Introduction to the Special Section on Large-Scale Optimization for Audio, Speech, and Language Processing

PATTERN recognition in audio, speech, and language processing requires estimation of parameters in statistical models via optimization criteria. Formulation of these optimization models is far from straightforward. For example, likelihood criteria usually are inadequate if the training data do not represent all possible variations in patterns. Significant progress in pattern recognition has been achieved by introducing discrimination criteria for training, but overtraining remains a danger. An important formulation device is a regularization term in the optimization objective that captures the prior information available about parameter values and their relationships.

The challenges are not over once an optimization problem has been formulated. Such problems are large in scale (requiring a great deal of data to define them) and have highly specialized structures. Conventional optimization algorithms must be adapted significantly to meet the challenges posed by these properties. Over the years, speech processing researchers have developed and applied such specialized methods as Baum-Welch, extended Baum-Welch, Rprop, and GIS.

To summarize the state of the art and to gain additional insights into formulation and solution of optimization models in speech and language processing, the IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING (TASL) has created this special issue. Five guest editors, assisted by a team of dedicated reviewers, were tasked with selecting the papers for this issue. An overview paper contributed by the guest editors was handled by Signal Processing Society VP Publications Mari Ostendorf (who was not a guest editor), and went through the regular TASLP review process. We summarize the six papers in this special issue below.

The overview paper "Optimization Algorithms and Applications for Speech and Language Processing," by Stephen J. Wright, Dimitri Kanevsky, Li Deng, Xiaodong He, Georg Heigold, and Haizhou Li, reviews the use of optimization formulations and algorithms in many aspects of the field, including machine translation, speaker/language recognition, and automatic speech recognition. Several approaches developed in the speech and language processing communities are described in a way that makes them more recognizable as optimization procedures.

An overview of how certain Hessian operations can be conveniently represented in a wide class of matrix optimization problems is presented in "Second Order Methods for Optimizing Convex Matrix Functions and Sparse Covariance Clustering," by Gillian Chin, Jorge Nocedal, Peder A. Olsen, and Steven J. Rennie. The authors point out that second-order information is accessible in many matrix optimization problems, and can be incorporated into optimization algorithms via efficient computation of Hessians and Hessian-vector products. Extensive numerical results illustrate the behavior of these novel algorithms.

Another convex optimization approach is the subject of the paper "A Difference of Convex Functions Approach to Large-Scale Log-Linear Model Estimation," by Theodoros Tsiligkaridis, Etienne Marcheret, and Vaibhava Goel. These authors represent a rational function of mixtures of exponentials as a difference of convex functions, for which the authors construct convex auxiliary functions amenable to optimization. They present convergence analysis of their algorithms and apply them to the task of optimizing a cross-entropy objective function for audio-frame classification.

Different kinds of optimization problems arise in deep learning procedures, such as the Deep Belief Networks (DBNs) that have had tremendous success for large-vocabulary continuous speech recognition tasks. The paper "Optimization Techniques to Improve Training Speed of Deep Belief Networks for Large Speech Tasks" by Tara N. Sainath, Brian Kingsbury, Hagen Soltau, and Bhuvana Ramabhadran explores a variety of optimization techniques to improve DBN training speed. These include parallelization of the gradient computation during cross-entropy and sequence training, and reducing the number of parameters in the network using a low-rank matrix factorization.

The paper "Active-Set Newton Algorithm for Overcomplete Non-Negative Representations of Audio" by Tuomas Virtanen, Jort F. Gemmeke, and Bhiksha Raj represents audio signals as linear combinations of atoms from large overcomplete dictionaries. It describes a second-order optimization algorithm for estimating non-negative weights given dictionaries with magnitude spectra observations. This algorithm is most beneficial when the overcomplete dictionary is large, and it is shown to converge to a sparse weight vector without imposing explicit sparsity constraints.

In "Large Vocabulary Speech Recognition on Parallel Architectures," authors Patrick Cardinal, Pierre Dumouchel, and Gilles Boulianne argue that while the speed of modern processors has remained constant over the last few years, integration capacity continues to follow Moore's law. Therefore, to be scalable, applications must be parallelized. The paper describes a parallel implementation of the A* decoder on a four-core processor with GPU, which produces a significant performance boost when compared to the Viterbi beam search.

In closing, we thank all the authors who submitted papers to this special issue for the hard work they put in—both those with accepted articles and those whose papers were not accepted. Our acceptance decisions were based not only on quality but also by our perception of fit to the theme, and by our desire to maintain a balance between different aspects of the theme. We thank all the reviewers for providing valuable feedback that improved the articles significantly. Finally, we are grateful to Li Deng, TASL's Editor-in-Chief, as well as to Mari Ostendorf for their support

Digital Object Identifier 10.1109/TASL.2013.2283631

and guidance throughout the process of preparing and finalizing these articles.

DIMITRI KANEVSKY, *Lead Guest Editor*
IBM Research
Yorktown Heights, NY 10598 USA

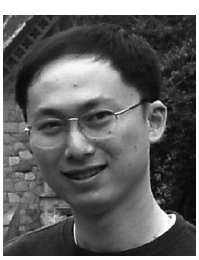
XIAODONG HE, *Guest Editor*
Microsoft Research
Redmond, WA 98052-6399 USA

GEORG HEIGOLD, *Guest Editor*
Google Research
Mountain View, CA 94043 USA

HAIZHOU LI, *Guest Editor*
Institute for Infocomm Research
Singapore, 138632

STEPHEN J. WRIGHT, *Guest Editor*
University of Wisconsin-Madison
Madison, WI 53706 USA

Dimitri Kanevsky received the M.S. degree in 1974 and the Ph.D. degree in 1977, both from the Moscow State University, Russia. Dimitri Kanevsky joined the Human Ability and Accessibility department at IBM T. J. Watson Research Center in 2013. Prior to joining the Human Ability and Accessibility department Dimitri was a Research staff member in the Speech and Language Algorithms Department, where he was responsible for developing the first Russian automatic speech recognition system, as well as key projects for embedding speech recognition in automobiles and broadcast transcription systems. Prior to working at IBM, he worked at a number of prestigious centers for higher mathematics, including the Max Planck Institute in Germany and the Institute for Advanced Studies in Princeton. He currently holds 178 US patents and was granted the title of Master Inventor IBM in 2002 , 2005 and 2010. His conversational biometrics based security patent was recognized by MIT, Technology Review, as one of five most influential patents for 2003. His contributions on Extended Baum-Welch algorithms for speech, embedding speech recognition in automobiles and his work on conversational biometrics were recognized as science accomplishments in 2002, 2004 and 2008 by the Director of Research at IBM. He organized a special session on Large Scale Optimization at ICASSP 2012 in Japan. He also organized a NIPS workshop on log-linear models in 2012 in Lake Tahoe, Nevada. In 2012 Dimitri was honored at the White House as a Champion of Change for his efforts to advance access to science, technology, engineering, and math. In 2012 he received Tan Chin Tuan Exchange Fellowship award from Nanyang Technological University for development sparse optimization methods and its application in speech.



Xiaodong He (M'03–SM'08) received the B.S. degree from Tsinghua University in 1996, the M.S. degree from the Chinese Academy of Sciences in 1999, and the Ph.D. degree from the University of Missouri-Columbia in 2003. He joined Microsoft in 2003, where he is currently a Researcher in the Conversational Systems Research Center of Microsoft Research Redmond. He also has been an Affiliate Professor in the Department of Electrical Engineering, University of Washington, Seattle, since 2012. His research interests include speech recognition, spoken language understanding, machine translation, natural language processing, information retrieval, and machine learning. He has authored/co-authored one book and 60 technical papers in these areas and is the speaker of the tutorial on Speech Translation at ICASSP13. In benchmark evaluations, he and his colleagues developed the MSR-NRC-SRI entry and the MSR entry which obtained No. 1 place in the 2008 NIST MT Evaluation and No. 1 place in the 2011 IWSLT Evaluation, all in Chinese-to-English translation, respectively. Dr. He serves as Associate Editor of *IEEE Signal Processing Magazine* since 2011 and Lead Guest Editor of IEEE J-STSP special issue on Statistical Learning Methods for Speech and Language Processing in 2010. He has served in the organizing committee of ICASSP13 as the Chair of Special Sessions, and in the program committees of various speech and language processing conferences.

Georg Heigold received the Diplom degree in physics from ETH Zurich, Switzerland, in 2000. He was a Software Engineer at De La Rue, Berne, Switzerland, from 2000 to 2003. From 2004 to 2010, he was with the Computer Science Department, RWTH Aachen University, Aachen, University. Since 2010, he has been a Research Scientist at Google, Mountain View, CA. His research interests include automatic speech recognition, discriminative training, and log-linear modeling. He organized a special session on Large Scale Optimization at ICASSP 2012 in Japan. He also organized a NIPS workshop on log-linear models in 2012 in Lake Tahoe, Nevada.

Haizhou Li (M91–SM01) received the B.Sc., M.Sc., and Ph.D. degree in electrical and electronic engineering from South China University of Technology, Guangzhou, China in 1984, 1987, and 1990 respectively. Dr. Li is currently the Principal Scientist and Department Head of Human Language Technology in the Institute for Infocomm Research, Singapore. Dr. Li has worked on speech and language technology in academia and industry since 1988. He taught in the University of Hong Kong (1988–1990), South China University of Technology (1990–1994), and Nanyang Technological University (2006-present). He was a Visiting Professor at CRIN in France (1994–1995). Dr. Li was appointed as Research Manager at the Apple-ISS Research Centre (1996–1998), Research Director in Lernout & Hauspie Asia Pacific (1999–2001), and Vice President in InfoTalk Corp., Ltd., (2001–2003). His current research interests include automatic speech recognition, speaker and language recognition, and natural language processing. Dr. Li has served as an Associate Editor of IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING, ACM Transactions on Speech and Language Processing, Computer Speech and Language, and PROCEEDINGS OF THE IEEE. He is an elected Member of IEEE Speech and Language Processing Technical Committee (2013-2015), the Vice President of the International Speech Communication Association (20132014), the President-Elect of Asia Pacific Signal and Information Processing Association (2013–2014). He was appointed the General Chair of ACL 2012 and INTERSPEECH 2014. Dr. Li was a recipient of the National Infocomm Award 2002 and the President's Technology Award 2013 in Singapore. He was named one the two Nokia Visiting Professors in 2009 by the Nokia Foundation.

Stephen J. Wright received the B. Sc. (Hons) in 1981 and the Ph.D. degree in 1984, both from the University of Queensland, Australia. After holding positions at North Carolina State University, Argonne National Laboratory, and the University of Chicago, he joined the Computer Sciences Department at the University of Wisconsin-Madison as a professor in 2001. His research interests include theory, algorithms, and applications of computational optimization. Dr Wright has been Chair of the Mathematical Optimization Society and has served on the editorial boards of SIAM Review, Mathematical Programming, Series A and B, the SIAM Journal on Optimization, the SIAM Journal on Scientific Computing, and other journals. He also serves on the Board of Trustees of the Society for Industrial and Applied Mathematics (SIAM) and is a Fellow of SIAM. With coauthors Mario Figueiredo and Robert Nowak, he won the Best Paper prize from the IEEE Signal Processing Society in 2011.