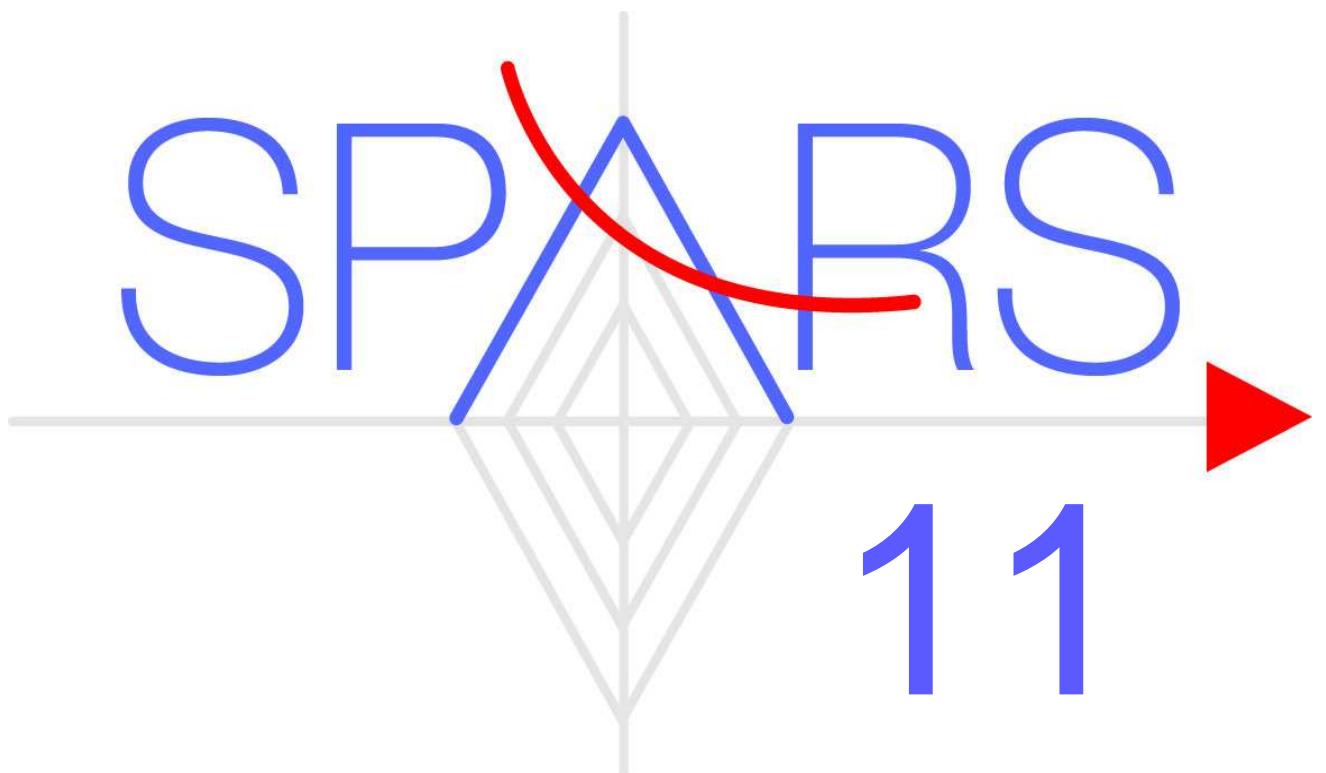


4th Workshop on Signal Processing with Adaptive Sparse Structured Representations

June 27-30, 2011

Edinburgh



Proceedings



Workshop: Signal Processing with Adaptive Sparse Structured Representations

June 27-30, 2011

Edinburgh, UK

Foreword

It is with great pleasure that we welcome you to the beautiful city of Edinburgh, the historical home of sampling theory¹, for the 4th Workshop on Signal Processing with Adaptive Sparse Structured Representations: SPARS '11.

Sparse models have already been applied with outstanding success in signal and image processing as well as in machine learning. In machine learning they provide a powerful method for model order selection within regression and classification problems (e.g. Lasso). While in signal processing they have led to many algorithms for de-noising, compression (e.g. jpeg2000), de-blurring and more.

In particular, these techniques are at the core of compressed sensing, an emerging approach which proposes a radically new viewpoint on signal acquisition compared to Shannon sampling. There are also strong connections between sparse signal models and kernel methods, whose algorithmic success on large datasets relies deeply on sparsity.

The aim of this workshop is to bring together different work in this area from the applied mathematics, signal processing and machine learning communities. Both theoretical developments and practical applications will be discussed. Although each community is generally aware of the others' work we hope that such a meeting will provide an excellent opportunity for dialog between the communities.

As with any workshop of this type there is a great deal of work required to make it happen. We would like to take this opportunity to thank the International Centre for Mathematical Sciences (ICMS) for not only managing the workshop for us but also for substantially funding it – therefore making the extremely low registration fees possible. We would also like to thank our other sponsors: the UK Engineering and Science Research Council (EPSRC), and the London Mathematics Society (LMS) for financial assistance; and INRIA Rennes for the use of their website for the abstract submissions.

The other group of people without whom the conference would not happen is our team of PhD students and post-doctoral researchers at the Edinburgh Centre for Compressed Sensing (E-CoS). Beyond their usual roles in E-CoS they have been assigned various unenviable tasks to make sure that the workshop runs as smoothly as possible. For this we thank them.

Our final thanks go to our magnificent line up of plenary speakers. Despite high demand we have been able to secure this world leading set of speakers from across the globe.

We sincerely hope that everyone will enjoy this workshop and that it will prove to be both enlightening and fun.

Coralia Cartis
Mike Davies
Jared Tanner

¹E. T. Whittaker, “On the Functions Which are Represented by the Expansions of the Interpolation Theory”, Proc. Royal Soc. Edinburgh, Sec. A, vol.35, pp. 181–194, 1915

Committees

Organisers

Coralia Cartis - School of Mathematics, University of Edinburgh, UK

Mike Davies - School of Engineering & Electronics, University of Edinburgh, UK

Jared Tanner - School of Mathematics, University of Edinburgh, UK

Steering Committee

Laurent Daudet - Universit Paris Diderot, France

Stéphane Canu - INSA de Rouen, France

Mike Davies - University of Edinburgh, UK

Jalal Fadili - GREYC-ENSICAEN, France

Remi Gribonval - Centre de Recherche INRIA Rennes,

France

Mark Plumbley - Queen Mary University of London, UK

Scott Rickard - UCD CASL & University College Dublin, Ireland

Jared Tanner - University of Edinburgh, UK

Bruno Torrésani - LATP, CMI, Université de Provence, France

Pierre Vandergheynst - Ecole Polytechnique Fédérale de Lausanne, Switzerland

Technical Program Committee (abstract referees)

Coralia Cartis - University of Edinburgh, UK

Laurent Daudet - Universit Paris Diderot, France

Mike Davies - University of Edinburgh, UK

Michael Elad - Technion, Israel

Jalal Fadili - GREYC-ENSICAEN, France

Mario Figueiredo - Instituto Superior Tecnico, Portugal

Remi Gribonval - Centre de Recherche INRIA Rennes, France

Gabriel Peyré - Université Paris-Dauphine, France

Justin Romberg - Georgia Tech, USA

Jared Tanner - University of Edinburgh, UK

Bruno Torrésani - LATP, CMI, Université de Provence, France

Joel Tropp - California Institute of Technology, USA

Edinburgh Compressed Sensing Group

Jared Tanner - University of Edinburgh, UK

Mike Davies - University of Edinburgh, UK

Coralia Cartis - University of Edinburgh, UK

Peter Richtarik - University of Edinburgh, UK

Natalia Bochkina - University of Edinburgh, UK

Paolo Favaro - Heriot-Watt University & University of Edinburgh, UK

Mehrdad Yaghoobi - University of Edinburgh, UK

Martin Lotz - University of Edinburgh, UK

Gabriel Rilling - University of Edinburgh, UK

Michael Lexa - University of Edinburgh, UK

Fabien Millioz - University of Edinburgh, UK

Pavel Zhlobich - University of Edinburgh, UK

Andrew Thompson - University of Edinburgh, UK

Bubacarr Bah - University of Edinburgh, UK

Ke Wei - University of Edinburgh, UK

Chunli Guo - University of Edinburgh, UK

Shaun Kelly - University of Edinburgh, UK

Martin Takac - University of Edinburgh, UK

Jeffrey Blanchard - Grinnell College, USA

Thomas Blumensath - University of Oxford, UK

Technical Program

	Monday 27	Tuesday 28	Wednesday 29	Thursday 30
09:00-09:50	Registration	Francis Bach	David L. Donoho	Joel A. Tropp
09:50-10:20	Welcome (10:00)	Coffee Break	Coffee Break	Coffee Break
10:20-11:20	Yi Ma	#5 Classification & Clustering #6 Structured Sparsity 1	#11 CS Theory #12 Sparsity Applications	#17 Dictionary Learning #18 A to D Conversions
11:20-11:50	Coffee Break	Break	Break	Break
11:50-12:50	#1 Sparsity Theory #2 SAR Imaging	#7 Random Matrix Theory #8 Structured Sparsity 2	#13 Generalized CS #14 Estimation & Detection	#19 Low Dimensional & Analysis Sparse Model Learning #20 Performance Evaluations
12:50-14:30	Lunch	Lunch	Lunch	Lunch
14:30-15:20	David J. Brady	Remi Gribonval	Martin Vetterli	Stephen Wright
15:20-16:00	Posters A & Coffee	Posters A & Coffee	Posters B & Coffee	Posters B & Coffee
16:00-17:00	#3 Medical Imaging #4 Sparse Approx. & CS Algorithms 1	#9 Analysis Framework #10 Dynamical & Time-varying Systems	#15 Generalized Sampling Techniques #16 Sparse Approx. & CS Algorithms 2	#21 PCA/ICA/BSS #22 Sparse Filter Design
Evening Wine	Reception, (18:00-21:00)		Whisky Tasting Excursion, £20 cost (18:30-22:00)	

Note: Plenary talks and sessions with odd numbers take place in the “Main Auditorium” (in the Queen Mother Conference Centre) and sessions with even numbers take place in the “Great Hall”.

Abstracts

Plenary Talks: Monday 27

10:20-11:20

<i>TILT and RASL: For Low-Rank Structures in Images and Data</i>	10
Yi Ma	

14:30-15:20

<i>Coding for Multiplex Optical Imaging</i>	12
David J. Brady	

Plenary Talks: Tuesday 28

09:00-09:50

<i>Strutured Sparsity-Inducing Norms through Submodular Functions</i>	13
Francis Bach	

14:30-15:20

<i>Sparsity & Co.: An Overview of Analysis vs Synthesis in Low-Dimensional Signal Models</i>	14
Rémi Gribonval	

Plenary Talks: Wednesday 29

09:00-09:50

<i>Precise Optimality Results in Compressed Sensing</i>	15
David L. Donoho	

14:30-15:20

<i>Sampling in the Age of Sparsity</i>	16
Martin Vetterli	

Plenary Talks: Thursday 30

09:00-09:50

<i>Finding structure with randomness</i>	17
Joel A. Tropp	

14:30-15:20

<i>Gradient Algorithms for Regularized Optimization</i>	18
Stephen Wright	

Contributed Talks: Monday 27

#1 Sparsity Theory (11:50-12:50)

<i>Optimally Sparse Frames</i>	19
Peter Casazza, Andreas Heinecke, Felix Krahmer, Gitta Kutyniok	

<i>Lagrangian Biduality of the ℓ_0- and ℓ_1-Minimization Problems</i>	20
Dheeraj Singaraju, Allen Yang, Shankar Sastry, Roberto Tron, Ehsan Elhamifar	

<i>Signal Recovery Via ℓ_p Minimization: Analysis using Restricted Isometry Property</i>	21
Shisheng Huang, Jubo Zhu, Fengxia Yan, Meihua Xie, Zelong Wang	

#2 SAR Imaging (11:50-12:50)

<i>Compressed sensing for joint ground imaging and target indication with airborne radar</i>	22
Ludger Prünte	

<i>Automatic target recognition from highly incomplete SAR data</i>	23
Chaoran Du, Gabriel Rilling, Mike Davies, Bernard Mulgrew	

<i>Tomographic SAR Inversion via Sparse Reconstruction</i>	24
Xiao Xiang Zhu, Richard Bamler	
#3 Medical Imaging (16:00-17:00)	
<i>On the efficiency of proximal methods for CBCT and PET reconstruction with sparsity constraint</i>	
Sandrine Anthoine, Jean-François Aujol, Yannick Boursier, Mélot Clothilde	25
<i>Reliable Small-object Reconstruction from Sparse Views in X-ray Computed Tomography</i>	
Jakob Heide Joergensen, Emil Y. Sidky, Xiaochuan Pan	26
<i>Near-optimal undersampling and reconstruction for MRI carotid blood flow measurement based on support splitting</i>	
Gabriel Rilling, Yuehui Tao, Mike E. Davies, Ian Marshall	27
#4 Sparse Approximation and Compressed Sensing Algorithms 1 (16:00-17:00)	
<i>Denoising signal represented by mixtures of multivariate Gaussians in a time-frequency dictionary</i>	
Emilie Villaron, Sandrine Anthoine, Bruno Torresani	28
<i>Efficiency of Randomized Coordinate Descent Methods on Minimization Problems with a Composite Objective Function</i>	
Martin Takac, Peter Richtárik	29
<i>Robust sparse recovery with non-negativity constraints</i>	
Martin Slawski, Matthias Hein	30
Contributed Talks: Tuesday 28	
#5 Classification and Clustering (10:20-11:20)	
<i>Sparse Subspace Clustering</i>	
Ehsan Elhamifar, Rene Vidal	31
<i>Subspace Clustering by Rank Minimization</i>	
Paolo Favaro, Avinash Ravichandran, Rene Vidal	32
<i>Multiscale Geometric Dictionaries for Point-cloud Data</i>	
Guangliang Chen, Mauro Maggioni	33
#6 Structured Sparsity 1 (10:20-11:20)	
<i>Modeling Statistical Dependencies in Sparse Representations</i>	
Tomer Faktor, Yonina C. Eldar, Michael Elad	34
<i>A source localization approach based on structured sparsity for broadband far-field sources</i>	
Aris Gretsistas, Mark Plumley	35
<i>Sparsity with sign-coherent groups of variables via the cooperative-Lasso</i>	
Julien Chiquet, Yves Grandvalet, Camille Charbonnier	36
#7 Random Matrix Theory (11:50-12:50)	
<i>Tail bounds for all eigenvalues of a sum of random matrices</i>	
Alex Gittens, Joel Tropp	37
<i>Random Projections are Nearly Isometric For Parametric Functions Too</i>	
William Mantzel, Justin Romberg	38
<i>Concentration Inequalities and Isometry Properties for Compressive Block Diagonal Matrices</i>	
Han Lun Yap, Jae Young Park, Armin Eftekhari, Christopher Rozell, Michael Wakin	39
#8 Structured Sparsity 2 (11:50-12:50)	
<i>Sparse Anisotropic Triangulations and Image Estimation</i>	
Laurent Demaret	40
<i>Compressive Sensing with Biorthogonal Wavelets via Structured Sparsity</i>	
Marco Duarte, Richard Baraniuk	41
<i>A convex approach for structured wavelet sparsity patterns</i>	
Nikhil Rao, Robert Nowak, Stephen Wright, Nick Kingsbury	42
#9 Analysis Framework (16:00-17:00)	
<i>Hybrid Synthesis-Analysis Frame-Based Regularization: A Criterion and an Algorithm</i>	
Manya Afonso, José Bioucas-Dias, Mário Figueiredo	43
<i>Cosparsé Analysis Modeling</i>	
Sangnam Nam, Michael E. Davies, Michael Elad, Rémi Gribonval	44

<i>Implications for compressed sensing of a new sampling theorem on the sphere</i>	
Jason McEwen, Gilles Puy, Jean-Philippe Thiran, Pierre Vandergheynst, Dimitri Van De Ville	45
#10 Dynamical and Time-varying Systems (16:00-17:00)	
<i>Compressive Sensing for Gaussian Dynamic Signals</i>	
Wei Dai, Dino Sejdinovic, Olgica Milenkovic	46
<i>Simultaneous Estimation of Sparse Signals and Systems at Sub-Nyquist Rates</i>	
Hojjat Akhondi Asl, Pier Luigi Dragotti	47
<i>A Hierarchical Re-weighted-ℓ_1 Approach for Dynamic Sparse Signal Estimation</i>	
Adam Charles, Christopher Rozell	48
Contributed Talks: Wednesday 29	
#11 Compressed Sensing Theory (10:20-11:20)	
<i>Weighted L_p Constraints in Noisy Compressed Sensing</i>	
Laurent Jacques, David Hammond, Jalal Fadili	49
<i>Spread Spectrum for Universal Compressive Sampling</i>	
Gilles Puy, Pierre Vandergheynst, Rémi Gribonval, Yves Wiaux	50
<i>On Bounds of Restricted Isometry Constants for Gaussian Random Matrices</i>	
Bubacarr Bah, Jared Tanner	51
#12 Sparsity Applications (10:20-11:20)	
<i>Towards Optimal Data Acquisition in Diffuse Optical Tomography: Analysis of Illumination Patterns</i>	
Marta Betcke, Simon Arridge	52
<i>Recent evidence of sparse coding in neural systems</i>	
Christopher Rozell, Mengchen Zhu	53
<i>Sparse Detection in the Chirplet Transform</i>	
Fabien Millioz, Mike Davies	54
#13 Generalized Compressed Sensing (11:50-12:50)	
<i>Riemannian optimization for rank minimization problems</i>	
Bart Vandereycken	55
<i>The degrees of freedom of the Lasso in underdetermined linear regression models</i>	
Maher Kachour, Jalal Fadili, Christophe Chesneau, Charles Dossal, Gabriel Peyré	56
<i>Guaranteed recovery of a low-rank and joint-sparse matrix from incomplete and noisy measurements</i>	
Mohammad Golbabaei, Pierre Vandergheynst	57
#14 Estimation and Detection (11:50-12:50)	
<i>Message-Passing Estimation from Quantized Samples</i>	
Ulugbek Kamilov, Vivek Goyal, Sundeep Rangan	58
<i>Ambiguity Sparse Processes</i>	
Sofia Olhede	59
<i>Sparseness-based non-parametric detection and estimation of random signals in noise</i>	
Dominique Pasto, Abdourrahmane Atto	60
#15 Generalized Sampling Techniques (16:00-17:00)	
<i>Reconstruction and Cancellation of Sampled Multiband Signals Using Discrete Prolate Spheroidal Sequences</i>	
Mark Davenport, Michael Wakin	61
<i>Exponential Reproducing Kernels for Sparse Sampling</i>	
Jose Antonio Uriagu, Pier Luigi Dragotti, Thierry Blu	62
<i>Generalized sampling and infinite-dimensional compressed sensing</i>	
Anders Hansen, Ben Adcock	63
#16 Sparse Approximation and Compressed Sensing Algorithms 2 (16:00-17:00)	
<i>A Lower Complexity Bound for ℓ_1-regularized Least-squares Problems using a Certain Class of Algorithms</i>	
Tobias Lindström Jensen	64
<i>A New Recovery Analysis of Iterative Hard Thresholding for Compressed Sensing</i>	
Andrew Thompson, Coralia Cartis	65

<i>Recipes for Hard Thresholding Methods</i>	66
Anastasios Kyrillidis, Volkan Cevher	
Contributed Talks: Thursday 30	
#17 Dictionary Learning (10:20-11:20)	
<i>Local optimality of dictionary learning algorithms</i>	67
Boris Mailh�, Mark Plumley	
<i>Approximate Message Passing for Bilinear Models</i>	68
Philip Schniter, Volkan Cevher	
<i>Structure-Aware Non-Negative Dictionary Learning</i>	69
Ken O'Hanlon, Mark Plumley	
#18 Analogue to Digital Conversions (10:20-11:20)	
<i>Multi-Channel Analog-to-Digital (A/D) Conversion using Fewer A/D Converters than Channels</i>	70
Ahmed H. Tewfik, Youngchun Kim, B. Vikrham Gowreesunker	
<i>Practical Design of a Random Demodulation Sub-Nyquist ADC</i>	71
Stephen Becker, Juhwan Yoo, Mathew Loh, Azita Emami-Neyestanak, Emmanuel Cand�s .	
<i>Compressive Spectral Estimation Can Lead to Improved Resolution/Complexity Tradeoffs</i>	72
Michael Lexa, Mike Davies, John Thompson	
#19 Low Dimensional and Analysis Sparse Model Learning (11:50-12:50)	
<i>K-SVD Dictionary-Learning for Analysis Sparse Models</i>	73
Ron Rubinstein, Michael Elad	
<i>Analysis Operator Learning for Overcomplete Cosparse Representations</i>	74
Mehrdad Yaghoobi, Sangnam Nam, Remi Gribonval, Mike E. Davies	
<i>Learning hybrid linear models via sparse recovery</i>	75
Eva Dyer, Aswin Sankaranarayanan, Richard Baraniuk	
#20 Performance Evaluations (11:50-12:50)	
<i>Evaluating Dictionary Learning for Sparse Representation Algorithms using SMALLbox</i>	76
Ivan Damjanovic, Matthew Davies, Mark Plumley	
<i>A Reproducible Research Framework for Audio Inpainting</i>	77
Amir Adler, Valentin Emiya, Maria G. Jafari, Michael Elad, R�mi Gribonval	
<i>GPU Accelerated Greedy Algorithms for Sparse Approximation</i>	78
Jeffrey Blanchard, Jared Tanner	
#21 PCA/ICA/BSS (16:00-17:00)	
<i>Two Proposals for Robust PCA Using Semidefinite Programming</i>	79
Michael McCoy, Joel Tropp	
<i>Blind Source Separation of Compressively Sensed Signals</i>	80
Martin Kleinsteuber, Hao Shen	
<i>Finding Sparse Approximations to Extreme Eigenvectors: Generalized Power Method for Sparse PCA and Extensions</i>	81
Peter Richt�rik, Michel Journee, Yurii Nesterov, Rodolphe Sepulchre	
#22 Sparse Filter Design (16:00-17:00)	
<i>Stable Embeddings of Time Series Data</i>	82
Han Lun Yap, Christopher Rozell	
<i>Estimating multiple filters from stereo mixtures: a double sparsity approach</i>	83
Simon Arberet, Prasad Sudhakar, R�mi Gribonval	
<i>Well-posedness of the frequency permutation problem in sparse filter estimation with l_p minimization</i>	84
Alexis Benichoux, Prasad Sudhakar, R�mi Gribonval	
Posters A	
<i>Optical wave field reconstruction based on nonlocal transform-domain sparse regularization for phase and amplitude</i>	85
Vladimir Katkovnik, Jaakkko Astola	
<i>Efficient sparse representation based classification using hierarchically structured dictionaries</i>	86
Jort Gemmeke	

<i>Sparse Object-Based Audio Coding Using Non-Negative Matrix Factorization of Spikegrams</i>	87
Ramin Picheva, Hossein Najaf-Zadeh, Frederic Mustiere, Christopher Srinivasa, Hassan Lahdili	
<i>Recovery of Compressively Sampled Sparse Signals using Cyclic Matching Pursuit</i>	88
Bob Sturm, Mads Christensen	
<i>Structured and soft ! Boltzmann machine and mean-field approximation for structured sparse representations</i>	89
Angélique Drémeau, Laurent Daudet	
<i>BM3D-frame sparse image modeling and decoupling of inverse and denoising for image deblurring</i>	90
Aram Danielyan, Vladimir Katkovnik, Karen Egiazarian	
<i>Super-resolution and reconstruction of far-field ghost imaging via sparsity constraints</i>	91
Wenlin Gong, Shensheng Han	
<i>Fast compressive terahertz imaging</i>	92
Hao Shen, Lu Gan, Nathan Newman, Yaochun Shen	
<i>Dictionary Learning: Application to ECG Denoising</i>	93
Anastasia Zakharova, Olivier Laligant, Christophe Stolz	
<i>Unsupervised Learning of View-Condition Invariant Sparse Representation for Image Category Classification</i>	94
Hui Ka Yu	
<i>Joint localisation and identification of acoustical sources with structured-sparsity priors</i>	95
Gilles Chardon, Laurent Daudet	
<i>An Alternating Direction Algorithm for (Overlapping) Group Regularization</i>	96
Mário Figueiredo, José Bioucas-Dias	
<i>Sparse Approximation of the Neonatal EEG</i>	97
Vladimir Matic, Maarten De Vos, Bogdan Mijovic, Sabine Van Huffel	
<i>Inversion of 2-D images to estimate densities in \mathbb{R}^3</i>	98
Dalia Chakrabarty, Fabio Rigat	
<i>Constrained Non-Negative Matrix Factorization for source separation in Raman Spectroscopy</i>	99
Hérald Rabeson	
<i>Sparse Templates-Based Shape Representation for Image Segmentation</i>	100
Stefania Petra, Dirk Breitenreicher, Jan Lellmann, Christoph Schnörr	
<i>Wyner-Ziv Coding for Distributed Compressive Sensing</i>	101
Kezhi Li, Su Gao, Cong Ling, Lu Gan	
<i>Methods for Training Adaptive Dictionary in Underdetermined Speech Separation</i>	102
Tao Xu, Wenwu Wang	
<i>Analysis of Subsampled Circulant Matrices for Imaging</i>	103
Matthew Turner, Lina Xu, Wotao Yin, Kevin Kelly	
<i>A new BCI Classification Method based on EEG Sparse Representation</i>	104
Younghak Shin, Seungchan Lee, Heung-No Lee	
<i>A Realistic Distributed Compressive Sensing Framework for Multiple Wireless Sensor Networks</i>	105
Oliver James, Heung-No Lee	

Posters B

<i>Sparse Phase Retrieval</i>	106
Shiro Ikeda, Hidetoshi Kono	
<i>Probabilistic models which enforce sparsity</i>	107
Ali Mohammad-Djafari	
<i>Greedy Algorithms for Sparse Total Least Squares</i>	108
Bogdan Dumitrescu	
<i>Super-resolution based on Sparsity Priori</i>	109
Hui Wang, Shensheng Han, Mikhail I. Kolobov	
<i>Fast Compressive Sensing Recovery with Transform-based Sampling</i>	110
Hung-Wei Chen, Chun-Shien Lu, Soo-Chang Pei	
<i>Feature Selection in Carotid Artery Segmentation Process based on Learning Machines</i>	111
Rosa-María Menchón-Lar, Consuelo Bastida-Jumilla, Juan Morales Sánchez, José-Luis Sancho-Gómez	

<i>Best Basis Matching Pursuit</i>	112
Tianyao Huang, Yimin Liu, Huadong Meng, Xiqin Wang	112
<i>Adaptive Algorithm for Online Identification and Recovering of Jointly Sparse Signals</i>	
Roi Amel, Arie Feuer	113
<i>Primal-Dual TV Reconstruction in Refractive Deflectometry</i>	
Adriana González, Laurent Jacques, Emmanuel Foumouo, Philippe Antoine	114
<i>Greedy decompositions with randomly varying time-frequency subdictionaries</i>	
Manuel Moussallam, Gael Richard, Laurent Daudet	115
<i>A Sparsity based Regularization Algorithm with Automatic Parameter Estimation</i>	
Damiana Lazzaro	116
<i>An unsupervised iterative shrinkage/thresholding algorithm for sparse expansion in a union of dictionaries.</i>	
Matthieu Kowalski, Thomas Rodet	117
<i>An Infeasible-Point Subgradient Algorithm and a Computational Solver Comparison for ℓ_1-Minimization</i>	
Andreas Tillmann, Dirk Lorenz, Marc Pfetsch	118
<i>On the relation between perceptrons and non-negative matrix factorization</i>	
Hugo Van hamme	119
<i>Recovery of finite alphabet signals from incomplete measurements</i>	
Si Mohamed Aziz Sbaï, Abdeldjalil Aïssa-El-Bey, Dominique Pastor	120
<i>Adding Dynamic Smoothing to Mixture Mosaicing Synthesis</i>	
Graham Coleman, Jordi Bonada, Esteban Maestre	121
<i>Block-Sparse Recovery via Convex Optimization</i>	
Ehsan Elhamifar, Rene Vidal	122
<i>Performance limits of the measurements on Compressive Sensing for Multiple Sensor System</i>	
Sangjun Park, Hwanchol Jang, Heung-No Lee	123
<i>Message Passing Aided Least Square Recovery for Compressive Sensing</i>	
Jaewook Kang, Heung-No Lee, Kiseon Kim	124
<i>Matrix-free Interior Point Method for Compressed Sensing Problems</i>	
Kimonas Fountoulakis, Jacek Gondzio	125
<i>A Block-Based Approach to Adaptively Bias the Weights of Adaptive Filters</i>	
Luis Azpícueta-Ruiz, Jerónimo Arenas-García	126

TILT and RASL: For Low-Rank Structures in Images and Data

Yi Ma

ECE Department, UIUC and VC Group, Microsoft Research Asia

Abstract—In this talk, we will introduce two fundamental computational tools, namely TILT and RASL, for extracting rich low-rank structures in images and videos, respectively. Both tools utilize the same transformed Robust PCA model for the visual data:

$$D \circ \tau = A + E \quad (1)$$

and use practically the same algorithm for extracting the low-rank structures A from the visual data D , despite image domain transformation τ and sparse corruptions E . We will show how these two seemingly simple tools can help unleash tremendous information in images and videos that we used to struggle to get. We believe these new tools will bring disruptive changes to many challenging tasks in computer vision and image processing, including feature extraction, image correspondence or alignment, 3D reconstruction, and object recognition, etc.

Yi Ma is the research manager of the Visual Computing group at Microsoft Research Asia in Beijing since January 2009. He is also an associate professor at the Electrical & Computer Engineering Department of the University of Illinois at Urbana-Champaign. His main research interest is in computer vision, high-dimensional data analysis, and systems theory. He is the first author of the popular vision textbook “[An Invitation to 3-D Vision](#),” published by Springer in 2003. Yi Ma received two Bachelor’s degree in Automation and Applied Mathematics from Tsinghua University (Beijing, China) in 1995, a Master of Science degree in EECS in 1997, a Master of Arts degree in Mathematics in 2000, and a PhD degree in EECS in 2000, all from the University of California at Berkeley. Yi Ma received the David Marr Best Paper Prize at the International Conference on Computer Vision 1999, the Longuet-Higgins Best Paper Prize at the European Conference on Computer Vision 2004, and the Sang Uk Lee Best Student Paper Award with his students at the Asian Conference on Computer Vision in 2009. He also received the CAREER Award from the National Science Foundation in 2004 and the Young Investigator Award from the Office of Naval Research in 2005. He is an associate editor of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) and the International Journal of Computer Vision (IJCV). He has served as the chief guest editor for special issues for the Proceedings of IEEE and the IEEE Signal Processing Magazine. He will also serve as Program Chair for ICCV 2013 in Sydney, Australia. He is a senior member of IEEE and a member of ACM, SIAM, and ASEE.

This is joint work with John Wright of Columbia, Emmanuel Candes of Stanford, and my students Zhengdong Zhang, Xiao Liang, Yigang Peng of Tsinghua, Arvind Ganesh of UIUC.

Coding for Multiplex Optical Imaging

David J. Brady

Duke Imaging and Spectroscopy Program, Department of Electrical and Computer Engineering
Duke University, Durham, North Carolina 20291-0291
www.disp.duke.edu

Abstract—Efficient sampling of sparse signals requires measurement of linear feature projections. “Weighing design” consists of selecting projection coefficients to satisfy mathematical and physical objectives. This paper reviews the weighing design problem for applications in spectral imaging, focal tomography and holography.

I. INTRODUCTION

We consider linear measurement systems described by the forward model $\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}$ where \mathbf{g} is measurement data, \mathbf{H} is the measurement operator, \mathbf{f} is the object state and \mathbf{n} is noise. The goal of these systems is to estimate, e.g. to image, \mathbf{f} given \mathbf{g} . If \mathbf{H} is not the identity operator then the system takes “multiplex measurements”. While radar and computed tomography aficionados may consider the multiplex designation redundant, from 1840 until 1950 imager design emphasized physical design for focal transformations. The advent of digital computers and electronic detectors changed this goal, but even after 60 years the ensuing revolution is still evolving. Compressed sensing theory focuses particular attention on Shannon’s work at the start of this revolution. This paper considers the implications of compressed sensing on two results in measurement theory from the same era, specifically multiplex spectroscopy [1] and holography [2].

II. SPECTROSCOPY AND SPECTRAL IMAGING

For over half a century, weighing design for multiplex spectroscopy focused on linear estimators of \mathbf{f} given \mathbf{g} . While Harwit and Sloan acknowledge in the seminal work on this approach [3] that biased estimators may achieve better results, very little work on coding for nonlinear estimators appeared before 2000. An important exception appears in work on computed tomographic imaging spectrometers, which applied convex optimization to multiplex spectral imaging [4]. The goal of this work was to overcome a “missing cone” Radon projections. More recently, my group has shown that Golay-style coded apertures eliminate the missing cone and that compressed sensing theory may be applied to estimate full 3D data cubes from coded 2D snapshots [5].

While it is clear that significant advantages arise from the combination of coded projections and constrained optimization, optimal codes for these systems are currently unknown. This is in sharp contrast to previous theory for linear estimators, which showed Hadamard codes to be optimal for additive noise and identity operators to be optimal for Poisson noise. As my talk describes using both simulated and experimental data, pseudo-random codes may outperform identity and Hadamard codes when combined with modern regularization and optimization strategies.

III. FOCAL TOMOGRAPHY

Major practical successes in compressed sensing have arisen in applications where it is physically impossible to implement \mathbf{H} as identity matrix. Spectral imaging is one such example, others arise in various multidimensional tomographies. Natural imaging of 3D scenes is tomographic problem of particular interest. Historically, focal imaging systems are map 2D object planes to 2D image planes.

This model has been preferred because focal recording devices (e.g. film and detector arrays) are confined to 2D surfaces. With the advent of computational imaging, however, the physical structure of the measurement system need not be tied to the physical structure of the image. Specifically, one should be able to implement multiplex codes that enable direct estimation of 3D objects from snapshot data. While adhoc tomographic recording strategies using camera arrays or pupil coding strategies have been attempted to achieve this goal, systematic studies of codes for native 3D optical imaging are just beginning. Image space coding strategies similar to those used in coded aperture spectral imaging are particularly attractive for this challenge.

IV. HOLOGRAPHY

Optical imaging inherently combines analog signal processing in optical elements with digital image formation. Quasi-focal design with compact kernel support is essential to reasonable rank measurement on natural fields. Imagers using laser illumination, in contrast, may achieve reasonable rank measurement operators with unbounded sampling kernels. This allows lensless imaging over large apertures. Unfortunately, natural objects reflect laser light diffusely, meaning that a random phase is added to the reflected field in each image pixel. Such specular images are not sparse on any basis. This difficulty may be overcome by estimating the magnitude of the scattering cross section of each pixel, which forms a compressible image. Under this approach one seeks to invert transformed statistics of measurement data to estimate a particular set of object statistics [6]. Weighing design for this application consists of selecting both the raw sampling structure and the synthetic statistics taken as intermediate indicators of the object state. Coding for this application introduces new challenges and opportunities and suggests novel statistical definitions for the concept of compressive sampling.

REFERENCES

- [1] M. Golay, “Multislit spectroscopy,” *J. Opt. Soc. Amer.*, vol. 39, pp. 437–444, 1949.
- [2] D. Gabor, “A new microscopic principle,” *Nature*, vol. 161, pp. 777–778, 1948.
- [3] M. Harwit and N. J. A. Sloane, *Hadamard transform optics*. Academic Press, 1979.
- [4] A. K. Brodzik and J. M. Mooney, “Convex projections algorithm for restoration of limited-angle chromotomographic images,” *J. Opt. Soc. Am. A*, vol. 16, no. 2, pp. 246–257, 1999.
- [5] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, “Single-shot compressive spectral imaging with a dual-disperser architecture,” *Opt. Express*, vol. 15, no. 21, pp. 14 013–14 027, 2007.
- [6] K. Choi, R. Horisaki, J. Hahn, S. Lim, D. L. Marks, T. J. Schulz, and D. J. Brady, “Compressive holography of diffuse objects,” *Appl. Opt.*, vol. 49, no. 34, pp. H1–H10, Dec 2010.

Strutured Sparsity-Inducing Norms through Submodular Functions

Francis Bach

INRIA - Ecole Normale Supérieure
Paris, France

Abstract—Sparse methods for supervised learning aim at finding good linear predictors from as few variables as possible, i.e., with small cardinality of their supports. This combinatorial selection problem is often turned into a convex optimization problem by replacing the cardinality function by its convex envelope (tightest convex lower bound), in this case the ℓ_1 -norm. In this work, we investigate more general set-functions than the cardinality, that may incorporate prior knowledge or structural constraints which are common in many applications: namely, we show that for nondecreasing submodular set-functions, the corresponding convex envelope can be obtained from its Lovász extension, a common tool in submodular analysis. This defines a family of polyhedral norms, for which we provide generic algorithmic tools (subgradients and proximal operators) and theoretical results (conditions for support recovery or high-dimensional inference). By selecting specific submodular functions, we can give a new interpretation to known norms, such as those based on rank-statistics or grouped norms with potentially overlapping groups; we also define new norms, in particular ones that can be used as non-factorial priors for supervised learning.

The concept of parsimony is central in many scientific domains. In the context of statistics, signal processing or machine learning, it takes the form of variable or feature selection problems, and is commonly used in two situations: First, to make the model or the prediction more interpretable or cheaper to use, i.e., even if the underlying problem does not admit sparse solutions, one looks for the best sparse approximation. Second, sparsity can also be used given prior knowledge that the model should be sparse. In these two situations, reducing parsimony to finding models with low cardinality turns out to be limiting, and structured parsimony has emerged as a fruitful practical extension, with applications to image processing, text processing or bioinformatics (see, e.g., [1], [2], [3], [4], [5], [6], [7]). For example, in [4], structured sparsity is used to encode prior knowledge regarding network relationship between genes, while in [6], it is used as an alternative to structured non-parametric Bayesian process based priors for topic models.

Most of the work based on convex optimization and the design of dedicated sparsity-inducing norms has focused mainly on the specific allowed set of sparsity patterns [1], [2], [4], [6]: if $w \in \mathbf{R}^p$ denotes the predictor we aim to estimate, and $\text{Supp}(w)$ denotes its support, then these norms are designed so that penalizing with these norms only leads to supports from a given family of allowed patterns. In this paper, we instead follow the approach of [8], [3] and consider specific penalty functions $F(\text{Supp}(w))$ of the support set, which go beyond the cardinality function, but are not limited or designed to only forbid certain sparsity patterns. These may also lead to restricted sets of supports but their interpretation in terms of an *explicit* penalty on the support leads to additional insights into the behavior of structured sparsity-inducing norms. While direct greedy approaches (i.e., forward selection) to the problem are considered in [8], [3], we provide convex relaxations to the function $w \mapsto F(\text{Supp}(w))$, which extend the traditional link between the ℓ_1 -norm and the cardinality

function.

This is done for a particular ensemble of set-functions F , namely *nondecreasing submodular functions*. Submodular functions may be seen as the set-function equivalent of convex functions, and exhibit many interesting properties—see [9] for a tutorial on submodular analysis and [10], [11] for other applications to machine learning. In this presentation, we will present the following contributions:

— We make explicit links between submodularity and sparsity by showing that the convex envelope of the function $w \mapsto F(\text{Supp}(w))$ on the ℓ_∞ -ball may be readily obtained from the Lovász extension of the submodular function.

— We provide generic algorithmic tools, i.e., subgradients and proximal operators, as well as theoretical guarantees, i.e., conditions for support recovery or high-dimensional inference, that extend classical results for the ℓ_1 -norm and show that many norms may be tackled by the exact same analysis and algorithms.

— By selecting specific submodular functions, we recover and give a new interpretation to known norms, such as those based on rank-statistics or grouped norms with potentially overlapping groups [1], [2], [7], and we define new norms, in particular ones that can be used as non-factorial priors for supervised learning. These are illustrated on simulation experiments, where they outperform related greedy approaches [3].

For more details, see [12].

REFERENCES

- [1] P. Zhao, G. Rocha, and B. Yu, “Grouped and hierarchical model selection through composite absolute penalties,” *Annals of Statistics*, vol. 37, no. 6A, pp. 3468–3497, 2009.
- [2] R. Jenatton, J. Audibert, and F. Bach, “Structured variable selection with sparsity-inducing norms,” arXiv:0904.3523, Tech. Rep., 2009.
- [3] J. Huang, T. Zhang, and D. Metaxas, “Learning with structured sparsity,” in *Proc. ICML*, 2009.
- [4] L. Jacob, G. Obozinski, and J.-P. Vert, “Group Lasso with overlaps and graph Lasso,” in *Proc. ICML*, 2009.
- [5] S. Kim and E. Xing, “Tree-guided group Lasso for multi-task regression with structured sparsity,” in *Proc. ICML*, 2010.
- [6] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, “Proximal methods for sparse hierarchical dictionary learning,” in *Proc. ICML*, 2010.
- [7] J. Mairal, R. Jenatton, G. Obozinski, and F. Bach, “Network flow algorithms for structured sparsity,” in *Adv. NIPS*, 2010.
- [8] J. Haupt and R. Nowak, “Signal reconstruction from noisy random projections,” *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.
- [9] F. Bach, “Convex analysis and optimization with submodular functions: a tutorial,” HAL, Tech. Rep. 00527714, 2010.
- [10] A. Krause and C. Guestrin, “Near-optimal nonmyopic value of information in graphical models,” in *Proc. UAI*, 2005.
- [11] Y. Kawahara, K. Nagano, K. Tsuda, and J. Bilmes, “Submodularity cuts and applications,” in *Adv. NIPS*, 2009.
- [12] F. Bach, “Structured sparsity-inducing norms through submodular functions,” in *Advances in Neural Information Processing Systems*, 2010.

Sparsity & Co.: An Overview of Analysis *vs* Synthesis in Low-Dimensional Signal Models

R. Gribonval

Centre INRIA Rennes - Bretagne Atlantique

Campus de Beaulieu, 35042 Rennes Cedex

France

Email: remi.gribonval@inria.fr

Abstract—In the past decade there has been a great interest in a synthesis-based model for signals, based on sparse and redundant representations. Such a model assumes that the signal of interest can be composed as a linear combination of few columns from a given matrix (the dictionary). An alternative *analysis-based* model can be envisioned, where an analysis operator multiplies the signal, leading to a *cosparse* outcome. How similar are the two signal models ? The answer obviously depends on the dictionary/operator pair, and on the measure of (co)sparsity.

For dictionaries in Hilbert spaces that are frames, the canonical dual is arguably the most natural associated analysis operator. When the frame is *localized*, the canonical frame coefficients provide a near sparsest expansion for several ℓ^p sparseness measures, $p \leq 1$. However, for frames which are not localized, this no longer holds true: the sparsest synthesis coefficients may differ significantly from the canonical coefficients.

In general the sparsest synthesis coefficients may also depend strongly on the choice of the sparseness measure, but this dependency vanishes for dictionaries with a *null space property* and signals that are combinations of sufficiently few columns from the dictionary. This uniqueness result, together with algorithmic guarantees, is at the basis of a number of signal reconstruction approaches for generic linear inverse problems (e.g., compressed sensing, inpainting, source separation, etc.).

Is there a similar uniqueness property when the data to be reconstructed is *cosparse* rather than *sparse* ? Can one derive cosparse regularization algorithms with performance guarantees ? Existing empirical evidence in the litterature suggests that a positive answer is likely. In recent work we propose a uniqueness result for the solution of linear inverse problems under a cosparse hypothesis, based on properties of the analysis operator and the measurement matrix. Unlike with the synthesis model, where recovery guarantees usually require the linear independence of sets of few columns from the dictionary, our results suggest that linear dependencies between rows of the analysis operators may be desirable.

ACKNOWLEDGMENT

This overview will present results obtained in joint work with M. Nielsen [1], S. Nam, M. Elad, M. Davies [2]. The author acknowledges the support by the European Community's FP7-FET program, SMALL project, under grant agreement no. 225913.

REFERENCES

- [1] R. Gribonval and M. Nielsen, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure," *Applied and Computational Harmonic Analysis*, vol. 22, no. 3, pp. 335–355, May 2007. [Online]. Available: <http://www.math.auc.dk/research/reports/R-2003-16.pdf>
- [2] S. Nam, M. Davies, M. Elad, and R. Gribonval, "Cosparse analysis modeling - Uniqueness and algorithms," in *Acoustics, Speech and Signal Processing, 2011. ICASSP 2011. IEEE International Conference on*, Prague, Czech Republic, May 2011. [Online]. Available: <http://hal.inria.fr/inria-00557933/en>

Precise Optimality Results in Compressed Sensing

David Donoho
Department of Statistics
Stanford University

Abstract—Of the many papers on compressed sensing and sparse recovery to date, a large fraction concern qualitative phenomena, where for example certain phenomena are observed “for sufficiently sparse signals” and, while empirically it is clear that there is a sharp transition in observable behavior as sparsity crosses a threshold, much existing published research uses methods that are often unable to pinpoint the transition point precisely. Of course, for engineering work, one would like to have precise knowledge of the limits of compressed sensing, rather than just qualitative knowledge.

Other results promise stability of certain recovery procedures with unspecified stability constants C . Again, precise evaluations would be more useful.

I will describe recent work giving precise asymptotic results on mean squared error and other characteristics, of a range of recovery procedures in a range of high-dimensional problems from sparse regression and compressed sensing; these include results for LASSO, group LASSO, and nonconvex sparsity penalty methods. A key application of such precise formulas is their use in deriving precise optimality results which were not known previously, and to our knowledge are not available by other methods.

Approximate message passing, and ideas from minimax statistical decision theory as well of statistical physics, are the key ingredients to the results I will focus on. This is joint work over several papers with several co-authors, including Andrea Montanari, Iain Johnstone, and Arian Maleki.

I will also try to discuss precise results and methods of Tanner, of Blanchard, Cartis, and Tanner, of Weiyu Xu and Hassibi, and of Stojnic.

REFERENCES

- [1] M. Bayati and A. Montanari, *The dynamics of message passing on dense graphs, with applications to compressed sensing*, IEEE Trans. on Inform. Theory (2010), arXiv:1001.3448.
- [2] M. Bayati and A. Montanari, *The LASSO risk for gaussian matrices*, arXiv:1008.2581, 2010.
- [3] J. D. Blanchard, C. Cartis, and J. Tanner, *The restricted isometry property and ℓ_q -regularization: Phase transitions for sparse approximation*, SIAM Review 2011.
- [4] D. L. Donoho and J. Tanner. *Precise Undersampling Theorems*. Proceedings of the IEEE. June 2010, **98**:6, 913-924.
- [5] D. L. Donoho, A. Maleki, and A. Montanari, *Message Passing Algorithms for Compressed Sensing*, Proceedings of the National Academy of Sciences **106** (2009), 18914–18919.
- [6] D.L. Donoho, A. Maleki, and A. Montanari. *The Noise Sensitivity Phase Transition in Compressed Sensing*, arXiv:1004.1218, 2010.
- [7] D.L. Donoho, A. Maleki, and A. Montanari, *Compressed Sensing Over ℓ_p -balls: Minimax Mean Squared Error*, arXiv:1103.1943v2, 2011.
- [8] A. Maleki, *Approximate Message Passing Algorithms for Compressed Sensing*, Ph.D. Thesis, Stanford University, 2010.
- [9] M. Stojnic. *Various thresholds for ℓ_1 -optimization in compressed sensing*. ArXiv. <http://arxiv.org/abs/0907.3666>. 2009.
- [10] Weiyu Xu; Hassibi, B.; Compressed sensing over the Grassmann manifold: A unified analytical framework Communication, Control, and Computing, 2008 46th Annual Allerton Conference; 23-26 Sept. 2008; 562 - 567

Sampling in the Age of Sparsity

Martin Vetterli

Ecole Polytechnique Fdrale de Lausanne, Switzerland and University of California, Berkeley, USA

Abstract—Sampling is a central topic not just in signal processing and communications, but in all fields where the world is analog, but computation is digital. This includes sensing, simulating, and rendering the real world, estimating parameters, or using analog channels.

The question of sampling is very simple: when is there a one-to-one relationship between a continuous-time function and adequately acquired samples of this function? Sampling has a rich history, dating back to Whittaker, Nyquist, Kotelnikov, Shannon and others, and is an active area of contemporary research with fascinating new results.

Classic results are on bandlimited functions, where taking measurements at the Nyquist rate is sufficient for perfect reconstruction. These results were extended to shift-invariant and multiscale spaces during the development of wavelets. All these methods are based on subspace structures, and on linear approximation. Irregular sampling, with known sampling times, relies on the theory of frames. These classic results can be used to derive sampling theorems related to PDE's, to mobile sensing and as well as to sampling based on timing information.

Recently, nonlinear sampling methods have appeared. Nonlinear approximation in wavelet spaces is powerful for approximation and compression. This indicates that functions that are sparse in a basis (but not necessarily on a fixed subspace) can be represented efficiently. The idea is even more general than sparsity in a basis, as pointed out in the framework of signals with finite rate of innovation. Such signals are nonbandlimited continuous-time signals, but with a parametric representation having a finite number of degrees of freedom per unit of time. This leads to sharp results on sampling and reconstruction of such sparse continuous-time signals, leading to sampling at Occam's rate.

Among nonlinear methods, compressed sensing and compressive sampling, have generated a lot of attention. This is a discrete time, finite dimensional set up, with strong results on possible recovery by relaxing the ℓ_0 into ℓ_1 optimization, or using greedy algorithms. These methods have the advantage of unstructured measurement matrices (actually, typically random ones) and therefore a certain universality, at the cost of some redundancy. We compare the two approaches, highlighting differences, similarities, and respective advantages.

We finish by looking at selected applications in practical signal processing and communication problems. These cover wideband communications, noise removal, distributed sampling, and superresolution imaging, to name a few. In particular, we describe a recent result on multichannel sampling with unknown shifts, which leads to an efficient superresolution imaging method.

He works on signal processing and communications, in particular, sampling, wavelets, multirate signal processing for communications, theory and applications, image and video compression, joint source-channel coding, self-organized communication systems and sensor networks and inverse problems like acoustic tomography. Martin Vetterli has published about 150 journal papers on the subjects.

His work won him numerous prizes, like best paper awards from EURASIP in 1984 and of the IEEE Signal Processing Society in 1991, 1996 and 2006, the Swiss National Latsis Prize in 1996, the SPIE Presidential award in 1999, and the IEEE Signal Processing Technical Achievement Award in 2001, the IEEE Signal Processing Society Award in 2010. He is a Fellow of IEEE, of ACM and EURASIP, and was a member of the Swiss Council on Science and Technology (2000-2004) and is an ISI highly cited researcher in engineering.

He is the co-author of three textbooks, with J. Kovacevic, "Wavelets and Subband Coding" (Prentice-Hall, 1995), with P. Prandoni, Signal Processing for Communications, (PPUR, 2008) and with J. Kovacevic and V. Goyal, of the forthcoming book Fourier and Wavelet Signal Processing" (2010).

Martin Vetterli got his Engineering degree from Eidgenoessische Technische Hochschule Zuerich (ETHZ), his MS from Stanford University and his Doctorate from Ecole Polytechnique Fdrale de Lausanne (EPFL).

He was an Associate Professor in EE at Columbia University in New York, and a Full Professor in EECS at the University of California at Berkeley before joining the Communication Systems Division of EPFL. He held several positions at EPFL, including Chair of Communication Systems, and founding director of the National Center on Mobile Information and Communication systems. He was Vice-President of EPFL, in charge of institutional affairs from 2004 to 2011. He currently is Dean of the Computer and Communication Sciences School of EPFL.

Joint work with T.Blu (CUHK), Y.Lu (Harvard), D.Gontier (ENSEPF), Y.Barbotin, A.Hormati, M.Kolundzija, J.Ranieri, J.Unnikrishnan (EPFL)

Finding Structure with Randomness

Nathan Halko and Per-Gunnar Martinsson
 Applied Mathematics
 University of Colorado at Boulder
 Boulder, CO 80309
 Email: nathan.halko@colorado.edu
 Email: per-gunnar.martinsson@colorado.edu

Joel A. Tropp
 Computing and Mathematical Sciences
 California Institute of Technology
 Pasadena, CA 9125
 Email: jtropp@cms.caltech.edu

Abstract—The purpose of this research is to make the case that randomized algorithms provide a powerful tool for constructing approximate matrix factorizations. These techniques are simple and effective, sometimes remarkably so. Compared with standard deterministic algorithms, the randomized methods are often faster and—perhaps surprisingly—more robust. Furthermore, they can produce factorizations that are accurate to any specified tolerance above machine precision, which allows the user to trade accuracy for speed if desired. In short, this work describes how randomized methods interact with classical techniques to yield effective, modern algorithms supported by detailed theoretical guarantees.

This extended abstract is drawn from the paper [1].

The task of computing a low-rank approximation to a matrix \mathbf{A} can be split into two computational stages. The first is to construct a low-dimensional subspace that captures the action of the matrix. The second is to restrict the matrix to the subspace and then compute a standard factorization (QR, SVD, etc.) of the reduced matrix.

Stage A: Compute an approximate basis for the range of the input matrix \mathbf{A} . In other words, we require a matrix \mathbf{Q} for which

$$\mathbf{Q} \text{ has orthonormal columns and } \mathbf{A} \approx \mathbf{Q}\mathbf{Q}^*\mathbf{A}. \quad (1)$$

Stage B: Given \mathbf{Q} that satisfies (1), we use \mathbf{Q} to help compute a standard factorization (QR, SVD, etc.) of \mathbf{A} .

The task in Stage A can be executed very efficiently with random sampling methods, while Stage B can be completed with well-established deterministic methods.

We focus on one formulation of the problem described in Stage A. Given a matrix \mathbf{A} , a target rank k , and an oversampling parameter p , we seek a matrix \mathbf{Q} with $k + p$ orthonormal columns such that

$$\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^*\mathbf{A}\| \approx \min_{\text{rank}(\mathbf{X}) \leq k} \|\mathbf{A} - \mathbf{X}\|. \quad (2)$$

Although there exists a minimizer \mathbf{Q} that solves the fixed rank problem for $p = 0$, the opportunity to use a small number of additional columns provides a flexibility that is crucial for the effectiveness of the computational methods we discuss.

The box labeled “Proto-Algorithm” describes, without computational details, an approach to solving (2). This simple algorithm is by no means new. It is essentially the first step of a subspace iteration with a random initial subspace [2, §7.3.2]. The novelty comes from the additional observation that the initial subspace should have a slightly higher dimension than the invariant subspace we are trying to approximate. With this revision, it is often the case that *no further iteration is required* to obtain a high-quality solution to (2). We believe this idea can be traced to [3], [4], [5].

A principal goal of this research is to provide a detailed analysis of the performance of the algorithm. This investigation produces precise error bounds, expressed in terms of the singular values of the input matrix. Let us offer a taste of these results.

PROTO-ALGORITHM

Given an $m \times n$ matrix \mathbf{A} , a target rank k , and an oversampling parameter p , this procedure computes an $m \times (k + p)$ matrix \mathbf{Q} whose columns are orthonormal and whose range approximates the range of \mathbf{A} .

- 1 Draw a random $n \times (k + p)$ test matrix Ω .
- 2 Form the matrix product $\mathbf{Y} = \mathbf{A}\Omega$.
- 3 Construct a matrix \mathbf{Q} whose columns form an orthonormal basis for the range of \mathbf{Y} .

Theorem. Suppose that \mathbf{A} is a real $m \times n$ matrix. Select a target rank $k \geq 2$ and an oversampling parameter $p \geq 2$, where $k + p \leq \min\{m, n\}$. Execute the proto-algorithm with a standard Gaussian test matrix to obtain an $m \times (k + p)$ matrix \mathbf{Q} with orthonormal columns. Then

$$\mathbb{E} \|\mathbf{A} - \mathbf{Q}\mathbf{Q}^*\mathbf{A}\| \leq \left[1 + \frac{4\sqrt{k+p}}{p-1} \cdot \sqrt{\min\{m, n\}} \right] \sigma_{k+1}, \quad (3)$$

where \mathbb{E} denotes expectation with respect to the random test matrix and σ_{k+1} is the $(k + 1)$ th singular value of \mathbf{A} .

The term σ_{k+1} appearing in (3) is the smallest possible error achievable with any basis matrix \mathbf{Q} with k columns. The theorem asserts that, on average, the algorithm produces a basis whose error lies within a small polynomial factor of the theoretical minimum.

ACKNOWLEDGMENT

NH and PGM were supported in part by NSF awards #0748488 and #0610097. JAT was supported in part by ONR award #N000140810883.

REFERENCES

- [1] N. Halko, P.-G. Martinsson, and J. A. Tropp, “Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions,” *SIAM Rev.*, vol. 53, no. 2, pp. 217–288, June 2011.
- [2] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed., ser. Johns Hopkins Studies in the Mathematical Sciences. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
- [3] T. Sarlós, “Improved approximation algorithms for large matrices via random projections,” in *Proc. 47th Ann. IEEE Symp. Foundations of Computer Science (FOCS)*, 2006, pp. 143–152.
- [4] P.-G. Martinsson, V. Rokhlin, and M. Tygert, “A randomized algorithm for the approximation of matrices,” Yale Univ., New Haven, CT, Computer Science Dept. Tech. Report 1361, 2006.
- [5] C. H. Papadimitriou, P. Raghavan, H. Tamaki, and S. Vempala, “Latent semantic indexing: A probabilistic analysis,” *J. Comput. System Sci.*, vol. 61, no. 2, pp. 217–235, 2000. [Online]. Available: <http://www.sciencedirect.com/science/article/B6WJ0-45FC93J-W/2/1a6dfbe012f6fe2fcf927db62e2da5e2>

Gradient Algorithms for Regularized Optimization

Stephen Wright
Computer Sciences Department
University of Wisconsin
1210 W. Dayton Street
Madison, WI 53706, USA
Email: swright@cs.wisc.edu

Abstract—In a typical formulation for regularized optimization problems, a weighted regularization term (usually simple and nonsmooth) is added to the underlying objective, with the purpose of inducing a particular kind of structure in the solution. The talk discusses several approaches for minimizing such functions, focusing on the case of large-scale problems in which the regularizer has a separable structure. The classic example of a separable regularizer is the ℓ_1 norm, which induces sparsity in the solution vector.

I. INTRODUCTION

One formulation of a regularized version of the optimization problem $\min_x f(x)$ (where $f : \mathbb{R}^n \rightarrow \mathbb{R}$) is

$$f(x) + \tau c(x), \quad (1)$$

where c is a convex (usually nonsmooth) function and $\tau > 0$ is the regularization parameter. The regularizer c is chosen to induce desired structure in the solution x . For example, the choice $c(x) = \|x\|_1$ is known to cause sparsity in the solution of (1), while if c is a total variation norm for an image vector x , adjoining elements of the solution of (1) tend to have the same values. Besides image processing, this formulation appears in compressed sensing, LASSO, regularized logistic regression, among many other applications.

We discuss iterative approaches for solving (1) which have one feature in common: while forming some sort of approximation to the underlying objective f , they treat c explicitly. This basic strategy makes sense because c is often a simple, separable function. We discuss variants of this approach and their relevance in several classes of applications.

II. PROX-LINEAR FRAMEWORK

The prox-linear framework uses subproblems in which f is replaced by a linear approximation about the current iterate, and a quadratic term is introduced to penalize long steps:

$$d_k := \arg \min_d \nabla f(x_k)^T d + \tau c(x_k + d) + \frac{1}{2\alpha_k} \|d\|^2, \quad (2)$$

and setting $x_{k+1} = x_k + d_k$. The parameter α_k can be manipulated in the manner of a step length to ensure sufficient decrease at each iteration, or over a sequence of iterations. The approach has appeared in the literature repeatedly in various guises; for a description and analysis motivated by compressed sensing, see [6].

III. VARIATIONS

A block-coordinate variant of (2) is obtained by fixing most components of d in (2) to be zero, thus reducing the dimension of the subproblem (2) and requiring evaluation of the gradient ∇f only for the “active” components of d — those that are allowed to vary from zero. Provided that the active components are not coupled with inactive components in the regularizer c , the subproblem generally remains easy to solve. Convergence can be proved provided that each component occasionally takes its turn at being active. This approach

is described in [5], [7]. Manifold identification properties can also be proved for this approach. In the case of $c(x) = \|x\|_1$, these results take the form that the nonzero components of x_k eventually occur in the same locations as the nonzeros of the solution x^* of (1).

Manifold identification properties are particularly relevant for the next enhancement discussed: reduced Newton methods, in which second-order information is used to enhance the search direction on the active manifold. Such an approach was proposed by [4] in the context of regularized logistic regression, and later analyzed by [7] in a more general setting. In some contexts, sampling can be used to obtain an approximate Hessian cheaply; see [1].

Finally, we discuss the *regularized dual averaging* approach in which exact gradients $\nabla f(x_k)$ are replaced by cheap sampled approximations, possibly based on a random sample of a small subset of the available data. A subproblem similar to (2) is formulated but with $\nabla f(x_k)$ replaced by the average of all gradients encountered so far and the prox-term penalizing deviation from the *initial* iterate. A sublinear convergence rate is proved in [3], [8]. Manifold identification properties are described in [2], opening the possibility of a “second-phase” algorithmic strategy in which a different algorithm is invoked when the active manifold has been identified with some level of confidence. Computational experience with this strategy on regularized regression problems will be presented in the talk.

ACKNOWLEDGMENT

The speaker gladly acknowledges collaborations with Rob Nowak, Mario Figueiredo, Sangkyun Lee, and others.

REFERENCES

- [1] R. H. Byrd, G. M. Chin, W. Neveitt, and J. Nocedal, “On the use of stochastic Hessian information in unconstrained optimization,” Technical Report, Northwestern University, June 2010.
- [2] S. Lee and S. J. Wright, “Manifold identification of dual averaging methods for regularized stochastic online learning,” to appear in Proceedings of ICML 2011.
- [3] Y. Nesterov, “Primal-dual subgradient methods for convex programs,” *Mathematical Programming, Series B* 120 (2009), pp. 221–259.
- [4] W. Shi, G. Wahba, S. J. Wright, K. Lee, R. Klein, and B. Klein, “LASSO-Patternsearch algorithm with application to ophthalmology data,” *Statistics and its Interface* 1 (2008), pp. 137–153.
- [5] P. Tseng and S. Yun, “A coordinate gradient descent method for non-smooth separable minimization,” *Mathematical Programming, Series B* 117 (2009), pp. 387–423.
- [6] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, “Sparse reconstruction by separable approximation,” *IEEE Transactions on Signal Processing* 57 (2009), pp. 2479–2493.
- [7] S. J. Wright, “Accelerated block-coordinate relaxation for regularized optimization,” Technical report, University of Wisconsin-Madison, August 2010.
- [8] L. Xiao, “Dual averaging methods for regularized stochastic learning and online optimization,” *Journal of Machine Learning Research* 11 (2010), pp. 2543–2596.

Optimally Sparse Frames

Peter G. Casazza
 Department of Mathematics
 University of Missouri
 Columbia, MO 65211, USA
 Email: casazzap@missouri.edu

Andreas Heinecke
 Department of Mathematics
 University of Missouri
 Columbia, MO 65211, USA
 Email: ah343@mizzou.edu

Felix Krahmer
 Hausdorff Center for Mathematics
 University of Bonn
 53115 Bonn, Germany
 Email: krahmer@uni-bonn.de

Gitta Kutyniok
 University of Osnabrück
 Institute of Mathematics
 49069 Osnabrück, Germany
 Email: kutyniok@uos.de

Abstract—Aiming at low-complexity frame decompositions, we introduce and study the notion of a *sparse frame*, which is a frame whose elements have a sparse representation in a given orthonormal basis. We provide an algorithmic construction to compute frames with desired frame operators, in particular, including tight frames, and prove that this construction indeed generates optimally sparse frames.

I. INTRODUCTION

Frames have established themselves as a means to derive redundant, yet stable decompositions of a signal for analysis or transmission, while also promoting sparse expansions. However, when the signal dimension is large, the computation of the frame measurements of a signal typically requires a large number of additions and multiplications, and this makes a frame decomposition intractable in applications with limited computing budget.

To tackle this problem, we propose sparsity of a frame as a new paradigm, thereby reducing the number of required additions and multiplications when computing frame measurements significantly.

II. SPARSITY: A NEW PARADIGM FOR FRAME CONSTRUCTIONS

A. Sparse Frames

We begin by proclaiming the following definition for a sparse frame:

Definition 2.1: Let $(e_j)_{j=1}^n$ be an orthonormal basis for \mathbb{R}^n , and let $(\varphi_i)_{i=1}^N$ be a frame for \mathbb{R}^n . Then $(\varphi_i)_{i=1}^N$ is called *k-sparse with respect to* $(e_j)_{j=1}^n$, if there exists a $n \times N$ -matrix C such that

$$(\varphi_1 | \cdots | \varphi_N) = (e_1 | \cdots | e_n) \cdot C \quad \text{and} \quad \|C\|_0 \leq k. \quad (1)$$

Notice that in the special case of $(e_j)_{j=1}^n$ being the standard unit basis, the sparsity of a frame equals the number of non-zero entries of its frame vectors.

B. A Notion of Optimality

We next state a notion of optimality, which will typically be considered within a particular class of frames.

Definition 2.2: Let \mathcal{F} be a class of frames for \mathbb{R}^n , let $(\varphi_i)_{i=1}^N \in \mathcal{F}$, and let $(e_j)_{j=1}^n$ be an orthonormal basis for \mathbb{R}^n . Then $(\varphi_i)_{i=1}^N$ is called *optimally sparse in \mathcal{F} with respect to* $(e_j)_{j=1}^n$, if $(\varphi_i)_{i=1}^N$ is k_1 -sparse with respect to $(e_j)_{j=1}^n$ and there does not exist $(\psi_i)_{i=1}^N \in \mathcal{F}$ which is k_2 -sparse with respect to $(e_j)_{j=1}^n$ with $k_2 < k_1$.

The class interesting to us later on is $\mathcal{F}(N, \{\lambda_i\}_{i=1}^n)$, which is the set of all unit norm frames $(\varphi_i)_{i=1}^N$ in \mathbb{R}^n whose frame operator has eigenvalues $\lambda_1, \dots, \lambda_n$.

C. A Novel Structural Property of Synthesis Matrices

Aiming for determining the maximally achievable sparsity for such a class $\mathcal{F}(N, \{\lambda_i\}_{i=1}^n)$, we first need to introduce a particular measure associated with the set of eigenvalues $\{\lambda_i\}_{i=1}^n$. This measure indicates the maximal number of partial sums which are an integer; here one maximizes over all reorderings of the eigenvalues.

Definition 2.3: A finite sequence of real values $\lambda_1, \dots, \lambda_n$ is *ordered blockwise*, if for any permutation π of $\{1, \dots, n\}$ the set of partial sums $\{\sum_{j=1}^s \lambda_{\pi(j)} : s = 1, \dots, n\}$ contains at least as many integers as the set $\{\sum_{j=1}^s \lambda_{\pi(j)} : s = 1, \dots, n\}$. The *maximal block number* of a finite sequence of real values $\lambda_1, \dots, \lambda_n$, denoted by $\mu(\lambda_1, \dots, \lambda_n)$, is the number of integers in $\{\sum_{j=1}^s \lambda_{\sigma(j)} : s = 1, \dots, n\}$, where σ is a permutation of $\{1, \dots, n\}$ such that $\lambda_{\sigma(1)}, \dots, \lambda_{\sigma(n)}$ is ordered blockwise.

As an example, consider the tight-frame-case $\lambda = \lambda_1 = \dots = \lambda_n$, whose maximal block number is $\nu(\lambda, \dots, \lambda) = \gcd(\lambda, n)$.

III. MAIN RESULT

A. The Spectral Tetris Algorithm

The so-called Spectral Tetris algorithm was first introduced in [3] as an algorithm to generate unit norm tight frames for any number of frame vectors N , say, and for any ambient dimension n provided that $\frac{N}{n} \geq 2$. An extension to the construction of unit norm frames having a desired frame operator associated with eigenvalues $\lambda_1, \dots, \lambda_n \geq 2$ satisfying $\sum_{j=1}^n \lambda_j = N$ was then introduced and analyzed in [1] – in fact, an even more general algorithm for the construction of fusion frames was stated therein.

Our main theorem provides a lower bound for the achievable sparsity for a given number of frame vectors and a given frame operator, and also shows that this algorithm indeed generates optimally sparse frames. For stating this result, we will denote the frame constructed by Spectral Tetris applied to the number of frame vectors N and the sequence of eigenvalues $\lambda_1, \dots, \lambda_n$ by $\text{STF}(N; \lambda_1, \dots, \lambda_n)$.

Theorem 3.1 ([2]): Let $n, N > 0$, and let the real values $\lambda_1, \dots, \lambda_n \geq 2$ be ordered blockwise and satisfy $\sum_{j=1}^n \lambda_j = N$. Then the following hold.

- (i) Any frame in $\mathcal{F}(N, \{\lambda_i\}_{i=1}^n)$ has sparsity at least $N + 2(n - \mu(\lambda_1, \dots, \lambda_n))$ with respect to any orthonormal basis.
- (ii) The frame $\text{STF}(N; \lambda_1, \dots, \lambda_n)$ is $N + 2(n - \mu(\lambda_1, \dots, \lambda_n))$ -sparse with respect to the standard unit vector basis, i.e., it is optimally sparse.

ACKNOWLEDGMENT

The first and second author were supported by the grant AFOSR F1ATA00183G003, NSF 1008183, and DTRA/ NSF 1042701. The third author also acknowledges the support of the Hausdorff Center for Mathematics. The fourth author acknowledges support by DFG Grant SPP-1324, KU 1446/13 and DFG Grant, KU 1446/14.

REFERENCES

- [1] R. Calderbank, P. Casazza, A. Heinecke, G. Kutyniok, and A. Pezeshki, *Sparse fusion frames: Existence and construction*, Adv. Comput. Math., to appear.
- [2] P. Casazza, A. Heinecke, F. Krahmer, and G. Kutyniok, *Optimally sparse frames*, preprint.
- [3] P. Casazza, M. Fickus, D. Mixon, Y. Wang, and Z. Zhou, *Constructing tight fusion frames*, Appl. Comput. Harmon. Anal. **30** (2011), 175–187.

Lagrangian Biduality of the ℓ_0 and ℓ_1 -Minimization Problems

Dheeraj Singaraju, Allen Y. Yang and Shankar Sastry
 University of California, Berkeley
 Berkeley, CA 94720

Roberto Tron and Ehsan Elhamifar
 Johns Hopkins University
 Baltimore, MD 21218

I. INTRODUCTION

The last decade has seen a renewed interest in the problem of estimating the sparsest solution in an underdetermined system of equations $A\mathbf{x} = \mathbf{b}$, called ℓ_0 -minimization (ℓ_0 -min):

$$(P_0) \quad \mathbf{x}_0 = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_0 \quad \text{s.t. } A\mathbf{x} = \mathbf{b} \in \mathbb{R}^m, \quad (1)$$

where $A \in \mathbb{R}^{m \times n}$ ($m << n$), and $\|\cdot\|_0$ is the ℓ_0 -semi-norm or the counting norm. The problem of computing \mathbf{x}_0 is known to be NP-hard in general. However, it was observed empirically that the solution to (1) can often be obtained by solving the following convex relaxation, known as ℓ_1 -minimization (ℓ_1 -min):

$$(P_1) \quad \mathbf{x}_1 = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_1 \quad \text{s.t. } A\mathbf{x} = \mathbf{b}. \quad (2)$$

Recently, compressive sensing theory has investigated the equivalence of the solutions of (P_0) and (P_1) by characterizing the set of k -sparse vectors \mathbf{x}_0 that can be recovered by solving (2) with $\mathbf{b} = A\mathbf{x}_0$ [1], [2]. As pointed out in [3], the numerical verification of most conditions for equivalence is not computationally tractable. The work [3] further derived sufficient conditions to verify when *all* the possible k -sparse solutions can be recovered by solving (2), with an emphasis on the numerical feasibility of the verifications. However, it is well known that given a matrix A , it may be possible to recover only a subset of all the possible k -sparse solutions [2].

We believe that there is a need to obtain a *certificate of optimality* of \mathbf{x}_1 , which answers the question: Is $\mathbf{x}_1 = \mathbf{x}_0$? Specifically, it is of interest to produce a *per-instance* certificate of optimality for any candidate solution obtained at runtime by solving (2), rather than certificates for all the possible k -sparse solutions.

Contributions. We present a novel primal-dual analysis of (P_0) . We propose to use the optimal value of the Lagrangian dual function of (P_0) to obtain a non-trivial lower bound for the sparsity of \mathbf{x}_0 . Interestingly, maximizing the Lagrangian dual of (P_0) is equivalent to ℓ_1 -min with additional constraints. Moreover, our analysis can be applied to other problems which involve minimization of the ℓ_0 -semi-norm, such as *Sparse PCA*, to interpret convex relations of the original NP-hard problems as maximizing their Lagrangian duals.

II. PRIMAL-DUAL ANALYSIS OF ℓ_0 -MIN

In this work, we consider the following modified ℓ_0 -min problem:

$$(P_0^*) \quad \mathbf{x}_0^* = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_0 \quad \text{s.t. } A\mathbf{x} = \mathbf{b} \text{ and } \|\mathbf{x}\|_\infty \leq M, \quad (3)$$

and its Lagrangian dual:

$$\begin{aligned} (D_0^*) \quad & \{\delta_1^*, \delta_2^*\} = \operatorname{argmax}_{\{\delta_1 \in \mathbb{R}^n, \delta_2 \in \mathbb{R}^n\}} \left[\mathbf{1}^\top \min \{\mathbf{0}, \mathbf{1} - \delta_1\} + \delta_2^\top \mathbf{b} \right], \\ & \text{s.t. } -\frac{1}{M} \delta_1 \leq A^\top \delta_2 \leq \frac{1}{M} \delta_1 \text{ and } \delta_1 \geq \mathbf{0}. \end{aligned} \quad (4)$$

Notice that if (P_0) has a unique solution \mathbf{x}_0 , we can choose any finite positive valued $M \geq \|\mathbf{x}_0\|_\infty$ to ensure that $\mathbf{x}_0^* = \mathbf{x}_0$. If (P_0)

does not have a unique solution, we may still choose a finite valued $M > 0$ to regularize the desired solution. The constraint $M \geq \|\mathbf{x}_0^*\|_\infty$ is also referred to as the *box constraint*.

Our main result gives a *biduality* relation between (P_0^*) and the following ℓ_1 -min problem with the box constraint:

$$(P_1^*) \quad \mathbf{x}_1^* = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{M} \|\mathbf{x}\|_1 \quad \text{s.t. } A\mathbf{x} = \mathbf{b} \text{ and } \|\mathbf{x}\|_\infty \leq M, \quad (5)$$

where it must be noted that \mathbf{x}_1^* is not necessarily equal to \mathbf{x}_1 .

Theorem 1. (P_1^*) is the Lagrangian dual of (D_0^*) , i.e., it is the Lagrangian bidual (dual of the dual) of (P_0^*) .

It must be noted that the duality gaps of (P_0^*) and (P_1^*) with respect to their dual (D_0^*) are non-zero and zero, respectively.

Corollary 1. Since solving (P_1^*) is equivalent to maximizing the Lagrangian dual function in (D_0^*) , we have $\frac{1}{M} \|\mathbf{x}_1^*\|_1 \leq \|\mathbf{x}_0^*\|_0$.

Corollary 2. Let $M_0 = \|\mathbf{x}_0\|_\infty$, $M_1 = \|\mathbf{x}_1\|_\infty$ and let M be the constant used in (5). We then have (a) solving (P_1^*) with any M that satisfies $M \geq \max\{M_1, M_0\}$ is equivalent to solving (P_1) , and (b) if $M_1 < M_0$, we cannot recover \mathbf{x}_0^* by solving (5) with $M = M_0$.

III. SIMULATION RESULTS

We randomly generate entries of $A \in \mathbb{R}^{128 \times 256}$ and $\mathbf{x}_0 \in \mathbb{R}^{256}$ from a Gaussian distribution with unit variance. The sparsity of \mathbf{x}_0 is varied from 1 to 64. We solve (P_1^*) with $M = M_0, 5M_0$ and $10M_0$ to obtain upper and lower bounds for $\|\mathbf{x}_0\|_0$, as $\|\mathbf{x}_1^*\|_0$ and $\frac{1}{M} \|\mathbf{x}_1^*\|_1$, respectively. Figure 1 shows the results of our simulations.

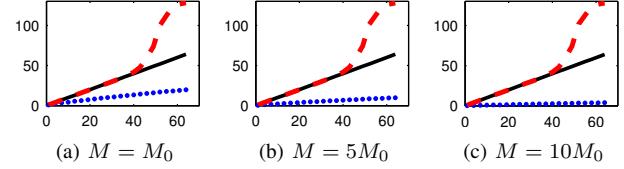


Fig. 1. x-axis: $\|\mathbf{x}_0\|_0$ - sparsity of \mathbf{x}_0 . y-axis: mean values (over 100 trials) of the upper bound (red dashed line), lower bound (blue dotted line) and true value (black solid line) for $\|\mathbf{x}_0\|_0$.

Our lower bounds are tight for extremely sparse \mathbf{x}_0 and are more conservative as the number of non-zero entries in \mathbf{x}_0 increases. These bounds are tighter when the value of M is closer to M_0 . Furthermore, we observed in our simulations that with the same notation as in Corollary 2, if $M_1 \geq M$, then in some cases, we can recover \mathbf{x}_0^* by solving (P_1^*) with $M = M_0$, but not by solving (P_1) .

REFERENCES

- [1] E. Candès. Compressive Sampling. In *Proceedings of the International Congress of Mathematicians*, 2006.
- [2] D. Donoho. For Most Large Underdetermined Systems of Linear Equations, the minimal ℓ_1 -norm near-solution approximates the sparsest near-solution. *Communications on Pure and Applied Mathematics*. 2006.
- [3] A. Iouditski, F. K. Karzan, and A. Nemirovski. Verifiable conditions of ℓ_1 -recovery of sparse signals with sign restrictions. *ArXiv e-prints*, 2009.

Signal Recovery Via ℓ_p Minimization: Analysis using Restricted Isometry Property

Shisheng Huang, Jubo Zhu, Fengxia Yan, Meihua Xie, Zelong Wang, Bo Lin

Department of Mathematics and Systems, College of Science,
National University of Defense Technology, Changsha, 410073, China
Email: huangss@nudt.edu.cn.

Compressed sensing or compressive sensing (CS) is a new protocol to sample signals at a rate proportional to their information content rather than their bandwidth[1], [2], [3], [4]. In a discrete setting, let $\chi^N(K) = \{x \in \mathbb{R}^N : \|x\|_0 \leq K\}$ denote the family of at most K -sparse vectors in \mathbb{R}^N , where $\|x\|_0$ counts the number of nonzero entries of x . The basic theory of CS asserts that one can recover a K -sparse signal $x \in \chi^N(K)$ from relatively few incomplete measurements $y = \Phi x \in \mathbb{R}^M$ for a carefully chosen sampling matrix Φ by solving the ℓ_1 -minimization problem

$$\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{subject to} \quad y = \Phi x \quad (1)$$

where $\|x\|_1$ denotes the ℓ_1 -norm of x . The celebrated works by Candès, et al.[1], [2], [3] and Cai, et al.[5] have shown that if the sampling matrix Φ satisfies the restricted isometry property (RIP) with some order, then solving the convex optimization (1) can recover an arbitrary K -sparse signal x exactly.

There is another strategy which recovers sparse signal by solving the ℓ_p ($0 < p < 1$) minimization[6], [7]

$$\min_{x \in \mathbb{R}^N} \|x\|_p \quad \text{subject to} \quad y = \Phi x \quad (2)$$

where $\|x\|_p = (\sum_{i=1}^N |x_i|^p)^{1/p}$ is the standard ℓ_p quasi-norm of vector x . Furthermore, we can recover the sparse signal by solving

$$\min_{x \in \mathbb{R}^N} \|x\|_p \quad \text{subject to} \quad \|y - \Phi x\|_2 \leq \epsilon \quad (3)$$

in the noisy case, where ϵ represents the noise level.

In this paper, we focus on sparse signal recovery via the ℓ_p minimization, along with the analysis of its performance using RIP.

Theorem 1 Let $x \in \chi^N(K)$ be a K -sparse signal, and $y = \Phi x$ be the linear measurement vector. Let K_1 be a positive integer and

$$\eta = \frac{K^{1/p-1/2}}{K_1^{1/p-1/2}} + \sqrt{\frac{K_1}{K}} > 0$$

Then under the condition

$$\delta_K + \eta \theta_{K,K_1} < 1 \quad (4)$$

solving the ℓ_p minimization problem (2) with $0 < p < 1$ can recover x exactly. In particular, the condition (4) becomes $\delta_{2K} < 1/2$ as $p \rightarrow 0$.

Theorem 2 Let $x \in \chi^N(K)$ be a K -sparse signal, and $y = \Phi x + e$ be the linear measurement vector with $\|e\|_2 \leq \epsilon$. Let K_1 be a positive integer and

$$\eta = \frac{K^{1/p-1/2}}{K_1^{1/p-1/2}} + \sqrt{\frac{K_1}{K}} > 0$$

Then under the condition

$$\delta_K + \eta \theta_{K,K_1} < 1 \quad (5)$$

solving the ℓ_p minimization problem (3) with $0 < p < 1$ recovers x^* satisfying

$$\|x^* - x\|_2 \leq \frac{2\sqrt{2}\sqrt{1+\delta_K}}{1-\delta_K-\eta\theta_{K,K_1}} \epsilon$$

Theorem 3 Let $x \in \mathbb{R}^N$ be an arbitrary signal, and $y = \Phi x + e$ be the linear measurement vector with $\|e\|_2 \leq \epsilon$. Let K_1 be positive integer and

$$\eta = \frac{K^{1/p-1/2}}{K_1^{1/p-1/2}} + \sqrt{\frac{K_1}{K}} > 0$$

Then under the condition

$$\delta_K + \eta \theta_{K,K_1} < 1 \quad (6)$$

solving the ℓ_p minimization problem (3) with $0 < p < 1$ recovers x^* satisfying

$$\begin{aligned} \|x^* - x\|_2 \leq & \frac{2\sqrt{2}\sqrt{1+\delta_K}}{1-\delta_K-\eta\theta_{K,K_1}} \epsilon \\ & + \frac{2\sqrt{2}\theta_{K,K_1}\|x - x_K\|_p}{K_1^{1/p-1/2}(1-\delta_K-\eta\theta_{K,K_1})} \end{aligned} \quad (7)$$

where x_K is the best K -term approximant of x .

According to theorems 1-3, our results show that the ℓ_p minimization can recover sparse signal with good performance provided that the sampling matrix Φ satisfies the RIP with parameter $\delta_{2K} < 1/2$.

In a recent paper, Davies and Gribonval construct examples shown that if $\delta_{2K} \geq 1/\sqrt{2}$, exact recovery of certain K -sparse signal using (1) or (2) can fail in the noiseless case[8]. Blanchard, Cartis and Tanner also discuss the sharpness of the RIC bounds in compressed sensing[9]. We can see that there still has a room for improvement.

REFERENCES

- [1] E. Candès and T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory, 2005, 51(12): 4203-4215.
- [2] E. Candès, J. Romberg and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Commun. Pure Appl. Math., 2006, 59(8): 1207-1223.
- [3] E. Candès, The restricted isometry property and its implications for compressed sensing, C. R. Acad. Sci. Ser. I, 2008, 346: 589-592.
- [4] D. Donoho, Compressed sensing, IEEE Trans. Inf. Theory, 2006, 52(4): 1289-1306.
- [5] T. Cai, L. Wang, and G. Xu, New bounds for restricted isometry constants, IEEE Trans. Inf. Theory, 2010, 56(9): 4388-4394.
- [6] R. Chartrand, V. Staneva, Restricted isometry properties and nonconvex compressive sensing, Inverse Problems, 2008, 24, 035020.
- [7] S. Foucart, and M. Lai, Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q \leq 1$, Applied and Computational Harmonic Analysis, 2009, 26: 395-407.
- [8] M. Davies, R. Gribonval, Restricted isometry constants where ℓ_q sparse recovery can fail for $0 < p \leq 1$, IEEE Trans. Inf. Theory, 2009, 55(5): 2203-2214.
- [9] J. Blanchard, C. Cartis, J. Tanner, Compressed sensing: how sharp is the restricted isometry property?, SIAM Review, 2011, 53(1): 105-125.

Compressed sensing for joint ground imaging and target indication with airborne radar

Ludger Prünne

Fraunhofer Institute for High Frequency Physics and Radar Techniques FHR; Neuenahrer Str. 20, 53343 Wachtberg, Germany

I. INTRODUCTION

The detection of moving vehicles on land surface using airborne or spaceborne radar systems (ground moving target indication, GMTI) is an important topic for military as well as civil applications. This task is often combined with imaging the area via synthetic aperture radar (SAR) [1]. It is desirable to execute this radar operation mode concurrently with other radar observation assignments.

During the last years sparsity and compressed sensing (CS) has come into focus of the radar community, cf. e.g. [2] or [3]. Sparsity principles and CS have also been applied to GMTI problems (cf. [4]), mostly for reconstruction of covariance matrices (cf. [5]) in classical GMTI. Here we apply CS directly for jointly imaging moving and non moving targets. This method indicates the velocity of each scatterer – i.e. of each non zero element of the scenery. The multichannel data are measured in the classical stripmap geometry.

II. MEASUREMENT AND DATA STRUCTURE

The principles of airborne multichannel stripmap radar are as follows: Pulses are transmitted from equidistant positions on a linear flightpath and waves scattered back from earth are received via n antennas – arranged in direction of flight with distances much higher than the distance the platform covers between two pulses – measuring amplitude, phase and delay time. Due to the finite width of the footprint, every target is illuminated by several pulses.

In consequence – after some preprocessing – a single non moving scatterer at $(0, \rho)$ in an otherwise empty scenery generates in pulse p and channel j a signal that is non zero in distance $r = \sqrt{\rho^2 + x_{p,j}^2}$, only. Here $x_{p,j}$ indicates the position of the j th receiving channel at pulse p , representable via $x_{p,j} = cp + c_j$ with constants c and c_j depending on the recording setup. The phase of the received signal is proportional to the distance r , so it is in first approximation a chirp. Altogether we obtain a complex signal following approximately

$$s(j, r, p) = a\delta\left(r - \sqrt{\rho^2 + (cp - c_j)^2}\right) D_{r,p,j} e^{(-i4\pi/\lambda)(cp - c_j)^2/\rho}$$

with amplitude $a \in \mathbb{C}$, $D \in \mathbb{R}^+$ indicating the transmitted energy in direction of the scatterer and wavelength λ , δ denotes the Kronecker delta. The signal of a moving scatterer differs in shape and phase.

The combination between the multichannel pointwise measurements in r and p and chirp like responses of the scatterers is suitable for applying CS, since the coherence between chirps and point basis is low. So we define the measurement matrix Φ by shifted versions of s and its analogons in different velocities. We considered here discretization claims (cf. e.g. [6]) as e.g. the gap between two velocities as well as the distance between two imaging points. Additionally we represented Φ by a linear operator suitable for fast computational methods, necessary due to the high amount of data. To hold the restricted isometry property for a number of scatterers as high as possible, we intended to treat the non moving ones after merging by wavelet transform. Alternatively we suppress them by subtracting channels, shifted according to the different c_j .

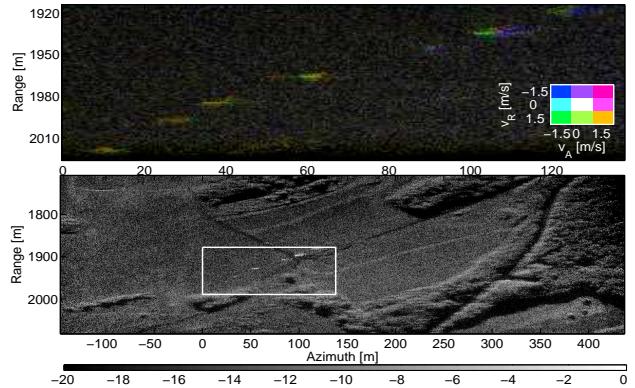


Fig. 1. Upper: Reconstruction of moving targets with velocity and amplitude coding. Lower: CS-SAR image of a wider scenery.

III. NUMERICAL RESULTS WITH REAL RADAR DATA

Our numerical results are obtained by solving the noisy basis pursuit problem $\{\min \|x\|_1 \text{ such that } \|y - \Phi x\|_2 < \sigma\}$ – with the data y , reconstruction x and σ estimated from the known signal to noise ratio – using the SPGL1 algorithm (cf. [7]). The data have been recorded by a 4 channel airborne device, imaging a scenery with several vehicles driving with approximately the same velocity. Experimental results are depicted in Figure 1. The lower one is computed using a pointwise representation for all scatterers without considering their velocities, the scenery is clearly recognizable like in a classical SAR image. In the upper one the detail marked below is considered by suppressing the non moving scatterers as described above. Here direction and velocity of moving scatterers is coded in color, so the convoy is clearly imaged and detectable. This result holds also for regarding a fraction of the data or for simulated data.

REFERENCES

- [1] C. H. Gierull, "Statistical analysis of multilook SAR interferograms for CFAR detection of ground moving targets," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 4, pp. 691–701, Apr 2004.
- [2] J. H. G. Ender, "On compressive sensing applied to radar," *Signal Processing*, vol. 90, no. 5, pp. 1402–1414, May 2010.
- [3] V. M. Patel, G. R. Easley, D. M. Healy Jr, and R. Chellappa, "Compressed synthetic aperture radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 244–254, Apr 2010.
- [4] I. Stojanovic and W. C. Karl, "Imaging of moving targets with multi-static SAR using an overcomplete dictionary," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 1, pp. 164–176, Feb 2010.
- [5] K. Sun, H. Zhang, G. Li, H. Meng, and X. Wang, "Airborne radar STAP using sparse recovery of clutter spectrum," ArXiv:1008.4185v1, 2010.
- [6] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, Oct 2010, in press.
- [7] E. van den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

Automatic target recognition from highly incomplete SAR data

Chaoran Du, Gabriel Rilling, Mike Davies, and Bernard Mulgrew

School of Engineering, University of Edinburgh, Edinburgh, UK

I. INTRODUCTION

In a variety of synthetic aperture radar (SAR) applications only partial data are available. The conventional SAR imaging approaches lead to severe artifacts which dramatically degrade the image quality, making further processing such as target detection and classification difficult. Considering the fact that image reconstruction from incomplete data can be viewed as an underdetermined inverse problem, we here apply compressed sensing (CS) related techniques to realize automatic target recognition (ATR) from partial SAR data, with or without image reconstruction. The impact of various subsampling patterns on ATR performance is also investigated.

II. CS FOR SAR

SAR images cannot be accurately recovered by CS approaches because of the speckle noise, which endows the images with a high entropy. However, SAR images have a very high dynamic range in many situations due to the presence of a few very bright objects, which typically are associated to man-made structures. Such objects generally occupy a small fraction of the image and their corresponding pixels have much larger magnitudes than the background pixels. The facts that a) SAR data are samples of the spatial Fourier transform of the reflectivity field, b) the bright objects are sparse in the image domain, and c) background pixels have much lower magnitudes suggest that the bright objects can be effectively reconstructed from partial data by using CS approaches.

III. ATR FROM INCOMPLETE SAR DATA

A. With image formation:

The test image is first reconstructed from partial SAR data by solving a constrained ℓ_1 norm minimization problem. Then, the mean-squared error (MSE) classifier is utilized for ATR [1]. The MSE classifier is a nearest neighbor classifier, and it compares normalized images in magnitude. This is because variations in intensity may occur for different SAR acquisition geometries. Also, before normalizing the images we set to zero all but the largest N_b pixels. This is because typically the brightest pixels are located within the target part, and the darker pixels constitute the clutter and target shadow.

B. Without image formation:

The smashed-filter (SF) classifier for compressive classification [2] is adopted, which operates directly on observed data. It is similar to the MSE classifier except it compares in the data domain. Note that the comparison is made between the observed data and data corresponding to normalized complex-valued reference images since there is no test image reconstruction.

IV. SIMULATIONS

Images of three types of targets from the MSTAR database are used in simulations, and two independent sets of images at different elevation angles are adopted as the test and reference images. Three patterns subsampling SAR data along the aperture are considered because of their various applications and simple implementation for existing hardware. As shown in Fig. 1, the patterns are: a)

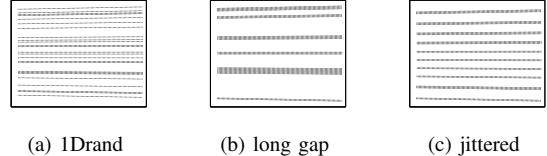


Fig. 1. Sample subsampling patterns with 25% SAR data.

1Drand (subsampling uniformly at random), b) *long gap* (randomly subsampling with the long average gap length constraint), and c) *jittered* (similar to periodically subsampling but the locations and lengths of gaps are slightly jittered). The CS reconstructed images are generated by using the SPGL1 algorithm. For comparison, the test images are also recovered by applying the conventional polar format algorithm (PFA), assuming that all the missing data are zeros. The ATR performance of different scenarios presented in Table I is evaluated by the probability of correct classification Pr_{cc} .

V. DISCUSSION

Compared with conventional SAR imaging methods, CS reconstructed images lead to significantly improved ATR performance since the dominant scatterers can be efficiently recovered. High recognition rates can even be achieved when only a small percentage of data are available. The SF classifier is feasible in theory, but its application in practice is difficult because it can only compare the distances between complex-valued SAR images. This makes it sensitive to the phases of images, which vary too much with even a small change in the observation angle or distance. Although the similarity between the MSE and SF classifier is discussed in [2], the MSE classifier performs much better here because it can compare images in magnitude. The jittered subsampling pattern performs the worst, and a likely explanation is that its similarity to the periodic sampling results in stronger aliasing effect.

TABLE I
ATR PERFORMANCE OF DIFFERENT SCENARIOS (Pr_{cc})

data amount	subsampling pattern	MSE ($N_b = 300$) CS	SF PFA
25% data	1Drand	93.9%	84.7%
	long gap	90.0%	68.3%
	jittered	83.8%	54.1%
10% data	1Drand	76.4%	55.1%
	long gap	75.8%	50.2%
	jittered	58.3%	40.7%

REFERENCES

- [1] M. Cetin, W. C. Karl, and D. A. Castanon, "Feature enhancement and ATR performance using nonquadratic optimization-based SAR imaging," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1375 – 1395, Oct. 2003.
- [2] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Computat. Imag. V*, vol. 6498, San Jose, CA, Jan. 2007, pp. 142–153.

Tomographic SAR Inversion via Sparse Reconstruction

Xiao Xiang Zhu⁽¹⁾

(1) Lehrstuhl für Methodik der Fernerkundung,
Technische Universität München,
Arcisstrasse 21, 80333 Munich, Germany
Email: xiaoxiang.zhu@bv.tum.de

Richard Bamler^(1,2)

(2) Remote Sensing Technology Institute (IMF),
German Aerospace Center (DLR),
Oberpfaffenhofen, D-82234 Wessling, Germany

A conventional spaceborne or airborne Synthetic Aperture Radar (SAR) maps the three-dimensional (3-D) reflectivity distribution of a scene to be imaged into the 2-D azimuth-range ($x - r$) plane. This can be seen as a projection along the third radar coordinate, elevation (s). x , r , and s form an orthogonal coordinate system specific to the particular SAR imaging geometry. This projection particularly handicaps the interpretation of SAR images of 1) volumetric scatterers and 2) of urban areas and man-made objects, i.e. objects with constructive elements oriented at steeper angles than the local incidence angle.

SAR tomography (TomoSAR) extends the synthetic aperture principle of SAR into the elevation direction for 3-D imaging [1]. It uses acquisitions from slightly different viewing angles to reconstruct for every azimuth-range ($x - r$) pixel the reflectivity function along the elevation direction s . It is essentially a spectral analysis problem. Differential SAR tomography (D-TomoSAR) [2], also referred to as 4-D focusing, obtains a 4-D (space-time) map of scatterers by estimating both the elevation and the motion parameters of multiple scatterers inside an azimuth-range pixel.

Modern SAR sensors, e.g. the German TerraSAR-X satellite, provide a very high spatial resolution (VHR) of up to 1 m. This resolution is particularly helpful when it comes to interferometric and tomographic imaging of buildings and urban infrastructure. Although the azimuth-range resolution of this class of very high resolution data reaches sub-meter values, the tight orbit control of modern sensors limits the elevation aperture size and, hence, leads to a low tomographic elevation resolution of typically 30 m, i.e. 10-50 times less than that in azimuth or range. This very unsatisfactorily anisotropic resolution element calls for robust super-resolution (SR) TomoSAR algorithms and it also renders the signal sparse in elevation, i.e. only a few point-like reflections are expected per azimuth-range cell. In order to maintain the original resolution for urban infrastructure imaging, these algorithms should not require averaging in azimuth and range.

Considering the sparsity of the signal in elevation, the compressive sensing (CS) [3] [4] approach to TomoSAR was outlined in [5] where the SR capability of tomographic SAR inversion via L_1 norm regularization and its robustness on elevation estimation against phase noise have been proven using TS-X high resolution spotlight data.

As described in [3], L_1 norm minimization gives the sparsest solution if (and only if) our sparse mapping matrix fulfills the Restricted Isometry Property (RIP) and incoherence properties. However, for our application TomoSAR, RIP and incoherence are violated for several reasons. First the mapping matrix \mathbf{R} is pre-determined by the measurement system (the elevation aperture sampling pattern) and may not be optimum. Second, the reflectivity profile to be reconstructed is often sampled much more densely than the elevation resolution unit in order to allow for good resolution and scatterer positioning accuracy. The small sampling distance renders \mathbf{R} over-complete, reduces RIP and increases coherence. This fact may introduce outliers

in the estimates. In addition, as detailed in [6], the L_1 approximation introduces systematic amplitude biases. Those artifacts are not critical when the aim is only to reconstruct a reasonable reflectivity profile. However, high-quality sparse tomographic SAR inversion requires the estimation of the number of scatterers, as well as the amplitude, phase, and elevation of each scatterer. Hence, special care must be taken of these nuisance artifacts.

In this paper, we proposed a compressive sensing (CS) based "Scale-down by L_1 norm Minimization, Model selection, and Estimation Reconstruction" (SL1MMER, pronounced "slimmer") algorithm to improve the CS estimator and correct for these two deficiencies. SL1MMER combines the advantage of compressive sensing sparse reconstruction (e.g. SR properties and high point localization accuracy) and amplitude and phase estimation accuracy of linear estimation, and hence gives reliable estimation of the number of scatterers, elevation, motion parameters, amplitude and phase of each scatterer. Furthermore, a practical demonstration of the super-resolution of SL1MMER for SAR tomographic reconstruction is provided with a tremendously increased proportion of detected double scatterers from 20% of the conventional linear estimator to 38%.

A systematic performance assessment of the proposed SL1MMER algorithm will be presented in the final paper regarding the elevation estimation accuracy, super-resolution power and robustness. Compared to the Cramér-Rao lower bound, both numeric results and an analytic approximation of the elevation estimation accuracy are provided. It is shown that SL1MMER is an *efficient* estimator. The SR factors are found by extensive simulations. These establish fundamental bounds for super-resolution of spectral estimators. The achievable SR factors of SL1MMER in the typical parameter range of tomographic SAR are found to be promising and are in the order 1.5~25. The minimal number of acquisitions required for a robust estimation are derived and given by explicit formulas.

REFERENCES

- [1] G. Fornaro, F. Serafino, and F. Soldovieri, "Three-dimensional focusing with multipass SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 3, pp. 507–517, 2003.
- [2] F. Lombardini, "Differential tomography: a new framework for SAR interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 37–44, 2005.
- [3] E. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, vol. 3. Citeseer, 2006, p. 14331452.
- [4] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [5] X. Zhu and R. Bamler, "Tomographic SAR Inversion by L_1 -Norm Regularization—The Compressive Sensing Approach," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 10, pp. 3839–3846, 2010.
- [6] ———, "Super-Resolution Power and Robustness of Compressive Sensing for Spectral Estimation with Application to Spaceborne Tomographic SAR," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. accepted, no. pp, 2011.

On the efficiency of proximal methods for CBCT and PET reconstruction with sparsity constraint

Sandrine Anthoine*, Jean-François Aujol**, Yannick Boursier[§] and Clothilde Mélot*

* Aix-Marseille Université, Laboratoire d'Analyse, de Topologie et Probabilités, CNRS, Marseille, France

** Université Bordeaux 1, Institut de Mathématiques de Bordeaux, Talence, France

[§] Aix-Marseille Université, Centre de Physique des Particules de Marseille, CNRS/IN2P3, Marseille, France

I. INTRODUCTION

Cone Beam Computerized Tomography (CBCT) and Positron Emission Tomography (PET) Scans are medical imaging devices that respectively provide anatomical and metabolic complementary information on the patient. X-ray absorption is an intrinsic physical property of biological tissues, but the dose delivery necessary to get an image can theoretically be lowered by improving the detection efficiency of the scanner. Similarly, the dose of active radiotracer injected to the patient before a PET-Scan as well as the duration of the exam have to be lowered. The models considered come directly from the physics of the new generation acquisition devices, and take into account the specificity of the (Poisson) noise.

For the CBCT modality, we denote by $\mu \in \mathbb{R}^{I_1}$ the unknown attenuation vector indexed by $i \in \{1, \dots, I_1\}$ and $\mathbf{y} \in \mathbb{R}^{J_1}$ the measurements indexed by $j \in \{1, \dots, J_1\}$ ($J_1 \ll I_1$). More precisely, if the CBCT camera contains M pixels, a tomographic set of measurements \mathbf{y} is obtained with Θ angles of projection so that $J_1 = M\Theta$. For a monochromatic beam of X-ray, the Beer-Lambert law provides the following acquisition model in a discrete setting : $y_j \sim \mathcal{P}(z_j \exp(-[\mathbf{A}\mu]_j))$, where $\mathcal{P}(\lambda)$ is a Poisson distribution with parameter λ and z_j stands for the number of photons emitted by the source within the solid angle relative to the pixel j . The linear operator \mathbf{A} , called the system matrix, is a numerical implementation of the operators of projection that fully describes the geometry of the acquisition system. The coefficient $a_{i,j}$ of \mathbf{A} typically characterizes the probability that any event occurring on a photon in pixel i will be detected on pixel j . We model the measurements as independently distributed pure Poisson random variables since new generation photon-counting detectors are not affected by dark noise classically modeled by additive Gaussian noise.

We model the data acquisition of PET-Scan in a similar way so that $w_j \sim \mathcal{P}([\mathbf{Bv}]_j)$, where $\mathbf{v} \in \mathbb{R}^{I_2}$ denotes the concentration activity vector to reconstruct, $\mathbf{w} \in \mathbb{R}^{J_2}$ the vector of measurements and \mathbf{B} the system matrix which describes the full properties of the PET-Scan ($J_2 \ll I_2$). Since the Poisson likelihood reads: $P(Y = y|X = x) = \frac{x^y}{y!} \exp(-x)$, the negative log-likelihood \mathcal{L} for each modality is:

$$\mathcal{L}_{CT}(\mu) = \sum_{j=1}^{J_1} \left\{ y_j [\mathbf{A}\mu]_j + z_j \exp(-[\mathbf{A}\mu]_j) \right\} \quad (1)$$

$$\mathcal{L}_{PET}(\mathbf{v}) = \sum_{j=1}^{J_2} \left\{ [\mathbf{Bv}]_j - w_j \log_e([\mathbf{Bv}]_j) \right\} \quad (2)$$

with the notation: $\log_\epsilon(x) = \log(x + \epsilon)$. Since these problems are ill-posed, we add a regularization term J to the data fidelity term \mathcal{L}

and we consider the following problems:

$$\hat{\mu} = \arg \min_{\mu \geq 0} \mathcal{L}_{CT}(\mu) + J(\mu) \quad (3)$$

$$\hat{\mathbf{v}} = \arg \min_{\mathbf{v} \geq 0} \mathcal{L}_{PET}(\mathbf{v}) + J(\mathbf{v}) \quad (4)$$

II. ALGORITHMS AND RESULTS

We propose various fast numerical schemes to compute the solution, depending on the regularization choice. Regularizations based on Total Variation norm and sparsity-inducing ℓ_1 -norm on a tight frame (wavelets, curvelets, etc.) have been investigated. In particular, we show that a new algorithm recently introduced by A. Chambolle and T. Pock is well suited in the PET case when considering non differentiable regularizations. Numerical experiments on simulations and real data for several level of X-ray dose (for CBCT) and radio-tracer dose (for PET) indicate that the proposed algorithms compare favorably with respect to well-established methods in tomography. First results are displayed on Figure 1.

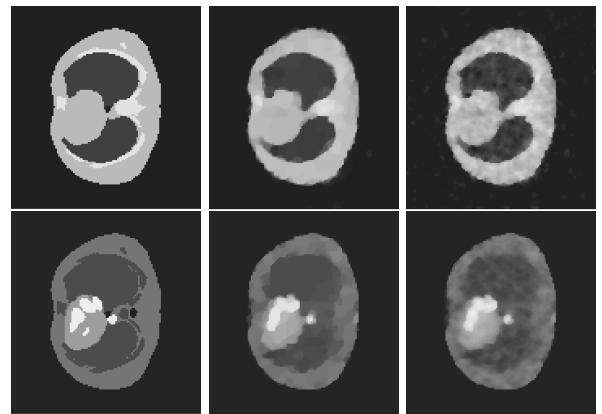


Fig. 1. CBCT (top row) : ground truth (left), reconstruction for 1000 photon counts with a TV reg. (accelerated Forward-Backward splitting) (middle), state-of-the-art (right). PET (bottom row) : ground truth (left), reconstruction pour 2000000 gamma counts with a TV reg. (Chambolle-Pock algo.) (middle), state-of-the-art (right).

REFERENCES

- [1] Z. Harmany, R. Marcia, and R. Willett, “This is SPIRAL-TAP: Sparse poisson intensity reconstruction algorithms theory and practice,” *arXiv:1005.4274*, 2010.
- [2] P.L. Combettes and V. Wajs, “Signal recovery by proximal forward-backward splitting,” *SIAM J. on Multi. Model. and Simu.*, vol. 4, no. 4, 2005.
- [3] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *JMIV*, to appear.
- [4] I. Daubechies, M. Defrise, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Com. P. & A. Math.*, vol. 57, 2004.
- [5] A. Beck and M. Teboulle, “Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems,” *IEEE TIP*, vol. 18(11), 2009.

Reliable Small-object Reconstruction from Sparse Views in X-ray Computed Tomography

Jakob H. Jørgensen*, Emil Y. Sidky† and Xiaochuan Pan†

*Department of Informatics and Mathematical Modeling, Technical University of Denmark,

Richard Petersens Plads, Building 321, 2800 Kongens Lyngby, Denmark. Email: jakj@imm.dtu.dk

†Department of Radiology, University of Chicago, 5841 S. Maryland Ave., Chicago IL, 60637. Email: {sidky,xpan}@uchicago.edu

I. BACKGROUND

Techniques based on Compressive Sensing (CS) are being developed for application in Magnetic Resonance Imaging (MRI) and X-ray Computed Tomography (CT) [1]. CS-inspired Total Variation (TV)-minimization algorithms have been demonstrated to produce accurate CT images from less data than required by standard Filtered Back Projection [2], [3], by exploiting approximate sparsity in the gradient of cross sections of the human body. The amount of data in CT, i.e., the number of measurements, is intrinsically tied to the X-ray dose delivered to the patient. As even a single diagnostic CT scan can lead to significantly increased risk of radiation-induced cancer [4], it is clear that a major reason to pursue CS-inspired CT algorithms is the potential for low-dose X-ray imaging.

One practical consideration for medical imaging based on CS is the extremely large system models involved; in CT, for example, image arrays with 10^9 voxels are standard. Such large systems are challenging to solve accurately in acceptable time. Complicating this issue is the fact that clinically relevant features are often very small – occupying only a few voxels. As result both global and pointwise convergence of algorithms solving CS-based optimization problems may have clinical impact. To demonstrate this issue we examine a realistic simulation of CT for breast cancer screening.

II. THE PRESENT STUDY

Breast CT imaging is being considered as a replacement for mammography in screening for early-stage diagnosing of breast cancer. One particular indicator of breast cancer is formation of *microcalcifications* – very small, highly attenuating calcium deposits. For screening, low-dose imaging is pertinent to minimize accumulated X-ray dose, while accurate and reliable microcalcification shape and attenuation reconstruction is crucial for precise diagnosing.

We consider nonnegativity-constrained TV-regularized image reconstruction in order to exploit gradient sparsity to compensate for the few-view projection data. We investigate in simulation studies, both using ideal data from a discrete model and more realistic data from a continuous model, as well as real CT scanner data, requirements on the number of views, noise level and choice of regularization parameter for accurate reconstruction of small objects.

One concern, in particular for small objects, is that TV-minimization is contrast reducing [5], but this can to some extent be controlled by choice of regularization parameter.

Another practical concern, that we address in the study, is the choice of stopping rule in the minimization algorithm. Two commonly used stopping rules consist of requiring sufficiently low data residual norm or norm difference between successive iterates, but as we demonstrate, these choices can be unreliable for ensuring sufficiently accurate reconstructions. On the other hand rigorous optimization theory-based stopping criteria, such as the KKT conditions, may lead to impractical running times for real data, where dimensions of the CT system matrix of $10^9 \times 10^9$ are not uncommon.

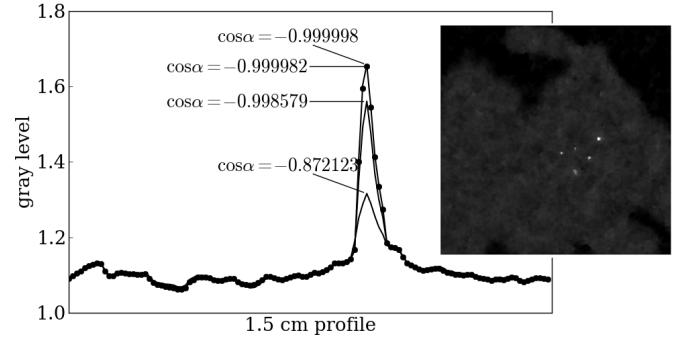


Fig. 1. Profiles through single microcalcification for reconstructions increasingly close to satisfying optimality condition $\cos \alpha = -1$. Inset: 150^2 pixel region of interest of full 2048^2 pixel reconstruction for $\cos \alpha = -0.999998$.

We compare the use of different stopping rules for small object reconstruction, including KKT conditions, the gradient map criterion [6] and the cosine alpha criterion suggested for constrained TV-minimization in CT [3]. As an example we consider a 2048^2 pixel reconstruction of a breast phantom with five small microcalcifications from 64-view data with a 1024-bin detector subject to a realistic noise level. We show four reconstructions increasingly close to satisfying the optimality condition $\cos \alpha = -1$ from [3], where α is the angle between the TV-gradient and the data residual gradient.

We observe a non-uniform convergence across the image, in the sense that although a given iteration may be accurate in most of the image, some features, in particular small objects, may not be. Such inadequate reconstructions can have significant clinical implications, and we discuss strategies to ensure reliable reconstruction.

REFERENCES

- [1] E. J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE TRANSACTIONS ON INFORMATION THEORY*, vol. 52, no. 2, pp. 489–509, 2006.
- [2] E. Y. Sidky, C.-M. Kao, and X. Pan, “Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT,” *JOURNAL OF X-RAY SCIENCE AND TECHNOLOGY*, vol. 14, no. 2, pp. 119–139, 2006.
- [3] E. Y. Sidky and X. Pan, “Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization,” *PHYSICS IN MEDICINE AND BIOLOGY*, vol. 53, no. 17, pp. 4777–4807, 2008.
- [4] D. J. Brenner and E. J. Hall, “Current concepts - Computed tomography - An increasing source of radiation exposure,” *NEW ENGLAND JOURNAL OF MEDICINE*, vol. 357, no. 22, pp. 2277–2284, 2007.
- [5] D. Strong and T. Chan, “Edge-preserving and scale-dependent properties of total variation regularization,” *INVERSE PROBLEMS*, vol. 19, no. 6, pp. S165–S187, 2003.
- [6] L. Vandenberghe, “Optimization methods for large-scale systems,” 2009, Lecture Notes. www.ee.ucla.edu/~vandenbe/ee236c.html.

Near-optimal undersampling and reconstruction for MRI carotid blood flow measurement based on support splitting

Gabriel Rilling*, Yuehui Tao†, Mike E. Davies* and Ian Marshall†

*Institute for Digital Communications, School of Engineering, University of Edinburgh, United Kingdom

†Medical Physics, University of Edinburgh, United Kingdom

I. INTRODUCTION

Magnetic resonance imaging (MRI) raw data is acquired as a set of *lines* in the Fourier domain (aka k -space), typically lines covering all k_x locations for a discrete set of k_y locations. For dynamic MRI, several time frames are recorded to form a movie. Since each line is sampled at the Nyquist rate in the x direction, the dynamic MRI reconstruction can be performed slice by slice for each x location independently. Optimal undersampling strategies for dynamic MRI are generally based on tiling of the signal support in the 2D Fourier domain (y, f), where f is the frequency with respect to time (t) [1], [2]. This is the 2D equivalent of the standard 1D sampling theory. The achievable acceleration, compared to Nyquist sampling of each time frame, is limited by how packable the (y, f) support is with respect to tiling. We here consider a hybrid strategy in which the support is split into two parts that are individually much more packable than the full support. This allows us to reach higher acceleration factors. Compressed sensing (CS) related ideas are then used to detect the support and reconstruct the full signal.

II. PHASE CONTRAST CAROTID BLOOD FLOW MRI

Measurement of carotid blood flow can be achieved via MRI using a technique referred to as “velocity encoding”. This involves acquiring two sets of time frames, applying velocity encoding on one and using the other one as a reference. The velocity information can then be obtained as the phase difference between the two sets of frames. For a typical (y, t) slice at a given x going through the right common carotid artery (RCCA), the signal support in the (y, f) domain can be modelled as a cross (see Fig. 1). The support can be viewed as the combination of a static part with only DC frequency content and a dynamic part that is localized in a small region of interest (ROI) corresponding to the RCCA. Importantly, the location of the dynamic ROI is not known a priori.

III. SPLIT SUPPORT UNDERSAMPLING STRATEGY

We consider a sampling strategy composed of two sampling patterns as shown in Fig. 2. Pattern (a) is adapted to the dynamic part of the support. Assuming the width of the band is B pixels, B parallel lines allow the reconstruction of the dynamic part, provided its location can be detected. Pattern (a) is a multi-coset sampling pattern containing two such sets of parallel lines, which guarantees the detection [3]. Pattern (b) is adapted to the static part. Each k_y location is sampled once, which results in standard Nyquist sampling of that part.

IV. DETECTION OF THE DYNAMIC PART AND SIGNAL RECONSTRUCTION

Assuming an upper bound B on the size of the dynamic ROI is known, the detection of the ROI can be achieved by an exhaustive search strategy assuming the ROI support is made of one *block* of pixels. If the image has N lines, only $N - B + 1$ supports need to

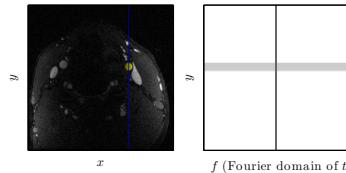


Fig. 1. Left: Carotid slice with the RCCA marked in yellow. An x slice going through the RCCA is marked in blue. Right: signal support model: static signal (dark gray) and dynamic ROI (light gray).

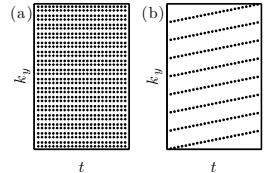


Fig. 2. Sampling patterns (5 \times acceleration). (a) pattern adapted to the dynamic part. (b) pattern adapted to the static part.

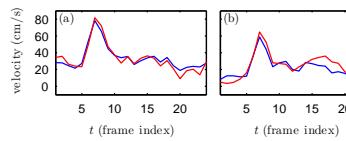


Fig. 3. Velocity estimation (Wiener deconvolution) (red) and reference (blue) for two representative voxels. (a) voxel in the center of the RCCA. (b) voxel close to the artery wall.

be tested which is not tractable. Once the dynamic ROI has been detected, the full static and dynamic (y, f) signal support is known and the full signal can be reconstructed by fitting a signal with that (y, f) support to the measurements from patterns (a) and (b) in the least squares sense. Also a Wiener deconvolution type of solution can be obtained assuming slightly more a priori knowledge on the (y, f) signal and the measurement noise.

The proposed methodology has been simulated by undersampling data from fully sampled acquisitions. Typical performance for two representative voxels is shown in Fig. 3. The accuracy of the estimation is limited by the noise in the data but the general behaviour is well-preserved.

V. POSSIBLE EXTENSIONS

Two simple extensions can be considered. First it is possible to detect and reconstruct multiple dynamic ROIs using more sets of parallel lines in pattern (a) [3]. Second, a low-pass support can be considered instead of the static support. This would allow the application to situations where the MRI signal has one (or several) main dynamic ROI and some less dynamic content otherwise, but not necessarily static.

REFERENCES

- [1] B. Madore, G. H. Glover, and N. J. Pelc, “Unaliasing by fourier-encoding the overlaps using the temporal dimension (unfold), applied to cardiac imaging and fmri,” *Magnetic Resonance in Medicine*, vol. 42, pp. 813–828, 1999.
- [2] N. Aggarwal and Y. Bresler, “Patient-adapted reconstruction and acquisition dynamic imaging method (PARADIGM) for MRI,” *Inverse Problems*, vol. 24, 2008.
- [3] P. Feng, “Universal minimum-rate sampling and spectrum-blind reconstruction for multiband signals,” Ph.D. dissertation, University of Illinois, 1998.

Denoising signal represented by mixtures of multivariate Gaussians in a time-frequency dictionary

Emilie Villaron, Sandrine Anthoine and Bruno Torrésani

Université de Provence - CNRS

Laboratoire d'Analyse, Topologie et Probabilités, UMR CNRS 6632

39 rue Joliot-Curie, 13453 Marseille Cedex 13

Email: {villaron, anthoine, torresan}@cmi.univ-mrs.fr

Abstract—In this paper, we derive an algorithm to denoise a signal generated via its synthesis coefficients on a redundant time-frequency dictionary, the coefficients following a multivariate mixture of Gaussians models. The crux of the problem is that correlations reside both in the model of the coefficients and naturally in the dictionary. We propose and prove the convergence of a “Majorization-Minimization” algorithm to solve this problem. We also advise a particular structure for the dictionary and covariance of the noise that allow to solve efficiently the matrix inversions needed in the algorithm.

I. INTRODUCTION

Signal denoising algorithm based on redundant decompositions over a dictionary usually assume that the coefficients are decorrelated. Although this yield interesting results, this hypothesis may not always be realistic, e.g. for signals propagating in a complex environment. With applications to the analysis of neurophysiological and BCI signals in mind, we tackle the denoising problem when the synthesis coefficients are modeled as multivariate mixture (which would model different brain states).

We propose a denoising algorithm in the case of mixture of correlated Gaussians when the covariances are known and show its convergence. Furthermore, we argue its efficiency when the dictionary has some translation-invariance properties which we define.

II. THE PROBLEM

In \mathbb{R}^R , consider a dictionary represented by its matrix $\Phi \in \mathbb{R}^{R \times T}$ (with $T \geq R$) and observations of the form

$$s = \Phi y + b \quad (1)$$

where $b \in \mathbb{R}^R$ is a zero-mean Gaussian noise with covariance Σ_0 , and $y \in \mathbb{R}^T$ is a random vector whose distribution is a multivariate zero-mean Gaussian mixture. We denote by $\Sigma_1, \dots, \Sigma_K \in \mathbb{R}^{T \times T}$ the covariance matrices and p_1, \dots, p_K the mixing parameters so that the density of y can be written :

$$p(y) = \sum_{k=1}^K p_k \times ((2\pi)^T \det \Sigma_k)^{-\frac{1}{2}} \exp \left(-\frac{1}{2} y^* \Sigma_k^{-1} y \right). \quad (2)$$

The maximisation of the log-likelihood associated with this model and an observed signal s leads to the optimization problem

$$\hat{y} = \arg \min_{y \in \mathbb{R}^T} \frac{1}{2} (s - \Phi y)^* \Sigma_0^{-1} (s - \Phi y) - \log p(y) = \arg \min_{y \in \mathbb{R}^T} L(y), \quad (3)$$

to be numerically solved. The correlations between synthesis coefficients prevents us from using usual approaches to do so.

III. ESTIMATION

In order to address Problem (3), we introduce an “MM” approach (for Majorization-Minimization, see for example [2]) which relies on minimizing a majorizer of L .

Let us set $C(y) = -\log(p(y))$, and $A = \frac{1}{2} \sum_{k=1}^K p_k \Sigma_k^{-1}$. Using Jensen's inequality, one can find a quadratic function Q_t majorizing the negative-log-likelihood L and being tangent to L at the current estimation point y_t . Its minimum can be analytically found and general theorems on MM algorithms [3] lead to:

Proposition 1: Denote $M = (\Phi^* \Sigma_0^{-1} \Phi + 2(A + \lambda I_T))$ and $V_t = \nabla C(y_t) - 2(A + \lambda I_T)y_t$. The iteration

$$y_t \longmapsto y_{t+1} = M^{-1} \left(\Phi^* \Sigma_0^{-1} s - V_t \right) \quad (4)$$

converges to the minimum of L .

Let us note that this update requires the inversion of the matrix $M \in \mathbb{R}^{T \times T}$, where T can be large. Simplifying hypotheses can help alleviate this problem in real situations.

IV. TRANSLATION-INVARIANT DICTIONARIES

To invert M efficiently even when L and T are large, we consider cases where the structure of A , Σ_0 and Φ renders M block circulant. This is in particular the case when 1) Σ_0 is circulant; 2) A is block-diagonal with identical Hermitian blocks and 3) we use doubly indexed dictionaries that are translation-invariant in the sense:

Definition 1: A dictionary Φ is invariant by (circular) translations if the columns of Φ verify

$$\phi_\lambda[k] = \phi_{m,n}[k] = \phi_{0,n}[k-m], \quad m = 0, \dots, M-1, \quad n = 0, \dots, N-1.$$

Note that in this case the Gram matrix of the dictionary is block-circulant (i.e. $(\Phi^* \Phi)_{\lambda\lambda'} = \langle \phi_{m,n}, \phi_{m',n'} \rangle = \langle \phi_{m-m',n}, \phi_{0,n'} \rangle$). Examples of such dictionaries are: translation-invariant wavelet frames, real Gabor dictionaries that may be translation invariant in time or in frequency. M being block-circulant, it may be diagonalized using the Fourier transform [1], yielding $M = \mathbf{F}^* \mathbf{P} \mathbf{F}$ with \mathbf{P} invertible block-diagonal and \mathbf{F} being the Kronecker product of the standard Fourier transform and the identity. Eq. (4) is then solved via

$$\mathbf{P} z = x_t, \quad \text{where } z = \mathbf{F} y \text{ and } x_t = \mathbf{F} \left(\frac{1}{\sigma_0^2} \Phi^* s - V_t \right). \quad (5)$$

V. CONCLUSIONS

This paper shows that denoising signals on a redundant dictionary taking into account both the correlation of the coefficients and that of the noise is possible when the structure of the dictionary and noise are compatible. Here the considered model is a mixture of correlated Gaussians but this work may be extended to other similar models.

REFERENCES

- [1] T. De Mazancourt et D. Gerlic, The inverse of a block-circulant matrix, *IEEE Trans. Ant. and Prop.*, **31**:5 (1983).
- [2] M. Figueiredo et al. Majorization-minimization algorithms for wavelet-based image restoration, *IEEE Trans. Im. Proc.*, **16**:12 (2007).
- [3] D.R. Hunter et K. Lange, A Tutorial on MM algorithms, *The American Statistician* **58** (2004).

Efficiency of Randomized Coordinate Descent Methods on Minimization Problems with a Composite Objective Function

Martin Takáč

School of Mathematics

The University of Edinburgh

Email: M.Takac@sms.ed.ac.uk

Peter Richtárik

School of Mathematics

The University of Edinburgh

Email: Peter.Richtarik@ed.ac.uk

Abstract—We develop a randomized block-coordinate descent method for minimizing the sum of a smooth and a simple nonsmooth block-separable convex function and prove that it obtains an ϵ -accurate solution with probability at least $1 - \rho$ in at most $O((2n/\epsilon) \log(1/\rho))$ iterations, where n is the dimension of the problem. This extends recent results of Nesterov [2], which cover the smooth case, to composite minimization, and improves the complexity by a factor of 4. In the smooth case we give a much simplified analysis. Finally, we demonstrate numerically that the algorithm is able to solve various ℓ_1 -regularized optimization problems with a billion variables.

I. INTRODUCTION

We consider the unconstrained convex optimization problem

$$\min_{x \in \mathbb{R}^N} F(x) \stackrel{\text{def}}{=} f(x) + \Psi(x), \quad (1)$$

where f is smooth and Ψ is block-separable. By x^* we denote an arbitrary optimal solution of (1) and by F^* the optimal value.

A. Block structure

Let $(\mathcal{U}_1, \dots, \mathcal{U}_n)$ be a block decomposition of (a column permutation of) the $N \times N$ identity matrix, with $\mathcal{U}_i \in \mathbb{R}^{N \times N_i}$ and $\sum_{i=1}^n N_i = N$. Any $x \in \mathbb{R}^N$ can then be represented as $x = \sum_{i=1}^n \mathcal{U}_i x^{(i)}$, where $x^{(i)} \in \mathbb{R}^{N_i}$, and we will write $x = (x^{(1)}; \dots; x^{(n)})$. Let $\|\cdot\|_{(i)}, \|\cdot\|_{(i)}^*$ be a pair of conjugate Euclidean norms in \mathbb{R}^{N_i} .

Smoothness of f means that the gradient of $t \mapsto f(x + \mathcal{U}_i t)$ is Lipschitz at $t = 0$, uniformly in x for all i , with constants $L_i > 0$:

$$\|\mathcal{U}_i^T [f'(x + \mathcal{U}_i t) - f'(x)]\|_{(i)}^* \leq L_i \|t\|_{(i)}, \quad x \in \mathbb{R}^N, \quad t \in \mathbb{R}^{N_i}. \quad (2)$$

Block separability of Ψ means that $\Psi(x) = \sum_{i=1}^n \Psi_i(x^{(i)})$.

B. Examples of Ψ

- **Unconstrained smooth minimization:** $\Psi(x) \equiv 0$. Iteration complexity analysis in this case was done in [2]. Our results (not in this abstract) are slightly better and analysis much simpler.
- **Block-constrained smooth minimization:** $\Psi_i(x) \equiv$ indicator function of some convex set in \mathbb{R}^{N_i} .
- **ℓ_1 -regularized minimization:** $\Psi(x) \equiv \lambda \|x\|_1$. In machine learning, this helps to prevent model over-fitting [1] and in compressed sensing this is used to recover sparse signals [3].

II. THE ALGORITHM AND ITS ITERATION COMPLEXITY

Let us define a norm on \mathbb{R}^N by $\|x\|_L = (\sum_{i=1}^n L_i \|x^{(i)}\|_{(i)}^2)^{\frac{1}{2}}$.

Theorem 1. Choose $x_0 \in \mathbb{R}^N$ and $0 < \epsilon < 2\mathcal{R}_L^2(x_0)$, where $\mathcal{R}_L^2(x_0) = \max_x \{\max_{x^*} \|x - x^*\|_L^2 : F(x) \leq F(x_0)\}$. Further, pick $\rho \in (0, 1)$ and let

$$k \geq \frac{2n\mathcal{R}_L^2(x_0)}{\epsilon} \log \left(\frac{F(x_0) - F^*}{\rho\epsilon} \right).$$

If x_k is the random vector generated by Algorithm 1, then $\text{Prob}(F(x_k) - F^* \leq \epsilon) \geq 1 - \rho$.

Algorithm 1 Uniform Coordinate Descent for Composite Functions

for $k = 0, 1, 2, \dots$ **iterate**

 Choose $i_k = i \in \{1, 2, \dots, n\}$ with probability $\frac{1}{n}$

$$T^{(i)} = \arg \min_{t \in \mathbb{R}^{N_i}} \langle \nabla f(x_k), \mathcal{U}_i t \rangle + \frac{L_i}{2} \|t\|_{(i)}^2 + \Psi(x_k + \mathcal{U}_i t)$$

$$x_{k+1} = x_k + \mathcal{U}_i T^{(i)}$$

III. NUMERICAL RESULTS

We will apply Algorithm 1 to random instance of (1) with

$$f(x) = \frac{1}{2} \|Ax - b\|_2^2, \quad \Psi(x) = \|x\|_1, \quad (3)$$

where $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, $N = n$.

In the first table below we present duration time (in seconds) of n iterations of Algorithm 1 applied to problem (1), (3) with a sparse solution x^* and random sparse matrix A . By $\|\cdot\|_0$ we denote number of nonzero elements.

$\ x^*\ _0$	$\ A\ _0 = 10^8$		$\ A\ _0 = 10^9$	
	$10^7 \times 10^6$	$10^8 \times 10^7$	$10^7 \times 10^6$	$10^8 \times 10^7$
16×10^2	5.89	11.04	46.28	70.48
16×10^3	5.83	11.59	46.07	59.03
16×10^4	4.28	8.64	46.93	77.44

Let us remark that $n = 10^7$ iterations in case when $m = 10^8$ and A has a billion nonzeros are executed in about 1 minute. In order to get a solution with accuracy $\epsilon = 10^{-5}$, one needs approximately $40 \times n$ iterations. In the next table we illustrate, on a random problem with $m = 10^7$, $n = 10^6$, $\|A\|_0 = 10^8$ and $\|x^*\|_0 = 16 \times 10^2$, the typical behavior of the method in reducing the gap $F(x_k) - F^*$.

k/n	$F(x_k) - F^*$	$\ x_k\ _0$	time [sec.]
0.0010	$< 10^{-16}$	857	0.01
15.2320	$< 10^{-10}$	997944	65.19
20.6150	$< 10^8$	978761	88.25
25.9120	$< 10^6$	763314	110.94
30.6620	$< 10^4$	57991	131.25
35.0520	$< 10^2$	2538	150.02
38.2650	$< 10^0$	1633	163.75
40.9880	$< 10^{-1}$	1604	175.38
42.7140	$< 10^{-4}$	1600	182.77
44.8600	$< 10^{-6}$	1600	191.94

REFERENCES

- [1] K.-W. Chang, C.-J. Hsieh, and C.-J. Lin. Coordinate descent method for large-scale l2-loss linear support vector machines. *Journal of Machine Learning Research*, 9:1369–1398, 2008.
- [2] Y. Nesterov. Efficiency of coordinate descent methods on huge-scale optimization problems. CORE Discussion Paper 2010/2.
- [3] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo. Sparse reconstruction by separable approximation. *Trans. Sig. Proc.*, 57:2479–2493, July 2009.

Robust sparse recovery with non-negativity constraints

Martin Slawski and Matthias Hein
Department of Computer Science, Saarland University
Email: {ms, hein}@cs.uni-saarland.de

Abstract—It has been established recently that sparse non-negative signals can be recovered using non-negativity constraints only. This result is obtained within an idealized setting of exact sparsity and absence of noise. We propose non-negative least squares – without any regularization – followed by thresholding for the noisy case. We develop conditions under which one can prove a finite sample result for support recovery and tackle the case of an approximately sparse target. Under weaker conditions, we show that non-negative least squares is consistent for prediction. As illustration, we present a feature extraction problem from Proteomics.

I. INTRODUCTION

In various applications, the sparse target $\beta^* \in \mathbb{R}^p$ to be recovered is known to be non-negative. Several recent papers discuss to what extent this additional prior knowledge may simplify the problem of recovering β^* from n , $n < p$, uncorrupted linear measurements $y = X\beta^*$. In [1], [2], [3], it is pointed out that ℓ_1 -minimization is no longer needed if the set $A = \{\beta : y = X\beta, \beta \succeq 0\}$ is a singleton. Donoho and Tanner [2] study the faces of the cone $X\mathbb{R}_+^p$ generated by the columns of X , showing that for random matrices with entries from a symmetric distribution, A fails to be a singleton with high probability if $n < 2p$ already for $s = 0$, where $s = |S|$, $S = \{j : \beta_j^* > 0\}$. On the other hand, they show that with X as the concatenation of a row of ones and a random Gaussian matrix \tilde{X} , the faces of $X\mathbb{R}_+^p$ are in a one-to-one relation with those of $\tilde{X}T^{p-1}$, where T^{p-1} is the standard simplex in \mathbb{R}^p , i.e. A is a singleton if and only if $\operatorname{argmin}_{\beta \in \tilde{A}} \mathbf{1}^\top \beta$, $\tilde{A} = \{\beta : \tilde{X}\beta^* = \tilde{X}\beta, \beta \succeq 0\}$ is. A similar result is shown in [3] with \tilde{X} replaced by a random binary matrix. In [4], we have generalized these two positive results to concatenations of random isotropic sub-Gaussian matrices and a row of ones as well as to random matrices with entries from a sub-Gaussian distribution on \mathbb{R}_+ . A major shortcoming of these results is that they are derived within a little realistic noise-free setting, and it is unclear how they can be transferred to the noisy case. Contradicting the well-established paradigm in statistics suggesting that a regularizer is necessary to prevent over-adaptation to noise, we show that such a transfer is indeed possible.

II. SPARSE RECOVERY FOR THE NOISY CASE

A. Approach

In [4], we assume that $y = X\beta^* + \varepsilon$, where ε is zero-mean sub-Gaussian noise with parameter σ . We suggest to find a minimizer $\hat{\beta}$ of the non-negative least squares (NNLS) criterion $\min_{\beta \geq 0} \|y - X\beta\|_2^2$ first, and to estimate the support S of β^* by $\hat{S}(\lambda) = \{j : \hat{\beta}_j(\lambda) > 0\}$, where $\hat{\beta}(\lambda)$ is obtained by hard thresholding $\hat{\beta}$ with threshold $\lambda \geq 0$, i.e. all components of $\hat{\beta}$ smaller than λ are set to zero.

B. Key condition and main result

In the noiseless case, S can be recovered if $X_S\mathbb{R}_+^s$ is a face of $X\mathbb{R}_+^p$, i.e. there exists a hyperplane separating the cone generated by the columns of the support $\{X_j\}_{j \in S}$ from the cone generated by

the columns of the off-support $\{X_j\}_{j \in S^c}$. For the noisy case, we employ a quantitative notion of separation captured by the constant

$$\widehat{\tau}(S) = \max_{\tau, w: \|w\|_2 \leq 1} \tau \text{ s.t. } X_S^\top w = 0, n^{-1/2} X_{S^c}^\top w \succeq \tau \mathbf{1}.$$

From convex duality, it is easy to see that $\widehat{\tau}(S)$ equals the distance of the subspace spanned by X_S and the simplex generated by X_{S^c} . Based on this relation, we investigate how $\widehat{\tau}(S)$ scales in dependency of n, p, s . We find that $\widehat{\tau}^2(S)$ is of the order s^{-1} minus a random deviation term for the random designs well-suited for sparse recovery in the noiseless case as mentioned in Section 1. A brief, qualitative version of our main result is as follows.

Theorem. Set $\lambda > \frac{2\sigma}{\widehat{\tau}^2(S)} \sqrt{\frac{2 \log p}{n}}$. If $\min_{j \in S} \beta_j^* > \tilde{\lambda}$, $\tilde{\lambda} = \lambda C(S)$, for a constant $C(S)$, $\hat{\beta}(\lambda)$ satisfies $\|\hat{\beta}(\lambda) - \beta^*\|_\infty \leq \tilde{\lambda}$, and $\hat{S}(\lambda) = S$, with high probability.

III. APPROXIMATELY SPARSE TARGETS

Using a lower bound on $\widehat{\tau}(S)$ again, we can bound the reconstruction error as long as β^* is concentrated on components in S .

IV. PREDICTION CONSISTENCY

We show that for a broad classes of non-negative designs, NNLS possesses a ‘self-regularizing property’ which prevents over-adaption to noise. For these designs, the mean square prediction error $n^{-1} \|X\hat{\beta} - X\beta^*\|_2^2$ is upper bounded by a term of order $O(\|\beta^*\|_1 \sqrt{\log p/n})$, a result resembling that obtained in [5] for ℓ_1 -regularized least squares.

V. APPLICATION

An important challenge in the analysis of protein mass spectrometry data is to extract peptide masses from a raw spectrum. In [6], this is formulated as a sparse recovery problem with non-negativity constraints in the presence of heteroscedastic noise. It is demonstrated that NNLS plus thresholding with a locally adaptive threshold outperforms standard sparse recovery methods.

REFERENCES

- [1] A. Bruckstein, M. Elad, and M. Zibulevsky, “On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations,” *IEEE Trans. Inf. Theory*, vol. 54, pp. 4813–4820, 2008.
- [2] D. Donoho and J. Tanner, “Counting the faces of randomly-projected hypercubes and orthants, with applications,” *Disc. Comp. Geometry*, vol. 43, pp. 522–541, 2010.
- [3] M. Wang, W. Xu, and A. Tang, “A unique nonnegative solution to an undetermined system: from vectors to matrices,” *IEEE Trans. Signal Proc.*, vol. 59, pp. 1007–1016, 2011.
- [4] M. Slawski and M. Hein, “Non-negative least squares for sparse recovery in the presence of noise,” In preparation, 2011.
- [5] E. Greenshtein and Y. Ritov, “Persistence in high-dimensional linear predictor selection and the virtue of overparametrization,” *Bernoulli*, vol. 6, pp. 971–988., 2004.
- [6] M. Slawski, R. Hussong, A. Tholey, T. Jakoby, B. Gregorius, A. Hildebrandt, and M. Hein, “Peak pattern deconvolution for Protein Mass Spectrometry by Non-Negative Least Squares/Least Absolute Deviation template matching,” submitted, 2011.

Sparse Subspace Clustering

Ehsan Elhamifar
Johns Hopkins University

René Vidal
Johns Hopkins University

Abstract—We propose a new approach to subspace clustering based on sparse representation. We exploit the fact that each data point in a union of subspaces can always be written as a sparse linear or affine combination of points in its own subspace. This allows us to build a similarity matrix, from which the segmentation of the data can be easily obtained using spectral clustering. We show that under mild assumptions on the principal angles between subspaces and the distribution of the data, the sparsest representation can be found efficiently by solving a (convex) ℓ_1 optimization problem. Our work extends the sparse representation theory from one to multiple subspaces without the assumption of uniqueness of the representation. Also, our approach has the following advantages over the state of the art: it is computationally efficient, it requires no initialization, can deal with both linear and affine subspaces, can handle points near the intersections, noise, outliers, and missing data. We also show that our algorithm significantly outperforms existing motion segmentation algorithms on 167 sequences.

I. INTRODUCTION

Subspace clustering is an important problem with numerous applications in image processing (e.g., image representation and compression) and computer vision (e.g., image/motion/video segmentation). Given a set of points drawn from a union of linear or affine subspaces, the task is to find the number of subspaces, their dimensions, a basis for each subspace, and the segmentation of the data. Over the past years, several subspace clustering algorithms have been proposed (see [1]). Among them, methods based on sparse representation [2], [3] are gaining significant prominence, because of their ability to handle noise, outliers and missing information. This paper discusses the sparse subspace clustering (SSC) algorithm, which is a subspace clustering method based on the sparse representation theory.

II. SPARSE SUBSPACE CLUSTERING

Let $\{\mathbf{y}_i\}_{i=1}^N$ be a collection of $N = \sum_{i=1}^n N_i$ points drawn from arrangement of n linear subspaces of \mathbb{R}^D , $\{\mathcal{S}_i\}_{i=1}^n$, of dimensions $\{d_i \ll D\}_{i=1}^n$. Let the columns of $\mathbf{Y}_i \in \mathbb{R}^{D \times N_i}$ denote the N_i points drawn from subspace \mathcal{S}_i and let $\mathbf{Y} = [\mathbf{Y}_1, \dots, \mathbf{Y}_n]^T \Gamma$ be the matrix containing all the data points, where $\Gamma \in \mathbb{R}^{N \times N}$ is an unknown permutation matrix which specifies the segmentation of data. We assume that we do not know *a priori* the bases for each one of the subspaces nor do we know which data points belong to which subspace. The *subspace clustering* problem refers to the problem of finding the number of subspaces, their dimensions, a basis for each subspace, and the segmentation of data from the matrix \mathbf{Y} alone.

SSC is based on the observation that each data point in a subspace can always be written as a linear combination of all the other data points. However, the *sparsest* representation is obtained when the point is written as a linear combination of points in its own subspace. Given a sparse representation for each data point, the sparse coefficients are used to build a similarity matrix from which the segmentation of data is obtained by spectral clustering.

In the following theorem we prove that for disjoint subspaces (each pair of subspaces intersect only at the origin), under appropriate condition on the principal angles between subspaces and distribution of data, the ℓ_1 minimization finds the sparse representation of each data point as a linear combination of points from the same subspace.

Algorithm 1 Sparse Subspace Clustering (SSC)

Input: A set of points $\{\mathbf{y}_i\}_{i=1}^N$ lying in n subspaces $\{\mathcal{S}_i\}_{i=1}^n$.

1: For every point \mathbf{y}_i , solve the following optimization problem:

$$\min \|\mathbf{c}_i\|_1 \quad \text{subject to} \quad \mathbf{y}_i = \mathbf{X}_i \mathbf{c}_i \quad (1)$$

where $\mathbf{X}_i = [\mathbf{y}_1, \dots, \mathbf{y}_{i-1}, \mathbf{y}_{i+1}, \dots, \mathbf{y}_N]$.

2: Form a similarity graph with N nodes representing the N data points. Connect node i , representing \mathbf{y}_i , to node $j \neq i$, representing \mathbf{y}_j , by edge weights equal to $|c_{ij}| + |c_{ji}|$.

3: Form the Laplacian matrix $\mathbf{L} \in \mathbb{R}^{N \times N}$ and apply K-means to the n smallest eigenvectors of \mathbf{L} .

Output: Segmentation of the data: $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$.

Theorem 1: Given N data points drawn from n subspaces $\{\mathcal{S}_i\}_{i=1}^n$ of dimensions $\{d_i\}_{i=1}^n$, let \mathbf{Y}_i denote the data points on \mathcal{S}_i and $\widehat{\mathbf{Y}}_i$ denote the data points on the other subspaces. Let \mathbb{W}_i be the set of all full rank submatrices $\check{\mathbf{Y}}_i \in \mathbb{R}^{D \times d_i}$ of \mathbf{Y}_i . If the sufficient condition

$$\max_{\check{\mathbf{Y}}_i \in \mathbb{W}_i} \sigma_{d_i}(\check{\mathbf{Y}}_i) > \sqrt{d_i} \max_{j \neq i} \cos(\theta_{ij}) \quad (2)$$

is satisfied for all $i \in \{1, \dots, n\}$, then for every nonzero $\mathbf{y} \in \mathcal{S}_i$, the solution to the following optimization problem

$$\begin{bmatrix} \mathbf{c}_i^* \\ \widehat{\mathbf{c}}_i^* \end{bmatrix} = \operatorname{argmin} \left\| \begin{bmatrix} \mathbf{c}_i \\ \widehat{\mathbf{c}}_i \end{bmatrix} \right\|_1 \quad \text{subject to } \mathbf{y} = [\mathbf{Y}_i, \widehat{\mathbf{Y}}_i] \begin{bmatrix} \mathbf{c}_i \\ \widehat{\mathbf{c}}_i \end{bmatrix} \quad (3)$$

gives the sparse subspace solution with $\mathbf{c}_i^* \neq 0$ and $\widehat{\mathbf{c}}_i^* = 0$.

III. APPLICATION TO MOTION SEGMENTATION

We apply SSC to the problem of separating a video sequence into multiple spatiotemporal regions corresponding to different rigid-body motions in the scene. Under the affine projection model, the motion segmentation problem can be cast as clustering a collection of point trajectories according to multiple affine subspaces. Table I compares SCC with other subspace clustering methods on the Hopkins155 motion database, a database of 155 sequences of two and three motions, available online at <http://www.vision.jhu.edu/data/hopkins155>. Clearly, SSC outperforms state-of-the-art methods.

TABLE I
CLASSIFICATION ERRORS (%) FOR SEQUENCES WITH 2 MOTIONS

	GPCA	LLMC	LSA	SCC	RANSAC	MSL	ALC	SSC
Checkerboard	6.09	3.96	2.57	1.30	6.52	4.46	1.55	1.12
Traffic	1.41	3.53	5.43	1.07	2.55	2.23	1.59	0.02
Articulated	2.88	6.48	4.10	3.68	7.25	7.23	10.70	0.62
All	4.59	4.08	3.45	1.46	5.56	4.14	2.40	0.75

REFERENCES

- [1] R. Vidal, “Subspace clustering,” *SPM*, vol. 28, no. 2, pp. 52–68, 2011.
- [2] E. Elhamifar and R. Vidal, “Sparse subspace clustering,” in *CVPR*, 2009.
- [3] ———, “Clustering disjoint subspaces via sparse representation,” in *IEEE ICASSP*, 2010.

Subspace Clustering by Rank Minimization

Paolo Favaro
Heriot-Watt University

Avinash Ravichandran
University of California Los Angeles

René Vidal
Johns Hopkins University

Abstract—We consider the problem of fitting multiple subspaces to a cloud of points drawn from the subspaces and corrupted by noise/outliers. We propose a rank minimization approach that decomposes the corrupted data matrix as the sum of a clean dictionary and a matrix of noise/outliers. By constraining the dictionary elements to be expressible as a linear combination of each other, we formulate the problem as one of minimizing the nuclear norm of the matrix of linear combinations. For noisy data, this problem can be solved in closed form by applying a polynomial thresholding to the SVD of the data. For one subspace, our framework reduces to classical PCA. For multiple subspaces, our framework provides an affinity matrix that can be used to cluster the data according to the subspaces. For data corrupted by outliers, we use an augmented Lagrangian approach, which requires a combination of our proposed polynomial thresholding operator with the more traditional shrinkage-thresholding operator.

I. INTRODUCTION

Subspace estimation and clustering are very important problems with widespread applications in computer vision and pattern recognition. This has motivated the development of a number of techniques based on sparse representation theory and rank minimization [1], [2], [3], [4]. For instance, [3] shows that a point in a union of independent subspaces admits a sparse representation with respect to the dictionary formed by all other data points, such that the nonzero coefficients correspond to other points in the same subspace. Moreover, the nonzero coefficients can be obtained as the solution of

$$\min_{C,E} \|C\|_1 + \frac{\alpha}{2} \|E\|_F^2 \quad \text{s.t. } D = DC + E \quad \text{and} \quad \text{diag}(C) = 0, \quad (1)$$

where D is the data matrix, E represents the noise and C is the matrix of coefficients. These nonzero coefficients are then used to cluster the data according to the multiple subspaces. A very similar approach is presented in [4]. The major difference is that a low-rank representation is used in lieu of the sparsest representation, i.e.

$$\min_C \|C\|_* + \alpha \|E\|_{2,1} \quad \text{s.t. } D = DC + E, \quad (2)$$

where $\|E\|_{2,1} = \sum_{k=1}^N \sqrt{\sum_{j=1}^N |E_{jk}|^2}$ is the $\ell_{2,1}$ norm of E .

II. SUBSPACE CLUSTERING IN THE PRESENCE OF NOISE

In this section, we propose the following rank minimization approach to subspace clustering in the presence of noise:

$$\min_{A,C,E} \|C\|_* + \frac{\alpha}{2} \|E\|_F^2 \quad \text{s.t. } A = AC \quad \text{and} \quad D = A + E. \quad (3)$$

While in principle this problem appears to be very similar to those in (1) and (2), there are a number of key differences. First, rather than expressing the noisy data as a linear combination of itself + noise, i.e., $D = DC + E$, we search for a clean dictionary A , which it is self-expressive, i.e., $A = AC$. We then assume that the data is obtained by adding noise to the clean dictionary, i.e., $D = A + E$. Thus, our method searches simultaneously for a clean dictionary, the sparse coefficients and the noise. Second, the main difference with (1) is that the ℓ_1 norm of C is replaced by the nuclear norm, and the main difference with (2) is that the $\ell_{2,1}$ norm of E is replaced by the Frobenius norm. As we will show, these changes result in a key difference between our method and the state of the art: while the

solutions to (1) and (2) require convex optimization, the solution to (3) can be computed in closed form from the SVD of D . The proof of this result will be done in three steps. In Lemma 1 we will relax the constraint $A = AC$ and add a penalty $\frac{\tau}{2} \|A - AC\|_F^2$ to the cost. We will show that the optimal solution for C , with A kept fixed, can be obtained in closed form from the SVD of A . Since the optimal E is $D - A$, we will not consider the term $\frac{\alpha}{2} \|E\|_F^2$. Then, in Lemma 2 we will optimize the relaxed cost over both A and C and show that the optimal A can be obtained in closed form by applying a polynomial thresholding to the SVD of D . Finally, in Lemma 3 we will show that the solution to (3) is given by classical PCA, except that the number of principal components can be automatically determined.

Lemma 1: Let $A = UAV^T$ be the SVD of A . The optimal solution to $\min_C \|C\|_* + \frac{\tau}{2} \|A - AC\|_F^2$ is $\widehat{C} = V_1(I - \frac{1}{\tau}\Lambda_1^{-2})V_1^T$, where $U = [U_1 \ U_2]$, $\Lambda = \text{diag}(\Lambda_1, \Lambda_2)$ and $V = [V_1 \ V_2]$ are partitioned according to $\mathbf{I}_1 = \{i : \lambda_i > 1/\sqrt{\tau}\}$ and $\mathbf{I}_2 = \{i : \lambda_i \leq 1/\sqrt{\tau}\}$.

Lemma 2: Let $D = U\Sigma V^T$ be the SVD of the data matrix D . The optimal solution to $\min_{A,C} \|C\|_* + \frac{\tau}{2} \|A - AC\|_F^2 + \frac{\alpha}{2} \|D - A\|_F^2$ is given by $\widehat{A} = U\Lambda V^T$ and $\widehat{C} = V_1(I - \frac{1}{\tau}\Lambda_1^{-2})V_1^T$, where each entry of $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ is obtained from one entry of $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ as the solution to

$$\sigma = \psi(\lambda) = \begin{cases} \lambda + \frac{1}{\alpha\tau}\lambda^{-3} & \text{if } \lambda > 1/\sqrt{\tau} \\ \lambda + \frac{\tau}{\alpha}\lambda & \text{if } \lambda \leq 1/\sqrt{\tau} \end{cases}, \quad (4)$$

that minimizes the cost, and $U = [U_1 \ U_2]$, $\Lambda = \text{diag}(\Lambda_1, \Lambda_2)$ and $V = [V_1 \ V_2]$ are partitioned according to \mathbf{I}_1 and \mathbf{I}_2 .

Lemma 3: Let $D = U\Sigma V^T$ be the SVD of the data matrix D . The optimal solution to $\min_{A,C} \|C\|_* + \frac{\alpha}{2} \|D - A\|_F^2$ s.t. $A = AC$ is given by $\widehat{A} = U_1\Sigma_1 V_1^T$ and $\widehat{C} = V_1 V_1^T$, where Σ_1 , U_1 and V_1 correspond to the top $r = \arg \min_k k + \frac{\alpha}{2} \sum_{i>k} \sigma_i^2$ singular values and singular vectors of D , respectively.

III. SUBSPACE CLUSTERING IN THE PRESENCE OF OUTLIERS

In this section, we propose the following rank minimization approach to subspace clustering in the presence of outliers:

$$\min_{A,C,E} \|C\|_* + \frac{\alpha}{2} \|D - A - E\|_F^2 + \langle Y, D - A - E \rangle + \gamma \|E\|_1 \quad (5)$$

It follows from Lemma 3 that the optimal solution for C and A such that $A = AC$ is $A = U_1\Lambda_1 V_1^T$ and $C = V_1 V_1^T$, where V_1 corresponds to the singular values of $D - E + \alpha^{-1}Y$ larger than $\sqrt{2/\alpha}$. Given A and C the solution for E is obtained by shrinkage thresholding of $D - A + \alpha^{-1}Y$. The algorithm proceeds by alternating these two steps.

REFERENCES

- [1] B. Recht, M. Fazel, and P. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.
- [2] E. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis,” *Journal of the ACM*, (Submitted) 2010.
- [3] E. Elhamifar and R. Vidal, “Sparse subspace clustering,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [4] G. Liu, Z. Lin, and Y. Yu, “Robust subspace segmentation by low-rank representation,” in *International Conference on Machine Learning*, 2010.

Multiscale Geometric Dictionaries for Point-cloud Data

Guangliang Chen¹, Mauro Maggioni^{1,2}

¹Mathematics and ²Compute Science Departments, Duke University, PO Box 90320, Durham, NC 27708, USA
Emails: {mauro, glchen}@math.duke.edu

Abstract—We develop a novel geometric multiresolution analysis for analyzing intrinsically low-dimensional point clouds in high-dimensional spaces, modeled as samples from a d -dimensional set \mathcal{M} (in particular, a manifold) embedded in \mathbb{R}^D , in the regime $d \ll D$. This type of situation has been recognized as important in various applications, such as the analysis of sounds, images, and gene arrays. In this paper we construct data-dependent multiscale dictionaries that aim at efficient encoding and manipulating of the data. Unlike existing constructions, our construction is fast, and so are the algorithms that map data points to dictionary coefficients and vice versa. In addition, data points have a guaranteed sparsity in terms of the dictionary.

I. INTRODUCTION

Data sets are often modeled as point clouds in \mathbb{R}^D , for D large, but having some interesting low-dimensional structure, for example that of a d -dimensional manifold \mathcal{M} , with $d \ll D$. When \mathcal{M} is simply a linear subspace, one may exploit this assumption for encoding efficiently the data by projecting onto a dictionary of d vectors in \mathbb{R}^D (found by SVD), at a cost $(n + D)d$ for n data points. When \mathcal{M} is nonlinear, there are no “explicit” constructions of dictionaries that achieve a similar efficiency: typically one uses either random dictionaries, or dictionaries obtained by black-box optimization. Such constructions (e.g. [1], [3], [4]), which typically cast the sparsity requirement as an optimization problem, suffer from many local minima and lack of theoretical guarantees. In this paper we construct data-dependent dictionaries based on a *geometric multiresolution analysis (GMRA)* of the data, inspired by multiscale techniques in geometric measure theory, to remedy the above deficiencies.

II. GEOMETRIC WAVELETS

Assume we have n samples drawn i.i.d. from a d -dimensional compact Riemannian manifold $\mathcal{M} \subset \mathbb{R}^D$ according to the natural volume measure $d\text{vol}$ on \mathcal{M} . We use such training data to present how to construct geometric wavelets, though our construction easily extends to any point-cloud data, by using locally adaptive dimensions.

Multiscale decomposition. We start by constructing a multiscale nested partition of \mathcal{M} into dyadic cells $\{C_{j,k}\}_{k \in \Gamma_j, 0 \leq j \leq J}$ in \mathbb{R}^D . There is a natural tree \mathcal{T} associated to the family: For any $j \in \mathbb{Z}$ and $k \in \Gamma_j$, we let $\text{children}(j, k) = \{k' \in \Gamma_{j+1} : C_{j+1, k'} \subseteq C_{j,k}\}$.

Multiscale SVD. For every $C_{j,k}$ we define the mean (in \mathbb{R}^D) by $\bar{c}_{j,k} := \mathbb{E}[x|x \in C_{j,k}]$ and the covariance by $\text{cov}_{j,k} = \mathbb{E}[(x - \bar{c}_{j,k})(x - \bar{c}_{j,k})^*|x \in C_{j,k}]$. Let the rank- d SVD of $\text{cov}_{j,k}$ be $\text{cov}_{j,k} = \Phi_{j,k} \Sigma_{j,k} \Phi_{j,k}^*$. The subspace spanned by the columns of $\Phi_{j,k}$, and then translated to pass through $\bar{c}_{j,k}$, $\langle \Phi_{j,k}, \cdot \rangle + \bar{c}_{j,k}$, is an approximate tangent space to \mathcal{M} at location $\bar{c}_{j,k}$ and scale 2^{-j} . We define the coarse approximations, at scale j , to the manifold \mathcal{M} and to any point $x \in \mathcal{M}$, as follows:

$$\mathcal{M}_j := \cup_{k \in \Gamma_j} P_{j,k}(C_{j,k}), \quad x_j := P_{j,k}(x), \quad x \in C_{j,k}, \quad (1)$$

where $P_{j,k}$ is the associated affine projection to $C_{j,k}$.

Multiscale geometric wavelets. We can then introduce our wavelet encoding of the difference between \mathcal{M}_j and \mathcal{M}_{j+1} , for $j < J$.

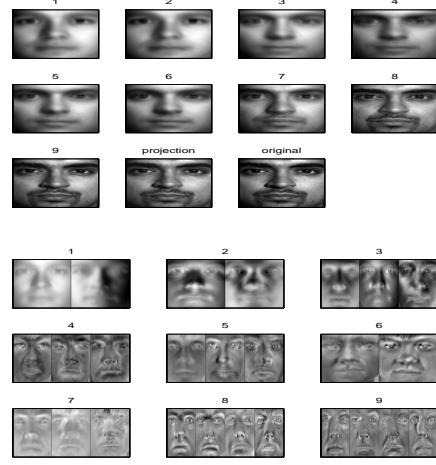


Fig. 1. We apply the GMRA to 2414 (cropped) face images from 38 human subjects in fixed frontal pose under varying illumination angles. This figure shows the multiscale approximations, from coarse to fine, of a data point (top), and the corresponding subset of dictionary elements, arranged in a multiscale fashion (bottom).

These operators are low-dimensional “detail” operators analogous to the wavelet projections in wavelet theory, and satisfy, by construction,

$$P_{\mathcal{M}_{j+1}}(x) = P_{\mathcal{M}_j}(x) + Q_{\mathcal{M}_{j+1}}(x), \quad \forall x \in \mathcal{M}. \quad (2)$$

Geometric Wavelet Transforms (GWT). Given a GMRA structure, we may compute a discrete Forward GWT for a point $x \in \mathcal{M}$ that maps it to a sequence of wavelet coefficient vectors:

$$q_x = (q_{J,x}, q_{J-1,x}, \dots, q_{1,x}, q_{0,x}) \quad (3)$$

where $q_{j,x} := \Psi_{j,x}^*(x_j - c_{j,x})$. Note that, for a fixed precision $\epsilon > 0$, q_x has a maximum possible length $(1 + \frac{1}{2} \log_2 \frac{1}{\epsilon})d$, which is independent of D and nearly optimal in d [2].

Sparsity The geometric wavelet dictionary may be constructed efficiently and is associated with efficient direct and inverse transforms. Depending on the geometric regularity of the data, it provides sparse (compressible) representations for data points.

ACKNOWLEDGMENT

We thank NSF (DMS 0650413, CCF 0808847, IIS 0803293), ONR N00014-07-1-0625, and the Sloan Foundation for partial support.

REFERENCES

- [1] M. Aharon, M. Elad, and A. M. Bruckstein. The K-SVD algorithm. In *SPARSE’05*.
- [2] W. K. Allard, G. Chen, and M. Maggioni. Multiscale geometric methods for data sets II: Geometric wavelets. Submitted.
- [3] A. Szlam and G. Sapiro. Discriminative k-metrics. In *ICML ’09*.
- [4] M. Zhou, H. Chen, J. Paisley, L. Ren, G. Sapiro, and L. Carin. Nonparametric Bayesian dictionary learning for sparse image representations. In *NIPS ’09*.

Modeling Statistical Dependencies in Sparse Representations

Tomer Faktor

The Electrical-Engineering Department
The Technion, Haifa 32000, Israel
Email: tomerfa@tx.technion.ac.il

Yonina C. Eldar

The Electrical-Engineering Department
The Technion, Haifa 32000, Israel
Email: yonina@ee.technion.ac.il

Michael Elad

The Computer-Science Department
The Technion, Haifa 32000, Israel
Email: elad@cs.technion.ac.il

Abstract—In this work we consider a Bayesian setting for sparse representation modeling and use the Boltzmann Machine (BM) to statistically model dependencies in the representation vector. We show that the exact MAP/MMSE estimation for the general case is computationally complex, and we propose a greedy approximation for both. Considering the special case where the dictionary is unitary, we derive a message-passing algorithm that leads to an exact MAP estimation. We further show that the MPM estimation improves over the MAP one for this case. Finally, when the model parameters defining the underlying graph are unknown, we suggest an algorithm that learns these parameters directly from the data using a Maximum-Pseudo-Likelihood approach.

I. BACKGROUND

The classic assumption of independence between the dictionary atoms in sparse representation modeling is often wrong. In this work we introduce such a structure to the representation vector in a flexible and adaptive manner. We consider a signal y that is built as $y = Ax + e$, where A is a dictionary of size n -by- m , x is a sparse representation vector and e is additive white Gaussian noise with variance σ_e^2 . We denote the sparsity pattern of x by $S \in \{-1, 1\}^m$ ($S_i = 1$ implies $x_i \neq 0$). We assume a Gaussian distribution with zero mean and variance $\sigma_{x,i}^2$ for each nonzero representation coefficient x_i . The core of our model lies in the prior distribution we place on the sparsity pattern. We follow the suggestion of [1], [2] and assume that the sparsity pattern is modeled by a Boltzmann machine (BM),

$$\Pr(S) = \frac{1}{Z} \exp \left(b^T S + \frac{1}{2} S^T W S \right). \quad (1)$$

This is a convenient graphical model for describing statistical dependencies between a set of binary random variables. The BM distribution can be easily represented by an MRF - a bias b_i is associated with a node i and a nonzero entry W_{ij} in the interaction matrix results in an edge connecting nodes i and j with the specified weight. The BM can serve as a powerful prior on the support in the signal model, as it can achieve sparsity and at the same time capture statistical dependencies and independencies in the sparsity pattern. Using the BM as a model for the support, several questions naturally arise: how to perform pursuit for finding the sparse representation, and how to find the model parameters W, b . In our work we address these two questions, as described below¹.

II. OUR WORK CONTRIBUTION

We adopt a Bayesian point of view, aiming to recover x from y . For general dependency models, we show that exact MAP and MMSE estimation of the sparse representation becomes computationally complex. To simplify the computations, we suggest using a greedy approach, which approximates these estimators and is suitable for any set of model parameters. For approximate MAP estimation we suggest an OMP-like algorithm, that starts with an empty support, and adds one entry at a time by greedily maximizing the posterior $\Pr(S|y)$. Once the support is found, an oracle formula is used to

estimate the non-zero entries in the representation. We also consider a thresholding-like version of this algorithm.

For the MMSE approximation we imitate the Random-OMP algorithm [3]: Instead of adding to the support the element that maximizes the posterior in each iteration, we make a random choice with probabilities proportional to this posterior's marginals. Second, we perform several runs of this algorithm and average the resulting sparse representations to obtain the final estimate for x .

When the dictionary is square and unitary, the BM distribution is a conjugate prior of the estimation problem. Based on this observation, we get that finding the MAP estimator for S becomes an inference task on a modified BM. We propose to handle this NP-hard computational task by assuming a banded interaction matrix W , and using an efficient message passing algorithm for obtaining the exact MAP estimate. Furthermore, we explore the MSE for various estimators and develop an expression for the error of a Bayesian estimator, based on the single-node marginals of the posterior. We observe that the Maximum Posterior Marginal (MPM) estimator for the support leads to optimal performance (in terms of MSE) among all estimators that are based on a single support. Finally, we develop an expression for the MMSE estimator and show how it can be evaluated and the gap between this and the MPM.

Finally, we study the problem of learning the Boltzmann parameters of the underlying graph from a set of supports. We develop an efficient algorithm which is based on a maximum pseudo-likelihood (MPL) approach and uses the sequential subspace optimization (SESOP) method for solving it [4]. We demonstrate the effectiveness of our proposed adaptive BM-based approach, by performing denoising experiments on image patches using a fixed DCT dictionary and learned BM parameters. The results show an improvement of $\sim 1[dB]$ over plain OMP denoising on these patches.

REFERENCES

- [1] P.J. Garrigues and B.A. Olshausen, Learning horizontal connections in a sparse coding model of natural images, in *Advances in Neural Information Processing Systems* 20, pp. 505-512. 2008.
- [2] V. Cevher, M.F. Duarte, C. Hedge, and R.G. Baraniuk, Sparse signal recovery using Markov random fields, in *Advances in Neural Information Processing Systems* 21, pp. 257-264. 2009.
- [3] M. Elad and I. Yavneh, A plurality of sparse representations is better than the sparsest one alone, *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4701-4714, Oct. 2009.
- [4] G. Narkiss and M. Zibulevsky, Sequential subspace optimization method for large-scale unconstrained optimization, Tech. Rep., EE Dept., Technion Israel Institute of Technology, Haifa, Israel, 2005.
- [5] T. Faktor, Y.C. Eldar, and M. Elad, Exploiting statistical dependencies in sparse representations for signal recovery, Submitted to *IEEE Trans. on Signal Processing*.
- [6] T. Faktor, Y.C. Eldar, and M. Elad, Denoising of image patches via sparse representations with learned statistical dependencies, ICASSP, Prague, Czech Republic, May 2011.

¹Parts of this work are reported in [5], [6] and others are new.

A source localization approach based on structured sparsity for broadband far-field sources

Aris Gretzistas and Mark D. Plumley

School of Electronic Engineering and

Computer Science

Queen Mary University of London

Mile End Road, London, E1 4NS, UK

Email: {name}.{surname}@eecs.qmul.ac.uk

Abstract—In this work, we present a source localization method for broadband sources impinging from the far-field on a uniform linear array (ULA) of sensors, based on a group sparsity structure over the different frequency bands and the joint sparsity structure over several time snapshots.

I. SPARSITY BASED NARROWBAND SOURCE LOCALIZATION

Source localization has been an active research field, playing fundamental role in many signal processing areas such as radar, sonar, seismology and acoustic tracking. The problem may also be referred to as Direction-of-arrival (DOA) estimation, when arrays of sensors are employed, and its main objective is to give an estimate of the spatial energy spectrum and therefore determine the number and location of the sources of energy corresponding to the peaks of the spatial spectrum.

The emerging field of sparse representations and compressed sensing (CS) has given renewed interest to the problem of source localization. The concept of spatial sparsity was first introduced in [1], where it was shown that the source localization problem can be cast as a sparse representations problem in a redundant dictionary and the ℓ_1 -SVD method was proposed for the recovery of the inverse system.

In [2], the dictionary is generated after the discretization of the angular space, containing the impulse responses for each potential direction of arrival. The dictionary is overcomplete meaning that the number of sensors is smaller than the number of possible locations (DOAs). Therefore, considering a linear array of M sensors, the array output can be expressed as:

$$\mathbf{y}(t) = \Phi \mathbf{s}(t) + \mathbf{n}(t) \quad (1)$$

where $\mathbf{s}(t)$ is a k -sparse vector of length N , $\mathbf{y}(t)$ is a $M \times 1$ measurements vector of the array of sensors and $\mathbf{n}(t)$ is the additive noise vector. Φ is a redundant dictionary as $M < N$, which contains the array responses of the potential angles of arrival:

$$\Phi = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_N)]. \quad (2)$$

The sparse solution to the above underdetermined system of linear equations can be approximated using convex optimization (e.g. Lasso or BPDN) or greedy approaches (e.g. OMP, CoSaMP, IHT etc.).

However, this single snapshot approach requires that the inverse problem should be solved at each time index separately. By taking multiple snapshots and assuming that the sources are not moving, as described in [2], we can formulate the source localization problem as a multiple measurement vector (MMV) joint sparse recovery problem, exploiting the fact that the sources will share a common support over all time snapshots:

$$\mathbf{Y} = \Phi \mathbf{S} + \mathbf{N} \quad (3)$$

where \mathbf{Y} and \mathbf{N} are $M \times L$ matrices and \mathbf{S} is a $N \times L$ matrix with L equals the number of time snapshots. Therefore, by enforcing only spatial sparsity and not temporal one can use convex optimization (e.g. ℓ_1/ℓ_2 minimization) or greedy methods (e.g. SOMP) to recover the joint sparse entries.

II. PROPOSED APPROACH FOR BROADBAND SOURCES

The described source localization methods of the previous section assume that the impinging sources on the array of sensors are narrowband centered at a carrier frequency. In the case of broadband sources, we first need to transform the data into the time-frequency domain and assume that each frequency bin contains the energy of k narrowband sources. Also, in the broadband scenario the manifold matrices (or redundant dictionaries) are different for different frequency bands due to the fact that the central carrier frequency varies. Subsequently, a naive approach could be to use the joint sparsity model for several time snapshots at each frequency bin ω_q :

$$\mathbf{Y}_q = \Phi_q \mathbf{S}_q + \mathbf{N}_q. \quad (4)$$

However, assuming that the sources are stationary, we expect that the sparse support will be common for all different frequency bins. Therefore, the proposed approach by interleaving the values of the solution vector appropriately, exploits the additional structure of grouped sparsity over all frequency bins. This proves to be quite beneficial as the proposed approach suppresses the undesirable effects of spatial aliasing.

More specifically, experiments showed that by choosing the spacing of the linear array to be much larger than half the wavelength of the highest frequency, the algorithm can still achieve super-resolution recovering the correct support, corresponding to the unknown DOAs, over all frequencies. Therefore, for the specific setting the algorithm outperforms the single frequency bin spatial sparsity approach as well as other conventional array processing methods such as MUSIC and Capon's beamformer.

ACKNOWLEDGMENT

This research is supported by ESPRC Leadership Fellowship EP/G007144/1 and EU FET-Open Project FP7-ICT-225913 “SMALL”.

REFERENCES

- [1] D. Malioutov, M. Cetin, and A.S. Willsky, *A sparse signal reconstruction perspective for source localization with sensor arrays*, Signal Processing, IEEE Transactions on, vol. 53, no. 8, pp. 3010 - 3022, aug. 2005.
- [2] A. Gretzistas and M.D. Plumley, *A multichannel spatial compressed sensing approach for direction of arrival estimation*, Proceedings of the 9th international conference on Latent variable analysis and signal separation, pp. 458 - 465, 2010.

Sparsity with sign-coherent groups of variables via the cooperative-Lasso

Julien Chiquet

and Camille Charbonnier

CNRS UMR 8071 & Université d'Évry
Laboratoire Statistique et Génome
Email: julien.chiquet@genopole.cnrs.fr

Yves Grandvalet

Université de Technologie de Compiègne
CNRS UMR 6599 Heudiasyc

I. INTRODUCTION

Penalization methods that build on the ℓ_1 -norm, referred to as Lasso procedures, are now widely used to tackle simultaneously signal estimation and feature selection in high-dimensional problems. In this framework, providing the signal with an *a priori* meaningful group structure beyond mere signal sparsity is an efficient way to improve performances.

Here we propose a new regularization term which builds upon the assumption that groups are *sign-coherent*, namely that coefficients within a group are either all null, non-negative or non-positive. Thanks to this apparently strong assumption on the signal structure we gain in flexibility in the inference of sparsity patterns. As a result our method acquires robustness to possible miss-specifications in the group structure compared to the group-Lasso [1], without paying the price of an additional penalty term as for the sparse group-Lasso [2].

This framework is large enough to cover settings where groups are derived *a priori* from the experimental design or defined empirically from the correlation structure among covariates. In both cases, the aim is to gather redundant or consonant variables in the same group, so as to force them to collaborate instead of conflicting with each other. Multitask datasets, where observations are split into tasks with distinct but supposedly close underlying signals, particularly fit this framework (see [3] in the context of Gaussian Graphical model inference). Note that flexibility in the sparsity pattern is essential in this multitask setting, allowing some tasks not to follow the overall pattern punctually.

II. COOPERATIVE-LASSO

Let Y be the response random variable from the exponential family that we want to predict from a size- p random vector $X = (X^1, \dots, X^p)$, assuming the existence of a function g and a parameter β^* such that $\mathbb{E}(Y|X) = g(X\beta^*)$. We assume that covariates $X = (X^1, \dots, X^p)$ are partitioned into K groups \mathcal{G}_k , $k = 1, \dots, K$ such that the true support of β^* , namely the set of non zero coefficients $\mathcal{S} = \{j \in \{1, \dots, p\}, \beta_j^* \neq 0\}$, is linked to this group structure. We observe a sample of size n which we represent by a size n vector \mathbf{y} and a size $n \times p$ matrix \mathbf{X} .

For any vector v , let \mathbf{v}^+ and \mathbf{v}^- be the componentwise positive and negative parts of \mathbf{v} . We call *coop-norm* of \mathbf{v} the sum of group-norms on \mathbf{v}^+ and \mathbf{v}^- .

$$\|\mathbf{v}\|_{\text{coop}} = \sum_{k=1}^K (\|\mathbf{v}_{\mathcal{G}_k}^+\|_2 + \|\mathbf{v}_{\mathcal{G}_k}^-\|_2)$$

Denote by $\ell(\mathbf{y}, \mathbf{X}; \beta)$ the negative log-likelihood. The coop-Lasso estimate of β^* is

$$\hat{\beta}^{\text{coop}} = \arg \min_{\beta \in \mathbb{R}^p} \ell(\mathbf{y}, \mathbf{X}; \beta) + \lambda \|\beta\|_{\text{coop}},$$

where $\lambda \geq 0$ is a tuning parameter common to all groups.

To solve this problem, we develop an algorithm built upon the subdifferential calculus approach originally proposed by Osborne, Presnell and Turlach [4] for the Lasso and adapted to the group-Lasso by Roth and Fischer [5].

We prove estimation and selection consistency of this estimator in the context of linear regression under variants of the so-called irrepresentable condition. In order to provide tools for the choice of λ , we also derive AIC and BIC criteria thanks to an estimation of its effective degrees of freedom in the linear regression setup.

III. ROBUST MICROARRAY GENE SELECTION

We analyse the dataset proposed by Hess *et al.* [6] providing gene expression profiles of patients treated with chemotherapy prior to surgery, classified as presenting either a pathologic complete response (pCR) or a residual disease (not-pCR). The objective of this dataset was to extract a small set of genes that best predict the response to preoperative chemotherapy and be able to prevent patients that would probably not benefit from chemotherapy from undergoing such a harsh treatment.

Each gene is covered by a set of sibling probes measuring the expression of different but possibly overlapping parts of the gene that, except in cases of alternative splicing, should provide redundant signals. The usual processing of this type of data is based upon individual probe measurements, roughly assimilated to genes in the final interpretation step. Here we take the gene level into account right from the statistical analysis, requiring sign-coherent effects of probes related to same genes. Requesting an overall consensus on the sign of effects at the gene level supports biological coherence while allowing for potential null effects within a group leaves room for potential alternative splicing divergences.

REFERENCES

- [1] Yuan, M. and Lin, Y., Model Selection and Estimation in Regression with Grouped Variables, *Journal of the Royal Statistical Society, Series B*, vol. 68, 2006.
- [2] Friedman, J. and Hastie, T. and Tibshirani, R., A note on the group Lasso and a sparse group Lasso, *arxiv preprint*, 2010.
- [3] Chiquet, J. and Grandvalet, Y. and Ambroise, C., Inferring multiple graphical structures, *Statistic and Computing*, 2011.
- [4] Osborne, M. R. and Presnell, B. and Turlach, B. A., On the LASSO and its dual, *Journal of Computational and Graphical Statistics*, vol.9, 2000.
- [5] Roth, V. and Fischer, B., The Group-Lasso for Generalized Linear Models: Uniqueness of Solutions and Efficient Algorithms, *ICML '08: Proceedings of the 25th international conference on Machine Learning*, 2008.
- [6] Hess *et al*, Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with Paclitaxel and Fluorouracil, Doxorubicin, and Cyclophosphamide in breast cancer, *Journal of Clinical Oncology*, vol. 24, 2006.

Tail bounds for all eigenvalues of a sum of random matrices

Alex Gittens

Applied and Computational Mathematics
California Institute of Technology
Pasadena, California
Email: gittens@caltech.edu

Joel Tropp

Applied and Computational Mathematics
California Institute of Technology
Pasadena, California
Email: jtropp@acm.caltech.edu

The general-purpose tools developed for quantifying the spectra of random matrices: the non-commutative Khintchine inequality [1], a lemma due to Rudelson [2], ϵ -net arguments [1], and the Laplace transform approach due to Ahlswede and Winter [3], give information on only the extreme singular values.

We introduce a simple technique, based upon the variational characterization of the eigenvalues of self-adjoint matrices and the Laplace transform machinery, for bounding *all* eigenvalues. To demonstrate the power of the variational Laplace transform, we investigate the spectra of matrices formed by sampling columns from a matrix with orthogonal rows and bound the number of samples needed to estimate eigenvalues of the covariance matrix of a Gaussian vector to within relative precision.

I. THE VARIATIONAL LAPLACE TRANSFORM

Consider $\mathbf{X} = \sum_j \mathbf{X}_j$, a sum of independent, random, self-adjoint matrices. Let

$$\mathbb{V}_d^n = \{\mathbf{V} \in \mathbb{C}^{n \times d} : \mathbf{V}^* \mathbf{V} = \mathbf{I}\}$$

denote the collection of orthonormal bases for the d -dimensional subspaces of \mathbb{C}^n .

Using the Courant-Fischer theorem to modify the Laplace transform machinery in [4], we find that

$$\begin{aligned} \mathbb{P}\{\lambda_k(\mathbf{X}) \geq t\} &\leq \inf_{\theta > 0} \min_{\mathbf{V} \in \mathbb{V}_{n-k+1}^n} \left\{ e^{-\theta t} \cdot \mathbb{E} \operatorname{tr} e^{\theta \mathbf{V}^* \mathbf{X} \mathbf{V}} \right\} \\ &\leq \inf_{\theta > 0} \min_{\mathbf{V} \in \mathbb{V}_{n-k+1}^n} \left[e^{-\theta t} \cdot \operatorname{tr} \exp \left\{ \sum_j \theta \mathbf{V}^* \mathbf{A}_j \mathbf{V} \right\} \right], \end{aligned} \quad (1)$$

where the *deterministic* matrices \mathbf{A}_j are chosen to satisfy the relation $\mathbb{E} e^{\mathbf{X}_j} \leq e^{\mathbf{A}_j}$.

II. CHERNOFF AND BENNETT INEQUALITIES

The variational Laplace transform (1) is applied by constructing appropriate \mathbf{A}_j and using a convenient choice of \mathbf{V} .

If the matrices \mathbf{X}_j are all positive semidefinite, let \mathbf{V}_+ in \mathbb{V}_{n-k+1}^n satisfy

$$\lambda_k(\mathbb{E} \mathbf{X}) = \lambda_{\max}(\mathbf{V}_+^* (\mathbb{E} \mathbf{X}) \mathbf{V}_+).$$

Then the following Chernoff-type bound holds:

$$\mathbb{P}\{\lambda_k(\mathbf{X}) \geq (1 + \delta)\lambda_k(\mathbb{E} \mathbf{X})\} \leq (n - k + 1) \cdot \left[\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right]^{\lambda_k(\mathbb{E} \mathbf{X}) / R(\mathbf{V}_+)}$$

for $\delta > 0$. Here, $R(\mathbf{V}_+)$ quantifies the concentration of the summands in the invariant subspace determined by \mathbf{V}_+ .

If the summands \mathbf{X}_j are self-adjoint and we are given the variance

$$\sigma_k^2 = \lambda_k \left(\sum_j \mathbb{E} \mathbf{X}_j^2 \right),$$

then the following Bernstein-type inequality holds:

$$\mathbb{P}\{\lambda_k(\mathbf{X}_j) \geq t\} \leq (n - k + 1) \cdot \exp \left\{ \frac{-t^2/2}{\sigma_k^2 + R(\mathbf{V}_+)t/3} \right\}.$$

Chernoff and Bernstein bounds on the lower tails of λ_k are also derived.

III. COLUMN SUBSAMPLING OF MATRICES WITH ORTHONORMAL ROWS

Let \mathbf{U} be a matrix with orthonormal rows. Sample from its columns by right multiplication with a diagonal matrix whose entries are independent $\text{Bern}(p)$ random variables, to form $\widehat{\mathbf{U}} = \mathbf{U}\mathbf{D}$. How does the spectrum of $\widehat{\mathbf{U}}$ behave? We apply our Chernoff bounds to estimate the probability that the k th singular value of $\widehat{\mathbf{U}}$ deviates either above or below \sqrt{p} , and find that the probability of deviation is controlled by a coherence-like quantity τ_k satisfying

$$\tau_k \leq \min_{|I| \leq k} \max_j \sum_{i \in I} u_{ij}^2.$$

IV. COVARIANCE ESTIMATION TO RELATIVE PRECISION

Draw i.i.d samples $\{\boldsymbol{\eta}_j\}_{j=1}^n \subset \mathbb{R}^p$ from a $\mathcal{N}(\mathbf{0}, \mathbf{C})$ distribution. The classical covariance estimation problem (how many samples are needed to ensure that the empirical covariance estimator has a fixed relative accuracy in the spectral norm?) has been studied extensively, and it is known that, for many distributions of interest, $O(p)$ samples suffice [5].

We investigate for the first time, using our Bernstein bounds, the question of how many samples are needed to ensure that individual eigenvalues are estimated to relative precision. We find that $O(k \log k)$ and $O((p - k + 1) \log(p - k + 1))$ samples are needed, respectively, to ensure that with high probability the k th eigenvalue is not underestimated or overestimated. Although we prove this result only for Gaussian vectors, the argument can be extended to other distributions.

REFERENCES

- [1] R. Vershynin, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2011, ch. Introduction to the non-asymptotic analysis of random matrices. [Online]. Available: <http://www-personal.umich.edu/~romanv/papers/non-asymptotic-rmt-plain.pdf>
- [2] M. Rudelson, “Random Vectors in the Isotropic Position,” *J. Funct. Anal.*, vol. 164, no. 1, pp. 60–72, 1999.
- [3] R. Ahlswede and A. Winter, “Strong converse for identification via quantum channels,” *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 569–579, 2002.
- [4] J. Tropp, “User-Friendly Tail Bounds for Sums of Random Matrices,” 2011, preprint. [Online]. Available: <http://arxiv.org/abs/1004.4389>
- [5] R. Adamczak, A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann, “Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles,” *J. Amer. Math. Soc.*, vol. 23, no. 2, pp. 535–561, 2010.

Random Projections are Nearly Isometric For Parametric Functions Too

William Mantzel and Justin Romberg
[willem, jrom]@gatech.edu
Georgia Institute of Technology

There has been substantial interest in dimension-reducing random projections over the last decade as a way to preserve enough structure to solve a regularized inverse problem (1) or exploit this structure for fast matching (3) or to quickly characterize a linear system by its approximate dimension-reduced counterpart.

This body of research tells us with overwhelming probability that a random projection of any finite union of subspaces is roughly isometric, a key property used in establishing recovery bounds.

Perhaps not surprisingly, there is also a wide variety of classes of parametric functions in L_2 for which these properties also apply. For this work, we explore the class of parametric functions $\mathcal{F} = \{f_\tau : \tau \in S\}$ with parameter $\tau \in S \subset \mathbb{R}^P$ for some bounded set S . We further impose some Lipschitz-like property on \mathcal{F} : $\|f_{\tau_1} - f_{\tau_2}\|_{L_2} \leq \|\tau_1 - \tau_2\|_2$ though this work may be trivially extended to a union of sets S_k that each obey $\|f_{\tau_1} - f_{\tau_2}\|_{L_2} \leq B_k \|\tau_1 - \tau_2\|_2$ for a different Lipschitz constant B_k for each set S_k .

Consider the problem of parametric estimation:

$$\bar{f} = \arg \min_{f \in \mathcal{F}} \|f - h\|$$

The randomly projected version is then:

$$\hat{f} = \arg \min_{f \in \mathcal{F}} \|\Phi(f - h)\| \quad (1)$$

for some random operator $\Phi : L_2(\mathbb{R}^D) \rightarrow \mathbb{R}^M$.

For the sake of this work, the operator is defined as:

$$[\Phi f]_m = \langle G_m, \Psi^T f \rangle$$

where $G_m[n] \sim N(0, 1/M)$ is a sequence of i.i.d. Gaussian random variables and $\Psi^T : L_2(\mathbb{R}^D) \rightarrow \ell_2$ is some arbitrary orthonormal transform (e.g. a wavelet analysis operator) that need not be a sparsifying transform. Note that $E[\|\Phi f\|^2] = \|f\|^2$.

The main result is that the \hat{f} that minimizes the compressed formulation will be characteristically similar to the deterministic minimizer \bar{f} in the following sense:

$$\frac{1}{r} \|\hat{f} - h\| - C_P \epsilon \sqrt{2 \log r} \leq \|\Phi(\hat{f} - h)\| \leq \|\Phi(\bar{f} - h)\| \leq (1 + \delta) \|\bar{f} - h\|, \quad (2)$$

with probability at least:

$$1 - e^{-M c_0(\delta)} - 2MN(S, \epsilon)r^{-M}, \quad (3)$$

for some constant C_P (e.g. $C_P = 4\sqrt{P}$), for any chosen $r > 1$, and where $N(S, \epsilon)$ is the number of points τ_n needed to cover S with radius at most ϵ (something like $|S| \epsilon^{-P}$).

This approach scales well as the modeling error decreases. In particular, when $h \in \mathcal{F}$, we have $\|\hat{f} - h\| \leq 2C_P r^{3/2} \epsilon$ for all positive upper bounds with arbitrarily high probability by choosing $\epsilon = r^{-2} \rightarrow 0$ (i.e. $\hat{f} = h$ in probability).

The second and third inequalities of Eq. 2 follow from the characterization of the minimizing \hat{f} and the Chernoff bound, respectively.

A brief sketch of the first inequality follows. For any such τ_n in the $N(S, \epsilon)$ covering, we have:

$$P \left\{ \sup_{\tau \in B(\tau_n, \epsilon)} \|\Phi(f_\tau - f_{\tau_n})\| > C_P \epsilon u \right\} \leq M \exp(-Mu^2/2),$$

which we state without proof. After utilizing $u = \sqrt{2 \log r}$, we then have for all $\tau \in B(\tau_n, \epsilon)$:

$$\|\Phi(f_\tau - h)\| \geq \|\Phi(f_{\tau_n} - h)\| - \|\Phi(f_\tau - f_{\tau_n})\| \geq \frac{1}{r} \|\hat{f} - h\| - C_P \epsilon \sqrt{2 \log r} \quad (4)$$

with probability at least:

$$1 - r^{-M} - Mr^{-M}$$

where the first term comes from a small lower bound on a chi-square random variable:

$$P \left\{ \|\Phi f\|^2 < \frac{1}{\lambda^2} \|f\|^2 \right\} \leq \lambda^{-M}.$$

Because Eq. 4 is true for all $f_\tau \in \mathcal{F}$ with probability at least

$$1 - 2MN(S, \epsilon)r^{-M},$$

it must also be true for \hat{f} as desired.

A straightforward application is the matched filtering problem, also considered in work by Eftekhari et al. (2). We take the class $\mathcal{F} = \{f_\tau = f_0(t - \tau) : \tau \in S\}$ of bounded shifts of some Lipschitz-continuous unit-norm base function f_0 where we again impose the condition that $\|f_{\tau_1} - f_{\tau_2}\|_{L_2} \leq \|\tau_1 - \tau_2\|_2$. Here, $P = D$ and we will require $M = O(P + \log(|S|))$ measurements for accurate recovery.

This work could be viewed as a “1-sparse” solution to the continuous case of compressed sensing on parametric functions and opens up many questions about the viability of the applicability of compressed sensing to functions that are some finite weighted sum of some infinite dictionary of parameterized basis functions.

REFERENCES

- [1] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.
- [2] A. Eftekhari, J. Romberg, and M. Wakin. Matched Filtering from Limited Frequency Samples. *Arxiv preprint arXiv:1101.2713*, 2011.
- [3] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. 2004.

Concentration Inequalities and Isometry Properties for Compressive Block Diagonal Matrices

Han Lun Yap,^g Jae Young Park,^m Armin Eftekhari,^c Christopher J. Rozell,^g and Michael B. Wakin^c

^g School of Electrical and Computer Engineering, Georgia Institute of Technology

^m Department of Electrical Engineering and Computer Science, University of Michigan

^c Division of Engineering, Colorado School of Mines

Abstract—In this talk, we survey our recent analysis of randomized, compressive block diagonal matrices. We present concentration of measure bounds which indicate that (unlike dense i.i.d. random matrices) the probability of norm preservation actually depends on the signal being measured. We discuss implications of this fact in various compressive signal processing applications. We also present an RIP bound for block diagonal matrices and explain that in the best case—for signals that are sparse in the frequency domain—these matrices perform nearly as well as dense i.i.d. random matrices despite having many fewer nonzero entries.

EXTENDED ABSTRACT

The analysis of randomized compressive linear operators often relies on quantifying the likelihood that a random matrix will preserve the norm of a signal after multiplication. For example, a standard *concentration of measure bound* [1] states that for a fixed signal $x \in \mathbb{R}^N$, if Φ is an $M \times N$ matrix populated with independent and identically distributed (i.i.d.) random entries drawn from a suitable distribution, the probability that $|\|\Phi x\|_2^2 - \|x\|_2^2|$ will exceed a small fraction of $\|x\|_2^2$ decays exponentially in the number of measurements M . From this one can also prove that if $M = O(K \log(N/K))$, then with high probability the *Restricted Isometry Property* (RIP) will hold, ensuring that $\|\Phi x\|_2^2 \approx \|x\|_2^2$ uniformly across all K -sparse signals x . Such results have immediate applications in proving the Johnson-Lindenstrauss (JL) lemma, establishing signal recovery bounds in Compressive Sensing (CS), etc.

Unfortunately, dense random matrices with i.i.d. entries are often either impractical because of the resources required to store and work with a large unstructured matrix, or unrealistic as models of acquisition devices with architectural constraints preventing global data aggregation. In this talk, we will survey our recent analysis [2, 4] of randomized, compressive block diagonal matrices. We model a signal $x \in \mathbb{R}^{NJ}$ as being partitioned into J blocks $x_1, x_2, \dots, x_J \in \mathbb{R}^N$, and for each $j \in \{1, 2, \dots, J\}$, we suppose that a local measurement operator $\Phi_j : \mathbb{R}^N \rightarrow \mathbb{R}^{M_j}$ collects the measurements $y_j = \Phi_j x_j$. Concatenating all of the measurements into a vector $y \in \mathbb{R}^{\sum_j M_j}$, we then have

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{bmatrix}}_{y: (\sum_j M_j) \times 1} = \underbrace{\begin{bmatrix} \Phi_1 & & & \\ & \Phi_2 & & \\ & & \ddots & \\ & & & \Phi_J \end{bmatrix}}_{\Phi: (\sum_j M_j) \times NJ} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_J \end{bmatrix}}_{x: NJ \times 1}. \quad (1)$$

In some scenarios, the local measurement operator Φ_j may be unique for each block, and we say that the resulting Φ has a *Distinct Block Diagonal* (DBD) structure. In other scenarios it may be appropriate or necessary to repeat a single operator across all blocks (such that $\Phi_1 = \Phi_2 = \dots = \Phi_J$); we call the resulting Φ a *Repeated Block Diagonal* (RBD) matrix.

We will present concentration of measure bounds [2] both for DBD matrices populated with i.i.d. subgaussian random variables and for RBD matrices populated with i.i.d. Gaussian random variables. Our main results essentially state that the probability of concentration depends on the “diversity” of the component signals x_1, x_2, \dots, x_J being well-matched to the measurement matrix, where this notion of signal diversity depends on whether the matrix is DBD or RBD. Such nonuniform concentration behavior is markedly unlike that of i.i.d. dense matrices, for which concentration probabilities are signal agnostic. For the most favorable classes of signals, however, the concentration of measure probability for block diagonal matrices scales exactly as for an i.i.d. dense random matrix (that is, the failure probability decays exponentially in the total number of measurements). We will provide several examples of signal classes that are particularly favorable for measurement via DBD or RBD matrices; among these are signals having sparse representations in the frequency domain.

Our concentration of measure bounds have a number of immediate applications. We will present a modified version of the JL lemma appropriate for block diagonal matrices and explain how this lemma can be used to guarantee the performance of various compressive-domain signal inference and processing algorithms. We will also briefly explain how our concentration bounds for block diagonal matrices can be used as an analytical tool when studying the structured Toeplitz and observability matrices that arise in certain linear systems applications.

Unfortunately, it does not appear that one can couple our nonuniform concentration results with covering arguments to arrive at a compelling RIP bound for block diagonal matrices. Using tools from the theory of empirical processes [3], however, we have proved [4] that DBD matrices can indeed satisfy the RIP but that the requisite number of measurements depends on the *coherence of the basis* in which the signals are sparse. We will present this result and explain that for the best case signals—which again include those that are sparse in the frequency domain—these matrices perform nearly as well as dense i.i.d. random matrices despite having many fewer nonzero entries.

REFERENCES

- [1] D. Achlioptas. Database-friendly random projections. In *Proc. 20th ACM SIGMOD-SIGACT-SIGART Symp. Principles of Database Systems (PODS)*, pages 274–281, New York, NY, USA, 2001. ACM.
- [2] J. Y. Park, H. L. Yap, C. J. Rozell, and M. B. Wakin. Concentration of measure for block diagonal matrices with applications to compressive signal processing. 2010. Preprint.
- [3] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008.
- [4] H. L. Yap, A. Eftekhari, M. B. Wakin, and C. J. Rozell. The restricted isometry property for block diagonal matrices. In *Proc. Conf. Information Sciences and Systems (CISS)*, March 2011.

Sparse Anisotropic Triangulations and Image Estimation

Laurent Demaret

IBB, Helmholtz Zentrum, Munich, Germany

Email: laurent.demaret@helmholtz-muenchen.de

Abstract—In recent years several methods relying on anisotropic triangulations were proposed, allowing for extremely sparse representations of edge singularities in images. This contribution has two aims: existing techniques and theoretical results are surveyed and compared to other anisotropic methods, and a new image estimator based on continuous and piecewise linear splines over anisotropic Delaunay triangulations is introduced. We prove an abstract consistency result for this estimator as a first indication of the potential of anisotropic triangulations in the context of image denoising. Algorithmic and computational aspects will be also addressed.

I. MOTIVATION

One of the most important topics of recent research in signal processing is the design and analysis of non-linear methods to analyse and process signals with heterogeneous features.

The observation that a given signal class has a specific structure leads to the natural question : *How to model the prior knowledge available for this class ?* In spite of their diversity most approaches recently proposed to tackle this problem share a leading principle: the search for small modelling sets of functions which still contain the signals of interest.

A decisive impulse has been given by the introduction of the concept of sparsity in statistical and computational image processing [5]: the signal under consideration is assumed to be well approximated by a few coefficients of a fixed frame. When dealing with sharp edges in images, the design of modelling sets is driven by anisotropy: the representations should adapt locally to singularities along curves, as for instance curvelet representations do [1].

Anisotropic triangulations are based on very large dictionaries of geometrical atoms (see [4] for a survey on this topic). The resulting output is at the same time very sparse, adapted to the geometrical contents of the signal, and has very few oscillations.

In this contribution we propose to use anisotropic Delaunay triangulations in the context of edge-preserving image estimation.

II. PRINCIPLE

Let us introduce some useful notations. We consider a true (unknown) image $f : [0, 1]^2$ and assume noisy observations

$$f_i^\varepsilon = f_i + \sigma \varepsilon_i, i \in X \quad (1)$$

where $X = \{1, \dots, N\}^2$ is the set of indices, $(\varepsilon_i)_{i \in X}$ are iid $N(0, 1)$, σ is the noise level and $(f_i)_{i \in X}$ results from the discretization of f .

For a given set $Y \subset X$, let $\mathcal{D}(Y)$ denote its Delaunay triangulation and $S_{\mathcal{D}(Y)}$ the space of continuous linear splines on $\mathcal{D}(Y)$. For formal definitions and a discussion of uniqueness of Delaunay triangulations, we refer to our papers [3] and [4].

In this work we introduce the following abstract estimator based on anisotropic Delaunay triangulations:

$$\hat{f}_\gamma^N := \operatorname{argmin}_{u \in S_{\mathcal{D}(Y)}} \|f^\varepsilon - u\|_2 + \gamma |Y| \quad (2)$$

The first term controls fidelity to the observed data, while the second penalises the size of the triangulations used for estimating the signal.

We investigate some properties of estimators $\hat{f}_{\gamma_N}^N$ when $N \rightarrow \infty$: we prove conditions on γ_N for consistency and almost sure convergence rates of the estimation error for elementary classes of α -piecewise regular signals. The choice of γ_N depending on the noise level σ will be also discussed. The extension to other dimensions than 2, (relevant cases being dimensions 1 to 3) and for more general noise than in (1) (subgaussian) is straightforward.

III. ALGORITHM

The penalisation term in (2) is non-convex in u and in contrast to minimisation problems related to sparse representations over frames, finding a solution of (2) is equivalent to a search in a structured subset of a large dictionary. The convex relaxation machinery of the sparse framework (see [2]) is therefore difficult to apply. A heuristic to obtain an approximate solution of (2) is to use greedy algorithms. We introduce a modified version of our adaptive thinning algorithm proposed in [3] in the context of image compression. In a simplified formulation, our implementation relies on the following iteration:

$u^0 := f^\varepsilon = (f_i^\varepsilon)_{i \in X}, Y^0 = X$	
for $i=1, P$ (<i>P iterations</i>)	
$(u^{i+1}, Y^{i+1}) = \operatorname{argmin}_{\substack{u \in S_{\mathcal{D}(Y')} \\ Y' = Y^i \setminus \{x\}, x \in Y^{i+1}}} (\ f^\varepsilon - u\ _2 + J(u))$	
end	

Here, J denotes a suitable penalisation term. Compared to the algorithm in [3], this additional penalisation is required in order to prevent from solutions locally reproducing noise. This implies that J should be chosen such that the sparsity of the gradient of u for piecewise smooth signals is taken into account and is therefore not necessarily convex. We illustrate the discussion by numerical investigations, also addressing crucial issues like the choice of the number of iterations P .

REFERENCES

- [1] E.J. Candès and D.L. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise- C^2 singularities. *Comm. Pure Appl. Math.*, 57:219–266, 2004.
- [2] P.L. Combettes and V.R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4(4):1168–1200, 2005.
- [3] L. Demaret, N. Dyn, A. Iske, Image compression by linear splines over adaptive triangulations. *Signal Processing Journal* **86**(7), 1604–1616, 2006.
- [4] L. Demaret, A. Iske Anisotropic Triangulation Methods in Image Approximation in *Algorithms for Approximation*, E.H. Georgoulis, A. Iske, and J. Levesley (eds.), Springer-Verlag, Berlin, 47–68, 2010.
- [5] D. Donoho. Compressed sensing, *IEEE Transactions on Information Theory*, 52(4), 1289–1306, 2006

Compressive Sensing with Biorthogonal Wavelets via Structured Sparsity

Marco F. Duarte
 Department of Computer Science
 Duke University
 Durham, NC 27708

Richard G. Baraniuk
 Electrical and Computer Engineering
 Rice University
 Houston, TX 77005

Compressive sensing (CS) merges the operations of data acquisition and compression by measuring sparse or compressible signals via a linear dimensionality reduction and then recovering them using a sparse-approximation based algorithm. A signal is K -sparse if its coefficients in some transform contain only K nonzero values; a signal is compressible if its coefficients decay rapidly when sorted by magnitude. The standard CS theory assumes that the sparsifying transform is an orthogonal basis.

Recently, progress has been made on CS recovery using more general, non-orthogonal transform based on *frames*. A tight frame consists of an analysis frame $\bar{\Psi}$ and a synthesis (dual) frame Ψ such that $\Psi^T \bar{\Psi} = I$. A signal x is analyzed by finding its transform coefficients via $\theta = \bar{\Psi}x$ and synthesized via $x = \Psi\theta$. Currently, provable CS recovery in a frame can be accomplished when either (A1) the coherence of the frame (the maximum inner product between any two synthesis frame vectors) is low [1], or (A2) the signal has a sparse or compressible analysis coefficient vector $\theta = \Psi^T x$ [2].

An important set of CS applications revolves around image acquisition, where CS has been used to boost the resolution of digital cameras at exotic wavelengths, reduce the scan time in MRI scanners, and so on. The sparsifying transforms of choice for image compression have long been the biorthogonal wavelet bases (BWBs), which are non-redundant tight frames with the property that the roles of the analysis and synthesis frames are interchangeable (i.e., $\bar{\Psi}^T \Psi = \Psi^T \bar{\Psi} = I$). In contrast to orthogonal wavelet bases (OWBs), BWBs can have symmetrical basis elements that induce less distortion on image edges when the coefficients θ are sparsified by thresholding. Symmetrical elements also yield more predictable coefficients, which boosts compression performance [3].

Unfortunately, BWBs not always satisfy condition (A1). As an example, the CDF9/7 synthesis frame elements are far from orthogonal; indeed the coherence is approximately $\frac{1}{2}$ for a 512×512 2-D synthesis frame. As a result, attempts at CS recovery using greedy techniques fails miserably (see Fig. 1(b)). In contrast, since the analysis and synthesis frames are interchangeable, then the approach in [2] is equivalent to standard ℓ_1 -norm minimization, requiring $M = O(K \log(N/K))$ measurements.

In this paper, we will develop a new CS recovery technique for BWBs based on the notion of *structured sparsity* [4], which can provide near-optimal recovery from as few as $O(K)$ CS measurements. The particular model we apply is the quad-tree sparse/compressible model of [4], which is prevalent in BWB synthesis coefficient vectors for natural images. To provide recovery performance guarantees for signals with structured sparsity in a frame rather than a basis, we marry the concepts of the D-RIP [2], which requires near-isometry for signals with sparse synthesis coefficient vectors, with the structured RIP and RAMP [4] that restricts this near isometry only to signals with synthesis coefficient vectors that follow the quad-tree sparsity

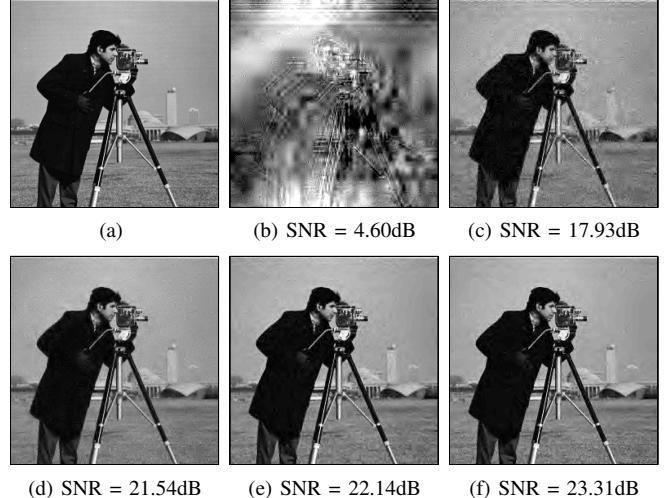


Fig. 1. (a) Original *Cameraman* image. Sparse recovery of the 512×512 *Cameraman* test image from $M = 60000$ noiselet measurements using: (b) CDF9/7 BWB and conventional CoSaMP [5] recovery; (c) D8 OWB and conventional CoSaMP; (d) CDF9/7 BWB and ℓ_1 -norm minimization; (e) D8 OWB and tree-structured CoSaMP [4]; (f) CDF9/7 BWB and tree-structured CoSaMP. The CoSaMP-based algorithms use $K = 10000$.

and compressibility models. The number of measurements needed in these cases is still $M = O(K)$. This class of signals includes the majority of the set of natural images, which can be shown to belong in a sufficiently smooth Besov space.

The benefits of structured sparse recovery in a BWB are clear from Fig. 1(f), which boasts both a higher recovery signal-to-noise ratio (SNR) and noticeably sharper edges and less ringing than the D8 OWB recovery in Fig. 1(c,e) or the CDF9/7 BWB recovery in Fig. 1(d). Our results can be easily extended to more general BWBs and redundant wavelet representations for smooth signals.

REFERENCES

- [1] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Info. Theory*, vol. 54, no. 5, pp. 2210–2219, May 2008.
- [2] E. J. Candès, D. Needell, Y. C. Eldar, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," May 2010, Preprint.
- [3] D. S. Taubman and M. W. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards and Practice*, Kluwer, 2001.
- [4] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Info. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
- [5] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2008.

A Convex Approach for Structured Wavelet Sparsity Patterns

Nikhil Rao*, Robert Nowak*, Stephen Wright*, Nick Kingsbury†

*University of Wisconsin- Madison
†University of Cambridge

I. INTRODUCTION

Hidden Markov Trees are used to model statistical dependencies of wavelet transform coefficients, providing a more accurate reconstruction as compared to independent coefficient-wise approaches like the lasso. However, in linear inverse problems such as compressed sensing and deblurring, the presence of the sensing (or blurring) matrix mixes up the dependencies, and the usual tree-based algorithms cannot be applied. Past work has dealt with this by resorting to greedy or suboptimal iterative reconstruction methods [1], [2], [5], [6]. Based on prior work in modeling DWT coefficients [4], we make the following key observations: 1) The sparsity patterns are highly structured, and so independent coefficient-wise thresholding techniques are suboptimal, and 2) The natural groupings of dependent coefficients overlap with each other. To overcome these drawbacks, we propose a modeling technique based on modeling groups of coefficients, and solve convex optimization problems that arise out of using appropriate penalties [3].

II. MODELING DWT COEFFICIENTS

We group DWT coefficients based on the observed statistical dependencies. We group parent-child pairs of coefficients across scale, to account for the inter scale dependencies, and pairs of adjacent coefficients in the same scale to account for intra-scale dependencies. Since the groups overlap, we use the recent overlapping group lasso formulation developed in [3], replicating the overlapping variables to decouple the groups, and use standard methods such as SpaRSA [7]. Another modeling strategy is to group the coefficients in hierarchies along paths from the root of the tree to the leaf (See Fig. 1)

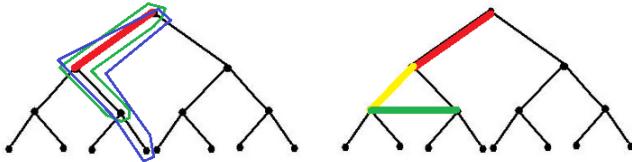


Fig. 1. Modeling DWT coefficients into groups. The rings (left) depict the hierarchical groups. The colored edges (right) depict the parent-child pairs in groups, and also the siblings in the same scale forming a group.

III. RESULTS

Our contribution is four-fold:

- We model the wavelet coefficients into groups, that mimic their statistical dependencies
- We use efficient convex optimization techniques to solve the recovery problem, using overlapping group sparsity penalties
- We develop new bounds for the number of iid gaussian measurements needed for accurate reconstruction of such group sparse signals
- We experimentally show that our technique performs better than the standard lasso, on both toy and real images, in both compressed sensing and image deblurring applications.

As an example of our work, Fig. 2 shows the reconstruction for a noisy version of the cameraman image (top). AWGN of variance 0.3 was added to the image, after normalization. We used compressive measurements to recover the image. The image was resized to size 64×64 , and we used 800 measurements to reconstruct it after vectorizing (length 4096). Fig 2 (bottom) shows the reconstruction in a deblurring task, with the original image blurred with a gaussian filter of variance 1. In the figure, OGlasso refers to the overlap group lasso with groups as in Fig 1 (right).

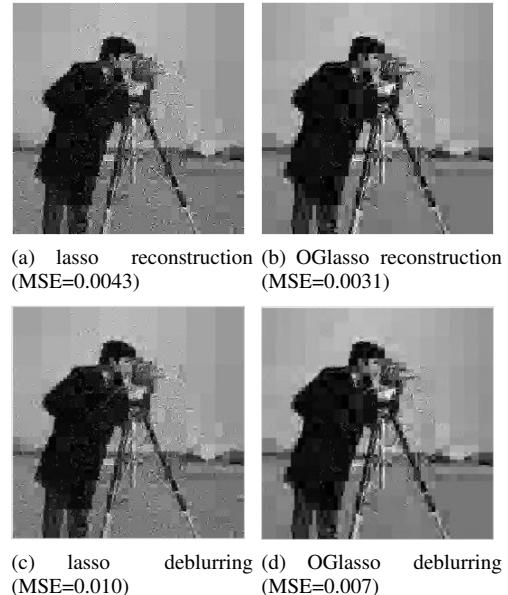


Fig. 2. Performance on the cameraman image

REFERENCES

- [1] La C. and Do M.N. Tree based orthogonal matching pursuit algorithm for signal reconstruction. *IEEE International Conference on Image Processing, Atlanta, GA.*, pages 1277 – 1280, Oct 2006.
- [2] Huang J., Zhang T., and Metaxas D. Learning with structured sparsity. *Technical report, arXiv:0903.3002. Preprint available at http://arxiv.org/pdf/0903.3002v2*, May 2009.
- [3] Jacob L., Obozinski G., and Vert J.P. Group lasso with overlap and graph lasso. *Proceedings of the 26th International Conference on machine Learning*, 2009.
- [4] Crouse M.S., Nowak R.D., and Baraniuk R.G. Wavelet based statistical signal processing using hidden markov models. *Transactions on Signal Processing*, 46(4):886–902, 1998.
- [5] Schniter P. Turbo reconstruction of structured sparse signals. *Proc. Conference on Information Sciences and Systems*, Mar 2010.
- [6] Baraniuk R.G., Cevher V., M.F. Duarte, and Hegde C. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 2010.
- [7] Wright S.J., Nowak R.D., and Figueiredo M.A.T. Sparse reconstruction by separable approximation. *Transactions on Signal Processing*, 57:2479–2493, 2009.

Hybrid Synthesis-Analysis Frame-Based Regularization: A Criterion and an Algorithm

Manya V. Afonso

Instituto de Sistemas e Robótica
Instituto Superior Técnico
1049-001 Lisboa, Portugal

José M. Bioucas-Dias and Mário A. T. Figueiredo

Instituto de Telecomunicações
Instituto Superior Técnico
1049-001 Lisboa, Portugal

I. INTRODUCTION

Consider the problem of estimating a signal/image \mathbf{x} from observations \mathbf{y} that follow the usual linear model $\mathbf{y} = \mathbf{B}\mathbf{x} + \mathbf{n}$, where \mathbf{B} represents a linear observation (e.g., convolution, compressive sensing) and \mathbf{n} is white Gaussian noise. Most frame-based approaches to regularize this inverse problem fall in one of two classes [1], [2]: (i) *synthesis* formulations, which are based on representing the unknown image as $\mathbf{x} = \mathbf{W}\beta$, where \mathbf{W} is the synthesis operator of a (tight) frame, and β is the vector of representation coefficients, to be estimated by solving the unconstrained convex problem

$$\min_{\beta} \frac{1}{2} \|\mathbf{y} - \mathbf{B}\mathbf{W}\beta\|_2^2 + \tau \phi(\beta) \quad (1)$$

(or a constrained version thereof [5]), where ϕ is a convex sparsity-inducing regularizer (typically, the ℓ_1 norm) and τ its weight; (ii) *analysis* formulations, which estimate the image itself (not its representation coefficients) by solving

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{B}\mathbf{x}\|_2^2 + \tau \psi(\mathbf{P}\mathbf{x}), \quad (2)$$

where \mathbf{P} is the analysis operator of a (tight) frame and ψ a convex sparsity-inducing regularizer (usually, also the ℓ_1 norm). If \mathbf{W} is an orthogonal frame, $\mathbf{P} = \mathbf{W}^{-1}$, and $\phi = \psi$, (1) and (2) are equivalent [1]; in general, namely for overcomplete frames, they are not equivalent. Although some debate and research have focused on comparing the two approaches [2], there is no consensus on which of the two is to be preferred. In this paper, we merge the two formulations, by proposing a hybrid synthesis-analysis criterion and an alternating direction algorithm for solving the resulting optimization problem.

II. PROPOSED APPROACH

Our hybrid synthesis-analysis criterion is embodied in an unconstrained problem, where the regularizer term is the sum of the synthesis and analysis regularizers from (1) and (2),

$$\min_{\beta} \frac{1}{2} \|\mathbf{y} - \mathbf{B}\mathbf{W}\beta\|_2^2 + \tau_1 \phi(\beta) + \tau_2 \psi(\mathbf{P}\mathbf{W}\beta), \quad (3)$$

where \mathbf{W} and \mathbf{P} are, respectively, the synthesis and analysis operators of two different tight frames (or of the same tight frame; notice that, even in this case, $\mathbf{P}\mathbf{W} \neq \mathbf{I}$). A different hybrid synthesis-analysis (called *balanced*) formulation was recently proposed [3]; however, it requires the analysis and synthesis operators to be of the same frame, thus it is less general.

We attack problem (3) using the variant of the *alternating direction method of multipliers* (ADMM, [4]) that we have proposed in [5] for problems involving the sum of an arbitrary number of convex terms. Each iteration of the algorithm involves applying the Moreau proximity operators of ϕ and ψ (which, if both are ℓ_1 norms, correspond to soft thresholdings), and a least squares minimization, which

is efficiently solved, under the following assumptions: \mathbf{W} and \mathbf{P} are, respectively, the synthesis and analysis operators of two Parseval frames ($\mathbf{W}\mathbf{W}^H = \mathbf{I}$ and $\mathbf{P}^H\mathbf{P} = \mathbf{I}$), for which fast transforms exist; \mathbf{B} models a periodic convolution, a subsampling (*i.e.*, we have an inpainting problem), or a partially observed Fourier transform (*i.e.*, one of the classical compressive imaging problems). Finally, we show that sufficient conditions for convergence are satisfied.

III. EXPERIMENTS AND CONCLUSIONS

We compare the hybrid formulation with pure synthesis and analysis criteria (solved via the algorithm from [6]), on several benchmark image deconvolution and reconstruction problems (see details of the problems in [6]). For \mathbf{W} , we use a 4-level redundant Haar frame; for \mathbf{P} , we adopt a 4-level redundant Daubechies-4 frame. Both ϕ and ψ are ℓ_1 norms. To sidestep the issue of adjusting the regularization weights, we simply hand-tune them for maximal ISNR (improvement in SNR); of course, this is inapplicable in practice. Since there is no space in this extended abstract for detailed results, we present a summary of the conclusions drawn from the experiments:

- The analysis and hybrid approaches clearly outperform the synthesis approach in terms of ISNR.
- The synthesis approach reaches its best ISNR faster (by a factor of $5 \sim 10$) than the analysis approach.
- The hybrid approach mildly outperforms the analysis approach in terms of ISNR.
- The hybrid approach reaches its best ISNR faster (by a factor of $2 \sim 3$) than the analysis approach.

Summarizing, the hybrid approach (efficiently handled by the proposed algorithm) yields the best speed/ISNR trade-off: it is preferable to the pure analysis criterion, since it is faster; it is preferable over the synthesis criterion, as it achieves a clearly better ISNR. Of course, these conclusions are based on a limited set of experiments; more work is needed to fully assess the relative merits of these approaches.

REFERENCES

- [1] M. Elad, P. Milanfar, and R. Rubinstein, “Analysis versus synthesis in signal priors,” *Inverse Problems*, vol. 23, pp. 947–968, 2007.
- [2] I. Selesnick and M. Figueiredo, “Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors,” in *Proceedings of SPIE*, vol. 7446 (Wavelets XIII), 2009.
- [3] Z. Shen, K.-C. Toh, and S. Yun, “An accelerated proximal gradient algorithm for frame based image restorations via the balanced approach,” *SIAM Jour. Imaging Sciences*, 2010, to appear.
- [4] J. Eckstein and D. Bertsekas, “On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators,” *Mathematical Programming*, vol. 55, no. 3, pp. 293–318, 1992.
- [5] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, “An augmented Lagrangian based method for the constrained formulation of imaging inverse problems,” *IEEE Trans. Image Proc.*, vol. 20, pp. 681–695, 2011.
- [6] ———, “Fast image recovery using variable splitting and constrained optimization,” *IEEE Trans. Image Proc.*, vol. 19, pp. 2345–2356, 2010.

Cosparse Analysis Modeling

Sangnam Nam*, Michael E. Davies†, Michael Elad‡ and Rémi Gribonval*

*INRIA Rennes – Bretagne Atlantique, France

†IDCOM & Joint Research Institute for Signal and Image Processing, Edinburgh University, UK

‡Department of Computer Science, The Technion – Israel Institute of Technology, Israel

Abstract—In the past decade there has been a great interest in a synthesis-based model for signals, based on sparse and redundant representations. This work considers an alternative *analysis-based* model, where an analysis operator multiplies the signal, leading to a *cosparse* outcome. We consider this analysis model, in the context of a generic missing data problem. Our work proposes a uniqueness result for the solution of this problem, based on properties of the analysis operator and the measurement matrix. A new greedy algorithm for solving the missing data problem is proposed along with theoretical study of the success of the algorithm and experimental results.

I. INTRODUCTION

Given a set of incomplete linear observation $\mathbf{y} = \mathbf{M}\mathbf{x}_0 \in \mathbb{R}^m$ of a signal $\mathbf{x}_0 \in \mathbb{R}^d$, $m < d$, the assumption that \mathbf{x}_0 admits a *sparse* representation \mathbf{z}_0 in some *synthesis dictionary* \mathbf{D} is known to be of significant help in recovering the original signal \mathbf{x}_0 . Indeed, it is now well understood that under incoherence assumptions on the matrix \mathbf{MD} , one can recover vectors \mathbf{x}_0 with sufficiently sparse representations by solving the optimization problem:

$$\hat{\mathbf{x}}_S := \mathbf{D}\hat{\mathbf{z}}; \quad \hat{\mathbf{z}} := \arg \min_{\mathbf{z}} \|\mathbf{z}\|_\tau \text{ subject to } \mathbf{y} = \mathbf{M}\mathbf{D}\mathbf{z} \quad (1)$$

for $0 \leq \tau \leq 1$.

An alternative to (1) which has also been used successfully in practice is to consider the *analysis* ℓ_τ -optimization [2], [6], [7]:

$$\hat{\mathbf{x}}_A := \arg \min_{\mathbf{x}} \|\Omega \mathbf{x}\|_\tau \text{ subject to } \mathbf{y} = \mathbf{M}\mathbf{x}, \quad (2)$$

where $\Omega : \mathbb{R}^d \rightarrow \mathbb{R}^p$ is an *analysis operator*. Typically the dimensions are $m < d \leq p, n$.

The fact that \mathbf{z}_0 contains few zeros, i.e., is *sparse* may be thought of as the principal reason why one can recover the so-called *sparse* signals via (1). We show that while the optimization (2) has similar look to (1), a different model, which we name the *cosparse analysis model*, is more closely linked to (2) than the sparse synthesis model. In particular, contrary to the sparse model, we are more interested in the signals \mathbf{x}_0 whose analysis representation $\Omega \mathbf{x}_0$ contains *many zeros*. We call such signals *cosparse* and the quantity $\ell := p - \|\Omega \mathbf{x}_0\|_0$ the *cosparsity*.

II. UNIQUENESS

Based on the existing work [1], [3], we establish [4] the uniqueness of cosparse signals in the context of linear inverse problems above. The result we have derived has simple forms for two particular classes of analysis operators: Ω that is in general position, which means that the rows of Ω has no non-trivial linear dependencies, and the popular 2D TV analysis operators Ω that consists of all the vertical and horizontal one-step differences in a 2D image. For these two types of Ω , we have:

The authors acknowledge the support by the European Community's FP7-FET program, SMALL project, under grant agreement no. 225913.

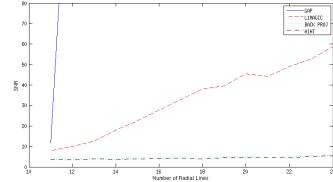


Fig. 1. SNR vs Number of radial observation lines in the Shepp Logan phantom recovery. The line for GAP is clipped because the SNR was over 150 from 12 radial lines.

Proposition 1. 1) Let Ω be in general position. Then, the problem $\mathbf{y} = \mathbf{M}\mathbf{x}$ has at most one ℓ -cosparsely solution if and only if $m \geq 2(d - \ell)$.

2) Let Ω be the 2D TV analysis operator. Then, the problem $\mathbf{y} = \mathbf{M}\mathbf{x}$ has at most one ℓ -cosparsely solution if $m + \ell \geq 2d$.

III. ALGORITHM, THEORY, AND EXPERIMENTAL RESULT

With the uniqueness property established, we propose a new greedy algorithm which aims to recover cosparsely signals based on incomplete linear observations. This algorithm, named the Greedy Analysis Pursuit (GAP), may be considered as the counterpart of the Orthogonal Matching Pursuit (OMP) in the sparse model. However, the GAP tries to detect the elements *outside* the locations of the zeros of analysis representations, this way carving its way towards the index set of zeros in the end.

We then provide a theoretical condition that guarantees the success of both the GAP and the analysis ℓ_1 -minimization in cosparsely signal recovery. Finally, we run some synthetic experiments to demonstrate the effectiveness of the proposed algorithm. Interestingly, we observe that GAP performs better than the analysis ℓ_1 -minimization in the given tasks. In particular, Fig. 1 shows SNR vs the number of radial observation lines in the Shepp Logan phantom recovery problem.

REFERENCES

- [1] T. Blumensath and M. E. Davies. Sampling theorems for signals from the union of finite-dimensional linear subspaces. *IEEE Trans. Inform. Theory*, 55(4):1872–1882, 2009.
- [2] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3):947–968, June 2007.
- [3] Y. M. Lu and M. N. Do. Sampling signals from a union of subspaces. *IEEE Signal Proc. Mag.*, pages 41–47, mar 2008.
- [4] S. Nam, M. E. Davies, M. Elad, and R. Gribonval. The cosparsely analysis model and algorithms. Preprint, 2011.
- [5] S. Nam, M. Davies, M. Elad, and R. Gribonval. Cosparsely analysis modeling - Uniqueness and algorithms. In *Proc. ICASSP 2011*.
- [6] J. Portilla. Image restoration through l0 analysis-based sparse optimization in tight frames. *Proc. of the 16th IEEE Int. Conf. on Image Proc.*, pages 3865–3868, 2009.
- [7] I. W. Selesnick and M. A. T. Figueiredo. Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors. In *Proc. of SPIE*, 7446 (Wavelets XIII), August 2009.

Implications for compressed sensing of a new sampling theorem on the sphere

Jason D. McEwen*, Gilles Puy*, Jean-Philippe Thiran*, Pierre Vandergheynst*, Dimitri Van De Ville*† and Yves Wiaux*†

* Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

† University of Geneva (UniGE), CH-1211 Geneva, Switzerland

Sampling theorems on the sphere state that all the information of a continuous band-limited signal on the sphere may be contained in a discrete set of samples. For an equiangular sampling of the sphere, the Driscoll & Healy (DH) [1] sampling theorem has become the standard, requiring $\sim 4L^2$ samples on the sphere to represent exactly a signal band-limited in its spherical harmonic decomposition at L . Recently, a new sampling theorem on an equiangular grid has been developed by McEwen & Wiaux (MW) [2], requiring only $\sim 2L^2$ samples to represent exactly a band-limited signal, thereby redefining Nyquist rate sampling on the sphere. No sampling theorem on the sphere reaches the optimal number of samples suggested by the L^2 dimension of a band-limited signal in harmonic space (although the MW sampling theorem comes closest to this bound). A reduction by a factor of two in the number of samples required to represent a band-limited signal on the sphere between the DH and MW sampling theorems has important implications for compressed sensing.

Compressed sensing on the sphere has been studied recently for signals sparse in harmonic space [3], where a discrete grid on the sphere is not required. However, for signals sparse in the spatial domain (or in its gradient) a discrete grid on the sphere is essential. A reduction in the number of samples of the grid required to represent a band-limited signal improves both the dimensionality and sparsity of the signal, which in turn affects the quality of reconstruction.

We illustrate the impact of the number of samples of the DH and MW sampling theorems with an inpainting problem, where measurements are made in the spatial domain (as dictated by many applications). A test signal sparse in its gradient is constructed from a binary Earth map, smoothed to give a signal band-limited at $L = 32$. We first solve the total variation (TV) inpainting problem directly on the sphere:

$$\mathbf{x}^* = \arg \max_{\mathbf{x}} \|\mathbf{x}\|_{\text{TV}} \text{ such that } \|\mathbf{y} - \Phi \mathbf{x}\|_2 \leq \epsilon, \quad (1)$$

where M noisy measurements \mathbf{y} of the signal \mathbf{x} are made. The measurement operator Φ represents a random masking of the signal. The TV norm $\|\cdot\|_{\text{TV}}$ is defined to approximate the continuous TV norm on the sphere and thus includes the quadrature weights of the adopted sampling theorem, regularising the gradient computed on the sphere. However, as discussed, the dimensionality of the signal \mathbf{x} is optimal in harmonic space. Consequently, we reduce the dimensionality of our problem by recovering the harmonic coefficients $\hat{\mathbf{x}}$ directly:

$$\hat{\mathbf{x}}^* = \arg \max_{\hat{\mathbf{x}}} \|\Psi \hat{\mathbf{x}}\|_{\text{TV}} \text{ such that } \|\mathbf{y} - \Phi \Psi \hat{\mathbf{x}}\|_2 \leq \epsilon, \quad (2)$$

where Ψ represents the inverse spherical harmonic transform; the signal on the sphere is recovered by $\mathbf{x}^* = \Psi \hat{\mathbf{x}}^*$. For this problem the dimensionality of the signal directly recovered $\hat{\mathbf{x}}$ is identical for both sampling theorems, however sparsity in the spatial domain remains superior (*i.e.* fewer non-zero values) for the MW sampling theorem.

This work is supported by CIBM of the Geneva and Lausanne Universities, EPFL, and the SNSF, Leenaards and Louis-Jeantet foundations.

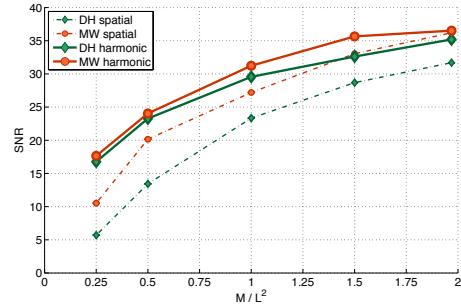
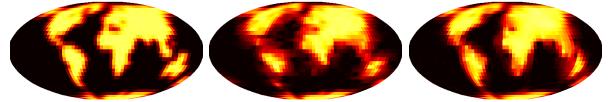


Fig. 1. Reconstruction performance for the DH and MW sampling theorems



(a) Ground truth (b) DH reconstruction (c) MW reconstruction

Fig. 2. Reconstructed Earth topographic data for $M/L^2 = 1/2$

Reconstruction performance is plotted in Fig. 1 when solving the inpainting problem in the spatial (1) and harmonic (2) domains, for both sampling theorems (averaged over ten simulations of random measurement operators and independent and identically distributed Gaussian noise). Strictly speaking, compressed sensing corresponds to the range $M/L^2 < 1$ when considering the harmonic representation of the signal. Nevertheless, we extend our tests to $M/L^2 \sim 2$, corresponding to the equivalent of Nyquist rate sampling on the MW grid. In all cases the superior performance of the MW sampling theorem is evident. In Fig. 2 we show example reconstructions, where the superior quality of the MW reconstruction is again clear.

Although recovering the signal in the harmonic domain is more effective, it is also computationally more demanding. At present we are thus limited to low band-limits. To solve the convex optimisation problem in the harmonic domain both the inverse spherical harmonic transform and its adjoint operator are required. A fast inverse spherical harmonic transform exists [2], from which a fast adjoint operator follows directly. The application of fast inverse and adjoint operators is the focus of ongoing research and will allow compressed sensing problems on the sphere to be tackled effectively at much higher band-limits.

REFERENCES

- [1] J. R. Driscoll and D. M. J. Healy, "Computing Fourier transforms and convolutions on the sphere," *Advances in Applied Mathematics*, vol. 15, pp. 202–250, 1994.
- [2] J. D. McEwen and Y. Wiaux, "A novel sampling theorem on the sphere," *IEEE Trans. Sig. Proc.*, submitted, 2011.
- [3] H. Rauhut and R. Ward, "Sparse recovery for spherical harmonic expansions," *ArXiv:1102.4097*, 2011.

Compressive Sensing for Gaussian Dynamic Signals

Wei Dai, Dino Sejdinovic, and Olgica Milenkovic

Emails: wei.dai1@imperial.ac.uk, d.sejdinovic@bristol.ac.uk, milenkov@uiuc.edu

Consider a linear dynamic system described by the update equations $\mathbf{x}_t = \Psi_t \mathbf{x}_{t-1} + \mathbf{u}_t$, $\mathbf{y}_t = \Phi_t \mathbf{x}_t + \mathbf{v}_t$. Here, $\mathbf{x}_t \in \mathbb{R}^n$ represents the state vector of the system, $\mathbf{y}_t \in \mathbb{R}^m$ denotes the measurement vector, $\mathbf{u}_t \in \mathbb{R}^n$ and $\mathbf{v}_t \in \mathbb{R}^m$ are Gaussian innovation vectors with $\mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_u)$ and $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_v)$, respectively. The subscript $t = 1, 2, \dots$ describes the time instances at which the signal is observed. Suppose that the statistics of \mathbf{x}_{t-1} are known, and given by $\mathbf{x}_{t-1} \sim \mathcal{N}(\hat{\mathbf{x}}_{t-1}, \Sigma_{t-1})$. Given \mathbf{y}_t , Ψ_t and Φ_t , the MAP estimate of \mathbf{x}_t , denoted by $\hat{\mathbf{x}}_t$, coincides with the corresponding MMSE estimate.

Now suppose that one has prior information that \mathbf{x}_t is K -sparse. The MAP estimator that takes the sparsity assumption into consideration is given by

$$\hat{\mathbf{x}}_t = \arg \max_{\mathbf{x}: \|\mathbf{x}\|_0 \leq K} p_{\mathbf{X}_t | \mathbf{Y}_t, \mathbf{X}_{t-1}}(\mathbf{x} | \mathbf{y}_t, \hat{\mathbf{x}}_{t-1}),$$

where the pseudo-norm $\|\cdot\|_0$ counts the number of non-zero entries of its argument. Let $\mathbf{A}_t = 2(\Sigma_u^{-1} + \Phi_t^T \Sigma_v^{-1} \Phi_t)$, $\mathbf{b}_t = -2(\Sigma_u^{-1} \Psi_t \hat{\mathbf{x}}_{t-1} + \Phi_t^T \Sigma_v^{-1} \mathbf{y}_t)$, and $f_t(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A}_t \mathbf{x} + \mathbf{b}_t^T \mathbf{x}$. It can be verified that the sparse MAP estimator is equivalent to

$$\hat{\mathbf{x}}_t = \arg \min_{\mathbf{x}: \|\mathbf{x}\|_0 \leq K} f_t(\mathbf{x}). \quad (1)$$

At the first glance, the Gaussian sparse modelling looks arbitrary. The common strategy for dynamic CS usually involves certain sparsity-promoting distributions, which often result in reconstructions with high computational complexity and weak performance guarantees. Note that Gaussian modelling has been successfully applied to dynamic signal processing, and that in many applications, e.g. MRI imaging, the dynamic signal at each time instance is sparse. Our model combines the advantages of both Gaussian and sparse modelling and renders good performance guarantees.

It is NP-hard to solve the optimization problem (1). We therefore propose a practical greedy algorithm to solve (1). It is based on the well-known subspace pursuit (SP) algorithm for standard compressive sensing, and therefore termed SP-MAP. The details are described in Algorithm 1. *It can be proved* that the proposed SP-MAP algorithm coincides the standard SP algorithm when $\Sigma_u = \sigma_u^2 \mathbf{I}$, $\Sigma_v = \mathbf{I}$ and $\sigma_u^2 \rightarrow \infty$. The *performance guarantees* of the proposed SP-MAP algorithm are based on RIP like conditions and will be detailed in the full version of this abstract.

We performed extensive numerical simulations to test our approach for K -sparse dynamical signals. In order to generate a sparse Gaussian dynamic signal, we use the model $\mathbf{x}_t = \mathcal{T}_k(\Phi_t \mathbf{x}_{t-1} + \mathbf{u}_t)$, where the nonlinear mapping $\mathcal{T}_K(\mathbf{x})$ produces a vector that agrees with \mathbf{x} in the K largest magnitude entries, and has all other coordinates equal to zero.⁴⁶

Algorithm 1 The SP-MAP Algorithm

Let ℓ_{\max} be the maximum iterations at each time instance. Let $\hat{\mathbf{x}}_0 = \mathbf{0}$. At time instance t , perform the following operations. Initialization:

- 1) Define $\mathbf{x}'_t = \Psi_t \hat{\mathbf{x}}_{t-1}$, $\mathbf{A} = 2(\Sigma_u^{-1} + \Phi_t^T \Sigma_v^{-1} \Phi_t)$ and $\mathbf{b} = -2(\Sigma_u^{-1} \mathbf{x}'_t + \Phi_t^T \Sigma_v^{-1} \mathbf{y}_t)$.
- 2) Let $\ell = 0$. Let $\hat{\mathbf{x}}_t = -\mathbf{A}^{-1} \mathbf{b}$. Let \mathcal{K} be the set of the K indices corresponding to the largest $\mathbf{A}_{i,i} |\hat{\mathbf{x}}_{t,i}|^2$'s, $i \in [n]$. Define $\hat{\mathbf{x}}_t^{(\ell)}$ such that $\hat{\mathbf{x}}_{t,\mathcal{K}^c}^{(\ell)} = \mathbf{0}$ and $\hat{\mathbf{x}}_{t,\mathcal{K}}^{(\ell)} = -\mathbf{A}_{\mathcal{K},\mathcal{K}}^{-1} \mathbf{b}_{\mathcal{K}}$.
- 3) Let $\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_t^{(\ell)}$. Compute $f^{(\ell)} = \frac{1}{2} \hat{\mathbf{x}}_t^T \mathbf{A} \hat{\mathbf{x}}_t + \mathbf{b}^T \hat{\mathbf{x}}_t$.

Iterations:

- 1) Let $\ell = \ell + 1$.
 - 2) For every $i \notin \mathcal{K}$, compute $\Delta_i = ((\hat{\mathbf{x}}_{t,\mathcal{K}}, \mathbf{A}_{\mathcal{K},i}) + \mathbf{b}_i)^2 / \mathbf{A}_{i,i}$. Let \mathcal{K}_Δ be the set of the K indices corresponding to the largest Δ_i 's, $i \in \mathcal{K}^c$.
 - 3) Let $\tilde{\mathcal{K}} = \mathcal{K} \cup \mathcal{K}_\Delta$. Define $\tilde{\mathbf{x}}_t$ such that $\tilde{\mathbf{x}}_{t,\tilde{\mathcal{K}}^c} = \mathbf{0}$ and $\tilde{\mathbf{x}}_{t,\tilde{\mathcal{K}}} = -\mathbf{A}_{\tilde{\mathcal{K}},\tilde{\mathcal{K}}}^{-1} \mathbf{b}_{\tilde{\mathcal{K}}}$. For every $i \in \tilde{\mathcal{K}}$, compute $\Delta_i = \mathbf{A}_{i,i} \tilde{\mathbf{x}}_{t,i}^2$.
 - 4) Let \mathcal{K} be the set of the K indices corresponding to the largest Δ_i 's, $i \in \tilde{\mathcal{K}}$. Define $\hat{\mathbf{x}}_t^{(\ell)}$ such that $\hat{\mathbf{x}}_{t,\mathcal{K}^c}^{(\ell)} = \mathbf{0}$ and $\hat{\mathbf{x}}_{t,\mathcal{K}}^{(\ell)} = -\mathbf{A}_{\mathcal{K},\mathcal{K}}^{-1} \mathbf{b}_{\mathcal{K}}$. Compute $f^{(\ell)} = \frac{1}{2} \hat{\mathbf{x}}_t^{(\ell)} \mathbf{A} \hat{\mathbf{x}}_t^{(\ell)} + \mathbf{b}^T \hat{\mathbf{x}}_t^{(\ell)}$.
 - 5) If $f^{(\ell)} > f^{(\ell-1)}$, quit the iterations.
 - 6) Let $\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_t^{(\ell)}$. If $\ell \geq \ell_{\max}$, quit the iterations. Otherwise, go to Step 1 for the next iteration.
-

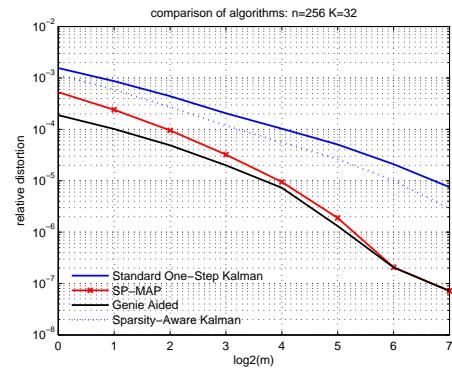


Figure 1. Comparison of reconstruction algorithms.

Figure 1 compares the proposed SP-MAP algorithm with other algorithms designed for dynamic CS. According to the simulation results, the SP-MAP algorithm outperforms others and it performs very close to the genie-aided approach when the number of samples per time instance is sufficient.

References are omitted due to the space limitation.

Simultaneous Estimation of Sparse Signals and Systems at Sub-Nyquist Rates

Hojjat Akhondi Asl and Pier Luigi Dragotti

EXTENDED ABSTRACT

In this work, we consider the problem of system identification based on a sparse sampling system. Unlike standard techniques for system identification which require the sampling rate to be at or above the Nyquist rate, we use sparse sampling techniques to identify the system at sub-Nyquist sampling rates. We propose a novel algorithm for simultaneous estimation of sparse signals along with system identification using the theories of finite rate of innovation (FRI) sampling [3], [1]. Specifically, we will divide the estimation problem into two stages where we first assume that the input sparse signal is known, so that the problem simplifies to a system identification problem only and then in the second stage, we consider the problem of simultaneously estimating the input sparse signal and also the linear system, known as blind system identification, and propose a novel iterative algorithm for that setup. We will show that, based on our numerical simulations, the solution to the second problem is normally convergent.

System Identification with Known Input Signal

For this scenario, as shown on Figure 1, a two-channel system is proposed for sampling the input sparse signal with and without the unknown system. In the figure, $g(x)$ represents the known input signal, $\psi(x)$ represents the unknown system to be identified, $\varphi(x)$ represents the pre-defined sampling kernel which we assume to be purely imaginary E-splines [2] in both channels, T represents the sampling interval and s_k represent the samples. In the first channel, the input signal is

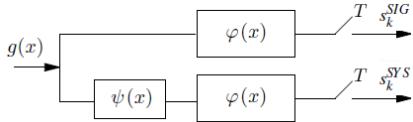


Fig. 1. System identification setup with known input signal

directly sampled with the kernel $\varphi(x)$ and given the obtained samples which we denote with s_k^{SIG} , the exponential moments of the input signal, denoted with τ_m^{SIG} , are calculated [1]. In the second channel, the same input signal is fed through the unknown system $\psi(x)$ and then sampled with the same sampling kernel. Same as in the first channel, given the samples s_k^{SYS} , the exponential moments τ_m^{SYS} are calculated.

With purely imaginary E-spline sampling kernel employed, by dividing the exponential moments obtained from the two channels, it can be shown that the Fourier transform of the unknown function can be obtained. Given the partial Fourier

The authors are with the C&SP Group, Electrical and Electronic Engineering, Imperial College London, Exhibition Road, London SW7 2AZ, England. Tel: +44 (0) 20 759-46192. E-mails: {hojjat.akhondi-asl03 and p.dragotti@imperial.ac.uk}.

transform of the unknown system, there will be an inverse problem to solve for the unknown parameters of the unknown system. In our work, we show for cases such as finite impulse response (FIR) filters (e.g. acoustic room impulse response estimation or line echo cancelation), B-splines (e.g. camera lens calibration) and E-splines (e.g. estimation of the electronic components of a finite order electronic circuit), we can solve the above inverse problem and identify the system. It should be pointed out that the above method works regardless of the structure of the input signal.

Blind System Identification

When both the signal and the system are unknown, the previous solution cannot be used directly and the problem is in general more involved. However, a recursive version of the discussed method can be utilized to estimate both the input sparse signal and the unknown system.

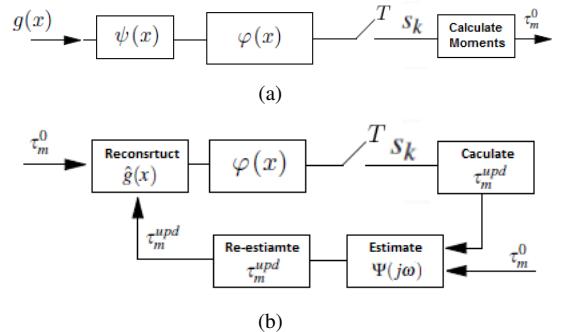


Fig. 2. The setup proposed for recursive estimation

In our work we assume that the input sparse signal is a stream of Diracs with unknown locations and amplitudes. As shown in Figure 2(a), the unknown input signal is fed to the unknown system $\psi(x)$ and then is sampled with our pre-specified purely imaginary E-spline sampling kernel. The annihilating filter method [3], [1] is directly applied to the exponential moments τ_m^0 and an initial estimate of the input signal is obtained, denoted as $\hat{g}(x)$ (Figure 2(b)). The estimated signal $\hat{g}(x)$ is recursively fed back to sampling kernel and its corresponding updated exponential moments are calculated at each recursion, denoted with τ_m^{upd} . By dividing the updated exponential moments τ_m^{upd} and the initial measurements τ_m^0 , an estimate of the Fourier transform of the unknown system is obtained. From this estimate, the unknown parameters of the unknown system are estimated and the measurements τ_m^{upd} are re-calculated. Our empirical results show that by applying the above method recursively, the estimations converge to the actual input signal $g(x)$ and the unknown function $\psi(x)$.

REFERENCES

- [1] P.L. Dragotti, M. Vetterli, and T. Blu. Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix. *In IEEE Transactions on Signal Processing*, 55(5):1741–1757, May 2007.
- [2] M. Unser and T. Blu. Cardinal Exponential Splines: Part I-theory and filtering algorithms. *In IEEE Transactions on Signal Processing*, 53(4):1425–1438, April 2005.
- [3] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *In IEEE Transactions on Signal Processing*, 50(6):1417–1428, June 2002.

A Hierarchical Re-weighted- ℓ_1 Approach for Dynamic Sparse Signal Estimation

Adam Charles and Christopher Rozell

School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, GA, 30332-0250
Email: {acharles6, crozell}@gatech.edu

Index Terms—Dynamic Systems, State Estimation, Compressive Sensing, Hierarchical Models, Bayesian Analysis

Compressive sensing results have allowed accurate reconstruction of highly undersampled signals by leveraging known signal structure [1]. Recently, there has been a push to extend these results into an area of great interest for a large number of fields: the estimation of dynamically changing signals [2]–[5]. If known, or even partially known dynamics are transforming a state, then past observations should be able to be incorporated into the estimation process of a state at any given time in order to increase the accuracy of the estimation. Typically a dynamical state $\mathbf{x}_n \in \mathbb{R}^N$ is assumed to evolve with some approximately known dynamics $f_n(\cdot)$ as

$$\mathbf{x}_n = f_n(\mathbf{x}_{n-1}) + \boldsymbol{\nu}_n, \quad (1)$$

where $\boldsymbol{\nu}_n$ is called the innovations and can be interpreted as the limitation of our knowledge of the system dynamics. Given a set of linear measurements at each iteration,

$$\mathbf{y}_n = \mathbf{G}_n \mathbf{x}_n + \boldsymbol{\epsilon}_n, \quad (2)$$

where $\mathbf{y}_n, \boldsymbol{\epsilon}_n \in \mathbb{R}^M$ are the measurements and measurement error, respectively, we wish to estimate the underlying evolving state. More specifically, we wish to recover the current state at each time step as best as possible given all previous measurements. In previous work [4], we explore a framework in which propagating first order statistics and utilization of appropriate ℓ_1 norms allow for accurate estimation when the state, the innovations or both are sparse.

In least-squares based state estimation, however, higher order statistics are propagated in order to obtain more accurate estimates at each iteration. For instance in the case of the Kalman filter, which arises when under assumptions of linearity in the modeled dynamics and Gaussian statistics in the innovations and measurement noise, a covariance matrix is propagated along with the mean to obtain an optimal estimate. In this work, we expand on the previously introduced framework in order to include similar higher order statistics by introducing a hierarchical model inspired by the reweighted ℓ_1 sparse inference method first proposed in [6]. We use previous information in a way similar to [7] in that we are leveraging the weightings $\boldsymbol{\Lambda} = \text{diag}(\lambda_i)$ in the optimization

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \|\mathbf{y} - \mathbf{G}\mathbf{x}\|_2^2 + \|\boldsymbol{\Lambda}\mathbf{x}\|_1 \quad (3)$$

in order to propagate information about our prediction and our confidence thereof of the next state. By using a Gamma prior over each element of $\boldsymbol{\lambda}$ in a Bayesian setting, we determine the expectation-maximization (EM) update equations in order to determine \mathbf{x}_n and $\boldsymbol{\lambda}_n$ at each iteration to be

$$\boldsymbol{\lambda}^t[i] = \frac{2}{|\mathbf{x}^{t-1}[i]| + f_n(\mathbf{x}_{n-1})[i] + \beta} \quad (4)$$

$$\mathbf{x}_n^t = \arg \min_{\mathbf{x}} \left[\|\mathbf{y}_n - \mathbf{G}_n \mathbf{x}\|_2^2 + \sum_i \boldsymbol{\lambda}^t[i] |\mathbf{x}[i]| \right] \quad (5)$$

where β is a small positive value which ensures stability in the $\boldsymbol{\lambda}$ values and t indicates the EM iteration. The EM algorithm run to convergence, which typically occurs for $10 \leq t \leq 30$.

We show improvements on simulated data using the adaptation of the second order variables over similar first order estimation programs in both the steady state relative mean squared error (rMSE) and the robustness. For example at sampling rates below CS recovery limits, steady state errors can be reduced from 2.48% using first order methods to 0.67% with the re-weighted model. Additionally, up to 30% of the signal sparsity locations can be erroneous and the re-weighted model continues to outperform both time-independent basis pursuit de-noising as well as the first order models.

REFERENCES

- [1] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans on Information Theory*, vol. 52, no. 2, Feb 2006.
- [2] N. Vaswani, “Kalman filtered compressed sensing,” *Proc of ICIP 2008*, pp. 893–896, 2008.
- [3] J. Ziniel, L. C. Potter, and P. Schniter, “Tracking and smoothing of time-varying sparse signals via approximate belief propagation,” *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, 2010.
- [4] A. Charles, M. S. Asif, J. Romberg, and C. Rozell, “Sparsity penalties in dynamical system estimation,” *Proc of the CISS*, March 2011.
- [5] M. S. Asif, A. Charles, J. Romberg, and C. Rozell, “Estimation and dynamic updating of time-varying signals with sparse variations,” *ICASSP*, 2011.
- [6] E. Candes, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec 2004, special Issue on Sparsity.
- [7] M. A. Khajehnejad, W. Xu, S. Avestimehr, and B. Hassibi, “Weighted ℓ_1 minimization for sparse recovery with prior information,” <http://arxiv.org/abs/0901.2912v1>, 2009.

Weighted ℓ_p Constraints in Noisy Compressed Sensing

Laurent Jacques*, David Kenric Hammond† and Jalal Fadili‡

*ICTEAM/ELEN, Université catholique de Louvain (UCL), Belgium. †Neuroinformatics Center, University of Oregon, USA.

‡GREYC CNRS-ENSICAEN-Université de Caen, 14050 Caen, France.

The advent of the Compressed Sensing (CS) paradigm revitalizes the way signals are acquired. In a nutshell, CS shows how sparse or compressible signals in a given basis can be reconstructed from fewer linear measurements than the ambient dimension N of the signal space [1], [5]. The gist of this approach lies in the use of a sensing matrix sufficiently incoherent from the signal sparsity basis. This happens with high probability for a large class of random matrix constructions as soon as the number of measurements M is higher than “few multiples” of the signal sparsity K . For instance, for Random Gaussian matrices, when $M = O(K \log N/K)$.

As a matter of fact, actual signal acquisition systems are often (if not always) corrupted by noise with various distribution laws (*e.g.*, Gaussian, Poisson, uniform or impulsive noises). Sensing techniques relying on CS cannot escape this, which motivates the study of CS reconstruction robustness in the presence of such perturbations.

In this work, we consider the corruption of CS measurements by heteroscedastic noise following a Generalized Gaussian Distribution (GGD). More precisely, the sensing model of a sparse (or compressible) signal $\mathbf{x} \in \mathbb{R}^N$ in the canonical basis is

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{n}, \quad (1)$$

where $\Phi \in \mathbb{R}^{M \times N}$ stands for the sensing matrix, and the noise $\mathbf{n} \in \mathbb{R}^M$ follows a (central) GGD. That is, $n_i \sim \text{GGD}(0; \alpha_i, p) \propto \exp -|t/\alpha_i|^p$, for a shape parameter $p > 1$ and a scale $\alpha_i > 0$. This extends [4] where only the case $\alpha_i = \alpha > 0$ is analyzed.

In order to reconstruct the signal, we propose an adaptation of the Basis Pursuit DeNoise (BPDN) program [5] for GGD noises called Basis Pursuit for Generalized Gaussian Noise (BPGGN):

$$\mathbf{x}^* = \arg \min_{\mathbf{u}} \|\mathbf{u}\|_1 \text{ s.t. } \|\mathbf{y} - \Phi \mathbf{u}\|_{p, \mathbf{w}} \leq \epsilon_{p, \mathbf{w}}. \quad (\text{BPGGN})$$

This solver proceeds as before by minimizing the (sparsity promoting) ℓ_1 -norm of the signal under a weighted $\ell_{p, \mathbf{w}}$ -norm fidelity term (with $p \geq 1$) adjusted so that $\epsilon_{p, \mathbf{w}}^p$ provides an upper bound (with high probability) on $\|\mathbf{n}\|_{p, \mathbf{w}}^p := \sum_i |w_i n_i|^p$. Our implicit objective is to adjust the weights $\mathbf{w} = (w_1, \dots, w_M) \in \mathbb{R}_+^M$ to the noise characteristics for minimizing the final reconstruction error.

For $p = 2$, $\mathbf{w} = \mathbf{1} := (1, \dots, 1)$, BPGGN reduces to BPDN which is $\ell_2 - \ell_1$ *instance optimal* when Φ satisfies the *Restricted Isometry Property* [5]. In other words, if for some normalization $\mu > 0$, $\mu \|\Phi \mathbf{v}\|$ is “close” to $\|\mathbf{v}\|$ for any K -sparse vector $\mathbf{v} \in \Sigma_K = \{\mathbf{u} : \#\{i : u_i \neq 0\} \leq K\}$, and if the signal \mathbf{x} is (with high probability) a feasible point of the BPDN constraint, then

$$\|\mathbf{x}^* - \mathbf{x}\| \leq C e_0(\mathbf{x}) + D \epsilon_{2,1}/\mu.$$

In this error bound, $C > 0$ and $D > 0$ are dependent on Φ and K only, $e_0(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_K\|_1/\sqrt{K}$ and \mathbf{x}_K is the best K -term approximation of \mathbf{x} [5].

Generalizing what is described in [4], it is possible to characterize the BPGGN stability for any $\mathbf{w} \in \mathbb{R}_+^M$ and $p \geq 2$ if the sensing matrix is well behaved with respect to the $\ell_{p, \mathbf{w}}$ -norm. Specifically, we ask the sensing matrix Φ to respect the generalized Restricted Isometry Property RIP($\ell_{p, \mathbf{w}}, \ell_2 | K, \delta, \mu$) at order $K \in \mathbb{N}$, radius $0 \leq \delta < 1$ and for a normalization $\mu > 0$. That is, for all $\mathbf{x} \in \Sigma_K$,

$$(1 - \delta)^{1/2} \|\mathbf{x}\| \leq \frac{1}{\mu} \|\Phi \mathbf{x}\|_{p, \mathbf{w}} \leq (1 + \delta)^{1/2} \|\mathbf{x}\|. \quad (2)$$

Notice that RIP($\ell_{2,1}, \ell_2 | K, \delta, \mu$) is the common RIP [3], [5].

We can prove that, with very high (controllable) probability, a random matrix $\Phi \sim \mathcal{N}^{M \times N}(0, 1)$, *i.e.*, with $\Phi_{ij} \sim_{\text{iid}} \mathcal{N}(0, 1)$, 49

is RIP($\ell_{p, \mathbf{w}}, \ell_2 | K, \delta, \mu$) as soon as $M^{2/p} = O(K \log N/K)$ and $\mu = \mathbb{E}\|\xi\|_{p, \mathbf{w}}$, for $\xi \sim \mathcal{N}^{M \times 1}(0, 1)$.

We also show that if Φ is a RIP($\ell_{p, \mathbf{w}}, \ell_2 | s, \delta_s, \mu$) matrix for $s \in \{K, 2K, 3K\}$ and $2 \leq p < \infty$, and if $\epsilon_{p, \mathbf{w}}$ guarantees that \mathbf{x} is a feasible point of the BPGGN constraint, then

$$\|\mathbf{x}^* - \mathbf{x}\| \leq A_p e_0(K) + B_p \mu^{-1} \epsilon_{p, \mathbf{w}}, \quad (3)$$

for values $A_p(\Phi, K) = \frac{2(1+C_p-\delta_{2K})}{1-\delta_{2K}-C_p}$, $B_p(\Phi, K) = \frac{4\sqrt{1+\delta_{2K}}}{1-\delta_{2K}-C_p}$, $C_p = O(\sqrt{(\delta_{2K} + \delta_{3K})(p-2)})$ as $p \gg 2$ and $C_p = \delta_{3K} + O(p-2)$ as $p \rightarrow 2$.

This results is interesting for at least the following two situations. First, in the case $p = 2$, for heteroscedastic Gaussian noise variance $\sigma_i^2 = \alpha_i^2/2$, (3) implies that the reconstruction error may be reduced through “cleaning” \mathbf{n} by setting $w_i = 1/\sigma_i$. For a Gaussian matrix $\Phi \sim \mathcal{N}^{M \times N}(0, 1)$, we have $\mu = \mathbb{E}\|\xi\|_{2, \mathbf{w}} \simeq \|\mathbf{w}\|$. Without cleaning (*i.e.*, $\mathbf{w} = \mathbf{1}$), the term $\mu^{-1} \epsilon_{2, \mathbf{w}}$ in (3) is close to $(\sum_i \sigma_i^2)^{1/2}/\sqrt{M}$. Setting $w_i = 1/\sigma_i$, $w_i n_i \sim \mathcal{N}(0, 1)$ and we get $\mu^{-1} \epsilon_{2, \mathbf{w}} \simeq \sqrt{M}/(\sum_i \sigma_i^{-2})^{1/2}$. This second quantity is always smaller than the first. Indeed, $M/(\sum_i \sigma_i^{-2}) = (\sum_i \sigma_i^{-2}/M)^{-1} \leq \sum_i (\sigma_i^{-2})^{-1}/M = \sum_i \sigma_i^2/M$ since the function $1/t$ is convex on \mathbb{R}_+ and $\sum_i 1/M = 1$.

Second, (3) also applies to the context of non-uniform measurement quantization [2], that is, when

$$\mathbf{y} = \mathcal{Q}[\Phi \mathbf{x}] = \Phi \mathbf{x} + \mathbf{n},$$

where, for each vector component, the scalar quantizer $\mathcal{Q}[t]$ maps $t \in \mathbb{R}$ on the level $\tau_k \in \mathbb{R}$ iff t belongs to the quantization bin $\mathcal{R}_k = [t_k, t_{k+1}) \ni \tau_k$, with thresholds $t_0 < t_1 < \dots$. When an *oracle* tells us on which side of each level an unquantized measurement is, we show that the noise \mathbf{n} in (1), which is uniform on each bin, can be approximated as a (half) GGD noise with arbitrary high p . However, since a matrix $\Phi \sim \mathcal{N}^{M \times N}(0, 1)$ is RIP($\ell_{p, \mathbf{w}}, \ell_2 | K, \delta, \mu$) if $M^{2/p} = O(K \log N/K)$, a trade-off must be found between perfect noise modeling and reconstruction controllability. If this RIP holds, writing k_i as the bin label of $(\Phi \mathbf{x})_i$, setting $w_i = 1/|\tau_{k_i} - t_{k_i+1}|$ if $(\Phi \mathbf{x})_i > \tau_{k_i}$ and $w_i = 1/|\tau_{k_i} - t_{k_i}|$ otherwise (thanks to the oracle), we found that

$$\mu^{-1} \epsilon_{p, \mathbf{w}} \leq C (\rho_p \sqrt{p+1})^{-1}. \quad (4)$$

assuming $M^{-1/p} \|\mathbf{w}\|_p \geq \rho_p$ for large M (with $\rho_p \geq \min_i w_i$).

Implementing BPGGN with a monotone operator splitting method [6] and making the weights sign-sensitive, we observe numerically this $O(1/\sqrt{p+1})$ error reduction when CS measurements are quantized with a Lloyd-Max quantizer and without invoking any oracle.

REFERENCES

- [1] D. Donoho, “Compressed Sensing,” *IEEE Tr. Inf. Th.*, **52**(4), pp. 1289–1306, 2006.
- [2] W. Dai, H. V. Pham, and O. Milenkovic, “Distortion-rate functions for quantized compressive sensing,” *IEEE Inf. Th. Work. (ITW)*, pp. 171–175, 2009.
- [3] J. Laska, P. Boufounos, M. Davenport and R. Baraniuk, “Democracy in action: Quantization, saturation, and compressive sensing”, *App. Comp. and Harm. Anal. (ACHA)*, 2011 (to appear).
- [4] L. Jacques, D. Hammond, and M. Fadili, “Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine.” *IEEE Tr. Inf. Th.*, **57**(1), pp. 559–571, 2011.
- [5] E. Candès, “The restricted isometry property and its implications for compressed sensing,” *Compte Rendus Acad. Sciences, Paris, Serie I*, **346**, pp. 589–592, 2008.
- [6] P. Combettes and J. Pesquet, “A Proximal Decomposition Method for Solving Convex Variational Inverse Problems,” *Inverse Problems*, **24**, p. 27, Dec. 2008.

Spread Spectrum for Universal Compressive Sampling

Gilles Puy*, Pierre Vandergheynst*, Rémi Gribonval†, Yves Wiaux*‡

*Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

†Centre de Recherche INRIA Rennes-Bretagne Atlantique, F-35042 Rennes cedex, France

‡University of Geneva (UniGE), CH-1211 Geneva, Switzerland

Abstract—We propose a *universal* and *efficient* compressive sampling strategy based on the use of a spread spectrum technique. The method essentially consists in a random pre-modulation of the signal of interest followed by projections onto randomly selected vectors of an orthonormal basis. The effectiveness of the technique is induced by a decrease of coherence between the sparsity and the sensing bases. The sensing scheme is *universal* for a family of sensing bases in the sense that the number of measurements needed for accurate recovery is optimal and independent of the sparsity matrix. It is also *efficient* as sensing matrices with fast matrix multiplication algorithms can be used. These results are confirmed experimentally through analyses of the phase transition of the ℓ_1 -minimization problem.

I. SPREAD SPECTRUM TECHNIQUE

Let $x \in \mathbb{C}^N$ be an s -sparse digital signals in an orthonormal basis $\Psi = (\psi_1, \dots, \psi_N) \in \mathbb{C}^{N \times N}$ and $\alpha \in \mathbb{C}^N$ be its decomposition in this basis: $\alpha = \Psi^* x$. The spread spectrum technique consists in a pre-modulation of the original signal x by a wide-band signal $c = (c_l)_{1 \leq l \leq N} \in \mathbb{C}^N$, with $|c_l| = 1$ and random phases, and a projection onto m randomly selected vectors of another orthonormal basis $\Phi = (\phi_1, \dots, \phi_N) \in \mathbb{C}^{N \times N}$ [2]. The indices $\Omega = \{l_1, \dots, l_m\}$ of the selected vectors are chosen independently and uniformly at random from $\{1, \dots, N\}$. We denote Φ_Ω^* the $m \times N$ matrix made of the selected rows of Φ^* . The measurement vector $y \in \mathbb{C}^m$ thus reads as

$$y = A_\Omega \alpha \text{ with } A_\Omega = \Phi_\Omega^* C \Psi \in \mathbb{C}^{m \times N}. \quad (1)$$

In the above equation, the matrix $C \in \mathbb{C}^{N \times N}$ stands for the diagonal matrix associated to the sequence c . Finally, we aim at recovering α by solving the ℓ_1 -minimization problem

$$\arg \min_{\bar{\alpha} \in \mathbb{C}^N} \|\bar{\alpha}\|_1 \text{ subject to } y = A_\Omega \bar{\alpha}. \quad (2)$$

II. REDUCING THE MUTUAL COHERENCE BY PRE-MODULATION

In the absence of pre-modulation, i.e. when C is reduced to the identity matrix, the compressive sampling theory already demonstrates that a small number $m \ll N$ of random measurements is sufficient for an accurate and stable reconstruction of α [1]. However, the recovery conditions depend on the mutual coherence $\mu = \max_{1 \leq i, j \leq N} |\langle \phi_i, \psi_j \rangle|$ between Φ and Ψ . The performance is optimal when the bases are perfectly incoherent, i.e. $\mu = N^{-1/2}$, and unavoidably decreases when μ increases.

The spread spectrum technique proposed in this work significantly reduces the mutual coherence μ towards its optimal value [2]. In the presence of a digital pre-modulation by a random Rademacher or Steinhaus sequence $c \in \mathbb{C}^N$, the mutual coherence $\mu = \max_{1 \leq i, j \leq N} |\langle \phi_i, \psi_j \rangle|$ is essentially bounded by the *modulus-coherence* $\beta^2(\Phi, \Psi) = \max_{1 \leq i, j \leq N} \sum_{k=1}^N |\phi_{ki}^* \psi_{kj}|^2$. Indeed, we can show that the mutual coherence μ satisfies

$$N^{-1/2} \leq \mu \leq \beta(\Phi, \Psi) \sqrt{2 \log(2N^2/\epsilon)}, \quad (3)$$

with probability at least $1 - \epsilon$.

This work is supported by the CIBM of the Geneva, Lausanne Universities and EPFL, by the Leenaards and Louis-Jeantet foundations, by the SNSF (grant PP00P2-123438), by the EU FET-Open project FP7-ICT-225913-SMALL, and by the EPFL-Merck Serono Alliance award.

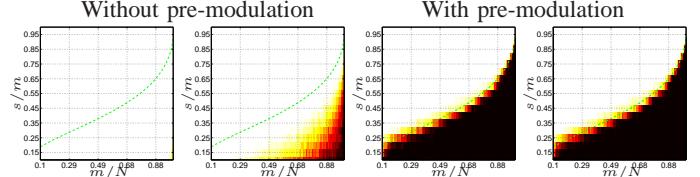


Fig. 1. Phase transition of the ℓ_1 -minimization problem for sparse signals in the Fourier basis and random selection of **Fourier** (first and third panels) or **Hadamard** (second and fourth panels) measurements without and with random modulation. The dashed green line indicates the weak phase transition of Donoho-Tanner [3] and the color bar goes from white to black indicating a probability of recovery from 0 to 1. The domain of recovery becomes optimal with the spread spectrum technique.

Definition 1. (*Universal sensing basis*) An orthonormal basis $\Phi \in \mathbb{C}^{N \times N}$ is called a universal sensing basis if all its entries ϕ_{ki} , $1 \leq k, i \leq N$, are of equal complex magnitude.

For universal sensing bases Φ , e.g. the Fourier, Hadamard, or noiselet transform, we have $\beta(\Phi, \Psi) = N^{-1/2}$ whatever the sparsity matrix Ψ . The mutual coherence μ is thus equal to its optimal value, up to a logarithmic factor, whatever the sparsity matrix considered!

III. SPREAD SPECTRUM UNIVERSALITY

Theorem 1. Let $c \in \mathbb{C}^N$, with $N > 3$, be a random Rademacher or Steinhaus sequence and y satisfying equation (1). For universal sensing bases $\Phi \in \mathbb{C}^{N \times N}$ and for a universal constant $C > 0$, if $m \geq C s \log^8(N)$, then α is the unique minimizer of the ℓ_1 -minimization problem (2) with probability at least $1 - \mathcal{O}(N^{-\log^3(N)})$.

For universal sensing bases, the spread spectrum technique is thus *universal*: the recovery condition does not depend on the sparsity basis and the number of measurements needed to reconstruct sparse signals is optimal in the sense that it is reduced to the sparsity level s . The experimental study of the phase transitions of the ℓ_1 -minimization problem confirms this result (see Figure 1). The spread spectrum technique is also *efficient* as the pre-modulation only requires a sample-by-sample multiplication between x and c and fast matrix multiplication algorithms are available for several universal sensing bases such as the Fourier, Hadamard, or noiselet bases.

IV. CONCLUSION

We presented a *universal* and *efficient* compressive sampling strategy based on spread spectrum. For applications such as radio interferometry and MRI, this technique is of great interest to optimize the number of measurements needed for an accurate recovery [4], [5].

REFERENCES

- [1] Rauhut, “Compressive Sensing and Structured Random Matrices,” *Radon Series Comp. Appl. Math.*, vol. 9, pp.1-92, 2010.
- [2] Puy et al., “Universal and Efficient Compressed Sensing Strategy through Spread Spectrum Modulation,” *IEEE Trans. Signal Process.*, submitted, 2011.
- [3] Donoho et al., “Counting faces of randomly-projected polytopes when the projection radically lowers dimension,” *J. Amer. Math. Soc.*, vol. 22, pp. 1-53, 2009.
- [4] Wiaux et al., “Spread spectrum for imaging techniques in radio interferometry,” *Mon. Not. R. Astron. Soc.*, vol. 400, pp. 1029-1038, 2009.
- [5] Puy et al., “Accelerated MR imaging with spread spectrum encoding,” *ISMRM*, accepted, 2011.

On Bounds of Restricted Isometry Constants for Gaussian Random Matrices

Bubacarr Bah

School of Mathematics and Maxwell Institute
University of Edinburgh
Edinburgh, EH9 3JZ, United Kingdom
Email: b.bah@sms.ed.ac.uk

Jared Tanner

School of Mathematics and Maxwell Institute
University of Edinburgh
Edinburgh, EH9 3JZ, United Kingdom
Email: jared.tanner@ed.ac.uk

Abstract—Many of the theorems in Compressed Sensing (CS) rely upon the linear operator having suitable bounds on its restricted isometry constants (RIC). Due to the intractability of RICs for deterministic matrices, the focus has been on probabilistic bounds. This work is an effort to determine as accurate as possible bounds for Gaussian random matrices. The outcome is presented here in the form of improved RIC bounds. In addition, we present asymptotic approximations of the RIC bounds from which we deduce sampling theorems consistent with what is found in CS literature.

I. INTRODUCTION

For a matrix A of size $n \times N$, the upper and lower RICs of A , U_k and L_k respectively, over all k -sparse vectors, x , are the smallest U_k and L_k that satisfy, [2], [3],

$$(1 - L_k) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + U_k) \|x\|_2^2. \quad (1)$$

Unfortunately, computing the RICs of a matrix A is in general NP-hard. Consequently, the research community is actively computing probabilistic bounds for various random matrix ensembles. Amongst other approaches, our work aims at computing accurate RICs for Gaussian random matrices, in part as a model for i.i.d. mean zero ensembles, [1], [2], [4].

II. IMPROVED RIC BOUNDS

The first set of RIC bounds for the Gaussian ensemble were derived in [4] using a union bound over all $\binom{N}{k}$ submatrices and bounding the singular values of each submatrix using concentration of measure bounds. This was improved in [2] by similarly using a union bound over all $\binom{N}{k}$ submatrices, but with more accurate bounds on the probability density function of Wishart matrices. We achieved further improvement by grouping submatrices with overlapping support sets, say, A_K and $A_{K'}$ with $|K \cap K'| \gg 1$, for which we expect the singular values to be highly correlated.

These bounds were derived in the *linear-growth asymptotics*:

$$k/n \rightarrow \rho \text{ and } n/N \rightarrow \delta \text{ for } (\delta, \rho) \in (0, 1)^2 \text{ as } (k, n, N) \rightarrow \infty. \quad (2)$$

Our improved bounds are stated in Theorem 2.1 and their comparison to bounds of [2], \mathcal{U}^{BCT} and \mathcal{L}^{BCT} , and to empirically observed lower bounds, $U(k, n, N)$ and $L(k, n, N)$, are shown in Figure 1.

Theorem 2.1: Let A be a matrix of size $n \times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0, 1/n)$. For any fixed $\epsilon > 0$, in the linear-growth asymptotic,

$$\mathbf{P}(L_k < \mathcal{L}(\delta, \rho) + \epsilon) \rightarrow 1 \text{ and } \mathbf{P}(U_k < \mathcal{U}(\delta, \rho) + \epsilon) \rightarrow 1 \quad (3)$$

exponentially in n .

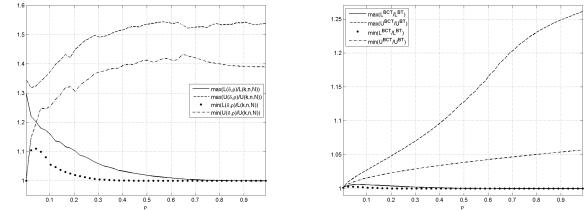


Fig. 1. As a function of ρ and taken over $\delta \in [0.05, 0.95]$ for each ρ ; Left panel: sharpness ratios, $\frac{\mathcal{U}(\delta, \rho)}{U(k, n, N)}$ and $\frac{\mathcal{L}(\delta, \rho)}{L(k, n, N)}$; Right panel: improvement ratios, $\frac{\mathcal{U}^{BCT}(\delta, \rho)}{\mathcal{U}(\delta, \rho)}$ and $\frac{\mathcal{L}^{BCT}(\delta, \rho)}{\mathcal{L}(\delta, \rho)}$.

III. ASYMPTOTICS OF RIC BOUNDS

We asymptotically approximated the RIC bounds with simpler functions for small values of δ and ρ , close to 0. Naturally three cases lend themselves to these kind of analysis: (i) fixing ρ and letting $\delta \rightarrow 0$; (ii) fixing ρ and letting $\delta \rightarrow 0$; and (iii) setting ρ as a function of δ parameterize by γ and letting both $\rho, \delta \rightarrow 0$. The third case, being the most interesting, leads to the following theorem.

Theorem 3.1: Let $\rho_\gamma(\delta) = [-\gamma \log(\delta)]^{-1}$ and let $\tilde{\mathcal{U}}(\delta, \rho_\gamma(\delta))$ and $\tilde{\mathcal{L}}(\delta, \rho_\gamma(\delta))$ be the approximations of $\mathcal{U}(\delta, \rho)$ and $\mathcal{L}(\delta, \rho)$ respectively. For a fixed γ as $\delta \rightarrow 0$,

$$\tilde{\mathcal{U}}(\delta, \rho_\gamma(\delta)) = [-2\rho \log(\delta^2 \rho^3)]^{\frac{1}{2}} - \frac{2}{3}\rho \log(\delta^2 \rho^3); \quad (4)$$

$$\tilde{\mathcal{L}}(\delta, \rho_\gamma(\delta)) = [-2\rho \log(\delta^2 \rho^3)]^{\frac{1}{2}} + \frac{2}{3}\rho \log(\delta^2 \rho^3). \quad (5)$$

Consequent to our good asymptotic approximations the following sampling theorem can be deduced from Theorem 3.1.

Corollary 3.2: Given a sensing matrix, A , of size $n \times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0, 1/n)$, in the limit as $n/N \rightarrow 0$ a sufficient condition for recovery for CS algorithms is $n \geq \gamma k \log(N/n)$ measurements with $\gamma = 36$ for l_1 -minimization, $\gamma = 93$ for IHT, $\gamma = 272$ for Subspace Pursuit and $\gamma = 365$ for CoSaMP.

REFERENCES

- [1] B. Bah and J. Tanner *Improved bounds on restricted isometry constants for gaussian matrices*, SIAM J. on Matrix Analysis, Vol. 31(5) (2010) 2882-2898.
- [2] J. D. Blanchard, C. Cartis, and J. Tanner, *Compressed Sensing: How sharp is the RIP?*, SIAM Review, Vol. 53(1) (2011) 105-125.
- [3] E. J. Candès, *The restricted isometry property and its implications for compressed sensing*, C. R. Math. Acad. Sci. Paris, Vol. 346(9-10) (2008) 589-592.
- [4] E. J. Candès and T. Tao *Decoding by linear programming*, IEEE Trans. Inform. Theory, Vol. 51(12) (2005) 4203-4215.

Towards Optimal Data Acquisition in Diffuse Optical Tomography: Analysis of Illumination Patterns

Marta M. Betcke and Simon R. Arridge

Department of Computer Science
University College London
WC1E 6BT London, UK

Email: m.betcke@ucl.ac.uk, s.arridge@cs.ucl.ac.uk

I. DIFFUSE OPTICAL TOMOGRAPHY

In diffuse optical tomography (DOT) [1], [2] the near infrared light is used to probe the optical properties of the tissue such as absorption and scattering. The value of those parameters can be related to oxygenation levels of the tissue and hence provides a functional imaging modality. Two main applications of DOT are neonatal brain imaging and breast imaging.

Transport of light through tissue is described by the Boltzmann transport equation. In DOT, the scattering is assumed to be the dominant process, and the so call diffusion approximation holds

$$\begin{aligned} -\nabla \kappa(r) \nabla \phi(r, \omega) + \mu_a(r) \phi(r, \omega) + \frac{i\omega}{c} \phi(r, \omega) &= 0, \quad r \in \Omega \quad (1) \\ \mathcal{B}^- \phi(r, \omega) &= J^-, \quad r \in \partial\Omega \\ \mathcal{B}^+ \phi(r, \omega) &= J^+, \quad r \in \partial\Omega, \end{aligned}$$

where κ and μ_a are the space dependent diffusion and absorption coefficients, respectively, Ω is the considered domain and \mathcal{B}^\pm denote appropriate boundary conditions. We assume that the boundary conditions specify a unique solution of (1).

The inward photon current, J^+ , travels through tissue undergoing scattering and absorption according to optical parameters and gives rise to the outgoing current, J^- at the boundary $\partial\Omega$, defining a map from the parameter space X to the space of measurable boundary currents Z , $\mathcal{F} : X \rightarrow Z$.

The change of the solution ϕ of (1) at the boundary Ω , due to the change of optical parameters up to first order is the Fréchet derivative of the forward map \mathcal{F} , $r \in \partial\Omega$

$$\begin{aligned} \frac{\partial \mathcal{F}(r, \omega)}{\partial \mu_a(r')} &= -\psi(r, r', \omega) \phi(r', \omega), \quad r' \in \Omega \\ \frac{\partial \mathcal{F}(r, \omega)}{\partial \kappa(r')} &= -\nabla \psi(r, r', \omega) \cdot \nabla \phi(r', \omega), \quad r' \in \Omega, \end{aligned}$$

where ψ is the solution to the equation adjoint to (1).

The inverse problem in DOT is to recover the optical parameters μ_a and κ from the boundary measurements, which amounts to inverting the map \mathcal{F} . We note that since both the direct and adjoint fields depend on the parameters to estimate, \mathcal{F} is not linear. Hence it is typically tackled with some type of Gauss Newton method, which involves solution of a linearized problem at each iteration.

II. ANALYSIS OF ILLUMINATION PATTERNS

Linearizing the forward map, amounts to computing Fréchet derivatives for different inward boundary currents J^+ (illumination patterns), which each gives rise to an outward current J^- . In praxis only limited number of currents J^+ can be applied and any only a finite set of measurements can be taken to sample J^- . In the modern DOT systems, the measurements are usually taken by a camera with a specified aperture providing highly resolved measurements of J^- .

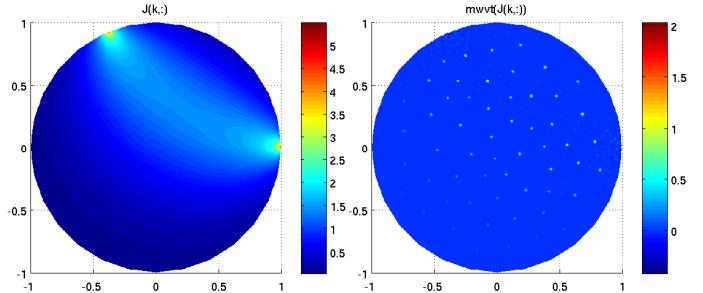


Fig. 1. Fréchet derivative (left) and its wavelet coefficients (right) for a point source and point detector on the boundary of the circle. The wavelet coefficients are plotted at the mesh nodes corresponding to the particular wavelet.

The quality and speed limiting factor of the acquisition process is application of the inward boundary currents. This leads to a problem of choice of an optimal set of illumination patterns i.e. the boundary conditions for (1), to acquire the measurable information with the least possible number of different inward boundary currents.

In this contributions we propose a method of analysis of the linearized problem in dependence of the illumination patterns based on a wavelet transform of Fréchet derivatives of the forward map \mathcal{F} . Fréchet derivative for one source illumination pattern gives rise to two rows (one for the real and one for the imaginary part) in the system matrix of the linearized problem. Figure 1 shows an example of Fréchet derivative (in fact $\Re(\partial\mathcal{F}/\partial\mu_a)$) for circular domain in \mathbb{R}^2 . Due to smoothness the Fréchet derivatives are highly compressible in wavelet basis, which hence provides an effective representation of the information. As the diffusion equation (1) is usually solved with finite elements, we apply wavelets constructed directly on the finite element mesh [3]. The compressed form of the Fréchet derivatives is then used to infer the dependencies between different illumination patterns and to arrive at an optimal set of illumination patterns. An optimal pattern set should be able to sense all the measurable wavelet coefficients on the domain but have the minimal redundancy between measurements. An example of an optimality criteria is the condition number of the system matrix for a fixed number of applied illumination patterns.

REFERENCES

- [1] S.R. Arridge, *Inverse Problems*, 15, pages 41–93, 1999
- [2] S.R. Arridge and J.C. Schotland, *Optical tomography: forward and inverse problems* Inverse Problems 25(12), pages 123010, 2009.
- [3] W. Dahmen, and R. Stevenson, *Element-by-Element Construction of Wavelets Satisfying Stability and Moment Conditions*, SIAM J. Numer. Anal. 37(1), pages 319–352, 1999.

Recent evidence of sparse coding in neural systems

Christopher J. Rozell and Mengchen Zhu
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332-0250
Email: {crozell,mczhu}@gatech.edu

I. INTRODUCTION

Neural systems are continually confronted with a wealth of high-dimensional data produced by their sensory environments. It is this stream of apparently complex data that is used to understand and navigate the world around us. While the sheer datarate of these sensory signals may appear overwhelming at first glance, the statistics of this data are highly structured and the actual information content is much lower than the high ambient dimension of the raw data itself. Indeed, as research has shown, models based on sparsity in some basis can allow significant improvements in many applications.

There exists a long history of proposing sparsity models as a way that neural systems could represent the low-dimensional structure in sensory data. The results receiving the most attention in this area are undoubtedly the work of Olshausen & Field [1]. In this pioneering experiment, the authors showed that by assuming only a neural coding model based on sparse approximation and applied to the statistics of natural images, the optimal dictionary (i.e., overcomplete basis set) is a set of localized, oriented, bandpass functions (similar to a Gabor wavelet system). This has been of great interest to the computational neuroscience community because the response properties of cells in the primary visual cortex are often characterized with (qualitatively) similar functions, lending credence to the hypothesis that the neural system may be optimized to represent information in a sparse code. In a fruitful demonstration of the power of interdisciplinary work, the signal processing community has since made use of these ideas from neuroscience (as well as adding their own significant advances), and it is now routine to think of using unsupervised learning to determine optimal dictionaries for representing new signal classes.

II. RECENT EVIDENCE

Unfortunately, the results from [1] represent only circumstantial evidence for the sparse coding hypothesis in neuroscience. Despite 15 years passing since the publication of these results, we still lack direct confirmation that sparse approximation is a significant coding principle in neural systems (due at least in part to technological barriers that are just being advanced by neurophysiologists). In the interim, there has been a growing body of (also circumstantial) work lending more evidence to support this hypothesis. Here we will survey a collection of questions that have been asked and at least partially answered in the neuroscience community about the validity of the sparse coding hypothesis, with the aim of trying to induce more interaction between research communities.

- *Do the learned dictionaries quantitatively match the measured response properties of cells in visual cortex?* While the qualitative match is undeniable, the results of Olshausen & Field (and related results by Bell & Sejnowski using ICA [2]) actually fail to match the measured response properties in some quantitative ways. However, recent refinements of this model have shown that the fits can be made much more quantitative through a

combination of increasing the overcompleteness of the dictionary (biology is estimated to be 25-50 times overcomplete), and using an inference scheme that induces more “hard sparseness” (e.g., more coefficients that are exactly zero) [3]. This lends an interesting aspect to the discussion about heuristic greedy algorithms that achieve many zeros and convex relaxations that may have difficulty driving coefficients to zero in practice.

- *Do neural responses actually look sparse?* Traditionally, many neurophysiology experiments were performed with artificial stimuli (e.g., sinusoidal gratings) where it would be difficult to answer this question. In recent experiments using natural stimuli, there is accumulating evidence of response patterns appearing more sparse than predicted by classic models (e.g. [4]).
- *Could neural systems solve the non-smooth optimizations we use in sparse approximation?* A plethora of algorithms for ℓ_1 minimization have appeared in the literature recently, for obvious reasons, all of them are designed to operate on a digital computer with a centralized CPU. We have recently introduced a dynamical system that provably solves these optimizations using computational primitives appropriate for neural architectures [5].
- *What about all the other nonlinear response properties I hear about in neural systems?* It is true that a broad variety of nonlinear response properties reported by physiologist. For example, in the visual cortex there have been many effects reported in the literature known as non-classical receptive field effects (nCRFs), where essentially the cell responds in a nonlinear way to a modulatory stimulus that would not otherwise drive the cell. These effects have all been modeled individually, or described with a collection of ideas such as adaptive gain control or predictive coding. We have recently performed a host of simulated physiology experiments on a sparse coding model and shown that nearly all of these nonlinear effects appear as simply emergent effects from this single coding rule (both at the level of individual cells and cell populations) [6].

REFERENCES

- [1] B. Olshausen and D. Field, “Sparse coding with an overcomplete basis set: A strategy employed by V1?” *Vision research*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [2] A. Bell and T. Sejnowski, “The “independent components” of natural scenes are edge filters,” *Vision research*, vol. 37, no. 23, p. 3327, 1997.
- [3] M. Rehn and F. Sommer, “A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields,” *Journal of Comp. Neuro.*, vol. 22, no. 2, pp. 135–146, 2007.
- [4] W. Vinje and J. Gallant, “Sparse coding and decorrelation in primary visual cortex during natural vision,” *Science*, vol. 287, no. 5456, 2000.
- [5] C. Rozell, D. Johnson, R. Baraniuk, and B. Olshausen, “Sparse coding via thresholding and local competition in neural circuits,” *Neural Computation*, vol. 20, no. 10, pp. 2526–2563, 2008.
- [6] M. Zhu and C. J. Rozell, “Population characteristics and interpretations of ncrf effects emerging from sparse coding,” in *Computational and Systems Neuroscience (Cosyne) Meeting*, 2011.

Sparse Detection in the Chirplet Transform

Fabien Millioz

School of Electrical and Electronics
University of Edinburgh
Edinburgh, United Kingdom
Email: fmillioz@ed.ac.uk

Mike Davies

School of Electrical and Electronics
University of Edinburgh
Edinburgh, United Kingdom
Email: mike.davies@ed.ac.uk

I. INTRODUCTION

This paper first investigates the choice of parameters of the chirplet transform, leading to good detection properties. Then we propose a cheap iterative detection algorithm with a single chirplet transform computation, which avoids over-detection due to the redundancy of the chirplet transform.

II. CHIRPLET TRANSFORM

The discrete chirplet transform [1] $C[n, k, d]$ of a signal $x[m]$ is determined by

$$C[n, k, d] = \sum_{m=-M}^{+M} x[n+m] \phi_M[m] e^{-j2\pi\frac{1}{2L}d_{max}m^2} e^{-j2\pi m\frac{k}{K}}, \quad (1)$$

with n, k are the time and frequency indices respectively, K the number of frequencies. There are $2L+1$ different chirprates, ranging from $-Ld_{max}$ to $+Ld_{max}$. The smoothing window $\phi_M[m]$ has $2M+1$ points, normalised such that its maximal value is 1. This transform may be viewed as a collection of *chirped* Short Time Fourier Transforms.

The stationary phase approximation [2] leads to an approximation of the chirplet transform of a chirp of chirprate d_0 , centred on time index n and frequency $(k_0 + d_0 n)K$,

$$|C[n, k, d]| \approx A \sqrt{\frac{1}{|d_0 - d|}} \Phi_{|d_0 - d|MK} [k - (k_0 + d_0 n)K] \quad (2)$$

However, when the chirp's chirprate d_0 is equal to the analysing chirplet's chirprate d , the chirplet transform is equivalent to a Fourier transform: $|C[n, k, d_0]| \approx A|\Phi_M[k - (k_0 + d_0 n)K]|$ (3)

where $\Phi_M[k]$ is the discrete Fourier transform of $\phi_M[m]$. Note that the maximum of $\Phi_M[k]$ is $\Phi_M[0] = \sum \phi_M[m]$.

Figure 1 illustrates the validity of these approximations. The blue and green curves are the values approximated by (3) and (2) respectively, the red curve is the value of $|C[0, 0, d]|$.

The transition from approximation (2) to (3) corresponds to the value of $|d_0 - d| = \frac{1}{\Phi_M[0]^2} = \Delta_{d_0}$ such that these two approximations are equal.

The energy of the chirplet coefficient is maximal if $|d_0 - d| < \Delta_{d_0}$. Consequently, by choosing the parameters of the chirplet transform such that the chirprate step is less than $2\Delta_{d_0}$, we are assured to get a good detection for any chirp.

III. MAXIMUM CHIRPLET TRANSFORM

To simplify the detection problem, we define the Maximum Chirplet Transform (MCT) $D[n, k]$ at a given time-frequency point $[n, k]$, containing all maxima of the square modulus of the chirplet coefficients along the chirprates

$$D[n, k] = \max_d |C[n, k, d]|^2. \quad (4)$$

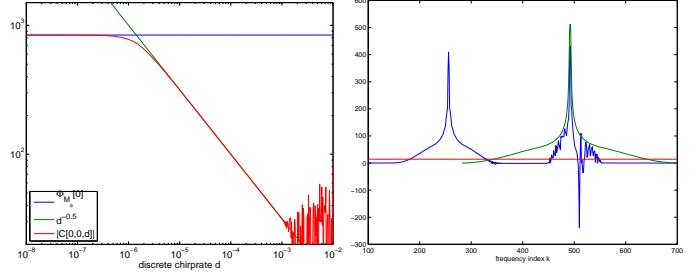


Fig. 1. Maximum value of $|C[0, 0, d]|$ Fig. 2. Illustration of a detection step. of a wave and its approximations.

We choose to use a detection based on a Neyman-Pearson approach, which provides a detection threshold. However, this swrwxrion provides a high number of detection coefficients in the MCT due to the redundancy of the chirplet transform.

Considering all possible spectral windows for any chirp, we can define a *upper-bound spectral window*, such that for any analysed chirp, the upper-bound spectral window normalised and centred on the highest MCT value is greater or equal than all MCT spectrum points.

We propose an iterative detection, using a single chirplet transform computation: at each iteration, the upper-bound spectral window centred on the point of highest magnitude is subtracted from the MCT spectrum, and a new iteration is run over this new spectrum. The iterations stop when all spectrum coefficients are below the detection threshold t . Contrary to the Matching Pursuit approach, the spectrum does not need to be re-computed at each iteration.

Figure 2 illustrates a first iteration of such a detection. In blue, the MCT spectrum, in green the upper-bound window, and in red the detection threshold t .

IV. CONCLUSION

The minimal chirprate step of the chirplet transform leading to good detection properties have been investigated in this paper. A detection method, based on iterative subtraction of the Maximum Chirplet transform spectrum has been proposed, selecting only a few chirplet coefficients in spite of the redundancy of the chirplet transform.

REFERENCES

- [1] S. Mann and S. Haykin, "The chirplet transform: A generalization of gabor's logon transform," Calgary, Canada, Jun. 1991.
- [2] E. Chassande-Mottin and P. Flandrin, "On the stationary phase approximation of chirp spectra," in Proc. of the IEEE Int. Symp. on Time-Frequency and Time-Scale Analysis, Pittsburgh (US), 1998, pp. 117–120.

Riemannian optimization for rank minimization problems

Bart Vandereycken

Seminar for Applied Mathematics, ETH Zürich, Switzerland
Email: bart.vandereycken@sam.math.ethz.ch

Abstract—We present an acceleration strategy based on smooth Riemannian optimization for rank minimization problems. By assuming that the rank of the minimizer is known, the original rank minimization can be formulated as a smooth optimization problem on the manifold of fixed-rank matrices. We show that these low-dimensional problems can be solved very efficiently. We apply our framework to large-scale Lyapunov equations and low-rank matrix completion, and compare to the state of the art.

I. INTRODUCTION

Let $\epsilon \geq 0$ be a tolerance and $\mathcal{L} : R^{n \times m} \rightarrow R^{n \times m}$ a linear operator. Then, given $C \in R^{n \times m}$, the noisy version of the rank minimization problem [1]

$$\begin{aligned} & \text{minimize} && \text{rank}(X) \\ & \text{subject to} && X \in R^{n \times m}, \|\mathcal{L}(X) - C\|_F < \epsilon \end{aligned} \quad (1)$$

can be turned into a smooth optimization problem,

$$\begin{aligned} & \text{minimize} && f(X) := \|\mathcal{L}(X) - C\|_F^2, \\ & \text{subject to} && X \in \mathcal{M}_k := \{X \in R^{n \times m} : \text{rank}(X) = k\}, \end{aligned} \quad (2)$$

provided that the rank k of the optimizer in (1) is known.

It is well known that the set \mathcal{M}_k is a smooth submanifold embedded in $R^{n \times m}$. Since the objective function is smooth also, problem (2) is a smooth Riemannian optimization problem, which turns out to be significantly easier to solve than (1).

We present numerical algorithms to solve (2) based on the framework of retraction-based optimization in [2]. The numerical algorithms heavily exploit the smoothness of \mathcal{M}_k and are generalizations of classical optimization algorithms on Euclidean space, like Newton's method and non-linear CG.

We consider two different applications: large-scale Lyapunov matrix equations for PDEs and low-rank matrix completion.

II. LYAPUNOV MATRIX EQUATION

Solving the Lyapunov equation coincides with solving (2) for

$$\mathcal{L} : R^{n \times n} \rightarrow R^{n \times n}, X \mapsto AXM^T + MXA^T$$

where $A \succ 0$ and $M \succ 0$ are given p.s.d. matrices. In case of large-scale applications [3], matrix A corresponds to a discretized PDE and M is a mass matrix. For some important PDEs, the solution $X := \mathcal{L}^{-1}(C)$ admits a very good low-rank approximation.

By using $f(X) = \text{tr}(AXM) - \text{tr}(XC)$ as objective function, it is shown in [4] that one actually minimizes the energy norm of the error. Now, we can apply Riemannian optimization to obtain a low-rank approximation to the solution of Lyapunov equation.

Since the discretized PDEs are very large and highly ill-conditioned, the optimization algorithms need to be preconditioned. By exploiting the low-rank structure of the tangent spaces and the specific form of the \mathcal{L} , it is possible to efficiently precondition the Riemannian Hessian of $f(X)$ by \mathcal{L} , restricted to the tangent space at X . In practice, this results in a mesh-independent convergence for the truncated CG-Newton method.

In the table below, the results of a numerical experiment are listed to solve large-scale Lyapunov equations for a relative residual of 10^{-6} . The matrix A corresponds to a discretized 2D Poisson equation, and $M = I$. A trust-region version of Newton's method was preconditioned as explained above.

size n	500^2	1000^2	1500^2
time (s.)	40	175	443
rank(X)	12	12	12

III. LOW-RANK MATRIX COMPLETION

In the context of low-rank matrix completion, the objective function in (2) becomes $f(X) := \|P_\Omega(X - A)\|_F^2$ where A is a given (numerically) low-rank matrix only known on a subset Ω through

$$P_\Omega : R^{n \times m} \rightarrow R^{n \times m}, X_{ij} \mapsto \begin{cases} X_{ij}, & (i, j) \in \Omega, \\ 0 & \text{otherwise.} \end{cases}$$

In case the exact recovery of (1) can be formulated in terms of the restricted isometry property [1], the operator P_Ω can be assumed to be well-conditioned on \mathcal{M}_k . In this case, preconditioning is unnecessary for large-scale problems.

Below, the experimental results are shown for the completion of an $n \times n$ random matrix (i.i.d. Gaussian) of rank 50 with $|\Omega| = m$. We compared our nonlinear CG and Newton methods with two well-known implementations, SVT [5] and inexact ALM [6]. The relative tolerance was 10^{-4} . Clearly, exploiting the knowledge of the rank of the minimizer by optimizing on a manifold can greatly reduce the time needed to solve (2).

n	m/n^2	CG	TR Newton	SVT	in. ALM
		(seconds/nb. of iterations)			
5000	0.10	40/18	99/27	930/107	320/45
10 000	0.05	106/20	251/30	2048/110	1230/70

REFERENCES

- [1] B. Recht, M. Fazel, and P. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.
- [2] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton University Press, 2008.
- [3] P. Benner, V. Mehrmann, and D. Sorensen, Eds., *Dimension Reduction of Large-Scale Systems*. Springer-Verlag, 2005.
- [4] B. Vandereycken and S. Vandewalle, “A Riemannian optimization approach for computing low-rank solutions of Lyapunov equations,” *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 5, pp. 2553–2579, 2010.
- [5] J.-F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [6] Z. Lin, M. Chen, L. Wu, and Y. Ma, “The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices,” University of Illinois, Urbana, Department of Electrical and Computer Engineering, UIUC Technical Report UILU-ENG-09-2215, 2009.

The degrees of freedom of the Lasso in underdetermined linear regression models

Maher Kachour and Jalal Fadili
GREYC, ENSICAEN

Christophe Chesneau
LMNO Université de Caen

Charles Dossal
IMB Université Bordeaux 1

Gabriel Peyré
Ceremade Université Paris-Dauphine

Abstract—In this paper, we investigate the degrees of freedom (df) of penalized ℓ_1 minimization (also known as the Lasso) for an underdetermined linear regression model. We show that under a suitable condition on the design matrix, the number of nonzero coefficients of the Lasso solution is an unbiased estimate for the degrees of freedom. An effective estimator of the number of degrees of freedom may have several applications including an objectively guided choice of the regularization parameter in the Lasso through the SURE framework.

Index Terms—Lasso, degrees of freedom, SURE.

I. INTRODUCTION

Consider the following linear regression model

$$y = Ax^0 + \varepsilon, \quad (1)$$

where $y \in \mathbb{R}^n$ is the response vector, $A = (a_1, \dots, a_p) \in \mathbb{R}^{n \times p}$ is a deterministic design matrix with $n < p$, $x^0 \in \mathbb{R}^p$ is the unknown regression vector, and $\varepsilon \in \mathbb{R}^n$ is the noise vector whose entries are i.i.d. $\mathcal{N}(0, \sigma^2)$. The goal is to solve (1) when the solution is assumed to be sparse. Towards this goal, a now popular estimator is the Lasso [4]. The Lasso estimate amounts to solving the following convex problem

$$\min_{x \in \mathbb{R}^p} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1, \quad (2)$$

where $\lambda > 0$ is the regularization or tuning parameter. In the last years, there has been a huge amount of work where efforts have focused on investigating the theoretical guarantees of the Lasso as a sparse recovery procedure from noisy measurements in the underdetermined case $n < p$.

Degrees of freedom df is a familiar phrase in statistics. In over-determined linear regression df is the number of estimated predictors. Degrees of freedom is often used to quantify the model complexity of a statistical modeling procedure (e.g. it corresponds to the penalty term of model selection criteria such as AIC and BIC). However, generally speaking, there is no exact correspondence between the degrees of freedom df and the number of parameters in the model. On the other hand, the degrees of freedom plays an important role for an objective selection of the tuning parameter.

Let us denote by \hat{x} any estimator of x^0 which depends on y and let $\hat{y} = A\hat{x}$. Since $y \sim \mathcal{N}(Ax^0, \sigma^2 \mathbf{I})$, according to [2], the degrees of freedom of \hat{y} is

$$df(\hat{y}) = \sum_{i=1}^n \frac{\text{cov}(\hat{y}_i, y_i)}{\sigma^2}. \quad (3)$$

If \hat{y} is almost differentiable, Stein's lemma [3] yields the following unbiased estimator of df

$$\hat{df}(\hat{y}) = \text{div } \hat{y} = \sum_{i=1}^n \frac{\partial \hat{y}_i}{\partial y_i}. \quad (4)$$

Contributions Let $\hat{\mu}_\lambda = \hat{\mu}_\lambda(y) = A\hat{x}_\lambda(y)$ be the Lasso response vector, where $\hat{x}_\lambda(y)$ is a solution of the Lasso problem (2). In the overdetermined case, i.e. $n > p$, $\hat{x}_\lambda(y)$ is unique, and the authors in

[5] showed that for any given λ the number of non-zero coefficients of \hat{x}_λ is an unbiased estimator of the degrees of freedom of the Lasso. Though their proof contains a gap. The contribution of this paper is to extend their result to the underdetermined case where the Lasso solution is not unique. To ensure the uniqueness of the solution, we introduce the condition (UC) on the design matrix.

II. MAIN RESULTS

Let $z \in \mathbb{R}^p$, $S \subseteq \{1, 2, \dots, p\}$ and $|S|$ its cardinality. We denote by A_S the submatrix $A_S = [\dots, a_j, \dots]_{j \in S}$, where a_j is the j th column of A and the pseudo-inverse $(A_S^t A_S)^{-1} A_S^t$ of A_S is denoted A_S^+ . Let z_j be the j th component of z . Similarly, we define $z_S = (\dots, z_j, \dots)_{j \in S}$ for z . Let $\text{supp}(z) = \{j : z_j \neq 0\}$ be the support or the active set of z .

Definition 1 (Condition (UC) [1]): A matrix A satisfies condition (UC) if, for all subsets $I \subset \{1, \dots, p\}$ with $|I| \leq n$, such that $(a_i)_{i \in I}$ are linearly independent, for all indices $j \notin I$ and all vectors $V \in \{-1, 1\}^{|I|}$,

$$|\langle a_j, (A_I^+)^t V \rangle| \neq 1. \quad (5)$$

Theorem 1: Suppose that A satisfies condition (UC). For any $y \in \mathbb{R}^n$, there exists a finite set of values λ , denoted by $\{\lambda_m\}$, for which we have

$$\max_{j \notin I} |\langle a_j, y - A\hat{x}_\lambda(y) \rangle| = \lambda, \quad (6)$$

where $I = \text{supp}(\hat{x}(\lambda))$ and $\hat{x}_\lambda(y)$ is the solution of the Lasso. Furthermore, if $\lambda \in]0, \|A^t y\|_\infty[\setminus \{\lambda_m\}$, then

$$\max_{j \notin I} |\langle a_j, y - A\hat{x}_\lambda(y) \rangle| < \lambda. \quad (7)$$

Theorem 2: Suppose that A satisfies condition (UC). For any $y \in \mathbb{R}^n$, and all values of λ for which (7) is satisfied, we have

- The Lasso response $\hat{\mu}_\lambda(y) = A\hat{x}_\lambda(y)$ is a uniformly Lipschitz function of y ;
- The support and vector sign of the Lasso solution are locally constant with respect to y , and consequently

$$\text{div } \hat{\mu}_\lambda(y) = |\text{supp}(\hat{x}_\lambda(y))|. \quad (8)$$

That is, using Stein's lemma [3] and the divergence formula (8), the number of non-zero coefficients of \hat{x}_λ is an unbiased estimator of the degrees of freedom of the Lasso.

REFERENCES

- [1] Dossal, C (2007). A necessary and sufficient condition for exact recovery by ℓ_1 minimization. Technical report, HAL-00164738:1.
- [2] Efron, B. (1981). How biased is the apparent error rate of a prediction rule. J. Amer. Statist. Assoc. vol. 81 pp. 461?470.
- [3] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. Ann. Statist. 9 1135-1151.
- [4] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. J. Roy. Statist. Soc. Ser. B 58(1) 267-288.
- [5] Zou, H., Hastie, T. and Tibshirani, R. (2007). On the "degrees of freedom" of the Lasso. Ann. Statist. Vol. 35, No. 5. 2173-2192.

Guaranteed recovery of a low-rank and joint-sparse matrix from incomplete and noisy measurements

Mohammad Golbabaei and Pierre Vandergheynst
 Signal Processing Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
 Email: {mohammad.golbabaei, pierre.vandergheynst}@epfl.ch

I. PROBLEM STATEMENT

Suppose you are given a matrix $X \in \mathbb{R}^{n_1 \times n_2}$ with rank $r \ll \min(n_1, n_2)$. Moreover, assume this matrix has sparse nonzero elements so that, due to the column-wise dependencies, they are all supported on $k \ll n_1$ number of rows (it can also be column-wise supported). This matrix wont have many degrees of freedom; If one knows the position of those k nonzero rows, the corresponding submatrix contains only $(k + n_2 - r)r$ degrees of freedom.

Provided by the enormous developments in areas of compressed sensing and low rank-matrix recovery [1][2][3][4], one may wonder if it is possible to acquire the whole matrix elements from a very few number of non-adaptive linear measurements. In this regard, three questions immediately follow; what should be those measurements? How to design a computationally tractable algorithm to recover this matrix from those possibly noisy measurements? And finally, how to evaluate the performance i.e., how many measurements do we need to recover exact low-rank and sparse matrix, and does the algorithm performs stable with respect to matrices that are approximately low-rank or not exactly joint-sparse but *compressible*? This paper attempts to answer the questions above.

II. PRIOR ARTS

Recently a few papers consider *rank awareness* in data joint-recovery from multiple measurement vectors (MMV) [5] [6]. More precisely, sparse MMV inverse problem (also known as simultaneous sparse approximation), focuses on recovering a joint-sparse matrix X from a set of measurements $Y \in \mathbb{R}^{m \times n_2}$ acquired as $Y = AX$. There, $A \in \mathbb{R}^{m \times n_2}$ is the measurement matrix that is unique for compressive sampling signals of all the n_2 channels (columns of X). Davis *et al.* [5] proposed a specific rank-aware *greedy* algorithm, that in case of using a random i.i.d. Gaussian A , is able to recover (with high probability) an *exact* k -joint-sparse and rank- r X from its noiseless MMV, if the total number of measurements scales as,

$$m = n_2 \bar{m} \gtrsim \mathcal{O}(n_2 k (\log n_1 / r + 1)). \quad (1)$$

III. ORIGINALITY OF OUR WORK

Our work contrasts with prior arts in three main aspects:

1- Let us define the linear map $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ and model our sampling mechanism by $y = \mathcal{A}(X) + z$, for a noise vector $z \in \mathbb{R}^m$. As we can see, this measurement scheme is able to model more general cases than a uniform sampling matrix for all the channels e.g., in *distributed* compressed sensing scenarios, each channel can be sampled by an independent measurement matrix (rather than a unique one), or even in non-distributed cases where the sampling matrix is designed so that each measurement reflects a global average behavior of the whole matrix rather than a local specific channel.

2- Our recovery algorithm is different and is based on the following convex minimization,

$$\begin{aligned} & \arg \min_X \|X\|_{2,1} + \lambda \|X\|_* \\ & \text{subject to } \|y - \mathcal{A}(X)\|_2 \leq \epsilon. \end{aligned} \quad (2)$$

The $l_{2,1}$ mixed-norm is defined as $\|X\|_{2,1} := \sum_i (\sum_j X_{i,j}^2)^{1/2}$ and the nuclear norm $\|X\|_*$ is the sum of the singular values of X .

3- Our performance analysis, guarantees *stability* of our recovery approach against noisy measurements, non-exact sparse and approximately low-rank data matrices. We prove that, if our measurement system satisfies a *specific* restricted isometry property (RIP), the solution of (2), stably recovers *all* joint-sparse and low-rank matrices. In particular, we show that, for certain random measurement schemes, the number of measurements m sufficient for stable recovery scales as,

$$m \geq \mathcal{O}\left(k(r + \log(n_1/k)) + n_2 r\right). \quad (3)$$

Regarding rank of the data matrix, our bound is of a different nature than (1) i.e., the lower the rank, less measurements are required. Indeed, in many multichannel signal applications, where (due to the structure behind) a huge data matrix turns out to have a low-rank ($r \ll k \ll n_2$), our approach outperforms those in the state-of-the-art, reflecting the importance of a good design for the measurements \mathcal{A} together with the recovery approach benefiting those structures (i.e., joint-sparse and low-rank).

In the rest of this paper, we develop an algorithm to solve (2) using proximal splitting methods [7]. A number of simulations on synthetic data as well as an interesting important application in Hyperspectral imaging, demonstrate a massive saving of the number of measurements required to recover data, compared to the existing methods.

REFERENCES

- [1] D.L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] E. J. Candes, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Pure Appl. Math.*, vol. 59, pp. 1207–1223, 2005.
- [3] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.
- [4] E. J . Candes and Y. Plan, “Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements..,” *IEEE Transactions on Information Theory*, 2009.
- [5] Mike E. Davies and Yonina C. Eldar, “Rank awareness in joint sparse recovery,” *CoRR*, vol. abs/1004.4529, 2010.
- [6] Jongmin Kim, Ok Kyun Lee, and Jong Chul Ye, “Compressive music: A missing link between compressive sensing and array signal processing,” *CoRR*, vol. abs/1004.4398, 2010.
- [7] P. L. Combettes and J. C. Pesquet, “Proximal splitting methods in signal processing,” in: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer-Verlag, vol. 49, pp. 185–212, 2011.

Message-Passing Estimation from Quantized Samples

Ulugbek Kamilov
 École Polytechnique Fédérale de Lausanne
 Email: ulugbek.kamilov@epfl.ch

Vivek K Goyal
 Massachusetts Institute of Technology
 Email: vgoyal@mit.edu

Sundeep Rangan
 Polytechnic Institute of New York University
 Email: srangan@poly.edu

Abstract—Recently, relaxed belief propagation and approximate message passing have been extended to apply to problems with general separable output channels rather than only to problems with additive Gaussian noise. We apply these to estimation of signals from quantized samples with minimum mean-squared error. This provides a remarkably effective estimation technique in three settings: an oversampled dense signal; an undersampled sparse signal; and any signal when the quantizer is not regular. The error performance can be accurately predicted and tracked through the state evolution formalism. We use state evolution to optimize quantizers and discuss several empirical properties of the optimal quantizers.

I. OVERVIEW

Estimation of a signal from quantized samples arises both from the discretization in digital acquisition devices and the quantization performed for compression. An example in which treating quantization with care is warranted is analog-to-digital conversion, where the advantage from oversampling is increased by replacing conventional linear estimation with nonlinear estimation procedures [1]–[3]. Sophisticated approaches are also helpful when using sparsity or compressibility to reconstruct an undersampled signal [4]–[6].

A rather general abstraction is to consider $y = Q(Ax)$, where $x \in \mathbb{R}^n$ is a signal of interest, $A \in \mathbb{R}^{m \times n}$ is a linear mixing matrix, and $Q : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is a quantizer. We will limit our attention here to scalar quantizers, meaning that Q is separable into m scalar quantizers $q_i : \mathbb{R} \rightarrow \mathcal{Y} \subset \mathbb{R}$ with \mathcal{Y} countable.

Implementation of belief propagation (BP) for estimation of a continuous-valued quantity requires discretization of densities; this is inexact and leads to high computational complexity. To handle quantization without any heuristic additive noise model and with low complexity, we use a recently-developed Gaussian-approximated BP algorithm, called *relaxed belief propagation* [7], [8], which extends earlier methods [9], [10] to nonlinear output channels.

Our first main contribution is to demonstrate that relaxed BP provides significantly-improved performance over traditional methods for estimating from quantized samples. Gaussian approximations of BP have previously been shown to be effective in a range of applications; the extension to general output channels [7], [8] is essential to our application.

Our second main contribution concerns the quantizer design. When quantizer outputs are used as an input to a nonlinear estimation algorithm, minimizing the mean-squared error (MSE) between quantizer input and output is not necessarily equivalent to minimizing the MSE of the final reconstruction. We use the fact that the MSE under large random mixing matrices A can be predicted accurately from a set of simple state evolution (SE) equations [8], [11]. Then, by modeling the quantizer as a part of the measurement channel, we use the SE formalism to optimize the quantizer to asymptotically minimize distortions after the reconstruction by relaxed BP.

II. SIMULATION EXAMPLE

Form A from i.i.d. Gaussian random variables, i.e., $A_{ai} \sim \mathcal{N}(0, 1/m)$; and assume i.i.d. Gaussian noise with variance $\sigma^2 =$

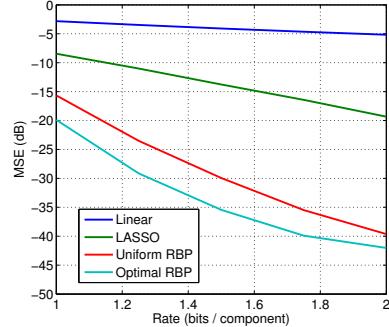


Fig. 1: Performance comparison.

10^{-5} perturbs measurements before quantization. The signal \mathbf{x} is generated with i.i.d. elements from the Gauss-Bernoulli distribution

$$\mathbf{x}_i \sim \begin{cases} \mathcal{N}(0, 10), & \text{with probability 0.1;} \\ 0, & \text{with probability 0.9.} \end{cases}$$

Figure 1 presents a comparison of reconstruction distortions and confirms (a) the advantage of relaxed BP estimation; and (b) the advantage of optimizing quantizers using the SE equations. The quantization rate is varied from 1 to 2 bits per component of \mathbf{x} , and for each quantization rate, we optimize quantizers for the MSE of the *measurements* (labeled “Uniform RBP”) and for MSE of the *reconstruction via relaxed BP* (labeled “Optimal RBP”). The figure also plots the MSE for linear MMSE estimation and lasso, both assuming the uniform quantizer that minimizes MSE of the measurements. Lasso performance was predicted by state evolution equations in [8], with the regularization parameter optimized. Relaxed BP offers dramatically better performance—more than 10 dB improvement at low rates. At higher rates, relaxed BP performance saturates due to the Gaussian noise at the quantizer input. Furthermore, optimizing the quantizer for the relaxed BP reconstruction improves performance by more than 4 dB for many rates. See also [12].

REFERENCES

- [1] N. T. Thao and M. Vetterli, *IEEE T. Signal Process.*, 42:519–531, 1994.
- [2] S. Rangan and V. K. Goyal, *IEEE T. Inform. Theory*, 47:457–464, 2001.
- [3] A. M. Powell, *Appl. Comput. Harm. Anal.*, 29:251–271, 2010.
- [4] A. Zymnis, S. Boyd, and E. Candès, *IEEE Signal Process. Lett.*, 17:149–152, 2010.
- [5] L. Jacques, D. K. Hammond, and J. M. Fadili, *IEEE T. Inform. Theory*, 57:559–571, 2011.
- [6] J. N. Laska, *et al.*, *Appl. Comput. Harm. Anal.*, 30, 2011.
- [7] S. Rangan, *Proc. Conf. on Inform. Sci. & Sys.*, Mar. 2010.
- [8] ———, arXiv:1010.5141v1 [cs.IT], Oct. 2010.
- [9] D. Guo and C.-C. Wang, *Proc. IEEE Inform. Theory Workshop*, Oct. 2006, pp. 194–198.
- [10] M. Bayati and A. Montanari, *IEEE T. Inform. Theory*, 57:764–785, 2011.
- [11] D. Guo and C.-C. Wang, *Proc. IEEE Int. Symp. Inform. Theory*, Jun. 2007, pp. 946–950.
- [12] U. Kamilov *et al.*, arXiv:1102.4652v1 [cs.IT], Feb. 2011.

Ambiguity Sparse Processes

Sofia Olhede, UCL

Abstract—A class of nonstationary time series is proposed of series that exhibit sparsity. The class is specified in the ambiguity domain, and is defined for the sampling of the observed data. Unlike traditional classes of nonstationary processes there is no implicit definition of local stability uniformly across the frequencies of the signal, and no tacit reference to a given representation, such as a class of wavelets or a given short-time Fourier transform. The properties of the class are derived and an Empirical Bayes method of estimation is introduced that is able to estimate the covariance of aggregations of inhomogeneous signals.

I. INTRODUCTION

A traditional assumption necessary for much of time series methodology to work is that of stationarity. Unfortunately many signals observed in real applications do not satisfy this constraint, and so since the 1940s theory and methods for nonstationary processes have been developed, see the discussions in [1].

In statistics there has been a focus on methods using the short-time Fourier transform [2], [3], or the wavelet transform [4]. A problem with such methods is that the analysis becomes strongly representation dependent. We think that it is natural to define a nonstationary process in terms of the given sampling, and to use the full observed bandwidth of the observations, arguing that the Ambiguity Function (AF) [5] of the process should be sparse. This is in contrast with both local Fourier methods and underspread processes [6], which automatically smooth out most of the ambiguity function.

Given the statistical properties of an ambiguity sparse process, an Empirical Bayes estimation procedure is suitable for the estimation of its second order structure, and this can be represented using any chosen (bilinear) time-frequency representation.

II. AN AMBIGUITY SPARSE PROCESS

Assume that $\{X_n\}$ is a zero-mean harmonizable process sampled at $t_n = n\Delta t$, with sampling period Δt , where a sample size N is collected and let $\{Z_n\}$ be the analytic signal constructed from $\{X_n\}$. Assume that $M_\tau(t_n) = E(Z_n Z_{n-\tau}^*)$, is the auto-covariance of the analytic signal and define the ambiguity function to be

$$A_\tau(\nu) = \Delta t \sum_{n=-\infty}^{\infty} M_\tau(t_n) e^{-2i\pi\nu t_n}. \quad (1)$$

An important characteristic of the ambiguity function is its support in the (τ, ν) plane as this characterises the process under analysis.

Definition 1: Ambiguity Sparse Process

A second order real-valued time series $\{X_n = X(n\Delta t)\}$ is denoted *Ambiguity Sparse* at sampling $N, \Delta t$ if its AF can be represented for $K \in \mathbb{N}$ in the form $A_\tau(\nu) = \sum_{k=1}^K A_\tau^{(k)}(\nu)$,

$$A_\tau^{(k)}(\nu) = \frac{\mathcal{B}^{(k)}(\nu, \tau/N)}{\left[\Delta t^2 (\nu - \nu_0^{(k)})^2 + (\tau - \tau_0^{(k)})^2 / N^2 \right]^{\delta^{(k)}}, \quad (2)}$$

with $\mathcal{B}^{(k)}(\nu, u)$ a smooth function near $(\nu_0^{(k)}, \tau_0^{(k)})$, taking a non-zero value at this point, with $\frac{3}{4} > \delta^{(k)} > \frac{1}{4}$.

III. INFERENCE

With the model of Eqn. (2), the sample ambiguity function is concentrated near the points $\{(\nu_0^{(k)}, \tau_0^{(k)})\}$, as would be expected, and is small in magnitude away from these points. To avoid excessive smoothing, an Empirical Bayes method [7] is used to shrink the observed ambiguity function and determine the estimated ambiguity function. This can then be transformed back into a representation of the autocovariance sequence of the process, that corresponds to a variable bandwidth smoothing of the sample autocovariance sequence, see also [8]. The estimated autocovariance sequence can be corrected into a valid covariance matrix, and any chosen bilinear representation [9] that is suitable to represent the process can be calculated. The performance of the method depends on the degree of sparsity of the ambiguity function and can be interpreted as a composite likelihood method.

IV. DISCUSSION

Time-frequency representations of nonstationary time series is a very well researched area. Because of the variety of nonstationary signals many different forms of representation of the covariance have been proposed, and it is possible to find classes of signals [1] so that any given representation is unsuitable. For this reason the basic object to estimate is the autocovariance of the analytic signal of the process, and then various representations can be formed that illustrate different characteristics of the process. It is important to avoid placing uniform smoothness assumptions on the evolution of the autocovariance sequence, especially in exploratory data analysis. We complement the interesting work in [10], as our methods are applicable to a large class of harmonizable processes that do not necessarily have sparse Wigner distributions. In our work the estimation of the second order structure is separated from its representation, and automatically adapted to the variable smoothness of the signal. This means that uniform and heavy-handed smoothing can be avoided, and the full bandwidth of the signal understood.

REFERENCES

- [1] R. M. Loynes, “On the concept of the spectrum for non-stationary processes,” *J. Roy. Stat. Soc. B*, vol. 30, pp. 1–30, 1968.
- [2] M. B. Priestley, “Evolutionary spectra and non-stationary processes,” *Journal of the Royal Statistical Society, B*, vol. 27, pp. 204–237, 1965.
- [3] R. Dahlhaus, “Fitting time series models to nonstationary processes,” *The Annals of Statistics*, vol. 25, pp. 1–37, 1997.
- [4] G. Nason, *Wavelet Methods in Statistics with R*. Berlin: Springer, 2008.
- [5] S. Olhede, “Learning the ambiguity function,” Department of Statistical Science, UCL, arxiv.org/abs/1103.3932, Tech. Rep. 310, 2011.
- [6] G. Matz and F. Hlawatsch, “Nonstationary spectral analysis based on time-frequency operator symbols,” *IEEE Trans. on Information Theory*, vol. 52, pp. 1067–1086, 2006.
- [7] I. M. Johnstone and B. W. Silverman, “Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences,” *The Annals of Statistics*, vol. 32, pp. 1594–1649, 2004.
- [8] A. M. Sayeed, “Optimal kernels for nonstationary spectral estimation,” *IEEE Trans. Signal Proc.*, vol. 43, pp. 478–491, 1995.
- [9] L. Cohen, *Time-frequency analysis: Theory and applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995.
- [10] P. Flandrin and P. Borgnat, “Time-frequency energy distributions meet compressed sensing,” *IEEE Trans. Signal Proc.*, vol. 58, pp. 2974–2982, 2010.

Sparseness-based non-parametric detection and estimation of random signals in noise

Dominique Pastor
Institut Télécom, Télécom Bretagne,
UMR CNRS 3192 Lab-STICC,
Université européenne de Bretagne,
Email: dominique.pastor@telecom-bretagne.eu

Abdourrahmane M. Atto
Université de Bordeaux,
UMR CNRS 5218 IMS,
IPB, ENSEIRB-MATMECA,
Email: abdou.atto@ims-bordeaux.fr

Abstract—The detection of random signals with unknown distributions and occurrences in additive and independent standard Gaussian noise can be performed on the basis of a weak probabilistic definition of sparseness. A sparseness-based estimator of the noise standard deviation can be derived from this definition. It outperforms standard robust estimators, for large signal probabilities of occurrence. The sparseness model and its theoretical applications are also commented in connection with statistical properties of wavelet transforms of wide-sense stationary random processes. Links between sparseness and the problem of testing the norm of a random signal are presented and discussed as well.

I. THE PROBLEM

We address the fundamental statistical signal processing problem of detecting some signal or signal coefficient with unknown probability distribution in additive standard Gaussian noise with possibly unknown standard deviation. The decision is performed from either the noisy observations or the coefficients observed in a transform domain where the signal is assumed to obey a sparseness model discussed below. We address the very general case of a random signal with unknown distribution for the following reasons. First, the deterministic assumption on the signal is an oversimplification with regard to physics and the random model should generally be preferred. Second, in applications based on passive sensors (Electronic Support Measure, spectrum sensing, among others), so little may be known about the signal or most of its describing parameters [1] that the signal probability distribution can be partially or definitely unknown.

II. SPARSENESS-BASED DETECTION AND ESTIMATION

The non-parametric tests established in [2] guarantee an error probability upper bound for the detection of those signals whose probability of presence does not exceed $p^* \leq 1/2$ and whose norm is above (a-s) some positive lower bound. Such constraints on the signal probability of occurrence and norm specify a sparseness model for random signals. This model actually bounds our lack of prior knowledge. It is said to be weak because it involves the case of probabilities of presence possibly equal to one half. In contrast, standard sparsity models derived from [3] are stronger since they correspond to $p^* \ll 1/2$. The sparseness model deriving from [2] applies to non-parametric estimation by wavelet shrinkage, whereas standard sparsity models concern deterministic signals even for detection problems [4].

The tests of [2] require prior knowledge of the noise standard deviation. Sparseness is then instrumental to design a new estimator of the noise standard deviation when signals have unknown distributions and probabilities of presence in noise [5]. This sparseness-based estimator outperforms standard robust ones, when p^* is large and even equal to 1/2. Indeed, robust estimators may fail in estimating the noise standard deviation in presence of

too many signals acting as outliers among the noise data, whereas our model covers such situations.

The results summarized above not only comply with the sparse nature of wavelet transforms for signal representations but also with statistical properties of wavelet packets. Specifically, coefficients returned by wavelet packet transforms of wide-sense stationary random processes tend to be Gaussian uncorrelated when the resolution level and the order of the decomposition filters are both large enough [6].

A perspective of the results summarized above is then the design of unsupervised algorithms capable of detecting, estimating and acquiring statistical knowledge about random signals that obey our sparseness model and whose distributions and occurrences are initially unknown.

III. SIGNAL NORM TESTING (SNT) AND SPARSENESS

Signal norm testing (SNT) is the problem of deciding whether a random signal norm exceeds some specified value or not, when the signal has unknown probability distribution in additive and independent standard Gaussian noise [7]. The crux in the approach is the invariance of the noise probability distribution. An optimality criterion, based on this invariance only, is introduced to design SNT tests. Sparse SNT (SSNT) will then be proposed. In particular, SSNT of a random signal whose norm has bimodal and/or heavy-tailed distribution is akin to the sparseness-based detection problem of [2]. Some applications of SSNT to signal and image processing will also be provided.

REFERENCES

- [1] T. Kailath and H. V. Poor, "Detection of stochastic processes," *IEEE Transactions on Information Theory*, vol. 44, pp. 2230 – 2259, 1998.
- [2] D. Pastor, R. Gay, and A. Gronenboom, "A sharp upper bound for the probability of error of likelihood ratio test for detecting signals in white gaussian noise," *IEEE Transactions on Information Theory*, vol. 48, no. 1, pp. 228–238, January 2002.
- [3] D. Donoho and I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425 – 455, August 1994.
- [4] Z. Wang, G. Arce, and B. Sadler, "Subspace compressive detection for sparse signals," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*, April 2008, pp. 3873 – 3876.
- [5] D. Pastor and F-X. Socheleau, "Robust estimation of noise standard deviation in presence of signals with unknown distributions and occurrences," in *revision, IEEE Transactions on Information Theory*, 2011.
- [6] A. Atto and D. Pastor, "Central limit theorems for wavelet packet decompositions of stationary random processes," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 896 – 901, February 2010.
- [7] D. Pastor, "Signal norm testing in additive and independent standard gaussian noise," RR - 2011 01 - SC, Institut Télécom, Télécom Bretagne, Lab-STICC UMR CNRS 3192, Tech. Rep., 2011.

Reconstruction and Cancellation of Sampled Multiband Signals Using Discrete Prolate Spheroidal Sequences

Mark A. Davenport
Stanford University
markad@stanford.edu

Michael B. Wakin
Colorado School of Mines
mwakin@mines.edu

Abstract—There remains a significant gap between the discrete, finite-dimensional compressive sensing (CS) framework and the problem of acquiring a continuous-time signal. In this talk, we will discuss how sparse representations for multiband signals can be incorporated into the CS framework through the use of Discrete Prolate Spheroidal Sequences (DPSS’s). DPSS’s form a highly efficient basis for sampled bandlimited functions; by modulating and merging DPSS bases, one obtains a sparse representation for sampled multiband signals. We will discuss the use of DPSS bases for both signal recovery and the cancellation of strong narrowband interferers from compressive samples.

EXTENDED ABSTRACT

In many respects, the core theory of compressive sensing (CS) is now well-settled. Given a suitable number of compressive measurements $y = \Phi x$ of a finite-dimensional vector x , one can recover x exactly if x can be expressed in some dictionary Ψ as $x = \Psi\alpha$ where α is exactly sparse. If α is not exactly sparse, then one can recover an approximation to x , and there exist provably efficient and robust algorithms for performing this recovery.

However, although one of the primary motivations for CS is to simplify the way that high-bandwidth signals are *sampled*, there remains a significant gap between the discrete, finite CS framework and the problem of acquiring a continuous-time signal. Previous work has attempted to bridge this gap by employing two very different strategies. First, in [11] the authors operate directly within the CS framework by employing the simple (but somewhat unrealistic) assumption that the analog signal being sampled is comprised of a sparse linear combination of pure tones with frequencies restricted to a harmonic grid. The advantage of this assumption is that it ensures a finite-dimensional sparse representation for x if one chooses Ψ to be the DFT basis. Alternatively, other authors have considered a more realistic signal model—the class of *multiband signals* built from sums of narrowband, bandpass signals—but have performed their analysis largely outside of the standard CS framework [4, 8].

In this talk, we will discuss how sparse representations for multiband signals can be incorporated directly into the CS framework through the use of Discrete Prolate Spheroidal Sequences (DPSS’s) [10]. First introduced by Slepian in 1978, the DPSS’s can be viewed (and derived) as the discrete-time, finite-length sequences whose Discrete-Time Fourier Transform (DTFT) is most concentrated within a given bandwidth. Most significantly, one can show that for a given sequence of length N and bandlimit $W \in (0, \frac{1}{2})$, the first $\approx 2NW$ DPSS functions form a basis that will capture virtually all of the energy in any length- N sample vector arising from the uniform sampling of a *bandlimited* analog signal. We will expand upon this fact in our talk and explain how, by modulating DPSS’s from the baseband to a carrier frequency f_c , one obtains a basis for sample vectors arising from the uniform sampling of *bandpass* analog signals. Merging collections of modulated DPSS’s, one then obtains bases for sample vectors arising from the uniform sampling of *multiband* analog signals.

We will discuss the role that such DPSS bases can have in CS. One natural application is in the recovery of windows of multiband signals from the sort of compressive measurements that arise in nonuniform sampling [1] or random demodulation [7] CS architectures. The DPSS bases enjoy a tremendous advantage over the DFT for this purpose; while the DFT representation for a multiband signal is not sparse (it is not even compressible!), the DPSS representation for a multiband signal is almost perfectly sparse and indeed reflects the fundamental information level. We will discuss ongoing work in developing DPSS-based recovery algorithms for CS. Our work on this front differs from [5, 6, 9] in that we consider discrete-time vectors that arise from sampling analog signals with arbitrary multiband spectra.

A second application of the DPSS bases in compressive signal processing involves the cancellation of strong narrowband interferers from a set of compressive samples. Building on the work in [2, 3], we will explain how such interferers can easily be cancelled by orthogonalizing a measurement vector against the DPSS subspace, and we will demonstrate that various signal inference problems can be solved with a high degree of accuracy after the cancellation of an interferer many times stronger than the signal itself.

REFERENCES

- [1] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, 2008.
- [2] M. Davenport, P. Boufounos, and R. Baraniuk. Compressive domain interference cancellation. In *Proc. Work. Struc. Parc. Rep. Adap. Signaux (SPARS)*, Saint-Malo, France, Apr. 2009.
- [3] M. Davenport, P. Boufounos, M. Wakin, and R. Baraniuk. Signal processing with compressive measurements. *IEEE J. Select. Top. Signal Processing*, 4(2):445–460, 2010.
- [4] P. Feng and Y. Bresler. Spectrum-blind minimum-rate sampling and reconstruction of multiband signals. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Atlanta, GA, May 1996.
- [5] L. Gosse. Compressed sensing with preconditioning for sparse recovery with subsampled matrices of Slepian prolate functions. Preprint, 2010.
- [6] S. Izquierdo and J.D. Lakey. Time-frequency localization and sampling of multiband signals. *Acta Appl. Math.*, 107(1):399–435, 2009.
- [7] S. Kirolos, J. Laska, M. Wakin, M. Duarte, D. Baron, T. Ragheb, Y. Massoud, and R. Baraniuk. Analog-to-information conversion via random demodulation. In *Proc. IEEE Dallas Circuits and Systems Work. (DCAS)*, Dallas, TX, Oct. 2006.
- [8] M. Mishali and Y. Eldar. Blind multi-band signal reconstruction: Compressed sensing for analog signals. *IEEE Trans. Signal Processing*, 57(3):993–1009, 2009.
- [9] S. Senay, L.F. Chaparro, and L. Durak. Reconstruction of nonuniformly sampled time-limited signals using prolate spheroidal wave functions. *Signal Processing*, 89(12):2585–2595, 2009.
- [10] D. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V – The discrete case. *Bell Systems Tech. J.*, 57:1371–1430, 1978.
- [11] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk. Beyond Nyquist: Efficient sampling of sparse, bandlimited signals. *IEEE Trans. Inform. Theory*, 56(1):520–544, 2010.

Exponential Reproducing Kernels for Sparse Sampling

Jose Antonio Uriaguén

Imperial College of London

jose.uriaguén08@imperial.ac.uk

Pier Luigi Dragotti

Imperial College of London

p.dragotti@imperial.ac.uk

Thierry Blu

The Chinese University of Hong Kong

thierry.blu@m4x.org

Abstract—The theory of Finite Rate of Innovation (FRI) broadened the traditional sampling paradigm to certain classes of parametric signals. In this paper we review the ideal FRI sampling scheme and some techniques to combat noise. We then present alternative and more effective denoising methods for the case of exponential reproducing kernels.

I. INTRODUCTION

In [1] and [2] it was shown how certain classes of non-bandlimited signals can be sampled and perfectly reconstructed. These signals can be completely characterised by their rate of innovation. In the presence of noise, the ideal approaches become unstable and alternative methods are required [3]. This paper focuses on the optimal use of exponential reproducing kernels introduced in [2] for the noisy scenario.

II. SAMPLING SIGNALS WITH FRI

Consider a stream of K Diracs at locations t_k , with amplitudes a_k and of duration τ seconds. If we sample the signal with an exponential reproducing kernel $\varphi(-\frac{t}{T})$ we obtain the measurements $y_n = \langle x(t), \varphi(\frac{t}{T} - n) \rangle$, for $n = 0, 1, \dots, N - 1$. Here N is the number of samples and we use a sampling period $T = \frac{\tau}{N}$.

An exponential reproducing kernel is any function $\varphi(t)$ that satisfies $\sum_{n \in \mathbb{Z}} c_{m,0} e^{\alpha_m(n-t)} \varphi(t-n) = 1$ with $\alpha_m \in \mathbb{C}$ for appropriate coefficients $c_{m,n} = c_{m,0} e^{\alpha_m n}$. Equivalently we can write

$$c_{m,0} \int_{-\infty}^{\infty} e^{-\alpha_m t} \varphi(t) dt = 1. \quad (1)$$

Furthermore, any composite function of the form $\varphi(t) = \gamma(t) * \beta_{\tilde{\alpha}_P}(t)$, where $\beta_{\tilde{\alpha}_P}(t)$ is an E-Spline [4], is able to reproduce the set $e^{\alpha_m t}$, $m = 0, 1, \dots, P$.

Reconstructing the input is a two step process [2]. First, the samples y_n are linearly combined to get the new measurements $s_m = \sum_{n=0}^{N-1} c_{m,n} y_n$. These are equivalent to a power series involving the locations t_k and amplitudes a_k for $\alpha_m = \alpha_0 + m\lambda$. Second, the unknown parameters can be retrieved using the classical Prony's method. The key ingredient is the annihilating filter, for which the following holds [3]:

$$\mathbf{Sh} = 0 \quad (2)$$

i.e. the Toeplitz matrix \mathbf{S} is rank deficient. Note that we require $P \geq 2K - 1$.

III. WORKING IN THE PRESENCE OF NOISE

When the sampling process is not ideal we obtain a corrupted version of the measurements $\hat{y}_n = y_n + \epsilon_n$. The Toeplitz matrix of (2) then becomes $\hat{\mathbf{S}} = \mathbf{S} + \mathbf{B}$ and is no longer rank deficient. When the noise term \mathbf{B} is additive white Gaussian (AWGN) it is reasonable to look for a solution that minimises $\|\hat{\mathbf{S}}\mathbf{h}\|^2$ s.t. $\|\mathbf{h}\| = 1$ [3]. This is a classical total-least-square (TLS) problem that can be solved using singular value decomposition (SVD). The solution is further improved by denoising $\hat{\mathbf{S}}$ using, for instance, Cadzow algorithm.

Jose Antonio Uriaguén is sponsored by the non-profit organisation “Fundación Caja Madrid” — Pier Luigi Dragotti is in part supported by a Global Research Award from the Royal Academy of Engineering.

Modified TLS and E-Splines

For exponential reproducing kernels \mathbf{B} is due to coloured noise. In order for SVD to provide a reliable separation of the signal and noise subspaces it becomes necessary to “pre-whiten” the noise. If we know the covariance matrix of the noise \mathbf{R} up to a constant factor λ , we can factor it: $\mathbf{R} = \lambda \mathbf{B}^* \mathbf{B} = \mathbf{Q}^T \mathbf{Q}$ and recover the appropriate subspaces by considering the SVD of $\hat{\mathbf{S}}' = \hat{\mathbf{S}} \mathbf{Q}^{-1}$.

It is also possible to control the term \mathbf{B} by designing an appropriate sampling kernel. Consider the matrix \mathbf{C} of size $(P+1) \times N$ with coefficients $c_{m,n}$ at locations (m, n) . If we want the noise to be white we need the matrix \mathbf{C} to have orthonormal rows. This is achieved by making them orthogonal with $\alpha_m = j\omega_m = j\frac{2\pi m}{N}$ and then orthonormal by setting $|c_{m,0}| = 1$, which is achieved using (1):

$$|\hat{\varphi}(\omega_m)| = |\hat{\gamma}(\omega_m) \hat{\beta}_{\tilde{\alpha}_P}(\omega_m)| = 1, \quad (3)$$

where $\hat{\varphi}(\cdot)$ is the Fourier transform of $\varphi(t)$. Among the kernels satisfying (3), we are interested in the one with the shortest support. This kernel can be formed as a linear combination of various derivatives of the original E-Spline. It is a variation of the maximal-order minimal-support kernels of [5] and is still able to reproduce exponentials. Now, solving the problem in the Fourier domain we only need to determine a polynomial that interpolates $(\omega_m, |\hat{\beta}_{\tilde{\alpha}_P}(\omega_m)|^{-1})$.

IV. SIMULATION RESULTS

Fig. 1 shows the modified E-Spline kernels ('ME') have the best performance, which improves with increasing order P . The modified Cadzow algorithm ('MC') marginally beats the original ('C').

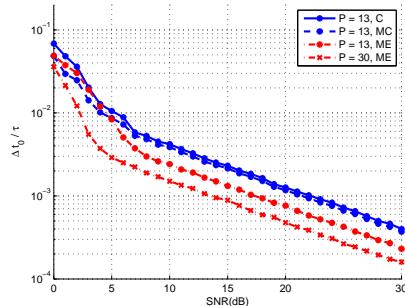


Figure 1. Retrieval of $K = 2$ Diracs in the presence of noise. We use $\tau = 1$ seconds, $N = 31$ samples and average over 1000 realisations.

REFERENCES

- [1] M. Vetterli, P. Marziliano, and T. Blu, “Sampling signals with finite rate of innovation,” *IEEE Transactions on Signal Processing*, vol. 50, pp. 1417–1428, 2002.
- [2] P. L. Dragotti, M. Vetterli, and T. Blu, “Sampling Moments and Reconstructing Signals of Finite Rate of Innovation: Shannon Meets Strang-Fix,” *IEEE Transactions on Signal Processing*, vol. 55 (5), pp. 1741–1757, 2007.
- [3] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, L. Coulot, “Sparse Sampling of Signal Innovations,” *IEEE Signal Processing Magazine*, vol. 25 (2), pp. 31–40, 2008.
- [4] M. Unser and T. Blu, “Cardinal Exponential Splines: Part I — Theory and Filtering Algorithms,” *IEEE Transactions on Signal Processing*, vol. 53, pp. 1425–1438, 2005.
- [5] T. Blu, P. Thevenaz, and M. Unser, “MOMS: maximal-order interpolation of minimal support,” *IEEE Transactions on Image Processing*, vol. 10, pp. 1069–1080, 2001.

Generalized sampling and infinite-dimensional compressed sensing

Ben Adcock

Simon Fraser University, Burnaby, Canada
Email: ben.adcock@sfu.ca

Anders C. Hansen

University of Cambridge, Cambridge, UK
Email: A.Hansen@damtp.cam.ac.uk

Abstract—We will discuss a generalization of the Shannon Sampling Theorem that allows for reconstruction of signals in arbitrary bases in a completely stable way. When extra information is available, such as sparsity or compressibility of the signal in a particular bases, one may reduce the number of samples dramatically. This is done via Compressed Sensing techniques, however, the usual finite-dimensional framework is not sufficient. To overcome this obstacle I'll introduce the concept of Infinite-Dimensional Compressed Sensing.

I. THE SHANNON SAMPLING THEOREM

The well known Shannon Sampling Theorem states that if

$$f = \mathcal{F}g, \quad g \in L^2(\mathbb{R}),$$

(note that \mathcal{F} is the Fourier Transform) and $\text{supp}(g) \subset [-T, T]$ for some $T > 0$, then both f and g can be reconstructed from point samples of f . In particular, if $\epsilon \leq \frac{1}{2T}$ (the Nyquist rate) then

$$f(t) = \sum_{k=-\infty}^{\infty} f(k\epsilon) \text{sinc}\left(\frac{t+k\epsilon}{\epsilon}\right), \quad L^2 \text{ and unif. conv.,} \quad (1)$$

$$g = \epsilon \sum_{k=-\infty}^{\infty} f(k\epsilon) e^{2\pi i k \cdot}, \quad L^2 \text{ convergence.} \quad (2)$$

In practice, one cannot process nor acquire the infinite amount of information $\{f(k\epsilon)\}_{k \in \mathbb{Z}}$ that is needed to fully reconstruct f and g and thus one must resort to forming, for some $N \in \mathbb{N}$, the approximations

$$f_N = \sum_{k=-N}^N f(k\epsilon) \text{sinc}\left(\frac{t+k\epsilon}{\epsilon}\right), \quad g_N = \epsilon \sum_{k=-N}^N f(k\epsilon) e^{2\pi i k \cdot}.$$

The question on how well these functions approximate f and g is related to the speed of convergence of the series in (1) and (2). Which again is related to how suitable the functions $\{\text{sinc}((\cdot + k\epsilon)/(\epsilon))\}_{k \in \mathbb{Z}}$ and $\{e^{2\pi i k \cdot}\}_{k \in \mathbb{Z}}$ are in series expansions of f and g . In particular, there may be L^2 functions $\{\varphi_k\}_{k \in \mathbb{N}}$ and coefficients $\{\beta_k\}_{k \in \mathbb{N}}$ such that the series

$$f = \sum_{k \in \mathbb{N}} \beta_k \mathcal{F}\varphi_k, \quad g = \sum_{k \in \mathbb{N}} \beta_k \varphi_k$$

converge faster than the series in (1). There are therefore two important questions to ask:

- (i) Can one obtain the coefficients $\{\beta_k\}_{k \in \mathbb{N}}$ (or at least approximations to them) in a stable manner, based on the same sampling information $\{f(k\epsilon)\}_{k \in \mathbb{N}}$, and will this yield better approximations to f and g ?
- (ii) Can one subsample from $\{f(\epsilon k)\}_{k \in \mathbb{N}}$ (e.g. not sampling at the Nyquist rate) and still get recovery of $\{\beta_k\}_{k \in \mathbb{N}}$ and hence f and g ?

The final answer to the first question YES! and can be summarized in the following generalization of the Shannon Sampling Theorem below.

The answer to the second question is also YES! (given some extra requirements on the signals f and g). This is done via the concept of Infinite-Dimensional Compressed Sensing.

Theorem I.1. Let \mathcal{F} denote the Fourier transform on $L^2(\mathbb{R}^d)$. Suppose that $\{\varphi_j\}_{j \in \mathbb{N}}$ is an orthonormal set in $L^2(\mathbb{R}^d)$ such that there exists a $T > 0$ with $\text{supp}(\varphi_j) \subset [-T, T]^d$ for all $j \in \mathbb{N}$. For $\epsilon > 0$, let $\rho : \mathbb{N} \rightarrow (\epsilon\mathbb{Z})^d$ be a bijection. Define the infinite matrix

$$U = \begin{pmatrix} u_{11} & u_{12} & u_{13} & \dots \\ u_{21} & u_{22} & u_{23} & \dots \\ u_{31} & u_{32} & u_{33} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad u_{ij} = (\mathcal{F}\varphi_j)(\rho(i)). \quad (3)$$

Then, for $\epsilon \leq \frac{1}{2T}$, we have that $\epsilon^{d/2} U$ is an isometry. Also, set

$$f = \mathcal{F}g, \quad g = \sum_{j=1}^{\infty} \beta_j \varphi_j \in L^2(\mathbb{R}^N),$$

and let (for $l \in \mathbb{N}$) P_l denote the projection onto $\text{span}\{e_1, \dots, e_l\}$. Then, for every $K \in \mathbb{N}$ there is an $n \in \mathbb{N}$ such that, for all $N \geq n$, the solution to

$$A \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \tilde{\beta}_3 \\ \vdots \\ \tilde{\beta}_K \end{pmatrix} = P_K U^* P_N \begin{pmatrix} f(\rho(1)) \\ f(\rho(2)) \\ f(\rho(3)) \\ \vdots \\ f(\rho(K)) \end{pmatrix}, \quad A = P_K U^* P_N U P_K|_{P_K l^2(\mathbb{N})}, \quad (4)$$

is unique. If

$$\tilde{g}_{K,N} = \sum_{j=1}^K \tilde{\beta}_j \varphi_j, \quad \tilde{f}_{K,N} = \sum_{j=1}^K \tilde{\beta}_j \mathcal{F}\varphi_j,$$

then

$$\|g - \tilde{g}_{K,N}\|_{L^2(\mathbb{R}^d)} \leq (1 + C_{K,N}) \|P_K^\perp \beta\|_{l^2(\mathbb{N})}, \quad \beta = \{\beta_1, \beta_2, \dots\},$$

$$\|f - \tilde{f}_{K,N}\|_{L^\infty(\mathbb{R}^d)} \leq (2T)^{d/2} (1 + C_{K,N}) \|P_K^\perp \beta\|_{l^2(\mathbb{N})},$$

where, for fixed K , the constant $C_{K,N} \rightarrow 0$ as $N \rightarrow \infty$.

The results can be found in [1], [2], [3], and the ideas stem from [4].

REFERENCES

- [1] B. Adcock and A. C. Hansen. A generalized sampling theorem for stable reconstructions in arbitrary bases. *Submitted*, 2010.
- [2] B. Adcock and A. C. Hansen. Stable reconstructions in Hilbert spaces and the resolution of the gibbs phenomenon. *Submitted*, 2010.
- [3] A. C. Hansen. Generalized sampling and infinite dimensional compressed sensing. *Submitted*, 2011.
- [4] A. C. Hansen. On the solvability complexity index, the n -pseudospectrum and approximations of spectra of operators. *J. Amer. Math. Soc.*, 24(1):81–124, 2011.

A Lower Complexity Bound for ℓ_1 -regularized Least-squares Problems using a Certain Class of Algorithms

Tobias Lindstrøm Jensen

Department of Electronic Systems, Aalborg University, Aalborg, Denmark, Email: tlj@es.aau.dk

Abstract—The l_1 -regularized least-squares problem have received broad attention the last couple of years. This have resulted in numerous approaches for supplying reliable solvers which combines both well known methods and recently developed techniques for efficient computations. We define a class of algorithms which is not as restrictive as classic black-box algorithms and hence includes most of the recently proposed methods. We show that it is not possible to obtain a worst-case convergence rate better than $\mathcal{O}\left(\frac{1}{k^2}\right)$ for $2k \leq s \leq n$, where k is the iteration counter, s is the size of the support, and n is the dimension.

I. INTRODUCTION

In this extended abstract we will show a lower bound on the convergence for a certain class of algorithms applied to the well studied ℓ_1 -regularized least-squares, $\ell_1 - \ell_2$ -problem or basis pursuit denoising/LASSO in Lagrange form. We will take an approach inspired by [1], see also [2], but we will not necessarily restrict us to the standard black-box assumption. Instead we provide a similar rule which however defines a broad class of algorithms including many popular algorithms.

Consider the well known convex problem of the form

$$\text{minimize } f(x) = \frac{1}{2}x^T Qx - c^T x + \gamma \|x\|_1 \quad (1)$$

for $x, c \in \mathbf{R}^n$, and $Q \in \mathbf{R}^{n \times n}$. Let $g(x) = \frac{1}{2}x^T Qx - c^T x$. Note that $\frac{1}{2}\|Ax - b\|_2^2 = \frac{1}{2}x^T A^T Ax - b^T Ax + b^T b$, such that for minimization over x we have $Q = A^T A$ and $c = A^T b$.

A. A Class of Algorithms \mathcal{M}

Let \mathcal{M} be a class of iterative algorithms with $m_f \in \mathcal{M}$ and denote the support as $\text{supp}(x) = \{ i \mid x_i \neq 0 \}$. The function m_f generate iterates $x^{(k)}$ according to

$$x^{(k)} = m_f(x^{(k-1)}; \nabla g(x^{(k-1)})), \quad k = 1, \dots \quad (2)$$

with

$$\text{supp}(x^{(k)}) \in \text{supp}(x^{(k-1)}) \bigcup \text{supp}(\nabla g(x^{(k-1)})) \quad (3)$$

Note that the function m_f is allowed to make other operations than that governed by $x^{(k-1)}$ and $\nabla g(x^{(k-1)})$ based on its knowledge of the function f , i.e., the iterative method may not satisfy the standard black-box assumption [2], [1], but *may* use a black-box assumption on g , see [3].

B. Results

By inspection it can shown that a wide range of algorithms belongs to \mathcal{M} , such as GPSR [4], IST [5], [6], FISTA [7], [3] FPC [8] and FPC-AS [9], parallel ℓ_1 -regularized least-squares [10], homotopy methods [11] and l_1 _ls [12] if k is the accumulated iteration counter for the preconditioned conjugate gradient method.

By construction a specific function with closed form solution, we provide the following theorem.

Theorem 1.1: For any k , $1 \leq k = \frac{1}{2}|S|$, $|S| \leq n$ and $S = \text{supp}(x^*)$, there exist a function $f(x) = \frac{1}{2}x^T Qx - c^T x + \gamma \|x\|_1$, $Q \succeq 0$, $x \in \mathbf{R}^n$ such that for any $m_f \in \mathcal{M}$ with $x^{(0)} = 0$,

$$f(x^{(k)}) - f^* \geq \frac{1}{6} \frac{\|x^{(0)} - x^*\|_2^2}{(k+1)^2}. \quad (4)$$

Note that sometimes $x^{(0)} = A^T b = c$ is used, which is the same as $x^{(1)}$ if $x^{(0)} = 0$. That is, using $x^{(0)} = A^T b$ compared to $x^{(0)} = 0$ only corresponds to a shift of one iteration.

II. DISCUSSIONS

We note that we do not assume $Q \succeq \mu$, and this result does therefore not conflict with the linear rate of convergence provided in [8]. The theorem show that it is not possible to provide algorithms with better worst-case complexity than $\mathcal{O}\left(\frac{1}{k^2}\right)$ without making further assumptions on the problem and/or algorithm. The result provided is then constructive in the sense that it renders functions and algorithm to avoid, visible to the algorithm designer. It is necessary to steer clear of these functions and/or algorithms if the designer is to provide algorithms with better worst-case iteration complexity.

REFERENCES

- [1] Y. Nesterov, *Introductory Lectures on Convex Optimization, A Basic Course*. Kluwer Academic Publishers, 2004.
- [2] A. S. Nemirovskii and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization*. John Wiley & Sons, Ltd., 1983.
- [3] Y. Nesterov, “Gradient methods for minimizing composite objective function,” Université catholique de Louvain, Center for Operations Research and Econometrics (CORE), 2007, no 2007076.
- [4] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, “Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems,” *IEEE J. Sel. Top. Sign. Proces.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.
- [5] M. A. T. Figueiredo and R. D. Nowak, “An EM algorithm for wavelet-based image restoration,” *IEEE Trans. Image Process.*, vol. 12, pp. 906–916, 2003.
- [6] I. Daubechies, M. Defrise, and C. D. Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Commun. Pure Appl. Math.*, vol. 57, pp. 1413–1457, 2005.
- [7] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J. Imag. Sci.*, vol. 2, pp. 183–202, 2009.
- [8] E. Hale, W. Yin, and Y. Zhang, “Fixed-point continuation for ℓ_1 -minimization: Methodology and convergence,” *SIAM J. Optim.*, vol. 19, no. 3, pp. 1107–1130.
- [9] Z. Wen, W. Yin, D. Goldfarb, and Y. Zhang, “A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization and continuation,” *SIAM J. Sci. Comput.*, vol. 32, no. 4, pp. 1832–1857, 2010.
- [10] R. Raina, A. Madhavan, and A. Y. Ng, “Large-scale deep unsupervised learning using graphics processors,” in *Proc. Int. Conf. on Machine Learning (ICML)*, 2009, pp. 873–880.
- [11] M. R. Osborne, B. Presnell, and B. A. Turlach, “A new approach to variable selection in least squares problems,” *IMA J Numer. Anal.*, vol. 20, no. 3, pp. 389–403, July 2000.
- [12] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, “An interior-point method for large-scale l_1 -regularized least squares,” *IEEE J. Sel. Top. Sign. Proces.*, vol. 1, no. 4, pp. 606–617, 2007.

A New Recovery Analysis of Iterative Hard Thresholding for Compressed Sensing

Coralia Cartis

School of Mathematics and the Maxwell Institute
 University of Edinburgh, Edinburgh EH9 3JZ
 coralia.cartis@ed.ac.uk

Andrew Thompson

School of Mathematics and the Maxwell Institute
 University of Edinburgh, Edinburgh, EH9 3JZ
 a.thompson-8@sms.ed.ac.uk

We consider a standard Compressed Sensing model in which we seek to recover a k -sparse signal $x \in \mathbb{R}^N$ from n linear measurements $b = Ax$, where $k \leq n \leq N$. Since the introduction of CS in 2004, many algorithms have been developed to solve this problem. Because of the paradoxical nature of CS – exact reconstruction from undersampled measurements – it is crucial for the acceptance of an algorithm that rigorous worst-case analysis verifies the degree of undersampling the algorithm permits. This aim can be accomplished by means of the phase transition framework in which we let $(k, n, N) \rightarrow \infty$, while preserving the proportions $\delta = n/N$ and $\rho = k/n$ [1].

We provide a new worst-case analysis for one of these recovery algorithms, Iterative Hard Thresholding (IHT) [3]. While previous recovery results analysed progress of the algorithm from one iteration to the next by means of the Restricted Isometry Property (RIP) [2], [5], we take a different approach. We derive two conditions for general measurement matrices: firstly, by analysing the fixed points of IHT we obtain a condition guaranteeing at most one fixed point (namely the original signal). Secondly, we give an improved condition guaranteeing convergence to some fixed point. If both conditions are satisfied, it follows that we have guaranteed recovery of the original signal.

Provided we make the assumption that the measurement matrix A and the original signal x are independent, the fixed point condition is especially amenable to statistical analysis. For the specific case of Gaussian measurement matrices, such an analysis allows us to derive a quantitative phase transition for exact recovery, which gives a substantial improvement over previous results [1].

We also extend the consideration to a variant of IHT with variable step-size, Normalized Iterative Hard Thresholding (NIHT) [4]. A similar analysis in this case yields a further improvement on the phase transition for Gaussian measurement matrices. This is in fact the first time worst-case guarantees for NIHT have been quantified in this way. Figure 1 illustrates the latter results, showing that recovery is guaranteed asymptotically, with high probability on the draw of A , for (δ, ρ) values falling below the respective curves.

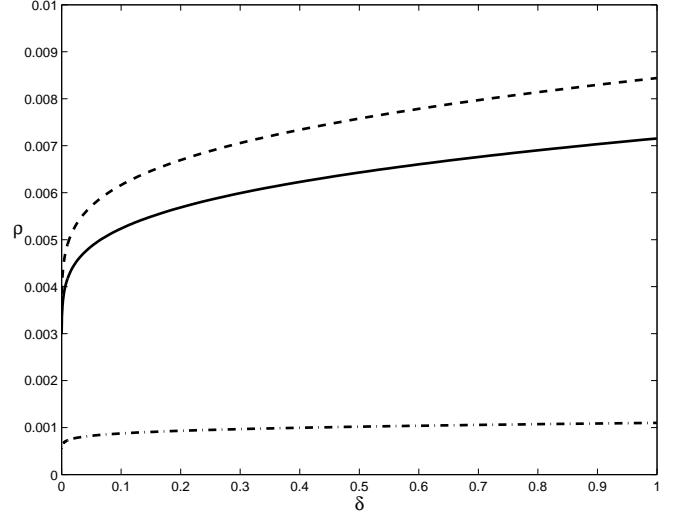


Fig. 1. Exact recovery phase transitions for IHT (unbroken), NIHT (dashed), and the previous RIP analysis of IHT [1] (dash-dot); recovery is guaranteed (asymptotically) below the curve.

A more realistic model from a practical point of view is one in which the original signal x is only approximately k -sparse, and where the measurements are corrupted by noise. We extend our results in both of these directions, proving that, with high probability, for (δ, ρ) values below the same phase transition curve as for exact recovery, the error in approximation is below some multiple of the unrecoverable energy of the system. We explicitly quantify this stability factor for both IHT and NIHT.

REFERENCES

- [1] J. Blanchard, C. Cartis, J. Tanner, and A. Thompson. Greedy phase transitions for sparse approximation algorithms. *Appl. Comput. Harmon. Anal.*, 30(2):188–203, 2011.
- [2] T. Blumensath and M. Davies. Iterative hard thresholding for compressed sensing. *Appl. Comput. Harmon. Anal.*, 27(3):265–274, 2009.
- [3] T. Blumensath and M.E. Davies. Iterative thresholding for sparse approximations. *J. Fourier Anal. and Appl.*, 14(5):629–654, 2008.
- [4] T. Blumensath and M.E. Davies. Normalized iterative hard thresholding: guaranteed stability and performance. *IEEE J. of Selected Topics in Sig. Proc.*, 4(2):298–309, 2010.
- [5] R. Garg and R. Khandekar. Gradient descent with sparsification: An iterative algorithm for sparse recovery with restricted isometry property. In *ICML, Montreal, Canada*, 2009.

Recipes for Hard Thresholding Methods

Anastasios Kyrillidis¹ and Volkan Cevher^{1,2}

¹Laboratory for Information and Inference Systems
Ecole Polytechnique Federale de Lausanne
²Idiap Research Institute
{anastasios.kyrrillidis,volkan.cevher}@epfl.ch

I. MOTIVATION

Iterative Hard Thresholding (IHT) algorithm is a popular sparse recovery method, known for its simplicity, ease of implementation, and low computational complexity per iteration. The algorithm can be described by the following recursion: $x_{i+1} = H_K(x_i + \mu\Phi^T(u - \Phi x_i))$, where i is the iteration count, $H_K(\cdot)$ is the projection onto K -sparse signals, $u \in \mathbb{R}^M$ is the observation vector, $\Phi \in \mathbb{R}^{M \times N}$ ($M \ll N$) is the measurement matrix and μ is a step-size quantity. IHT method is theoretically well-investigated [1].

In this paper, we describe several modular building blocks to derive IHT variants with faster convergence, reduced computational complexity and better phase transition performance.

II. BUILDING BLOCKS

Adaptive step-size selection: Given x_i is K -sparse, [2] observes that the support set of the new estimate x_{i+1} is included in the set $S_i = \text{supp}(x_i) \cup \text{supp}(H_K(\nabla_{\mathcal{I} \setminus \text{supp}(x_i)} f(x_i)))$, where $\mathcal{I} \setminus \text{supp}(x_i)$ is the set of indices of the non-zero elements outside the set $\text{supp}(x_i)$. We propose to use this key information to select a step-size μ at each iteration in closed form as the minimizer of the objective function $f(x_i)$: $\mu = \frac{\|\nabla_{S_i} f(x_i)\|^2}{\|\Phi \nabla_{S_i} f(x_i)\|^2}$ [2]. This adaptive step-size selection, however, results in more restrictive isometry constants:

Lemma 1: Let δ_{3k} be the smallest number such that $(1 - \delta_{3k})\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_{3k})\|x\|_2^2$ is satisfied for all $3k$ -sparse vectors. Then, in noiseless case, the IHT and HTP [6] methods with adaptive μ selection converge towards the true k -sparse signal if $\delta_{3k} < 0.123$ and $\delta_{3k} < 0.2448$, respectively. ■

We also compare and contrast the other alternative methods for step-size selection [3].

Memory: Iterative algorithms can use memory (i.e., previous estimates or gradients) to provide momentum in convergence. The success of the memory-based approaches depends on the iteration dependent momentum terms combining the previous estimates and/or the gradients. We consider both adaptive and non-adaptive proposals for memory-based acceleration, and investigate their effects on the algorithmic approximation guarantees. We illustrate that memory size plays significant role on the convergence speed; by keeping track a history of previous computations, we can reduce the total run-time of sparse approximation [4], [5].

Gradient updates on restricted support sets: We also investigate the impact of greedy gradient updates on restricted supports in conjunction with the other building blocks. Such updates enhance the well-characterized (F)HTP, SP, and CoSaMP algorithms.

III. EXPERIMENTAL RESULTS AND DISCUSSION

Figure 1 highlights the differences in convergence speed between various IHT methods on synthetic data. We use the naming convention from [2]. Our codes are available at <http://lions.epfl.ch/software>. Convergence plots provide empirical evidence for our claims on faster

convergence and reduced complexity. We also theoretically analyze the computational trade-offs in detail.

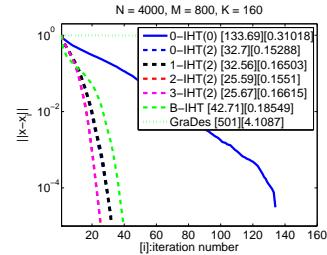


Fig. 1: Average error per iteration - [avg. # of iterations][avg. execution time] - Algorithms: IHT(0): adaptive μ on S_i , IHT(2): IHT(0) + gradient update on $\text{supp}(x_{i+1})$, B-IHT [7], GraDes [8].

Lemma 1 improves δ_{3k} condition constants presented in [2] for the corresponding IHT algorithm. Furthermore, we deduce that μ selection may deteriorate restricted isometry bounds but has a major impact in stability and convergence speed. Figure 2 depicts a representative example.

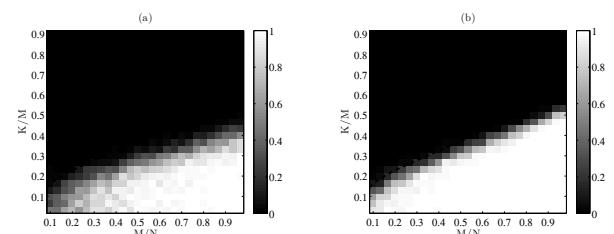


Fig. 2: Probability of exact recovery with sparse measurement matrices Φ . (a) HTP with $\mu = 1$, (b) HTP with our adaptive μ selection.

REFERENCES

- [1] T. Blumensath and M.E. Davies, *Iterative hard thresholding for compressed sensing*, Applied and Computational Harmonic Analysis, vol. 27, no. 3, pp. 265-274, 2009.
- [2] V. Cevher, *On Accelerated hard thresholding methods for sparse approximation*, Technical Report, 2011.
- [3] T. Blumensath and M.E. Davies, *Normalized Iterative Hard Thresholding: Guaranteed Stability and Performance*, Selected Topics in Signal Processing, IEEE Journal of , vol.4, no.2, pp.298-309, April 2010.
- [4] V. Cevher, *An ALPS view of sparse recovery*, Technical Report, 2010.
- [5] K. Qiu and A. Dogandžić, *Double overrelaxation thresholding methods for sparse signal reconstruction*, Proc. 44th Annu. Conf. Inform. Sci. Syst., Princeton, NJ, Mar. 2010.
- [6] S. Foucart, *Hard thresholding pursuit: An algorithm for compressive sensing*, Preprint, 2010.
- [7] T. Blumensath, *Accelerated iterative hard thresholding*, Preprint, 2011.
- [8] R. Garg and R. Khandekar, *Gradient descent with sparsification: An iterative algorithm for sparse recovery with restricted isometry property*, In ICML. ACM, 2009.

Local optimality of dictionary learning algorithms

Boris Mailhé

Centre for Digital Music

School of Electronic Engineering Computer Science

Queen Mary University of London

boris.mailhe@eecs.qmul.ac.uk

Mark Plumley

Centre for Digital Music

School of Electronic Engineering Computer Science

Queen Mary University of London

mark.plumley@eecs.qmul.ac.uk

This work aims at providing theoretical arguments to compare dictionary learning algorithms. We focus on three algorithms: the Olshausen and Field algorithm (Ols-DLA) [1], MOD [2] and K-SVD [3]. We show that the stationary points of Ols-DLA and MOD are the critical points of the residual error energy cost function (i.e. points with null gradient, not necessarily local minima), whereas the set of K-SVD stationary points is strictly included in the critical point set. We provide an example of a point where Ols-DLA and MOD would stop whereas K-SVD can reach a better solution.

Let S be a $T \times N$ matrix of training data. We consider the following dictionary learning problem

$$\min_{\Phi, X} \|S - \Phi X\|_2^2 \quad (1)$$

with Φ a dictionary matrix of size $T \times A$ and X a sparse decomposition matrix of size $A \times N$ containing at most $K \ll T$ non-zero elements in each column.

Both MOD, K-SVD and Ols-DLA minimise the cost function (1) by alternating updates of the support of X , the values of the non-zero coefficients in X and the coefficients in Φ . MOD and Ols-DLA use the Orthogonal Matching Pursuit algorithm (OMP) [4] to estimate X when fixing Φ , then update the atoms (columns Φ_a of Φ) when fixing X :

- Ols-DLA: $\Phi \leftarrow \Phi - \alpha R X^*$, with $R = S - \Phi X$ the residual and α a fixed learning rate
- MOD: $\Phi = S X^+$, with X^+ the pseudo-inverse of X

K-SVD also uses OMP to estimate the support of X , but it then jointly updates the values of the dictionary and decomposition coefficients. For an atom Φ_a and the corresponding coefficient line X^a , it defines the patch matrix $P_{(a)} = R_{cosupp(a)} + \Phi_a * X^a$, with $cosupp(a) = \{n \in [1, N] | X_n^a \neq 0\}$. Then the atom is updated with the principal component of its patch matrix:

$$\Phi_a \leftarrow \arg \max_V V^* P_{(a)} P_{(a)}^* V \quad X^a \leftarrow V^* P_{(a)}$$

We investigate the stationary points of these algorithms to find whether they converge towards the same solutions. If one of those algorithms converges, then the decomposition support becomes stationary, so OMP only computes an orthogonal projection:

$$\forall n, X_n^{supp(n)} \leftarrow \Phi_{supp(n)}^+ S \quad (2)$$

with $supp(n) = \{a \in [1, A] | X_n^a \neq 0\}$.

If one differentiates the cost function (1),

$$\begin{aligned} & \|S - (\Phi + d\Phi)(X + dX)\|_2^2 - \|S - \Phi X\|_2^2 \\ &= -2\langle RX^*, d\Phi \rangle - 2\langle \Phi^* R, dX \rangle + o(\|dX\|_2 + \|d\Phi\|_2) \end{aligned}$$

one can easily show that the stationary conditions for both MOD and Ols-DLA are equivalent to setting RX^* and $\Phi^* R$ to 0: the stationary points of MOD and Ols-DLA are the critical points of the cost function. This equivalence does not hold for K-SVD. Instead, we have the following lemma:

Lemma 1. *The critical points of the cost function are the where points where each atom D_a is an eigenvector of the matrix $P_{(a)} P_{(a)}^*$ (and the decomposition X is orthogonal as described in Equation (2)).*

If an atom is an eigenvector not associated with the highest eigenvalue, it is stationary for MOD and Ols-DLA but not for K-SVD. For example, if we set

$$S = \begin{pmatrix} 2 & 2 \\ -1 & 1 \end{pmatrix} \quad \Phi = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad X = \begin{pmatrix} -1 & 1 \end{pmatrix}$$

then MOD and Ols-DLA would stop whereas K-SVD would find the best atom on the next iteration.

This result hints that Ols-DLA or MOD can be used as initialisations for K-SVD. We got promising results on random signals generated from a Gaussian dictionary: whereas both all algorithms only recover the exact dictionary in less than 10% of the cases, running Ols-DLA followed by K-SVD recovers the exact dictionary in more than 90% of the cases.¹

REFERENCES

- [1] B. A. Olshausen and D. J. Field, “Emergence of simple-cell receptive field properties by learning a sparse code for natural images,” *Nature*, vol. 381, pp. 607–609, jun 1996.
- [2] K. Engan, S. Aase, and J. Hakon Husoy, “Method of optimal directions for frame design,” in *Acoustics, Speech, and Signal Processing, 1999. ICASSP ’99. Proceedings., 1999 IEEE International Conference on*, vol. 5, 1999, pp. 2443 –2446 vol.5.
- [3] M. Aharon, M. Elad, and A. Bruckstein, “k -svd: An algorithm for designing overcomplete dictionaries for sparse representation,” *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311 –4322, nov 2006.
- [4] Y. Pati, R. Rezaifar, and P. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, Nov. 1993, pp. 40 –44 vol.1.

¹This work was supported by the EPSRC Project EP/G007144/1 Machine Listening using Sparse Representations and by the EU FET-Open project FP7-ICT-225913- SMALL.

Approximate Message Passing for Bilinear Models

Philip Schniter

Department of Electrical and
Computer Engineering
The Ohio State University
Email: schniter@ece.osu.edu

Volkan Cevher

Laboratory for Information and Inference Systems
Idiap Research Institute
Ecole Polytechnique Federale de Lausanne
Email: volkan.cevher@epfl.ch

I. EXTENDED SUMMARY

Problem: We consider the following bilinear model in the unknowns $\mathbf{X} \in \mathbb{R}^{N \times L}$ and $\Phi \in \mathbb{R}^{M \times N}$, which has applications in dictionary learning, matrix completion, collaborative filtering, compressive system calibration, compressive sensing with dictionary uncertainty, and Bayesian experimental design:

$$\mathbf{Y} = \mathcal{P}(\Phi \mathbf{X}) + \mathbf{W}. \quad (1)$$

In (1), \mathbf{Y} are known observations, $\mathcal{P}(\cdot)$ accomplishes element-wise selection or linear projection, and \mathbf{W} models additive perturbation. Please see [1] for further details.

Approach: We take a Bayesian approach to the inference problems (in particular, posterior estimation) that revolve around the bilinear model (1). In particular, we leverage the approximate message passing (AMP) framework of [2], [3] and extend it to the bilinear domain. Compared to Bayesian approaches that rely on Gibbs sampling methods or variational inference, the AMP framework allows us to fully exploit the *blessings-of-dimensionality* (e.g., the asymptotic normality and concentration-of-measures) to achieve salient advantages in computation and estimation accuracy. Our “turbo AMP” framework also allows us to characterize the impact of our message scheduling using extrinsic information transfer (EXIT) charts, originally developed to predict the convergence of turbo decoding.

Example Application: For concreteness, we describe the application of the bilinear model (1) to the *compressive system calibration problem*. Based on the theoretical premise of compressive sensing, a great deal of research has revolved around the design of sampling systems, such as Analog-to-Information receivers and Xampling. The sampling matrices in these systems are pre-designed with certain desired theoretical properties to guarantee recovery along with the constraints of hardware implementations. However, when implementing the mathematical “sampling” operation—here defined by the matrix Φ —in real hardware, one often introduces what are effectively perturbations on Φ that create an undesired gap between theoretical and practical system performance. As a means of closing this gap, we are interested in jointly learning the true matrix Φ while simultaneously recovering the signal \mathbf{X} .

Suppose, then, that our compressive sensing system produces a sequence of vector observations \mathbf{y}_l ($l = 1, \dots, L$; collectively referred as \mathbf{Y}) that correspond to a sequence of unknown sparse signals \mathbf{x}_l (collectively, \mathbf{X}). We assume that the signal coefficients $\{x_{jl}\}$ are drawn i.i.d from a (known) compressible prior $x_{jl} \sim p_x(\cdot)$, and we model the entries of the true (unknown) sampling matrix Φ as i.i.d Gaussian with variance μ^w and known mean $\bar{\Phi}$. For ease of description, we assume that μ^w is known, that the signals \mathbf{x}_l are canonically sparse, and that the projection operator $\mathcal{P}(\cdot)$ is identity. This calibration problem yields the factor graph in Fig. 1, to which

we apply *bilinear AMP* in order to generate (approximate) posterior marginals on the elements of Φ and \mathbf{X} .

This calibration problem can be interpreted as an instance of dictionary learning, whereby one seeks a sparsifying dictionary for some training data. In this setting, it is known that ℓ_1 -norm minimization can locally identify the correct dictionary (i.e., Φ) given $L = \mathcal{O}(N^3K)$ training samples, where K is the “sparsity” of \mathbf{x}_l [4]. We note, however, that the computational complexity of this approach is extremely demanding for large scale problems.

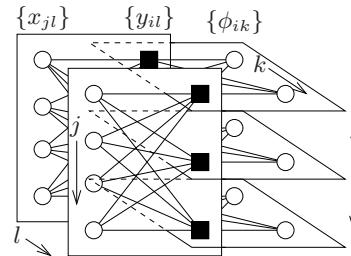


Fig. 1. An illustration of the factor graph for our message passing solution.

Preliminary Results: Figure 2 shows example results for the application of bilinear AMP to the calibration problem. The non-convexity of the problem is quite apparent from the plots. Here, to generate the signals, we used an i.i.d Bernoulli-Gaussian prior that generated zero-mean unit-variance active coefficients with probability K/N . The nominal sampling matrix $\bar{\Phi}$ was generated i.i.d Gaussian with zero mean and $1/M$ -variance, and true Φ was generated by perturbing $\bar{\Phi}$ with an additive noise of the same distribution.

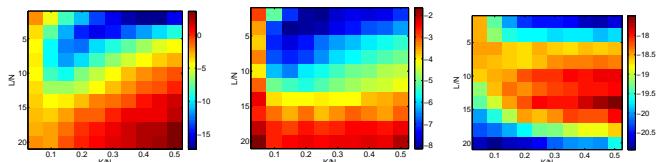


Fig. 2. Recovery errors in dB: the dictionary, the signals, and the data.

REFERENCES

- [1] L. Carin, R. Baraniuk, V. Cevher, D. Dunson, M. Jordan, G. Sapiro, and M. Wakin, “A Bayesian Approach to Learning Low-Dimensional Signal Models from Incomplete Measurements,” *IEEE SP Mag*, 2010.
- [2] D. L. Donoho, A. Maleki, and A. Montanari, “Message passing algorithms for compressed sensing,” *Proc. National Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.
- [3] S. Rangan, “Generalized approximate message passing for estimation with random linear mixing,” *arXiv:1010.5141*, Oct. 2010.
- [4] Q. Geng, H. Wang, and J. Wright, “On the Local Correctness of L^1 Minimization for Dictionary Learning,” *arXiv:1101.5672*, 2011.

Structure Aware Non-Negative Dictionary Learning

Ken O'Hanlon and Mark D. Plumley
School of Electronic Engineering and Computer Science
Queen Mary University of London
Mile End Road, London, E1 4NS, UK
Email: {ken.hanlon, mark.plumley}@eecs.qmul.ac.uk

Abstract—Blind spectrogram decompositions are commonly used for tasks such as automatic transcription. In musical spectrograms, signal elements are highly overlapping in both the time and frequency domains, presenting difficulties to the decomposition method. The harmonic structure prevalent in tonal music signals has been exploited in decomposition methods with positive results. The performance of blind decompositions for transcription tasks has been shown to be dependent on the relationship of the learning order to the number of sources in the signal. We consider structure-aware dictionary learning methods, which have prior knowledge of the structure and cardinality of the sources.

I. INTRODUCTION

Blind decompositions have been used for musical signal processing tasks. Often the signal is represented by a magnitude spectrogram S , and the decomposition seeks, a dictionary of atoms D and a matrix T consisting of the time support vectors for these atoms, such that

$$S \approx DT \quad (1)$$

where $S \in \mathbb{R}_+^{M \times N}$, $D \in \mathbb{R}_+^{M \times K}$, $T \in \mathbb{R}_+^{K \times N}$, M is the number of frequency bins, N is the number of time bins and K is the learning order of the decomposition.

The most commonly used method to perform this decomposition is Non-negative Matrix Factorization (NMF). Sparse dictionary learning methods, such as the non-negative variant of the K-SVD algorithm (NN-K-SVD) [1] have also been used. These methods are constrained by the selected learning order and encounter similar difficulties when applied to musical signals, such as single-spike atoms and dual-source atoms. We also observe the disappearance in T of sources with infrequent support or relatively low energy, particularly if there exists in the dictionary an atom representing a highly coherent source e.g. a higher or lower octave of the same note.

Musical signals contain many harmonic elements and this structure can be exploited. Harmonic atoms were proposed as groups of related Gabor atoms, sparse coded with Matching Pursuit for pitch tracking [3]. Similar works have used dictionaries learnt offline or harmonic atoms with constrained spectral envelopes. Harmonic signal decomposition methods have also been proposed, such as a state-of-the-art transcription method [2] using Bayesian harmonic NMF.

II. STRUCTURE-AWARE DICTIONARY LEARNING

We consider spectrogram decomposition with a priori knowledge of the number and individual structure of atoms, which we refer to as structure-aware dictionary learning. This knowledge is encoded in a binary matrix, $I \in \mathbb{R}_+^{M \times K}$, which indicates the harmonic peaks and sidelobes of each note. The signal decomposition now becomes

$$S \approx D'T \quad \text{where} \quad D'_{m,k} = I_{m,k} D_{m,k}. \quad (2)$$

We have implemented structure-aware versions of NN-K-SVD and NMF [4]. These differ only in filtering by I . In SA-NN-K-SVD, the atoms are filtered after their individual SVD updates. In SA-NMF, D is filtered after its multiplicative update. Using spectrograms

composed from a dictionary of synthetic harmonic atoms, experiments were performed in which we attempted to recover the original dictionary. It was found that dictionary recovery is significantly enhanced and accelerated using the structure-aware methods.

It was also observed that SA-NN-K-SVD outperformed SA-NMF, depending on the sparse coder used. Further experiments were performed with different spectrogram parameters to compare the structure-aware methods. First we skewed the distribution of the atoms. Again we found that SA-NMF was outperformed by the SA-NN-K-SVD. In another experiment, we randomised the shape of the atoms. This time we found that SA-NMF outperformed the SA-NN-K-SVD. These results led us to derive a method we refer to as SANNSMUDL (Structure Aware Non-Negative Sparse Multiplicative Update Dictionary Learning) using sparse coding to update T , and a multiplicative update for D . Results to date indicate that this method improves dictionary recovery relative to the SA-NN-K-SVD and SA-NMF.

III. CONCLUSION

We have derived a method which performs better than other methods in our experimental setup. The spectrograms were synthesized to be highly overlapping and we believe that this method may be generalizable when the atomic structure supports are relatively sparse. We aim to further test this method with transcription tasks, building on work presented in [5]. Further work will include learning structure from signals, so as to inform this method.

ACKNOWLEDGMENT

This research is supported by ESPRC Leadership Fellowship EP/G007144/1 and EU FET-Open Project FP7-ICT-225913 "SMALL".

REFERENCES

- [1] M. Aharon, M. Elad, and A. M. Bruckstein. K-SVD and its non-negative variant for dictionary design. In *Proceedings of the SPIE conference wavelets*, volume 5914, pages 327–339, July 2005.
- [2] N. Bertin, R. Badeau, and E. Vincent. Enforcing harmonicity and smoothness in bayesian non-negative matrix factorization applied to polyphonic music transcription. *IEEE Transactions on Audio, Speech and Language Processing*, 18(3):538–549, 2010.
- [3] R. Gribonval and E. Bacry. Harmonic decomposition of audio signals with matching pursuit. *IEEE Transactions in Signal Processing*, 51:101–111, 2003.
- [4] K. O'Hanlon and M. D. Plumley. Structure-aware dictionary learning with harmonic atoms. submitted to EUSIPCO 2011, 2011.
- [5] K. O'Hanlon and M. D. Plumley. Structured non-negative dictionary learning for music transcription. poster available online at http://small-project.eu/small-london-workshop/posters/SMAL_poster_3.pdf/view, Jan 2011.

Multi-Channel Analog-to-Digital (A/D) Conversion using Fewer A/D Converters than Channels

Ahmed H. Tewfik, Youngchun Kim
 Department of Electrical and Computer Engineering
 The University of Texas at Austin
 Email: {y-kim,tewfik}@austin.utexas.edu

B. Vikrham Gowreesunker
 Systems and Application R&D Center
 Texas Instruments Inc.
 Email: vikrham@ti.com

I. INTRODUCTION

Many applications in signal processing such as audio, physiological signals, and brain machine interfaces, require digitizing analog signals from multiple channels. However, designers of such systems are often faced with restrictions that limit their ability to use multiple analog-to-digital (A/D) converters. We approach the problem of multi-channel A/D conversion with the unique concept of using fewer A/D converters than channels. To the best of our knowledge, no such approach has been previously proposed.

A naïve approach to the problem involves modulating the analog signals so that they occupy non overlapping frequency bands and digitizing the sum of the modulated signal. The main drawback of such an approach is that it increases the frequency of operation of the sigma delta A/D converter, adding to power consumption. If several signals are multiplexed using such an approach, or if the bandwidth of the underlying signals is large, such an approach may lead to unfeasible frequency of operation for the sigma delta A/D converter.

In our solution to this problem, we deliberately mix the channels in such a fashion that we can later separate them. Given M channels of analog data, we generate N mixtures of the analog data such that $N < M$. The A/D conversion is done on the N mixtures. Finally the mixtures are separated into M digitized channels. We show that perfect separation of the input signals after A/D conversion is possible if all input signals are known to have sparse representations involving no more than a fixed number of atoms drawn from a known dictionary. Mixing is done by modulating and spreading some of the input signals so that the total bandwidth of the mixture is slightly larger than that of the original input signals. Under such a scenario, signals can be separated using any method for sparse signal representation. We quantify the amount of bandwidth expansion needed to achieve signal separation and also discuss the design of spreading sequences and dictionaries.

II. METHOD DESCRIPTION

We now describe one approach to mixing the signals, based on bandwidth expansion where we deliberately introduce redundancy in the mixture. Without loss of generality, we illustrate one example with two input analog channels and one A/D converter. Given two signals, $S_1(t)$ and $S_2(t)$, each with a bandwidth of interest of B Hz, we first pre-condition the signals followed by mixing and then A/D. The signals are first low-pass filtered to B Hz, and then a modulation is applied to one of the signals (say $S_2(t)$) so it is now shifted in the frequency space and spread over a wide frequency band. $S_1(t)$ and the modulated version of $S_2(t)$ are mixed together before an oversampled A/D is applied to the mixture. The modulation and mixing are illustrated in figure 1. By using a theoretical construction similar to the one used to establish the restricted isometry property in compressed sensing, we establish that by slightly expanding the

bandwidth of the mixture relative to that of the underlying input signals, it is possible to recover the underlying signals exactly after A/D conversion assuming that we have an algorithm that can recover the sparsest representation of any given signal. Specifically, for the two signal case illustrated here, we represent the digitized mixture signal using a union of the dictionary matched to the input signals and a spread version of that dictionary. Atoms in the spread version are obtained by spreading in discrete time the atoms in the original dictionary using the spreading sequence corresponding to the one applied to the input signal in the analog domain. Assuming that we can find the exact sparse representation of the mixture, signal separation is achieved by identifying the dictionary atoms selected to represent the mixture. Coefficients corresponding to spread atoms are associated with the signal that was spread before mixing. All other coefficients are associated with the signal that was not. We provide several examples of A/D conversion using synthetic sinusoidal data. We also present experimental results corresponding to digitizing two independent audio signals with a 3 KHz bandwidth that demonstrate an ability to achieve signal to noise ratios in the 55-60 dB range using a second order sigma delta A/D converter and the method of [1] and the references therein, for computing sparse signal representations.

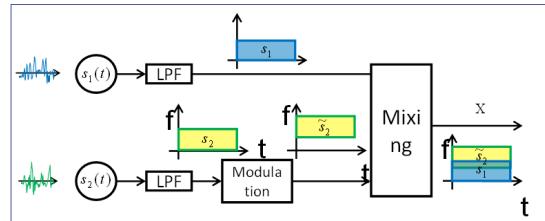


Fig. 1. Mixing analog signals using modulation.

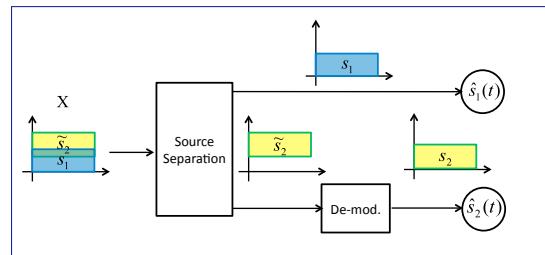


Fig. 2. Un-mixing analog signals using source separation and demodulation

REFERENCES

- [1] B.V. Gowreesunker and A. Tewfik, "Learning Sparse Representation using Iterative Subspace Identification," *Signal Processing, IEEE Transactions on*, vol. 58, no. 6, pp. 3055 -3065, june 2010.
- [2] A. Liutkus, R. Badeau, and G. Richard, "Informed Source Separation Using Latent Components," *(LVA/ICA'10)*

Practical Design of a Random Demodulation Sub-Nyquist ADC

Stephen Becker
 Applied & Computational Mathematics
 California Institute of Technology
 Pasadena, California 91125
 Email: sbecker@caltech.edu

Juhwan Yoo
 Mathew Loh
 Azita Emami–Neyestanak
 Electrical Engineering
 California Institute of Technology
 Pasadena, California 91125
 Email: {juhwan,rui,azita}@caltech.edu

Emmanuel Candès
 Mathematics Department
 Statistics Department
 Stanford University
 Stanford, California 94305
 Email: candes@stanford.edu

I. INTRODUCTION

The new field of compressed sensing (CS) [1] has stirred interest in designing hardware that samples at the *information rate*. After CS was introduced in 2004, hardware devices were immediately proposed [2]. CS has been used to speed up traditional imaging, such as MRI, and it has increased interest in general sparse approximation techniques; however, as of 2011, there are few hardware devices that use the principles of CS to implement incoherent measurements. There are two related reasons for this: most mathematicians who propose measurement schemes lack the understanding of what the engineering bottlenecks are, and secondly, the engineering requirements are quite non-standard since, for example, the SNR of the system is no longer a simple calculation.

One type of CS device that has been proposed is the random modulation pre-integrator (RMPI) [3]. The RMPI samples a wide-band signal, with up to 2.5 GHz bandwidth, using 8 independent channels, each with a 50 MHz ADC, so the total sampling rate is $12.5 \times$ lower than the Shannon-Nyquist limit. Because the low-rate sampling induces aliasing, each channel uses a pseudo-random bit sequence (PRBS) to spread the spectrum of the incoming signal, followed by an integration. See Figure 1.

We present an implementation of the RMPI in 90 nm CMOS. This is not just an abstract academic idea, but a real ADC device with 8 ENOB, and required significant engineering achievements. Our results are highly useful for teams working on other compressed sensing architectures. Specifically, we discuss

- **Design choices.** There are subtle issues involved in choosing the number of channels, the periodicity of the PRBS, and the integrator. Central to the design is a departure from the simplistic mathematical models.
- **Robust calibration.** Because the design does not exactly follow a model, it is necessary to calibrate the system in order to characterize its step response. In practice, calibration is difficult because the input signal is not known exactly. To overcome this, we introduce a method for phase-blind calibration.
- **Reconstruction techniques.** To maximize the performance of the system, many reconstruction techniques are needed. We discuss ℓ_1 analysis and synthesis formulations, how to perform reweighting, choice of dictionary, and windowing. We also alleviate fears raised by [4] that the system is sensitive to “off-grid” frequencies.

II. PREVIOUS WORK

RMPI devices have been studied by several groups [5]–[7], but a high-bandwidth device has yet to be manufactured and shown to

be functional. A related design, the modulated wide-band converter (MBC) [4], which follows the Xampling methodology, has been fabricated, and is predicted to reconstruct banded signals of up to 120 MHz from a 1 GHz bandwidth range using only 280 MHz overall sampling rate. The hardware prototype has been tested up to 1.6 MHz of bandwidth. Because the signal model consists of inputs with time-invariant statistics, it is possible to achieve arbitrarily high SNR of the frequency support by collecting more time samples.

Our goal is even more ambitious: to recover highly non-stationary signals, such as radar pulses which only occur for 200 ns. Because there are so few measurements, simple techniques such as OMP are not sufficient, and we use powerful variants of ℓ_1 recovery.

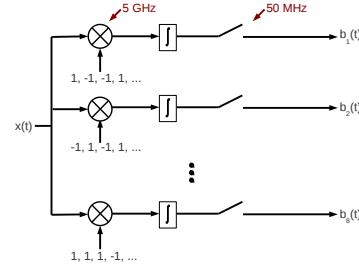


Fig. 1. Diagram of the multi-channel RMPI.

REFERENCES

- [1] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [2] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, “An architecture for compressive imaging,” in *Proc. Int'l. Conf. Image Proc.*, Atlanta, 2006.
- [3] W. Skones, B. Oyama, S. Stearns, J. Romberg, and E. Candès, “Analog to information (A-to-I), technical and management proposal,” August 2005, in response to DARPA BAA 05-35.
- [4] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, “Xampling: Analog to digital at sub-Nyquist rates,” *IET Cir. Dev. and Systems*, vol. 5, no. 1, pp. 8–20, Jan. 2011.
- [5] T. Ragheb, J. Laska, H. Nejati, S. Kirolos, R. Baraniuk, and Y. Massoud, “A prototype hardware for random demodulation based compressive analog-to-digital conversion,” in *51st Midwest symp. on circuits and systems, MWSCAS*, 2008, pp. 37–40.
- [6] X. Chen, Z. Yu, S. Hoyos, B. M. Sadler, and J. Silva-Martinez, “A sub-Nyquist rate sampling receiver exploiting compressive sensing,” *IEEE Trans. Circuits and Systems I: Regular Papers*, vol. 58, no. 3, pp. 507–520, March 2011.
- [7] J. A. Tropp, J. N. Laska, M. F. Duarte, J. Romberg, and R. G. Baraniuk, “Beyond Nyquist: Efficient sampling of sparse bandlimited signals,” *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 520–544, 2010.

Compressive Spectral Estimation Can Lead to Improved Resolution/Complexity Tradeoffs

Michael A. Lexa, Mike E. Davies, John S. Thompson
 Institute for Digital Communications
 The University of Edinburgh, Edinburgh, UK

SUMMARY

Compressed sensing (CS) has primarily two modes of acquiring measurements of sparse signals. One is by taking inner product measurements described by an underdetermined linear system of equations $y = Ax$, where $y \in \mathbb{R}^m$ represents the measurements gathered about a sparse signal $x \in \mathbb{R}^n$ of interest. In this setting, the matrix $A \in \mathbb{R}^{m \times n}$ is chosen to possess a particular property, namely the restricted isometry property, and the measurements are acquired by computing inner products between x and the rows of A . Alternatively, one can acquire CS measurements by sampling x at random locations (random point evaluations). In this case, an underdetermined linear system also relates the measurements to a higher dimensional representation, but the measurements are *acquired* differently—random samples are not acquired as inner products.

This work concerns a structured form of random sampling and proposes new method to directly recover finite resolution power spectral density (PSD) estimates of spectrally-sparse wide-sense stationary random processes. This method produces estimates at arbitrarily low sampling rates and can achieve better tradeoffs between system complexity (as defined below) and resolution than existing methods.

PSD estimate. Let $x(t)$ be a real valued, zero-mean wide-sense stationary random process with power spectral density function $P_{xx}(\omega)$. Assume $P_{xx}(\omega)$ is bandlimited to $W/2$ Hz and spectrally sparse, i.e., assume its support has Lebesgue measure that is small relative to the overall bandwidth.

For a fixed time interval $1/W$ and for a suitable positive integer L , consider sampling $x(t)$ at time instants $t = (nL + c_i)/W$ for $1 \leq i \leq q$, $n \in \mathbb{Z}^+$, where the time offsets c_i are distinct, positive real numbers less than L . Such a scheme exhibits an average sampling rate of qW/L Hz and can be implemented as a multichannel system where channel i shifts $x(t)$ by c_i/W and then samples uniformly at W/L Hz [1].

Let $r_{y_a y_b}(k)$ denote the cross correlation function of channels a and b . It can be then be shown that $r_{y_a y_b}(k)$ is related to $P_{xx}(\omega)$ through the linear equation,

$$r_{y_a y_b}(k) * h_{a,b}(k) \Big|_{k=0} = \frac{1}{2\pi} \sum_m e^{-i\frac{2\pi}{L}(c_a - c_b)m} \int_{-\pi W/L}^{\pi W/L} P_{xx}(\omega - 2\pi \frac{W}{L}m) d\omega,$$

where $*$ denotes convolution and $h_{a,b}(k)$ is the impulse response of an ideal fractional delay digital filter with delay

$(c_a - c_b)/W$. Denoting the integrals by $P_{xx}(m)$ and letting l index the $\binom{q}{2}+1$ combinations of pairs (a, b) (including $a = b$), we can form the linear system of equations, $u = \Psi v$, where the elements of $v \in \mathbb{R}^L$ are $P_{xx}(m)$, $[\Psi]_{l,m} = \frac{1}{2\pi} e^{-i\frac{2\pi}{L}(c_a - c_b)m}$, and $u \in \mathbb{R}^{q(q-1)/2+1}$ contains elements of the above convolutions evaluated at zero. $P_{xx}(m)$ equals the power of $x(t)$ in the m th spectral segment of $P_{xx}(\omega)$ of width $\pi W/L$. Collectively, the set $\{P_{xx}(m)\}$ forms a finite resolution approximation to $P_{xx}(\omega)$. The resolution is determined by L , the period of the random sampling pattern; larger L implies finer resolution. Based on the reasonable assumption that more channels implies higher hardware complexity, we take q , i.e. the number of channels, to be a measure of system complexity.

Improved tradeoffs through sparsity. Generally speaking, the above linear system can only be uniquely solved if $q(q-1)/2 + 1 \geq L$. But with the freedom to independently choose q and L , this inequality can be met at arbitrarily low sampling rates and at arbitrarily high resolutions. With fixed hardware complexity (fixed q), this relation poses a limitation on the resolution and sampling rate. For example, with $q = 8$ and $W = 1$ GHz, resolution is constrained to 125 MHz, a resolution too low for some applications in spectrum scanning.

With the assumption that $P_{xx}(\omega)$ is spectrally sparse, CS algorithms can tremendously improve this tradeoff, providing greater resolution for a given level of complexity. Because PSDs are nonnegative by definition and because the row of Ψ corresponding to $c_a - c_a$ is a row of 1's, we can avoid the more computationally intensive CS algorithms, like ℓ_1 minimisation, and simply seek a nonnegative least squares solution [2].

Note that even if the CS reconstruction is exact, what we recover is a stochastic estimate of the true PSD, regardless of the resolution. We are in essence trying to discover an average behaviour of a random process from (ultimately) a finite number of CS samples, and CS recovery algorithms have little bearing on this basic fact.

These results have direct application to radio frequency electromagnetic scanning problems and have potential application to radio interferometry. There is also a potential to extend this estimation approach to purely random sampling.

REFERENCES

- [1] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," *Proc. IEEE Inter. Conf. on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1688–1691, May 1996.
- [2] D. Donoho and J. Tanner, "Counting the faces of randomly-projected hypercubes and orthants, with applications," *Discrete and Computational Geometry*, vol. 43, no. 3, pp. 522–541, 2010.

K-SVD Dictionary-Learning for Analysis Sparse Models

Ron Rubinstein and Michael Elad

The Computer Science Department

The Technion – Israel Institute of Technology, Haifa 32000, Israel

Email: [ronrubin,elad]@cs.technion.ac.il

Abstract—The synthesis-based sparse representation model for signals has drawn a considerable interest in the past decade. Such a model assumes that the signal of interest can be decomposed as a linear combination of a few columns from a given dictionary. In our work we concentrate on an alternative, analysis-based model, where an analysis operator multiplies the signal, leading to a sparse outcome. Our goal in this work is to learn the analysis operator from a set of signal examples, and the approach taken is parallel and similar to the one adopted by the K-SVD algorithm. We demonstrate the effectiveness of the algorithm in several experiments, treating synthetic data and real images, showing a successful and meaningful recovery of the analysis operator.

I. BACKGROUND

Signal models are fundamental for handling various processing tasks, such as denoising, solving inverse problems, compression, sampling, and more. Among the many ways we have to model signals, one approach that has found a great popularity in the past decade is the synthesis-based sparse representation model. In this model, a signal $\mathbf{x} \in \mathbb{R}^d$ is modeled as being the outcome of the multiplication $\mathbf{x} = \mathbf{D}\alpha$, where $\mathbf{D} \in \mathbb{R}^{d \times m}$ is a dictionary – its columns are signal prototypes (atoms) that we use to compose the signal. We typically consider a redundant dictionary with $m > d$. The vector $\alpha \in \mathbb{R}^m$ is the signal’s representation, and a fundamental feature in this model is the expectation that it is sparse, i.e. $\|\alpha\|_0 = k \ll d$. This implies that the signals we work on can be composed as linear combinations of a few atoms from the dictionary [1].

The vast work on this model studied problems such as ways to estimate the representation from corrupted signals, theoretical guarantees for such estimates to recover an outcome that is close to the true solution, and ways to learn the dictionary \mathbf{D} from signal examples. Two popular techniques for this task are the MOD and K-SVD algorithms [2], [3], [4].

While the *synthesis* model has been intensively studied, there is an *analysis* viewpoint to sparse representations that has been left aside. The analysis model relies on a linear operator (matrix) $\Omega : \mathbb{R}^{p \times d}$. The key property of this model is our expectation that the coefficient vector $\Omega\mathbf{x} \in \mathbb{R}^p$ is expected to be sparse with ℓ zeros. These zeroes describe the subspace this signal belongs to. Assuming that $\text{spark}(\Omega^T) = d + 1$ (i.e. every set of d rows from Ω are linearly independent), the signals we model reside in a union of $(d - \ell)$ -dimensional subspaces. While this may sound similar to the synthesis counterpart approach, it is in-fact very different. Interestingly, relatively little is known about the analysis model, and little attention has been given to it in recent literature, compared to the synthesis model (see [6]).

In this paper we focus on the analysis model, and in particular, the development of an algorithm that would learn the analysis operator Ω from a set of examples $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$, so that the analysis coefficients $\Omega\mathbf{X}$ are sparse. Very little is known about this problem, and only recently work has started on this task [7], [8]. In this work we propose a novel algorithm that is parallel to the K-SVD in its rationale and computational steps. More on this work contribution is given hereafter.

II. OUR WORK CONTRIBUTION

Given the training set \mathbf{X} , we assume that every example is a noisy version of a pure analysis signal. Thus, $\mathbf{x}_i = \mathbf{z}_i + \mathbf{e}_i$, where \mathbf{e}_i is an additive noise $\|\mathbf{e}_i\|_2 \leq \epsilon$, and \mathbf{z}_i satisfies $\|\Omega\mathbf{z}_i\| = p - \ell$. Thus, our goal is to use the given set of examples to find both the clean signals $\{\mathbf{z}_i\}_i$ and the operator Ω , by solving the following optimization task:

$$\min_{\{\mathbf{z}_i\}_i, \Omega} \sum_i \|\mathbf{x}_i - \mathbf{z}_i\|_2^2 \quad \text{s.t.} \quad \|\Omega\mathbf{z}_i\|_0 = p - \ell. \quad (1)$$

Notice the resemblance between this goal and the one used for the synthesis model,

$$\min_{\{\alpha_i\}_i, \mathbf{D}} \sum_i \|\mathbf{x}_i - \mathbf{D}\alpha_i\|_2^2 \quad \text{s.t.} \quad \|\alpha_i\|_0 = k.$$

Similar to the solution adopted in the synthesis case, the solution of (1) is obtained in our work by iterating between an update of $\{\mathbf{z}_i\}_i$ and an update of Ω . Given the current Ω , the clean signals \mathbf{z}_i are found by a novel sparse-coding algorithm that greedily gather the zeros of the vector $\Omega\mathbf{z}_i$. Fixing these signals, the update of Ω is done row-by-row, by gathering for each row all the examples that are believed to be orthogonal to it, forming a matrix and computing the singular-vector that corresponds to its smallest singular-value. This resembles the synthesis K-SVD approach (with the difference that in the synthesis model we take the singular vector that corresponds to the largest singular-value). We demonstrate the effectiveness of the algorithm in several synthetic experiments and tests on natural images, showing a successful and meaningful recovery of the analysis operator in all these cases.

ACKNOWLEDGMENT

This work was supported by the European Commissions FP7-FET program, SMALL project (grant agreement no. 225913).

REFERENCES

- [1] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.
- [2] K. Engan, S. Aase, and J. Hakon-Husoy, Method of optimal directions for frame design, In Proceedings of ICASSP, pages 2443–2446, 1999.
- [3] M. Aharon, M. Elad, and A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, *IEEE Transactions on Signal Processing*, 54(11):4311–4322, November 2006.
- [4] R. Rubinstein, A. Bruckstein, and M. Elad, Dictionaries for sparse representation modeling, *IEEE Proceedings*, 98(6):1045–1057, April 2010.
- [5] M. Elad, P. Milanfar, and R. Rubinstein, Analysis versus synthesis in signal priors, *IOP Inverse Problems*, 23(3):947–968, June 2007.
- [6] S. Nam, M. Davies, M. Elad, and R. Gribonval, Cosaprsse analysis modeling – uniqueness and algorithms, In Proceedings of ICASSP - to appear, May 2010.
- [7] B. Ophir, M. Elad, N. Bertin, and M.D. Plumley, Sequential minimal eigenvalues – an approach to analysis dictionary learning, In Proceedings of EUSIPCO - to appear, September 2010.
- [8] M. Yaghoobi, S. Nam, R. Gribonval, and M.E. Davies, Analysis operator learning for overcomplete cosparse representations, In Proceedings of EUSIPCO - to appear, September 2010.

Analysis Operator Learning for Overcomplete Cosparse Representations

Mehrdad Yaghoobi[†], Sangnam Nam[‡], Remi Gribonval[‡] and Mike E. Davies[†]

[†] IDCom, AGB, KB, EH9 3JL, UK , Tel: +44 131 6505565 Fax: +44 131 6506554

[‡] INRIA, Centre Inria Rennes - Bretagne Atlantique, 35042 Rennes Cedex, France

E-mails: yaghoobi@ieee.org, sangnam.nam@inria.fr, remi.gribonval@inria.fr and mike.davies@ed.ac.uk

Abstract—We consider the problem of learning low-dimensional signal models from a collection of training samples. The mainstream approach would be to learn an overcomplete *dictionary* to approximate the training samples using sparse synthesis coefficients. This famous sparse model has a less well known counterpart, in analysis form, called the cosparse analysis model. In this new model, signals are characterized by their parsimony in a transformed domain using an overcomplete *analysis operator*. We propose to learn an analysis operator from a training corpus using a constrained optimization program based on L1 optimization. We derive a practical learning algorithm, based on projected subgradients, and demonstrate its ability to robustly recover a ground truth analysis operator, provided the training set is of sufficient size. A local optimality condition is derived, providing preliminary theoretical support for the well-posedness of the learning problem under appropriate conditions.

I. COSPARSITY AND COSPARSE ANALYSIS MODEL

In the cosparsity model [1], signal $\mathbf{y} \in \mathbb{R}^m$ is characterized by its parsimony in a transformed domain, using a given overcomplete transform $\Omega \in \mathbb{R}^{n \times m}$, $n > m$, called the *analysis operator*. In this setting, the concept of sparsity is slightly different to the standard definition of sparsity, as the number of zero elements in $\mathbf{z} = \Omega\mathbf{y}$, $p = n - \|\mathbf{z}\|_0$ has a more important role in analyzing the model, and it has been named *cosparsity* [1].

II. ANALYSIS OPERATOR LEARNING (AOL)

When a set of samples $\mathbf{Y} = [\mathbf{y}_i]_{i \in \mathcal{I}}$, is given, a question is how can we choose a suitable analysis operator Ω , which provides the highest cosparsity for \mathbf{Y} ? This is the central problem we consider.

The standard approach for many similar model adaptation problems, is to define a relevant optimization problem such that its optimal solution promotes maximal sparsity of $\mathbf{Z} := \Omega\mathbf{Y}$. A convex sparsity promoting penalty $f(\Omega)$ is the sum of absolute values of \mathbf{Z} , i.e. $f(\Omega) = \|\Omega\mathbf{Y}\|_1$. Unconstrained minimization of $f(\Omega)$ has some trivial solutions: a solution for such a minimization problem is $\Omega = \mathbf{0}$! A suggestion to exclude such trivial solutions is to restrict the solution set to an admissible set \mathcal{C} and reformulate AOL as, e.g.

$$\min_{\Omega} \|\Omega\mathbf{Y}\|_1 \text{ s. t. } \Omega \in \mathcal{C} \quad (1)$$

It is crucial to make a clever choice of the constraint for the problem (1) to exclude such deficient solutions. After explaining why some standard constraints are not enough, we propose a combined constraint, which is the Uniform Normalized Tight Frame (UNTF) [2].

III. PROJECTED SUBGRADIENT ALGORITHM FOR AOL

Subgradient methods have often been used to minimize convex objectives, when the solution is sought only with a few significant figures. As the problem is here constrained, we use the projected

This work is supported by EU FP7, FET-Open grant number 225913 and EPSRC grant EP/F039697/1.

subgradient method. The subgradient of the objective is simply $\partial f(\Omega) = \overline{\text{sgn}}(\Omega\mathbf{Y})\mathbf{Y}^T$, where $\overline{\text{sgn}}$ is the extended sign function.

Projection of an operator onto the set of uniform normalized frames can be easily found by renormalizing the columns of the operator. Projection of a full rank matrix onto the tight frame manifold is also easy and can be done using a singular value decomposition of the linear operator [3].

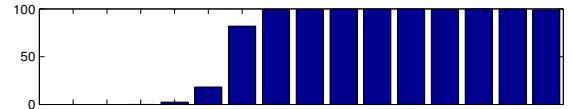
A point on the intersection of the uniformly normalized set and the set of tight-frames, which is the proposed UNTF constraint set, can often be found by alternatingly projecting onto these sets. Note that, there is no guarantee for convergence to an UNTF using this method, but this technique practically works very well [3]. As the projected subgradient continuously changes the current point, which needs to be projected onto the UNTF's, we only use a single pair of projections at each iteration of the algorithm. In practice the solutions seem to converge to UNTF's. A pseudocode of this algorithm is as follows,

```

1: initialization:  $k = 1$ ,  $K_{max}$ ,  $\Omega^{[0]} = \mathbf{0}$ ,  $\Omega^{[1]} = \Omega_{in}$ ,  $\gamma$ ,  $\epsilon \ll 1$ 
2: while  $\epsilon \leq \|\Omega^{[k]} - \Omega^{[k-1]}\|_F$  and  $k \leq K_{max}$  do
3:    $\Omega_G = \partial f(\Omega^{[k]})$ 
4:    $\Omega^{[k+1]} = \mathcal{P}_{TF} \left\{ \mathcal{P}_{UN} \left\{ \Omega^{[k]} - \gamma \Omega_G \right\} \right\}$ 
5:    $k = k + 1$ 
6: end while
7: output:  $\Omega_{out} = \Omega^{[k-1]}$ .
```

IV. EMPIRICAL EVIDENCE

A pseudo-random UNTF operator $\Omega_0 \in \mathbb{R}^{24 \times 16}$ was used to generate $l = 768$ training samples, with different cosparsities, by randomly selecting a normal vector in the orthogonal complement space of p randomly selected rows of Ω_0 . We started the simulation with a different pseudo-random admissible Ω_{in} , iterated 50000 times, and the average recovery of the rows of Ω_0 , for different cosparsities and 100 trials, is shown below as a function of the cosparsity of the signals.



REFERENCES

- [1] S. Nam, M. Davies, M. Elad, and R. Gribonval, “Cosparse analysis modeling- uniqueness and algorithms,” in *IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011.
- [2] M. Yaghoobi, S. Nam, R. Gribonval, and M. Davies, “Analysis operator learning for overcomplete cosparse representations,” submitted to the European Signal Processing Conference (EUSIPCO), August, 2011.
- [3] J. Tropp, I. Dhillon, R. Heath Jr., and T. Strohmer, “Designing structural tight frames via an alternating projection method,” *IEEE Trans. on Information Theory*, vol. 51, no. 1, pp. 188–209, 2005.

Learning hybrid linear models via sparse recovery

Eva L. Dyer, Aswin C. Sankaranarayanan, Richard G. Baraniuk
Dept. of Electrical and Computer Engineering, Rice University, Houston, TX 77005

Abstract—We introduce new methods to tackle the problem of *hybrid linear learning*—learning the number and dimensions of the subspaces present in a collection of high-dimensional data and then determining a basis or overcomplete dictionary that spans each of the subspaces. To do this, we pose this problem as the estimation of a set of points on the Grassmannian manifold $\mathcal{G}(k, n)$, i.e., the collection of all k -dimensional subspaces in \mathbb{R}^n . In order to estimate the subspaces present in the data, we couple the use of consensus methods for robust parameter estimation with sparse recovery for intelligent selection of sample sets. We demonstrate that the coupling of these two ideas is essential for: (1) extending the performance of current subspace learning methods to settings where shared structures exist between subspaces, as well as (2) providing concrete guarantees that specify when consensus will be formed on the true subspace structures present in the data in polynomial time. We demonstrate the utility of these methods for segmenting multispectral images and learning block-sparse dictionaries.

I. INTRODUCTION

Linear and nonlinear dimensionality reduction techniques (PCA and Isomap respectively), are typically founded on the assumption that global—and in some cases smooth—geometric structure exists amongst a collection of high-dimensional point cloud data. However, in many settings where ensembles of data arise from either heterogeneous sources or are measured under variable observational conditions, the resulting data exhibit geometric structure that cannot be succinctly described by a global model. Instead, datasets of this nature admit *hybrid geometric structure* that are best described by a union of distinct linear and/or nonlinear structures.

One important instance of a hybrid geometric model is referred to as a *hybrid linear model* (HLM). As the name suggests, a HLM can be used to model data that live on a union of linear structures; in the case where each of these structures passes through the origin, we say that these signals live on a *union of subspaces*. In addition to k -sparse signals that live on a combinatorial union of $\binom{n}{k}$ k -dimensional subspaces, a wide variety of datasets have been shown to manifest union of subspace models (or live on unions of affine planes), including scenes under various illumination conditions [1], image ensembles [2], and the trajectories of multiple moving objects [3].

Union of subspace models have been employed in a wide-range of applications in signal processing, machine learning, and computer vision, including the representation and denoising of sparse signals and also in morphological components analysis for signal separation. Even more recently, HLM's have been leveraged for the sampling and recovery of structured sparse signals [4], [5] from incomplete or compressive measurements. However, in all of these settings, the generative model for the data must be known *a priori*, i.e., the collection of orthonormal bases or overcomplete dictionaries that gave rise to the data must be known.

In this work, we tackle the problem of learning HLMs directly from point cloud data that are assumed to lie on a finite union of subspaces. We introduce new methods for learning HLMs or learning the number and dimensions of the subspaces present in a collection of point cloud data and then determining a basis or overcomplete dictionary that spans each of the subspaces. Whereas previous formulations of the subspace learning problem aim to segment the data into disjoint sets

according to the subspace membership of each point, we pose the problem of learning HLMs as the estimation of a set of points on the Grassmannian manifold $\mathcal{G}(k, n)$, where each point on the Grassmannian manifold represents a k -dimensional subspace in \mathbb{R}^n .

To find a robust estimate of the subspaces present in data, we employ a consensus-based approach [6] which attempts to find agreement (consensus) upon a subspace estimates across multiple sample sets selected from the data. In contrast to previous consensus-based approaches for subspace learning that either employ sample sets selected at random [7] or more sophisticated sample selection procedures [8], we propose the use of sparse recovery for efficient sample set selection. This is done by finding a sparse representation of each of our data points with respect to the overcomplete dictionary consisting of the remaining points in the dataset as in [9]. Following this subset selection procedure, we determine the mapping of the span of each of the support set onto $\mathcal{G}(k, n)$ and then look for the estimates that agree across multiple points in the dataset.

In order to study our proposed approach, we extend standard analyses for sparse approximation algorithms [10] to the case where sparse representations must be formed from a union of overcomplete sub-dictionaries where each sub-dictionary spans a low-dimensional subspace. We show that under certain conditions on the principal angles between the subspaces in our HLM, we can guarantee that both orthogonal matching pursuit (OMP) and basis pursuit (BP) will recover a sufficient number of sample sets that will in turn yield correct estimates of the true subspaces that generated our data.

Although our proposed method wed two different existing ideas that are currently employed in subspace learning, we demonstrate that the coupling of these two ideas is essential for: (1) extending the performance of state-of-the-art subspace clustering methods [9] to settings where shared/overlapping structures exist between subspaces, as well as (2) providing guarantees that describe when consensus will be formed in polynomial time. Following our analysis of sample set selection with sparse recovery methods, we demonstrate the utility of these methods for segmenting multispectral images and learning block-sparse dictionaries.

REFERENCES

- [1] R. Basri and D. Jacobs, *Lambertian Reflectance and Linear Subspaces*. IEEE Trans. PAMI, 25 (2), pp. 218–233, Feb 2003.
- [2] R. Garg, H. Du, et al., *The Dimensionality of Scene Appearance*. ICCV 2009.
- [3] J. Yan and M. Pollefeys, *A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate*. ECCV 2006.
- [4] T. Blumensath, M. Davies, *Sampling theorems for signals from the union of finite-dimensional linear subspaces*. IEEE Trans. on Info. Theory, 55 (4), April 2009.
- [5] R. Baraniuk, V. Cevher, M. Duarte and C. Hegde, *Model-Based Compressive Sensing*. IEEE Trans. on Info. Theory, 56, pp. 1982–2001, April 2010.
- [6] M. Fischler and R. Bolles, *Random Sample Consensus*. Comm. of the ACM, 24, pp. 381395, June 1981.
- [7] A. Yang, S. Rao, and Y. Ma, *Robust Statistical Estimation and Segmentation of Multiple Subspaces*. CVPR Workshop 2006.
- [8] T. Zhang, A. Szlam, Y. Wang, G. Lerman. *Hybrid Linear Modeling via Local Best-fit Flats*. Arxiv preprint, 2010.
- [9] E. Elhamifar and R. Vidal, *Clustering disjoint subspaces via sparse representation*. ICASSP, 2010.
- [10] J.A. Tropp, *Greed is good: Algorithmic results for sparse approximation*. IEEE Trans. Inform. Theory, vol. 50, num. 10, pp. 2231-2242, Oct. 2004.

Evaluating Dictionary Learning for Sparse Representation Algorithms using SMALLbox

Ivan Damnjanovic, Matthew E. P. Davies and Mark D. Plumbley
School of Electronic Engineering and Computer Science
Queen Mary University of London
Mile End Road, London, E1 4NS, UK
Email: {name}.{surname}@eecs.qmul.ac.uk

Abstract—SMALLbox is an open source MATLAB toolbox aiming at becoming a testing ground for the exploration of new provably good methods to obtain inherently data-driven sparse models, which are able to cope with large-scale and complicated data.

I. SMALLBOX - EVALUATION FRAMEWORK

The field of sparse representations has gained a huge interest in recent years, in particular in applications such as compressed sensing, image de-noising and source separation. We are witnessing a growing number of sparse representation algorithms that are becoming freely available in the research community [1-2]. This growth raised a necessity for an environment for proper testing and benchmarking. The SPARCO framework [3] partially addresses this problem by providing a large collection of imaging, signal processing, compressed sensing, and geophysics sparse reconstruction problems for testing these algorithms. It also includes a large library of operators that can be used to create new test problems.

Sparse representation approaches find the sparse solution in a given dictionary, but give suboptimal solution in many scenarios in which no suitable model is known. Many algorithms exist that aim to solve the sparse representation dictionary learning problem [4-5]. The main driving force for this work is the lack of a toolbox such as SPARCO for dictionary learning problems. Recognising the need of the community for such a toolbox, we set out to design SMALLbox - a MATLAB toolbox with three main aims:

- to enable an easy way of comparing dictionary learning algorithms,
- to provide a unifying API that will enable interoperability and re-use of already available toolboxes for sparse representation and dictionary learning,
- to aid the reproducible research effort in sparse signal representations and dictionary learning.

To enable re-use of already developed problems from SPARCO, the main interoperability is given through the “Problem” structure which in SMALLbox can be defined either as a sparse representation or dictionary learning problem. In generating a problem, some of the utilities can be used to decode a dataset and prepare a test signal or a training set for dictionary learning. The dictionaries can be either defined or learned using dictionary learning algorithms. In the former case, they can be given as implicit dictionaries, as a combination of the given operators and structures, or explicitly in the form of a dictionary matrix. In the latter case, they are learned from training data. Once the dictionary is set in the problem, the problem is ready to be solved by one of the sparse representation algorithms.

SMALLbox has been designed to enable an easy exchange of information and a comparison of different modules developed through a unified API structure. The structure was made to fulfil two main

goals. The first goal is to separate a typical sparse signal processing problem into three meaningful units:

- a) problem specification (preparing data for learning the structures, representation and reconstruction),
- b) dictionary learning (using a prepared training set to learn the natural structures in the data) and
- c) sparse representation (representing the signal with a pre-specified or learned dictionary).

The second goal is to provide a seamless connection between the three types of modules and ease of communication of data between the problem, dictionary learning and sparse representation parts of the structure. To achieve these goals, SMALLbox provides a “glue” structure to allow algorithms from different toolboxes to be used with a common API.

The SMALLbox evaluation framework is implemented as a MATLAB toolbox, which can be downloaded from <http://small-project.eu> and is in the form of an archive containing the SMALLbox directory structure and necessary MATLAB scripts. To enable easy comparison with the existing state-of-the-art algorithms, installation scripts will download third party toolboxes as required. In addition, the code is well documented with examples giving step-by-step instructions of how to implement new problems or introduce new sparse-representation and dictionary learning algorithms to the toolbox. These examples are built upon the set of test problems already implemented in SMALLbox. Since SMALLbox is an evaluation framework of the EU FET SMALL project, more problems, solvers and dictionary learning techniques that are developed will be included in SMALLbox as the project proceeds.

ACKNOWLEDGMENT

This research is supported by EU FET-Open ProjectFP7-ICT-225913 “SMALL”, EPSRC Platform Grant EP/045235/1, and EPSRC “Sustainable Software for Digital Music and Audio Research” Grant EP/H043101/1.

REFERENCES

- [1] D. Donoho, V. Stodden and Y. Tsaig, *Sparselab*, 2007, <http://sparselab.stanford.edu/>
- [2] T. Blumensath and M. E. Davies, *Gradient pursuits*, In IEEE Transactions on Signal Processing, vol. 56, no. 6, pp. 2370-2382, June 2008.
- [3] E. v. Berg, M. P. Friedlander, G. Hennenfent, F. Herrmann, R. Saab and O. Yilmaz *Sparco: A testing framework for sparse reconstruction*, In ACM Trans. on Mathematical Software, 354:1-16, February 2009.
- [4] R. Rubinstein, M. Zibulevsky and M. Elad, *Double Sparsity: Learning Sparse Dictionaries for Sparse Signal Approximation*, In IEEE Transactions on Signal Processing, Vol. 58, No. 3, Pages 1553-1564, March 2010.
- [5] K. Skretting and K. Engang, *Recursive Least Squares Dictionary Learning Algorithm*, In IEEE Transactions on Signal Processing, Vol 58, no4, 2010.

A Reproducible Research Framework for Audio Inpainting

Amir Adler and Michael Elad
Computer Science Department - The Technion
Haifa 32000, Israel
E-mail: {adleram,elad}@cs.technion.ac.il

Valentin Emiya and Rémi Gribonval
INRIA Rennes - Bretagne Atlantique
35042 Rennes Cedex, France
Email: firstname.lastname@inria.fr

Maria G. Jafari and Mark D. Plumbley
Queen Mary University of London
London E1 4NS, U.K.
Email: firstname.lastname@elec.qmul.ac.uk

Abstract—We introduce a unified framework for the restoration of distorted audio data, leveraging the Image Inpainting concept and covering existing audio applications. In this framework, termed Audio Inpainting, the distorted data is considered missing and its location is assumed to be known. We further introduce baseline approaches based on sparse representations.

For this new audio inpainting concept, we provide reproducible-research tools including: the handling of audio inpainting tasks as inverse problems, embedded in a frame-based scheme similar to patch-based image processing; several experimental settings; speech and music material; OMP-like algorithms, with two dictionaries, for general audio inpainting or specifically-enhanced declipping.

I. INTRODUCTION

Inpainting is a task proposed in the field of image processing: a set of missing pixels is reconstructed from the other reliable pixels of the image. Inpainting can be generalized as a problem of missing data estimation and techniques for image inpainting can be adapted to inpainting of other kinds of signals: one observes a partial set of reliable data while the remaining unreliable data is considered missing and is estimated from the reliable data. In particular, we consider Audio Inpainting [1] as a general task that covers a family of audio applications, including click removal, declipping, packet loss concealment and several applications for the restoration of time-frequency coefficients. We present works for audio inpainting in the time-domain [1], [2] and provide contributions on how to process audio signals in this context, which applicative scenarios and benchmarks are worth addressing and how sparse representations can solve those problems efficiently.

II. AUDIO INPAINTING IN TIME DOMAIN

A. Global and local formulation of Audio Inpainting

Let us consider a vector $\mathbf{s} \in \mathbb{R}^L$ of audio data. We only observe a subset of reliable samples $\mathbf{y}^r = \mathbf{M}^r \mathbf{s}$, where $\mathbf{y}^r \in \mathbb{R}^{L'}$, $L' < L$ and \mathbf{M}^r is the so-called measurement matrix obtained from the $L \times L$ identity matrix by selecting the rows associated with the observed reliable coefficients in \mathbf{s} . The audio inpainting problem is defined as the recovery of the original signal \mathbf{s} based on the knowledge of:

- 1) the reliable data \mathbf{y}^r ,
- 2) the support of the missing data (or, equivalently, \mathbf{M}^r),
- 3) additional information about the observed signal,
- 4) and, optionally, information about the missing data (e.g. in the case of clipping below).

As in many audio processing tasks and similarly to patch-based image processing, the signal can be locally modeled and processed: it is segmented into frames; each frame is then inpainted; the full restored signal is finally synthesized using an overlap-add method. Thus, the above global formulation of the inpainting problem can be straightforwardly translated locally at the frame level.

B. Audio Inpainting Problems

We propose **several scenarios or Problems** in which new inpainting algorithms can be compared against existing ones. They are related to speech or music restoration in different applications.

1) *Isolate-sample-to-large-hole Problem*: audio signals are degraded by periodically removing N_{miss} samples and performance are assessed as a function of N_{miss} . Small values of N_{miss} represent the click removal problem while large values of N_{miss} are simulating the packet loss concealment problem.

2) *Missing-sample-topology Problem*: for a fixed number of missing samples N_{miss} in a frame, a segments of b consecutive missing samples must be inpainted, where $a \times b = N_{\text{miss}}$. The performance is then reported as a function of the hole size b .

3) *Declipping Problem*: the missing samples are those beyond the clipping level θ_{clip} , such that the observation at time t is $\mathbf{y}^r(t) = \mathbf{s}(t)$ if $|\mathbf{s}(t)| < \theta_{\text{clip}}$, $\mathbf{y}^r(t) = \text{sign}(\mathbf{s}(t)) \theta_{\text{clip}}$ otherwise.

III. BASELINE DICTIONARIES AND SOLVERS

We propose sparsity-based approaches to address the Audio Inpainting problems described in Section II-B. **Two dictionaries** known to provide good models for audio waveforms are used: a discrete cosine transform dictionary, where phases are locked, and a free-phase Gabor dictionary. As a Solver, the inpainting version of the **OMP algorithm** is used to inpaint audio frames. We propose an enhancement for audio declipping, where the missing samples are constrained to have an amplitude beyond the clipping level.

IV. MATERIAL FOR REPRODUCIBLE RESEARCH

For reproducible-research purposes, we provided GPL Matlab code and Creative Commons data related to the presented works and arranged in a Problems/Dictionaries/Solvers architecture as in [3]:

- a series of Problems described in Section II-B, including experiment generation, result display and speech and music datasets;
- an analysis/synthesis scheme to address the Problems by just inserting any frame-level inpainting solver (see Section II-A);
- the Dictionaries and Solvers proposed in Section III.

ACKNOWLEDGMENT

This work was supported by the EU Framework 7 FET-Open project FP7-ICT-225913-SMALL: Sparse Models, Algorithms and Learning for Large-Scale data.

REFERENCES

- [1] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, “Audio Inpainting,” submitted to *IEEE Trans. Audio, Speech, and Lang. Proc.*, 2011, <http://hal.inria.fr/inria-00577079/en/>.
- [2] A. Adler, V. Emiya, M. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, “A Constrained Matching Pursuit Approach to Audio Declipping,” in *Proc. of ICASSP*, Prague, Czech Republic, May 2011.
- [3] I. Damnjanovic, M. Davies, and M. D. Plumbley, “SMALLbox - an evaluation framework for sparse representations and dictionary learning algorithms,” in *Proc. of LVA/ICA*, 2010, pp. 418–425.

GPU Accelerated Greedy Algorithms for Sparse Approximation

Jeffrey D. Blanchard

Department of Mathematics and Statistics
Grinnell College, Grinnell, Iowa 50112
jeff@math.grinnell.edu

Jared Tanner

School of Mathematics and the Maxwell Institute
University of Edinburgh, Edinburgh EH9 2PH
jared.tanner@ed.ac.uk

I. INTRODUCTION

Let $x \in \mathbb{R}^N$ be a target vector with at most k nonzero entries. We wish to recover the k -sparse vector x from the measurements $y = Ax \in \mathbb{R}^n$ where A is an $n \times N$ matrix. This can be framed as finding the solution to the intractable, combinatorial problem

$$\min \|x\|_0 \text{ subject to } y = Ax. \quad (1)$$

where $\|\cdot\|_0$ is the nonzero-counting measure. Under certain conditions, even simple iterative support recovery algorithms will return the exact solution to (1). In these situations, one wishes to employ an algorithm with guaranteed recovery capabilities but with low computational complexity. It is also well-known that linear programming can be used to solve the ℓ_1 -minimization problem as a convex relaxation of (1). While ℓ_1 -minimization has better theoretical and empirical guarantees on recovery than do greedy algorithms, it is often reported to be computationally more expensive.

II. GPU ACCELERATED GREEDY ALGORITHMS

With the introduction of graphical processing units (GPU) specifically designed for high performance computing, the computational burdens of solving (1) have been dramatically reduced. Lee and Wright [1] utilized this massively parallelized architecture to accelerate the SpaRSA algorithm which iteratively solves the ℓ_1 -minimization problem in lieu of (1). Building off of their work, we have implemented three greedy algorithms in this heterogeneous CPU-GPU computing environment, namely Hard Thresholding, Iterative Hard Thresholding (IHT) [2] and Normalized IHT (NIHT) [3]. These GPU-accelerated greedy algorithms running on an Nvidia Tesla C2050 demonstrate speedups of over 50 times a standard implementation executing on a state-of-the-art 6-core Intel Xeon 5650 CPU. The parallelized matrix multiplication transfers the computational burden to the support set identification step in each iteration of a greedy algorithm; these GPU-based algorithms employ modifications of the standard support identification techniques to exploit the advantages of the GPU.

III. EMPIRICAL WEAK PHASE TRANSITIONS

A motivating factor in developing the GPU accelerated greedy algorithms is the ability to perform large scale testing. Most greedy algorithms have a theoretical recovery guarantee based on the *restricted isometry property*. Comparing the efficacy of these theoretical results in terms of a strong phase transition curve, which separates the unit square into a region of guaranteed recovery and a region where recovery is not guaranteed, was performed in [4], [5]. These strong phase transitions are too pessimistic and these algorithms have resisted a formal average case analysis. An empirical weak phase transition was found in [6] although either the problem dimensions were small ($N \leq 4000$) or the number of tests was small (10 tests for

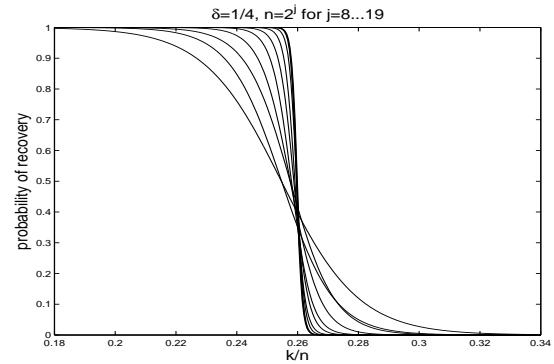


Fig. 1. Empirical Weak Phase Transition for NIHT using a random selection of rows from the DCT with $\delta = 0.25$.

$N = 2^{19}$). With the GPU-accelerated greedy algorithms, large-scale testing with large problem dimensions is now possible and reveals the behavior of the algorithms on large problems. For example, the problem dimensions can be scaled up to accurately determine the value of ρ which is the weak phase transition point; this value of ρ is higher than the 50% success point found by testing smaller problem dimensions. This is demonstrated in Fig. 1 when A is a random subset of rows from a DCT.

ACKNOWLEDGMENT

We thank NVIDIA corporation, in particular Chandra Chejji of NVIDIA Academic Research Programs, for the donation of the GPU hardware used to perform these calculations. J.D.B. was a National Science Foundation International Research Fellow at the University of Edinburgh funded by the award NSF OISE 0854991. J.T. acknowledges support from the Philip Leverhulme Foundation.

REFERENCES

- [1] E. Lee and S. Wright, "Implementing algorithms for signal and image reconstruction on graphical processing units," 2008.
- [2] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.
- [3] ———, "Normalised iterative hard thresholding: guaranteed stability and performance," *IEEE Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 298–309, 2010.
- [4] J. D. Blanchard, C. Cartis, and J. Tanner, "Compressed sensing: How sharp is the restricted isometry property?" *SIAM Review*, vol. 53, no. 1, pp. 105–125, 2011.
- [5] J. D. Blanchard, C. Cartis, J. Tanner, and A. Thompson, "Phase transitions for greedy sparse approximation algorithms," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 188–203, 2011.
- [6] D. L. Donoho and A. Maleki, "Optimally tuned iterative thresholding algorithms for compressed sensing," *IEEE Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 330–341, 2010.

Two Proposals for Robust PCA Using Semidefinite Programming

(Extended Abstract)

Michael McCoy

Applied and Computational Mathematics
The California Institute of Technology
Pasadena, CA 91125
Email: mccoy@caltech.edu

Joel A. Tropp

Applied and Computational Mathematics
The California Institute of Technology
Pasadena, CA 91125
Email: jtropp@acm.caltech.edu

I. INTRODUCTION

Principal component analysis (PCA) [1] is a common method for identifying structure in high-dimensional data. As principal components are defined by directions of high variance in the observations, PCA is highly sensitive to outliers in the data. This observation has led to many approaches to robust PCA [2]; unfortunately, these proposals are often based on intractable optimization problems or lack a principled foundation.

Here, we propose new approaches to the robust PCA problem that can be solved using semidefinite programs. The first method, *maximum mean absolute deviation rounding* (MDR), takes a greedy approach to the problem of finding directions of maximum spread. Our second proposal, the *low-leverage decomposition* (LLD), uses a convex optimization problem to split the observed data into a low-leverage matrix and a corruption matrix.

II. MAXIMUM MEAN ABSOLUTE DEVIATION ROUNDING

Suppose we have n observations \mathbf{x}_i , each of dimension p . Assuming that the data is centered, the top principal component is defined to be the vector that maximizes the empirical variance of the data, that is

$$\mathbf{v}_{\text{PCA}} = \arg \max_{\|\mathbf{v}\|_2=1} \sum_{i=1}^n |\langle \mathbf{x}_i, \mathbf{v} \rangle|^2 \quad (1)$$

Equivalently, \mathbf{v}_{PCA} is the dominant right singular vector of the $n \times p$ matrix \mathbf{X} whose rows correspond to the observations \mathbf{x}_i .

The squared inner-product in (1) may give overwhelming weight to outlying observations. Our proposal therefore replaces the squared inner-product with the magnitude of the inner-product:

$$\mathbf{v}_{\text{MD}} = \arg \max_{\|\mathbf{v}\|_2=1} \sum_{i=1}^n |\langle \mathbf{x}_i, \mathbf{v} \rangle|. \quad (2)$$

This type of approach to robust PCA has been studied in many works, e.g. [3], but there are no known algorithms for computing \mathbf{v}_{MD} with guarantees of efficiency or approximation quality. By recasting our proposal as an operator norm, we show that it is indeed computationally hard to compute even the *value* of the maximum in (2).

Despite the intractability of computing \mathbf{v}_{MD} exactly, we show that a semidefinite relaxation of the problem gives an good upper bound on the maximum value in (2), and moreover we give an efficient randomized method that determines a unit-norm vector \mathbf{v}_{MDR} such that

$$\sum_i |\langle \mathbf{x}_i, \mathbf{v}_{\text{MDR}} \rangle| \geq (1 - \varepsilon) \sqrt{\frac{2}{\pi} \sum_{i=1}^n |\langle \mathbf{x}_i, \mathbf{v}_{\text{MD}} \rangle|}$$

for any $\varepsilon > 0$, except with arbitrarily small probability. We also show how to compute additional components using a greedy orthogonal restriction method, and show that a state-of-the-art algorithm [4] can solve the problem in practice.

III. LOW-LEVERAGE DECOMPOSITION

Our second proposal comes from the formulation of the robust PCA problem as an optimal low-rank model for the data. Using well-known group sparsity and low-rank heuristics, we define a decomposition of the data matrix \mathbf{X} whose rows are the observations \mathbf{x}_i as the optimal point of the convex optimization program

$$\begin{aligned} & \text{minimize} && \sum_i \sigma_i(\mathbf{P}) + \gamma \sum_j \|c_j\| \\ & \text{subject to} && \mathbf{P} + \mathbf{C} = \mathbf{X} \end{aligned} \quad (3)$$

where the vectors c_i are the rows of \mathbf{C} and $\sigma_i(\mathbf{P})$ is the i th singular value of the matrix \mathbf{P} . This semidefinite program is similar to the rank-sparsity decomposition introduced in [5].

We show that the optimal point $(\mathbf{P}_*, \mathbf{C}_*)$ in (3) has the property that the leverage scores of the recovered observations \mathbf{P}_* are bounded above by γ^2 . Additionally, we describe an *alternating direction method of multipliers* (ADMM) algorithm that provides good results for practical computation of the LLD [6].

We conclude our work with some numerical examples that compare several popular robust PCA approaches against MDA and LLD. This talk is based on work of the authors [7].

REFERENCES

- [1] H. Hotelling, “Analysis of a complex of statistical variables into principal components,” *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933.
- [2] P. J. Huber and E. Ronchetti, *Robust statistics*, 2nd ed. Hoboken, New Jersey: Wiley, 2009.
- [3] R. A. Maronna, “Principal Components and Orthogonal Regression Based on Robust Scales,” *Technometrics*, vol. 47, no. 3, pp. 264–273, August 2005.
- [4] S. Burer and R. D. C. Monteiro, “A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization,” *Math. Program.*, vol. 95, no. 2, pp. 329–357, 2003.
- [5] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, “Rank-Sparsity Incoherence for Matrix Decomposition,” *preprint*, June 2009, arXiv:0906.2220.
- [6] Z. Lin, M. Chen, L. Wu, and Y. Ma, “The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices,” *Math. Program.*, *submitted*, 2009, arXiv:1009.5055.
- [7] M. McCoy and J. A. Tropp, “Two Proposals for Robust PCA using Semidefinite Programming,” *ArXiv e-prints*, Dec. 2010. [Online]. Available: <http://arxiv.org/abs/1012.1086>

Blind Source Separation of Compressively Sensed Signals

Martin Kleinsteuber* and Hao Shen*
 Geometric Optimization and Machine Learning Group
 Technische Universität München, Munich, Germany
 Email: kleinsteuber@tum.de, hao.shen@tum.de

Abstract—We present an approach to simultaneously separate and reconstruct signals from a compressively sensed linear mixture. We assume that the signals have a common sparse representation. The approach combines classical Compressive Sensing (CS) theory with a linear mixing model. Since Blind Source Separation (BSS) from a linear mixture is only possible up to permutation and scaling, factoring out these ambiguities leads to the problem of ℓ^1 -minimization over the so-called oblique manifold. We discuss the occurring cost function and propose a geometric conjugate subgradient method to solve the problem.

I. INTRODUCTION

The problem of recovering signals from only the mixed observations without knowing the priori information of both the source signals and the mixing process is often referred to as *Blind Source Separation* (BSS), cf. [1]. Different BSS methods are used in various challenging data analysis applications, such as functional Magnetic Resonance Imaging (fMRI) analysis and microarray analysis. In order to achieve reasonable performance, prominent methods, e.g. Independent Component Analysis (ICA), usually require a large number of observations [2]. Unfortunately, the availability of a large amount of data samples can not be guaranteed in many real applications, due to either cost or time issues.

The theory of compressed sensing (CS), cf. [3] shows that, when a signal is *sparse* (or *compressible*) with respect to some basis, only a small number of samples suffice for exact (or approximate) recovery. It is interesting to know that the concept of sparsity has also been used as a separation criterion in the context of BSS [4]. Although a family of efficient algorithms in the probabilistic framework are proposed therein, the scenario with compressively sensed samples has not been studied and thus differs from our approach. In this work, the authors are interested in separating sparse signals which are compressively sampled.

II. PROBLEM DESCRIPTION

For the sake of convenience of presentation, signals are represented as column vectors, instead of the conventional row vectors. The instantaneous linear BSS model is given as follows

$$Y = SA, \quad (1)$$

where $S = [s_1, \dots, s_m] \in \mathbb{R}^{n \times m}$ denotes the data matrix of m sources with n samples ($m \ll n$), $A = [a_1, \dots, a_k] \in \mathbb{R}^{m \times k}$ is the mixing matrix of full rank, and $Y = [y_1, \dots, y_k] \in \mathbb{R}^{n \times k}$ represents the k linear mixtures of S . Here, we consider the scenarios with $m \geq k$, i.e., the number of observed mixtures is *less than or equal to* the number of sources. The task of standard BSS is to estimate the sources S , given only the mixtures Y . We refer to [5] for more details.

We assume that all sources $s_i \in \mathbb{R}^n$, for $i = 1, \dots, m$, have sparse representations with respect to the same basis, i.e., given $\Psi \in \mathbb{R}^{n \times n}$ a basis of \mathbb{R}^n , referred to as *representation basis*, each source s_i is assumed to have a q_i -sparse representation with respect to Ψ , denoted by $x_i \in \mathbb{R}^n$, i.e.

$$s_i = \Psi x_i, \quad (2)$$

or more compactly as

$$S = \Psi X, \quad (3)$$

where $X = [x_1, \dots, x_m] \in \mathbb{R}^{n \times m}$.

Now let us take one step further to compressively sample each mixture $y_i \in \mathbb{R}^n$ individually by a *sampling basis* $\Phi_i \in \mathbb{R}^{p_i \times n}$ for $i = 1, \dots, k$. Then, a compressively sensed observation $\hat{y}_i \in \mathbb{R}^{p_i}$ of the i -th mixture is constructed as

$$\hat{y}_i = \Phi_i y_i = \Phi_i \Psi X a_i. \quad (4)$$

We refer to (4) as the *compressively sensed BSS* (CS-BSS) model.

The task of our work is then formulated as follows: Given the common presentation basis $\Psi \in \mathbb{R}^{n \times n}$ and the compressively sensed observations $\hat{y}_i \in \mathbb{R}^{p_i}$, for $i = 1, \dots, k$, together with their corresponding sampling bases $\Phi_i \in \mathbb{R}^{p_i \times n}$, estimate the mixing matrix $A \in \mathbb{R}^{m \times k}$ and the sparse representations $X \in \mathbb{R}^{n \times m}$. Following the well-known argument that the mixing matrix A is identifiable only up to a column-wise scaling and permutation, without loss of generality, we restrict the mixing matrix A onto the $m \times k$ oblique manifold $\mathcal{OB}(m, k)$, which is defined as

$$\mathcal{OB}(m, k) := \left\{ A \in \mathbb{R}^{m \times k} \mid \text{rk}(A) = k, \text{ddiag}(A^\top A) = I_k \right\}, \quad (5)$$

where I_k is the $k \times k$ identity matrix, and $\text{ddiag}(Z)$ forms a diagonal matrix, whose diagonal entries are those of Z .

It is unavoidable that, in real applications, observations \hat{y}_i are contaminated by noise. In other words, the equalities defined in (4) do not hold in general. In the sense of least squares error, we propose the following cost function

$$f: \mathcal{OB}(m, k) \times \mathbb{R}^{n \times m} \rightarrow \mathbb{R},$$

$$f(A, X) := \|X\|_1 + \sum_{i=1}^k \lambda_i \|\Phi_i \Psi X a_i - \hat{y}_i\|_2^2, \quad (6)$$

where the scalars $\lambda_i \in \mathbb{R}^+$ weigh the reconstruction error of each mixture individually, and balance these errors against the sparsity term $\|X\|_1$. In this work, we provide an analysis of the cost function (6) and propose a geometric conjugate gradient method. The performance of our proposed approach is investigated by numerical experiments.

REFERENCES

- [1] S. Haykin, *Unsupervised Adaptive Filtering. Vol. 1: Blind Source Separation*. Wiley-Interscience, 2000.
- [2] S. Bermejo, “Finite sample effects in higher order statistics contrast functions for sequential blind source separation,” *IEEE Signal Processing Letters*, vol. 12, no. 6, pp. 481–484, 2005.
- [3] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.
- [4] M. Zibulevsky and B. A. Pearlmutter, “Blind source separation by sparse decomposition in a signal dictionary,” *Neural Computation*, vol. 13, no. 4, pp. 863–882, 2001.
- [5] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

Finding Sparse Approximations to Extreme Eigenvectors: Generalized Power Method for Sparse PCA and Extensions

Peter Richtárik

School of Mathematics, The University of Edinburgh

Email: Peter.Richtarik@ed.ac.uk

Abstract—In the first part of this work, based on [2], we develop a new approach to sparse principal component analysis (sparse PCA). We propose four optimization formulations of the problem, aimed at extracting one or several sparse dominant components. While the initial formulations involve nonconvex functions, we rewrite them into the form of an optimization program involving maximization of a convex function on a compact set and propose and analyze a simple gradient method for solving it (generalized power method). We demonstrate numerically on a set of random and gene expression test problems that our approach outperforms existing algorithms both in quality of the obtained solution and in speed.

A natural extension of the ideas above allows us to construct a method for finding, simultaneously, *jointly sparse approximations* to the eigenvectors associated with the largest and smallest eigenvalues of a symmetric psd matrix. This problem is equivalent to the Compressed Sensing problem of finding bounds on the asymmetric Restricted Isometry constants with the additional new requirement for the respective sparse eigenvectors to be supported on the same set. We prove a result on the emergence of joint sparsity in the iterates of the method and show that in the non-penalized case, the iterates are identical to the normalized gradients of the iterates of the Cauchy steepest descent method applied to minimizing a convex quadratic function [1].

I. PRELIMINARIES

Let $A = [a_1, \dots, a_n] \in \mathbb{R}^{p \times n}$, with $p \ll n$. Let $\bar{\lambda}$ (resp. $\underline{\lambda}$) be the largest (resp. smallest) eigenvalue of $S = A^T A$. Fix $\gamma > 0$.

II. GENERALIZED POWER METHOD FOR SPARSE PCA

For simplicity, we focus here on the problem of finding a sparse approximation z_* to the eigenvector of S “corresponding” to $\bar{\lambda}$. That is, we seek a sparse unit-norm vector $z_* \in \mathbb{R}^n$ such that $\|Az_*\|_2$ is large. Consider the following optimization problem:

$$\max\{\|Az\|_2 - \gamma\|z\|_1 : \|z\|_2 \leq 1\}. \quad (1)$$

It turns out that the optimal solution z_* of (1) is given by

$$z_* = z/\|z\|_2, \quad z^{(i)} = \text{sign}(a_i^T x)[|a_i^T x| - \gamma]_+, \quad i = 1, \dots, n,$$

where x is solves the smooth convex maximization problem

$$\max_{\|x\|_2 \leq 1} \sum_{i=1}^n [|a_i^T x| - \gamma]_+^2. \quad (2)$$

Note that since $p \ll n$, the dimension of the search space is decreased enormously. It is easy to show that $\gamma \geq \|a_i\|_2 \Rightarrow z_*^{(i)} = 0$, and hence γ controls sparsity of the solution.

For problems of type (2), i.e., for maximization of a convex function f over a compact set Q , we propose the following simple gradient method: Choose $x_0 \in Q$ and for $k \geq 0$ iterate:

$$x_{k+1} \in \arg \max\{f(x_k) + \langle f'(x_k), y - x_k \rangle : y \in Q\} \quad (\text{GPM})$$

This is our main convergence result:

Theorem 1 ([2]). *Let f be convex, Q compact and $\{x_i\}$ be the iterates produced by GPM. Then*

$$\min_{0 \leq i \leq k} \max_{y \in Q} \langle f'(x_i), y - x_i \rangle \leq \frac{\max f^* - f(x_0)}{k+1}.$$

If, in addition, f is strongly convex with parameter $\sigma_f > 0$, the convex hull of Q is strongly convex with parameter σ_Q , and we define $\delta_f = \min\{\|s\|^* : s \in \partial f(x), x \in Q\}$, then

$$\sum_{k=0}^{\infty} \|x_{k+1} - x_k\|^2 \leq \frac{2(\max f - f(x_0))}{\sigma_Q \delta_f + \sigma_f}.$$

III. JOINTLY SPARSE MIN AND MAX EIGENVECTORS

Consider the following optimization problem:

$$\max\{x^T S y - \gamma\|(x, y)\|_1 : \|x\|_2 = \|y\|_2 = 1, x^T y = 0\}. \quad (3)$$

If $\gamma = 0$, the optimal value of (3) is $\frac{1}{2}(\bar{\lambda} - \underline{\lambda})$, and if x^*, y^* are the optimal solutions, then $p = (x^* + y^*)/\sqrt{2}$ and $q = (x^* - y^*)/\sqrt{2}$ are the maximal and minimal eigenvectors of S , respectively. Below we give a method for (approximately) solving (3) for $\gamma > 0$ and show that γ induces *joint sparsity* in x and y . Hence, the method is able to identify a small principal submatrix of S whose extreme eigenvalues are a good approximation to $\bar{\lambda}$ and $\underline{\lambda}$.

Let $y_\gamma(x)$ (resp. $x_\gamma(y)$) be the optimal solution of (3) for fixed x (resp. y). Fix unit-norm x_0 and consider the following method:

$$y_k = y_\gamma(x_k), \quad x_{k+1} = x_\gamma(y_k). \quad (\text{ADM})$$

Theorem 2. *Let $w \in \mathbb{R}^n$ with $\|w\|_2 = 1$, $u = Sw$, $L = \{tw : t \in \mathbb{R}\}$, $B = \{s : \|s + u\|_\infty \leq \gamma\}$ and*

$$\text{Opt} \stackrel{\text{def}}{=} \max\{u^T z - \gamma\|z\|_1 : \|z\|_2 = 1, w^T z = 0\}. \quad (4)$$

If L does not pass through the interior of B , then the solution of (4) is given by $z = d/\|d\|_2$, $\text{Opt} = \sqrt{\omega(t^)} = \|d\|_2$, where*

$$t^* \in \arg \min_t [\omega(t)] \stackrel{\text{def}}{=} \sum_{i=1}^n (|u^{(i)} + tw^{(i)}| - \gamma)_+^2,$$

$$d^{(i)} = \text{sign}(u^{(i)} + t^* w^{(i)})[|u^{(i)} + t^* w^{(i)}| - \gamma]_+, \quad i = 1, \dots, n.$$

This result gives conditions under which the operations in (ADM) can be performed efficiently (in a closed form).

Let $u_k = Sx_k$. We further show that

1) *validity result*: if $\gamma \leq \sqrt{\|u_0\|_2^2 - (u^T x_0)^2}/(\|x_0\|_1 + \sqrt{n})$, then the condition of Theorem 2 will hold for all ADM iterates,

2) *joint sparsity result*: any of the conditions (i) $\|A^T a_i\|_2 \leq \gamma$, $x_k^{(i)} = 0$, (ii) $|x_k^{(i)}| \leq (\gamma - |u_k^{(i)}|)/\sqrt{\gamma^2(n-4) + 2\gamma\|u_k\|_1 + \|u_k\|_2^2}$, implies $y_k^{(i)} = 0$.

REFERENCES

- [1] H. Akaike. On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method. *Annals of the Institute of Statistical Mathematics*, 11:1–16, 1959.
- [2] M. Journée, Y. Nesterov, P. Richtárik, and R. Sepulchre. Generalized power method for sparse principal component analysis. *Journal of Machine Learning Research*, 11:517–553, 2010.

Stable Embeddings of Time Series Data

Han Lun Yap and Christopher J. Rozell

School of Electrical and Computer Engineering

Georgia Institute of Technology, Atlanta, Georgia 30332–0250

Email: {yaphanlun, crozell}@gatech.edu

Sparsity models have revolutionized signal processing in several ways, including exciting results in the field of compressed sensing (CS). This notion of exploiting low-dimensional structure in high-dimensional signals have been successful for manifold-modeled signals as well [1]. One thus wonders if these ideas can further be extended to characterizing systems rather than just acquiring signals?

Suppose we have a dynamical system whose internal (often high-dimensional) system state $x(t) \in \mathbb{R}^N$ is only indirectly observed via a one-dimensional time series of measurements produced through an observation function $s(t) = h(x(t))$, where $h : \mathbb{R}^N \rightarrow \mathbb{R}$. Surprisingly, when the dynamical system has low-dimensional structure because the state is confined to an attractor \mathcal{M} of dimension d in the state space, Takens' Embedding Theorem [2] shows that information about the hidden state of this system can be preserved in the time series output data $s(t)$. Specifically, Takens defined the *delay coordinate map* $F : \mathbb{R}^N \rightarrow \mathbb{R}^M$ as a mapping of the state vector $x(t)$ to a point in the *reconstruction space* (\mathbb{R}^M) by taking M uniformly spaced samples of the past time series (with sampling interval T_s) and concatenating them into a single vector, $F(x(t)) = [s(t) \ s(t - T_s) \ \dots \ s(t - (M - 1)T_s)]^T$. Takens' main result states that (under a few conditions on T_s) for almost every smooth observation function $h(\cdot)$, F is an *embedding*¹ of \mathcal{M} when $M > 2d$. However, this guarantees that only the *topology* of the attractor is preserved, but not its *geometry*. Thus in the presence of noise, the robustness of any processing performed in the reconstruction space (e.g., dimensionality estimation) cannot be guaranteed.

Recent work in CS has highlighted the importance of well-conditioned measurement operators $\tilde{F} \in \mathbb{R}^{M \times N}$ to ensure the geometry of a low-dimensional signal family \mathcal{M} is preserved. In effect, if \tilde{F} satisfies the *Restricted Isometry Property* (RIP) of order d , which basically ensures the *stable embedding*² of d -sparse vectors into a lower dimensional space, then robust recovery of these sparse vectors from their measurements can be guaranteed. Here we present work done in [3], where we extend this notion by establishing sufficient conditions whereby the delay coordinate map F is a stable embedding of the state space attractor for linear systems with linear observations

functions³. The results we obtain contrast with the standard CS results in three principle ways. First, the conditioning of F cannot always be improved by taking more measurements, as some system/observation pairs will have a fundamental limit in how well the system geometry can be preserved. Second, the necessary number of measurements scales with the dimension of the attractor d but is independent of the dimension of the ambient space N . Third, the total number of measurements may in fact have to be larger than the system dimension ($M > N$) in order to make a particular conditioning guarantee.

To avoid these high-dimensional measurements, previous work has proposed filtering (typically lowpass) the time series data to obtain measurement vectors of a smaller size [2]. To be precise, if F is a delay coordinate map with M delays, then the filtering operation is represented by a matrix $B \in \mathbb{R}^{m \times M}$ such that the resulting measurement vectors of the system state $x(t)$ can be written as $B \cdot F(x(t))$ and we call the map $H = B \cdot F : \mathbb{R}^N \rightarrow \mathbb{R}^m$ the *filtered delay coordinate map*. We show that if B satisfies the RIP of order $O(d)$, where d is the dimension of the system attractor⁴ \mathcal{M} , then B is a stable embedding of $F(\mathcal{M})$. This comes from recent results in [4], which shows that if a matrix B satisfies the RIP of order $O(d)$, then by randomizing the signs of the columns of B , it also ensures a stable embedding of a manifold of dimension d . Thus even if we require M to be large to ensure a stable embedding of \mathcal{M} with a certain conditioning, further filtering the time series data with a well-chosen B ensures that H is also a stable embedding of \mathcal{M} with approximately the same conditioning but possibly with significantly fewer measurements m .

REFERENCES

- [1] R. G. Baraniuk and M. B. Wakin, “Random projections of smooth manifolds,” *Found. of Comp. Math.*, vol. 9, no. 1, pp. 51–77, 2009.
- [2] T. Sauer, J. A. Yorke, and M. Casdagli, “Embedology,” *J. Stat. Phys.*, vol. 65, no. 3/4, pp. 579–616, 1991.
- [3] H. L. Yap and C. J. Rozell, “Stable Takens’ Embeddings for Linear Dynamical Systems,” March 2011, submitted.
- [4] H. L. Yap, M. B. Wakin, and C. J. Rozell, “Stable manifold embeddings with operators satisfying the restricted isometry property,” in *Proc. Conf. Information Sciences and Systems (CISS)*, March 2011.

³A general stable embedding result for nonlinear dynamical systems is obviously of great interest. Our study on linear systems hopes to elucidate some of the unique issues that arise when trying to stabilize the embeddings of dynamical systems, helping to pave the way for extensions to nonlinear systems.

⁴Here, we have to assume that the system attractors is a low-dimensional manifold.

¹An *embedding* is a *one-to-one immersion*.

²We say that \tilde{F} is a stable embedding of \mathcal{M} of conditioning ϵ if for all $x, y \in \mathcal{M}$, $(1 - \epsilon) \leq \frac{\|\tilde{F}(x - y)\|_2}{\|x - y\|_2} \leq (1 + \epsilon)$.

Estimating multiple filters from stereo mixtures: a double sparsity approach

Simon Arberet
EPFL

simon.arberet@epfl.ch

Prasad Sudhakar and Rémi Gribonval
INRIA Rennes - Bretagne Atlantique
{firstname.lastname}@inria.fr

Abstract—We consider the problem of estimating multiple filters from convolutive mixtures of several unknown sources. We propose to exploit both the time-frequency (TF) sparsity of the sources and the sparsity of the mixing filters. Our framework consists of: a) a clustering step to group the TF points where only one source is active, for each source; b) a convex optimisation step, to estimate the filters using TF cross-relations that capture linear constraints satisfied by the unknown filters. Experiments demonstrate that the approach is well suited for the estimation of sufficiently sparse filters.

I. INTRODUCTION AND NOTATIONS

Given two convolutive mixtures $x_i = \sum_{j=1}^N a_{ij} * s_j$, $i = 1, 2$, we wish to estimate the mixing filters a_{ij} from the mixtures without the knowledge of the sources s_j .

II. CROSS-RELATIONS FOR BLIND FILTER ESTIMATION

In the single source setting and in the absence of noise, the so-called time-domain cross-relation holds. A traditional method to solve for the filters using it is to minimise $\|x_2 * a_1 - x_1 * a_2\|_2$ with a normalisation constraint on the filters [1] (as there is only one source, the source index is dropped on the filters). Denoting $\mathbf{B} := \mathcal{B}[x_1, x_2]$ a matrix built by concatenating Toeplitz matrices derived from the observed mixtures, this leads to the minimisation of $\|\mathbf{B} \cdot \mathbf{a}\|_2$ subject to $\|\mathbf{a}\|_2 = 1$ where \mathbf{a} is a concatenation of the vectorized unknown filters. The normalisation $\|\mathbf{a}\|_2 = 1$ is to avoid the trivial zero-vector solution. It can be replaced by $\|\mathbf{a}\|_1 = 1$ to seek sparse filters [2]. However, these approaches are non-convex and suffer from a shift ambiguity of the solution. Instead, we propose the following convex optimisation problem

$$\min_{\mathbf{a}} \|\mathbf{a}\|_1 \text{ s.t. } \|\mathbf{B} \cdot \mathbf{a}\|_2 \leq \epsilon \text{ and } \mathbf{a}_1(t_0) = 1 \quad (1)$$

where t_0 is an arbitrarily chosen time index. We show that the new problem no longer suffers from a shift ambiguity.

This work was supported in part by the French Agence Nationale de la Recherche (ANR), project ECHANGE (ANR-08-EMER-006) and by the EU FET-Open project FP7-ICT-225913-SMALL.

III. MULTIPLE SPARSE FILTER ESTIMATION

In the presence of multiple sources, the time-domain cross-relation does not hold anymore. We extend the cross-relation approach to multiple sources, assuming that: the sources are sparse in the TF domain; we know large enough TF regions where each source is the only one contributing to the mixtures.

Cross-relations in the TF domain. We propose two TF formulations (narrowband and wideband [3]) of the cross-relation. They result in an optimisation problem similar to (1) with a new matrix \mathbf{B}_{nb} or \mathbf{B}_{wb} , built from TF representations of the mixture. Each row of these matrices corresponds to a point in the TF plane.

Filter estimation from partial TF information. Assuming that the sources are mutually disjoint in the TF plane, we propose to build for each source a matrix extracted from \mathbf{B}_{nb} (resp. \mathbf{B}_{wb}) by keeping only the rows indexed by the set Ω_j of TF points where the j-th source is the only active one. We then solve the resulting optimisation problem to estimate the filters.

IV. EXPERIMENTS

The proposed framework combines a TF clustering step, to detect the regions Ω_j , with a convex optimisation step, to estimate the sparse filters associated to each source. An experimental evaluation of the proposed approach with real audio data shows that our approach outperforms standard ICA approaches for filter estimation when the filters are sufficiently sparse.

REFERENCES

- [1] G. Xu, H. Liu, L. Tong, and T. Kailath, “A least-squares approach to blind channel identification,” *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2982–2993, 1995.
- [2] A. Aissa-El-Bey and K. Abed-Meraim, “Blind simo channel identification using a sparsity criterion,” in *Proc. of SPAWC*, 2008, pp. 271 – 275.
- [3] S. Arberet, P. Sudhakar, and R. Gribonval, “A wideband doubly-sparse approach for MITO sparse filter estimation,” in *ICASSP’11*, 2011.

Well-posedness of the frequency permutation problem in sparse filter estimation with ℓ_p minimization

Alexis Benichoux¹, Prasad Sudhakar², Rémi Gribonval²

¹Université de Rennes I & IRISA UMR6074 - ²Centre de recherche INRIA Campus de Beaulieu 35042 Rennes Cedex

Abstract—A well-known issue in blind convolutive source separation is that the sources and filters are at best identifiable up to an arbitrary scaling and permutation at each frequency bin. We propose to exploit the sparsity of the filters as a consistency measure for correcting such permutations. We show that the permutation is well-posed, up to a global permutation, under appropriate sparsity hypotheses on the filters. A greedy combinatorial algorithm is proposed for permutation recovery. Its empirical performance shows that the time-domain sparsity of the filters allows to recover permutations much beyond theoretical predictions.

I. CONTEXT

Let $x_i[t]$ be M mixtures of N source signals $s_j[t]$, resulting from the convolution with a filter $a_{ij}[t]$ of length L such that:

$$x_i[t] = \sum_{j=1}^N (a_{ij} * s_j)[t], \quad 1 \leq i \leq M. \quad (1)$$

We consider the problem of estimating the matrix of filters $\mathbf{A} = (a_{i,j})$ from the mixtures, without knowledge about the sources. A standard approach is to formulate the problem in the Fourier domain: one needs to estimate $a_{ij}[\omega]$. This suffers from a well known ambiguity : without further assumption on either $a_{ij}[t]$ or $s_j[t]$, one can at best hope to find an estimation $\tilde{\mathbf{A}} = (\tilde{a}_{i,j})$ where for every frequency $\omega \leq L$ we have

$$\tilde{a}_{i,j}[\omega] = \lambda_j[\omega] a_{i\sigma_\omega(j)}[\omega], \quad (2)$$

with λ_j a scaling ambiguity and σ_ω a permutation ambiguity. Several methods [1] to exploit properties of either \mathbf{S} or \mathbf{A} solve these. Our focus here is on the use of the sparsity of \mathbf{A} in the time domain to find $\sigma_1 \dots \sigma_L \in \mathfrak{S}_N$, assuming the scaling $\lambda \in \mathbb{C}^L$ is solved. Of course we can at best hope to obtain uniqueness up to a global permutation of the columns of \mathbf{A} . We exploit [2, Th.6.2a] the ℓ_p quasi-norm $\|\mathbf{A}\|_p^p := \sum_{i,j} |a_{ij}[t]|^p$, $0 \leq p \leq 1$, as a consistency measure to solve the permutations.

II. THEORETICAL GUARANTEES

If the filters a_{ij} have disjoint supports, without further sparsity hypothesis, we show that permutations can only increase the ℓ_p norm.

Theorem 1 ([3]): Let $\Gamma_{ij} \subset \{1, \dots, L\}$ be the time domain support of a_{ij} . Suppose that for all i and $j_1 \neq j_2$ we have $\Gamma_{i,j_1} \cap \Gamma_{i,j_2} = \emptyset$. Then for $0 \leq p \leq 1$ we have $\|\mathbf{A}\|_p^p \leq \|\tilde{\mathbf{A}}\|_p^p$.

To obtain uniqueness guarantees, we now introduce assumptions on the sparsity $k := \max_{i,j} \|a_{i,j}\|_0$. We measure the permutation error for $0 \leq p \leq 1$ with

$$\Delta_p := \min_{\pi \in \mathfrak{S}_N} \max_{i,j} \left\| \{a_{i\pi(j)}[\omega] - \tilde{a}_{ij}[\omega]\}_{1 \leq \omega \leq L} \right\|_p. \quad (3)$$

For sparse filters, the true filters are the sparsest among all filters incurring sufficiently few permutations. The skilled reader will rightly sense the role of the ℓ_0 Fourier-Dirac uncertainty principle [4] in the following result.

This work was supported by the EU FET-Open project FP7-ICT-225913-SMALL

Theorem 2 ([3]): (i) If $1 \leq \Delta_0 \leq L/2k$, then $\|\mathbf{A}\|_0 \leq \|\tilde{\mathbf{A}}\|_0$.
(ii) If $\|\mathbf{A}\|_0 \geq \|\tilde{\mathbf{A}}\|_0$, then $\Delta_0 \geq \frac{L}{2k}$.

For prime L , the results hold with $L+1-2k$ instead of $\frac{L}{2k}$. The equality case implies that the filters are pathologically related to Dirac combs of step $\frac{L}{2k}$.

III. A COMBINATORIAL ALGORITHM

We perform minimisation iteratively by considering one frequency bin $1 \leq \omega \leq \frac{L}{2}$ at a time and choosing a permutation (in a combinatorial fashion) that minimises the ℓ^p norm locally, while keeping the other bins fixed. To preserve $a_{ij} = \tilde{a}_{ij}$, the same permutation is applied on the corresponding mirror frequency $L+1-\omega$. This iterative procedure is repeated over all frequency bins till the ℓ^p norms of the filters converges.

We conservatively consider the filters as successfully recovered when the SNR of the permutation corrected time-domain filters exceeds 200dB. Fig. 1 shows the phase transition diagram for filter recovery using the proposed algorithm for the number of sources $N = 4$, number of channels $M = 3$, length of individual filters $L = 1024$ and $p = 1$. White indicates guaranteed success, black is guaranteed failure.

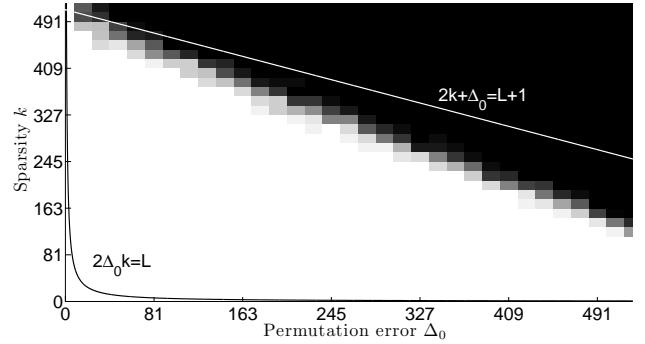


Fig. 1. Phase transition diagram for filter recovery by ℓ^1 minimisation.

The guarantees of Theorem 2 are delimited by the black line in general, and the white line if L is prime. We observe a phase transition close to the prime length case.

REFERENCES

- [1] M. Pedersen, J. Larsen, U. Kjems, and L. Parra, “A survey of convolutive blind source separation methods,” *Multichannel Speech Processing Handbook*.
- [2] P. Sudhakar, “Sparse models and convex optimisation for convolutive blind source separation,” Ph.D. dissertation, 2011.
- [3] P. Sudhakar, A. Benichoux, and R. Gribonval, “Well-posedness of the frequency permutation problem in sparse filter estimation with ℓ_p minimization,” *in preparation*.
- [4] M. Elad and A. Bruckstein, “A generalized uncertainty principle and sparse representation in pairs of bases,” *Information Theory, IEEE Transactions on*, vol. 48, no. 9, pp. 2558–2567, 2002.

Optical wave field reconstruction based on nonlocal transform-domain sparse regularization for phase and amplitude

Vladimir Katkovnik and Jaakko Astola

Department of Signal Processing, Tampere University of Technology (TUT), Tampere, Finland,
 vladimir.katkovnik@tut.fi, jaakko.astola@tut.fi

I. INTRODUCTION

Let us consider the following problem basic for optical wave field reconstruction. A wave field \mathbf{u}_0 at the *object* plane is modeled as complex-valued one characterized by phase and amplitude transmittance. The wave field propagation (blur operator) from the object to the parallel sensor plane is defined by the Rayleigh-Sommerfeld diffraction integral. The problem is to reconstruct the object plane distribution for both phase and amplitude from noisy complex-valued observations given at the sensor plane. The methods conventional in optics give rather blurred reconstructions and exhibit pronounced "waves", "wiggles" and "ringings". The optical diffraction defining the fundamental limitations on the spatial resolution of reconstructions is one of the main sources of these artefacts. In this paper we propose and develop a variational inverse imaging technique with the main motivation to wipe out the mentioned artefacts and obtain crisp imaging.

II. MAIN RESULTS

Being in line with the general formalism of the compressive sensing (CS), in particular for optical setup [1], [2], we propose an approach and algorithm which are different from the main stream in this field in three basic aspects. First, to deal with the complex-valued wave fields comprehensively we use modeling and regularization which are separate for phase and amplitude. The following equations link amplitude and phase of the object wave field with the corresponding transform (spectral) representations:

$$\text{mod}(\mathbf{u}_0) = \Psi_A \cdot \theta_A, \quad \text{angle}(\mathbf{u}_0) = \Psi_\varphi \cdot \theta_\varphi, \quad (1)$$

$$\theta_A = \Phi_A \cdot \text{mod}(\mathbf{u}_0), \quad \theta_\varphi = \Phi_\varphi \cdot \text{angle}(\mathbf{u}_0), \quad (2)$$

where θ_A and θ_φ are amplitude and phase spectra, respectively. The synthesis and analysis matrices $\Psi_A, \Phi_A, \Psi_\varphi, \Phi_\varphi$ are shown with the indices A and φ for amplitude and phase. The operations $\text{mod}(\mathbf{u}_0)$ and $\text{angle}(\mathbf{u}_0)$ applied to a vector give the vectors of amplitude and phase values. The equations (1) define what is called the synthesis giving the signal, amplitude ($\text{mod}(\mathbf{u}_0)$) and phase ($\text{angle}(\mathbf{u}_0)$), from the spectra θ_A and θ_φ . Contrary to it the analysis equations (2) give the spectra for the amplitude and for the phase of the object distribution.

Second, for modeling of phase and amplitude which can be spatially varying and continuous or discontinuous we use special bases functions, known as the BM3D-frames [3]. These frames provide rich overcomplete (large size) sets of functions which are data adaptive and nonlocal. BM3D filtering and BM3D-frames are recognized as a very efficient tool for various imaging problems [4], [5].

Third, while the conventional CS techniques use a single objective function to be optimized the algorithm developed in this paper

is based on a vector constrained optimization with two objective function minimized alternatively. It searches for a fixed-point giving a balance between two quality measures defined by the objective functions. This vector minimization decouples the inverse and filtering operations and results in the iterative algorithm simple in implementation and very efficient [4]. Convergence *in-small* to the fixed point is proved for this algorithm.

III. EXPERIMENTS

In Fig. 1 we show an example of the phase reconstruction obtained by the conventional technique (left column) and by the proposed algorithm (right column). The second row shows the cross-sections of the images shown in the first row. A nearly ideal chessboard phase modulation in the object plane is reconstructed by the proposed algorithm while the conventional technique gives the reconstruction severely damaged by multiple artifacts.

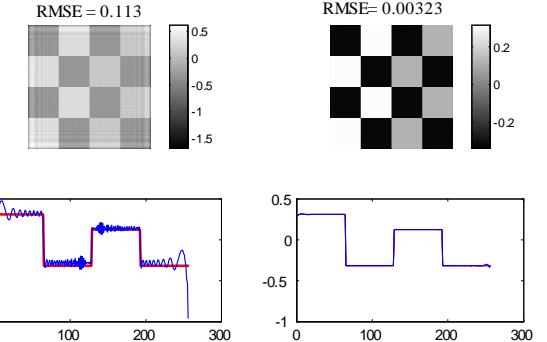


Fig. 1. Phase reconstruction for object with phase modulation. Left column obtained by the regularized inverse algorithm and the right column is obtained by the proposed algorithm with sparse amplitude and phase filtering.

REFERENCES

- [1] S. Gazit, A. Szameit, Y. C. Eldar, M. Segev, "Super-resolution and reconstruction of sparse sub-wavelength images," Optics Express **17**, 23920-23946 (2009). (2010).
- [2] D. J. Brady, K. Choi, D. L. Marks, R. Horisaki and S. Lim, "Compressive holography," Optics Express, **17**, 13040-13049 (2009).
- [3] A. Danielyan, V. Katkovnik and K. Egiazarian, "Image deblurring by augmented Lagrangian with BM3D frame prior," Workshop on Information Theoretic Methods in Science and Engineering (2010).
- [4] V. Katkovnik, A. Danielyan and K. Egiazarian, "Decoupled inverse and denoising for image deblurring: variational BM3D-frame technique," submitted to ICIP-2011 (2011).
- [5] K. Dabov, A. Foi, V. Katkovnik, and Egiazarian, K., "Image denoising by sparse 3D transform-domain collaborative filtering," IEEE Transactions on Image Processing, vol. **16**, 2080 - 2095 (2007).

Efficient sparse representation based classification using hierarchically structured dictionaries

Jort F. Gemmeke

Department of Linguistics, Radboud University, Nijmegen, The Netherlands.
Email: jgemmeke@amadana.nl

I. INTRODUCTION

Recently, it has been proposed to use sparse representation based classification (SRC) [1] for automatic speech recognition (ASR) [2]. In ASR with SRC, speech features are represented as a sparse linear combination of speech *exemplars*: speech features extracted from a training database token. With the exemplar dictionary atoms associated with classes, classification is done by using the weights of the activated dictionary atoms as evidence for the class of the observed speech token. In ASR however, the employed databases contain millions tokens, which makes it infeasible to use all available training tokens for the exemplar dictionary. Therefore, the employed dictionary is a subset of a few thousand atoms.

In this work, an iterative algorithm is proposed that can efficiently use an hierarchically structured dictionary containing hundreds of thousands of atoms. The algorithm works by, on each iteration, replacing dictionary atoms that have an increasing weight by their child-nodes. The overall size of the dictionary is kept down by merging atoms that have a decreasing weight.

II. METHOD

In previous work [2], a variant of Lee and Seungs iterative NMF algorithm [3] was used to obtain sparse representations:

$$\mathbf{x}_{i+1} \leftarrow \mathbf{x}_i \cdot * (\mathbf{A}^T (\mathbf{y} \cdot / (\mathbf{A} \mathbf{x}_i))) \cdot / (\mathbf{A}^T \mathbf{1} + \lambda). \quad (1)$$

with $\cdot *$ and $\cdot /$ denoting element-wise multiplication and division, respectively. The observed speech feature vector \mathbf{y} is of length E and the dictionary \mathbf{A} has dimensions $E \times N$. The sparse representation \mathbf{x}_i has length N and is indexed by iteration counter $i \in [1, I]$. The vector $\mathbf{1}$ is an all-one vector of length E . Applying update rule (1) minimises the generalised Kullback-Leibler (KL) divergence between \mathbf{y} and $\mathbf{A}\mathbf{x}$, with an L_1 norm controlling the sparsity through the constant λ .

The dictionary \mathbf{A} is a small subset of the complete dictionary $\hat{\mathbf{A}}$, the collection of \hat{N} exemplar tokens that comprises all available training material. In this work, first a hierarchical ordering of $\hat{\mathbf{A}}$ is found, so that each exemplar is either a leaf node, or is a parent of two exemplars. This hierarchy is obtained through a variant of hierarchical agglomerative clustering (HAC), which iteratively merges the two closest exemplars until all exemplars are merged. The difference with conventional HAC is that after merging, the parent node is represented by one of the child nodes, rather than by the mean of the two exemplars. This approach, reminiscent of K-medoid clustering, ensures an efficient clustering strategy (as distances only need to be computed once) and ensures that at any point in the hierarchy, the cluster nodes are still exemplars with associated state labels.

For the first iteration of our proposed method, an initial dictionary \mathbf{A}_0 is determined by taking the top $N_0 \ll \hat{N}$ exemplars as determined by the hierarchy. Then, (1) is applied with this initial dictionary \mathbf{A}_0 and initial sparse representation \mathbf{x}_0 to obtain the sparse representation \mathbf{x}_1 . The change in exemplar activation is determined as $\Delta\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_0$. If for a certain exemplar its $\Delta x > 0$, and it is not a leaf node, its child exemplar is added from the dictionary $\hat{\mathbf{A}}$ to the

TABLE I
RECOGNITION ACCURACIES FOR VARIOUS DICTIONARY SIZES

	Dictionary size N				
proposed	2000	4000	8000	16000	32000
95.9	77.9	83.7	89.9	92.3	95.5

dictionary \mathbf{A}_0 (recall that its other child is the exemplar itself). Both the exemplar and the added exemplar get weight $0.5 * x_1 > 0$. After processing all activated exemplars in this fashion, the dictionary \mathbf{A}_1 is obtained. This procedure is repeated on every subsequent iteration.

If the number of exemplars in the dictionary \mathbf{A} exceeds some threshold $N_i > M$, exemplars with $\Delta x < 0$ are sought. For these exemplars, it is checked whether the other child exemplar of its parent is also in the dictionary \mathbf{A} , and whether that child also has $\Delta x < 0$. If so, the two children are merged: effectively, one exemplar is removed as parent nodes are represented by one of the child nodes. The weights of the two exemplars are summed.

Both the merging and the splitting procedure are symmetric: for example if at a later iteration a previously merged exemplar obtains $\Delta x > 0$, it will be splitted again to restore the previous situation. Note that the only computational overhead introduced by this procedure is the book-keeping of merging and splitting exemplars, and this is made efficient by using lookup tables which list the parent and child nodes of each exemplar at each point in the hierarchy.

III. RESULTS

Using the experimental setup described in [2], the method is evaluated on the digit recognition task TIDIGITS. The parameters used were $E = 690$, $N_0 = 2000$, $M = 4000$, $I = 600$, $\hat{N} = 408\,066$. HAC was done using a euclidean distance measure on log-compressed features.

The proposed method was compared to recognition with several fixed dictionary sizes $N \in \{2000, 4000, 8000, 16000, 32000\}$. The results in Table I show that with these settings, the proposed method is able to perform at least as good as fixed dictionaries containing 32000 exemplars. Timing experiments (not shown) confirmed that the algorithm does not perform substantially slower than using a fixed dictionary of size $2000 - 4000$.

ACKNOWLEDGMENT

Tuomas Virtanen is acknowledged for helpful discussions. The research of Jort F. Gemmeke was supported by the Dutch-Flemish STEVIN project MIDAS and by IWT project ALADIN.

REFERENCES

- [1] J. Wright, A. Y. Yang, A. Ganesh, S. Shankar Sastry and Yi Ma, "Robust Face Recognition via Sparse Representation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–222, February 2009.
- [2] J. F. Gemmeke, T. Virtanen, and A. Hurmalainen, "Exemplar-based sparse representations for noise robust automatic speech recognition," accepted for publication in *IEEE Transactions on Audio, Speech and Language processing*, 2010. [Available online]: www.amadana.nl
- [3] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Neural Information Processing Systems*, April 2001, pp. 556–562.

Sparse Object-Based Audio Coding Using Non-Negative Matrix Factorization of Spikegrams

I. INTRODUCTION

A sparse audio representation applied to audio source coding is previously proposed in [1]. In that approach the audio signal is projected onto a set of gammatone/gammachirp kernels that generates sparse representations dubbed as spikegrams. Addressing each spike individually in the previously proposed approach is very costly in terms of bits when audio coding applications are considered. To reduce the overall bitrate, a technique based on frequency episode discovery is proposed in [1]. Other techniques based on graph theory have also been proposed in [6]. In this paper, we outline a novel approach based on NMF-2D with sparsity [2] to extract the component and projection matrices. These two matrices are then quantized, arithmetically coded, and sent to the receiver. The receiver multiplies the two matrices and generates a resynthesized spikegram that can be used to generate the original audio signal. Our work has some similarities with the approach proposed by Nikunen and Virtanen [3]. However, Nikunen and Virtanen used spectrograms and only applied the NMF technique to code the amplitude of the signal and not to the non-positive phase information. Since phase information is not coded by NMF, the overall bitrate cannot be reduced below a certain point in [3]. On the other hand, our proposed spikegram contains both phase and amplitude information, and there is no need to send phase as side information as proposed in [3]. Furthermore, Nikunen and Virtanen used the standard NMF as proposed by Lee and Seung [4], which we found less optimal for coding purposes than the NMF-2D. The fact that repetitive patterns in an audio signal span in the 2-D frequency plane, warrants the use of NMF-2D which would optimally resolve audio objects on both axes. Moreover, since our spikegrams are sparse it is much easier to impose a sparseness constraint on the NMF than in the case of a spectrogram where the information is spread in a relatively more uniform way in both time and frequency. Finally, it can be shown [5] that the mean-squared error between the original audio signal and the reconstructed audio signal is inversely proportional to the redundancy (overcompleteness) of the representation for a given error in the representation domain. Therefore, the mean-squared error in the temporal domain is lower for our spikegram (overcomplete representation) compared to the spectrogram (orthogonal representation) used by other researchers for a given error in the representation domain. Results on different audio signals show that our approach is able to code audio signals at 30+ dB with a bitrate around 85 kbps. Informal listening tests also show that the quality of the resynthesized audio signals is near transparent. Preliminary results show that these bitrates can be further reduced by noise shaping, bandwidth extension, etc. These results are a first step toward an object-based universal audio coder.

II. DESCRIPTION OF THE APPROACH

Figure 1 shows the block diagram of the proposed approach. A perception-based sparse representation called “spikegram” is first generated by projecting the audio signal onto gammatone/gammachirp kernels. Since some of the coefficients are negative, we apply an invertible transform to the representation to create a

non-negative representation and keep only rows that have non-zero elements. We then apply the NMF-2D, where the decomposition is done as following:

$$\mathbf{V} \approx \mathbf{\Lambda} = \sum_{\tau} \sum_{\phi} \mathbf{W}^{\tau} \mathbf{H}^{\phi} \quad (1)$$

where $\downarrow \phi$ denotes the downward shift operator which moves each element in the matrix ϕ rows down, and $\rightarrow \tau$ denotes the right shift operator which moves each element in the matrix τ columns to the right. \mathbf{W} contains the basis vectors, \mathbf{H} is the projection matrix, \mathbf{V} is the transformed spikegram (audio signal representation), and $\mathbf{\Lambda}$ is the approximate reconstruction of the transformed spikegram.

In order to guarantee a certain audio quality, we propose an approach to adaptively modify the NMF-2D parameters such as sparseness, length of temporal and frequency shifts. The cost function of the NMF-2D is also perceptually shaped, since the standard mean-squared error cost of the NMF-2D penalizes high frequencies. As shown in the figure, the elements of \mathbf{H} are vector quantized while the elements of \mathbf{W} are scalar quantized (since \mathbf{W} is a much smaller matrix compared to \mathbf{H} , the bit cost of using scalar quantization is negligible). Matrices \mathbf{H} and \mathbf{W} are then transmitted to the receiver and reconstructed to obtain the original audio signal. Results with different audio signals show that a bitrate of 85 kbps at 30+ dB can be achieved. Informal listening tests also confirm good quality of reconstruction.

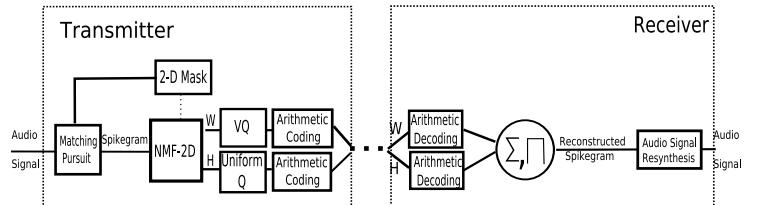


Fig. 1. Block diagram of the proposed audio coder based on NMF-2D

III. FUTURE WORK

A better noise shaping model introduced in the cost function (see [1]) as well as a method based on spectral bandwidth extension can further reduce the bitrate for the same audio quality.

REFERENCES

- [1] R. Pichevar et al., Auditory-Inspired Sparse Representation of Multimedia Signals with Applications to Audio Coding. *Speech Communication*, 2010
- [2] M. Morup and M. Schmidt, Sparse Non-Negative Matrix Factor 2-D Deconvolution. *Technical University of Denmark*, 2010
- [3] J. Nikunen and T. Virtanen, Object-Based Audio Coding Using Non-Negative Matrix Factorization for the Spectrogram Representation. *AES, London*, 2010
- [4] D. Lee and S. Seung, Learning the parts of objects with nonnegative matrix factorization. *Nature*, 1999
- [5] V. Goyal, Quantized Overcomplete Expansions: Analysis, Synthesis, and Algorithms. *UC Berkley*, 1995
- [6] C. Srinivasa, Master’s thesis (to appear). *University of Ottawa*, 2011

Recovery of Compressively Sampled Sparse Signals using Cyclic Matching Pursuit

Bob L. Sturm and Mads G. Christensen

Department of Architecture, Design and Media Technology

Aalborg University Copenhagen

Lautrupvang 15, 2750 Ballerup, Denmark

E-mail:{bst,mgc}@create.aau.dk

Abstract—We empirically show how applying a pure greedy algorithm cyclically can recover compressively sampled sparse signals as well as other more computationally complex approaches, such as orthogonal greedy algorithms, iterative thresholding, and ℓ_1 -minimization.

I. INTRODUCTION

Under certain conditions, we can recover a vector $\mathbf{x} \in \mathbb{R}^N$ from measurements $\mathbf{u} = \Phi\mathbf{x}$ created by a matrix with unit-norm columns $\Phi \in \mathbb{R}^{m \times N}$ ($N > m$). Here we focus on a cyclic application of the pure greedy algorithm matching pursuit (MP) [1]. Given the index set $\Omega_k \subset \Omega = \{1, 2, \dots, N\}$ (indexing the columns of Φ), MP augments this set by $\Omega_{k+1} = \Omega_k \cup \{n_k\}$ using

$$n_k = \arg \min_{n \in \Omega} \|\mathbf{r}_k - \langle \mathbf{r}_k, \varphi_n \rangle \varphi_n\|_2^2 = \arg \max_{n \in \Omega} |\langle \mathbf{r}_k, \varphi_n \rangle| \quad (1)$$

where φ_n is the n th column of Φ , $\mathbf{r}_k = \mathbf{u} - \Phi\mathbf{x}_k$ is the residual, and the n_k row of \mathbf{x}_{k+1} is defined

$$[\mathbf{x}_{k+1}]_{n_k} = [\mathbf{x}_k]_{n_k} + \langle \mathbf{r}_k, \varphi_{n_k} \rangle. \quad (2)$$

For initialization, $\Omega_0 = \emptyset$ and $\mathbf{x}_0 = \mathbf{0}$. Pure greedy algorithms like MP are successful only for the most trivial of cases, e.g., when Φ contains an orthogonal basis and \mathbf{x} activates only functions in that basis that are orthogonal to the rest of Φ [2], [3].

Cyclic MP (CMP) [4], [5] runs as MP at each iteration, but includes a model refinement. Define the i th value of $\Omega_k \subset \Omega = \{1, 2, \dots, N\}$, $\Omega_k(i)$. First for $i = 1$, CMP finds a replacement atom

$$n_i = \arg \min_{n \in \Omega} \|\mathbf{r}_{k \setminus i} - \langle \mathbf{r}_{k \setminus i}, \varphi_n \rangle \varphi_n\|_2^2 = \arg \max_{n \in \Omega} |\langle \mathbf{r}_{k \setminus i}, \varphi_n \rangle| \quad (3)$$

where $\mathbf{r}_{k \setminus i} = \mathbf{u} - [\Phi\mathbf{x}_k - \varphi_{\Omega_k(i)}[\mathbf{x}_k]_{\Omega_k(i)}]$. Then CMP updates Ω_k such that $\Omega_k(i) = n_i$, and the solution $[\mathbf{x}_k]_{n_i} = \langle \mathbf{r}_{k \setminus i}, \varphi_{n_i} \rangle$. Then CMP does the same for $i = 2$, up to k . After cycling through all atoms until some stopping criterion is met, CMP augments Ω_k as in MP, and refines the model again.

Figure 1 shows the probability of exact recovery ($\|\mathbf{x} - \hat{\mathbf{x}}\|_2 / \|\mathbf{x}\|_2 < 0.01$) for vectors of varying sparsity k with elements drawn from two distributions, for six undersampling ratios m/N with no noise, using both CMP and Orthogonal MP (OMP). For these experiments, we make $N = 400$, sample Φ from the uniform spherical ensemble, and average the results over 100 independent trials for each sparsity and number of measurements. In our implementation, we make CMP run the refinement procedure a max of five times, or until $\|\mathbf{r}'_k\|_2^2 / \|\mathbf{r}_k\|_2^2 > 0.999$, where \mathbf{r}'_k is the residual after refinement. It is clear that CMP can perform just as well as OMP at this task without matrix inversions. Our final work will include comparisons with other methods, such as iterative thresholding [6], ℓ_1 minimization [7], and two-stage thresholding [8], as well as an analysis of the algorithm.

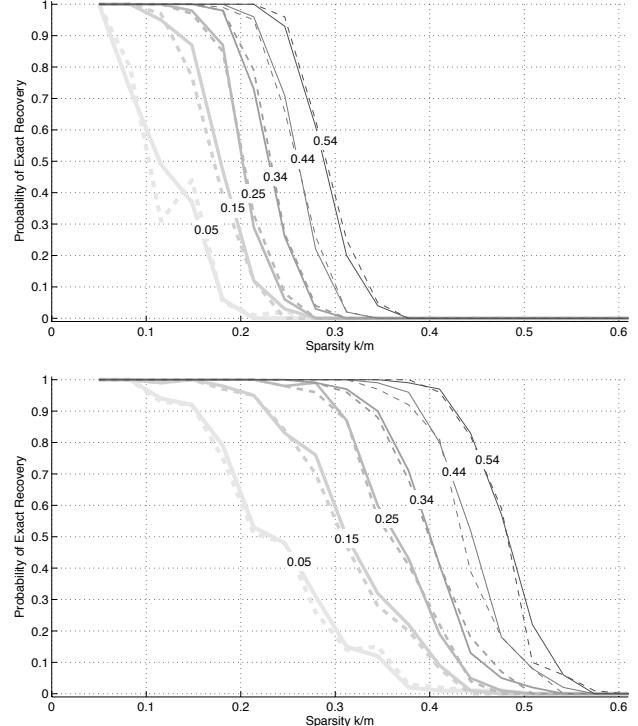


Fig. 1. Probability of exact recovery using CMP (solid) and OMP (dashed) at several undersampling values m/N (labeled). Top: Active elements distributed Constant Amplitude Random Signs [8]. Bottom: Active elements distributed Normal.

REFERENCES

- [1] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. Amsterdam: Academic Press, Elsevier, 2009.
- [2] G. Davis, S. Mallat, and M. Avellaneda, “Adaptive greedy approximations,” *J. Constr. Approx.*, vol. 13, no. 1, pp. 57–98, Jan. 1997.
- [3] R. A. DeVore and V. N. Temlyakov, “Some remarks on greedy algorithms,” *Adv. Comput. Math.*, vol. 5, pp. 173–187, 1996.
- [4] M. G. Christensen and S. H. Jensen, “The cyclic matching pursuit and its application to audio modeling and coding,” in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2007.
- [5] B. L. Sturm and M. Christensen, “Cyclic matching pursuit with multiscale time-frequency dictionaries,” in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2010.
- [6] T. Blumensath and M. E. Davies, “Normalized iterative hard thresholding: guaranteed stability and performance,” *IEEE J. Selected Topics Signal Process.*, vol. 4, no. 2, pp. 298–309, Apr. 2010.
- [7] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, Aug. 1998.
- [8] A. Maleki and D. L. Donoho, “Optimally tuned iterative reconstruction algorithms for compressed sensing,” *IEEE J. Selected Topics in Signal Process.*, vol. 4, no. 2, pp. 330–341, Apr. 2010.

Structured and soft ! Boltzmann machine and mean-field approximation for structured sparse representations

Angélique Drémeau ^(a,b)

^(a)Fondation Pierre-Gilles De Gennes pour la Recherche
29 rue d'Ulm, F-75005 Paris, France
Email: angelique.dremeau@espci.fr

Laurent Daudet ^(b)

^(b)Institut Langevin, ESPCI ParisTech, Univ. Paris Diderot, CNRS
UMR 7587, F-75005 Paris, France
Email: laurent.daudet@espci.fr

Abstract—We address the problem of *structured* sparse representation within a Bayesian framework. In particular, we consider a mean-field approximation for the estimation of the dependencies between atoms using a Boltzmann machine. This algorithm is shown to outperform the reference algorithm [1] with regard to their success criterion.

Index Terms—Structured sparse representations, Boltzmann machine, mean-field approximation.

I. INTRODUCTION

Recent contributions have emphasized the interest of considering *structures* between atoms selected in sparse representations (SR), for a wide range of dictionaries and classes of signals. This problem can be set into a Bayesian framework, e.g. Cevher *et al.* [2] and Faktor *et al.* [1]. Both use Boltzmann machines to model the dependencies between atoms, but differ in the prior model on the SR coefficients. In this paper, we consider a similar model as in [1].

Our observation model is $\mathbf{y} = \sum_{i=1}^M s_i x_i \mathbf{d}_i + \mathbf{n}$, where $\mathbf{s} \in \{0, 1\}^M$ is the SR support, $\mathbf{n} \sim \mathcal{N}(0, \sigma_n^2 \mathbf{I}_N)$ and \mathbf{I}_N the identity matrix. We suppose that $\forall i, p(x_i) = \mathcal{N}(0, \sigma_{x_i}^2)$ and \mathbf{s} is distributed according to a Boltzmann machine of parameters \mathbf{b} and \mathbf{W} ¹:

$$p(\mathbf{s}) \propto \exp(2\mathbf{b}^T \mathbf{s} + 2\mathbf{s}^T \mathbf{W}\mathbf{s} - 2\mathbf{1}_M^T \mathbf{W}\mathbf{s}), \quad (1)$$

where $\mathbf{1}_M = [1, \dots, 1]$ of length M .

II. STRUCTURED SOFT BAYESIAN PURSUIT ALGORITHM

Based on this model, we consider here the following *marginalized* maximum a posteriori (MAP) estimation problem:

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s} \in \{0, 1\}^M} \log p(\mathbf{s}|\mathbf{y}), \quad (2)$$

where $p(\mathbf{s}|\mathbf{y}) = \int_{\mathbf{x}} p(\mathbf{x}, \mathbf{s}|\mathbf{y}) d\mathbf{x}$. To tackle problem (2), a greedy algorithm could be used [1] to approach the solution with a succession of local decisions. In this paper, we alternatively propose a mean-field (MF) approximation of $p(\mathbf{x}, \mathbf{s}|\mathbf{y})$ which approximates $p(\mathbf{x}, \mathbf{s}|\mathbf{y})$ with a probability distribution, say $q(\mathbf{x}, \mathbf{s})$, constrained to have a “suitable” factorization while minimizing the Kullback-Leibler distance with $p(\mathbf{x}, \mathbf{s}|\mathbf{y})$. Here, $q(\mathbf{x}, \mathbf{s})$ is constrained to the structure:

$$q(\mathbf{x}, \mathbf{s}) = \prod_i q(x_i, s_i) = \prod_i q(x_i|s_i) q(s_i). \quad (3)$$

Then the minimization of the Kullback-Leibler distance subject to (3) can be performed by the “variational Bayes EM algorithm” (VB-EM) [3], which evaluates the $q(x_i, s_i)$'s by computing at each iteration²:

$$q(x_i|s_i) = \mathcal{N}(m(s_i), \Gamma(s_i)),$$

$$q(s_i) \propto \sqrt{\Gamma(s_i)} \exp\left(\frac{1}{2} \frac{m(s_i)^2}{\Gamma(s_i)}\right) \exp\left(2s_i(b_i + \sum_{j \neq i} w_{ij}(q(s_j=1)-1))\right)$$

$$\text{where } \Gamma(s_i) = \frac{\sigma_{x_i}^2 \sigma_n^2}{\sigma_n^2 + \sigma_{x_i}^2 s_i}, \quad m(s_i) = s_i \frac{\sigma_{x_i}^2}{\sigma_n^2 + \sigma_{x_i}^2 s_i} \langle \mathbf{r}_i \rangle^T \mathbf{d}_i, \\ \langle \mathbf{r}_i \rangle = \mathbf{y} - \sum_{j \neq i} q(s_j=1) m(s_j=1) \mathbf{d}_j.$$

¹This distribution is equal to the one used in [1], [2] with $\mathbf{s} \in \{-1, 1\}^M$.

²For a sake of clarity, we drop here the iteration indices.

Compared to [1], the proposed algorithm does not make any hard decision on the SR support at each iteration but rather updates posterior probabilities. In that way, it can be seen as a *soft* process. Both algorithms have similar complexities, of order M^2 per iteration.

Coming back to (2), $p(\mathbf{s}|\mathbf{y})$ is simplified as $p(\mathbf{s}|\mathbf{y}) \simeq \int_{\mathbf{x}} \prod_i q(x_i, s_i) d\mathbf{x} = \prod_i q(s_i)$. We finally obtain $\forall i \hat{s}_i = \arg \max_{s_i \in \{0, 1\}} \log q(s_i)$, which is solved by simple thresholding.

III. EXPERIMENTAL RESULTS

To assess the performance of the proposed algorithm, we follow the same methodology as in [1]. We generate a large number K of observations according to the model and estimate the ability of the algorithm to reconstruct the SR support via the probability

$$1 - \frac{1}{K} \sum_{k=1}^K \frac{\|\mathbf{s}^{(k)} \cap \hat{\mathbf{s}}^{(k)}\|_0}{\max(\|\mathbf{s}^{(k)}\|_0, \|\hat{\mathbf{s}}^{(k)}\|_0)}. \quad (4)$$

The data is generated with $N = 64$, $M = 256$, and a DCT dictionary. The Boltzmann parameters are drawn independently: the elements of \mathbf{b} from $\mathcal{N}(-2.5, 1)$ and the elements of \mathbf{W} from $\mathcal{U}[-0.1, 0.1]$. The standard deviations σ_{x_i} are *i.i.d.* realizations of $\mathcal{U}[15, 60]$. For each point of simulation, we run 500

trials. We adjust the final threshold at 0.25. The figure above compares 2 algorithms: “MAP-greedy”, proposed in [1] and “SSoBaP” (for Structured Soft Bayesian Pursuit algorithm), proposed here. For the performance criterion considered, we can see that “SSoBaP” outperforms “MAP-greedy” over a wide range of noise variances.

IV. CONCLUSION

In this paper, we have shown that a MF approximation together with a VB-EM algorithm is a promising and competitive approach for the estimation of structures between atoms. To the extent of the considered criterion, the resulting algorithm is shown to outperform the baseline algorithm [1]. Complementary results, involving other performance criteria and other state-of-the-art algorithms, will be added in the final paper to confirm the relevance of this approach.

REFERENCES

- [1] T. Faktor, Y. C. Eldar, and M. Elad, “Exploiting statistical dependencies in sparse representations for signal recovery,” Submitted to IEEE Trans. On Signal Processing.
- [2] V. Cevher, M. F. Duarte, C. Hegde, and R. G. Baraniuk, “Sparse signal recovery using markov random fields,” in *NIPS*, 2008.
- [3] M. J. Beal and Z. Ghahramani, “The variational bayesian em algorithm for incomplete data: with application to scoring graphical model structures,” *Bayesian Statistics*, vol. 7, pp. 453–463, 2003.

BM3D-frame sparse image modeling and decoupling of inverse and denoising for image deblurring

Aram Danielyan, Vladimir Katkovnik and Karen Egiazarian

Department of Signal Processing, Tampere University of Technology (TUT), Tampere, Finland,
e-mail: firstname.lastname@tut.fi.

I. INTRODUCTION

We consider a reconstruction of \mathbf{y} from observations $\mathbf{z} = \mathbf{A}\mathbf{y} + \sigma\boldsymbol{\varepsilon}$, where $\mathbf{z}, \mathbf{y} \in R^N$ are vectors representing correspondingly observed and true images, \mathbf{A} is a $N \times N$ blur matrix, $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}_{N \times 1}, \mathbf{I}_{N \times N})$ is a vector of i.i.d. standard Gaussian components, and σ is the standard deviation of the noise.

Contribution of this paper concerns two aspects of inverse image reconstruction. First, we use the *BM3D*-frames presented in [1] for sparse image modeling. Second, we formalize image reconstruction as a vector variational problem with two objective functions. This technique results in decoupling of inverse and filtering. Comparison versus the standard variational settings with a single objective function demonstrates a clear advantage of the decoupling. Overall, the achieved results numerically and visually are very good and mainly overcome the best competitive results in the field.

II. BM3D IMAGE MODELING

In detail discussion of BM3D modeling can be found in [2]. It is a nonlocal adaptive technique based on high-order groupwise models defined in 3D transform domain. It has been shown in [1] that provided a fixed grouping the BM3D analysis/synthesis can be given in the matrix form linking the image \mathbf{y} and its groupwise spectrum vector $\boldsymbol{\omega} \in R^M$ by the forward and backward transforms

$$\boldsymbol{\omega} = \Phi \cdot \mathbf{y}, \mathbf{y} = \Psi \cdot \boldsymbol{\omega}. \quad (1)$$

Proposition 1. The matrices $\Phi^T \Phi$ and $\Psi \Psi^T$ are diagonal with positive items; $\Psi \Phi = \mathbf{I}_{N \times N}$.

The last formula enables perfect reconstruction of the image \mathbf{y} from the groupwise spectrum $\boldsymbol{\omega}$. It follows from the proposition that Φ and Ψ^T are full column rank matrices. The rows of the full rank ($M \times N$) matrix Φ constitute a frame in R^N , and the columns of the full rank ($N \times M$) matrix Ψ constitute a frame dual to Φ . These frames are not tight, $\Phi^T \cdot \Phi \neq \mathbf{I}_{N \times N}$ and $\Psi^T \cdot \Psi \neq \mathbf{I}_{N \times N}$. In general $\Psi \neq (\Phi^T \Phi)^{-1} \Phi^T$, and Ψ is an alternative dual frame.

III. VARIATIONAL IMAGE DEBLURRING

For the above observation model with Gaussian i.i.d. noise we consider the following variational setting

$$(\hat{\boldsymbol{\omega}}, \hat{\mathbf{y}}) = \arg \min_{\boldsymbol{\omega}, \mathbf{y}} \left\{ \frac{1}{2\mu} \|\mathbf{z} - \mathbf{A}\mathbf{y}\|_2^2 + \tau \cdot \|\boldsymbol{\omega}\|_p \mid \boldsymbol{\omega} = \Phi\mathbf{y}, \mathbf{y} = \Psi\boldsymbol{\omega} \right\}, \quad (2)$$

where both the analysis and synthesis links between the image and spectrum are considered as constraints. For $p = 1$ and $p = 0$ (2) is defining respectively l_2-l_1 and l_2-l_0 optimization problems.

Let us replace the constrained minimization in (2) by an unconstrained one where the constraints are replaced by the quadratic penalties with positive weights γ_s . In this way we arrive to the

following objective function

$$\begin{aligned} \mathcal{L}(\mathbf{y}, \boldsymbol{\omega}) &= \frac{1}{2\mu} \|\mathbf{z} - \mathbf{A}\mathbf{y}\|_2^2 + \tau \cdot \|\boldsymbol{\omega}\|_p + \\ &\quad \frac{1}{2\gamma_1} \|\boldsymbol{\omega} - \Phi\mathbf{y}\|_2^2 + \frac{1}{2\gamma_2} \|\mathbf{y} - \Psi\boldsymbol{\omega}\|_2^2. \end{aligned} \quad (3)$$

This $\mathcal{L}(\mathbf{y}, \boldsymbol{\omega})$ is universal in the sense, that with $\gamma_1 \rightarrow \infty$ it corresponds to the *synthesis* approach and with $\gamma_2 \rightarrow \infty$ it corresponds to the *analysis* approach to image reconstruction. In general, with finite γ_1, γ_2 it defines a *combined synthesis/analysis* approach.

IV. MAIN RESULTS

Let us decompose (3) into the sum of two objective functions, $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$, where

$$\begin{aligned} \mathcal{L}_1(\mathbf{y}, \boldsymbol{\omega}) &\triangleq \frac{1}{2\mu} \|\mathbf{z} - \mathbf{A}\mathbf{y}\|_2^2 + \frac{1}{2\gamma_2} \|\mathbf{y} - \Psi\boldsymbol{\omega}\|_2^2, \\ \mathcal{L}_2(\mathbf{y}, \boldsymbol{\omega}) &\triangleq \tau \cdot \|\boldsymbol{\omega}\|_p + \frac{1}{2\gamma_1} \|\boldsymbol{\omega} - \Phi\mathbf{y}\|_2^2. \end{aligned} \quad (4)$$

We define a novel image deblurring algorithm using the following alternative minimization of \mathcal{L}_1 and \mathcal{L}_2 :

$$\begin{cases} \mathbf{y}_{t+1} = \arg \min_{\mathbf{y}} \mathcal{L}_1(\mathbf{y}, \boldsymbol{\omega}_t), \\ \boldsymbol{\omega}_{t+1} = \arg \min_{\boldsymbol{\omega}} \mathcal{L}_2(\mathbf{y}_{t+1}, \boldsymbol{\omega}), \end{cases} \quad t = 0, 1, \dots \quad (5)$$

In this algorithm instead of conventional for the variational approaches minimization of a single objective function \mathcal{L} we use an alternative minimization of two objective functions partial summands of \mathcal{L} . It is easy to notice that minimization of \mathcal{L}_1 on \mathbf{y} serves to inverse the blur operator, while minimization of \mathcal{L}_2 on $\boldsymbol{\omega}$ serves as a denoising operation. Thus, the proposed decomposition of \mathcal{L} corresponds to decoupling of deblurring and denoising.

The proposed algorithm is looking for a *fixed-point* $(\boldsymbol{\omega}^*, \mathbf{y}^*)$ defined as a solution of two equations:

$$\begin{cases} \mathbf{y}^* = \arg \min_{\mathbf{y}} \mathcal{L}_1(\mathbf{y}, \boldsymbol{\omega}^*), \\ \boldsymbol{\omega}^* = \arg \min_{\boldsymbol{\omega}} \mathcal{L}_2(\mathbf{y}^*, \boldsymbol{\omega}). \end{cases} \quad (6)$$

The following convergence result is proved for the algorithm (5).

Proposition 2. For any fixed $\mu, \gamma_1, \gamma_2, \tau$, the sequence $(\mathbf{y}_t, \boldsymbol{\omega}_t)$ generated by (5) converges to a fixed point $(\mathbf{y}^*, \boldsymbol{\omega}^*)$ of the equations (6) if it exists.

Extensive simulation experiments show a serious advantage of the developed algorithm over the best techniques in the field.

REFERENCES

- [1] A. Danielyan, V. Katkovnik and K. Egiazarian, "Image deblurring by augmented Lagrangian with BM3D frame prior," Workshop on Information Theoretic Methods in Science and Engineering (2010).
- [2] K. Dabov, A. Foi, V. Katkovnik, and Egiazarian, K., "Image denoising by sparse 3D transform-domain collaborative filtering," IEEE Transactions on Image Processing, vol. **16**, 2080 - 2095 (2007).
- [3] V. Katkovnik, A. Danielyan and K. Egiazarian, "Decoupled inverse and denoising for image deblurring: variational BM3D-frame technique," submitted to ICIP-2011 (2011).

Super-resolution and reconstruction of far-field ghost imaging via sparsity constraints

Wenlin Gong, and Shensheng Han

Key Laboratory for Quantum Optics and Center for Cold Atom Physics of CAS,
Shanghai Institute of Optics and Fine Mechanics,
Chinese Academy of Sciences, Shanghai 201800, China

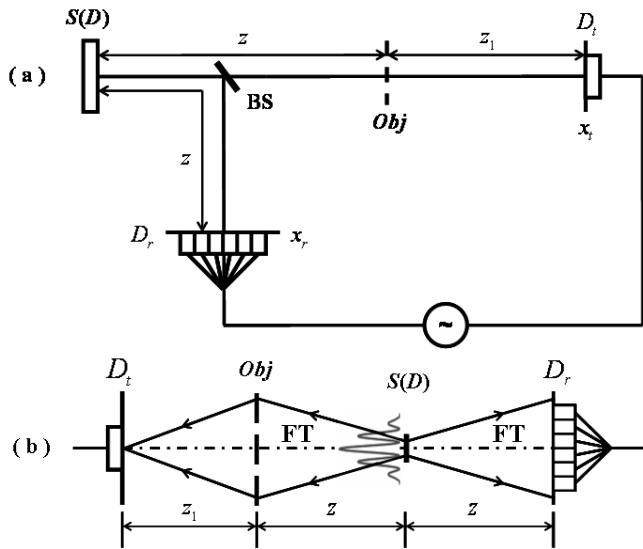


Fig. 1. (a). Standard schematic of lensless far-field GI with thermal light; (b). the physical explanation of far-field GI, the thermal source S shown in the scheme (a) acts as a phase conjugated mirror and a spatial low-pass filter because of its finite transverse size.

Abstract—For ghost imaging [1], [2], [3], [4], [5], the speckle's transverse size on the object plane is the system's diffraction limit and enhancing the resolution beyond this diffraction limit is generally called super-resolution [4], [5], [6]. When signals satisfied a certain sparsity conditions, Donoho has demonstrated mathematically that super-resolution restoration was possible [7]. By combining the sparse prior property of images with ghost imaging method, we demonstrated experimentally that super-resolution imaging can be nonlocally achieved in the far field applying a new sparse reconstruction method called gradient projection for sparse reconstruction (GPSR) algorithm [5], [6], [8], [9].

Fig. 1(a) presents the experimental schematic for lensless far-field ghost imaging, which is designed as the method mentioned in Ref. [5]. Fig. 1(b) describes the physical explanation of far-field GI and its resolution, which is discussed in detail in Ref. [6].

Fig. 2 presents experimental results of a double-slit recovered with ghost imaging (GI) and ghost imaging via sparsity constraints (GISC) methods in different collecting areas $L_1 \times L_1$, using the schematic shown in Fig. 1(a).

In conclusion, we have achieved super-resolution far-field

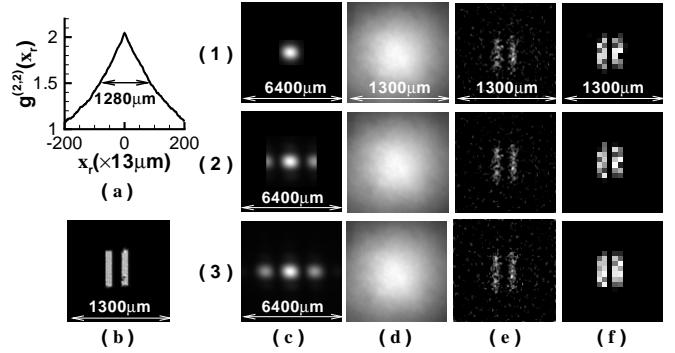


Fig. 2. Experimental reconstruction of a double-slit in different collecting areas with $z=1200\text{mm}$, $z_1=500\text{mm}$ and $D=0.6\text{mm}$ (the speckle's transverse size $\Delta x_s=1280\mu\text{m}$). (a). The cross-section curve of the speckle on the object plane obtained by measuring the second-order correlation function of light field on the reference detection plane (the curve's full-width at half-max (FWHM) is the diffraction limit of GI); (b). the object; (c). the object's diffraction patterns received by the test detector D_t ; (d). GI method (averaged 3000 measurements); (e) and (f) are GISC when the pixel-resolution of the camera D_r is $13\mu\text{m}$ and $65\mu\text{m}$, respectively (with 3000 and 500 measurements for (e)-(f), respectively). The collecting areas of the detector D_t shown in (1-3) are $1.6\text{mm} \times 1.6\text{mm}$, $3.2\text{mm} \times 3.2\text{mm}$, and $6.4\text{mm} \times 6.4\text{mm}$.

GI by combining GI method with the sparse prior property of images. We also show that Fourier-transform diffraction pattern of the object and its image in real-space can be obtained at the same time. This brand new far-field super-resolution imaging method will be very useful to microscopy in biology, material, medical sciences, and in the filed of remote sensing, etc.

The work was partly supported by the Hi-Tech Research and Development Program of China under Grant Project No. 2006AA12Z115, and Shanghai Natural Science Foundation under Grant Project No. 09JC1415000.

REFERENCES

- [1] J. Cheng and S. Han, Phys. Rev. Lett. **92**, 093903 (2004).
- [2] A. Gatti et al., Phys. Rev. Lett. **93**, 093602 (2004).
- [3] M. D'Angelo, and Y. H. Shih, Laser. Phys. Lett. **2**, 567-596 (2005).
- [4] F. Ferri et al., Phys. Rev. Lett. **94**, 183602 (2005).
- [5] W. Gong et al., Appl. Phys. Lett. **95**, 071110 (2009).
- [6] W. Gong and S. Han, “Super-resolution far-field ghost imaging via compressive sampling,” arXiv: Quant-Ph/0911.4750, (2009).
- [7] D. L. Donoho, Siam. J. Math. Anal. **23**, 1309-1331 (1992).
- [8] E. J. Candès and M. B. Wakin, IEEE Signal Process. Mag. **25**, 21 (2008), and references therein.
- [9] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, IEEE J. Sel. Top. in Sig. Proc. **1**, 586-597 (2007).

Fast Compressive Terahertz Imaging

H. Shen^a, L. Gan^b, N. Newman^a and Y. C. Shen^{a*}

^a Dept of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, UK

^b Electronic and Computer Engineering, Brunel University, Uxbridge UB8 3PH, UK

Terahertz pulsed imaging (TPI) modalities have numerous applications such as medical diagnosis, detection and chemical mapping of illicit drugs and explosives, and inspection of pharmaceutical tablet. However, as the majority of terahertz images were obtained in a pixel-by-pixel raster scan fashion, existing terahertz imaging systems have slow imaging speed. Recently, Chan *et al.* [1] first reported a new terahertz imaging system based on the concept of compressed sensing (CS) [2, 3] for high-speed image acquisition in which the number of measurements is much smaller than that of the total pixels in reconstructed images [2, 3]. In particular, experimental results in [1] suggested that only 300 measurements were used to obtain an image of 32×32 pixels with reasonable quality.

Despite its great potential, compressive TPI is still at the infant stage and much needs to be done before its practical applications. In this talk, we will present our work on the development of fast compressive TPI cameras from the signal processing perspective. In particular, we will focus on the design and implementation of efficient sampling operators. We will also highlight main challenges for reconstruction of terahertz images, especially for time-domain terahertz pulsed systems. Extensive hardware measurement and reconstruction results will be presented.

Fig. 1 shows the experimental arrangement for compressive TPI. The masks were used to modulate the terahertz waveforms. Recall that in [1], full random masks have been used. Although such a sampling operator is theoretically optimal, they require huge memory for storage and heavy computation complexity for reconstruction. Besides, due to the lack of spatial light modulator in terahertz imaging, the hardware implementation is complicated. In our work, we have investigated the construction of deterministic and structured random operators. Specifically, we have developed a 40×400 deterministic binary sampling operator [4] and Fig. 2 shows an example of experimental results for time-domain terahertz pulsed imaging. As can be seen, the Chinese character “big” can be reconstructed at different terahertz frequencies. To enable fast sampling, we have also proposed the use of a single rotating mask (a spin disk) for automatic and continuous implementation [5]. Such a design offers the advantages of compact design, easy computation and fast implementation with potentially video-rate sampling speed. As compared with conventional TPI, only 10%-20% of the pixels are required. Fig. 3 shows some experimental results where the 32×32 terahertz images “A”, “U”, and “H” were reconstructed using only 160 measurements. Our experimental results suggested that CS based TPI may have great potential in real-time imaging applications.

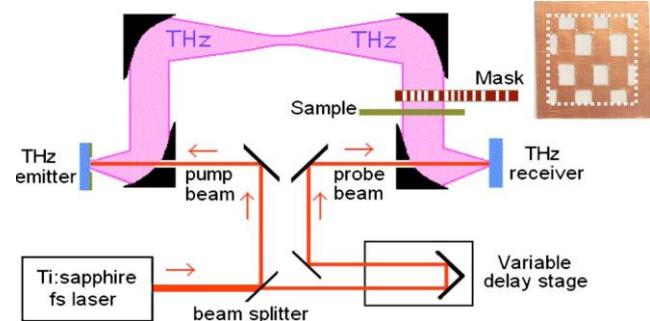


Figure 1. Experimental arrangement for terahertz pulsed imaging using CS. The inset shows one of 40 designed masks with the dotted line indicating the 40×40 mm² imaging area. The copper pixels are opaque to terahertz radiation while the white pixels are transparent to terahertz radiation. [4]

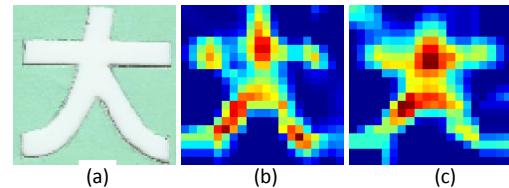


Figure 2. (a) Original 20×20 image of a Chinese character “big”. (b) Reconstructed image at 0.3 terahertz. (c) Reconstructed image at 1.0 terahertz. The measurement operator is the 40×400 deterministic sampling operator proposed in [4].

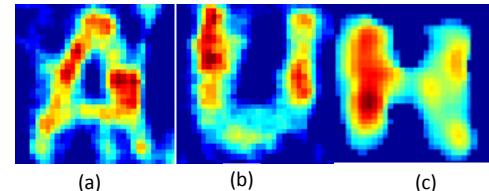


Figure 3. Reconstructed terahertz images shaped as the English characters (a) “A”, (b) “U”, and (c) “H” using 160 measurements from the spin disk implementation.

ACKNOWLEDGEMENT

The authors thank EPSRC laser loan pool for the laser system used in part of the work. NN would like to acknowledge funding from the EPSRC Vacation Bursary Scheme.

REFERENCES

- [1] W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman, *Appl. Phys. Lett.*, 93 (2008), 121105.
- [2] D. Donoho, *IEEE Trans. Inf. Theory*, 52, 1289, (2006).
- [3] E. Candes, J. Romberg, and T. Tao, *IEEE Trans. Inf. Theory*, 52, 489, (2006).
- [4] Y. C. Shen, L. Gan, M. Stringer, A. Burnett, K. Tych, H. Shen, J. E. Cunningham, E. P. J. Parrott, J. A. Zeitler, L. F. Gladden, E. H. Linfield, and A. G. Davies, *Appl. Phys. Lett.*, 95 (2009), 231112.
- [5] H. Shen, N. Newman, L. Gan, S. C. Zhong, Y. Huang and Y. C. Shen, 35th International Conference on IRMMW-THz, (2010).

Dictionary Learning: Application to ECG Denoising

Anastasia Zakharova, Olivier Laligant and Christophe Stoltz

University of Burgundy, Le2i laboratory

Le Creusot, France

Email: anastasia.a.zakharova@gmail.com, olivier.laligant(christophe.stoltz)@u-bourgogne.fr

Abstract—We propose a denoising method for ECG signals which is based on dictionary learning. On the preprocessing step, we obtain an overcomplete dictionary adapted to different types of ECG signals (choosing the training set in such a way that all the clinically important phenomena are included in it). We use then this dictionary for denoising. This method preserves the form of QRS complex and time localization of the signal that allows us to recognize an anomaly. We show that the proposed algorithm outperforms the algorithm of ECG denoising by sparse 2d-separable transform.

A. Dictionary Learning

A sparse representation is very useful in denoising because it improves the efficiency of the algorithm. Since a signal is said to be sparse in some dictionary D the choice of such a dictionary is crucial. Since it is not evident which dictionary is the best for ECG signals, we will learn a dictionary D which is particularly adapted to this type of signals; i.e., they are maximally sparse in it.

In order to learn the dictionary, we will use the same strategy as in [3]; that is, we solve a joint optimization problem

$$\min_{D \in \mathcal{C}, \alpha \in R^k} \frac{1}{2} \|x - D\alpha\|_2^2 + \lambda \|\alpha\|_1, \quad (1)$$

(with α being the decomposition coefficients, λ the regularization parameter) by alternating between the variables α and D ; while one of them is fixed, we minimize another one.

B. Simulation Results

As a training set to learn the dictionary, we used the database of ECG records obtained by the Creusot - Montceau-les-Mines hospital and we chose 14 signals taking a segment of 1000 samples from each of them in such a way that they represent the variety of clinically important phenomena.

We performed two kinds of simulation. First, we added to the signal randomly generated Gaussian noise with different variances. Then we applied the denoising algorithm and we studied the performance of the method by calculating the SNR of the noisy and reconstructed signal. The results are shown on Figure 2 where our method is compared to sparse 2d separable method and one can see that it performs better. Note that in [2] it was shown that sparse 2d separable algorithm outperforms the methods of soft thresholding [1] and extended Kalman smoother filtering [4].

The second simulation concerns the analysis of the ECG signal of a concrete patient. We apply the denoising algorithm to the pattern with visible noise and we compare the result with a similar pattern of the same patient which was not damaged with noise. As one can see on the Figure 2, the form of the denoised signal resembles a lot the form of signal with no noise while sparse 2d separable method fails to denoise this signal.

REFERENCES

- [1] D.L. Donoho, *Denoising by soft-thresholding*, IEEE Trans. on Information Theory, v. 41, p. 613–627, 1995.

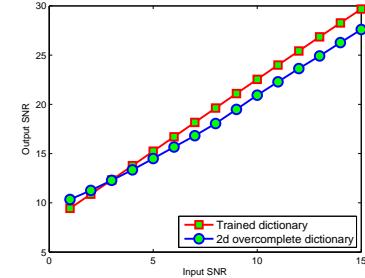


Fig. 1. Comparison of denoising methods in terms of SNR

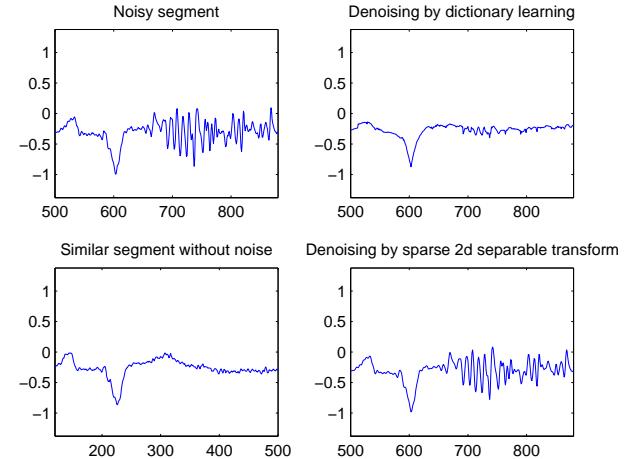


Fig. 2. Performance of the denoising algorithm on the signal with natural noise

- [2] A. Ghaffari, H. Palangi, M. Babaie-Zadeh, C. Jutten, *ECG denoising and compression by sparse 2D separable transform with overcomplete mixed dictionaries*, 2010.
[3] J. Mairal, F. Bach, J. Ponce, G. Sapiro, *Online learning for matrix factorization and sparse coding*, Journal of Machine Learning Research v.11, p.19–60, 2010.
[4] R. Sameni, M.B. Shamsollahi, C. Jutten, and G.D. Clifford, *A nonlinear Bayesian filtering framework for ECG denoising*, IEEE Trans. on Biomedical Engineering, v. 54, p.2171–2185, 2007.

Unsupervised Learning of View Condition Invariant Sparse Representation of Object Images by Blind Modeling

Ka-Yu Hui

Department of Information Engineering, The Chinese University of Hong Kong
Email: kyhui6@ie.cuhk.edu.hk

Abstract—Recently, sparse coding has been employed in natural image category classification problems and has produced state-of-the-art performance [1][2]. In this work, we present an unsupervised method for learning a view condition invariant representation for object images without explicit knowledge on the view conditions involved (“blind modeling”). The method requires only sufficient unlabeled image series, and can be used as a simple post-processing step to improve the performance of these state-of-the-art systems.

I. BACKGROUND

One of the core challenges of image category classification problems is the massive amount of variations that may affect how an object appears in an image. The said variations can be accounted by the physical variations of objects within categories (e.g. category of cars) coupled with large amounts of view conditions, which may include view angle, view distance, and light conditions. This makes it hard to relate limited amounts of labeled samples to a prohibitively larger set of potential test images. Thus state-of-the-art methods use different measures to reduce such variations. For example, they use local descriptors like SIFT (Scale-Invariant Feature Transform) [3] which extract features that are scale and rotation invariant from the image. They also use a technique called pooling to generate a translation-invariant representation for the image using the coded local descriptors. [4] provides a framework of these state-of-the-art systems for general sparse-coding research audience.

Despite these advances, simple transform invariances may not necessarily capture all the complicated effects of perspective and lighting on objects. This motivates a “blind modeling” of these real-world view conditions.

II. BASIC FORMULATION

Define $y \in \mathbb{R}^N$ as a certain representation of an observed object image. Our goal is to find a better sparse representation that is more invariant to effects of real-world view-conditions. We define our generative model of image as:

$$y = \mathcal{H}_h D x + \epsilon \quad (1)$$

where D is a latent dictionary with elements d_i and sparse representation $x \in \mathbb{R}^M$. Here \mathcal{H}_h is a certain random linear transform matrix $N \times N$ randomly drawn from a finite set of H linear transforms $\Delta = \{\mathcal{H}_0, \mathcal{H}_1 \dots \mathcal{H}_H\}$, which is assumed unknown and represents effects of real-world view conditions that we wish our recognition system would be invariant to. x is thus called a Δ -invariant representation of the object. ϵ would be a noise term.

With Δ being unknown (1) is not a amenable sparse recovery problem. Instead, define Δ -expanded dictionary Δ_D as $[\Delta_{d_1} \Delta_{d_2} \dots \Delta_{d_M}]$

with “view expansion” of element d_i $\Delta_{d_i} \triangleq [\mathcal{H}_1 d_i \ \mathcal{H}_2 d_i \ \dots \ \mathcal{H}_H d_i]$. This gives us an equivalent model

$$\begin{cases} y = \Delta_D x' + \epsilon \\ x' = x \otimes e_h \end{cases} \quad (2)$$

where e_h is an all-zero vector of length H except with the h^{th} element having a value of 1.

(2) now contains a more amenable sparse recovery problem. We can now recover dictionary C as a permuted Δ_D , provided that H and thus size of Δ_D is reasonably small for practical recovery. If it is as we assumed that y comes from some state-of-the-art systems in which basic levels of translation/scale/rotation invariance is already accounted for, then H will likely to be reasonable. If we can find a mapping function $f(k)$ which gives the identity of the “view expansion” Δ_{d_i} to which c_k , column k of C , belongs, then the Δ -invariant representation x may be recovered from its Δ -expanded representation x' .

III. CONTRIBUTION

Our work shows if we have sufficient amounts of relevant video footage containing unlabeled observations of objects we seek to detect and classify, i.e. image series that fits:

$$Y_s : y_t = \mathcal{H}_{ht} D x + \epsilon \text{ for } t = 1..T \quad (3)$$

along with x and thus x' being sufficiently sparse, $f(k)$ can indeed be found reliably using a simple clustering operation.

In our experiment, we will be using linear Spatial-Pyramid-Matching (linear-SPM) [1], a representative state-of-the-art system, as the baseline. Using the results from linear-SPM, we will apply our algorithm to produce a sparse representation with improved view-condition invariance and provide evalution.

The value of our proposed system is that the use of “blind modeling” spare us from the need to have exact knowledge of the complicated effects of real-world view conditions. This unsupervised process can theoretically be used to post-process any high-level representation vector of most state-of-the-art systems, and potentially be applied to other non-image domains.

REFERENCES

- [1] J. Yang, K. Yu, Y. Gong, and T. Huang, *Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification*, CVPR, 2009.
- [2] M. Ranzato, Y. Boureau, and Y. LeCun, *Sparse feature learning for deep belief networks*, NIPS, 2007
- [3] D. Lowe, *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision, 2004.
- [4] Y. Boureau, F. Bach, Y. LeCun, J. Ponce, *Learning mid-level features for recognition*, CVPR, 2010.

Joint localisation and identification of acoustical sources with structured-sparsity priors

Gilles Chardon and Laurent Daudet

Institut Langevin - ESPCI

10 rue Vauquelin, 75231 Paris CEDEX 05, France

firstname.lastname@espci.fr

Abstract—This work introduces a localisation and identification method for acoustical sources from array measurements. It is based on group-sparsity priors on the acoustical field, assumed to be produced by a small number of sources, but with unspecified directivity. The method is tested in a passive case, as well as an active setup, where reflectors are illuminated by a transducer array.

I. INTRODUCTION

Localisation of sources is a classical problem in array processing, for which numerous methods have been designed, including schemes based on sparsity priors [1]. However these methods only recover elementary sources, and cannot treat the case of complex directivities.

In this work, we develop a method aiming at jointly localising acoustical sources and identifying their directivities. It is based on a decomposition of the acoustical field on a dictionary of elementary sources (monopoles, dipoles, quadripoles, etc.) and group sparsity priors on the decomposition.

The method is tested in two cases :

- acquisition with a passive in audible range, where acoustical sources are to be localized from the field produced on a microphone array;
- underwater acquisition with an active ultrasonic array, when reflectors (wires) are to be characterized from the retrodiffused field after being illuminated by the transducer array.

II. SPARSITY MODEL

Sources are assumed to be sparse in space, and their directivities limited to low-order spherical harmonics (in this work, monopoles and dipoles, but the generalisation to higher-order harmonics is straightforward).

In the passive case, the harmonic field radiated at the point x_i by a source localised at the point x_i can be expressed as

$$p_{ij} = \alpha h_0(|x_i - x_j|) + \beta h_1(|x_i - x_j|) \sin(\theta_{ij}) + \gamma h_1(|x_i - x_j|) \cos(\theta_{ij})$$

where θ_{ij} is the angle between $\vec{x_i x_j}$ and a reference axis, h_0 and h_1 are Hankel functions of order 0 and 1. The vector of the field produced by a source on the array can be decomposed as a linear combination of three vectors, corresponding to the three first harmonics : $\mathbf{m}_i = (h_0(|x_i - x_j|))_j$, $\mathbf{d}_i = (h_1(|x_i - x_j|) \sin(\theta_{ij}))_j$, $\mathbf{d}'_i = (h_1(|x_i - x_j|) \cos(\theta_{ij}))_j$. The sources being sparse, the total field measured is a sum of a small set of such vectors. Formally the measurement vector can be decomposed as

$$\mathbf{p} = \mathbf{M}\mathbf{u}_m + \mathbf{D}\mathbf{u}_d + \mathbf{D}'\mathbf{u}_{d'}$$

where \mathbf{M} , \mathbf{D} , \mathbf{D}' are the dictionaries of monopoles and dipoles, and the vectors \mathbf{u}_m , \mathbf{u}_d and $\mathbf{u}_{d'}$ have identical supports.

In the active setup, a set of measurements is obtained, with the reflectors illuminated by a different transducer at each measurement. In this case, the field measured for a single source illuminated by the transducer k has the same expression than in the active case, but

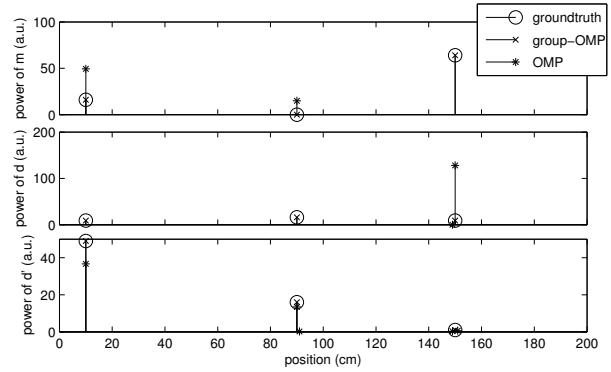


Fig. 1. Results of OMP and group-OMP for three active sources

the coefficients of the expansion α_k , β_k and γ_k are dependent on the relative positions of the transducers and the reflector, as well as the directivity of the emitting transducer. With multiple sources, as the reflectors are assumed to be fixed between the measurements, the total field measured keeps the same sparsity pattern. Here the measurement vector for an given illumination can be decomposed as

$$\mathbf{p}^k = \mathbf{M}\mathbf{u}_m^k + \mathbf{D}\mathbf{u}_d^k + \mathbf{D}'\mathbf{u}_{d'}^k$$

where the vectors \mathbf{u}_m^k , \mathbf{u}_d^k and $\mathbf{u}_{d'}^k$, for every k , have identical supports.

III. RECOVERY

Algorithms used to recover the sources include mixed norms minimisation [2], as well as the adaptation of Orthogonal Matching Pursuit (OMP) to group-sparsity, where, at each iteration, the triplet $(\mathbf{m}_i, \mathbf{d}_i, \mathbf{d}'_i)$ maximizing the norm of the orthogonal projection on $\text{span}(\mathbf{m}_i, \mathbf{d}_i, \mathbf{d}'_i)$ is selected.

Figure 1 shows a comparison of OMP and group-OMP in a simple case, with three active sources. OMP, with 9 iterations, recovers the localisations but fails at identifying the directivities, while group-OMP, with the same dictionary, recovers both.

Simulations and experimental results will be compared for both passive and active setup, with different algorithms.

REFERENCES

- [1] D. M. Malioutov, M. Cetin, and A. S. Willsky, “A sparse signal reconstruction perspective for source localization with sensor arrays.” *IEEE Transactions on Signal Processing*, pp. 3010–3022, 2005.
- [2] M. Kowalski, “Sparse Regression Using Mixed Norms,” *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 303–324, 09 2009.

An Alternating Direction Algorithm for (Overlapping) Group Regularization

Mário A. T. Figueiredo and José M. Bioucas-Dias
 Instituto de Telecomunicações, Instituto Superior Técnico, 1049-001 Lisboa, Portugal

I. INTRODUCTION

The sparsity-inducing nature of the ℓ_1 norm makes it one of the most popular convex regularizers in signal processing, statistics, and machine learning. Although independently proposed in several communities, it is best known by LASSO, the designation proposed in [9], where it was introduced as a variable selection tool. More recently, the interest in ℓ_1 regularization was boosted by its central role in *compressive sensing* [1], [2]. The non-differentiable nature of the ℓ_1 norm has stimulated a large amount of research on efficient algorithms for solving the optimization problems resulting from its use as a regularizer (see [10] and the many references therein).

In some scenarios, it makes sense to select/remove (disjoint) groups of variables, rather than individual ones, which is achieved using group-norms (e.g., [10], [11], [12]). Several researchers have extended this approach by allowing the groups to overlap, as a means of expressing preference for certain structural relationships (namely, hierarchies) among the selected/removed variables [5], [6], [7], [12].

The optimization problems that result from adopting group-structured regularizers with overlapping groups are considerably more challenging than those involving simple ℓ_p norms or group-norms with non-overlapping groups; the reason is that the overlaps destroy the separability between groups that crucially underlies the simplicity of the algorithms devised for the non-overlapping case.

In this paper, we propose tackling the optimization problems resulting from the adoption of group-structured regularizers by using a particular instance of the alternating direction method of multipliers (ADMM, [3]), recently introduced in [4], which involves no assumptions on the overlapping structure (or lack thereof) of the groups.

II. PROBLEM FORMULATION

We consider the optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + r(\mathbf{x}), \quad \text{with } r(\mathbf{x}) = \sum_{i=1}^k \lambda_i \phi_i(\mathbf{x}_{G_i}), \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{n \times p}$ is a matrix, $\mathbf{x}_{G_i} \in \mathbb{R}^{|G_i|}$ is the sub-vector of \mathbf{x} corresponding to the indices in the i -th group $G_i \subseteq \{1, \dots, p\}$, each ϕ_i is a convex function (the ℓ_2 norm, in group-LASSO [7], but other choices are possible), and $\lambda_1, \dots, \lambda_k$ are positive weights. No assumptions are made about group overlap: the groups may overlap or not; if they overlap, no special structure (such as a hierarchy) is assumed for the way they do so. If $k = p$, $G_i = \{i\}$, $\lambda_i = \lambda$, and $\phi_i(x_i) = |x_i|$, we recover standard ℓ_1 regularization.

A key component of most state-of-the-art algorithms for solving problem (1) is the so-called Moreau proximity operator of r [6], [10]:

$$\text{prox}_r(\mathbf{x}) = \arg \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{x} - \mathbf{u}\|_2^2 + r(\mathbf{u}). \quad (2)$$

In the absence of group overlapping, computing prox_r boils down to computing $\text{prox}_{\lambda_i \phi_i}$, for $i = 1, \dots, k$ [10]. With overlapping groups, prox_r can only be easily computed if the groups are hierarchically structured and for some choices of the ϕ_i (ℓ_1 , ℓ_2 , or ℓ_∞ norms) [5].

III. PROPOSED APPROACH

We propose addressing problem (1) by mapping it into the form

$$\min_{\mathbf{x} \in \mathbb{R}^p} \sum_{j=1}^m g_j(\mathbf{H}_j \mathbf{x}), \quad (3)$$

where the g_j are convex functions and the \mathbf{H}_j are matrices, as follows: $m = k+1$, $g_j = \lambda_j \phi_j$, for $j = 1, \dots, k$, $g_{k+1}(\mathbf{u}) = \|\mathbf{u} - \mathbf{y}\|_2^2$, $\mathbf{H}_{k+1} = \mathbf{A}$, and (for $j = 1, \dots, k$) \mathbf{H}_j is a $|G_j| \times p$ matrix with the subset of rows of the identity corresponding to group G_j . Then, we handle this problem using the algorithm proposed in [4] (which is an instance of ADMM). We show convergence of the algorithm, regardless of matrix \mathbf{A} , as long as any index $i \in \{1, \dots, p\}$ belongs to at least one group.

Each iteration of the algorithm involves computing the proximity operator of each function g_j and minimizing a quadratic function. For problems of moderate size, the Hessian of this function can be inverted only once, with cost $O(\min\{n, p\}^3)$, and this inverse used throughout the iterations. For large problems, we avoid the cost of the inversion by using to a (warm-started) Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, or a limited memory version thereof [8].

We report experiments on identifying hyperspectral signatures on large dictionaries (equipped with a non-hierarchical overlapping group structure); one of the regularizers in (1) is, in this case, the indicator of the probability simplex (on the full vector \mathbf{x}). To the best of our knowledge, this problem is out of the reach of other recent algorithms for overlapping group regularization.

REFERENCES

- [1] E. Candès, J. Romberg, T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information”, *IEEE-TIT*, vol. 52, pp. 489–509, 2006.
- [2] D. Donoho, “Compressed sensing”, *IEEE-TIT*, vol. 52, pp. 1289–1306, 2006.
- [3] J. Eckstein, D. Bertsekas, “On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators”, *Math. Program.*, vol. 5, pp. 293–318, 1992.
- [4] M. Figueiredo, J. Bioucas-Dias, “Restoration of Poissonian images using alternating direction optimization”, *IEEE-TIP*, vol. 19, pp. 3133–3145, 2010.
- [5] R. Jenatton, J.-Y. Audibert, F. Bach, “Structured variable selection with sparsity-inducing norms”, arXiv:0904.3523, 2009.
- [6] R. Jenatton, J. Mairal, G. Obozinski, F. Bach, “Proximal methods for sparse hierarchical dictionary learning”, *Proc. ICML*, 2010.
- [7] S. Kim and E. Xing, “Tree-guided group lasso for multi-task regression with structured sparsity”, *Proc. ICML*, 2010.
- [8] J. Nocedal, S. Wright, *Numerical optimization*, Springer, 2006.
- [9] R. Tibshirani, “Regression shrinkage and selection via the lasso”, *J. Roy. Stat. Soc. (B)*, vol. 58, pp. 267–288, 1996.
- [10] S. Wright, R. Nowak, M. Figueiredo, “Sparse reconstruction by separable approximation”, *IEEE-TSP*, vol. 57, pp. 2479–2493, 2009.
- [11] M. Yuan, Y. Lin, “Model selection and estimation in regression with grouped variables”, *J. Roy. Stat. Soc. (B)*, vol. 68, pp. 49–67, 2006.
- [12] P. Zhao, G. Rocha, B. Yu, “The composite absolute penalties family for grouped and hierarchical variable selection”, *Annals of Stat.*, vol. 37, pp. 3468–3497, 2009.

Sparse Approximation of the Neonatal EEG

Vladimir Matić

Department of Electrical Engineering
ESAT-SCD and IBBT
Katholieke Universiteit Leuven, Belgium
Email: vladimir.matic@esat.kuleuven.be

Maarten De Vos

Neuropsychology Lab
Oldenburg University, Germany
ESAT-SCD and IBBT
Katholieke Universiteit Leuven, Belgium

Bogdan Mijović

and Sabine Van Huffel
Department of Electrical Engineering
ESAT-SCD and IBBT
Katholieke Universiteit Leuven, Belgium

I. INTRODUCTION

At the Neonatal Intensive Care Units, continuous electroencephalographic (EEG) recordings are regularly performed for the assessment of hypoxic brain injuries of newborns. Nowadays, there is a tendency for the development of wireless EEG devices, that would decrease the amount of movement artifacts and provide a comfortable surrounding for the babies. One of the major issues is the large quantity of data that has to be transmitted over the wireless link - approximately 20 EEG channels with a sampling frequency (f_s) of 256Hz . This significantly affects the battery life, as the recordings should be continuous for a period of 48 up to 72 hours. We are investigating the applicability of the compressive sensing theory for this purpose. Therefore, finding the sparse approximation of the complex neonatal EEG morphology is the preliminary step of this work.

II. METHODS

In order to provide a sparse representation of the neonatal EEG signal, several bases have been explored, namely wavelets, Discrete Cosine Transform, Slepian and local cosine basis. However, none of these orthonormal bases have managed to provide an accurate approximation when retaining only a relatively low number of coefficients. It has been shown that an adult EEG signal can be sparsely approximated in an overcomplete Gabor dictionary [1] and that compressive sensing theory can be applied for the acquisition process. Following the same approach, neonatal EEG represented in Gabor dictionary yields to a nearly sparse decomposition as well.

For the reconstruction purpose several algorithms have been tested: OMP, BP and IHT. Due to its simplicity and fast and accurate performance with a relatively low number of measurements, IHT has been chosen for further experiments. The entries of the sensing matrix were chosen as i.i.d. Gaussian.

III. RESULTS

An overcomplete Gabor dictionary has been created with an atom length of 1024 samples and it consisted of 40.960 atoms. In that way, we can represent 4 seconds ($f_s = 256\text{Hz}$) of the EEG signal, with Signal-to-Error Ratio (SER) varying from 5% to 25%, and a Normalized RMSE (NRMSE) between 1% and 5% using 100 to 300 measurements (0.095 – 0.3 compression rate). As a preliminary study we processed 300 EEG segments, with different number of measurements. Significant number of these segments showed very complex morphology (background EEG) for which the reconstruction error was usually higher than for structured EEG patterns.

Obtained results suggest that we can accurately reconstruct EEG patterns with highly nonstationary dynamics with only 15 – 30% of measurements, whereas the structured and simpler wave patterns we can reconstruct with as little as 5 – 10% measurements. In that sense we can achieve a desired compression from 10 up to 30 percentage with respect to the reconstruction error.

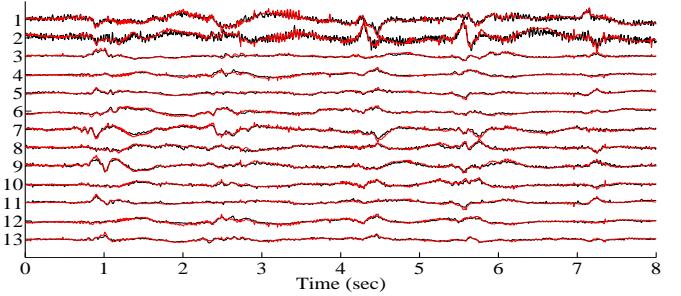


Fig. 1. The reconstruction of 13 channels of a moderately complex signal with 150 measurements (compression rate 14.65)

IV. DISCUSSION

The objective of previously presented approach was to examine whether neonatal EEG can be sparsely approximated in a redundant dictionary. As the accuracy of the reconstruction is concerned we can consider EEG as a combination of two separate parts: background EEG and structured EEG patterns. Background EEG usually represents ongoing, rather chaotic part of the EEG activity (burst and suppression intervals) which does not possess much informative features. Therefore, from clinical point of view, slightly higher reconstruction errors can be tolerated for these EEG segments. On the other hand, structured EEG patterns such as epileptic seizures are represented either by an oscillatory EEG behaviour or by recurrent spike train series. Occurrences of such morphologies are very important for clinicians and high accuracy in reconstruction is necessary. Fortunately, these EEG segments are morphologically simpler and can be accurately reconstructed with smaller number of measurements. In this work we have used a highly redundant Gabor dictionary. As the future work is concerned, parametric dictionary design should be incorporated, which will provide better tiling of the time frequency plane. In that way we can obtain a dictionary with higher incoherence and with higher exact recovery conditions [2]. In the sense of sparse approximation, we will investigate the possible application of algorithms for dictionary learning [3]. We hope that application of various dictionary learning techniques will enable us better sparsification and classification of the desired EEG patterns.

REFERENCES

- [1] S. Aviyente, *Compressed sensing framework for EEG compression*, in IEEE/SP SSP, Madison, 2007.
- [2] M. Yaghoobi, L. Daudet, and M. E. Davies, *Parametric dictionary design for sparse coding*, IEEE Transactions on Signal Processing, vol. 57, no. 12, 2009, pp. 4800-4810.
- [3] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, *Online learning for matrix factorization and sparse coding*, Journal of Machine Learning Research, vol. 11, 2010, pp. 19-60.

Inversion of 2-D images to estimate densities in \mathbf{R}^3

Dalia Chakrabarty
 Department of Statistics
 University of Warwick
 Coventry, CV4 7AL
 Email: d.chakrabarty@warwick.ac.uk

Fabio Rigat
 Department of Statistics
 University of Warwick
 Email: f.rigat@warwick.ac.uk

I. INTRODUCTION

The estimation of densities in \mathbf{R}^3 , given 2-D images involves the inversion of the exercise of projection of the convolution $\rho(\mathbf{x}) \circ \eta(\mathbf{x})$ of the unknown density $\rho(\mathbf{x})$ with a blurring or correction function $\eta(\mathbf{x})$, (which may or may not be an unknown), along a given orientation. Here, $\rho(\mathbf{x}) \in \mathbf{R}^3$ is piecewise continuous but may not be piecewise smooth, is non-negative and bounded in \mathbf{R}^3 . The correction function $\eta(\mathbf{x}) > 0$ is in general, $\eta : \mathbf{R}^3 \rightarrow \mathbf{R}^3$. The traditional implementation of the inverse Radon transform is rather limited if the orientation is not a measurable, (Panaretos 2009, Chakrabarty 2010, Chakrabarty et. al 2008) or if there is noise in the data or, as in the application we discuss, if the underlying 3-D structure is multimodal, with the isolated modes manifesting sharp boundaries and individual substructure - in which case a Gaussian mixture model is insufficient. We present a new methodology that allows for a non-parametric reconstruction of $\eta(\mathbf{x})$ and the heterogeneous $\rho(\mathbf{x})$ of a cuboidal slab of a given material sample \mathcal{S} , by performing an inversion of 2-D images recorded in electron scattering experiments.

Learning of 2 unknown functions from a single image is an ill-posed inverse problem which in the Bayesian approach that we adopt here, will admit only prior-driven solutions. The situation supplemented by only weak priors on $\rho(\mathbf{x})$ readily suggests enhancement of sparsity in the models and/or expansion of information domain. In fact, both are implemented in our methodology - we increase sparsity by invoking the inherent smoothness that is imposed by the data and expand information by suggesting the recording of multiple images at multiple values of a model parameter, namely beam energy.

In the experiments, electron beams of different energies E_k , $k = 1, \dots, N_{eng}$ are made incident at different points on \mathcal{S} , with a uniform distance δ between the i^{th} and $i + 1^{th}$ beam pointings, ($i = 1, \dots, N_{data}$) where δ is set by the relevant instrumentation. The atomistic interactions between the beam electrons and the material atoms causes the distribution of the electron mean-free paths to become pear-shaped, and the pear size increases with beam energy. In our model we approximate the pear formed at the k^{th} energy, at the i^{th} pointing, as a hemi-sphere centred on the i^{th} beam pointing $(x_1^{(i)}, x_2^{(i)})$, with (penetration depth $h_k^{(i)}$ equal to) radius $R0_k^{(i)}$, $\forall i, k$. The recorded 2-D radiation density from this pear is $I_k(x_1^{(i)}, x_2^{(i)})$. Then

$$I_k(x_1^{(i)}, x_2^{(i)}) = \frac{\int_0^{R0_k^{(i)}} [\int_0^{h_k^{(i)}} dz \int_{-\infty}^{\infty} \rho(R, z) \eta(z - x_3, \gamma, \theta) dx_3] 2\pi R dR}{\int_0^{R0_k^{(i)}} 2\pi R dR}, \quad (1)$$

where $R^2 := (x - x^{(i)})^2 + (y - y^{(i)})^2$, $R0_k^{(i)}$ is the maximal radius of this pear. An axisymmetric geometry is assumed for the density within each pear and we assume $\eta : X_3 \rightarrow X_3$.

The 3-D desity structure that gives rise to the data discussed above, is viewed as a tree with the i^{th} beam location identified

as the i^{th} node, ($i = 1, \dots, N_{data}$), and the k^{th} attribute at any node is the density in the inter-peer volume between the k^{th} and $k - 1^{th}$ pears, (0^{th} peer has zero volume, $k = 1, \dots, N_{eng}$), i.e. as $\rho_k^{(i)} := \rho(x_i, y_i, z)$, $z \in [h_{k-1}^{(i)}, h_k^{(i)}]$. In this work, we attempt an identification of the density tree by representation in terms of separable basis functions, the choice of which is motivated to reflect the smoothness that the data imposes. This case reflects the inter-nodal independence of attributes at the each node. We refer to this situation as Case I. If however, $\hat{\rho}(x_i, y_i, z)$ depends on beam pointings other than the i^{th} one, the set of basis functions are still separable in z but have richer inter-nodal dependence that can be modelled using nearest-neighbour contributions, the closed-form solutions of which have been identified, using Stoke's Theorem from differential geometry (Case II).

For Case I, we have the simple recursive relation $\alpha_k^{(i)} = \frac{I_k^{(i)} - I_{k-1}^{(i)}}{\int_{h_{k-1}^{(i)}}^{h_k^{(i)}} \eta(z) dz}$, $I_0^{(i)} = h_0^{(i)} = 0$, $\forall i$. The mean structure of the assumed Gaussian likelihood, is borrowed from this relation as the product $\alpha_k^{(i)} \int_{h_{k-1}^{(i)}}^{h_k^{(i)}} \eta(z) dz$. The variance is the noise in the data. This likelihood is then used to write the posterior probability for the unknowns, given the data, and truncated normal priors. This high dimensional posterior is sampled from using an adaptive Metropolis-Hastings (Hario et. al. 2005), to learn $\rho(x, y, z)$ and $\eta(z)$.

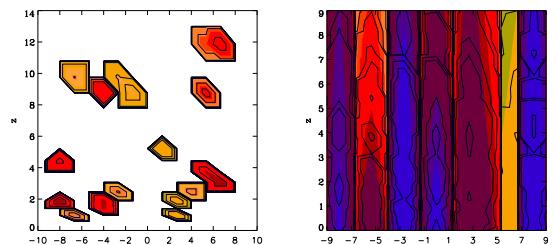


Fig. 1. Slice though the $Y = 0$ plane of the learnt density is shown as a contour plot in the $X - Z$ space for 2 different simulated data. The density estimate at the median of the inferred 90% credible region is showed in solid lines, superimposed on the true density (in filled colored contours).

REFERENCES

- [1] Chakrabarty, D., *Astronomy & Astrophysics*, 510, 45, 2010.
- [2] Chakrabarty, D., de Philippis, B. & Russell, H., *Astronomy & Astrophysics*, 487, 75, 2008.
- [3] H. Haario, E. Saksman & J. Tamminen, *Computational Statistics*, 20, 265, 2005.
- [4] Panaretos, V.M., *Annals of Statistics*, 37, 3272, 2009.

Constrained Non-Negative Matrix Factorization for source separation in Raman Spectroscopy

Hérald Rabeson

IFP Energies nouvelles

1 et 4 avenue du Bois-Préau, 92852 Rueil-Malmaison Cedex, France

herald.rabeson@ifpen.fr

Abstract—Raman spectroscopy is a powerful tool for quantitative analysis of mixtures, it is relatively fast and sensitive and it allows to follow the distribution of chemical species according to an evolution parameter. Non-negative matrix factorization (NMF), in a constrained version, is the source separation method chosen to estimate the chemical species and their concentrations. Influence of noise level, peak shifts or broadening are compared through Monte-Carlo simulations.

I. INTRODUCTION

For quantitative analysis of Raman spectra (Fig. 1), self-modelling curve resolution methods [1], [2], [3] have become the standard tools in the last 20 years. Nevertheless, the NMF approach, introduced in other fields [4], [5], has been used for different Raman applications during the last years such as detection of target spectrum [6] or separation of specific markers [7].

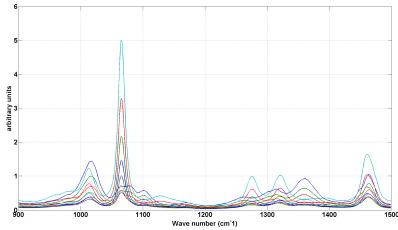


Fig. 1. Raman spectra mixtures

II. METHODS

The NMF problem consists in finding a factorization of a non-negative matrix \mathbf{V} of size $F \times N$ with non-negative matrices \mathbf{W} and \mathbf{H} of sizes $F \times K$ and $K \times N$, respectively:

$$\mathbf{V} \approx \mathbf{WH} \quad (1)$$

In our case, K is the number of chemical species supposed to be known here, F the number of experimental spectra and N the number of points per spectra. Standard NMF approach takes into account the non-negativity constraints of the data but the non-uniqueness of the solutions is an issue. Sparseness of the sources can be controlled as presented in [10]. Integrating additivity and non-negativity constraints can be solved using a bayesian approach if computational time is not a constraint [9]. In the present case, a constrained version of the original NMF is used using the formalism in [8]. Given a matrix \mathbf{V} , the problem is finding an approximation with non-negative matrices \mathbf{W} and \mathbf{H} by minimizing a cost function such as

$$f(\mathbf{W}, \mathbf{H}) = \|\mathbf{V} - \mathbf{WH}\|^2 + \alpha J_1(\mathbf{W}) + \beta J_2(\mathbf{H}) \quad (2)$$

As pure spectra of chemical species were available, they were used for the initialization of \mathbf{W} which is crucial. Nevertheless, these spectra do not always match with the experimental spectra of each species in the mixture solutions. Thus, a smooth regularization is ensured by setting $J_1(\mathbf{W}) = D_\alpha \|W - W_0\|^2$ following approaches in [8], [7].

III. RESULTS

The use of a constrained version NMF allowed to obtain more robust results (compared to standard NMF even in the case of specific initialization of \mathbf{W}) for both spectral sources and concentrations (Fig. 2), through different Monte-Carlo simulations on noise level, peak shifts or broadenings.

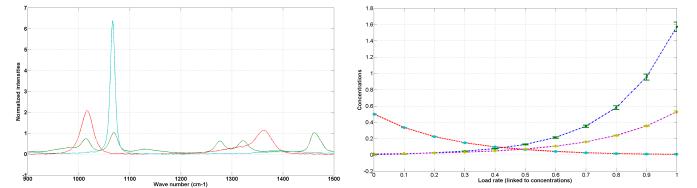


Fig. 2. Left: Pure (initial) reference spectra, corresponding to the columns of \mathbf{W} . Right: Mean estimated concentration evolutions (lines of \mathbf{H}) with associated standard deviations for each species, in the case of shifted peaks.

IV. CONCLUSION

With constrained NMF, a more robust quantitative analysis for Raman spectra is achieved. Use of chemical knowledge (such as additivity) on the concentration laws should further improve the accuracy of the method.

REFERENCES

- [1] R. Tauler, A. Izquierdo-Ridorsa, and E. Casasas. Simultaneous analysis of several spectroscopic titrations with self-modeling curve resolution. *Chemometrics and Intelligent Laboratory Systems*, 18:293–300, 1993.
- [2] W. Windig. Spectral data files for self-modelling curve resolution with examples using the simplisma approach. *Chemometrics and Intelligent Laboratory Systems*, 36:3–16, 1997.
- [3] M. Garrido, F. X. Rius, and M. S. Larrechi. Multivariate curve resolution-alternating least squares (mcr-als) applied to spectroscopic data from monitoring chemical reactions processes. *Analytical and bioanalytical chemistry*, 390:2059–2066, 2008.
- [4] P. Paatero and U. Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5:111–126, 1994.
- [5] D. Lee and H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.
- [6] H. Li, T. Adali, W. Wang, D. Emge, and A. Cichocki. Non-negative matrix factorization with orthogonality constraints and its application to raman spectroscopy. *J. VLSI Signal Process. Syst.*, 48:83–97, 2007.
- [7] A.-S. Montcuquet, L. Hervé, F. Navarro, J.-M. Dinten, and J. I. Mars. Nonnegative matrix factorization: a blind spectra separation method for *in vivo* fluorescent optical imaging. *J. Biomed. Opt.*, 15, 2010.
- [8] M. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics and Data Analysis*, pages 155–173, 2007.
- [9] N. Dobigeon, S. Moussaoui, J. Y. Tourneret, and C. Carteret. Bayesian separation of spectral sources under non-negativity and full additivity constraints. *Signal Processing*, 8:2657–2669, 2009.
- [10] P. O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 5:1457–1469, 2004.

Sparse Templates-Based Shape Representation for Image Segmentation

Stefania Petra, Dirk Breitenreicher, Jan Lellmann and Christoph Schnörr
 Image & Pattern Analysis Group
 University of Heidelberg, Germany
 Email: {petra,breitenreicher,lellmann,schnoerr}@math.uni-heidelberg.de

Abstract—We present a new approach to image segmentation based on sparse coverings of the image domain by shape templates. The basic idea is to formulate the segmentation problem as a sparse representation problem, utilizing new mathematical tools from ℓ_1 -minimization and compressed sensing.

METHODS AND RESULTS

Given a large set of shape templates and a pre-segmentation, we are required to segment a noisy image where objects may overlap by taking into account prior knowledge about the shape of the objects and their parts. The parts may be partially occluded and the location and nature of occlusion is unknown. This can be modeled as a sparse error that affects only a few pixels in the input image while the "true" segmentation is represented as a sparse linear combination of the entire shape-templates training set.

Unfortunately, the shape dictionary - built by stacking all the training shapes and their translations to all pixel positions as column vectors - is not incoherent, but a truly redundant dictionary. As a consequence, all currently available theoretical recovery conditions predict a poor performance of the ℓ_1 -minimization approach (i.e. exact recovery in the coefficient domain). However, we show empirically, see Fig. 2, that accurate recovery is possible for moderate sparsity of the basic templates and dense errors, similar to the work in [1].

We discuss the implications of these results on our application, and illustrate our approach on real world images, see Fig. 1, by numerical examples that employ large-scale convex programming.

ACKNOWLEDGMENT

The first author gratefully acknowledges support by the German Science Foundation (DFG), grant SCHN457/11-1.

REFERENCES

- [1] John Wright and Yi Ma. Dense error correction via ℓ_1 -minimization. *IEEE Trans. Inf. Theor.*, 56:3540–3560, July 2010.



Fig. 1. Separating chain links from the background and from each other by convex optimization in terms of a sparse covering of the image by shape templates. The dictionary of shape templates was generated from four templates by translation, rotation and scaling. The approach presented in this work copes with a significant amount of overlapping templates and occlusion. *Left to right:* Original image, pre-segmentation using a thresholded distance to the color red as foreground indicator, shape templates used for segmentation, and the final result.

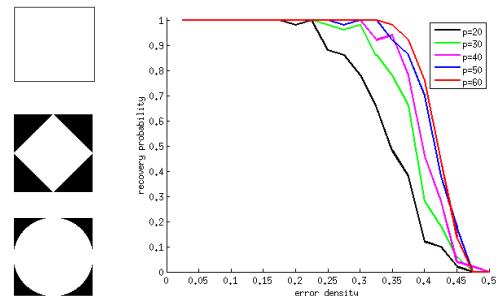


Fig. 2. The three templates (*left*) together with their translations build a dictionary. The true segmentation consisting of a sparse covering with only a few templates is recovered exactly. Recovery performance for increasing density of the error (including both salt and pepper noise and occlusion) improves with increasing image resolution (*right*).

Wyner-Ziv Coding for Distributed Compressive Sensing

Kezhi Li, Su Gao, Cong Ling

Department of Electrical and
Electronic Engineering

Imperial College London, UK

Email: {k.li08, su.gao06, c.ling}@imperial.ac.uk

Lu Gan

School of Engineering and Design

Brunel University, UK

Email: lu.gan@brunel.ac.uk

I. ABSTRACT

Distributed source coding (DSC)/compression, pioneered by Slepian and Wolf [1], and Wyner and Ziv [2] in 1970's, is an important topic of signal processing and information theory. DSC exploits the correlation of the input data by different nodes to reduce the transmission rate. In many cases, the coding can be as efficient as if the sensors were co-located and performed joint encoding. Compressed sensing (CS) [3], [4] has gained significant interest in many theoretical and applied areas because it permits simultaneous sensing and compression. Lately, a new framework called distributed compressed sensing [5], [6] has been introduced to exploit both intra- and inter-signal correlations of the distributed signals from a CS approach. The technique resembles DSC in both problem formulation and applications.

Since distributed CS is an analog technique, a fundamental open question is to find the best source coding scheme for the distributed CS samples. As a first step towards answering this question, this paper presents some initial results on distributed source coding in this context by exploiting the correlation among the CS samples at different sensors. The whole architecture of our system consists of two parts, which are distributed CS using the Toeplitz sensing matrix and DSC using nested lattices. The framework is depicted in Fig. 1.

The correlated sources are first processed separately using CS. Then the samples are sent into the second stage for DSC. After transmission through the lossless channel, at the receiver is the joint source decoder followed by CS recovery. We mostly use Wyner-Ziv coding [2] based on the nested lattice scheme, where only one source is encoded lossy and transmitted to the decoder, and the signal is reconstructed with a fidelity criterion under the assistance of the side information which is the other signal.

Extensive simulations have been carried out to examine the reconstruction performances of different images. For illustration purposes, we only present here the recovery result of a tank image "tank1", while our conclusions drawn are applicable to other input sources. The image size is $150 \times 330 = 49500$ pixels. And we use another image "tank2", which is slightly different as the side information for jointly decoding. The reconstruction result is shown in Fig. 2. The

compression rate of CS is 20% in this simulation. Of course, higher sampling rate means better recovery performance. The result (c) shows that the original image can be recovered with acceptable PSNR.

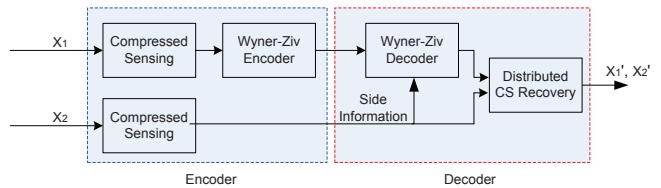


Fig. 1. Distributed CS based Wyner-Ziv coding.

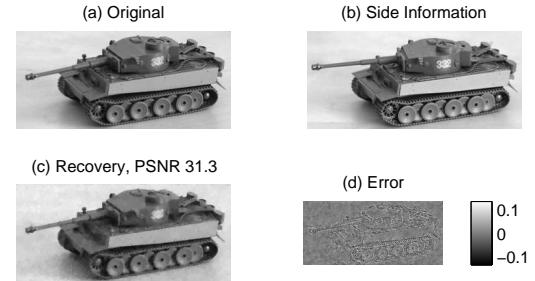


Fig. 2. (a) Original image "tank1". (b) Side information image "tank2". (c) Recovery result of proposed Wyner-Ziv Distributed CS. (d) The error between original image and recovery result.

REFERENCES

- [1] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471–480, 1973.
- [2] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, 1976.
- [3] D. L. Donoho, "Compressed sensing," vol. 52, pp. 1289–1306, July 2006.
- [4] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies," vol. 52, pp. 5406–5425, 2006.
- [5] M. F. D. S. S. Dror Baron, Michael B. Wakin and R. G. Baraniuk, "Distributed compressive sensing," 2005, preprint.
- [6] T. Do, Y. Chen, D. Nguyen, N. Nguyen, L. Gan, and T. Tran, "Distributed compressed video sensing," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, 7-10 2009, pp. 1393 –1396.

Methods for Training Adaptive Dictionary in Underdetermined Speech Separation

Tao Xu

Centre for Vision, Speech and Signal Processing
University of Surrey
Email: t.xu@surrey.ac.uk

Wenwu Wang

Centre for Vision, Speech and Signal Processing
University of Surrey
Email: w.wang@surrey.ac.uk

I. INTRODUCTION

Underdetermined speech separation is a challenging problem that has been studied extensively in recent years. A promising method to this problem is based on the so-called sparse signal representation. Using this technique, we have recently developed a multi-stage algorithm [1], where the source signals are recovered using a fixed dictionary obtained by e.g. the discrete cosine transform (DCT). In this abstract, instead of using the fixed dictionary, we present three methods for training adaptive dictionaries for the reconstruction of source signals, and compare their performance.

II. STRATEGIES FOR TRAINING THE ADAPTIVE DICTIONARY

Following our previous work [1], here we propose a separation system depicted in Figure 1 for the case of four sources and two mixtures. In this system, the mixing matrix is estimated in the transform domain by a clustering algorithm as in [1]. However, different from [1], the source signals are reconstructed from multiple adaptive dictionaries with each obtained by one of the alternative training strategies described below.

In the first strategy (STD), for each source, we train a dictionary. Therefore four different dictionaries are trained from the four original sources respectively. They are then combined to form a single dictionary matrix for separating the source in the following stages. In the second strategy (ESTD), the dictionaries are learned from the coarsely estimated sources which can be obtained, for example, from the traditional DCT based separation method as described in [1]. In the third strategy (MTD), a single dictionary is directly learned from the mixtures. In the upper part of Figure 1, the ESTD method is depicted.

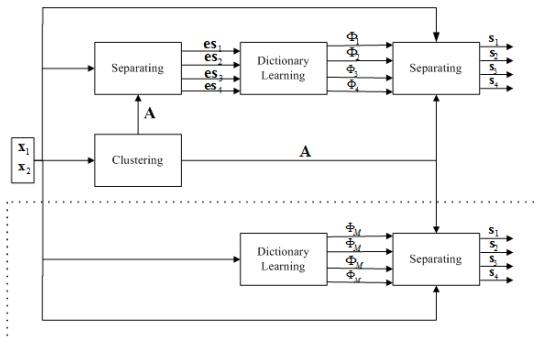


Fig. 1. The flow chart of the proposed system for separating four speech sources from two mixtures.

Firstly the sources are estimated from the mixtures by using e.g. the DCT. Secondly, the dictionaries are learned from these four coarsely separated sources, whose atoms are then used to reconstruct the sources at the second separating stage. In the dashed box, dictionary learning based on the MTD method, i.e. directly from the mixtures, is

illustrated. For the clustering and separating stages, the same method as in [1] has been used. In all the three dictionary learning strategies, the K-SVD algorithm [2] was used to obtain the dictionary atoms.

III. EXPERIMENTAL RESULTS

In this section, we evaluate the proposed algorithm by performing the experiments using four speech sources in the TIMIT database, which are English male (EM), English female (EF), Japanese female (JF) and Chinese female (CF) speech respectively. The sources have a duration of 5 seconds, sampled at 10 kHz. For objective quality assessment, we use the two global performance criteria defined in the BSSEVAL toolbox [3] to evaluate the estimated source signals, which are the signal to distortion ratio (SDR) and the source to interference ratio (SIR).

Based on the estimated mixing matrix obtained from the clustering stage, we can recover the four speech sources using the DCT dictionary and the adaptive dictionaries based on the STD, ESTD and MTD methods. The results are presented in Table I and II. From these

	DCT	STD	ESTD	MTD
EM speech	7.59	9.89	6.93	-1.41
EF speech	9.53	11.44	9.26	3.54
JF speech	2.73	7.38	2.13	-4.22
CF speech	14.59	15.05	14.13	8.91

TABLE I
SDR (IN DB) MEASURED FOR EACH ESTIMATED SPEECH SOURCE.

	DCT	STD	ESTD	MTD
EM speech	14.23	19.47	14.49	2.81
EF speech	11.35	30.49	11.35	5.25
JF speech	6.07	22.21	5.97	-2.12
CF speech	18.12	25.89	18.34	12.50

TABLE II
SIR (IN DB) MEASURED FOR EACH ESTIMATED SPEECH SOURCE.
tables, we can observe that the separation performance using STD trained dictionary is considerably better than using the DCT dictionary. Using the ESTD trained dictionary, the results are close to the DCT dictionary. However, it is difficult to obtain good results by using the dictionary learned from the mixtures, i.e. the MTD method. These results suggest that the properly learned dictionaries outperform the fixed dictionary in underdetermined speech separation.

REFERENCES

- [1] T. Xu and W. Wang, "A block-based compressed sensing method for underdetermined blind speech separation incorporating binary mask," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 2022 – 2025.
- [2] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [3] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. on Audio, Speech & Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.

Analysis of Subsampled Circulant Matrices for Imaging

Matthew A. Turner, Lina Xu, and Kevin F. Kelly
 Rice Quantum Institute
 Department of Electrical and Computer Engineering
 Rice University
 Houston, Texas, USA 77005
 Email: kkelly@rice.edu

Wotao Yin
 Department of Computational and Applied Mathematics
 Rice University
 Houston, Texas, USA 77005
 Email: wotao.yin@rice.edu

Abstract—Motivated by experiments with our homebuilt compressive imaging system, we numerically explore subsampled circulant matrices for compressive imaging, imaging based on the theory of compressive sensing, to show that different constructions of such matrices have distinct phase diagrams in so-called “undersampling phase space.” Such an investigation will be useful for guiding principles of instrument design where hardware constraints must be balanced with instrument performance.

I. INTRODUCTION

Recent work has detailed the promise and theory of circulant matrices as one possible strategy in compressive sensing [1], [2], where generally one is interested in solving an intentionally underdetermined system of equations represented by $y = \Phi x$. Here y are our observations, which in an optical system is a voltage registered digitally from an analog photodetector that converts light intensity to a corresponding electric potential. Each individual measurement can be represented as $y_i = \langle \phi_i, x \rangle$, where ϕ_i is the i^{th} reshaped row of Φ and x is the scene to be imaged [3]. Following the formalism of [1], we have constructed our measurement matrix Φ by taking m rows from an $N \times N$ circulant matrix Φ° , or mathematically,

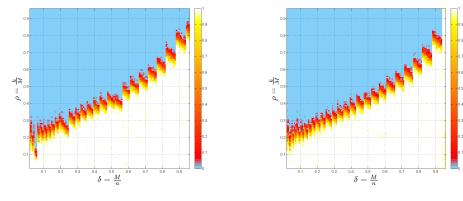
$$\Phi = \frac{1}{\sqrt{M}} R_\Omega \Phi^\circ, \quad (1)$$

where R_Ω selects M rows from Φ° according to the index set $\Omega \subset \{1, 2, \dots, n\}$ with cardinality equal to the number of measurements $|\Omega| = M$. In this work we show that for circulant matrices generated by a Bernoulli seed vector with entries from $\{0, 1\}$, there is a marked difference in the performance of such subsampled circulant measurement matrices depending on the method for selecting the index set Ω .

II. SUMMARY OF INVESTIGATION

During initial tests with random circulant matrices in our homebuilt compressive imaging system, we noticed that images taken with Ω built sequentially, i.e. $\Omega = \{1, 2, 3, \dots, M\}$, are lower quality than those taken with measurement matrices Φ built from randomly chosen indices in Ω . Recently Donoho and Tanner presented a method for empirically testing where in so-called “undersampling phase space” a certain reconstruction algorithm with specified measurement ensemble transitions from low probability of success to high probability of success [4]. We apply this method to our specific case of subsampled circulant matrices as measurement ensembles. With the CVX

convex optimization package [5], we solve both the noiseless basis pursuit and linear program problems over many points in undersampling phase space to explore the differences between the phase diagram for randomly subsampled circulant matrices the phase diagram for sequentially subsampled circulant matrices. We chose CVX for initial tests because it is suitably robust whereas iterative solvers we tested failed. We performed at least 28 trials on 15,876 evenly spaced points ($\delta = M/n, \rho = k/M$), where N is the length of the signal, m is the number of measurements, and k is the sparsity of the signal (see Figure). We further discuss the implications for instrument design in general.



(a) Sequentially subsampled (b) Randomly subsampled

Figure: Phase diagrams from solving BP with sequentially and randomly subsampled circulant matrices with $N_{\text{trials}} \geq 28$.

REFERENCES

- [1] H. Rauhut, J. K. Romberg, and J. A. Tropp, “Restricted isometries for partial random circulant matrices,” *arXiv*, vol. abs/1010.1847, 2010.
- [2] J. Romberg, “Compressive sensing by random convolution,” *Submitted to SIAM Journal on Imaging Sciences*, 2009.
- [3] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk, “A new compressive imaging camera architecture using optical-domain compression,” in *Proceedings of Computational Imaging IV at SPIE Electronic Imaging*, 2006.
- [4] D. L. Donoho and J. Tanner, “Precise undersampling theorems,” *Proceedings of the IEEE*, vol. 98, no. 6, June 2010.
- [5] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 1.21,” <http://cvxr.com/cvx>, Feb. 2011.

A New BCI Classification Method based on EEG Sparse Representation

Younghak Shin, Seungchan Lee and Heung-No Lee*

School of information and communications
 Gwangju Institute of Science and Technology (GIST)
 Gwangju, Republic of Korea, 500-712
 Email: shinyh, seungchan, heungno@gist.ac.kr

Abstract—Motor imagery based Brain Computer Interface (BCI) systems provide a new communication and control channel between the human and an external device with only imagination of limbs movements. Because Electroencephalogram (EEG) signals are very noisy and non-stationary, powerful classification methods are needed. We propose a new classification method based on sparse representation of EEG signals and ell-1 minimization. This method requires a well constructed dictionary. We show very high classification accuracy can be obtained by using our method. Moreover, our method shows improved accuracy over a well known LDA classification method.

I. INTRODUCTION

Motor imagery based EEG signals are very sensitive to noise and artifacts, for example caused by unwanted eye movements. Thus, powerful signal processing methods are needed. In this paper, we are interested in developing a new classification method for the BCI system. Using right hand 'R' and foot 'F' of motor imagery data sets, we propose a new sparse representation based classification (SRC) method. The SRC method is motivated from compressed sensing (CS) theory. SRC works by finding a sparse representation of the test signal in terms of the training signals included in the dictionary. To make a proper dictionary, we use a common spatial pattern (CSP) which has distinguishable property for different classes. CSP is a powerful signal processing technique suitable for EEG-based BCI system [1]. After CSP filtering, We use sensorimotor rhythms (Mu and Beta rhythm) as a feature of BCI system [2].

II. METHODS

Let N_t be the total training signals. We define a dictionary matrix $\mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, \dots, \mathbf{a}_{i,N_t}]$ for class $i = R, F$, where each column vector $\mathbf{a} \in \mathbb{R}^{m \times 1}$ is obtained by CSP filtering, FFT of the time domain signal in a training trial. By combining the two matrices, we form the complete dictionary, $\mathbf{A} := [\mathbf{A}_R; \mathbf{A}_F]$. We apply the same procedure done to obtain the columns of the dictionary to the test signal. Then, this test signal can be sparsely represented as a linear combination of some columns of \mathbf{A} . We can represent this as a matrix algebraic form: $\mathbf{y} = \mathbf{Ax}$.

We use certain FFT coefficients (Mu and Beta rhythms) as a feature, and the linear equation becomes under-determined ($m < 2N_t$). CS theory has shown that the ell-1 norm minimization can solve this under-determined system well in polynomial time [3]. Unlike the conventional ell-2 norm minimization, the ell-1 norm minimization gives a sparse representation result. In this study, we use the basis pursuit method, one of the standard linear programming methods [4].

III. RESULTS

We have analyzed five data sets, which have same 140 trials for each class. Figure 1 shows the SRC classification accuracy of all subjects. We can see the SRC method shows good performance when the number of training signals is large enough. For all subjects, average accuracy grows when the number of training signals increases.

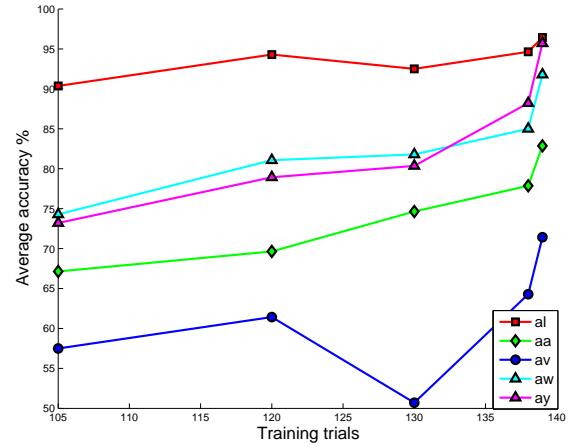


Fig. 1. Average accuracy of SRC with different number of training signals

IV. CONCLUSIONS

We apply the idea of sparse representation as a new classification method to the motor imagery based BCI. The sparse representation method needs a well designed dictionary composed of training data. We use the CSP filtering and the FFT to produce the columns of the dictionary. We have shown that a good classification result can be obtained by the proposed method. In addition, we have compared with the conventional approach such as linear discriminant analysis (LDA) method, which is well known for robust performance for the BCI system. Our result shows proposed method better than LDA.

V. ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (Do-Yak Research Program, NO. 2010-0017944).

REFERENCES

- [1] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, *Optimal spatial filtering of single trial EEG during imagined hand movement*, IEEE Trans. Rehabil. Eng., vol. 8, no. 4, pp. 441-446, Dec. 2000.
- [2] G. Pfurtscheller, Ch. Neuper, D. Flotzinger, M. Pregenzer, *EEG-based discrimination between imagination of right and left hand movement*, Electroencephalogr. Clin. Neurophysiol., vol. 103, pp. 642-651, 1997.
- [3] E. Cande's, J. Romberg, and T. Tao, *Stable Signal Recovery from Incomplete and Inaccurate Measurements*, Comm. Pure and Applied Math., vol. 59, no. 8, pp. 1207-1223, 2006.
- [4] S. Chen, D. Donoho, and M. Saunders, *Atomic Decomposition by Basis Pursuit*, SIAM Rev., vol. 43, no. 1, pp. 129-159, 2001.

A Realistic Distributed Compressive Sensing Framework for Multiple Wireless Sensor Networks

J. Oliver and Heung-No Lee*

Gwangju Institute of Science and Technology, South Korea. Email: {heungno,oliver}@gist.ac.kr *Corresponding Author

Abstract—In this paper, we propose a new compressive sensing framework for sensor networks. Unlike the conventional approaches, we consider the design of sensing matrix with the prior knowledge of the channel between the signals and the sensors. We determine that full or partial knowledge of the channel at sensors enables effective sensing matrix design and supports a good signal recovery. We discuss some of our key results and scope for our future research.

Index Terms—Compressive Sensing, Sensing matrix, Sensor networks.

I. INTRODUCTION

Compressive sensing (CS) is an emerging signal acquisition technique that recovers a sparse signal from few linear measurements [1]. Due to its popularity, CS is currently applied in many areas such as coding, signal processing and wireless sensor networks [2]. In this paper, we present a new CS framework for wireless sensor networks.

We consider a sensor network consisting of S sensors connected to a centralized fusion center. Each sensor measures a desired sparse signal and then compresses the sensed signal using a sensing matrix. The compressed measurements from different sensors are sent to the fusion center for joint recovery of the sparse signals. Unlike the conventional framework [3], we consider realistic scenarios in which there exists channel between the signal to be sensed and the sensor, such as in underwater acoustic systems and seismic sensor systems. We observe that the signals thus acquired have a lot of redundancy which can be handled efficiently by the proper design of sensing matrices using the prior knowledge of the channel.

In this paper, we consider a wireless sensor network having S number of sensors deployed at random locations. Let \mathbf{s} denote an K -sparse signal of length N ($K \ll N$). The signal received at the j -th sensor can be modeled by $\mathbf{x}_j = C_j \mathbf{s}$, where C_j is an $(N + L - 1) \times N$ matrix which models the delay dispersed channel between the intrinsic source to the j -th sensor. The sparse signal at the j -th sensor is then compressed by an $M \times (N + L - 1)$ random Gaussian matrix F_j [1] to obtain the linear measurement $\mathbf{y}_j = F_j \mathbf{x}_j$. The joint received vector obtained at the fusion center can then be modeled as

$$\mathbf{y} = FC\mathbf{s} + \mathbf{n} \quad (1)$$

where $\mathbf{y} = [\mathbf{y}_1^T, \dots, \mathbf{y}_S^T]^T$, F is a block diagonal matrix with F_j s as the diagonal entries and $C = [C_1^T \cdots C_S^T]^T$. The goal of the fusion center is to recover \mathbf{x}_j s and the intrinsic sparse signal \mathbf{s} , from \mathbf{y} .

One of the key challenges in our framework is the design of a good sensing matrix at each sensor. With the existence of the channel, the conventional Gaussian sensing matrix in (1) may not be enough to capture the maximum information because the sensing matrix design now has to depend on the characteristics of the channel. Therefore, we would like to utilize the channel information in the sensing matrix design. If the channel matrix C_j is exactly available at the j -th sensor, then one would like design the sensing matrix F_j based on C_j . At the fusion center we reconstruct the sparse signal by L_1 minimization with the recovery matrix $A = FC$. The incoherence of A , which should be low for good signal recovery, now depends on C_j s. Our

aim is to design good sensing matrices F_j s based on the exact or partial knowledge about C_j s such that the recovery matrix A behaves incoherently. In addition, it would be interesting to investigate the minimum measurements required for either exact or approximate recovery of the sparse signals under this realistic conditions.

II. DISCUSSIONS

We have carried out a preliminary investigation to determine the number measurements needed at the fusion center for exact signal recovery when the channel is known. Our preliminary study exposes a few surprising results obtained by incorporating additional channel information for sensing matrix design. Unlike the conventional theory which demands $O(k \log N)$ measurements for the unique L_1 solution, we show that, only sub-sparse measurements from each sensor is needed to obtain perfect L_1 signal recovery at the fusion center. This achievement is possible since we properly use the available channel information for signal acquisition. From our preliminary studies we found that as the number of sensors increases, the measurements needed for a given probability of recovery decrease.

III. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a compressive sensing framework with application to wireless sensor networks. In our framework, we have considered the design of sensing matrices to obtain a low coherent recovery matrix by making use of the prior knowledge of the channel. We would like to proceed in the following directions for our future research:

- Given the channel matrices C_j s exactly or partially, how to design good sensing matrices F_j s such that A is incoherent?
- What is the relationship between the channel parameters and the coherence of the recovery matrix?
- What is the condition for the unique L_0 solution? How is this condition related to the channel parameters?
- What is the equivalence relation for the existence of the unique L_1 solution? How does it depend on the channel parameters?
- What is the restricted isometry property (RIP) in this practical situation?
- How much information can we obtain from a sensor network given coverage and sensor density?
- How does the correlation among the sensors affect the information obtainable from the network?

Acknowledgement: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (Do-Yak Research Program, No. 2010-0017944)

REFERENCES

- [1] Richard Baraniuk, “Compressive sensing,” *IEEE Signal Processing Magazine*, vol. 24, no.4, pp. 118-121, 2007.
- [2] Emmanuel Candes and Michael Wakin, “An introduction to compressive sampling,” *IEEE Signal Processing Magazine*, vol. 25, no.2, pp. 21-30, 2008.
- [3] M. F. Duarte et al, “Distributed compressed sensing of jointly sparse signals,” in Proc. of 39-th ACSSC, CA, November 2005.

Sparse Phase Retrieval

Shiro Ikeda

The Institute of Statistical Mathematics,
Tachikawa, Tokyo 190-8562, Japan
shiro@ism.ac.jp

Hidetoshi Kono

Japan Atomic Energy Agency,
Kizugawa, Kyoto 619-0215, Japan
kono.hidetoshi@jaea.go.jp

Coherent X-ray Diffraction Imaging (CXDI) is a technique for the 2-dimensional (2D) and 3D reconstruction of nanoscale structures. The detector receives the photons scattered by the object, and ideally, the diffraction pattern gives the power spectrum of the electron density. Since we are only provided the power spectrum and the phase is lost, we need to retrieve the phase in order to reconstruct the structure from the diffraction image.

Let $f_{xy} \geq 0$ be the electron density of a molecule projected onto a 2D plane. We consider the discretized coordinate, $x, y = 1, \dots, M$ and ideal diffraction pattern is $|F(u, v)|^2$ where, $F(u, v)$ is the Fourier transform of $f(x, y)$ as follows,

$$F_{uv} = \frac{1}{M} \sum_{x,y} f_{xy} \exp\left(\frac{2\pi i(ux+vy)}{M}\right). \quad (1)$$

A widely used phase retrieval method is the hybrid input-output (HIO) method [1], [2]. The HIO method set a support region and assume $f_{xy} = 0$ outside the support and estimate the phase with an iteratively process. It effectively solves the problem for high signal-to-noise ratio measurements.

Recently, a new type of coherent beam, x-ray free electron lasers (XFELs), became available. This new technology can potentially provide a novel mean to determine the three-dimensional (3D) structure of biomolecules from the diffraction data of single molecules instead of conventional crystallography [3], [4]. One of the crucial processes of single molecule imaging is the phase retrieval from a very weak diffraction due to the "single" molecule. Figure 1a shows a simulated electron density of a biomolecule. If the power spectrum is obtained as in Fig. 1b, the HIO method will successfully reconstruct the 2D density. However, the simulated result of the diffraction pattern would not be any better than Fig. 2a. The diffraction image is so noisy that the HIO method does not even converge.

Here, we propose a new approach, the sparse phase retrieval (SPR) method [5], for retrieving phases of diffraction data which will be obtained by XFELs. Instead of assuming the support region, we use the Bayesian statistics as in [6], employing a sparse prior. Let N_{uv}

be the number of the photons detected at (u, v) of the detector. It is natural to assume each N_{uv} follows a Poisson distribution independently,

$$p(\mathbf{N}|\mathbf{F}) = p(\mathbf{N}|\mathbf{f}) = \prod_{uv} \frac{|F_{uv}|^{2N_{uv}} \exp(-|F_{uv}|^2)}{N_{uv}!}, \quad (2)$$

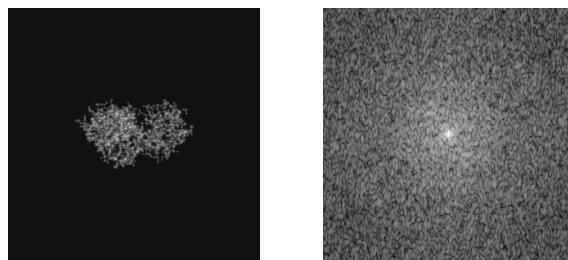
where, $\mathbf{N} = \{N_{uv}\}$, $\mathbf{F} = \{F_{uv}\}$, $\mathbf{f} = \{f_{xy}\}$, and the fact \mathbf{F} is a deterministic function of \mathbf{f} was used. Assuming a sparse prior of \mathbf{f} as $\pi(\mathbf{f}) \propto \prod_{xy} \exp(-\rho_{xy} f_{xy})$, where $\rho_{xy} \in \Re_+$, we compute the maximum a posteriori (MAP) estimator, and the SPR method computes the MAP estimate for the density reconstruction. The estimate $\hat{\mathbf{f}}$ is the maximizer of the following function,

$$\ell(\mathbf{f}|\mathbf{N}) = \sum_{uv} (N_{uv} \ln |F_{uv}|^2 - |F_{uv}|^2) - \sum_{xy} \rho_{xy} f_{xy}.$$

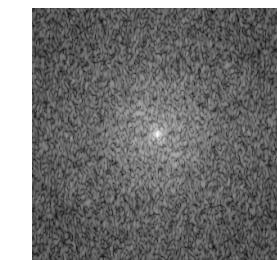
The sparse prior automatically sets many entries of \mathbf{f} equal to 0 without specifying the support region and the sparsity is adjusted by modifying ρ_{xy} . Figure 2b shows the density reconstructed by the SPR method. Compared to the HIO method, the SPR method gives better results under the noise. This is a new promising direction for phase retrieval in practice.

REFERENCES

- [1] J. Fienup, "Reconstruction of an object from the modulus of its Fourier transform," *Optics Letters*, vol. 3, no. 1, pp. 27–29, 1978.
- [2] J. Miao, D. Sayre, and H. Chapman, "Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects," *J. Opt. Soc. Am. A*, vol. 15, no. 6, pp. 1662–1669, 1998.
- [3] D. Sayre, "Prospects for long-wavelength X-ray microscopy and diffraction," in *Imaging Processes and Coherence in Physics*, ser. Springer Lecture Notes in Physics, M. Schlenker, *et al.*, Eds. Berlin: Springer, 1980, vol. 112, pp. 229–235.
- [4] K. Gaffney and H. Chapman, "Imaging atomic structure and dynamics with ultrafast x-ray scattering," *Science*, vol. 316, pp. 1444–1448, 2007.
- [5] S. Ikeda and H. Kono, "Phase retrieval from single biomolecule diffraction pattern," 2011, arXiv:1101.1442.
- [6] R. Irwan and R. G. Lane, "Phase retrieval with prior information," *J. Opt. Soc. Am. A*, vol. 15, no. 9, pp. 2302–2311, 1998.

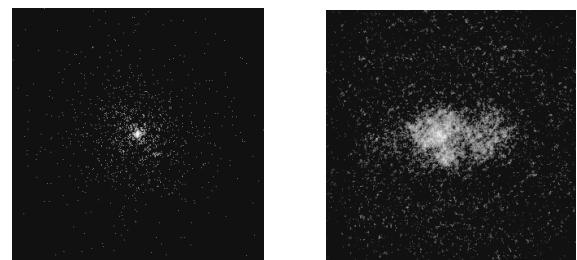


(a) Electron density of lysozyme.



(b) Ideal diffraction pattern.

Fig. 1. (a) 2D electron density of a protein, lysozyme. (b) An ideal 2D diffraction image of lysozyme without noise.



(a) Simulated diffraction pattern.

(b) Reconstructed density.

Fig. 2. (a) A simulated diffraction image of lysozyme under a realistic situation. (b) A reconstructed density images with SPR method.

Probabilistic models which enforce sparsity

Ali Mohammad-Djafari

Laboratoire des signaux et systèmes (L2S)

UMR 8506 CNRS-SUPELEC-UNIV PARIS SUD

SUPELEC, Plateau de Moulon, 91192 Gif-sur-Yvette, France Email: <http://djafari.free.fr>

Abstract—In this paper, we propose different prior modelling for signals and images which can be used in a Bayesian inference approach in many inverse problems in signal and image processing where we want to infer on sparse signals or images. The sparsity may be directly on the original space or in a transformed space. Here we consider it directly on the original space (impulsive signals). These models are either heavy tailed (Generalized Gaussian, Weibull, Student-t or Cauchy) or mixture models (Mixture of Gaussians, Bernouilli-Gaussian, Bernouilli-Gamma, Mixture of translated Gaussians,...). Depending on the prior model selected, the Bayesian computations (optimization for the Joint Maximum A Posteriori (MAP) estimate or MCMC or Variational Bayes Approximations (VBA) for Posterior Means (PM) or complete density estimation) may become more complex. We propose these models, discuss on different possible Bayesian estimators, drive the corresponding appropriate algorithms, and discuss on their corresponding relative complexities and performances. We then show some simulation results of the application of these methods to a deconvolution problem and to a sources separation of sparse signals.

I. INTRODUCTION

In many generic inverse problems in signal and image processing we want to infer on an unknown signal $f(t)$ or an unknown image $f(r)$ through an observed signal $g(s)$ or an observed image $g(s)$ related between them through an operator \mathcal{H} such as convolution $g = h * f$ or any other linear or non linear transformation $g = \mathcal{H}f$. When this relation is linear and we have discretized the problem, we arrive to the relation: $\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}$ where \mathbf{f} represents the unknowns, \mathbf{g} the observed data, $\boldsymbol{\epsilon}$ the errors of modelling and measurement and \mathbf{H} the matrix of the system response.

The Bayesian inference approach then is based on the posterior law:

$$p(\mathbf{f}|\mathbf{g}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{p(\mathbf{g}|\mathbf{f}, \boldsymbol{\theta}_1)p(\mathbf{f}|\boldsymbol{\theta}_2)}{p(\mathbf{g}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)} \quad (1)$$

where $p(\mathbf{g}|\mathbf{f}, \boldsymbol{\theta}_1)$ is the likelihood, $p(\mathbf{f}|\boldsymbol{\theta}_2)$ the prior model, $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ are their corresponding parameters (often called the hyperparameters of the problem) and $p(\mathbf{g}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is called the evidence of the model.

One of the main steps in the Bayesian approach is the prior modelling which has the role of translating our prior knowledge on the unknown signal or image in a probability law. Sparsity is one of the prior knowledges we may translate. The main objective of this paper is to see what are the different possibilities.

The second main step in the Bayesian approach is to do the computations: either optimization when the Maximum A posteriori (MAP) is selected or numerical approximations such as MCMC or the Bayesian Variational Approximations (BVA) when the Expected A posteriori (EAP) estimation is selected or when we want to explore the whole posterior probability law.

In this paper, we propose different prior modelling for signals and images which can be used in a Bayesian inference approach in many inverse problems in signal and image processing where we want to infer on sparse signals or images. The prior models discussed are the following: Generalized Gaussian (GG), Weibull (W) and Rayleigh (R), Student-t (St) and Cauchy (C), Mixture

of two Gaussians (MoG2), Bernouilli-Gaussian (BG), Bernouilli-Gamma (BGamma), Mixture of three Gaussians (MoG3) and Mixture of one Gaussian and two Gammas (MoGGammas). Some of these models are well-known, some others less. In general, we can classify them in two categories: i) Simple Non Gaussian models with heavy tails and ii) Mixture models with hidden variables which result to hierarchical models.

Depending on the prior model selected, the Bayesian computations (optimization for the MAP estimate or MCMC or VBA for EAP or complete density estimation) may become more complex. The second main objective of this paper is to discuss on the relative complexities and performances of the algorithms obtained with the proposed prior law.

The rest of the paper is organized as follows:

In section II, we present in details the proposed prior models and discuss their properties. For example, we will see that the Student-t model can be interpreted as an infinite mixture with a variance hidden variable or that the BG model can be considered as the degenerate case of a MoG2 where one of the variances go to zero. Also, we will examine the less known models of MoG3 and MoGGammas where the heavy tails are obtained by combining a centered Gaussian and two large variance non-centred Gaussians or Gammas.

In Section III, we examine the expression of the posterior laws that we obtain using these priors and discuss then on complexity of the Bayesian computation of the algorithms. In particular for the mixture models, we give details of the joint estimation of the signal and the hidden variable as well as the hyperparameters (parameters of the mixtures and the noise) for unsupervised cases. In particular, we examine the relative performances of MCMC and Variational Bayesian Approximation (VBA) methods.

In Section IV, we compare the performances of these algorithms in signal deconvolution, image restoration and Blind Sources Separation.

As a typical simulation result, we generated first a signal $f(t)$ using the BG model with parameters $\lambda = .1, v = 10$ over $N = 200$ samples. Then, we generated a PSF $h(t)$ with a Gaussian shape over $L = 11$ samples which is used to generate the observed data $g(t) = h(t) * f(t) + \epsilon(t)$ with a Gaussian noise $\epsilon(t)$ with variance $v_\epsilon = .5$. We then used these data with different algorithms to estimate \mathbf{f} , noted $\hat{\mathbf{f}}$ and when needed the hidden variables \mathbf{z} , noted $\hat{\mathbf{z}}$ and the hyperparameters $\boldsymbol{\theta}$ noted $\hat{\boldsymbol{\theta}}$. We then compare $\hat{\mathbf{f}}$ with \mathbf{f} , $\hat{\mathbf{z}}$ with \mathbf{z} and $\hat{\boldsymbol{\theta}}$ with $\boldsymbol{\theta}$.

Finally, in section V we show some applications. More detailed results will be on the final paper. A draft version is available on request.

REFERENCES

- [1] C. A. Bouman and K. D. Sauer, “A generalized Gaussian image model for edge-preserving MAP estimation,” vol. 2, pp. 296–310, July 1993.
- [2] J. J. Kormylo and J. M. Mendel, “Maximum-likelihood detection and estimation of Bernoulli-Gaussian processes,” vol. 28, pp. 482–488, 1982.
- [3] M. Lavielle, “Bayesian deconvolution of Bernoulli-Gaussian processes,” *Signal Processing*, vol. 33, pp. 67–79, 1993.
- [4] J. Idier, ed., *Approche bayésienne pour les problèmes inverses*. Paris: Traité IC2, Série traitement du signal et de l'image, Hermès, 2001.

Greedy Algorithms for Sparse Total Least Squares

Bogdan Dumitrescu

Department of Signal Processing, Tampere University of Technology
PO BOX 553, 33101 Tampere, Finland, e-mail: bogdan.dumitrescu@tut.fi

I. SPARSE TOTAL LEAST SQUARES

Consider the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$, with $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$. Its total least squares (TLS) solution is given by

$$\begin{aligned} \min & \quad \|[\mathbf{E} \ \mathbf{f}]\|_F \\ \text{s.t.} & \quad (\mathbf{A} + \mathbf{E})\mathbf{x} = \mathbf{b} + \mathbf{f} \end{aligned} \quad (1)$$

The solution results by taking $[\mathbf{E} \ \mathbf{f}]$ as the matrix with smallest Frobenius norm that added to $[\mathbf{A} \ \mathbf{b}]$ makes the result singular. As such, it is related to the smallest singular value of $[\mathbf{A} \ \mathbf{b}]$: at optimality we have $\|[\mathbf{E} \ \mathbf{f}]\|_F = \sigma_{\min}([\mathbf{A} \ \mathbf{b}])$.

Assume now that we search sparse TLS solutions $\mathbf{x} \in \mathbb{R}^n$ to $\mathbf{A}\mathbf{x} = \mathbf{b}$, having at most s nonzero elements ($\|\mathbf{x}\|_0 = s$). In principle, the sparse TLS problem can be solved in two steps.

1. Selection of nonzero elements. If \mathcal{I} is a set of s indices (the support), then denote $\mathbf{x}_{\mathcal{I}} \in \mathbb{R}^s$ the vector of nonzero elements of \mathbf{x} (and assume the other elements are zero) and $\mathbf{A}_{\mathcal{I}} \in \mathbb{R}^{m \times s}$ the matrix formed by the columns of \mathbf{A} with indices in \mathcal{I} . The support of the sparse TLS solution with s nonzeros is given by

$$\begin{aligned} \min_{\mathcal{I}} & \quad \sigma_{\min}([\mathbf{A}_{\mathcal{I}} \ \mathbf{b}]) \\ \text{s.t.} & \quad |\mathcal{I}| = s \end{aligned} \quad (2)$$

2. Once the support is determined, the solution $\mathbf{x}_{\mathcal{I}}$ results from solving the standard TLS problem (1) with $\mathbf{A}_{\mathcal{I}}$ instead of \mathbf{A} .

Problem (2) is hard; it can be solved exactly only by enumeration.

II. PREVIOUS WORK

To the best of our knowledge, the sparse TLS problem is discussed only in [1], where applications are also presented. Sparsity is promoted by replacing the criterion of (1) with $\|[\mathbf{E} \ \mathbf{f}]\|_F^2 + \lambda\|\mathbf{x}\|_1$. The problem is not convex and an algorithm based on alternating coordinate descent is proposed in [1].

The sparse TLS problem can be related to the computation of the lower restricted isometry property constant associated with a matrix \mathbf{D} , which is the smallest δ_p^{\min} such that $(1 - \delta_p^{\min})\|\mathbf{y}\|_2^2 \leq \|\mathbf{D}\mathbf{y}\|_2^2$, for all vectors \mathbf{y} with $\|\mathbf{y}\|_0 = p$. This amounts to finding the p columns of \mathbf{D} that form a matrix whose smallest singular value is minimum. In the sparse TLS case, we have $\mathbf{D} = [\mathbf{A} \ \mathbf{b}]$ and $p = s + 1$, but \mathbf{b} must always be one of the selected columns. The greedy algorithm from [2] is hence applicable, starting with \mathbf{b} and then searching for s columns of \mathbf{A} in the attempt of solving (2).

III. PROPOSED GREEDY ALGORITHM

The structure of the proposed greedy algorithm is

- Input: \mathbf{A} , \mathbf{b} , s
- 1. for $j = 1 : s$
 - 1.1. $\mathcal{I} = \{j\}$
 - 1.2. for $i = 2 : s$
 - 1.2.1. find "best" column \mathbf{a}_k , $k \notin \mathcal{I}$
 - 1.2.2. increase support: $\mathcal{I} \leftarrow \mathcal{I} \cup \{k\}$
 - 1.3. compute $v_j = \sigma_{\min}([\mathbf{A}_{\mathcal{I}} \ \mathbf{b}])$
- Output: support \mathcal{I} that gives the smallest v_j

Note that all columns of \mathbf{A} are tried for the first position in \mathcal{I} , but the other positions are filled using the standard greedy strategy. For the "best" column selection in 1.2.1 we have used three heuristics. Two are known (but have been used in a different context in [2]): i) minimum singular value (MSV): the column $k \notin \mathcal{I}$ for which $\sigma_{\min}([\mathbf{A}_{\mathcal{I} \cup \{k\}} \ \mathbf{b}])$ is minimum; ii) the choice from [2]. MSV is very slow and is used mainly for reference.

The third heuristic, named SAS, is our contribution and consists of choosing the column that makes the smallest angle with the subspace generated by the already selected columns and \mathbf{b} . Denoting $\mathcal{S}_{\mathcal{I}} = \text{Im}([\mathbf{A}_{\mathcal{I}} \ \mathbf{b}])$ the "best" column is found by

$$k = \underset{\ell \notin \mathcal{I}}{\text{argmax}} \frac{\text{proj}_{\mathcal{S}_{\mathcal{I}}} \mathbf{a}_{\ell}}{\|\mathbf{a}_{\ell}\|}. \quad (3)$$

The projection of a column on $\mathcal{S}_{\mathcal{I}}$ can be computed easily if a partial QR factorization (\mathcal{J} is the complement of \mathcal{I})

$$\mathbf{Q}^T [\mathbf{b} \ \mathbf{A}_{\mathcal{I}} \ \mathbf{A}_{\mathcal{J}}] = \begin{bmatrix} \mathbf{R} & \mathbf{P} \\ 0 & \mathbf{S} \end{bmatrix} \quad (4)$$

is available at iteration i of step 1.2, where $\mathbf{R} \in \mathbb{R}^{i \times i}$, $i = |\mathcal{I}| + 1$, is upper triangular and \mathbf{Q} is orthogonal. After the orthogonal transformation, the matrix \mathbf{P} contains the projections of columns \mathbf{a}_{ℓ} , $\ell \notin \mathcal{I}$, on $\mathcal{S}_{\mathcal{I}}$ (while \mathbf{S} contains the components orthogonal on $\mathcal{S}_{\mathcal{I}}$). Hence, the "best" column is that for which $\|\mathbf{p}_{\ell}\|/\|\mathbf{a}_{\ell}\|$ is maximum.

The SAS greedy algorithm for TLS can be organized as a QR factorization with pivoting. The complexity of step 1.2.1 is $O((m-i)(n-i))$, similar to that of the algorithm from [2].

IV. RESULTS

As a sample of our simulations, the table below shows the final value $\sigma_{\min}([\mathbf{A}_{\mathcal{I}} \ \mathbf{b}])$ averaged over 100 systems generated with $m = 100$, $n = 300$. For each sparsity s , a solution \mathbf{x} with $\|\mathbf{x}\|_0 = s$ is generated with elements from $\mathcal{N}(0, 1)$; \mathbf{A} has elements from $\mathcal{N}(0, 1/n)$ and $\mathbf{b} = \mathbf{A}\mathbf{x}$. Ideally, a sparse TLS algorithm should solve (2), getting $\sigma_{\min}([\mathbf{A}_{\mathcal{I}} \ \mathbf{b}]) = 0$; a nonzero value means that only an approximation has been found. The results show that for relatively low values of s , SAS is better or as good as other methods. Simulations with noise added to \mathbf{A} and \mathbf{b} have a similar relative behavior. In a simple Matlab implementation, the execution time is 3-5 times larger than that of [2].

s	5	10	15	20	25	30
MSV	0	0	0.0107	0.0353	0.0836	0.1039
[2]	0	0	0.0100	0.0339	0.1072	0.1225
SAS	0	0	0	0.0078	0.0495	0.1205
[1]	0.0001	0.0127	0.0423	0.0686	0.0887	0.1086

REFERENCES

- [1] H. Zhu, G. Leus, and G.B. Giannakis, "Sparsity-Cognizant Total Least Squares for Perturbed Compressive Sampling," *IEEE Trans. Signal Proc.*, 2011, to appear.
- [2] C. Dossal, G. Peyré, and J. Fadili, "A Numerical Exploration of Compressed Sampling Recovery," *Lin. Alg. Appl.*, vol. 432, pp. 1663–1679, 2010.

Super-resolution based on Sparsity Priori

Hui Wang^{1,2}, Shensheng Han¹, Mikhail I. Kolobov²

1. Key Laboratory for Quantum Optics and Center for Cold Atom Physics, Shanghai Institute of Optics and Fine Mechanics,
Chinese Academy of Sciences, Shanghai 201800, China
2. Laboratoire PhLAM, Université Lille-1, F-59655 Villeneuve d'Ascq cedex, France

Abstract—Super-resolution is always referred to be able to recover the object's Fourier transform spectrum exceeding Rayleigh resolution limit [1]. In some practical imaging problems, super-resolution is possible by taking advantage of priori knowledge of the object. Sparsity priori has long been considered in the efforts of super-resolution. It allows for exact image recovery from a number of samples much smaller than that required by the Nyquist/Shannon theorem, and is therefore expected to realize super-resolution recovery [2], [3], [4], [5]. CS, as a mathematical algorithm, is developed following mathematical interests and the recovery conditions are based on mathematical concepts. While Super-resolution, as a concept of imaging technique, calls for physical conditions, which imposes more priori information besides sparsity. In this paper, we impose sparsity together with other priori based on physical scheme and focus on the super-resolution behavior. This helps to relate the mathematical theorems to the physical quantities and makes good sense to the practical applications.

I. INTRODUCTION

In this paper, we consider simple sparse discrete object composed of several spikes which can take positive and negative values. This corresponds to the objects that carry both amplitude and phase information. Our imaging system is a simple 4-f scheme with a diaphragm of varying size on the Fourier spectrum plane which determines the optical resolution of the scheme. The measurements are also taken on the Fourier plane but are restricted to the area inside this diaphragm, which imposes another priori about sampling. From the point of imaging technique, there're two problems we care most. One is that given an imaging scheme, what kind of object can we recover. The other is given a sort of object, how to design a scheme that is able to recover it.

From the theoretical point of view, without loss of generality, the measurements could be set equispaced on the Fourier plane with separation equal to Nyquist sampling interval. Consequently, for a certain imaging scheme, there will be a fixed corresponding sensing matrix. Through analyzing the sensing matrix, it's possible to know the condition that guarantees the successful recovery and furthermore to classify the recoverable and unrecoverable objects. There're actually several theorems [2], [6], [7], [8] presenting conditions for the success recovery. These conditions can be classified into two groups [6]: one is for the uniqueness of the solution to L_0 minimization; the other is for that the unique solution to L_1 minimization (Basis Pursuit, BP) is equivalent to the solution to L_0 minimization. Comparisons are taken between these two groups to show that based on the same band-limited measurements, the different super-resolution behavior between the ideal optimized recovery (L_0 minimization) and the compromised but costless result (L_1 minimization).

From the practical point of view, to design or improve an imaging scheme, the matrix analysis above should be discussed together with other physical concepts, such as diffraction limit, super-resolution factor and Signal-to-Noise Ratio (SNR) et al.. For example, one important parameter to analyze the sensing matrix is its mutual coherence, which, in our scheme of low-pass system, is always related to the physical mutual coherence between two adjacent pixels, and further related to the diaphragm size, the diffraction limit and

super-resolution factor. This kind of connections between physical and mathematical concepts bring advantages to imaging applications, because it's generally more convenient to measure the physical quantities than to analyze a matrix. And by combining those recovery conditions with physical concepts, it helps to figure out the ultimate limit of the super-resolution for certain object.

We also analyze the influence of quantum fluctuations on super-resolution behavior. It has been shown that when taking another priori knowledge of the object (i.e. the object is finite), the ultimate limit of super-resolution is determined by the SNR in the imaging scheme [9], [10], [11]. However, if we take the priori of sparsity, the ultimate limit role seems to be replaced by other forms of Uncertainty Principle [2], [12]. In our simulations, we control the SNR by changing the light intensity on the object. And it's shown from both theory and simulation that the quantum fluctuations influence less on super-resolution when using the sparsity priori. In this paper we shall give the results of our simulations.

REFERENCES

- [1] David L. Donoho, "Superresolution via Sparsity Constraints" SIAM J. Math. Anal. **23**, 015 (1992).
- [2] David L. Donoho and Philip B. Stark, "Uncertainty Principles and Signal Recovery" SIAM J. Appl. Math. **49**, 906 (1989).
- [3] Snir Gazit, Alexander Szameit, Yonina C. Eldar, Mordechai Segev, "Super-resolution and reconstruction of sparse sub-wavelength images" Opt. Express. **17** 23920 (2009).
- [4] Yoav Shechtman, Snir Gazit, Alexander Szameit, Yonina C. Eldar, and Mordechai Segev, "Super-resolution and reconstruction of sparse images carried by incoherent light" Opt. Lett. **35**, 1148 (2010).
- [5] Wenlin Gong and Shensheng Han, "Super-resolution single-beam imaging via compressive sampling" arXiv:physics.optics/1005.1735v1 (2010).
- [6] David L. Donoho and Michael Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization" Proc. Nat. Acad. Sci. USA, vol. **100**, 2197 (2003).
- [7] Emmanuel J. Candès and Terence Tao, "Decoding by Linear Programming" IEEE Trans. Inform. Theory. **51**, 4203 (2005)
- [8] Emmanuel J. Candès and Terence Tao, "Near-Optimal Signal Recovery From Random Projections" IEEE Trans. Inform. Theory. **52**, 5406 (2006)
- [9] Mikhail I. Kolobov and Claude Fabre, "Quantum Limits on Optical Resolution" Phys. Rev. Lett. **85**, 3789 (2000).
- [10] Vladislav N. Beskrovnyy and Mikhail I. Kolobov, "Quantum limits of super-resolution in reconstruction of optical objects" Phys. Rev. A. **71**, 043802 (2005).
- [11] Mikhail I. Kolobov, "Quantum limits of superresolution for imaging discrete subwavelength structures" Opt. Express. **16**, 59 (2008).
- [12] Emmanuel J. Candès, Justin Romberg, Member, IEEE, and Terence Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information" IEEE Trans. Inform. Theory. **52**, 489 (2006)

Fast Compressive Sensing Recovery with Transform-based Sampling

Hung-Wei Chen

Inst. Info. Sci., Academia Sinica and
Graduate Inst. Comm. Eng., NTU, Taiwan
Email: hungwei@iis.sinica.edu.tw

Chun-Shien Lu

Institute of Information Science,
Academia Sinica, Taipei, Taiwan
Email: lcs@iis.sinica.edu.tw

Soo-Chang Pei

Graduate Inst. Comm. Eng.,
Nat'l Taiwan Uni., Taipei, Taiwan
Email: pei@cc.ee.ntu.edu.tw

Abstract—We present a fast compression sensing (CS) reconstruction algorithm with computation complexity $O(M^2)$, where M denotes the length of a measurement vector $Y = \phi X$ that is sampled from the signal X of length N via the sampling matrix ϕ with dimensionality $M \times N$. Our method has the following characteristics: (1) it is fast due to a closed-form solution is derived; (2) it is accurate because significant components of X can be reconstructed with higher priority via a sophisticated design of ϕ ; (3) thanks to (2), our method can better reconstruct a less sparse signal than the existing methods under the same measurement rate $\frac{M}{N}$.

I. INTRODUCTION

In the context of compressive sensing (CS) [1], the constraint of sparsity enables the possibility of sparse signal recovery from measurements (far) fewer than the original signal length. Moreover, the measurements generated from random projection of the original signal via a sampling matrix are equally weighted; *i.e.*, no one is more significant than the others. Thus, compressive sensing is inherently weaken in handling less sparse signals such as highly textured images. The problem here is that can we yield weighted measurements so that non-sparse or less sparse signals can be properly reconstructed than the existing CS recovery solutions?

In this paper, we present a sophisticated design of the sampling matrix ϕ that can directly capture “important” measurements. With these information, the quality original signal can be sparsely reconstructed based on the important (corresponding to low-frequency) components in some transformed domain. Thus, the qualities of reconstructed signals mimic those of JPEG compressed images.

II. PROPOSED METHOD

We start from the random projection, $Y = \phi X$, and observe that if important information of X can be sampled and stored in Y , then it is possible to reconstruct X with fewer important measurements.

For this, we introduce a linear operator T and impose it to random projection to obtain $T \circ Y = T \circ (\phi X)$, where \circ stands for a linear operation. This equation is further derived¹ based on the principle of linear operations [2] as:

$$T \circ Y = T \circ (\phi X) = (T \circ \phi)(T \circ X). \quad (1)$$

Eq. (1) indicates that if T is a transform operator, then $T \circ X$ is a transformed vector in some transform domain. In particular, the positions at lower frequencies in $T \circ X$ indicate important transformed coefficients and $T \circ Y$ indicates important measurements since they are linear combinations of significant transformed coefficients. For simplicity, the operator \circ will be omitted below.

In order to sample “important” transformed coefficients from TX and speed up recovery, we design a new sampling matrix, $(T\phi)^z$, by setting the last $N - M$ columns of $T\phi$ to be zeros. This implies that the non-zero columns of $(T\phi)^z$ form a full-rank matrix with rank M . Once $(T\phi)^z$ is built in the transform domain, it is inversely transformed back to the time/space domain and a sophisticated designed sampling matrix $\Phi = T^{-1}((T\phi)^z)$ is obtained.

¹The proof is omitted here due to space limit.

Now, Φ is stored in the sensors for the purpose of compressive sensing. We have the following derivations:

$$Y = \Phi X \Rightarrow TY = (T\Phi)(TX) = (T\phi)^z(TX). \quad (2)$$

Recall that the last $N - M$ columns of $(T\phi)^z$ are zeros. This means that we only sample the lower-frequency components in TX by discarding the remaining higher-frequency components. In order to speed up sparse signal recovery, let Φ^s denote the submatrix of dimensionality $M \times M$ by discarding the zero columns of $(T\phi)^z$, and let $(TX)^s$ denote the $M \times 1$ vector by discarding the last $N - M$ transformed coefficients. Therefore, we can derive:

$$TY = \Phi^s(TX)^s \Rightarrow (\Phi^s)^{-1}TY = (\Phi^s)^{-1}\Phi^s(TX)^s = (TX)^s. \quad (3)$$

It is evident that the signal X can be approximately and fast recovered if (i) Y is available via random projection in Eq. (2); (ii) Y is processed via Eq. (3); and (iii) $(\Phi^s)^{-1}TY$ is padded with $N - M$ zero values (to obtain TX) and inversely transformed via T^{-1} .

III. ANALYSIS AND RESULTS

The principle of our method is to preserve the top K -lowest frequency components of TX . Here, T is chosen to be a DCT operator. Thus, we have $M = K$ and the computation complexity of recovery is in the order of $O(M^2)$; *i.e.*, only one inverse matrix operation and two DCT operations are required.

In this paper, a 1D DCT structure is exploited to design Φ . The original signal X can be approximately reconstructed from as many measurements as the number of coefficients sampled via Eq. (2). We provide recovery comparison of some CS algorithms [1] under different measurement rates (MRs) in Table I². The exploitation of the simple structure inherent in the Haar wavelet is also studied in our framework.

TABLE I
RECOVERY COMPARISON OF CS ALGORITHMS FOR BARBARA IMAGE.

Methods	Metrics	MR (6.25%)	MR (12.5%)	MR (25.0%)
Our Method (DCT-based)	PSNR(dB) SSIM	22.20 0.59	23.78 0.67	26.27 0.81
Lasso	PSNR(dB) SSIM	16.82 0.33	20.31 0.51	23.91 0.71
OMP (Sparsify toolbox)	PSNR(dB) SSIM	17.62 0.34	19.86 0.48	22.53 0.65
Basis Pursuit	PSNR(dB) SSIM	16.82 0.33	20.31 0.51	23.91 0.71
StOMP (SparseLab toolbox)	PSNR(dB) SSIM	10.94 0.23	12.51 0.37	21.92 0.61

REFERENCES

- [1] <http://dsp.rice.edu/cs>
- [2] N. Merhav and V. Bhaskaran, “A transform domain approach to spatial domain image,” *HPL-94-116*, Technion City, Haifa 32000, Israel, 1994.

²Structural similarity (SSIM) indexing is also adopted for image quality evaluation, where $0 \leq \text{SSIM} \leq 1$. The bigger, the better.

Feature Selection in Carotid Artery Segmentation Process based on Learning Machines

Rosa María Menchón Lara, Consuelo Bastida Jumilla, Juan Morales Sánchez, and José Luis Sancho Gómez.
Universidad Politécnica de Cartagena, Departamento de Tecnologías de la Información y las Comunicaciones,
Plaza del Hospital, 1, 30202, Cartagena (Murcia), SPAIN. Email: josel.sancho@upct.es

ABSTRACT

Recent studies show that the thickening of the carotid artery wall is indicative of the corresponding hardening and thickening of coronary arteries. Physicians can determine the tendency of a patient to the atherosclerosis through a B-mode ultrasound scan of the common carotid artery (CCA). This is a non-invasive technique that allows to take a measure of the Intima-Media Thickness (IMT)[1]. The IMT is the distance between the lumen-intima interface and the media-adventitia interface of CCA's far wall. Currently, doctors measure the IMT by setting manually only a few points, which may distort the results. Image segmentation can detect the *IMT contour* throughout the artery length, which leads to better results and allows us to extract statistics such as the maximum, the minimum or the average IMT with more precision.

In this work, an efficient image segmentation technique is proposed. Segmentation is treated as a pattern recognition problem and is solved using a neural network ensemble (also called *committee machine*) to improve the accuracy achieved by a single net. In particular, the results from three *experts* are combined by a '*meta*' neural network. With the proper training, the proposed system is able to recognize the pixels belonging to the IMT contour. Once the networks are trained, the proposed method allows getting IMT measurements in an automatic way.

The networks in our system are *Multi-Layer Perceptrons* (MLP). These nets have been trained by means of the *Optimally-Pruned Extreme Learning Machine* (OP-ELM)[2], which is easy to use and allows faster learning than others such as *Backpropagation* (BP) algorithm. Furthermore, it is able to select the optimal network size. For this purpose, OP-ELM uses both *MultiResponse Sparse Regression* (MRSR)[3] algorithm and an efficient *Leave-One-Out* (LOO) criterion.

To perform the neural network training, we need a training set composed of ultrasound images and the associated desired outputs (*supervised learning*) called *target images*. The target images are binary images in which white pixels (with value '1') show the IMT boundaries. A windowing process is applied over the original image to obtain the training set. In our case, square windows are used varying the size of the window in the different experts of the committee machine. Thus, a square neighborhood is taken as input pattern for each pixel under study. In order to reduce the computational cost, a feature selection procedure has been applied. The *Least Angle Regression* (LARS)[4] algorithm has been used to provide

a ranking of input features, which are ordered according to their relevance to the classification task. Then, the analysis is performed in a stepwise manner by adding at each iteration a new feature and training the net with the OP-ELM algorithm. Finally, the network model (number of input features and hidden neurons) with lowest error is selected. This strategy is followed to design the artificial neural networks of our system and the results show that it is possible to reduce the dimensionality of the data (see Fig. 1).

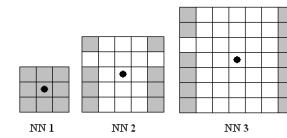


Fig. 1. Selected pixels to construct the input patterns to the three experts.

Figure 2 shows the preliminary results obtained by the system (combination of three neural networks). It can be seen that the obtained segmentation is satisfactory.

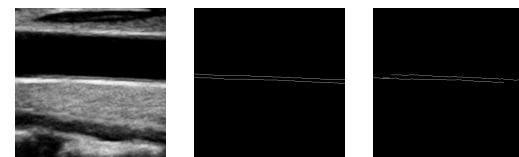


Fig. 2. Input image (left), target image (center), and output image (right).

ACKNOWLEDGEMENTS

This work is partially supported by the Spanish Ministerio de Ciencia e Innovación, under grant TEC2009-12675, by the UPCT (Inicialización a la Actividad Investigadora, 2010), and by the Séneca Foundation (09505/FPI/08).

REFERENCES

- [1] G. L. Burke, G. W. Evans, W. A. Riley, A. R. Sharrett, G. Howard, R. W. Barnes, W. Rosamond, R. S. Crow, P. M. Rautaharju, and G. Heiss, "Arterial wall thickness is associated with prevalent cardiovascular disease in middle-aged adults," *Stroke*, vol. 26, no. 3, pp. 386-391, Mar 1995.
- [2] Y. Miche, A. Sorjamaa, P. Bas, O. Simula, C. Jutten, and A. Lendasse, "Op-elm: Optimally pruned extreme learning machine," *IEEE Transactions on neural networks*, vol. 21, no. 1, pp. 158-162, 2010.
- [3] T. Similä and J. Tikka, "Multiresponse sparse regression with application to multidimensional scaling," in *Proceedings of the 15th International Conference on Artificial Neural Networks: Formal Models and Their Applications*, 2005, pp. 97-102.
- [4] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407-499, 2004.

Best Basis Matching Pursuit

Tianyao Huang, Yimin Liu, Huadong Meng, Xiqin Wang
 Department of Electronic Engineering, Tsinghua Univ., Beijing, China
 Email: huangty09@mails.tsinghua.edu.cn

Abstract—In grid-based Compressive Sensing (CS) approaches, the dictionary is built from a pre-defined grid. The bases off the grid points are left out of the dictionary. Some current Matching pursuit (MP) methods suffer from a degradation of performance when some off-grid bases exist. In this paper, a novel method namely Adaptively Sparse Recovery base on Constrained Total Least Squares (ASR-CTLS) is proposed to find the best bases even if they are off the grid. In the ASR-CTLS, the grid and the dictionary are adaptively updated with the CTLS technique. The convergence of the ASR-CTLS is theoretically analyzed, and numerical experiments on harmonic retrieval demonstrate the improvements of the ASR-CTLS.

I. SIGNAL MODEL AND ALGORITHM

Consider the grid-based CS model:

$$\mathbf{y} = \Phi(\mathbf{g})\mathbf{x} + \mathbf{w} \quad (1)$$

where \mathbf{y} is the measurement vector with N elements, and \mathbf{w} is the noise vector. \mathbf{x} is to be learned with D coefficients. Φ is built from the grid $\mathbf{g} = [g_1, g_2, \dots, g_D]$, which is generated by dividing a continuous space into D discrete grid points. For example, in harmonic retrieval we divide the frequency space and the Φ is the Fourier transform matrix. Off-grid basis emerges when the corresponding frequency point is not included in \mathbf{g} .

We don't often have enough priori to generate the perfect grid to guarantee that all of the nonzero elements in \mathbf{x} exactly lie on the grid points. So we cast the grid as an unknown parameter, and search the joint estimation of \mathbf{x} and \mathbf{g} via solving the optimum problem:

$$\hat{\mathbf{g}}, \hat{\mathbf{x}} = \arg \min \|\mathbf{x}\|_0, \text{ s.t. } \|\mathbf{y} - \Phi(\mathbf{g})\mathbf{x}\|_2^2 \leq \eta \quad (2)$$

where η is the noise power. In most cases, solving (2) is a complicated non-linear optimum problem. We introduce an iterative method.

Suppose $\hat{\mathbf{g}}^{(k)}$ is the estimate of \mathbf{g} , and is available after the k th iteration. To solve (2), loop between the following two equations:

$$\mathbf{x}_{\text{MP}} = \arg \min \|\mathbf{x}\|_0, \text{ s.t. } \left\| \mathbf{y} - \Phi(\hat{\mathbf{g}}^{(k)})\mathbf{x} \right\|_2^2 \leq \eta \quad (3)$$

$$\hat{\mathbf{g}}^{(k+1)}, \hat{\mathbf{x}}^{(k+1)} = \arg \min_{\mathbf{g}, \mathbf{x}} \|\mathbf{y} - \Phi(\mathbf{g})\mathbf{x}\|_2^2, \text{ s.t. supp}(\mathbf{x}) = \Lambda \quad (4)$$

Most MP methods presented before, e.g. OMP [2], CoSaMP [3], can be applied for (3) to obtain the sparsest solution \mathbf{x}_{MP} of \mathbf{x} . Notate $\Lambda = \text{supp}(\mathbf{x}_{\text{MP}})$ as the support set of \mathbf{x}_{MP} , and Λ is utilized in (4) to reduce the dimension.

The numerical solution to (4) can be obtained by three steps: estimate the mismatch in the grid with the CTLS technique; update the grid with the mismatch; estimate the \mathbf{x}_Λ with the projection onto the updated grid, where $(\cdot)_\Lambda$ as the elements indexed in Λ . The convergence of the solver for (4) can be theoretically guaranteed if the mapping $\Phi(\mathbf{g})$ is linear.

Define the mismatch in the grid as $\Delta\mathbf{g}_\Lambda = [\Delta g_1, \dots, \Delta g_{|\Lambda|}]^T$, thus $\mathbf{g}_\Lambda = \hat{\mathbf{g}}_\Lambda^{(k)} + \Delta\mathbf{g}_\Lambda$. Linearize the $\Phi_\Lambda(\mathbf{g}_\Lambda)$ at the local neighborhood of $\hat{\mathbf{g}}_\Lambda^{(k)}$ with Taylor expansion as

$$\Phi_\Lambda(\mathbf{g}_\Lambda) = \Phi_\Lambda(\hat{\mathbf{g}}_\Lambda^{(k)}) + \sum_{i=1}^{|\Lambda|} \mathbf{R}_i(\hat{\mathbf{g}}_\Lambda^{(k)}) \Delta g_i + \sum_{i=1}^{|\Lambda|} o(\Delta g_i^2) \quad (5)$$

where $o(\cdot)$ is the "big o" notation. Neglect $o(\Delta g_i^2)$, and substitute (5) into (1), thus we can apply the CTLS to estimate $\Delta\mathbf{g}_\Lambda$.

$$\begin{aligned} \widehat{\Delta\mathbf{g}}_\Lambda^{(k)} &= \arg \min_{\Delta\mathbf{g}_\Lambda, \mathbf{w}, \mathbf{x}_\Lambda} \left\| \begin{bmatrix} \Delta\mathbf{g}_\Lambda \\ \mathbf{w} \end{bmatrix} \right\|, \\ \text{s.t. } -\mathbf{y} + \left(\Phi_\Lambda(\hat{\mathbf{g}}_\Lambda^{(k)}) + \sum_i^{|\Lambda|} \mathbf{R}_i \Delta g_i \right) \mathbf{x}_\Lambda + \mathbf{w} &= \mathbf{0} \end{aligned} \quad (6)$$

The solution to (6) is derived in [1].

II. SIMULATIONS

We choose harmonic retrieval to illustrate the improvement of the proposed method. There are three sinusoids in observation. The amplitudes are $\alpha_1 = 20$, $\alpha_2 = 10$, $\alpha_3 = 1$; and the frequencies are $f_1 = 16.3/N$, $f_2 = 18.24/N$, $f_3 = 29.12/N$, where $N = 64$. Define $\text{SNR}_i = |\alpha_i|^2/\sigma^2$, where σ^2 is the noise power. The frequency space is uniformly divided into m grid points in xMP_m (OMP_{100N}, CoSaMP_{2N}, etc.), and N points in the ASR-CLTS. Fig.1 compares the Mean Square Errors (MSEs) of the frequency estimates. The results show that the ASR-CTLS obtains higher accuracy than OMP and CoSaMP especially while recovering the smallest sinusoid, and converges to the CR bound when the SNR is not less than -4dB.

REFERENCES

- [1] T. J. Abatzoglou, J. M. Mendel and G. A. Harada, "The constrained total least squares technique and its applications to harmonic superresolution," *Signal Processing, IEEE Transactions on*, vol. 39, pp. 1070-1087, 1991
- [2] M. A. Davenport and M. B. Wakin, "Analysis of Orthogonal Matching Pursuit Using the Restricted Isometry Property," *Information Theory, IEEE Transactions on*, vol. 56, pp. 4395-4401, 2010.
- [3] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, pp. 301-321, 2009.

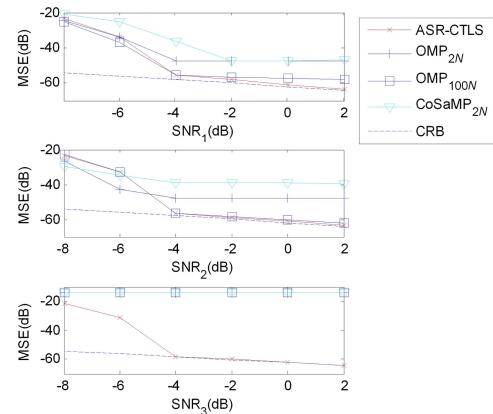


Fig. 1. The MSEs of the frequency estimates obtained in 500 independent Monte-Carlo trials. Three subplots are respectively dedicated to three sinusoids. CRB denotes the corresponding Cramer-Rao Bound.

Adaptive Algorithm for Online Identification and Recovering of Jointly Sparse Signals

Roy Amel

School of Electrical Engineering
Technion - Israel Institute of Technology
Haifa, Israel
Email: amelr2@tx.technion.ac.il

Arie Feuer

School of Electrical Engineering
Technion - Israel Institute of Technology
Haifa, Israel
Email: feuer@ee.technion.ac.il

Abstract—In this paper we present a novel method for online identification and recovering jointly sparse signals. This method can be useful in a number of applications such as blind sampling and reconstruction of multiband signals. There are several algorithms in the literature for solving jointly sparse vectors, but most of them are either greedy or not online algorithms (see e.g. [3]). Hence, introduce an inherent delay in the process. We introduce a novel online algorithm which is not greedy (in time) and improves the probability for success in identifying the joint support even in the presence of noise.

I. INTRODUCTION

In our work the model we look at is of the form

$$\mathbf{y}[t] = A\mathbf{x}[t] \quad (1)$$

Where $\mathbf{y}[t] \in \mathbb{R}^m$ represents the measured data, $\mathbf{x}[t] \in \mathbb{R}^n$ the signal we wish to reconstruct and A a known dictionary. It is assumed that $m < n$ and $\mathbf{x}[t]$ are jointly (in time) sparse. Namely, if $S[t]$ denotes the support of $\mathbf{x}[t]$, we have $S = \bigcup_t S[t]$ the joint support ($|S| \ll n$). This problem is referred to in the literature as Infinite Measurement Vectors (IMV) problem. Solving the problem at each t and letting $S[t-1] \subset S[t]$ leads to a greedy algorithm. Our approach is motivated by the ideas in [2]. Specifically, we solve at each time instance the problem:

$$(P1) \quad \min \|W[t]\mathbf{x}[t]\|_1 \text{ such that } \mathbf{y}[t] = A\mathbf{x}[t]$$

where $W[t]$ is diagonal matrix with non-negative entries and

$$W[t] = f(W[t-1], S[t-1]) \quad (2)$$

This type of algorithm is not greedy, by this we mean that it does not force $S[t-1] \subset S[t]$ and allows us make mistakes at a specific time, but still be able to, eventually, identify the right support with future data.

A. Noisy Environment

Under noisy environment conditions our model of work is changing to:

$$\mathbf{y}[t] = A\mathbf{x}[t] + \mathbf{v}[t] \quad (3)$$

where \mathbf{y} , A and \mathbf{x} are as above, $\mathbf{v}[t] \in \mathbb{R}^m$ is an additive noise. We assume that the noise is stationary, thus the way we cope with this problem is by measuring the noise variance, random several thousands of numbers under the same PDF and finally taking the largest one, we mark this number as C . Now with $\alpha \in [0.5, 1]$, we solve the next problem:

$$(P2) \quad \min \|W[t]\mathbf{x}[t]\|_1 \text{ such that } \|\mathbf{y}[t] - A\mathbf{x}[t]\|_\infty \leq \alpha \cdot C$$

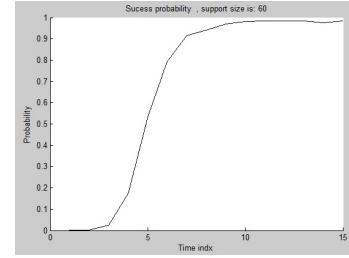


Figure 1. Success probability for true support evaluation with $n=200$, $m=100$ and $|S|=60$.

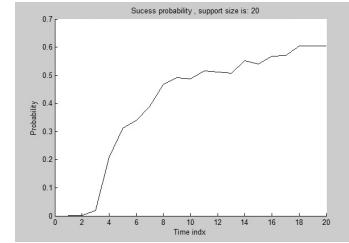


Figure 2. Success probability for true support evaluation with $n=200$, $m=100$, $|S|=20$ and $SNR = 30[dB]$

II. RESULTS

Figure 1 describes a simulation with $n=200$, $m=100$ and $|S|=60$. The simulation was repeated 250 times and the success ratio recorded. We note that after $t=15$ we get a very high probability of success even in this case where $|S|=60$. Moreover it seems that solving independently at each time instance will result in a close to zero probability of success.

Figure 2 describes a simulation result under noisy conditions with additive white gaussian noise. $n=200$, $m=100$, $|S|=20$ and $SNR = 30[dB]$. The simulation was repeated 250 times and the success ratio recorded. This result demonstrates the algorithm robustness to noise.

REFERENCES

- [1] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images", 2009, SIAM Review, Vol. 51, No. 1, pp. 34-81
- [2] E. J. Candès, M. B. Wakin, S. P. Boyd, "Enhancing Sparsity by Reweighted Minimization", J/ of Fourier Anal. And Appl. 14(5), Dec. 2008, pp. 877-905.
- [3] M. Mishali and Y.C. Eldar, " Blind Multiband Signal Reconstruction: Compressed Sensing for Analog Signals", IEEE Trans. SP, Vol. 57, No. 3, March 2009, pp. 903-1009.

Primal-Dual TV Reconstruction in Refractive Deflectometry

Adriana González*, Laurent Jacques†, Emmanuel Foumou* and Philippe Antoine*

*IMCN/NAPS, †ICTEAM/ELEN, Université catholique de Louvain, Belgium.

I. INTRODUCTION

Refractive deflectometry is a tomographic modality that measures light ray deflection when passing through transparent objects [1]. Combining multiple parallel light rays under various incident angles allows one to image the internal refractive-index distribution (or map) of complex materials (like optical fibers) while escaping from some limitations of interferometric systems (*e.g.*, instability to object vibrations, thickness measurement range).

II. FORWARD DEFLECTOMETRIC MODEL

Given a transparent material optically described by the refractive-index map $n : \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 \rightarrow n(\mathbf{x})$, a 2-D deflectometric measurement of n consists in measuring the deflection angle $\Delta(\tau, \theta)$ of a light ray of equation $\{\mathbf{x} : \mathbf{x} \cdot \mathbf{p}_\theta = \tau\}$, for $\tau \in \mathbb{R}$, $\theta \in [0, 2\pi]$, and $\mathbf{p}_\theta = (-\sin \theta, \cos \theta)$ perpendicular to the light ray direction $\mathbf{t}_\theta = (\cos \theta, \sin \theta)$. Mathematically, a first order linear approximation relates Δ to the Radon transform of the transverse gradient of n , that is:

$$\Delta(\tau, \theta) = \int_{\mathbb{R}^2} (\nabla n(\mathbf{x}) \cdot \mathbf{p}_\theta) \delta(\tau - \mathbf{x} \cdot \mathbf{p}_\theta) d^2 \mathbf{x}. \quad (1)$$

Since the Central Slice Theorem relates the 1-D Fourier Transform (FT) of an image Radon projections with the image 2-D FT [3], denoting \hat{n} as the 2-D FT of n , we can write

$$y(\omega, \theta) := \int_{\mathbb{R}} \Delta(\theta, \tau) e^{-i\tau\omega} d\tau = i\omega \hat{n}(\omega \mathbf{p}_\theta). \quad (2)$$

Alternatively, restricting (ω, θ) to $\mathbb{R}_+ \times [0, \pi]$ and setting $\mathbf{k} = (k_1, k_2) = \omega \mathbf{p}_\theta$ with $k = \|\mathbf{k}\| = \omega$, we have $\tilde{y}(\mathbf{k}) := y(\mathbf{R}\mathbf{k}) = i\mathbf{k} \hat{n}(\mathbf{k})$, with the $\pi/2$ rotation matrix \mathbf{R} (*i.e.*, $\mathbf{R}\mathbf{p}_\theta = \mathbf{t}_\theta$).

Therefore, assuming that the continuous refractive-index map n is approximated by $N = N_1 N_2$ values $\mathbf{n} \in \mathbb{R}^N$ arranged on a $N_1 \times N_2$ regular 2-D grid, the previous relations show that optical deflectometry can be associated to the forward linear model

$$\tilde{\mathbf{y}} = \mathbf{WSF}\mathbf{n} + \epsilon, \quad (3)$$

where $\tilde{\mathbf{y}} = (\tilde{y}(\mathbf{k}^1), \dots, \tilde{y}(\mathbf{k}^{M/2}))^T \in \mathbb{C}^{M/2} \simeq \mathbb{R}^M$ is the measurement vector on the observed frequency set $\mathcal{K} = \{\mathbf{k}^j\}_{1 \leq j \leq M/2}$, $\mathbf{F} \in \mathbb{C}^{N \times N}$ is the 2-D discrete FT, $\mathbf{S} \in \{0, 1\}^{M/2 \times N}$ is a binary matrix selecting \mathcal{K} in \mathbf{F} output (with $\mathbf{S}\mathbf{S}^* = \text{Id}$), and $\mathbf{W} \in i\mathbb{R}^{M/2 \times M/2} = i\text{diag}(k_1, \dots, k_{M/2})$. The vector $\epsilon \in \mathbb{C}^{M/2}$ is a (complex) Gaussian noise $\epsilon_j \sim \mathcal{N}(0, \sigma^2) + i\mathcal{N}(0, \sigma^2)$ with $\|\epsilon\|^2 < \varepsilon^2 := (M + c\sqrt{M})\sigma^2$ with high probability for $c = O(1)$.

III. INVERSE PROBLEM SOLVING

A realistic sampling of the Fourier plane as materialized by \mathbf{S} is obtained from a set of T radial lines associated to T different angular observations θ in (2). Most of the time, reconstructing \mathbf{n} from (3) is an ill-posed inverse problem since $M = M(T) < N$. However, actual refractive-index maps of transparent materials are composed of slowly varying areas separated by sharp boundaries (material interfaces). This inverse problem can therefore be regularized by assuming a Bounded Variation (BV) model of \mathbf{n} . In other words, we solve

$$\arg \min_{\mathbf{u}} \|\mathbf{u}\|_{TV} \text{ s.t. } \|\tilde{\mathbf{y}} - \mathbf{WSF}\mathbf{u}\| \leq \varepsilon, \quad (4)$$

where minimizing $\|\mathbf{u}\|_{TV} = \sum_j |(\nabla \mathbf{u})_j|$ promotes the BV (cartoon shaped) model while imposing data fidelity $\|\tilde{\mathbf{y}} - \mathbf{WSF}\mathbf{u}\| \leq \varepsilon$ [2].

Part of this work is funded by the DETROIT project (WIST3/SPW, Belgium).

In order to quantify the “ill-posedness” of (3), we simplify the sampling made by \mathbf{S} by picking uniformly at random $M/2$ complex frequencies on \mathcal{K} (*i.e.*, M real values) with $\mathcal{K} \subset \{\mathbf{k} : k_1 \geq 0\}$.

Practically, the convex minimization (4) can be recast as $\arg \min_{\mathbf{u}} \iota_C(\mathbf{Au}) + \|\mathbf{u}\|_{TV}$, where $\mathbf{A} = \mathbf{WSF} \in \mathbb{C}^{M \times N}$, $C = \{\mathbf{v} \in \mathbb{C}^M : \|\tilde{\mathbf{y}} - \mathbf{v}\| \leq \varepsilon\}$, and $\iota_C(\mathbf{v})$ is the indicator function equals to 0 if $\mathbf{v} \in C$ and ∞ otherwise. Due to the presence of the diagonal operator \mathbf{W} , a particularity of optical deflectometry compared to common tomographic techniques (*e.g.*, MRI or radio interferometry), the sensing matrix \mathbf{A} presents some unfavorable properties as it is not a tight frame ($\mathbf{A}\mathbf{A}^* \neq I$). The recently proposed algorithm by Chambolle and Pock (Algorithm 1 in [2]), which is based on a primal-dual formulation, relaxes the conditions on the operator, making this method suitable for solving (4) despite a non-differentiable objective.

IV. RESULTS

Chambolle-Pock (CP) method was analyzed using the well-known Shepp-Logan image (Fig. 1) which, being characterized by sharp edges and large smooth areas, allows simulating the refractive-index map characteristics. CP is compared with the Least-Squares (LS) Method, an intuitive way of solving linear systems using the operator pseudoinverse $\mathbf{A}^\dagger := \mathbf{A}^*(\mathbf{A}\mathbf{A}^*)^{-1} = \mathbf{F}^*\mathbf{S}^*\mathbf{W}^{-1}$ such that $[\arg \min_{\mathbf{u}} \|\mathbf{u}\| \text{ s.t. } \tilde{\mathbf{y}} = \mathbf{Au}] = \mathbf{A}^\dagger \tilde{\mathbf{y}}$. Fig. 1 shows the behavior of both algorithms regarding noise and the under-sampling ratio (M/N), presenting the mean SNR computed for 10 trials in the reconstruction of the map.

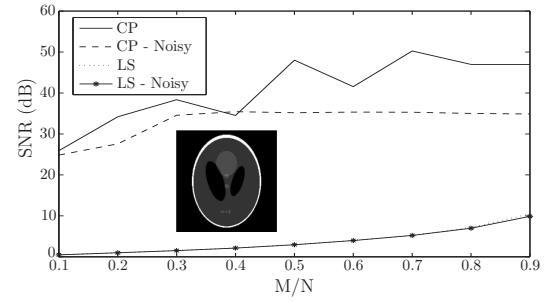


Fig. 1: SNR vs M/N .

V. CONCLUSION

Important results were obtained in refractive-index map reconstruction from (simulated) optical deflectometry. The inverse problem reconstructing \mathbf{n} from (3) is regularized by minimizing the Total Variation norm $\|\mathbf{n}\|_{TV}$ of the map (BV model). Thanks to the CP algorithm dealing with the untight operator \mathbf{A} , the method yields optimal reconstruction with high robustness to noise even for few measurements. This must be compared with the LS method that degrades rapidly for increasing noise level as the diagonal operator \mathbf{W} increases its impact on low frequencies. An open issue still remains in order to represent more accurately the data measured by the instrument. As Fourier Transform of Radon projections provides data in radial coordinates, further development must be done to find its correspondence in Cartesian coordinates.

REFERENCES

- [1] E. Foumouo, J.-L. Dewandel, L. Joannes, D. Beguin, L. Jacques, and P. Antoine, “Optical tomography based on phase-shifting schlieren deflectometry,” *Optics Letters*, **35**(22), pp. 3745–3747, 2010.
- [2] A. Chambolle and T. Pock, “A first-order primal dual algorithm for convex problems with applications to imaging,” *JMIV*, **40**(1), pp. 120–145, 2011.
- [3] R. N. Bracewell, “Numerical transforms,” *Science*, **248**(4956), pp. 697–704, 1990.

Greedy decompositions with randomly varying time-frequency subdictionaries

Manuel Moussallam , Gaël Richard

Institut Telecom - Telecom ParisTech - CNRS/LTCI
37, rue Dareau 75014 Paris, France

Laurent Daudet

Institut Langevin - ESPCI - Univ. Paris Diderot - CNRS UMR 7587
10, rue Vauquelin 75005 Paris, France

Introduction Greedy algorithms such as Matching Pursuit (MP) [1] and its variants (Orthogonal MP [2]) are widely used for sparse approximation. They are based on the repetition of two steps: 1: Select an atom in the dictionary and 2: Update residual. We are here interested in the first step. MP and OMP select the atom ϕ_{γ_i} in the dictionary Φ that maximizes the residuals projection $\phi_{\gamma_i} = \arg \max_{\phi_{\gamma} \in \Phi} |\langle R^{n-1} f, \phi_{\gamma} \rangle|$. There is always a tradeoff in the choice of the dictionary. If it is dense, we have a fast decay of the approximation error (as a function of the approximation order), but computations get cumbersome, and the cost of encoding each coefficient may become prohibitive for coding. If the dictionary is small (slight or no overcompleteness), computations are fast, the coding cost per coefficient is low, but more coefficients are needed. Other schemes ensure a better local fit of the selected atoms using a different correlation function [3]. Also, probabilistic approaches have been introduced [4] where successive runs with random sub-optimal atom selection are performed, then averaging yields a robust sparse approximation.

Here, we propose a different paradigm, that mitigates the drawbacks of using a large dictionary while keeping most of the benefits. We keep the standard correlation function but we randomly switch the subdictionary at each iteration, where the subdictionary is a subset of a large dictionary. Adaptive techniques have been proposed that first search in a *fixed* smaller subdictionary, and then find a local maximum in the large dictionary [1], [5]. Our approach keeps this dictionary subsampling paradigm, but the small subdictionaries are randomly alternated so as to maximize the probability that the large dictionary is evenly spanned during the process. The key point is that the choice of subdictionary is *not* adaptive, but is parametrized by a fixed pseudo-random sequence, also known by the decoder. In other words, we have the (theoretical) complexity of working with a small dictionary, and the small coding costs, but the whole large dictionary is spanned. In the following, we study the quality of the approximation: we prove the benefits of this method for synthetic signals, and show numerically that this behavior extends to the case of real signals.

Theoretical justification with sparse signal model Let assume an exact-sparse model $f = \sum_{j \in J} \alpha_j \phi_j$ with the $\{\phi_j\}$ a subset of a larger dictionary Φ verifying $\phi_j^T \phi_k \approx \delta_{j,k}$. Let X be a random variable from which are drawn the projections $X_k = |\langle f, \psi_k \rangle| \approx |\alpha_{j_0(k)} \langle \phi_{j_0(k)}, \psi_k \rangle|$ in the subdictionary $\Psi \subset \Phi$, where $j_0(k) = \arg \max_j |\langle \phi_j, \psi_k \rangle|$. Let us denote $X_{k:J}$ the k^{th} biggest X , then the residuals energy at the n^{th} iteration of a Matching Pursuit in Ψ is given by $\|R^n f\|^2 = \|f\|^2 - \sum_{k=0}^{n-1} X_{J-k:J}^2$ because the n biggest atoms have been selected one after another and subtracted. Now if Ψ is changed at iteration k , it is equivalent to redrawing the projections X_j . Since k components of f have already been (nearly) subtracted,

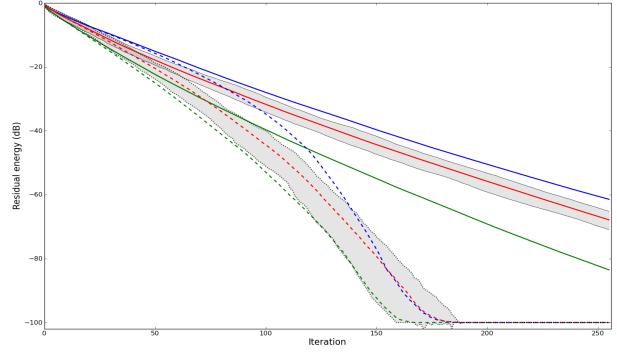


Fig. 1. Average residual energy decay for MP (continuous) and OMP (dashed) on full dictionary (green), fixed subdictionary (blue) and randomly varying subdictionaries (red - variance zone in light grey) for a 256-length noisy signal (20 runs)

only $J - k$ values are drawn. Atom selected at iteration k is the one that maximizes this $J - k$ length sequence, and the residuals energy if the dictionary changes at each iteration is described by $\|R^n f\|^2 = \|f\|^2 - \sum_{k=0}^{n-1} X_{J-k:J}^2$. Knowing Φ , one can derive a probabilistic model for X and using order statistics, prove that $\mathbb{E}(X_{J-k:J}^2) \geq \mathbb{E}(X_{J-k:J}^2)$, thus ensuring faster convergence of the new algorithm.

Conclusion The proposed algorithm appears to be suitable for sparse approximation of complex signals (though not for recovery). The potential benefits are in low bitrate compression, and we exhibit several sound examples were these advantages show off. The unsupervised nature of the algorithm and the randomness introduced in the atom selection makes it very easy to design worst-case scenarios for which the algorithm would converge slower than a pursuit over a fixed dictionary. However, on average, and with a small empirical variance, the proposed scheme appears to have the coding costs of the small dictionary with a decay rate close to the one on the large dictionary.

REFERENCES

- [1] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Sig. Proc.*, no. 12, December 1993.
- [2] Y. Pati, R. Rezaifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27 th Asilomar Conf. on Signals, Systems, and Computers*, 1993, pp. 40–44.
- [3] R. Gribonval, E. Bacry, S. Mallat, P. Depalle, and X. Rodet, "Analysis of sound signals with high resolution matching pursuit," *IEEE Symp. TITS*, pp. 125 –128, Jun. 1996.
- [4] S. E. Ferrando, E. J. Doolittle, B. A. J., and L. Bernal, "Probabilistic matching pursuit with gabor dictionaries," *Elsevier Sig. Proc.*, 2000.
- [5] R. Gribonval, "Fast matching pursuit with a multiscale dictionary of gaussian chirps," *IEEE Trans. Sig. Proc.*, no. 5, May 2001.

A Sparsity based Regularization Algorithm with Automatic Parameter Estimation

Damiana Lazzaro

Department of Mathematics, University of Bologna, Italy

Abstract—Regularization methods for the solution of ill-posed inverse problems can be successfully applied if a right estimation of the regularization parameter is known. While there exists a significant amount of research in the literature on the development of appropriate parameter selection methods for Tikhonov regularization [7], until now the L_1 regularization case has been considered in only a few very recent papers, [1], [3], [4], [5], [6] that focus only on Total Variation regularization. Among them, the only method that, to our knowledge, does not require any assumption on the noise level, has been proposed in[1] using variational distribution approximations. It updates the regularization parameter in an iterative manner and seem to be very effective for TV restoration problems when information about the perturbation level is not available. Nevertheless, in spite of the good performance of this method, its computational cost is still too high for real-time practical applications. In this work we consider the image deblurring problem and we evaluate its solution using a sparsity based regularization approach solved by means of the iterative forward-backward splitting method. The main contribution of this research is the proposal of a novel adaptive automatic rule for the estimation of the regularization parameter in L_1 -based restoration problems, without requiring any assumption about the perturbation process. This rule, developed in the context of the iterative forward-backward splitting method [2], exploits the information yielded by this approach to dynamically update the parameter value following the evolution of the objective functional. The iterative algorithm automatically stops, when the parameter has reached a seemly near optimal value. A large number of numerical experiments confirm that the proposed rule yields restoration results competitive with those of the best state-of-the-art algorithms.

- [4] Y. Lin, B. Wohlberg, and H. Guo, UPRE Method for Total Variation Parameter Selection, *Signal Processing*, Vol. 90, pp. 2546–2551, 2010.
- [5] J.P. Oliveira, J.M. Bioucas-Dias, and M.A.T. Figueiredo, Adaptive Total Variation Image Deblurring: A Majorization-Minimization Approach, *Signal Processing*, Vol. 89, pp. 1683–1693, 2009.
- [6] Y.W. Wen and A.M. Yip, Adaptive Parameter Selection for Total Variation Image Deconvolution, *Numer. Math.Theor. Meth. Appl.*, Vol. 2 No. 4, pp. 427–438, 2009.
- [7] C.R. Vogel, *Computational Methods for Inverse Problems*, SIAM Philadelphia, 2002.

REFERENCES

- [1] S.D. Babacan, R. Molina, and A.K. Katsaggelos , Parameter Estimation in TV Image Restoration using Variational Distribution Approximation, *IEEE Trans. Image Process.* Vol. 17, No. 3, pp. 326–339, 2008.
- [2] P.L. Combettes and V.R. Wajs , Signal Recovery by Proximal Forward-Backward Splitting, *SIAM Journal on Multiscale Modelling and Simulation*, Vol.4, No. 4, pp. 1168–1200, 2005.
- [3] H.Liao, F. Li, and M.K. Ng, Selection of Regularization Parameter in Total Variation Image Restoration, *Journal Optical Soc. Am. A*, Vol. 26, No. 11, pp. 2311–2320, 2009.

An unsupervised iterative shrinkage/thresholding algorithm for sparse expansion in a union of dictionnaries.

Matthieu Kowalski and Thomas Rodet
 Univ. Paris-Sud – Laboratoire des Signaux et Systèmes
 Email: {kowalski,rodet}@lss.supelec.fr

Abstract—We are interested in hybrid/morphological decomposition of signals which can be expressed as the sum of two sparse components in a well choosen frame and a residual – such as images (edges+textures) or audio signal (tonals + transients). Thanks to a Bernoulli-Gaussian prior on the synthesis coefficients, we derive an unsupervised algorithm in the spirit of ISTA with iteratively adapted thresholding/shrinkage. The model can be directly extended to joint-sparsity structure.

I. MODEL

Let $\mathbf{U} = \{\mathbf{u}_n \in \mathbb{C}^N\}_{n=1}^{N_1}$ and $\mathbf{V} = \{\mathbf{v}_m \in \mathbb{C}^N\}_{m=1}^{N_2}$ two frames of \mathbb{C}^N . Let Λ and Δ two finit subsets of \mathbb{N} . We are interrested in “signals $\mathbf{y} \in \mathbb{C}^N$ which admit an hybrid [1]/Morphological [3] expansion:

$$\mathbf{y} = \sum_{\lambda \in \Lambda} \alpha_\lambda \mathbf{u}_\lambda + \sum_{\delta \in \Delta} \beta_\delta \mathbf{v}_\delta + \mathbf{n}, \quad (1)$$

where $\mathbf{n} \in \mathbb{C}^N$ is some noise, and where $\alpha_\lambda, \beta_\delta \in \mathbb{C}$ are the *synthesis* coefficients. The sets Λ and Δ are called the significance maps: if $\lambda \in \Lambda$, then $\alpha_\lambda \neq 0$, and if $\lambda \notin \Lambda$, then $\alpha_\lambda = 0$.

We choose a Bernoulli-Complex Gaussian prior which naturally induces sparsity. Let us definine the indicator variables:

$$X_n = \begin{cases} 1 & \text{if } n \in \Lambda \\ 0 & \text{otherwise} \end{cases}, \quad \tilde{X}_m = \begin{cases} 1 & \text{if } m \in \Delta \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

with p and \tilde{p} the membership probabilities: $\forall n, p = p(X_n = 1), \forall m, \tilde{p} = p(\tilde{X}_m = 1)$. The corresponding model can then be written:

$$\mathbf{y} = \sum_{n=1}^{N_1} X_n \alpha_n \mathbf{u}_n + \sum_{m=1}^{N_2} \tilde{X}_m \beta_m \mathbf{v}_m + \mathbf{n}. \quad (3)$$

In the following, we restrain ourself to the significance map Λ , all the results can be directly applied to the map Δ with the Bernoulli-Gaussian prior:

$$p(\alpha_\lambda | X_\lambda) = (1 - X_\lambda)\delta_0 + X_\lambda \mathcal{CN}(0, \sigma^2, 0)$$

where the complex-Gaussian density are univariate.

We finally assume $\mathbf{n} \sim \mathcal{CN}(0, \sigma_0^2 \mathbf{I})$.

II. ALGORITHM DERIVATION

A. GEM

For the sake of simplicity, we rewrite (1):

$$\mathbf{y} = \mathbf{U}\boldsymbol{\alpha} + \mathbf{V}\boldsymbol{\beta} + \mathbf{n} = \Phi\boldsymbol{\theta} + \mathbf{n}, \quad (4)$$

with $\Phi = [\mathbf{U} \mathbf{V}]$ and $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$.

Inspired by [2], one can use a Generalized Expectation-Maximization (GEM) strategy in order to maximize the penalized likelihood. By introducing the hidden variable \mathbf{z} , we have

$$\mathbf{z} = \boldsymbol{\theta} + \mu \mathbf{n}_1 \quad \mathbf{y} = \Phi \mathbf{z} + \mathbf{n}_2, \quad (5)$$

with $\mathbf{n}_1 \sim \mathcal{CN}(0, \mathbf{I})$ and $\mathbf{n}_2 \sim \mathcal{CN}(0, \sigma_0^2 \mathbf{I} - \mu^2 \Phi \Phi^T)$ ($\mu^2 < \frac{\sigma_0}{\|\Phi \Phi^T\|}$).

a) *E-setp*: $\hat{\mathbf{z}}^t = \boldsymbol{\theta}^t + \frac{\mu^2}{\sigma_0^2} \Phi^* (\mathbf{y} - \Phi \boldsymbol{\theta}^t)$.
 b) *M-step*: the maps are first estimated by marginalization and maximization

$$\hat{\Lambda}^{t+1}, \hat{\Delta}^{t+1} = \arg \max_{\Lambda, \Delta} \mathbb{E}\{\log(p(\Lambda, \Delta | \mathbf{y}, \mathbf{z}, T)) | \mathbf{y}, \boldsymbol{\theta}^t, \Lambda^t, \Delta^t, T\},$$

then the coefficients $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}^{t+1} = \arg \max_{\boldsymbol{\theta}} \mathbb{E}\{\log(p(\boldsymbol{\theta} | \hat{\Lambda}^{t+1}, \hat{\Delta}^{t+1}, \mathbf{y}, \mathbf{z}, T)) | \mathbf{y}, \boldsymbol{\theta}^t, \Lambda^t, \Delta^t, T\}$$

B. Maps estimation

For the significance map Λ we obtain:

$$\hat{X}_\lambda^{t+1} = \begin{cases} 1 & \text{if } |\hat{z}_\lambda^t| > \sqrt{\frac{(\mu^2 + \sigma^2)\mu^2}{\sigma^2} \ln \left[\frac{1-p}{p} \frac{\mu^2 + \sigma^2}{\mu^2} \right]} \\ 0 & \text{if not} \end{cases}$$

Remark: with a joint-sparsity model on the map, we obtain a similar rule on the norm of groups of coefficients.

C. Coefficients estimation

We can then estimate the coefficients. If $\hat{X}_\lambda^{t+1} = 1$

$$\hat{\theta}_\lambda^{t+1} = \arg \max_{\theta_\lambda} \frac{\|\hat{z}_\lambda^t - \theta_\lambda\|}{2\mu^2} + \frac{\|\theta_\lambda\|^2}{2\sigma^2} = \frac{\hat{z}_\lambda^t}{1 + \mu^2/\sigma^2}, \quad (6)$$

and $\hat{\theta}_\lambda^{t+1} = 0$ if not.

III. RESULTS

We apply such a strategy to a xylophone signal with a union of Gabor dictionnaires. We find $\hat{\sigma}_0^2 = 0.001$ (output SNR $\simeq 16$ dB) for $\sigma_0^2 = 0.0009$ (input SNR $\simeq 12$ dB).

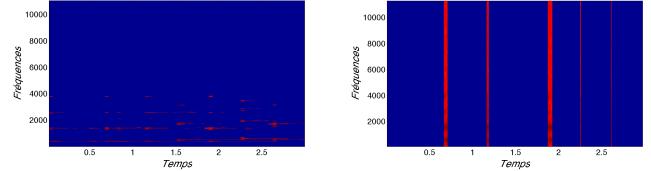


Fig. 1. Estimated Gabor coefficients of the two significance maps.

REFERENCES

- [1] L. Daudet and B. Torrésani. Hybrid representations for audiophonic signal encoding. *Signal Processing*, 82(11):1595–1617, 2002. Special issue on Image and Video Coding Beyond Standards.
- [2] M. Figueiredo and R. Nowak. An em algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8):906–916, 2003.
- [3] J.-L. Starck, M. Elad, and D. L. Donoho. Image decomposition via the combination of sparse representation and a variational approach. *IEEE Transaction on Image Processing*, 14(10):1570–1582, 2005.

An Infeasible-Point Subgradient Algorithm and a Computational Solver Comparison for ℓ_1 -Minimization

Andreas M. Tillmann*

Institute for Mathematical Optimization
Technische Universität Braunschweig
38106 Braunschweig, Germany
Email: a.tillmann@tu-bs.de

Dirk A. Lorenz

Institute for Analysis and Algebra
Technische Universität Braunschweig
38106 Braunschweig, Germany
Email: d.lorenz@tu-bs.de

Marc E. Pfetsch

Institute for Mathematical Optimization
Technische Universität Braunschweig
38106 Braunschweig, Germany
Email: m.pfetsch@tu-bs.de

Abstract—The ℓ_1 -minimization problem $\min\{\|x\|_1 : Ax = b\}$, also known as Basis Pursuit (BP), has become important in Compressed Sensing due to its ability to yield the sparsest solution of an underdetermined linear system $Ax = b$ under certain conditions. In the past few years, a lot of new algorithms solving (BP) or some (possibly regularized) variant of it have been developed. We contribute the ISAL1 to the available spectrum of solvers, which is a specialization to (BP) of a new infeasible-point subgradient algorithm (ISA) for solving convex constrained minimization problems. In this extension of the classical projected subgradient methods from nonsmooth optimization, the projections onto the feasible set need only be approximated, thus allowing for a potentially large reduction of the computational burden. In particular, in ISAL1, inexact projection onto $\{x | Ax = b\}$ by performing a fixed small number of conjugate gradient steps suffices to obtain convergence. Moreover, we will present results of an extensive computational comparison of various state-of-the-art ℓ_1 -solvers, also including ISAL1. Furthermore, we show how a new optimality check can speed up solvers and at the same time attain the true optimum (up to numerical precision).

I. INTRODUCTION

We propose a new infeasible-point subgradient algorithm (ISA) for the constrained minimization of convex functions. In contrast to typical projected subgradient schemes [1], the projection onto the feasible set is only approximated in the ISA. Hence, the iterates may be infeasible throughout the whole procedure, and still convergence can be achieved under certain conditions. This allows to tackle problems where computing the exact projections is expensive, especially for large-scale instances.

Here, we focus on one such problem: Finding the minimal- ℓ_1 -norm solution to an underdetermined linear system, i.e., (BP). This problem has become very important in the field of Compressed Sensing, because under certain conditions (e.g., the RIP), it allows for exact recovery of the minimum-support solution (ℓ_0 -minimizer) of $Ax = b$, which is generally \mathcal{NP} -hard to find.

Facing the vast choice of available (BP) solvers developed over the past years, one may wonder which one is “the best”? Of course, there are multiple answers (if any), depending on context, desired accuracy, etc. Here, we aim at an *exact* solution of (BP), without concern about special cases or related problems such as ℓ_0 -minimization. To this end, we conduct extensive numerical experiments with various prominent ℓ_1 -minimization solvers and the new ISAL1. We also present an easily implementable optimality check and demonstrate its usefulness by further computational experiments with several solvers.

II. NEW ALGORITHMS

The existing arsenal of solvers for (BP) includes methods based on various ideas, e.g., augmented Lagrangeans, interior-point schemes, spectral projected gradients or several penalized models. We propose to add the ISA modification of projected subgradient methods.

A. ISA and ISAL1

The ISA iteration consists mainly of the iterate update $x^{k+1} = P_X^{\varepsilon_k}(x^k - \alpha_k h^k)$, with stepsize α_k , subgradient h^k and the inexact projection operator $P_X^{\varepsilon_k}$ for the feasible set X with accuracy ε_k . We investigated the ISA for several stepsize choices and obtained convergence results depending on the behaviour of the series of projection accuracies (ε_k).

ISAL1 specializes the ISA for solving (BP). For this particular problem, projection onto $\{x | Ax = b\}$ amounts to solving a linear system. This can be done by applying the method of conjugate gradients (CG). For the inexact projections in ISAL1, we derive bounds on the CG residual norm which (among other things) guarantee the method’s convergence. In practice, convergence can still be obtained when only computing a fixed small number of CG steps.

B. Heuristic Support Evaluation

Our implementation of ISAL1 exhibited a typical drawback of subgradient methods: slow local convergence. This issue was successfully overcome by integrating a Heuristic Support Evaluation (HSE) scheme, which allows “jumping” to the exact optimal solution by roughly checking a well-known optimality criterion for (BP) (see, e.g., [2]) on the estimated true solution support.

III. COMPUTATIONAL COMPARISON OF ℓ_1 -SOLVERS

We carefully constructed a test set consisting of a wide range of (BP) instances, each with an optimal solution guaranteed to be unique (by employing the ERC or the Source Condition). We compare several well-known algorithms which can provably solve general (BP) instances, namely: SPGL1 (www.cs.ubc.ca/labs/scl/spgl1), YALL1 (yall1.blogs.rice.edu), ℓ_1 -MAGIC (www.acm.caltech.edu/l1magic), SolveBP of SparseLab (sparselab.stanford.edu) (employs PDCCO), L1-Homotopy (users.ece.gatech.edu/~sasif/homotopy), and ISAL1. As a reference, we also solved (BP) as a linear program, using the dual simplex method of CPLEX. Most available implementations seem to go for a balanced speed-accuracy trade-off by default, i.e., aiming at fast termination with a medium-accuracy result which hence not necessarily qualifies as an “exact solution”. Integrating the HSE however shows that without needing to change algorithmic parameters, one can achieve both a speed-up and highest accuracy at the same time. The potential of the HSE is supported by numerical experiments with several solvers.

REFERENCES

- [1] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*. Springer, 1985.
- [2] J.-J. Fuchs, “On sparse representations in arbitrary redundant bases,” *IEEE Transactions on Information Theory*, vol. 50, pp. 1341–1344, 2004.

* Supported by a DFG research grant.

On the relation between perceptrons and non-negative matrix factorization

Hugo Van hamme

Department ESAT, Katholieke Universiteit Leuven, Leuven, Belgium.

Email: hugo.vanhamme@esat.kuleuven.be

A classical result in the theory of artificial neural networks (ANNs) is Cybenko's theorem [1] which states that a perceptron with at least one hidden layer, sigmoidal output non-linearity and a sufficiently high, but finite, number of nodes can approximate any continuous and bounded function on a bounded domain within a given accuracy. This theorem has made multi-layer perceptrons (MLPs) a popular instrument for classification problems.

Perceptrons are inspired by the information processing in neurons [2], which makes them a trivial choice for modeling cognitive processes. Again inspired by the architecture of mammal brains, one would expect a several layers to be required in perceptron models. Deep MLP networks are however very difficult to train with the back-propagation algorithm. Supervisory information is needed and the fact that it cannot be provided for all hidden layers forms the crux of the problem.

In this paper, the parallels and differences between non-negative matrix factorization (NMF) and a single layer perceptron are discussed. NMF is capable of working without supervision but is equally capable of exploiting supervision information, which makes it very well suited to overcome the training problems in multilayer architectures. However, NMF additionally shows behavior that is observed in the brain: it performs lateral inhibition in a single layer and has masking properties.

NMF finds two factors $\mathbf{W} \in \mathbb{R}^{N \times R}$ and $\mathbf{H} \in \mathbb{R}^{R \times T}$, with positive or zero entries such that a data matrix $\mathbf{V} \in \mathbb{R}^{N \times T}$ with positive or zero entries is approximated by $\mathbf{W}\mathbf{H}$. The Frobenius norm and the (generalized) Kullback-Leibler divergence (KLD) are considered here as cost function to express the proximity of \mathbf{V} and $\mathbf{W}\mathbf{H}$. The weights \mathbf{W} are learned unsupervisedly by applying NMF to training data \mathbf{V} , where different training tokens occupy different columns in \mathbf{V} .

In this contribution the multiplicative update algorithms proposed in [3] are cast in the flow diagrams of Figure 1. When classifying a single test token, \mathbf{V} and \mathbf{H} become vectors \mathbf{v} and \mathbf{h} , while \mathbf{W} is assumed to be known from the training phase. Finding the NMF solution for \mathbf{h} is compared to node activation in an ANN. The left pane shows the single layer perceptron which maps \mathbf{v} to the hidden nodes \mathbf{y} through the weight matrix \mathbf{W} . A sigmoidal output non-linearity $\mathbf{h}(\mathbf{y})$ with diagonal structure is then applied. In NMF with Frobenius norm (middle), $\mathbf{h}(\mathbf{y})$ is replaced by the shaded box. The right pane shows the flow diagram for the KLD metric, which differs only from the middle pane by swapping the input projection \mathbf{W}^t and the element-wise division \oslash . Both NMF problems are shown to have lateral inhibition behavior, i.e. a hidden node will deactivate nodes with similar weights (or neighboring neurons in the brain). Physically, inhibition leads to categorical perception. Mathematically,

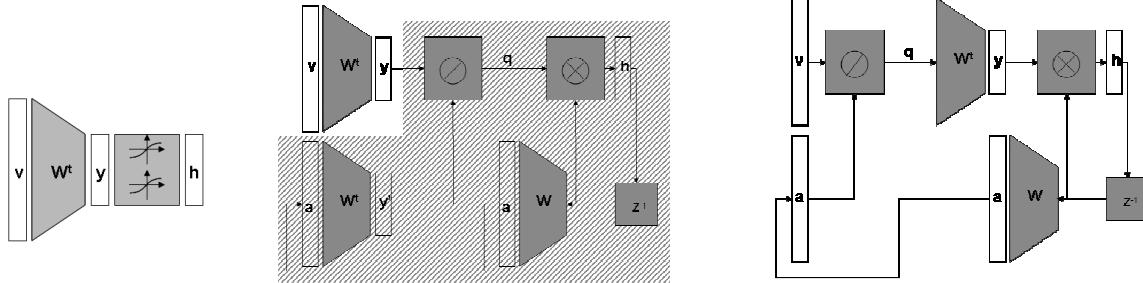


Fig. 1: A flow diagram representation of a single layer perceptron (left), NMF with Frobenius norm (middle) and NMF with KLD (right). \otimes and \oslash are element-wise multiplication and division respectively. “z⁻¹” is a memory over one iteration of the update formulae of [3].

tically, inhibition corresponds to the case where the minimal (Frobenius or KLD) cost is achieved on the constraint boundary.

Supervisory information can be introduced in NMF by augmenting \mathbf{V} with \mathbf{G} such that. $G_{ij} = 1$ if the i -th class is present in token j (and zero otherwise)

$$\begin{bmatrix} \mathbf{G} \\ \mathbf{V} \end{bmatrix} \approx \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix} \mathbf{H}$$

On test tokens, \mathbf{H} can be computed as before from \mathbf{V} and subsequently the class information \mathbf{G} can be estimated by forming $\mathbf{X}\mathbf{H}$. In other words, with supervision, \mathbf{X} acts like an output layer of a perceptron. With the previous result, NMF can be interpreted as an easily learnable perceptron with an output mapping that shows lateral inhibition, which adds to the cognitive motivation to use it as a building block for deep ANN structures.

Finally, the cognitive motivation is strengthened by observing that NMF also shows masking behavior in the sense that the sparsity pattern of \mathbf{H} is robust to perturbations in \mathbf{V} and not all changes to the input data affect \mathbf{H} and hence the classification. A first masking mechanism is through L_1 regularization of the cost function. Secondly, a positive bias term \mathbf{V}_0 can be added: find the best \mathbf{H} s.t. $\mathbf{V} \approx \mathbf{W}\mathbf{H} + \mathbf{V}_0$. The case where $\mathbf{V}_0 = \mathbf{W}\mathbf{H}_0$ for some \mathbf{H}_0 could model a reminiscent (decaying) neuron activation. The effect is that the cone of possible \mathbf{V} spanned by the model $\mathbf{W}\mathbf{H} + \mathbf{V}_0$ is moved away from the origin, which will typically activate more constraints at the optimal \mathbf{H} and can even lead to solutions that are not affected by the data \mathbf{V} at all (masking).

In summary, there are strong parallels between a single layer perceptron and NMF, the latter showing the additional cognitive properties of unsupervised learnability, lateral inhibition and masking.

REFERENCES

- [1] G. Cybenko, “Approximations by superpositions of sigmoidal functions,” in *Mathematics of Control, Signals, and Systems*, vol. 2, no. 4, pp. 303–314, 1989.
- [2] F. Rosenblatt, “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain,” in *Psychological Review*, vol. 65, pp. 386–408, 1958.
- [3] D. Lee, and H. Seung, “Algorithms for non-negative matrix factorization,” in *Advances in Neural Information Processing Systems*, vol. 13, pp. 556–562, 2001.

Recovery of finite alphabet signals from incomplete measurements

Si Mohamed Aziz Sbaï, Abdeldjalil Aïssa-El-Bey and Dominique Pastor
 Institut Télécom; Télécom Bretagne; UMR CNRS 3192 Lab-STICC, Technopôle Brest Iroise, France
 Université européenne de Bretagne, France

Abstract—This work presents a new framework for recovering finite alphabet signals. We address the problem of finding solutions to underdetermined systems of linear equations drawn from finite alphabet. We formulate this problem as a recovery of sparse signals from highly incomplete measurements. It is known that sparse solutions can be obtained by ℓ_1 minimization, through convex optimization. This relaxation procedure in our problem fails in recovering sparse solutions. However, the reconstruction of the finite alphabet signals is possible without exhibiting the sparse solutions. Empirical results show that this approach provides good recovery performance for random sensing matrices.

I. INTRODUCTION

Compressed sensing (CS) is a new concept that exploits sparsity of the signals in the acquisition process. The objective of CS is to reduce the number of the non adaptive measurements to be taken from signals, that is, the number of necessary measures required to reconstruct the signals. Recently, CS has attracted growing interests in a variety of fields, including source separation, radar, and communication.

In this study, instead of acquiring sparse signals, we wish to acquire signals drawn from a known finite alphabet. In this paper, we show that this problem can be expressed as a sparse recovering problem. The convex relaxation of this problem provides good recovery performance for random sensing matrices when a condition on the number of missing measurements holds.

II. PROBLEM FORMULATION

Suppose we are given $\mathbf{y} \in \mathbb{R}^m$ and a full-rank mixing matrix $\Phi \in \mathbb{R}^{m \times n}$ with $m < n$. The underdetermined linear system of equations $\mathbf{y} = \Phi\mathbf{x}$ has infinitely many solutions. The objective of this work is to find solutions drawn from a finite alphabet $\mathcal{A} = \{a_1, \dots, a_p\}$. Denote \mathbf{D} and \mathbf{J} the matrices in $\mathbb{R}^{m \times np}$ such that:

$$\mathbf{D} = \begin{pmatrix} \mathbf{a} & \mathbf{0}_p & \dots & \mathbf{0}_p \\ \mathbf{0}_p & \mathbf{a} & \dots & \mathbf{0}_p \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_p & \dots & \mathbf{0}_p & \mathbf{a} \end{pmatrix}^T \quad \mathbf{J} = \begin{pmatrix} \mathbf{1}_p & \mathbf{0}_p & \dots & \mathbf{0}_p \\ \mathbf{0}_p & \mathbf{1}_p & \dots & \mathbf{0}_p \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_p & \dots & \mathbf{0}_p & \mathbf{1}_p \end{pmatrix}^T$$

where $\mathbf{a} = (a_1, \dots, a_p)^T$ and $\mathbf{0}_p, \mathbf{1}_p$ are the column vectors of \mathbb{R}^p with respectively zero and one entries.

Finally, denote $\mathcal{S}(\mathbf{y}) := \{\mathbf{s} \in \mathbb{R}^{np} : \Phi\mathbf{D}\mathbf{s} = \mathbf{y} \text{ and } \mathbf{J}\mathbf{s} = \mathbf{1}_n\}$.

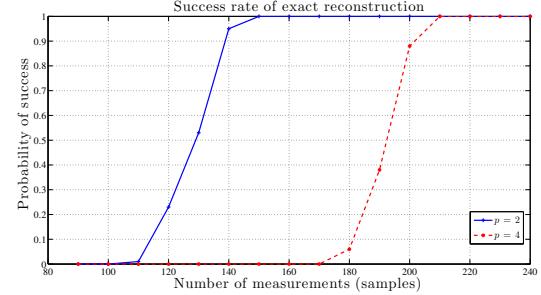
Lemma 2.1: Suppose $\mathbf{y} = \Phi\mathbf{x}$ has a unique solution \mathbf{f} in \mathcal{A}^n . Then, there exists a unique $\hat{\mathbf{s}} \in \mathbb{R}^{np}$ such that

$$\hat{\mathbf{s}} = \arg \min \|\mathbf{s}\|_0 \text{ s.t. } \mathbf{s} \in \mathcal{S}(\mathbf{y})$$

and $\mathbf{f} = \mathbf{D}\hat{\mathbf{s}}$.

As suggested by literature on sparse reconstruction [1], we propose to reconstruct \mathbf{f} from \mathbf{y} by choosing

$$\tilde{\mathbf{s}} = \arg \min \|\mathbf{s}\|_1 \text{ s.t. } \mathbf{s} \in \mathcal{S}(\mathbf{y})$$



and putting $\tilde{\mathbf{f}} = \mathbf{D}\hat{\mathbf{s}}$. The next section shows the performance of our approach for random sensing matrices.

III. EXPERIMENTAL RESULTS

Pertinence of the approach is experimentally assessed as follows: Consider signals with $n = 256$ samples randomly drawn from finite alphabet of cardinality p . The alphabet can be chosen arbitrarily. Given the number m of measurements, we sample the mixing matrix, for each iteration, with independent Gaussian entries and we compare the recovered $\tilde{\mathbf{f}}$ and the original \mathbf{f} signals. The recovery is regarded as successful if the relative error $\|\tilde{\mathbf{f}} - \mathbf{f}\|_2 / \|\mathbf{f}\|_2$ is less than 10^{-5} . For each m , we repeat 100 iterations of the experiment and average the results. The results are presented above for $p = 2$ and $p = 4$. Our experiments show that finite alphabet signals can be accurately recovered provided that the number m of measurements is above $\frac{n(p-1)}{p}$.

IV. DISCUSSIONS

Numerical simulations show that the minimum ℓ_1 norm solution enables recovery of finite alphabet signals for Gaussian matrices when $m > \frac{n(p-1)}{p}$. This result has not yet been proved. Work in progress on this proof involves kernels of random matrices.

The condition $m > \frac{n(p-1)}{p}$ can be rewritten in the form $kp < n$ where k is the number of missing measurements. Since similar recovery conditions can be found in [3], it can be wondered whether some uncertainty principle would not underly the approach proposed above.

On the other hand, the minimization problem in our approach can be related to the ℓ_1 -synthesis described in [2].

REFERENCES

- [1] S. S. Chen, D. L. Donoho and M. A. Saunders, *Atomic Decomposition by Basis Pursuit*, SIAM Review, Volume 43 (2001), 129-159.
- [2] E. J. Candes, Y. C. Eldar, D. Needell and P. Randall, *Compressed Sensing with Coherent and Redundant Dictionaries*, In Press, Applied and Computational Harmonic Analysis.
- [3] D. L. Donoho and P. B. Stark, *Uncertainty principles and signal recovery*, SIAM J. Appl. Math. 49 (1989), 906-931.

Adding Dynamic Smoothing to Mixture Mosaicing Synthesis

Graham Coleman
 Music Technology Group
 Universitat Pompeu Fabra
 Barcelona, Spain
 Email: graham.coleman@upf.edu

Jordi Bonada
 Music Technology Group
 Universitat Pompeu Fabra
 Barcelona, Spain
 Email: jordi.bonada@upf.edu

Esteban Maestre
 Center for Computer Research in Music and Acoustics
 Stanford University
 Stanford, USA
 Email: esteban@ccrma.stanford.edu

Abstract—Recent works in sound mosaicing synthesis [1], [2] have proposed algorithms that permit instantaneous mixtures of several sources atoms, based on sparse signal representation techniques. We propose combining l_1 regularization with linear dynamical smoothing as in the Kalman filter (also in [3], [4]) to promote desired transitions between atoms, while adapting the generic approach to the mixture mosaicing context. Furthermore, we modify the dynamics cost slightly to further promote sparse scores in the case of non-negativity. This is a work in progress in which we can present some sound examples, but for which the proposal is not fully validated.

I. INTRODUCTION

Mosaicing, a form of sample-based sound synthesis, consists in transforming and compositing disparate source sound segments from a database so that the result will match perceptual features (descriptors) of a target sequence. Classical methods [5], [6] considered matching a single source segment to a given target context (frame), while more recent methods [1], [2] consider sparse mixtures of multiple source segments at once.

Several criteria for these systems concern the dynamics—the changes from frame to frame—of the composition, or more abstractly, the sampling process that generates it. Do the descriptors of the source units change too much from frame to frame (continuity)? Do the transformation parameters applied to the sources change rapidly from frame to frame (transformation continuity)? Does the sampling process maintain a steady context within the source material by choosing contiguous blocks of material from the original source context, or does it jump around (contiguity)? By modeling dynamics we can search or sample sequences or mixtures that have desired properties above.

II. PROPOSAL

Given a matrix or linear operator D describing favored atom transitions from state to state, one way of generalizing it to mixtures of atoms is simply considering a form of linear dynamics where: $x_{t+1} = Dx_t + w_t$. x_t and x_{t+1} are mixture vectors for time steps t and $t+1$, w_t represents innovation, or deviance from expected dynamics.

Combining the smoothing version of the Kalman filter with an l_1 regularization term as in Basis Pursuit Denoising (BPDN) would give us the following program:

$$\min_x \sum_{t=1}^T \|Ax_t - b_t\|_2^2 + \lambda_1 \sum_{t=2}^T \|Dx_{t-1} - x_t\| + \lambda_2 \sum_{t=1}^T \|x_t\|_1 \quad (1)$$

Under the scheme given by Problem 2, if the transition matrix D gives a number of possibilities for a given atom, the most likely successor state (where the innovation cost is zero) will include nonzero weights on all of those possibilities. Therefore, when D includes many alternatives for transitions between atoms, the innovation cost and the sparsity cost are working against each other.

In our application, where weights are constrained to be non-negative, we propose modeling *alternatives* using an innovation cost where only positive innovation is penalized, that is having no cost when weights decrease (state is closer to sparsity than deterministic dynamics). We implement this by introducing a non-negative dummy variable y :

$$\min_x \sum_{t=1}^T \|Ax_t - b_t\|_2^2 + \lambda_1 \sum_{t=2}^T \|Dx_{t-1} - x_t - y_t\| + \lambda_2 \sum_{t=1}^T \|x_t\|_1 \quad (2)$$

where both x and y are constrained to be elementwise non-negative. In this scheme, successor states are not penalized for atom transition alternatives not taken, only for unlikely transitions that are taken.

III. OTHER APPROACHES

Several other approaches are also likely feasible for encouraging dynamics in synthesis. For one, we could extend the Kalman filter objective with a nonlinear model, which may render the objective function non-convex. In this case, heuristic methods based on convex relaxation such as DC Algorithms (DCA) could be used to find heuristic solutions quickly.

Sampling or Monte-Carlo approaches are also feasible. In particular, particle filters (Sequential Monte Carlo) have been used for tracking, and allow both nonlinear dynamics, and use non-parametric estimates for the states.

Finally, so called greedy signal decomposition methods could perhaps be adapted to account for dynamics. In practice this would be analogous to sampling in many ways. Perhaps a good example of this in image synthesis would be Ashikhmin [7].

REFERENCES

- [1] M. D. Hoffman, P. R. Cook, and D. M. Blei, “Bayesian spectral matching: Turning young MC into MC hammer via MCMC sampling.”
- [2] G. Coleman, E. Maestre, and J. Bonada, “Augmenting sound mosaicing with Descriptor-Driven transformation,” in *Proceedings of DAFX 2010*, Graz, Austria.
- [3] A. Charles, M. S. Asif, J. Romberg, and C. Rozell, “Sparsity penalties in dynamical system estimation.”
- [4] M. S. Asif, A. Charles, J. Romberg, and C. Rozell, “Estimation and dynamic updating of time-varying signals with sparse variations,” 2011.
- [5] A. Zils and F. Pachet, “Musical mosaicing,” *Digital Audio Effects (DAFx)*, 2001, here are the notes go for annotated bibliographies.
- [6] D. Schwarz, “Data-Driven concatenative sound synthesis,” Ph.D. dissertation, 2004. [Online]. Available: <http://recherche.ircam.fr/equipes/analyse-synthese/schwarz/thesis/>
- [7] M. Ashikhmin, “Synthesizing natural textures,” in *Proceedings of the 2001 symposium on Interactive 3D graphics*, ser. I3D ’01, 2001, p. 217–226, ACM ID: 364405.

Block-Sparse Recovery via Convex Optimization

Ehsan Elhamifar
Johns Hopkins University

René Vidal
Johns Hopkins University

Abstract—We consider the problem of recovering block-sparse signals, i.e. signals that can be written as linear combination of vectors drawn from a union of a few subspaces. To find a block-sparse representation of a signal, we consider two classes of non-convex programs based on minimizing a mixed ℓ_q/ℓ_0 quasi-norm ($q \geq 1$) and their convex ℓ_q/ℓ_1 relaxations. The first class directly penalizes the norm of the coefficient blocks, while the second one penalizes the norm of the reconstructed vectors from the blocks of the dictionary. For each class of convex programs, we provide conditions under which they are equivalent to the original non-convex programs. We apply our methods to classification tasks and obtain significant improvements relative to the state-of-the-art.

I. INTRODUCTION

The recovery of block-sparse signals involves solving a system of linear equations of the form

$$\mathbf{y} = \mathbf{B}\mathbf{c} = [B[1] \quad \dots \quad B[n]] \mathbf{c}, \quad (1)$$

where \mathbf{B} consists of n blocks $B[i] \in \mathbb{R}^{D \times m_i}$ whose atoms are generated by a $d_i \leq m_i$ dimensional subspace \mathcal{S}_i . The main difference with respect to classical sparse recovery is that the nonzero elements for the solution of (1) correspond to a few *blocks* rather than a few *elements* of \mathbf{B} . We say that a vector $\mathbf{c}^\top = [c[1]^\top \quad \dots \quad c[n]^\top]$ is *k-block-sparse*, if at most k blocks $c[i] \in \mathbb{R}^{m_i}$ are nonzero.

The problem of finding a representation of a signal \mathbf{y} that uses the minimum number of blocks of \mathbf{B} can be cast as

$$P_{\ell_q/\ell_0} : \min \sum_{i=1}^n I(\|\mathbf{c}[i]\|_q > 0) \quad \text{subj. } \mathbf{y} = \mathbf{B}\mathbf{c}, \quad (2)$$

where $I(\cdot)$ is the indicator function and $q \geq 0$. Since (2) is an NP-hard problem, we consider the following ℓ_1 relaxation of P_{ℓ_q/ℓ_0}

$$P_{\ell_q/\ell_1} : \min \sum_{i=1}^n \|\mathbf{c}[i]\|_q \quad \text{subj. } \mathbf{y} = \mathbf{B}\mathbf{c}, \quad (3)$$

which is a convex program for $q \geq 1$. We also propose an alternative approach based on solving the non-convex program for $q \geq 0$

$$P'_{\ell_q/\ell_0} : \min \sum_{i=1}^n I(\|\mathbf{B}[i]\mathbf{c}[i]\|_q > 0) \quad \text{subj. } \mathbf{y} = \mathbf{B}\mathbf{c}. \quad (4)$$

While P_{ℓ_q/ℓ_0} penalizes the norm of the coefficient blocks, P'_{ℓ_q/ℓ_0} penalizes the norm of the reconstructed vectors from the blocks. Since P'_{ℓ_q/ℓ_0} is NP-hard, for $q \geq 1$, we consider the ℓ_1 convex relaxation

$$P'_{\ell_q/\ell_1} : \min \sum_{i=1}^n \|\mathbf{B}[i]\mathbf{c}[i]\|_q \quad \text{subj. } \mathbf{y} = \mathbf{B}\mathbf{c}. \quad (5)$$

In what follows, we derive conditions under which the convex programs P_{ℓ_q/ℓ_1} and P'_{ℓ_q/ℓ_1} , respectively, are equivalent to P_{ℓ_q/ℓ_0} and P'_{ℓ_q/ℓ_0} for arbitrary $q \geq 1$. In doing so, we allow for an arbitrary number of atoms in each block of the dictionary, thus relaxing the assumption of uniqueness of the representation made by state-of-the-art methods, which restrict the blocks of a dictionary to have linearly independent atoms. To characterize the relation between blocks of a dictionary, we introduce the notion of mutual/cumulative subspace coherence, which can be thought of as natural extensions of mutual/cumulative coherence from one to multiple subspaces.

Definition 1: Mutual subspace coherence is defined as

$$\mu_S \triangleq \max_{i \neq j} \mu(\mathcal{S}_i, \mathcal{S}_j), \quad (6)$$

where $\mu(\mathcal{S}_i, \mathcal{S}_j)$ is the cosine of the smallest principal angle between subspaces \mathcal{S}_i and \mathcal{S}_j . *k-cumulative subspace coherence* is defined as

$$\zeta_k \triangleq \max_{\Lambda_k} \max_{i \notin \Lambda_k} \sum_{j \in \Lambda_k} \mu(\mathcal{S}_i, \mathcal{S}_j), \quad (7)$$

where Λ_k is a subset of k different elements from $\{1, \dots, n\}$.

To characterize the relation among atoms of a dictionary, we define the following notions.

Definition 2: For a dictionary \mathbf{B} , we define ϵ_q as the smallest constant such that for all i there exists a full column-rank submatrix $\bar{\mathbf{B}}[i] \in \mathbb{R}^{D \times d_i}$ of $\mathbf{B}[i]$ such that for all $\bar{\mathbf{c}}[i]$, we have

$$(1 - \epsilon_q) \|\bar{\mathbf{c}}[i]\|_q^2 \leq \|\bar{\mathbf{B}}[i]\bar{\mathbf{c}}[i]\|_2^2 \leq (1 + \epsilon_q) \|\bar{\mathbf{c}}[i]\|_q^2. \quad (8)$$

Define σ_q as the smallest constant such that for all i and $\mathbf{c}[i]$

$$\|\mathbf{B}[i]\mathbf{c}[i]\|_2^2 \leq \sigma_q \|\mathbf{c}[i]\|_q^2. \quad (9)$$

Roughly speaking, ϵ_q characterizes the best q -restricted isometry property among all submatrices of $\mathbf{B}[i]$ that span subspace \mathcal{S}_i . Also, from definition, we have $1 + \epsilon_q \leq \sigma_q$. We have the following result.

Theorem 1: For a signal that has a k -block-sparse representation in \mathbf{B} , the solution of P_{ℓ_q/ℓ_1} is equivalent to that of P_{ℓ_q/ℓ_0} , if

$$\sqrt{\sigma_q/(1 + \epsilon_q)} \zeta_k + \zeta_{k-1} < (1 - \epsilon_q)/(1 + \epsilon_q). \quad (10)$$

A stronger sufficient condition is given by

$$(k \sqrt{\sigma_q/(1 + \epsilon_q)} + k - 1) \mu_S < (1 - \epsilon_q)/(1 + \epsilon_q). \quad (11)$$

Similar conditions can be found for the equivalence between P'_{ℓ_q/ℓ_1} and P'_{ℓ_q/ℓ_0} for arbitrary $q \geq 1$ and for dictionaries with arbitrary number of atoms in each block. We refer the reader to [1] for details.

II. APPLICATION TO FACE CLASSIFICATION

Assume we have a training set from P classes, where each class consists of the data drawn from a few subspaces. That is,

$$\mathbf{B} = [\underbrace{\mathbf{B}[1] \quad \mathbf{B}[2] \quad \mathbf{B}[3]}_{\text{Class 1}} | \dots | \underbrace{\mathbf{B}[n-1] \quad \mathbf{B}[n]}_{\text{Class } P}] \quad (12)$$

Given a test data \mathbf{y} that belongs to one of the classes, the goal is to find the class to which the test example belongs. Since each class consists of a few blocks of the dictionary, the class of the test example can be obtained by finding the minimum number of blocks that reconstruct \mathbf{y} . Thus, the classification problem can be cast as a block-sparse recovery problem. We applied the proposed convex optimization programs for face classification on the Extended YaleB Database, which consists of a total of 2432 face images for 38 individuals corresponding to $P = 38$ classes. Our results improve the state-of-the-art classification results by 10% on the dataset.

REFERENCES

- [1] E. Elhamifar and R. Vidal, “Robust classification via structured sparse representation,” in *CVPR*, 2011.

Performance Limits of the Measurements on Compressive Sensing for Multiple Sensor System

Sangjun Park, Hwanchol Jang and Heung-No Lee*

Gwangju Institute of Science and Technology, South Korea. Email: {sjpark1,hcjang,heungno*}@gist.ac.kr

Abstract—A performance analysis of Multiple-Sensor-System(MSS) on a compressive sensing(CS)[1] w.r.t. the per-sensor-measurements(PSM) is studied. In the proposed MSS, sensors make measurements using CS and the decoder jointly recover signals from them. We obtain the upper bound on the recovery failure probability for given K -sparse signals, derive the relationship between PSM and the number of sensors(S) for the recovery. We examine the effect of SNR and S for the recovery. We use the concept of joint typicality proposed by Shannon[6]. We shows that PSM converges to the sparsity(K) as S increases for given K -sparse signals. Theoretical result is consistent with [3][4][5].

Index Terms—Compressive Sensing, Multiple Sensor System, Joint Recovery.

I. INTRODUCTION AND MOTIVATION

Multiple-Sensor-System (MSS) deploys many sensors to a limited region and uses them to measure the signal from a common information source in different locations. In MSS, high resolution signal can be obtained as many sensors are used to measure a common phenomenon from many places. However, the coverage areas of sensors may significantly overlap with each other as they are distributed in a limited region. This causes redundancy in the measurement signal. The transmission of the redundant signal to the fusion center is a significant communication costs. There is tradeoff between the resolution and the redundancy on the number of sensors. To work on this tradeoff relationship, we use the idea of the compressive sensing [1]. CS reduces the number of measurements while it recovers the signal perfectly. Using this technique, it is possible to reduce the redundancy and obtain high resolution simultaneously by reducing the per-sensor-measurements (PSM).

To investigate our problem, we propose to use an information theoretic tool, the concept of Jointly Typicality [6]. It was also used by Akcakaya and Tarokh [2] for the single sensor case. Using this tool, we can derive the upper bound on the failure probability as a function of PSM, the number of sensors, the sparsity and the noise variance.

Clearly, the MSS problem is different from a single sensor system in many aspects. For an appropriate modification of the tool for MSS problem, we should consider these differences. One big difference is the signal correlation among the sensors. For a successful extension, we use the inter-signal correlation in the system model and the decoder also takes advantage of this signal correlation for a signal recovery. To make the correlation model, we assume that each sensor has the same sparsity and shares the same support set which is the set of indices for the non-zero elements. Obviously, in the recovery, the decoder using this prior information gains benefits.

II. THEOREMS

Theorem 1: Let the rank of $\mathbf{F}_{s,J}$ be K for each s and J be any candidate set, $M > K$, $\sigma^2 = \min(\sum_{i \in I \setminus J} x_s(i)^2)$ over s , and $\delta > 0$. Then, $P\{\text{Fail}|\mathbf{x}\}$ converges to zero as the number of sensors increases.

Theorem 2: Let the rank of $\mathbf{F}_{s,J}$ be K for each s and J be any candidate set, $M > K$, $\sigma^2 = \min(\sum_{i \in I \setminus J} x_s(i)^2)$ over s , $\delta > 0$,

S_i be the number of sensors of the i^{th} MSS, σ_i^2 be the noise variance of the i^{th} MSS and $P_1\{\text{Fail}|\mathbf{x}\} \leq \gamma$. If the noise variance increases, i.e., $\sigma_1^2 < \sigma_2^2$, then, the sufficient condition for $P_2\{\text{Fail}|\mathbf{x}\} \leq \gamma$ is

$$S_2 \geq S_1 \max \left(\frac{f\left(\frac{\delta}{\sigma_1^2} \frac{M}{M-K}\right)}{f\left(\frac{\delta}{\sigma_2^2} \frac{M}{M-K}\right)}, \frac{g\left(\frac{\sigma_1^2}{\sigma_{\min,1}^2} + \frac{\delta}{\sigma_{\min,1}^2} \frac{M}{M-K}\right)}{g\left(\frac{\sigma_2^2}{\sigma_{\min,2}^2} + \frac{\delta}{\sigma_{\min,2}^2} \frac{M}{M-K}\right)} \right). \quad (1)$$

We note that $f(x) = \log(1+x) - x$, $g(x) = \log(x) - x + 1$, $\sigma_{\min,i}^2 \equiv \min(\sum_{j \in I \setminus J} x_s(j)^2) + \sigma_i^2$ over s and J , J denotes any subset with size K expect for I and I denotes the set whose entries are corresponding to indices of the nonzero elements in signal. All theorems will be explained in the next section.

III. CONTRIBUTIONS AND CONCLUSIONS

We use the described correlation model with noisy observation. First, we have found how many per-sensor-measurements (PSM) are needed for successful recovery in the MSS problem. As the number of sensors increases, how does PSM change? There is a limit we have found. We will show this behavior and will show how PSM depends on the sparsity. We have Theorem 1 that the infimum of PSM is the sparsity obtained as the number of sensors increases. Different from the results in [3], [4], [5], the work of ours gives analytical results. Our analysis works for a small number of sensors as well. Second, we have shown that the decoder which uses the prior information obtains benefit in terms of the Signal to Noise Ratio (SNR). Specifically, Theorem 2 tells us how the required SNR decreases as the number of sensors changes.

IV. ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (Do-Yak Research Program, NO. 2010-0017944).

REFERENCES

- [1] Dvaid L. Donoho, "Compressive sensing," *IEEE Trans. On Information Theory*, vol. 52, pp. 1289-1306, 2006.
- [2] Mehmet Akcakaya and Vahid Tarokh, "Shannon-Theoretic Limits on Noisy Compressive Sampling", *IEEE Trans. On Information Theory*, vol. 56, 2010.
- [3] S. Sarvotham et. al, "Analysis of the DCS one-stage greedy algorithm for common sparse supports, Technical Report TREE-0503: Rice University, Department of Electrical and Computer Engineering," Oct. 2005.
- [4] D.Baron et. al "Distributed compressive sensing," 2009.
- [5] Pablo Vinuelas-Peris and Antonio Artes-Rodriguez, "Bayesian Joint Recovery of Correlated Signals in Distributed Compressed Sensing," 2010 2nd International Workshop on Cognitive Information Processing
- [6] Thomas M. Cover and Joy A. Thomas, "Elements of Information Theory second Edition," 2006

Message Passing Aided Least Square Recovery for Compressive Sensing

Jaewook Kang, Heung-No Lee and Kiseon Kim

Gwangju Institute of Science and Technology, South Korea. Email: {jwkkang,heungno,kskim}@gist.ac.kr

I. INTRODUCTION AND MOTIVATION

Compressive sensing (CS) have got attention as a promising signal processing technique to reduce information rate of sparse signals [1]. One line of CS related researches are to devise low complexity recovery algorithms since the conventional L1-norm based recovery algorithms still have high computational complexity for practical applications. Recently, a few researchers have made an attempt to apply probabilistic message passing (PMP) ideas to CS recovery [2], [3] since PMP has provided a successful solution for low complexity decoding while showing suboptimal performance in channel coding problems, such as low-density parity check codes [4].

Motivated by such previous works, in this paper, we propose a new least square estimation (LSE) based CS recovery algorithm by applying PMP, called PMP-LSE. It is well known that CS recovery is basically an underdetermined system and it can be reformed as an overdetermined system with the support set information (SI). Therefore, in the proposed algorithm, PMP undertakes to find the SI of the signal to reform the recovery to an overdetermined case, and then LSE completes the recovery using the SI. Mainly, PMP-LSE has two strong benefits. First, PMP-LSE shows outstanding performance with noisy measurements by removing the noise effect from elements belonging to the non-support set. Second, PMP-LSE prevents the recovery from diverging. Under certain conditions, PMP based algorithms fails in the recovery due to divergence caused by a large number of iterations. In the algorithm, however, the possibility of the divergence highly decreases since PMP is only used to search the SI with a few iterations.

II. PROBLEM SETUP

We consider a sparse signal $\mathbf{x} \in \mathbb{R}^N$ whose sparsity is characterized by q , named sparsity rate. With the sparsity rate q , each element of \mathbf{x} belongs to the support set denoted by \mathbf{S} . Hence, $|\mathbf{S}|$ corresponds to Binomial random variable with $B(N, q)$. Let $\mathbf{x}_S \in \mathbb{R}^{|\mathbf{S}|}$ denote a vector consisting of nonzero elements belonging to \mathbf{S} , and assume that each element of \mathbf{x}_S follows Gaussian distribution with $N(0, \sigma_x^2)$. We also assume that the sensing matrix is a well-designed binary matrix, i.e., $\Phi \in \{0, 1\}^{M \times N}$, according to [5] such that the measurements $\mathbf{y} \in \mathbb{R}^M$ are generated by $\mathbf{y} = \Phi \mathbf{x}$. Then, noisy measurements $\mathbf{z} \in \mathbb{R}^M$ at the decoder are described as $\mathbf{z} = \mathbf{y} + \mathbf{n}$, where each element of $\mathbf{n} \in \mathbb{R}^M$ is Gaussian noise with $N(0, \sigma_n^2)$.

III. ALGORITHM

The algorithm is divided into two parts: PMP and LSE.

i) **PMP:** PMP consists of two kinds of probability calculations based on Bayesian rule: Variable to check message (VCM, $\mathbf{v}_{i \rightarrow j}$) and check to variable message (CVM, $\mathbf{c}_{j \rightarrow i}$) calculation where i and j indicate the index of elements of \mathbf{x} and \mathbf{z} , respectively.

$$\text{VCM} : \mathbf{v}_{i \rightarrow j}^l := p\{\mathbf{x}_i | \mathbf{z}\} = C^l p\{\mathbf{x}_i\} \times \prod_{k: \phi_{k,i}=1, k \neq i} \mathbf{c}_{k \rightarrow i}^{l-1} \quad (1)$$

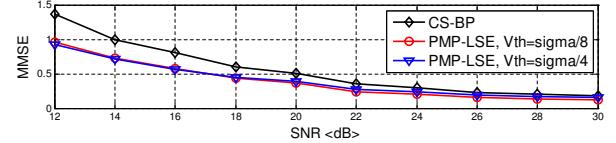


Fig. 1. MMSE performance of PMP-LSE ($N=100, M=80, q=0.1, N_{iter}=3$)

$$\begin{aligned} \text{CVM} : \mathbf{c}_{j \rightarrow i}^l &:= P\{z_j | x_i\} = P\{z_j - \sum_{k: \phi_{jk}=1, k \neq i} x_k + x_i | x_i\} \\ &= p\{z_j | \text{any } x_k : \phi_{jk} = 1\} * \mathbf{v}_{k_1 \rightarrow j}^l * \dots * \mathbf{v}_{k_{L_j-1} \rightarrow j}^l \end{aligned} \quad (2)$$

Here, l is the number of iteration, ϕ_{ji} is the (j, i) th element of Φ , L_j is the number of ones in j th row of Φ , and C^l is the normalization constant for l th VCM. And, $*$ indicates the convolution operation. At each iteration, PMP updates VCM and CVM by exchanging the probabilistic messages among the elements of \mathbf{x} and \mathbf{z} . After a few iterations, PMP distinguish the elements of the support set with a certain threshold denoted by V_{th} .

ii) **LSE:** Once the SI is given, \mathbf{x}_S is easily estimated only using the corresponding columns of Φ , denoted by Φ_S , i.e., $\mathbf{x}_S = (\Phi_S^T \Phi_S)^{-1} \Phi_S^T \mathbf{z}$. By combining the SI and \mathbf{x}_S , PMP-LSE completes to find the recovered signal $\hat{\mathbf{x}}$.

IV. NUMERICAL RESULTS

To demonstrate the performance, we simulated PMP-LSE with CS-BP [2]. Figure 1 plots the MMSE per elements as a function of SNR for variety of thresholds with three PMP iterations. Figure 1 shows that PMP-LSE outperforms CS-BP notably in low SNR region. The reason is that PMP-LSE prevents the corruption of zero elements from the noise effect by pre-detecting the support set using PMP.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MEST) (Haek Sim Research Program, NO. 2010-0026407)

REFERENCES

- [1] D.L. Donoho, "Compressive sensing," *IEEE Trans. On Information Theory*, vol. 52, pp. 1289-1306, 2006.
- [2] D. Baron, S. Sarvotham, and R. Baraniuk, "Bayesian Compressive Sensing via Belief Propagation," *IEEE Trans. Signal Process.*, Vol. 58, No. 1, pp. 269-280, Jan. 2010.
- [3] D.L. Donoho, A. Maleki and A. Montanari, "Message passing algorithms for compressed sensing: I. Motivation and construction," *Proc. IEEE ITW*, Cairo, Egypt, Jan. 2010.
- [4] R.G. Gallager, *Low-Density Parity Check Codes*, MIT Press: Cambridge, MA, 1963.
- [5] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity approaching irregular low-density parity check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619-637, Feb. 2001.

Matrix-free Interior Point Method for Compressed Sensing Problems

Kimonas Fountoulakis

School of Mathematics and Maxwell Institute
 The University of Edinburgh
 Mayfield Road, Edinburgh EH9 3JZ
 United Kingdom.
 Email: kf09@doc.ic.ac.uk

Jacek Gondzio

School of Mathematics and Maxwell Institute
 The University of Edinburgh
 Mayfield Road, Edinburgh EH9 3JZ
 United Kingdom.
 Email: J.Gondzio@ed.ac.uk

Abstract—We consider a class of optimization problems for sparse signal reconstruction which arise in the field of Compressed Sensing (CS). A plethora of approaches and solvers exist for such problems, for example GPRS, SparseLab, $\ell_1\text{-}\ell_s$, ℓ_1 magic, FPC_AS to mention a few.

Compressed Sensing applications lead to very well conditioned optimization problems and therefore can be solved easily by simple first-order methods. In this work we demonstrate that a second-order method such as an interior point algorithm can be specialized for the CS problems and offer a competitive alternative to the existing approaches. The new approach is based on the *Matrix-free Interior Point Method* [1] in which an iterative (Krylov-subspace) method is employed to compute an inexact Newton direction. The matrix-free IPM does not require an explicit storage of the constraint matrix but accesses it only to get the matrix-vector products. It is therefore well-suited for solving large scale problems because it can take full advantage of the low-complexity fast matrix-vector operations. A partial Cholesky preconditioner is employed to accelerate the convergence of the Krylov-subspace method. The computation of the preconditioner requires only matrix-vector products and fits into the matrix-free regime. Computational experience on the medium scale one-dimensional signals ($n = 4096$) confirms that the new approach is efficient and compares favourably with other state-of-the-art solvers.

I. INTRODUCTION

Interior point methods (IPMs) for linear and convex quadratic programming enjoy an unequalled worst-case complexity result. Indeed, they deliver a ε -accurate solution to such problems in $\mathcal{O}(\sqrt{n} \ln(1/\varepsilon))$ iterations, where n is the problem dimension. IPMs are usually applied to solve problems to a high degree of accuracy (small ε , say, 10^{-8}). However, it is straightforward to specialize them to work in a significantly less demanding environment such as that of Compressed Sensing where accuracy of $\varepsilon = 10^{-1}$ is often all the user wishes for.

The optimization problems arising in Compressed Sensing applications are very well-conditioned and therefore trivial from optimization point of view. This explains why the simple approaches based on projected gradient can solve these problems so efficiently. In this short note we argue that a specialized Interior Point Method (IPM), implemented in the HOPDM solver, can offer a competitive approach for optimization problems arising in Compressed Sensing applications.

II. FUNDAMENTALS OF MATRIX-FREE IPM

The interior point solver for convex quadratic programming needs to solve a particular weighted least-squares problem at each iteration. This is usually done by a direct approach which is based on the Cholesky factorization. In a relaxed environment in which only an approximate solution of the problem is requested it is advantageous to employ an iterative method such as for example the conjugate gradient algorithm to solve the underlying system of linear equations. An *exact* Newton method employed in the standard IPM is then replaced with an *inexact* one [1].

TABLE I: Comparison table

Solver	R. Gaussian.		Orth. R. Gaussian.		Part. Hadamard	
			CPU times (sec)			
	CPU	MSE	CPU	MSE	CPU	MSE
HOPDM	8.50	7.5e-3	12.36	1.5e-2	8.61	1.1e-2
GPSR 6	1.61	4.0e-4	1.15	1.9e-4	1.08	1.7e-4
FPC_AS	0.53	7.3e-4	0.75	1.3e-4	0.75	1.1e-4
PDCO	10.29	1.8e-2	11.72	8.2e-2	10.04	7.6e-3
$\ell_1\text{-}\ell_s$	8.60	4.0e-4	5.50	1.9e-4	4.47	1.7e-4

To achieve fast convergence of the iterative algorithm one needs to use a suitable preconditioner for the linear system. The preconditioner in the Matrix-Free IPM [2] is constructed in two steps. Firstly, the linear algebra subproblem is regularized and secondly a low-rank partial Cholesky factorization for this regularized system is computed. The process of computing the preconditioner does not require explicit access to the Jacobian matrix: only matrix-vector multiplications with this operator are needed and therefore the approach can take full advantage of the low-complexity matrix-vector operations which rely on the low-parametric representation of the Jacobian matrices.

III. COMPUTATIONAL EXPERIENCE

We consider CS problems analogous to the ones in [3]. The reconstruction of the sparse signals is achieved via random Gaussian (R. Gaussian.), orthonormalized random Gaussian (Orth. R. Gaussian.) and partial Hadamard (Part. Hadamard.) $m \times n$ sensing matrices.

The table (I) shows the computational time and the mean squared error ($MSE = \frac{\|\hat{x}-x\|^2}{n}$) of the reconstructed signals for the **matrix-free HOPDM** and some of the existing state-of-the-art solvers: **GPRS**, **FPC_AS**, **SparseLab (PDCO)**, $\ell_1\text{-}\ell_s$.

In the case of table (I), sparse signals $n = 4096$ with 160 randomly placed spikes were generated. The sensing matrices: Gaussian, orthonormalized Gaussian and partial Hadamard with $m = 1024$ and $n = 4096$ were used. Moreover, white noise with $\sigma^2 = 10^{-4}$ is added in the sampled signal and finally, the optimality tolerance for the termination criteria of each solver has been set to 10^{-2} .

REFERENCES

- [1] R. S. Dembo and S. C. Eisenstat and T. Steihaug, *Inexact Newton Methods*, SIAM Journal on Numerical Analysis 19 (1982) 400–408.
- [2] J. Gondzio, *Matrix-Free Interior Point Method*, Computational Optimization and Applications, Published online: October 15, 2010. DOI: 10.1007/s10589-010-9361-3
- [3] S. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinvesky, *A method for large-scale ℓ_1 -regularized least squares problems with applications in signal processing and statistics*, Tech. Report, Dept. of Electrical Engineering, Stanford University, 2007. Available at www.stanford.edu/~boyd/l1_ls.html

A Block-Based Approach to Adaptively Bias the Weights of Adaptive Filters

Luis A. Azpicueta-Ruiz

Dept. of Signal Theory and Communications

Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain.

Email: azpicueta@tsc.uc3m.es

Jerónimo Arenas-García

Dept. of Signal Theory and Communications

Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain.

Email: jarenas@tsc.uc3m.es

I. EXTENDED ABSTRACT

Adaptive filters are crucial in many signal processing applications. Recently, a simple configuration was presented to introduce a bias in the estimation of adaptive filters using an adaptively adjusted multiplicative factor $\alpha(n)$, showing important gains in terms of mean square error with respect to standard adaptive filter operation, mainly for low signal to noise ratios (see [1]).

In this paper, we modify that scheme to obtain further advantages by splitting the adaptive filter coefficients into M non-overlapping blocks, and employing a different scaling factor α_m with $m = 1, \dots, M$ for the coefficients in each block. In this way, bias vs variance compromise is managed independently in each block, allowing an enhancement if the energy of the unknown system is non-uniformly distributed, as it is the case of sparse identification.

The proposed scheme is based on a blockwise decomposition of an adaptive filter $\mathbf{w}(n)$ with length N in blocks of $P = N/M$ coefficients each. In order to implement a scheme able to selectively bias certain blocks of coefficients, we can multiply the corresponding coefficients by a shrinkage factor α_m to be adjusted according to mean squared error (MSE) performance. The output of the 'block-biased' scheme can then be obtained as

$$y_{BB}(n) = \sum_{m=1}^M \alpha_m \mathbf{w}_m^T(n) \mathbf{u}_m(n) = \sum_{m=1}^M \alpha_m y_m(n). \quad (1)$$

where $\mathbf{u}_m(n)$ and $\mathbf{w}_m(n)$ are the blocks of input vector and adaptive filter necessary to obtain the partial output $y_m(n)$ (see Fig. 1).

Note that for $M = 1$ this scheme is equivalent to that of [1]. However, when identifying an optimal solution $\mathbf{w}_o(n)$ with sparse structure under white noise conditions, the apparent SNR affecting filter weights will depend on each coefficient absolute value, thus justifying the use of different multiplicative factors α_m . As explained in [1] using $\alpha_m < 1$ decreases the variance of the estimation of the optimal solution coefficients in exchange of an increased bias, making it possible to reduce the overall MSE of the filter.

In this paper, a steady-state analysis is developed with the aim of finding the optimal scaling factors α_m^* with $m = 1, \dots, M$ that minimize the steady-state MSE of the proposed configuration. The conclusions of this analysis are that:

$$\alpha_m^* = \frac{1}{1 + \frac{\mathbb{E}\{\|\epsilon_m(\infty)\|_2^2\}}{\|\mathbf{w}_{o,m}\|_2^2}} \quad (2)$$

and the minimum mean square error of the proposed configuration for block-biased adaptive filtering will be

$$J_{ex,BB}^*(\infty) = \sigma_u^2 \sum_{m=1}^M \alpha_m^* \mathbb{E}\{\|\epsilon(\infty)\|_2^2\}, \quad (3)$$

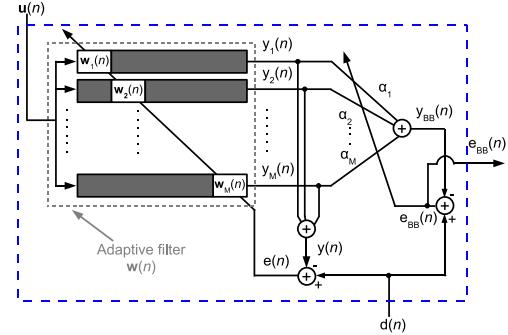


Fig. 1. Block diagram of the proposed scheme. Note that we could employ any kind of transversal adaptive filter.

which is less or equal than the MSE of the original unbiased filter, with the equality holding for $\alpha_m^* = 1 \forall m$.

In the paper we also present a practical algorithm for learning and adapting the value of the scaling factors, since adaptive $\alpha_m(n)$ learning rules are in general necessary to adapt to possible time-varying optimum solution \mathbf{w}_o or SNRs. Following [1], adaptive scheme from [2] (originally proposed for adaptive filter combinations) will be employed, where a normalized stochastic gradient algorithm is followed in order to minimize the power of the whole error of the scheme, i.e. $e_{BB}(n) = d(n) - y_{BB}(n)$.

Whole paper includes a set of experiments to compare the steady-state performance of our proposal and the optimum values resulted from the analysis. This comparison shows that the proposed scheme is able to approximate both the optimal steady-state value of mixing parameters and the optimal EMSE(∞). In addition, influence of number of blocks M is studied, since its adjustment imposes a compromise involving the gains of our proposal with respect to operation of a single adaptive filter, and computational cost. Experimental evaluation concludes with a study of the convergence properties of our scheme, showing an appropriate performance and reconvergence ability when SNR or unknown impulse response suddenly changes, without any *a priori* information about filtering scenario.

REFERENCES

- [1] M. Lázaro-Gredilla, L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, A. R. and J. Arenas-García, "Adaptively Biasing the Weights of Adaptive Filters," *IEEE Trans. Signal Process.*, vol. 58, pp. 3890–3895, Jul. 2010.
- [2] L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal and J. Arenas-García, "A normalized Adaptation Scheme for the convex combination of two adaptive filters," in *Proc. Intl. Conf. Acoustic, Speech, and Signal Process. (ICASSP)*, Las Vegas, NV, Apr. 2008, pp. 3301–3304.

Index

- Aïssa-El-Bey, Abdeldjalil, 120
Adcock, Ben, 63
Adler, Amir, 77
Afonso, Manya, 43
Amel, Roi, 113
Anthoine, Sandrine, 25, 28
Antoine, Philippe, 114
Arberet, Simon, 83
Arenas-García, Jerónimo, 126
Arridge, Simon, 52
Asl, Hojjat Akhondi, 47
Astola, Jaakko, 85
Atto, Abdourrahmane, 60
Aujol, Jean François, 25
Azpícueta-Ruiz, Luis, 126

Bach, Francis, 13
Bah, Bubacarr, 51
Bamler, Richard, 24
Baraniuk, Richard, 41, 75
Bastida-Jumilla, Consuelo, 111
Becker, Stephen, 71
Benichoux, Alexis, 84
Betcke, Marta, 52
Bioucas-Dias, José, 43, 96
Blanchard, Jeffrey, 78
Blu, Thierry, 62
Bonada, Jordi, 121
Boursier, Yannick, 25
Brady, David, 12
Breitenreicher, Dirk, 100

Candès, Emmanuel, 71
Cartis, Coralia, 65
Casazza, Peter, 19
Cevher, Volkan, 66, 68
Chakrabarty, Dalia, 98
Charbonnier, Camille, 36
Chardon, Gilles, 95
Charles, Adam, 48
Chen, Guangliang, 33
Chen, Hung-Wei, 110
Chesneau, Christophe, 56
Chiquet, Julien, 36
Christensen, Mads, 88
Clothilde, Mélot, 25
Coleman, Graham, 121

Dai, Wei, 46

Damjanovic, Ivan, 76
Danielyan, Aram, 90
Daudet, Laurent, 89, 95, 115
Davenport, Mark, 61
Davies, Matthew, 76
Davies, Mike E., 23, 27, 44, 54, 72, 74
De Vos, Maarten, 97
Demaret, Laurent, 40
Donoho, David, 15
Dossal, Charles, 56
Drémeau, Angélique, 89
Dragotti, Pier Luigi, 47, 62
Du, Chaoran, 23
Duarte, Marco, 41
Dumitrescu, Bogdan, 108
Dyer, Eva, 75

Eftekhari, Armin, 39
Egiazarian, Karen, 90
Elad, Michael, 34, 44, 73, 77
Eldar, Yonina C., 34
Elhamifar, Ehsan, 20, 31, 122
Emami-Neyestanak, Azita, 71
Emiya, Valentin, 77

Fadili, Jalal, 49, 56
Faktor, Tomer, 34
Favarro, Paolo, 32
Feuer, Arie, 113
Figueiredo, Mário, 43, 96
Foumouo, Emmanuel, 114
Fountoulakis, Kimonas, 125

Gan, Lu, 92, 101
Gao, Su, 101
Gemmeke, Jort, 86
Gittens, Alex, 37
Golbabaei, Mohammad, 57
Gondzio, Jacek, 125
Gong, Wenlin, 91
González, Adriana, 114
Gowreesunker, B. Vikrham, 70
Goyal, Vivek, 58
Grandvalet, Yves, 36
Gretsistas, Aris, 35
Gribonval, Rémi, 14, 44, 50, 74, 77, 83, 84

Halko, Nathan, 17
Hammond, David, 49

- Han, Shensheng, 91, 109
 Hansen, Anders, 63
 Heide Joergensen, Jakob, 26
 Hein, Matthias, 30
 Heinecke, Andreas, 19
 Huang, Shisheng, 21
 Huang, Tianyao, 112
 Ikeda, Shiro, 106
 Jacques, Laurent, 49, 114
 Jafari, Maria G., 77
 James, Oliver, 105
 Jang, Hwanchol, 123
 Jensen, Tobias Lindstrøm, 64
 Journee, Michel, 81
 Kachour, Maher, 56
 Kamilov, Ulugbek, 58
 Kang, Jaewook, 124
 Katkovnik, Vladimir, 85, 90
 Kelly, Kevin, 103
 Kim, Kiseon, 124
 Kim, Youngchun, 70
 Kingsbury, Nick, 42
 Kleinstuber, Martin, 80
 Kolobov, Mikhail I., 109
 Kono, Hidetoshi, 106
 Kowalski, Matthieu, 117
 Krahmer, Felix, 19
 Kutyniok, Gitta, 19
 Kyrrlidis, Anastasios, 66
 Lahdil, Hassan, 87
 Laligant, Olivier, 93
 Lazzaro, Damiana, 116
 Lee, Heung-No, 104, 105, 123, 124
 Lee, Seungchan, 104
 Lellmann, Jan, 100
 Lexa, Michael, 72
 Li, Kezhi, 101
 Ling, Cong, 101
 Liu, Yimin, 112
 Loh, Mathew, 71
 Lorenz, Dirk, 118
 Lu, Chun-Shien, 110
 Ma, Yi, 10
 Maestre, Esteban, 121
 Maggioni, Mauro, 33
 Mailhé, Boris, 67
 Mantzel, William, 38
 Marshall, Ian, 27
 Martinsson, Per-Gunnar, 17
 Matic, Vladimir, 97
 McCoy, Michael, 79
 McEwen, Jason, 45
 Menchón-Lara, Rosa-María, 111
 Meng, Huadong, 112
 Mijovic, Bogdan, 97
 Milenkovic, Olgica, 46
 Millioz, Fabien, 54
 Mohammad-Djafari, Ali, 107
 Moussallam, Manuel, 115
 Mulgrew, Bernard, 23
 Mustiere, Frederic, 87
 Najaf-Zadeh, Hossein, 87
 Nam, Sangnam, 44, 74
 Nesterov, Yurii, 81
 Newman, Nathan, 92
 Nowak, Robert, 42
 O'Hanlon, Ken, 69
 Olhede, Sofia, 59
 Pan, Xiaochuan, 26
 Park, Jae Young, 39
 Park, Sangjun, 123
 Pastor, Dominique, 60, 120
 Pei, Soo-Chang, 110
 Petra, Stefania, 100
 Peyré, Gabriel, 56
 Pfetsch, Marc, 118
 Pichevar, Ramin, 87
 Plumbley, Mark, 35, 67, 69, 76
 Prünste, Ludger, 22
 Puy, Gilles, 45, 50
 Rabeson, Hérald, 99
 Rangan, Sundeep, 58
 Rao, Nikhil, 42
 Ravichandran, Avinash, 32
 Richard, Gael, 115
 Richtárik, Peter, 29, 81
 Rigat, Fabio, 98
 Rilling, Gabriel, 23, 27
 Rodet, Thomas, 117
 Romberg, Justin, 38
 Rozell, Christopher, 39, 48, 53, 82
 Rubinstein, Ron, 73
 Sánchez, Juan Morales, 111
 Sancho-Gómez, José-Luis, 111
 Sankaranarayanan, Aswin, 75
 Sastry, Shankar, 20
 Sbaï, Si Mohamed Aziz, 120
 Schnör, Christoph, 100
 Schniter, Philip, 68
 Sejdinovic, Dino, 46
 Sepulchre, Rodolphe, 81
 She, Yaochun, 92
 Shen, Hao, 80, 92
 Shin, Younghak, 104
 Sidky, Emil Y., 26
 Singaraju, Dheeraj, 20

Slawski, Martin, 30
Srinivasa, Christopher, 87
Stolz, Christophe, 93
Sturm, Bob, 88
Sudhakar, Prasad, 83, 84

Takac, Martin, 29
Tanner, Jared, 51, 78
Tao, Yuehui, 27
Tewfik, Ahmed H., 70
Thiran, Jean Philippe, 45
Thompson, Andrew, 65
Thompson, John, 72
Tillmann, Andreas, 118
Torresani, Bruno, 28
Tron, Roberto, 20
Tropp, Joel, 17, 37, 79
Turner, Matthew, 103

Uriguen, Jose Antonio, 62

Van De Ville, Dimitri, 45
Van hamme, Hugo, 119
Van Huffel, Sabine, 97
Vandereycken, Bart, 55
Vanderghenst, Pierre, 45, 50, 57
Vetterli, Martin, 16
Vidal, Rene, 31, 32, 122
Villaron, Emilie, 28

Wakin, Michael, 39, 61
Wan, Xiqin, 112
Wan, Zelong, 21
Wang, Hui, 109
Wang, Wenwu, 102
Wiaux, Yves, 50
Wright, Stephen, 18, 42

Xie, Meihua, 21
Xu, Lina, 103
Xu, Tao, 102

Yaghoobi, Mehrdad, 74
Yan, Fengxia, 21
Yang, Allen, 20
Yap, Han Lun, 39, 82
Yin, Wotao, 103
Yoo, Juhwan, 71
Yu, Hui Ka, 94

Zakharova, Anastasia, 93
Zhu, Jubo, 21
Zhu, Mengchen, 53
Zhu, Xiao Xiang, 24

SPARS '11 Sponsors

