Vincenzo **Matta**, University of Salerno, Italy, vmatta@unisa.it
Cédric **Richard**, University of Nice Sophia-Antipolis, France, cedric.richard@unice.fr
Venkatesh **Saligrama**, Boston University, USA, srv@bu.edu
Ali H. **Sayed**, University of California, Los Angeles, USA, sayed@ucla.edu

Call for Papers

http://ssp2016.tsc.uc3m.es

2016 IEEE Statistical Signal Processing Workshop

26-29 June 2016, Palma de Mallorca, Spain

The 2016 IEEE Workshop on Statistical Signal Processing (SSP 2016) is the 19th of a series of unique meetings that bring members of the IEEE Signal Processing Society together with researchers from allied fields such as bioinformatics, communications, machine learning, and statistics.

The scientific program of SSP 2016 will include invited plenary talks, as well as regular and special sessions with contributed research papers. All submitted papers will be reviewed by experts and only a proportion will be accepted to maintain a high quality workshop. All accepted papers will be published on IEEE Xplore. The scope of the workshop includes basic theory, methods and algorithms, and applications in the following areas:

Theoretical Topics

- Adaptive systems and signal processing
- Detection and estimation theory
- Learning theory and pattern recognition
- Multivariate statistical analysis
- System identification and calibration
- Monte Carlo methods
- Network and graph analysis
- Random matrix theory
- Time-frequency and time-scale analysis
- Compressed sensing
- Point process estimation
- Stochastic filtering

Application Areas

- Bioinformatics and genomics
- Array processing, radar and sonar
- Communication systems and networks
- Sensor networks
- Information forensics and security
- Medical imaging
- Biomedical signal processing
- Preventive, social network analysis
- Smart grids and industrial applications
- Geoscience
- Astrophysics
- New methods, directions and applications

IEEE Signal Processing Society IEEE

Venue

Es Baluard Museu d'Art Modern i Contemporani, Palma de Mallorca, Spain

Paper Submission

Prospective authors are invited to submit full-length papers, with up to four pages for technical content including figures and references, using the templates and formatting guidelines posted on the website. All accepted papers must be presented at the workshop in order to be published in the proceedings. Best student paper awards, selected by a SSP committee, will be presented at the workshop.

Special Sessions

In addition to regular sessions, the workshop will also have a number of special sessions on topics of particular relevance. Prospective organizers of special sessions are invited to submit a proposal form, available on the workshop website, by e-mail to the Special Sessions Chair.

Important Dates

Submission of proposals for special sessions	**Nov 09, 2015**
Notification of acceptance of special sessions	**Nov 30, 2015**
Full paper submission deadline	**Feb 08, 2016**
Notificacion of acceptance	**April 04, 2016**
Camera ready papers due on	**April 18, 2016**

Organization

General Chairs:

Antonio Artés-Rodríguez (Universidad Carlos III de Madrid, Spain)

Joaquín Miguez (Universidad Carlos III de Madrid, Spain)

Technical Program Chairs:

Sergios Theodoridis (National and Kapodistrian University of Athens, Greece)

Konstantinos Slavakis (University of Minnesota, USA)

Finance Chair:

Matilde Sánchez-Fernández (Universidad Carlos III de Madrid, Spain)

Special Sessions Chair:

Mónica F. Bugallo (Stony Brook University, USA)

Local Arrangements Chairs:

Guillem Femenias (Universitat de les Illes Balears, Spain)

Felip Riera-Palou (Universitat de les Illes Balears, Spain)

Publications Chair

Pau Closas (CTTC, Spain)

General Chairs
Tsuhan Chen, Cornell Univ.
Ming-Ting Sun, Univ. Washington
Cha Zhang, Microsoft Research
Program Chairs
Philip Chou, Microsoft Research
Anthony Vetro, MERL
Max Mühlhäuser, TU Darmstadt
Lap-Pui Chau, NTU
Jenq-Neng Huang, Univ. Washington
Yung-Hsiang Lu, Purdue Univ.
Finance Chairs
Ying Li, IBM Research
Yi Wu, Intel Labs
Plenary Chairs
John Apostolopoulos, Cisco
Antonio Ortega, USC
Workshop Chairs
Pascal Frossard, EPFL
Ivana Tosic, Ricoh
Tutorial Chairs
Yap-Peng Tan, NTU
Lexing Xie, Australian Natl. Univ.
Special Session Chairs
Aljoscha Smolic, Disney Research
Luigi Atzoni, Univ. of Cagliari
Panel Chairs
Fernando Pereira, IST
Gene Cheung, NII
Award Chair
Chang Wen Chen, SUNY Buffalo
Industrial Program Chairs
Onur Guleryuz, Polytechnic Univ.
Ton Kalker, Huawei
Student Program Chairs
Jane Z. Wang, UBC
Ivan Bajić, Simon Fraser Univ.
Grand Challenge Chairs
Christian Timmerer, UNIKLU
Andrew Gallagher, Google
Demo/Expo Chairs
Jacob Chakareski, Univ. of Alabama
Qiong Liu, FXPAL
Local/Events Chairs
Zicheng Liu, Microsoft Research
Jue Wang, Adobe Research
Lu Xia, Amazon
Publicity Chairs
Kiyoharu Aizawa, Univ. of Tokyo
Maria Martini, KCOL
Sponsorship Chairs
Belle Tseng, Apple Inc.
Yen-Kuang Chen, Intel Research
Publication Chairs:
Junsong Yuan, NTU
Chia-Wen Lin, NTHU
Registration Chairs:
YingLi Tian, CUNY,
Yan Tong, Univ. of South Carolina
Web Chair
Jie Liang, Simon Fraser Univ.

CALL FOR PAPERS

IEEE International Conference on Multimedia and Expo (ICME) 2016

July 11-15, 2016 · Seattle, USA

With around 1000 submissions and 500 participants each year, the IEEE International Conference on Multimedia & Expo (ICME) has been the flagship multimedia conference sponsored by four IEEE societies since 2000. It serves as a forum to promote the exchange of the latest advances in multimedia technologies, systems, and applications from both the research and development perspectives of the circuits and systems, communications, computer, and signal processing communities. In 2016, an Exposition of multimedia products, prototypes and animations will be held in conjunction with the conference.

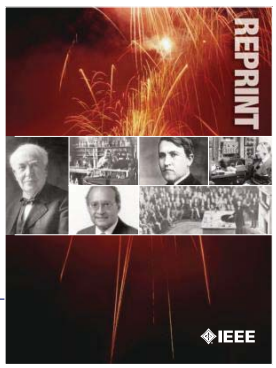
Authors are invited to submit a full paper (two-column format, 6 pages maximum) according to the guidelines available on the conference website at http://icme2016.org/. Only electronic submissions will be accepted. Topics of interest include, but are not limited to:

- Speech, audio, image, video, text and new sensor signal processing
- Signal processing for media integration
- 3D visualization and animation
- 3D imaging and 3DTV
- Virtual reality and augmented reality
- Multi-modal multimedia computing systems and human-machine interaction
- Multimedia communications and networking
- Media content analysis
- Multimedia quality assessment
- Multimedia security and content protection
- Multimedia databases and digital libraries
- Multimedia applications and services
- Multimedia standards and related issues

ICME 2016 aims to have high quality oral and poster presentations. Several awards sponsored by industry and institutions will be given out. Best papers will be presented in a single-track session to all participants. Accepted papers should be presented, or else they will not be included in the IEEE Xplore Library.

A number of Workshops will be organized by the sponsoring societies. To further foster new emerging topics, ICME 2016 also welcomes researchers, developers and practitioners to organize regular Workshops. Industrial exhibitions are held in parallel with the main conference. Proposals for Special Sessions, Tutorials, and Demos are also invited. Please visit the ICME 2016 website for submission details.

Special Session Proposals Due: October 1, 2015
Notification of Special Session Acceptance: October 17, 2015
Regular Paper Abstract Submission: November 30, 2015
Regular Paper Submission: December 4, 2015
Workshop Proposals Due: November 20, 2015
Notification of Workshop Proposal Acceptance: December 15, 2015
Panel/Tutorial Proposals Due: January 15, 2016
Notification of Panel/Tutorial Acceptance: February 29, 2016
Notification of Regular Paper Acceptance: March 11, 2016
Workshop & Demo Paper Submission: March 18, 2016
Notification of Workshop and Demo Paper Acceptance: April 22, 2016
Camera-Ready Papers Due: May 13, 2016
Exhibition Application: May 13, 2016
Conference Website: http://icme2016.org/



IEEE ORDER FORM FOR REPRINTS

Purchasing IEEE Papers in Print is easy, cost-effective and quick.

Complete this form, send via our secure fax (24 hours a day) to 732-981-8062 or mail it back to us.

PLEASE FILL OUT THE FOLLOWING

Author: _____

Publication Title: _____

Paper Title: _____

_____

RETURN THIS FORM TO:
IEEE Publishing Services
445 Hoes Lane
Piscataway, NJ 08855-1331

Email the Reprint Department at reprints@ieee.org for questions regarding this form

PLEASE SEND ME

■ 50 ■ 100 ■ 200 ■ 300 ■ 400 ■ 500 or _____(in multiples of 50) reprints.

■ YES ■ NO Self-covering/title page required. COVER PRICE: $74 per 100, $39 per 50.

■ $58.00 Air Freight must be added for all orders being shipped outside the U.S.

■ $21.50 must be added for all USA shipments to cover the cost of UPS shipping and handling.

PAYMENT

■ Check enclosed. Payable on a bank in the USA.

■ Charge my: ■ Visa ■ Mastercard ■ Amex ■ Diners Club

Account # _____ Exp. date _____

Cardholder's Name (please print): _____

_____

■ Bill me (you must attach a purchase order) Purchase Order Number _____

Send Reprints to:

_____

_____

_____

_____

Bill to address, if different:

_____

_____

_____

_____

Because information and papers are gathered from various sources, there may be a delay in receiving your reprint request. This is especially true with postconference publications. Please provide us with contact information if you would like notification of a delay of more than 12 weeks.

Telephone: _____ Fax: _____ Email Address: _____

2012 REPRINT PRICES (without covers)

Number of Text Pages

	1-4	5-8	9-12	13-16	17-20	21-24	25-28	29-32	33-36	37-40	41-44	45-48
50	$129	$213	$245	$248	$288	$340	$371	$408	$440	$477	$510	$543
100	$245	$425	$479	$495	$573	$680	$742	$817	$885	$953	$1021	$1088

Larger quantities can be ordered. Email reprints@ieee.org with specific details.

Tax Applies on shipments of regular reprints to CA, DC, FL, MI, NJ, NY, OH and Canada (GST Registration no. 12534188).
Prices are based on black & white printing. Please call us for full color price quote, if applicable.

2016 IEEE MEMBERSHIP APPLICATION

(students and graduate students must apply online)

IEEE *Advancing Technology for Humanity*

Start your membership immediately: Join online www.ieee.org/join

Please complete both sides of this form, typing or **printing in capital letters**. Use only English characters and abbreviate only if more than 40 characters and spaces per line. We regret that incomplete applications cannot be processed.

1 Name & Contact Information

Please PRINT your name as you want it to appear on your membership card and IEEE correspondence. As a key identifier for the IEEE database, circle your last/surname.

☐ Male ☐ Female Date of birth (Day/Month/Year) ____/____/____

Title First/Given Name Middle Last/Family Surname

▼ Primary Address ☐ Home ☐ Business (All IEEE mail sent here)

Street Address

City State/Province

Postal Code Country

Primary Phone

Primary E-mail

▼ Secondary Address ☐ Home ☐ Business

Company Name Department/Division

Street Address City State/Province

Postal Code Country

Secondary Phone

Secondary E-mail

To better serve our members and supplement member dues, your postal mailing address is made available to carefully selected organizations to provide you with information on technical services, continuing education, and conferences. Your e-mail address is not rented by IEEE. Please check box only if you do not want to receive these postal mailings to the selected address. ☐

2 Attestation

I have graduated from a three- to five-year academic program with a university-level degree.
☐ Yes ☐ No

This program is in one of the following fields of study:
☐ Engineering
☐ Computer Sciences and Information Technologies
☐ Physical Sciences
☐ Biological and Medical Sciences
☐ Mathematics
☐ Technical Communications, Education, Management, Law and Policy
☐ Other (please specify): _____

This academic institution or program is accredited in the country where the institution is located. ☐ Yes ☐ No ☐ Do not know

I have _____ years of professional experience in teaching, creating, developing, practicing, or managing within the following field:

☐ Engineering
☐ Computer Sciences and Information Technologies
☐ Physical Sciences
☐ Biological and Medical Sciences
☐ Mathematics
☐ Technical Communications, Education, Management, Law and Policy
☐ Other (please specify): _____

3 Please Tell Us About Yourself

Select the numbered option that best describes yourself. This information is used by IEEE magazines to verify their annual circulation. Please enter numbered selections in the boxes provided.

A. Primary line of business → ☐

1. Computers
2. Computer peripheral equipment
3. Software
4. Office and business machines
5. Test, measurement and instrumentation equipment
6. Communications systems and equipment
7. Navigation and guidance systems and equipment
8. Consumer electronics/appliances
9. Industrial equipment, controls and systems
10. ICs and microprocessors
11. Semiconductors, components, sub-assemblies, materials and supplies
12. Aircraft, missiles, space and ground support equipment
13. Oceanography and support equipment
14. Medical electronic equipment
15. OEM incorporating electronics in their end product (not elsewhere classified)
16. Independent and university research, test and design laboratories and consultants (not connected with a mfg. co.)
17. Government agencies and armed forces
18. Companies using and/or incorporating any electronic products in their manufacturing, processing, research or development activities
19. Telecommunications services, telephone (including cellular)
20. Broadcast services (TV, cable, radio)
21. Transportation services (airline, railroad, etc.)
22. Computer and communications and data processing services
23. Power production, generation, transmission and distribution
24. Other commercial users of electrical, electronic equipment and services (not elsewhere classified)
25. Distributor (reseller, wholesaler, retailer)
26. University, college/other educational institutions, libraries
27. Retired
28. Other_____

B. Principal job function → ☐

1. General and corporate management
2. Engineering management
3. Project engineering management
4. Research and development management
5. Design engineering management—analog
6. Design engineering management—digital
7. Research and development engineering
8. Design/development engineering—analog
9. Design/development engineering—digital
10. Hardware engineering
11. Software design/development
12. Computer science
13. Science/physics/mathematics
14. Engineering (not elsewhere specified)
15. Marketing/sales/purchasing
16. Consulting
17. Education/teaching
18. Retired
19. Other_____

C. Principal responsibility → ☐

1. Engineering and scientific management
2. Management other than engineering
3. Engineering design
4. Engineering
5. Software: science/mngmnt/engineering
6. Education/teaching
7. Consulting
8. Retired
9. Other_____

D. Title → ☐

1. Chairman of the Board/President/CEO
2. Owner/Partner
3. General Manager
4. VP Operations
5. VP Engineering/Dir. Engineering
6. Chief Engineer/Chief Scientist
7. Engineering Management
8. Scientific Management
9. Member of Technical Staff
10. Design Engineering Manager
11. Design Engineer
12. Hardware Engineer
13. Software Engineer
14. Computer Scientist
15. Dean/Professor/Instructor
16. Consultant
17. Retired
18. Other_____

Are you now or were you ever a member of IEEE?
☐ Yes ☐ No If yes, provide, if known:

Membership Number Grade Year Expired

4 Please Sign Your Application

I hereby apply for IEEE membership and agree to be governed by the IEEE Constitution, Bylaws, and Code of Ethics. I understand that IEEE will communicate with me regarding my individual membership and all related benefits. **Application must be signed.**

Signature Date

Over Please

5 Add IEEE Society Memberships (Optional)

The 39 IEEE Societies support your technical and professional interests. Many society memberships include a personal subscription to the core journal, magazine, or newsletter of that society. **For a complete list of everything included with your IEEE Society membership, visit www.ieee.org/join.** All prices are quoted in US dollars.

Please check ✔ the appropriate box.

		BETWEEN 16 AUG 2015-28 FEB 2016 PAY	BETWEEN 1 MAR 2016-15 AUG 2016 PAY
IEEE Aerospace and Electronic Systems ⌘ 🖂	AES010	25.00 ☐	12.50 ☐
IEEE Antennas and Propagation ⌘ 🖂	AP003	15.00 ☐	7.50 ☐
IEEE Broadcast Technology ▪ ⌘ 🖂	BT002	15.00 ☐	7.50 ☐
IEEE Circuits and Systems ▪ ⌘ 🖂	CAS004	22.00 ☐	11.00 ☐
IEEE Communications ⌘ 🖂	COM019	30.00 ☐	15.00 ☐
IEEE Components, Packaging, & Manu. Tech. ▪ ⌘ 🖂	CPMT021	15.00 ☐	7.50 ☐
IEEE Computational Intelligence ⌘ 🖂	CIS011	29.00 ☐	14.50 ☐
IEEE Computer ▪ ⌘ 🖂	C016	56.00 ☐	28.00 ☐
IEEE Consumer Electronics ⌘ 🖂	CE008	20.00 ☐	10.00 ☐
IEEE Control Systems ⌘ 🖂	CS023	25.00 ☐	12.50 ☐
IEEE Dielectrics and Electrical Insulation ⌘ 🖂	DEI032	26.00 ☐	13.00 ☐
IEEE Education ⌘ 🖂	E025	20.00 ☐	10.00 ☐
IEEE Electromagnetic Compatibility ▪ ⌘ 🖂	EMC027	31.00 ☐	15.50 ☐
IEEE Electron Devices ▪ ⌘ 🖂	ED015	18.00 ☐	9.00 ☐
IEEE Engineering in Medicine and Biology ⌘ 🖂	EMB018	40.00 ☐	20.00 ☐
IEEE Geoscience and Remote Sensing ⌘ 🖂	GRS029	19.00 ☐	9.50 ☐
IEEE Industrial Electronics ⌘ 🖂	IE013	9.00 ☐	4.50 ☐
IEEE Industry Applications ⌘ 🖂	IA034	20.00 ☐	10.00 ☐
IEEE Information Theory ▪ ⌘ 🖂	IT012	30.00 ☐	15.00 ☐
IEEE Instrumentation and Measurement ⌘ 🖂	IM009	29.00 ☐	14.50 ☐
IEEE Intelligent Transportation Systems ▪ ⌘ 🖂	ITSS038	35.00 ☐	17.50 ☐
IEEE Magnetics ▪ 🖂 ♦⌘	MAG033	26.00 ☐	13.00 ☐
IEEE Microwave Theory and Techniques ⌘ 🖂	MTT017	17.00 ☐	8.50 ☐
IEEE Nuclear and Plasma Sciences ▪ ⌘ 🖂	NPS005	35.00 ☐	17.50 ☐
IEEE Oceanic Engineering ▪ ⌘ 🖂	OE022	19.00 ☐	9.50 ☐
IEEE Photonics ▪ ⌘ 🖂	PHO036	34.00 ☐	17.00 ☐
IEEE Power Electronics ⌘ 🖂	PEL035	25.00 ☐	12.50 ☐
IEEE Power & Energy ▪ ⌘ 🖂	PE031	35.00 ☐	17.50 ☐
IEEE Product Safety Engineering ▪	PSE043	35.00 ☐	17.50 ☐
IEEE Professional Communication ▪ ⌘ 🖂	PC026	31.00 ☐	15.50 ☐
IEEE Reliability ▪ ⌘ 🖂	RL007	35.00 ☐	17.50 ☐
IEEE Robotics and Automation ⌘ 🖂	RA024	9.00 ☐	4.50 ☐
IEEE Signal Processing ▪ ⌘ 🖂	SP001	22.00 ☐	11.00 ☐
IEEE Social Implications of Technology ▪ ⌘ 🖂	SIT030	33.00 ☐	16.50 ☐
IEEE Solid-State Circuits ⌘ 🖂	SSC037	22.00 ☐	11.00 ☐
IEEE Systems, Man, & Cybernetics ⌘ 🖂	SMC028	12.00 ☐	6.00 ☐
IEEE Technology & Engineering Management ⌘ 🖂	TEM014	35.00 ☐	17.50 ☐
IEEE Ultrasonics, Ferroelectrics, & Frequency Control ▪ ⌘ 🖂	UFFC020	20.00 ☐	10.00 ☐
IEEE Vehicular Technology ▪ ⌘ 🖂	VT006	18.00 ☐	9.00 ☐

Legend—Society membership includes:

⌘ One or more Society publications
▪ Society newsletter
🖂 Online access to publication
♦ CD-ROM of selected society publications

Complete both sides of this form, sign, and return to:
IEEE MEMBERSHIP APPLICATION PROCESSING
445 HOES LN, PISCATAWAY, NJ 08854-4141 USA
or fax to +1 732 981 0225
or join online at www.ieee.org/join

Please reprint your full name here

6 2016 IEEE Membership Rates (student rates available online)

IEEE member dues and regional assessments are based on where you live and when you apply. Membership is based on the calendar year from 1 January through 31 December. All prices are quoted in US dollars.

Please check ✔ the appropriate box.

RESIDENCE	BETWEEN 16 AUG 2015-28 FEB 2016 PAY	BETWEEN 1 MAR 2016-15 AUG 2016 PAY
United States	$197.00 ☐	$98.50 ☐
Canada (GST)*	$173.35 ☐	$86.68 ☐
Canada (NB, NF and ON HST)*	$185.11 ☐	$92.56 ☐
Canada (Nova Scotia HST)*	$188.05 ☐	$94.03 ☐
Canada (PEI HST)*	$186.58 ☐	$93.29 ☐
Canada (GST and QST Quebec)	$188.01 ☐	$94.01 ☐
Africa, Europe, Middle East	$160.00 ☐	$80.00 ☐
Latin America	$151.00 ☐	$75.50 ☐
Asia, Pacific	$152.00 ☐	$76.00 ☐

*IEEE Canada Business No. 125634188

Minimum Income or Unemployed Provision

Applicants who certify that their prior year income did not exceed US$14,700 (or equivalent) or were not employed are granted 50% reduction in: full-year dues, regional assessment and fees for one IEEE Membership plus one Society Membership. If applicable, please check appropriate box and adjust payment accordingly. Student members are not eligible.

☐ I certify I earned less than US$14,700 in 2015
☐ I certify that I was unemployed in 2015

7 More Recommended Options

Proceedings of the IEEE ..... print $47.00 ☐ **or** online $41.00 ☐
Proceedings of the IEEE (print/online combination) ..... $57.00 ☐
IEEE Standards Association (IEEE-SA) ..... $53.00 ☐
IEEE Women in Engineering (WIE) ..... $25.00 ☐

8 Payment Amount

Please total the Membership dues, Society dues, and other amounts from this page:

IEEE Membership dues ⑥ ..... $_____
IEEE Society dues (optional) ⑤ ..... $_____
IEEE-SA/WIE dues (optional) ⑦ ..... $_____
Proceedings of the IEEE (optional) ⑦ ..... $_____
Canadian residents pay 5% GST or appropriate HST (BC—12%; NB, NF, ON-13%;NS-15%) on Society payments & publications only ..... TAX $_____

AMOUNT PAID ..... **TOTAL** $_____

Payment Method

All prices are quoted in US dollars. You may pay for IEEE membership by credit card (see below), check, or money order payable to IEEE, drawn on a US bank.

☐ Check ☐ VISA ☐ MasterCard ☐ Discover ☐ American Express ☐ Discover

Credit Card Number

MONTH / YEAR EXPIRATION DATE

CARDHOLDER'S 5-DIGIT ZIPCODE (BILLING STATEMENT ADDRESS) USA ONLY

Name as it appears on card

Signature

☐ Auto Renew my Memberships and Subscriptions (available when paying by credit card).
☐ I agree to the Terms and Conditions located at www.ieee.org/autorenew

9 Were You Referred to IEEE?

☐ Yes ☐ No If yes, provide the following:

Member Recruiter Name _____

IEEE Recruiter's Member Number (Required) _____

CAMPAIGN CODE	PROMO CODE

15-MEM-385 P 6/15

Information for Authors

(Updated/Effective January 2015)

For Transactions and Journals:

Authors are encouraged to submit manuscripts of Regular papers (papers which provide a complete disclosure of a technical premise), or Comment Correspondences (brief items that provide comment on a paper previously published in these TRANSACTIONS).

Submissions/resubmissions must be previously unpublished and may not be under consideration elsewhere.

Every manuscript must:

i. provide a clear statement of the problem and what the contribution of the work is to the relevant research community;
ii. state why this contribution is significant (what impact it will have);
iii. provide citation of the published literature most closely related to the manuscript; and
iv. state what is distinctive and new about the current manuscript relative to these previously published works.

By submission of your manuscript to these TRANSACTIONS, all listed authors have agreed to the authorship list and all the contents and confirm that the work is original and that figures, tables and other reported results accurately reflect the experimental work. In addition, the authors all acknowledge that they accept the rules established for publication of manuscripts, including agreement to pay all overlength page charges, color charges, and any other charges and fees associated with publication of the manuscript. Such charges are not negotiable and cannot be suspended. The corresponding author is responsible for obtaining consent from all co-authors and, if needed, from sponsors before submission.

In order to be considered for review, a paper must be within the scope of the journal and represent a novel contribution. A paper is a candidate for an Immediate Rejection if it is of limited novelty, e.g. a straightforward combination of theories and algorithms that are well established and are repeated on a known scenario. Experimental contributions will be rejected without review if there is insufficient experimental data. These TRANSACTIONS are published in English. Papers that have a large number of typographical and/or grammatical errors will also be rejected without review.

In addition to presenting a novel contribution, acceptable manuscripts must describe and cite related work in the field to put the contribution in context. Do not give theoretical derivations or algorithm descriptions that are easily found in the literature; merely cite the reference.

New and revised manuscripts should be prepared following the "Manuscript Submission" guidelines below, and submitted to the online manuscript system, ScholarOne Manuscripts. Do not send original submissions or revisions directly to the Editor-in-Chief or Associate Editors; they will access your manuscript electronically via the ScholarOne Manuscript system.

Manuscript Submission. Please follow the next steps.

1. *Account in ScholarOne Manuscripts.* If necessary, create an account in the on-line submission system ScholarOne Manuscripts. Please check first if you already have an existing account which is based on your e-mail address and may have been created for you when you reviewed or authored a previous paper.
2. *Electronic Manuscript.* Prepare a PDF file containing your manuscript in double-column, single-spaced format using a font size of 10 points or larger, having a margin of at least 1 inch on all sides. Upload this version of the manuscript as a PDF file "double.pdf" to the ScholarOne-Manuscripts site. Since many reviewers prefer a larger font, you are strongly encouraged to also submit a single-column, double-spaced version (11 point font or larger), which is easy to create with the templates provided **IEEE Author Digital Toolbox (http://www.ieee.org/publications_standards/publications/authors/authors_journals.html)**. Page length restrictions will be determined by the double-column version. Proofread your submission, confirming that all figures and equations are visible in your document before you "SUBMIT" your manuscript. Proofreading is critical; once you submit your manuscript, the manuscript cannot be changed in any way. You may also submit your manuscript as a .PDF or MS Word file. The system has the capability of converting your files to PDF, however it is your responsibility to confirm that the conversion is correct and there are no font or graphics issues prior to completing the submission process.
3. *EDICS (Not applicable to Journal of Selected Topics in Signal Processing).* All submissions must be classified by the author with an EDICS (Editors' Information Classification Scheme) selected from the list of EDICS published online at the at the publication's EDICS webpage (*please see the list below). Upon submission of a new manuscript, please choose the EDICS categories that best suit your manuscript. Failure to do so will likely result in a delay of the peer review process.
4. *Additional Documents for Review.* Please upload pdf versions of all items in the reference list that are not publicly available, such as unpublished (submitted) papers. Graphical abstracts and supplemental materials intended to appear with the final paper (see below) must also be uploaded for review at the time of the initial submission for consideration in the review process. Use short filenames without spaces or special characters. When the upload of each file is completed, you will be asked to provide a description of that file.
5. *Supplemental Materials.* IEEE Xplore can publish multimedia files (audio, images, video), datasets, and software (e.g. Matlab code) along with your paper. Alternatively, you can provide the links to such files in a README file that appears on Xplore along with your paper. For details, please see IEEE Author Digital Toolbox under "Multimedia." To make your work reproducible by others, these TRANSACTIONS encourages you to submit all files that can recreate the figures in your paper.
6. *Submission.* After uploading all files and proofreading them, submit your manuscript by clicking "Submit." A confirmation of the successful submission will open on screen containing the manuscript tracking number and will be followed with an e-mail confirmation to the corresponding and all contributing authors. Once you click "Submit," your manuscript cannot be changed in any way.
7. *Copyright Form and Consent Form.* By policy, IEEE owns the copyright to the technical contributions it publishes on behalf of the interests of the IEEE, its authors, and their employers; and to facilitate the appropriate reuse of this material by others. To comply with the IEEE copyright policies, authors are required to sign and submit a completed "IEEE Copyright and Consent Form" prior to publication by the IEEE. The IEEE recommends authors to use an effective electronic copyright form (eCF) tool within the ScholarOne Manuscripts system. You will be redirected to the "IEEE Electronic Copyright Form" wizard at the end of your original submission; please simply sign the eCF by typing your name at the proper location and click on the "Submit" button.

Comment Correspondence. Comment Correspondences provide brief comments on material previously published in these TRANSACTIONS. These items may not exceed 2 pages in double-column, single spaced format, using 9 point type, with margins of 1 inch minimum on all sides, and including: title, names and contact information for authors, abstract, text, references, and an appropriate number of illustrations and/or tables. Correspondence items are submitted in the same way as regular manuscripts (see "Manuscript Submission" above for instructions).

Authors may also submit manuscripts of overview articles, but note that these include an additional white paper approval process http://www.signalprocessingsociety.org/publications/overview-articles/. *[This does not apply to the Journal of Selected Topics in Signal Processing. Please contact the Editor-in-Chief.]*

Digital Object Identifier

Manuscript Length. For the initial submission of a regular paper, the manuscript may not exceed 13 double-column pages (10 point font), including title; names of authors and their complete contact information; abstract; text; all images, figures and tables, appendices and proofs; and all references. Supplemental materials and graphical abstracts are not included in the page count. For regular papers, the revised manuscript may not exceed 16 double-column pages (10 point font), including title; names of authors and their complete contact information; abstract; text; all images, figures and tables, appendices and proofs; and all references. For Overview Papers, the maximum length is double that for regular submissions at each stage (please reference http://www.signalprocessingsociety.org/publications/overview-articles/ for more information).

Note that any paper in excess of 10 pages will be subject to mandatory overlength page charges. Since changes recommended as a result of peer review may require additions to the manuscript, it is strongly recommended that you practice economy in preparing original submissions. Note: Papers submitted to the TRANSACTIONS ON MULTIMEDIA in excess of 8 pages will be subject to mandatory overlength page charges.

Exceptions to manuscript length requirements may, under extraordinary circumstances, be granted by the Editor-in-Chief. However, such exception does not obviate your requirement to pay any and all overlength or additional charges that attach to the manuscript.

Resubmission of Previously Rejected Manuscripts. Authors of manuscripts rejected from any journal are allowed to resubmit their manuscripts only once. At the time of submission, you will be asked whether your manuscript is a new submission or a resubmission of an earlier rejected manuscript. If it is a resubmission of a manuscript previously rejected by any journal, you are expected to submit supporting documents identifying the previous submission and detailing how your new version addresses all of the reviewers' comments. Papers that do not disclose connection to a previously rejected paper or that do not provide documentation as to changes made may be immediately rejected.

Author Misconduct. Author misconduct includes plagiarism, self-plagiarism, and research misconduct, including falsification or misrepresentation of results. All forms of misconduct are unacceptable and may result in sanctions and/or other corrective actions. Plagiarism includes copying someone else's work without appropriate credit, using someone else's work without clear delineation of citation, and the uncited reuse of an author's previously published work that also involves other authors. Self-plagiarism involves the verbatim copying or reuse of an authors own prior work without appropriate citation, including duplicate submission of a single journal manuscript to two different journals, and submission of two different journal manuscripts which overlap substantially in language or technical contribution. For more information on the definitions, investigation process, and corrective actions related to author misconduct, see the Signal Processing Society Policies and Procedures Manual, Section 6.1. http://www.signalprocessingsociety.org/about-sps/governance/policy-procedure/part-2. Author misconduct may also be actionable by the IEEE under the rules of Member Conduct.

Extensions of the Author's Prior Work. It is acceptable for conference papers to be used as the basis for a more fully developed journal submission. Still, authors are required to cite their related prior work; the papers cannot be identical; and the journal publication must include substantively novel aspects such as new experimental results and analysis or added theoretical work. The journal publication should clearly specify how the journal paper offers novel contributions when citing the prior work. Limited overlap with prior journal publications with a common author is allowed only if it is necessary for the readability of the paper, and the prior work must be cited as the primary source.

Submission Format. Authors are required to prepare manuscripts employing the on-line style files developed by IEEE, which include guidelines for abbreviations, mathematics, and graphics. All manuscripts accepted for publication will require the authors to make final submission employing these style files. The style files are available on the web at the **IEEE Author Digital Toolbox** under "Template for all TRANSACTIONS." (LaTeX and MS Word). Please note the following requirements about the abstract:

- The abstract must be a concise yet comprehensive reflection of what is in your article.
- The abstract must be self-contained, without abbreviations, footnotes, displayed equations, or references.
- The abstract must be between 150-250 words.
- The abstract should include a few keywords or phrases, as this will help readers to find it. Avoid over-repetition of such phrases as this can result in a page being rejected by search engines.

In addition to written abstracts, papers may include a graphical abstract; see http://www.ieee.org/publications_standards/publications/authors/authors_journals.html. for options and format requirements.

IEEE supports the publication of author names in the native language alongside the English versions of the names in the author list of an article. For more information, see "Author names in native languages" (http://www.ieee.org/publications_standards/publications/authors/auth_names_native_lang.pdf) on the IEEE Author Digital Toolbox page.

Open Access. The publication is a hybrid journal, allowing either Traditional manuscript submission or Open Access (author-pays OA) manuscript submission. Upon submission, if you choose to have your manuscript be an Open Access article, you commit to pay the discounted $1,750 OA fee if your manuscript is accepted for publication in order to enable unrestricted public access. Any other application charges (such as overlength page charge and/or charge for the use of color in the print format) will be billed separately once the manuscript formatting is complete but prior to the publication. If you would like your manuscript to be a Traditional submission, your article will be available to qualified subscribers and purchasers via IEEE Xplore. No OA payment is required for Traditional submission.

Page Charges.

Voluntary Page Charges. Upon acceptance of a manuscript for publication, the author(s) or his/her/their company or institution will be asked to pay a charge of $110 per page to cover part of the cost of publication of the first ten pages that comprise the standard length (two pages, in the case of Correspondences).

Mandatory Page Charges The author(s) or his/her/their company or institution will be billed $220 per each page in excess of the first ten published pages for regular papers and six published pages for correspondence items. (**NOTE: Papers accepted to IEEE TRANSACTIONS ON MULTIMEDIA in excess of 8 pages will be subject to mandatory overlength page charges.) These are mandatory page charges and the author(s) will be held responsible for them. They are not negotiable or voluntary. The author(s) signifies his willingness to pay these charges simply by submitting his/her/their manuscript to the TRANSACTIONS. The Publisher holds the right to withhold publication under any circumstance, as well as publication of the current or future submissions of authors who have outstanding mandatory page charge debt. No mandatory overlength page charges will be applied to overview articles in the Society's journals.

Color Charges. Color figures which appear in color only in the electronic (Xplore) version can be used free of charge. In this case, the figure will be printed in the hardcopy version in grayscale, and the author is responsible that the corresponding grayscale figure is intelligible. Color reproduction charges for print are the responsibility of the author. Details of the associated charges can be found on the IEEE Publications page.

Payment of fees on color reproduction is not negotiable or voluntary, and the author's agreement to publish the manuscript in these TRANSACTIONS is considered acceptance of this requirement.

*EDICS Webpages:

IEEE TRANSACTIONS ON SIGNAL PROCESSING:
http://www.signalprocessingsociety.org/publications/periodicals/tsp/TSP-EDICS/
IEEE TRANSACTIONS ON IMAGE PROCESSING:
http://www.signalprocessingsociety.org/publications/periodicals/image-processing/tip-edics/
IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE / ACM:
http://www.signalprocessingsociety.org/publications/periodicals/taslp/taslp-edics/
IEEE TRANSACTIONS ON INFORMATION, FORENSICS AND SECURITY:
http://www.signalprocessingsociety.org/publications/periodicals/forensics/forensics-edics/
IEEE TRANSACTIONS ON MULTIMEDIA:
http://www.signalprocessingsociety.org/tmm/tmm-edics/
IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING:
http://www.signalprocessingsociety.org/publications/periodicals/tci/tci-edics/
IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS:
http://www.signalprocessingsociety.org/publications/periodicals/tsipn/tsipn-edics/

2016 IEEE SIGNAL PROCESSING SOCIETY MEMBERSHIP APPLICATION

Mail to: **IEEE OPERATIONS CENTER, ATTN: Matthew Plotner, Member and Geographic Activities, 445 Hoes Lane, Piscataway, New Jersey 08854 USA**
or Fax to (732) 981-0225 (credit card payments only.)
For info call (732) 981-0060 or 1 (800) 678-IEEE or E-mail: new.membership@ieee.org

IEEE

1. PERSONAL INFORMATION

NAME AS IT SHOULD APPEAR ON IEEE MAILINGS: **SEND MAIL TO**: ☐ **Home Address** OR ☐ **Business/School Address**
If not indicated, mail will be sent to home address. Note: Enter your name as you wish it to appear on membership card and all correspondence.
PLEASE PRINT Do not exceed 40 characters or spaces per line. Abbreviate as needed. Please circle your last/surname as a key identifier for the IEEE database.

TITLE | FIRST OR GIVEN NAME | MIDDLE NAME | SURNAME/LAST NAME

HOME ADDRESS

CITY | STATE/PROVINCE | POSTAL CODE | COUNTRY

2.

Are you now or were you ever a member of IEEE? ☐ **Yes** ☐ **No**
If yes, please provide, if known:

MEMBERSHIP NUMBER_____

Grade_____**Year Membership Expired:**_____

3. BUSINESS/PROFESSIONAL INFORMATION

Company Name

Department/Division

Title/Position | Years in Current Position

Years in the Profession Since Graduation | ☐ PE State/Province

Street Address

City | State/Province | Postal Code | Country

4. EDUCATION

A baccalaureate degree from an IEEE recognized educational program assures assignment of "Member" grade. For others, additional information and references may be necessary for grade assignment.

A.

Baccalaureate Degree Received | Program/Course of Study

College/University | Campus

State/Province | Country | Mo./Yr. Degree Received

B.

Highest Technical Degree Received | Program/Course of Study

College/University | Campus

State/Province | Country | Mo./Yr. Degree Received

5.

Full signature of applicant

6. DEMOGRAPHIC INFORMATION – ALL APPLICANTS -

Date Of Birth _____ (Day / Month / Year)

☐ Male ☐ Female

7. CONTACT INFORMATION

Office Phone/Office Fax | Home Phone/Home Fax

Office E-Mail | Home E-Mail

8. 2016 IEEE MEMBER RATES

IEEE DUES Residence	16 Aug 14-28 Feb 15 Pay Full Year	1 Mar -15 Aug 15 Pay Half Year**
United States	$197.00 ☐	$98.50 ☐
Canada (incl. GST)	$173.35 ☐	$86.68 ☐
Canada (incl. HST for PEI)	$186.58 ☐	$93.29 ☐
Canada (incl. HST for Nova Scotia)	$188.05 ☐	$94.03 ☐
Canada (incl. HST for NB, NF and ON)	$185.11 ☐	$92.56 ☐
Canada (incl. GST and QST Quebec)	$188.01 ☐	$94.01 ☐
Africa, Europe, Middle East	$160.00 ☐	$80.00 ☐
Latin America	$151.00 ☐	$75.50 ☐
Asia, Pacific	$152.00 ☐	$76.00 ☐

Canadian Taxes (GST/HST): All supplies, which include dues, Society membership fees, online products and publications (except CD-ROM and DVD media), shipped to locations within Canada are subject to the GST of 5% or the HST of 12%, 13% or 15%, depending on the Province to which the materials are shipped. GST and HST do not apply to Regional Assessments. (IEEE Canadian Business Number 12563 4188 RT0001)
Value Added Tax (VAT) in the European Union: In accordance with the European Union Council Directives 2002/38/EC and 77/388/EEC amended by Council Regulation (EC)792/2002, IEEE is required to charge and collect VAT on electronic/digitized products sold to private consumers that reside in the European Union. The VAT rate applied is the EU member country standard rate where the consumer is resident. (IEEE's VAT registration number is EU826000081)
U.S. Sales Taxes: Please add applicable state and local sales and use tax on orders shipped to **Alabama, Arizona, California, Colorado, District of Columbia, Florida, Georgia, Illinois, Indiana, Kentucky, Massachusetts, Maryland, Michigan, Minnesota, Missouri, New Jersey, New Mexico, New York, North Carolina, Ohio, Oklahoma, West Virginia, Wisconsin.** Customers claiming a tax exemption must include an appropriate and properly completed tax-exemption certificate with their first order.

IEEE Signal Processing Society

2016 SPS MEMBER RATES

		16 Aug-28 Feb Pay Full Year	1 Mar-15 Aug Pay Half Year
Signal Processing Society Membership Fee*		$ 22.00 ☐	$ 11.00 ☐
Fee includes: **IEEE Signal Processing Magazine** (electronic and digital), **Inside Signal Proc. eNewsletter** (electronic) and **IEEE Signal Processing Society Content Gazette** (electronic).			
Add $17 to enhance SPS Membership and also receive:		$17.00 ☐	$ 8.50 ☐
IEEE Signal Processing Magazine (print) and **SPS Digital Library**: online access to Signal Processing Magazine, Signal Processing Letters, Journal of Selected Topics in Signal Processing, Trans. on Audio, Speech, and Language Processing, Trans. on Image Processing, Trans. on Information Forensics and Security and Trans. on Signal Processing.			
Publications available only with SPS membership:			
Signal Processing, IEEE Transactions on:	Print	$209.00 ☐	$104.50 ☐
Audio, Speech, and Lang. Proc., IEEE/ACM Trans. on:	Print	$160.00 ☐	$ 80.00 ☐
Image Processing, IEEE Transactions on:	Print	$207.00 ☐	$103.50 ☐
Information Forensics and Security, IEEE Trans. on:	Print	$179.00 ☐	$ 89.50 ☐
IEEE Journal of Selected Topics in Signal Processing:	Print	$176.00 ☐	$ 88.00 ☐
Affective Computing, IEEE Transactions on:	Electronic	$ 36.00 ☐	$ 18.00 ☐
Biomedical and Health Informatics, IEEE Journal of:	Print	$ 55.00 ☐	$ 27.50 ☐
	Electronic	$ 40.00 ☐	$ 20.00 ☐
	Print & Electronic	$ 65.00 ☐	$ 32.50 ☐
IEEE Cloud Computing	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
IEEE Trans. on Cognitive Comm. &Networking	Electronic	$ 27.00 ☐	$ 13.50 ☐
IEEE Trans. on Computational Imaging	Electronic	$ 28.00 ☐	$ 14.00 ☐
IEEE Trans. on Big Data	Electronic	$ 26.00 ☐	$ 13.00 ☐
IEEE Trans. on Molecular, Biological, & Multi-scale Communications	Electronic	$ 25.00 ☐	$ 12.50 ☐
IEEE Internet of Things Journal	Electronic	$ 26.00 ☐	$ 13.00 ☐
IEEE Trans. on Cloud Computing	Electronic	$ 43.00 ☐	$ 21.50 ☐
IEEE Trans. on Computational Social Systems	Electronic	$ 30.00 ☐	$ 15.00 ☐
IEEE Trans. on Signal & Info Proc. Over Networks	Electronic	$ 28.00 ☐	$ 14.00 ☐
IEEE Biometrics Compendium:	Online	$ 30.00 ☐	$ 15.00 ☐
Computing in Science & Engrg. Mag.:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
Medical Imaging, IEEE Transactions on:	Print	$ 74.00 ☐	$ 37.00 ☐
	Electronic	$ 53.00 ☐	$ 26.50 ☐
	Print & Electronic	$ 89.00 ☐	$ 44.50 ☐
Mobile Computing, IEEE Transactions on:	ELE/Print Abstract/CD-ROM	$ 41.00 ☐	$ 20.50 ☐
Multimedia, IEEE Transactions on:	Electronic	$ 43.00 ☐	$ 21.50 ☐
IEEE MultiMedia Magazine:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
Network Science and Engrg., IEEE Trans. on:	Electronic	$ 34.00 ☐	$ 17.00 ☐
IEEE Reviews in Biomedical Engineering:	Print	$ 25.00 ☐	$ 12.50 ☐
	Print & Electronic	$ 40.00 ☐	$ 20.00 ☐
IEEE Security and Privacy Magazine:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
IEEE Sensors Journal:	Electronic	$ 40.00 ☐	$ 20.00 ☐
Smart Grid, IEEE Transactions on:	Print	$100.00 ☐	$ 50.00 ☐
	Electronic	$ 40.00 ☐	$ 20.00 ☐
	Print & Electronic	$120.00 ☐	$ 60.00 ☐
Wireless Communications, IEEE Transactions on:	Print	$127.00 ☐	$ 63.50 ☐
	Electronic	$ 49.00 ☐	$ 24.50 ☐
	Print & Electronic	$127.00 ☐	$ 63.50 ☐
IEEE Wireless Communications Letters:	Electronic	$ 19.00 ☐	$ 9.50 ☐

*IEEE membership required or requested
Affiliate application to join SP Society only.

Amount Paid $_____

9.

IEEE Membership Affiliate Fee (See pricing in Section 8) $_____

Signal Processing Society Fees $_____

Canadian residents pay 5% GST or 13% HST
Reg. No. 125634188 on Society payment(s) & pubs only Tax $_____

AMOUNT PAID WITH APPLICATION TOTAL $_____
Prices subject to change without notice.

☐ **Check or money order enclosed Payable to IEEE on a U.S. Bank**
☐ **American Express** ☐ **VISA** ☐ **MasterCard**
☐ **Diners Club**

Exp. Date/ Mo./Yr.

Cardholder Zip Code Billing Statement Address/USA Only

Full signature of applicant using credit card | Date

10. WERE YOU REFERRED?

☐ Yes ☐ No If yes, please provide the follow information:
Member Recruiter Name:_____
IEEE Recruiter's Member Number (Required): _____

2016 IEEE SIGNAL PROCESSING SOCIETY STUDENT MEMBERSHIP APPLICATION

(Current and reinstating IEEE members joining SPS complete areas 1, 2, 8, 9.)

Mail to: **IEEE OPERATIONS CENTER, ATTN: Matthew Plotner, Member and Geographic Activities, 445 Hoes Lane, Piscataway, New Jersey 08854 USA**
or Fax to (732) 981-0225 (credit card payments only.)
For info call (732) 981-0060 or 1 (800) 678-IEEE or E-mail: new.membership@ieee.org

1. PERSONAL INFORMATION

NAME AS IT SHOULD APPEAR ON IEEE MAILINGS: **SEND MAIL TO**: ☐ **Home Address** OR ☐ **Business/School Address**
If not indicated, mail will be sent to home address. Note: Enter your name as you wish it to appear on membership card and all correspondence.
PLEASE PRINT Do not exceed 40 characters or spaces per line. Abbreviate as needed. Please circle your last/surname as a key identifier for the IEEE database.

TITLE | FIRST OR GIVEN NAME | MIDDLE NAME | SURNAME/LAST NAME

HOME ADDRESS

CITY | STATE/PROVINCE | POSTAL CODE | COUNTRY

2.

Are you now or were you ever a member of IEEE? ☐ **Yes** ☐ **No**
If yes, please provide, if known:

MEMBERSHIP NUMBER_____

Grade_____**Year Membership Expired:**_____

3. BUSINESS/PROFESSIONAL INFORMATION

Company Name

Department/Division

Title/Position | Years in Current Position

Years in the Profession Since Graduation | ☐ PE State/Province

Street Address

City | **State/Province** | **Postal Code** | **Country**

4. EDUCATION

A baccalaureate degree from an IEEE recognized educational program assures assignment of "Member" grade. For others, additional information and references may be necessary for grade assignment.

A.

Baccalaureate Degree Received | Program/Course of Study

College/University | Campus

State/Province | Country | Mo./Yr. Degree Received

B.

Highest Technical Degree Received | Program/Course of Study

College/University | Campus

State/Province | Country | Mo./Yr. Degree Received

5.

Full signature of applicant

6. DEMOGRAPHIC INFORMATION – ALL APPLICANTS -

Date Of Birth _____ ☐ Male ☐ Female
Day Month Year

7. CONTACT INFORMATION

Office Phone/Office Fax | Home Phone/Home Fax

Office E-Mail | Home E-Mail

8. 2016 IEEE STUDENT MEMBER RATES

IEEE DUES	16 Aug 14-28 Feb 15	1 Mar -15 Aug 15
Residence	Pay Full Year	Pay Half Year**
United States	$32.00 ☐	$16.00 ☐
Canada (incl. GST)	$33.60 ☐	$16.80 ☐
Canada (incl. HST for NB, NF, and ON)	$36.16 ☐	$18.08 ☐
Canada (incl. HST for Nova Scotia)	$36.80 ☐	$18.40 ☐
Canada (incl. HST for PEI)	$36.48 ☐	$18.24 ☐
Canada (incl. GST and QST Quebec)	$36.79 ☐	$18.40 ☐
Africa, Europe, Middle East, Latin America, Asia, Pacific	$27.00 ☐	$13.50 ☐

Canadian Taxes (GST/HST): All supplies, which include dues, Society membership fees, online products and publications (except CD-ROM and DVD media), shipped to locations within Canada are subject to the GST of 5% or the HST of 12%,13% or 15%, depending on the Province to which the materials are shipped. GST and HST do not apply to Regional Assessments. (IEEE Canadian Business Number 12563 4188 RT0001)

Value Added Tax (VAT) in the European Union: In accordance with the European Union Council Directives 2002/38/EC and 77/388/EEC amended by Council Regulation (EC)792/2002, IEEE is required to charge and collect VAT on electronic/digitized products sold to private consumers that reside in the European Union. The VAT rate applied is the EU member country standard rate where the consumer is resident. (IEEE's VAT registration number is EU826000081)

U.S. Sales Taxes: Please add applicable state and local sales and use tax on orders shipped to **Alabama, Arizona, California, Colorado, District of Columbia, Florida, Georgia, Illinois, Indiana, Kentucky, Massachusetts, Maryland, Michigan, Minnesota, Missouri, New Jersey, New Mexico, New York, North Carolina, Ohio, Oklahoma, West Virginia, Wisconsin.** Customers claiming a tax exemption must include an appropriate and properly completed tax-exemption certificate with their first order.

2016 SPS STUDENT MEMBER RATES

	16 Aug-28 Feb Pay Full Year	1 Mar-15 Aug Pay Half Year
Signal Processing Society Membership Fee*	$11.00 ☐	$ 5.50 ☐

Fee includes: **IEEE Signal Processing Magazine** (electronic and digital), **Inside Signal Processing eNewsletter** (electronic) and **IEEE Signal Processing Society Content Gazette** (electronic).

Add $8 to enhance SPS Membership and also receive: $ 9.00 ☐ $ 4.50 ☐
IEEE Signal Processing Society Magazine (print) and **SPS Digital Library**: online access to Signal Processing Magazine, Signal Processing Letters, Journal of Selected Topics in Signal Processing, Trans. on Audio, Speech, and Language Processing, Trans. on Image Processing, Trans. on Information Forensics and Security and Trans. on Signal Processing.

Publications available only with SPS membership:

Signal Processing, IEEE Transactions on:	Print	$105.00 ☐	$ 52.50 ☐
Audio, Speech, and Lang. Proc., IEEE/ACM Trans. on:	Print	$ 80.00 ☐	$ 40.00 ☐
Image Processing, IEEE Transactions on:	Print	$104.00 ☐	$ 52.00 ☐
Information Forensics and Security, IEEE Trans. on:	Print	$ 90.00 ☐	$ 45.00 ☐
IEEE Journal of Selected Topics in Signal Processing:	Print	$ 88.00 ☐	$ 44.00 ☐
Affective Computing, IEEE Transactions on:	Electronic	$ 18.00 ☐	$ 9.00 ☐
Biomedical and Health Informatics, IEEE Journal of:	Print	$ 28.00 ☐	$ 14.00 ☐
	Electronic	$ 20.00 ☐	$ 10.00 ☐
	Print & Electronic	$ 33.00 ☐	$ 16.50 ☐
IEEE Cloud Computing	Electronic and Digital	$ 20.00 ☐	$ 10.00 ☐
IEEE Trans. on Cognitive Comm. &Networking	Electronic	$ 14.00 ☐	$ 7.00 ☐
IEEE Trans. on Computational Imaging	Electronic	$ 14.00 ☐	$ 7.00 ☐
IEEE Trans. on Big Data	Electronic	$ 13.00 ☐	$ 6.50 ☐
IEEE Trans. on Molecular, Biological, & Multi-Scale Communications	Electronic	$ 13.00 ☐	$ 6.50 ☐
IEEE Internet of Things Journal	Electronic	$ 13.00 ☐	$ 6.50 ☐
IEEE Trans. on Cloud Computing	Electronic	$ 22.00 ☐	$ 11.00 ☐
IEEE Trans. on Computational Social Systems	Electronic	$ 15.00 ☐	$ 7.50 ☐
IEEE Trans. on Signal & Info Proc. Over Networks	Electronic	$ 14.00 ☐	$ 7.00 ☐
IEEE Biometrics Compendium:	Online	$ 15.00 ☐	$ 7.50 ☐
Computing in Science & Engrg. Mag.:	Electronic and Digital	$ 20.00 ☐	$ 10.00 ☐
	Print	$ 35.00 ☐	$ 17.50 ☐
Medical Imaging, IEEE Transactions on:	Print	$ 37.00 ☐	$ 18.50 ☐
	Electronic	$ 27.00 ☐	$ 13.50 ☐
	Print & Electronic	$ 45.00 ☐	$ 22.50 ☐
Mobile Computing, IEEE Transactions on:	ELE/Print Abstract/CD-ROM	$ 21.00 ☐	$ 10.50 ☐
Multimedia, IEEE Transactions on:	Electronic	$ 22.00 ☐	$ 11.00 ☐
IEEE MultiMedia Magazine:	Electronic and Digital	$ 20.00 ☐	$ 10.00 ☐
	Print	$ 35.00 ☐	$ 17.50 ☐
Network Science and Engrg., IEEE Trans. on:	Electronic	$ 17.00 ☐	$ 8.50 ☐
IEEE Reviews in Biomedical Engineering:	Print	$ 13.00 ☐	$ 6.50 ☐
	Print & Electronic	$ 20.00 ☐	$ 10.00 ☐
IEEE Security and Privacy Magazine:	Electronic and Digital	$ 20.00 ☐	$ 10.00 ☐
	Print	$ 35.00 ☐	$ 17.50 ☐
IEEE Sensors Journal:	Electronic	$ 20.00 ☐	$ 10.00 ☐
Smart Grid, IEEE Transactions on:	Print	$ 50.00 ☐	$ 25.00 ☐
	Electronic	$ 20.00 ☐	$ 10.00 ☐
	Print & Electronic	$ 60.00 ☐	$ 30.00 ☐
Wireless Communications, IEEE Transactions on:	Print	$ 64.00 ☐	$ 32.00 ☐
	Electronic	$ 25.00 ☐	$ 12.50 ☐
	Print & Electronic	$ 64.00 ☐	$ 32.00 ☐
IEEE Wireless Communications Letters:	Electronic	$ 10.00 ☐	$ 5.00 ☐

*IEEE membership required or requested
Affiliate application to join SP Society only.

Amount Paid $_____

9.

IEEE Membership Fee (See pricing in Section 8) $_____
Signal Processing Society Fees $_____

Canadian residents pay 5% GST or 13% HST
Reg. No. 125634188 on Society payment(s) & pubs only Tax $_____

AMOUNT PAID WITH APPLICATION TOTAL $_____
Prices subject to change without notice.

☐ **Check or money order enclosed Payable to IEEE on a U.S. Bank**
☐ **American Express** ☐ **VISA** ☐ **MasterCard** ☐ **Diners Club**

Exp. Date/ Mo./Yr.

Cardholder Zip Code Billing Statement Address/USA Only

Full signature of applicant using credit card | Date

10. WERE YOU REFERRED?

☐ Yes ☐ No If yes, please provide the follow information:
Member Recruiter Name:_____
IEEE Recruiter's Member Number (Required): _____

IEEE Signal Processing Society

2016 IEEE SIGNAL PROCESSING SOCIETY AFFILIATE MEMBERSHIP APPLICATION

Mail to: **IEEE OPERATIONS CENTER, ATTN: Matthew Plotner, Member and Geographic Activities, 445 Hoes Lane, Piscataway, New Jersey 08854 USA**
or Fax to (732) 981-0225 (credit card payments only.)
For info call (732) 981-0060 or 1 (800) 678-IEEE or E-mail: new.membership@ieee.org

1. PERSONAL INFORMATION

NAME AS IT SHOULD APPEAR ON IEEE MAILINGS: **SEND MAIL TO**: ☐ **Home Address** OR ☐ **Business/School Address**
If not indicated, mail will be sent to home address. Note: Enter your name as you wish it to appear on membership card and all correspondence.
PLEASE PRINT Do not exceed 40 characters or spaces per line. Abbreviate as needed. Please circle your last/surname as a key identifier for the IEEE database.

TITLE | FIRST OR GIVEN NAME | MIDDLE NAME | SURNAME/LAST NAME

HOME ADDRESS

CITY | STATE/PROVINCE | POSTAL CODE | COUNTRY

2

Are you now or were you ever a member of IEEE? ☐ **Yes** ☐ **No**
If yes, please provide, if known:

MEMBERSHIP NUMBER_____

Grade_____**Year Membership Expired:**_____

3. BUSINESS/PROFESSIONAL INFORMATION

Company Name

Department/Division

Title/Position | Years in Current Position

Years in the Profession Since Graduation | ☐ PE State/Province

Street Address

City | State/Province | Postal Code | Country

4. EDUCATION

A baccalaureate degree from an IEEE recognized educational program assures assignment of "Member" grade. For others, additional information and references may be necessary for grade assignment.

A.

Baccalaureate Degree Received | Program/Course of Study

College/University | Campus

State/Province | Country | Mo./Yr. Degree Received

B.

Highest Technical Degree Received | Program/Course of Study

College/University | Campus

State/Province | Country | Mo./Yr. Degree Received

5.

Full signature of applicant

6. DEMOGRAPHIC INFORMATION – ALL APPLICANTS -

Date Of Birth _____ ☐ Male ☐ Female
Day Month Year

7. CONTACT INFORMATION

Office Phone/Office Fax | Home Phone/Home Fax

Office E-Mail | Home E-Mail

8. 2016 IEEE AFFILIATE RATES

IEEE DUES Residence	16 Aug 14-28 Feb 15 Pay Full Year	1 Mar -15 Aug 15 Pay Half Year**
United States	$73.50 ☐	$36.75 ☐
Canada (incl. GST)	$77.18 ☐	$38.59 ☐
Canada (incl. HST for PEI)	$83.79 ☐	$41.89 ☐
Canada (incl. HST for Nova Scotia)	$84.58 ☐	$42.26 ☐
Canada (incl. HST for NB, NF and ON)	$83.06 ☐	$41.53 ☐
Canada (incl. GST and QST Quebec)	$84.51 ☐	$42.25 ☐
Africa, Europe, Middle East	$73.50 ☐	$36.75 ☐
Latin America	$73.50 ☐	$36.75 ☐
Asia, Pacific	$73.50 ☐	$36.75 ☐

Canadian Taxes (GST/HST): All supplies, which include dues, Society membership fees, online products and publications (except CD-ROM and DVD media), shipped to locations within Canada are subject to the GST of 5% or the HST of 13%, 14% or 15%, depending on the Province to which the materials are shipped. GST and HST do not apply to Regional Assessments. (IEEE Canadian Business Number 12563 4188 RT0001)

Value Added Tax (VAT) in the European Union: In accordance with the European Union Council Directives 2002/38/EC and 77/388/EEC amended by Council Regulation (EC)792/2002, IEEE is required to charge and collect VAT on electronic/digitized products sold to private consumers that reside in the European Union. The VAT rate applied is the EU member country standard rate where the consumer is resident. (IEEE's VAT registration number is EU826000081)

U.S. Sales Taxes: Please add applicable state and local sales and use tax on orders shipped to **Alabama, Arizona, California, Colorado, District of Columbia, Florida, Georgia, Illinois, Indiana, Kentucky, Massachusetts, Maryland, Michigan, Minnesota, Missouri, New Jersey, New Mexico, New York, North Carolina, Ohio, Oklahoma, West Virginia, Wisconsin.** Customers claiming a tax exemption must include an appropriate and properly completed tax-exemption certificate with their first order.

IEEE Signal Processing Society

2016 SPS MEMBER RATES

		16 Aug-28 Feb Pay Full Year	1 Mar-15 Aug Pay Half Year
Signal Processing Society Membership Fee*		$ 22.00 ☐	$ 11.00 ☐
Fee includes: **IEEE Signal Processing Magazine** (electronic and digital), **Inside Signal Proc. eNewsletter** (electronic) and **IEEE Signal Processing Society Content Gazette** (electronic).			
Add $17 to enhance SPS Membership and also receive:		$17.00 ☐	$ 8.50 ☐
IEEE Signal Processing Magazine (print) and **SPS Digital Library**: online access to Signal Processing Magazine, Signal Processing Letters, Journal of Selected Topics in Signal Processing, Trans. on Audio, Speech, and Language Processing, Trans. on Image Processing, Trans. on Information Forensics and Security and Trans. on Signal Processing.			
Publications available only with SPS membership:			
Signal Processing, IEEE Transactions on:	Print	$209.00 ☐	$104.50 ☐
Audio, Speech, and Lang. Proc., IEEE/ACM Trans. on:	Print	$160.00 ☐	$ 80.00 ☐
Image Processing, IEEE Transactions on:	Print	$207.00 ☐	$103.50 ☐
Information Forensics and Security, IEEE Trans. on:	Print	$179.00 ☐	$ 89.50 ☐
IEEE Journal of Selected Topics in Signal Processing:	Print	$176.00 ☐	$ 88.00 ☐
Affective Computing, IEEE Transactions on:	Electronic	$ 36.00 ☐	$ 18.00 ☐
Biomedical and Health Informatics, IEEE Journal of:	Print	$ 55.00 ☐	$ 27.50 ☐
	Electronic	$ 40.00 ☐	$ 20.00 ☐
	Print & Electronic	$ 65.00 ☐	$ 32.50 ☐
IEEE Cloud Computing	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
IEEE Trans. on Cognitive Comm. &Networking	Electronic	$ 27.00 ☐	$ 13.50 ☐
IEEE Trans. on Computational Imaging	Electronic	$ 28.00 ☐	$ 14.00 ☐
IEEE Trans. on Big Data	Electronic	$ 26.00 ☐	$ 13.00 ☐
IEEE Trans. on Molecular, Biological, & Multi-scale Communications	Electronic	$ 25.00 ☐	$ 12.50 ☐
IEEE Internet of Things Journal	Electronic	$ 26.00 ☐	$ 13.00 ☐
IEEE Trans. on Cloud Computing	Electronic	$ 43.00 ☐	$ 21.50 ☐
IEEE Trans. on Computational Social Systems	Electronic	$ 30.00 ☐	$ 15.00 ☐
IEEE Trans. on Signal & Info Proc. Over Networks	Electronic	$ 28.00 ☐	$ 14.00 ☐
IEEE Biometrics Compendium:	Online	$ 30.00 ☐	$ 15.00 ☐
Computing in Science & Engrg. Mag.:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
Medical Imaging, IEEE Transactions on:	Print	$ 74.00 ☐	$ 37.00 ☐
	Electronic	$ 53.00 ☐	$ 26.50 ☐
	Print & Electronic	$ 89.00 ☐	$ 44.50 ☐
Mobile Computing, IEEE Transactions on:	ELE/Print Abstract/CD-ROM	$ 41.00 ☐	$ 20.50 ☐
Multimedia, IEEE Transactions on:	Electronic	$ 43.00 ☐	$ 21.50 ☐
IEEE MultiMedia Magazine:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
Network Science and Engrg., IEEE Trans. on:	Electronic	$ 34.00 ☐	$ 17.00 ☐
IEEE Reviews in Biomedical Engineering:	Print	$ 25.00 ☐	$ 12.50 ☐
	Electronic	$ 25.00 ☐	$ 12.50 ☐
	Print & Electronic	$ 40.00 ☐	$ 20.00 ☐
IEEE Security and Privacy Magazine:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
IEEE Sensors Journal:	Electronic	$ 40.00 ☐	$ 20.00 ☐
Smart Grid, IEEE Transactions on:	Print	$100.00 ☐	$ 50.00 ☐
	Electronic	$ 40.00 ☐	$ 20.00 ☐
	Print & Electronic	$120.00 ☐	$ 60.00 ☐
Wireless Communications, IEEE Transactions on:	Print	$127.00 ☐	$ 63.50 ☐
	Electronic	$ 49.00 ☐	$ 24.50 ☐
	Print & Electronic	$127.00 ☐	$ 63.50 ☐
IEEE Wireless Communications Letters:	Electronic	$ 19.00 ☐	$ 9.50 ☐

*IEEE membership required or requested
Affiliate application to join SP Society only.

Amount Paid $_____

9.

IEEE Membership Affiliate Fee (See pricing in Section 8) $_____

Signal Processing Society Fees $_____

Canadian residents pay 5% GST or 13% HST
Reg. No. 125634188 on Society payment(s) & pubs only Tax $_____

AMOUNT PAID WITH APPLICATION TOTAL $_____

Prices subject to change without notice.

☐ **Check or money order enclosed Payable to IEEE on a U.S. Bank**
☐ **American Express** ☐ **VISA** ☐ **MasterCard**
☐ **Diners Club**

Exp. Date/ Mo./Yr.

Cardholder Zip Code Billing Statement Address/USA Only

Full signature of applicant using credit card | Date

10. WERE YOU REFERRED?

☐ Yes ☐ No If yes, please provide the follow information:
Member Recruiter Name:_____
IEEE Recruiter's Member Number (Required): _____

IEEE Signal Processing Society®

IEEE