

PROCEEDINGS OF THE
INTERNATIONAL BASP FRONTIERS
WORKSHOP 2013

January 27 - February 1, 2013

Villars-sur-Ollon, Switzerland

BASP FRONTIERS WORKSHOP 2013

Chairs

Dr. Y. Wiaux , EPFL, Switzerland

Dr. J. D. McEwen, University College London, UK

Technical Program Committee

Theoretical signal processing

Prof. M. Davies, University of Edinburgh, UK

Prof. M. Elad, Technion, Israel

Prof. J. Fadili, ENSICAEN, France

Prof. T. Strohmer, University of California, USA

Prof. J.-Ph. Thiran, EPFL, Switzerland

Prof. P. Vandergheynst, EPFL, Switzerland

Astrophysical signal processing

Dr. F. Baron, University of Michigan, USA

Dr. S. Bhatnagar, NRAO, USA

Dr. T. Cornwell, CSIRO, Australia

Prof. M. Hobson, University of Cambridge, UK

Prof. J. Peacock, University of Edinburgh, UK

Dr. J.-L. Starck, CEA, France

Biomedical signal processing

Dr. R. Deriche, INRIA Sophia Antipolis, France

Prof. M. Fink, ESPCI-ParisTech, France

Prof. M. Lustig, UC Berkeley, USA

Prof. L. Wald, MGH, USA

Local Organizing Committee

Dr. Y. Wiaux, EPFL, Switzerland

Prof. J.-Ph. Thiran, EPFL, Switzerland

Dr. J. D. McEwen, University College London, UK

Mrs. R. De Pietro, EPFL, Switzerland

Dr. A. Daducci, EPFL, Switzerland

Dr. R. Carrillo, EPFL, Switzerland

Mr. G. Puy, EPFL, Switzerland

Mr. B. Leistedt, University College London, UK

Foreword

Astronomy and biomedical sciences find common roots in their need to process acquired data into interpretable signals or images. In these applications of signal processing as for virtually all others, the complexity of data to be acquired and processed is constantly increasing, thus challenging signal processing theories. Data indeed come in larger volumes every day, can be multi-modal (referring to multiple imaging modalities), multi-spectral (referring to multiple imaging frequencies), scalar or tensor-valued (as polarization signals), living in high dimensional geometries, possibly non-Euclidean or discrete (as signals on the sphere or on graphs), etc.

As matter of fact, the astronomical and biomedical sciences communities are almost completely disconnected. Fostering contact and creating collaborations between these communities can shed new light on the problems of interest in each community and promote common research approaches for similar signal processing issues, thereby paving the way to new scientific advances. The international BASP Frontiers workshop was created to promote synergies between selected topics in theoretical, astronomical, and biomedical signal processing.

BASP Frontiers 2013 will take place in a very nice resort in the Swiss Alps named Villars-sur-Ollon, close to Lausanne and Lake Geneva. Everyone knows that the most fruitful discussions often take place after the sessions themselves, on the terrace, or during breakfast, lunch, or dinner. The winter atmosphere will further promote discussion and creativity.

Yves Wiaux & Jason McEwen
Workshop Chairs

Contents

General Interest	6
Michael A. Parker, <i>Higgs within reach: our understanding of the universe is about to change</i>	6
Theoretical Signal Processing	7
Anna Auria, Rafael E. Carrillo, and Yves Wiaux, <i>A convex optimization approach for image recovery from nonlinear measurements in optical interferometry</i>	7
Francis Bach, <i>Large-scale convex optimization for machine learning</i>	8
Richard Baraniuk, Aswin Sankaranarayanan, Christoph Studer and Thomas Goldstein, <i>Video Compressive Sensing</i>	9
Ayush Bhandari, <i>Sparse Sampling in Phase Space: Theory and Applications</i>	10
Cagdas Bilen, Gilles Puy, Remi Gribonval and Laurent Daudet, <i>Blind Sensor Calibration in Sparse Recovery</i>	11
Kristian Bredies, <i>Optimization for inverse problems with total generalized variation regularization and applications in medical imaging</i>	12
Yoram Bresler, <i>Learning Sparsifying Transforms</i>	13
Chunli Guo and Mike E. Davies, <i>Sample Allocation in Compressive Imaging</i>	14
Micha Feigin, Dan Feldman and Nir Sochen, <i>Learning Big (Image) Data via Coresets for Dictionaries</i>	15
Remi Gribonval, <i>Reconciling “priors” and “priors” without prejudice ?</i>	16
Rodney A. Kennedy, Parastoo Sadeghi, and Zubair Khalid, <i>Optimal signal processing on the 2-sphere: A general operator approach to signal recovery</i>	17
Matthieu Kowalski, Alexandre Gramfort, Pierre Weiss, and Sandrine Anthoine, <i>Solving mixed-norm estimate for the M/EEG inverse problem using an active-set strategy and proximal algorithm</i>	18
Felix Krahmer and Rachel Ward, <i>Beyond Incoherence: Stable and Robust Image Recovery from Variable Density Frequency Samples</i>	19
Boris Leistedt and Jason D. McEwen, <i>Flaglets: Exact Wavelets on the Ball</i>	20
Guillaume Madelin, Antonios Makr Mallis, and Frederick Poidevin, <i>Classification methods applied to astrophysical and MRI data</i>	21
Guillaume Obozinski and Francis Bach, <i>Convex relaxation for Combinatorial Penalties</i>	22
Gabriel Peyré, Samuel Vaiter, Charles Deledalle, and Jalal Fadili, <i>SURE-based Parameter Selection for Sparse Regularization of Inverse Problems</i>	23
Thomas Pock and Antonin Chambolle, <i>On a first-order primal-dual algorithm for convex optimization</i>	24
Jonathan Pokrass, Alexander M. Bronstein, Michael M. Bronstein, Pablo Sprechmann, and Guillermo Sapiro, <i>Sparse Modeling of Intrinsic Correspondences</i>	25
Gilles Puy and Pierre Vandergheynst, <i>Non-convex optimization for robust multi-view imaging</i>	26
Ramesh Raskar, <i>Femto-Photography and Looking Around the Corners</i>	27
Holger Rauhut, <i>Compressive Sensing with Structured Random Matrices</i>	28
Benjamin Ricaud and Bruno Torresani, <i>Refined support and entropic uncertainty inequalities</i>	29
Philip Schniter, <i>Adaptive compressive noncoherent change detection: An AMP-based approach</i>	30
Matthias W. Seeger, <i>Convergent and Scalable Algorithms for Expectation Propagation Approximate Bayesian Inference</i>	31
Eero Simoncelli, Chaitu Ekanadham and Daniel Tranchina, <i>Recovery of sparse translation-invariant signals with continuous basis pursuit</i>	32
Pierre Vandergheynst, <i>On localization and uncertainty in graph based representations</i>	33
Mehrdad Yaghoobi, Laurent Daudet, and Mike E. Davies, <i>Overcomplete Joint Sparsity Model for Dictionary Selection</i>	34

Astrophysical Signal Processing	35
Filipe B. Abdalla, <i>Signal processing techniques applied to future radio data</i>	35
Fabien Baron, Brian Kloppenborg, and John Monnier, <i>5D image reconstruction of stellar systems</i>	36
Nicolas Cantale, Frederic Courbin, and Georges Meylan, <i>2-Channels Finite Resolution Deconvolution</i>	37
Tobia D. Carozzi, <i>Information Capacity of Radio Interferometers</i>	38
Rafael E. Carrillo, Jason D. McEwen, and Yves Wiaux, <i>Sparsity averaging for radio-interferometric imaging</i>	39
Michael Hobson, <i>Accelerated Bayesian inference using nests and nets</i>	40
Daniel Jacobs, <i>Methods for detecting the 3D percolation of photons in the early universe.</i>	41
Brian Kloppenborg and Fabien Baron, <i>Accelerating optical interferometric image reconstruction and modeling using graphical processing units (GPUs)</i>	42
Anais Rassat, Jean-Luc Starck, Jalal Fadili and François-Xavier Dupe, <i>Application of Sparse inpainting to the Integrated Sachs-Wolfe Effect</i>	43
Ahmad Mour Sardarabadi and Alle-Jan van der Veen, <i>Application of Extended Factor Analysis for Interference Mitigation Using a reference array</i>	44
Oleg Smirnov, <i>Challenges of radio interferometric imaging in the SKA era</i>	45
Jean-Luc Starck, <i>Cosmic Microwave Background Sparse Recovery and Analysis</i>	46
Cyril Tasse, Joris van Zwieten, Ger van Diepen, Bas van der Tol, and Sanjay Bhatnagar, <i>Applying full polarization A-Projection to very-wide fields of view instruments: An imager for LOFAR</i>	47
Éric Thiébaud, <i>Optical Interferometric Imaging</i>	48
François Viallefond, <i>Towards a Theory of the Measurement Set</i>	49
Stefan J. Wijnholds, <i>Blind Self-Calibration of Sensor Arrays</i>	50
Stefan Wijnholds, Ahmad Mouri Sardarabadi, and Alle-Jan van der Veen, <i>Factor Analysis as a Tool for Signal Processing</i>	51
Laura Wolz, Filipe B. Abdalla, Rafael E. Carrillo, Yves Wiaux, and Jason D. McEwen, <i>The varying w spread spectrum effect for radio interferometric imaging</i>	52
Biomedical Signal Processing	53
Nikolaos Arvanitopoulos, Claudio Santelli, Matthias Seeger, and Sebastian Kozerke, <i>Divergence-free phase contrast MRI</i>	53
Claude Boccara, <i>Photo-acoustics and acousto-optics: mixing acoustics and optics for biomedical applications</i>	54
Davide Conigliaro, Paolo Manganotti, and Gloria Menegaz, <i>Early detection of epileptic seizures by entropy-based methods</i>	55
Alessandro Daducci, Anna Auria, Jean-Philippe Thiran, and Yves Wiaux, <i>ℓ_0-deconvolution for compressive diffusion MRI</i>	56
Maxime Descoteaux, <i>High angular resolution diffusion imaging (HARDI): from local reconstruction to fiber tracking</i>	57
Jack DiGiovanna, Gregoire Courtine, and Silvestro Micera, <i>Describing Neuronal Modulations with Bayesian Adaptive Regression Splines</i>	58
Mathias Fink, <i>Applications of ultrasonic time-reversal in biomedical imaging</i>	59
Luc Florack, Andrea Fuster, and Tom Dela Haije, <i>Riemann-Finsler Geometry and its Applications to Diffusion Magnetic Resonance Imaging</i>	60
Andrea Fuster, Tom Dela Haije, and Luc Florack, <i>On the Riemannian Rationale for Diffusion Tensor Imaging</i>	61
Foad Ghaderi, <i>Single Trial ERP detection Exploiting Sparsity in Time</i>	62
Alessandra Griffa, Alia Lemkaddem, Alessandro Daducci, and Jean-Philippe Thiran, <i>Local vs global tractography to study brain connectivity</i>	63

Diana Khabipova, Yves Wiaux, Rolf Gruetter and Jose P. Marques, <i>The importance of priors for l_2 regularization and total variation methods in quantitative susceptibility mapping</i> .	64
Theo Lasser, <i>Optical Coherent Imaging from Tissue to Molecule</i>	65
Denis Le Bihan, <i>Water diffusion MRI: what are we looking at?</i>	66
Chunlei Liu, <i>Imaging Tissue Magnetic Response with MRI</i>	67
Michael Lustig, <i>Rapid, Robust High Resolution MRI</i>	68
Abdallah G. Motaal, Rui Castro, Luc Florack, Klaas Nicolay, and Gustav Strijkers, <i>Accelerated Mouse Cardiac MR by Exploiting Retrospective Triggering Random Acquisition and CS Reconstruction</i>	69
Evren Ozarslan, <i>Tissue microstructure recovery with dMRI: Computational challenges</i>	70
Kawin Setsompop, <i>Ultra-Fast diffusion and resting-state fMRI imaging with Simultaneous Multi-Slice EPI and Q-space compressed sensing</i>	71
Yuchen Xie, Jiaqi Sun, Jeffery Ho, and Baba C. Vemuri, <i>Dictionary Learning on Riemannian Manifolds and Applications to EAP Reconstruction</i>	72
Maxim Zaitsev, <i>Magnetic Resonance Imaging with Non-Linear Spatial Encoding Magnetic Fields</i>	73
Mauro Zucchelli, Paolo Manganotti, and Gloria Menegaz, <i>Waveform decoding and detection in hdEEG</i>	74

Higgs within reach: our understanding of the universe is about to change.

Michael A.Parker*,

* Cavendish Laboratory University of Cambridge Cambridge CB3 0HE UK.

Abstract—We discuss the recent discovery of a Higgs-like boson at the Large Hadron Collider at CERN, and its implications for our understanding of the Standard Model of particles and fields.

The recent discovery of a new particle, consistent with the much sought Higgs boson marks a breakthrough in our understanding of the physical universe. Such a particle was first predicted in 1964, and plays a central part in the symmetry breaking mechanism which underlies the Standard Model of particles and their interactions. The theoretical motivation for the Higgs prediction, and its role in the Standard Model will be discussed.

The signal was discovered at the Large Hadron Collider in the presence of very high rates of background processes, using the advanced general purpose particle detectors built by the ATLAS and CMS collaborations. The detection systems and data reduction methods will be described. These use a variety of techniques to characterise the outgoing particles, allowing decay products consistent with those expected from the Higgs to be identified. The data reduction requires real-time processing in the detector front-end electronics and filtering in farms of processors before data storage. Massive distributed computing infrastructure has been deployed as a world-wide "Grid" to enable the collaborating scientists to analyse the multiple petabyte data samples.

The evidence for the new particle will be presented and open questions about its compatibility with the Standard Model predictions will be presented. The issue of the stability of the Higgs mass raises questions about potential new physics in the TeV energy range. Some current searches for such phenomena will be described.

A convex optimization approach for image recovery from nonlinear measurements in optical interferometry

Anna Auría*, Rafael E. Carrillo*, and Yves Wiaux*^{†‡}

* Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

[†] Department of Radiology and Medical Informatics, University of Geneva (UniGE), CH-1211 Geneva, Switzerland

[‡] Department of Radiology, Lausanne University Hospital (CHUV), CH-1011 Lausanne, Switzerland.

Abstract—Image recovery in optical interferometry is an ill-posed nonlinear inverse problem arising from incomplete power spectrum and bi-spectrum measurements. We formulate a linear version of the problem for the order-3 tensor formed by the tensor product of the signal with itself. This linear problem is regularized by standard convex ℓ_1 —relaxations of sparsity and low rank constraints and solved using the most advanced algorithms in convex optimization. We show preliminary results on small size synthetic images as a proof of concept.

Interferometry is a unique tool to image the sky at otherwise inaccessible resolutions. The set of visibilities measured provides an incomplete Fourier coverage of the brightness intensity as a function of angular direction, in vector form $x \in \mathbb{R}^N$ with components x_i . However, at optical wavelengths, the phase of the complex visibilities is affected by atmospheric turbulence. The measurable quantities are power spectrum data $|\hat{x}_i|^2$, and phases associated with the bi-spectrum $\hat{x}_i \cdot \hat{x}_j \cdot \hat{x}_k$ [1]. This poses a nonlinear inverse problem for image reconstruction, which is sensitive to the optimization strategy.

Generalizing the Phase Lift approach [2], we formulate a linear version of the problem for the order-3 tensor $\mathcal{X} = x \otimes x \otimes x \in \mathbb{R}^{N \times N \times N}$ with components \mathcal{X}_{ijk} , which arises from the tensor product of the signal with itself. The total flux is measured independently and we consider a normalized signal such that $\sum_i x_i = \hat{x}_0 = 1$, \hat{x}_0 representing the zero-frequency. Thus, the linear measurement model $y = \mathcal{A}(\mathcal{X}) \in \mathbb{C}^M$ encompasses both power spectrum and bi-spectrum measurements, for a measurement operator \mathcal{A} defined as a selection operator after Fourier transform along all tensor dimensions.

In this setting, prior information is essential to regularize the ill-posed inverse problem in the perspective of image reconstruction. Firstly, we adopt a sparsity prior on the tensor to acknowledge some signal sparsity $K \ll N$. While ℓ_0 —minimization would promote sparsity explicitly we adopt the common ℓ_1 —convex relaxation. Secondly, convex reality and positivity constraints are also enforced to acknowledge the fact that we deal with intensity images. Thirdly, in order to counter-balance the large increase of dimensionality when solving for \mathcal{X} instead of x , we rely explicitly on the fact that \mathcal{X} is formed by the tensor product of x , whose components sum to unity, so that summations over one or two indices respectively lead to the order-2 tensor $\mathcal{C}(\mathcal{X}) = x \otimes x \in \mathbb{R}^{N \times N}$, and to the signal x itself. We enforce the semi-definite positivity of $\mathcal{C}(\mathcal{X})$ and its rank-1 structure by resorting to nuclear norm minimization. The nuclear norm of an order-2 tensor is defined as the ℓ_1 —norm of its singular values vector. Its minimization should be understood as the convex relaxation of the minimization of the rank function counting the number of singular values. Finally, we also resort to a re-weighting scheme consisting in approaching both ℓ_0 —minimization on \mathcal{X} and rank minimization on $\mathcal{C}(\mathcal{X})$ by solving a sequence of weighted ℓ_1 and nuclear norm minimizations, each initialized to the solution of the previous problem. The fundamental symmetry of the tensor \mathcal{X} over index permutation is also enforced by ensuring that any operation performed preserves the symmetry. The weighted ℓ_1 and nuclear

norm minimization problem thus reads as:

$$\min_{\mathcal{X}} \|\mathcal{C}(\mathcal{X})\|_*^W + \lambda \|\mathcal{X}\|_1^W \quad \text{such that} \quad \|y - \mathcal{A}(\mathcal{X})\|_2 \leq \epsilon, \\ \text{and} \quad \mathcal{C}(\mathcal{X}) \succeq 0, \mathcal{X} \geq 0, \quad (1)$$

where the symbols $\|\cdot\|_*^W$ and $\|\cdot\|_1^W$ respectively denote weighted nuclear and ℓ_1 norms. In the weighted nuclear norm, the singular values of $\mathcal{C}(\mathcal{X})$ are essentially divided by their value at the previous iteration, in order to approximate the rank function. In the weighted ℓ_1 —norm, each tensor component \mathcal{X}_{ijk} is essentially divided by some robust estimation of its value from the previous iteration, in order to promote ℓ_0 —minimization. This estimation is obtained by symmetrized sums over two dimensions in order to promote structure in the tensor sparsity. In the first iteration, no weighting is applied. A non-weighted ℓ_1 —norm is not a meaningful prior function as the tensor values sum to unity. We thus set $\lambda = 0$ at the first iteration.

We solve this complex problem taking advantage of the versatility of convex optimization, using a combination of the Douglas-Rachford and dual forward-backward algorithms. The solution is low rank and we extract x as the principal eigenvector of $\mathcal{C}(\mathcal{X})$. We want to highlight the fact that in this framework the results do not depend on the initialization of the algorithm, in stark contrast with common non-convex approaches.

Figure 1 shows an example of reconstruction of a 16×16 synthetic image from $M = 0.75N$ measurements affected by 30dB of input noise, along with a phase transition diagram for random 8×8 images, representing the probability of good reconstruction in the sparsity-undersampling plane in a noiseless setting. In both cases equal numbers of random power and bi-spectrum data are considered.

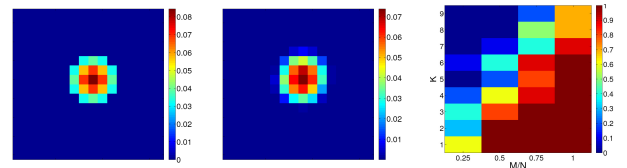


Figure 1. Left and centre: 16×16 image and reconstruction with $M = 0.75N$; SNR= 37.2dB. Right: phase transition diagram for 8×8 images.

The principal drawback of this approach is the dimension of the problem, leading to computation time and memory requirements issues. Advanced algorithmic, coding and hardware solutions need to be investigated in this perspective.

REFERENCES

- [1] E. Thiébaud and J. Giovannelli, “Image reconstruction in optical interferometry,” *IEEE Signal Processing Magazine*, vol. 21, pp. 97–109, 2010.
- [2] E. Candès, T. Strohmer, and V. Voroninski, “Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming,” *Communications on Pure and Applied Mathematics*, 2011, preprint in arXiv:1109.4499v1.

Large-scale convex optimization for machine learning

Francis Bach*

* INRIA - Sierra Project-team Département d'Informatique de l'Ecole Normale Supérieure Paris, France

Abstract—Many machine learning and signal processing problems are traditionally cast as convex optimization problems. A common difficulty in solving these problems is the size of the data, where there are many observations ("large n ") and each of these is large ("large p "). In this setting, online algorithms which pass over the data only once, are usually preferred over batch algorithms, which require multiple passes over the data. In this talk, I will present several recent results, showing that in the ideal infinite-data setting, online learning algorithms based on stochastic approximation should be preferred, but that in the practical finite-data setting, an appropriate combination of batch and online algorithms leads to unexpected behaviors, such as a linear convergence rate with an iteration cost similar to stochastic gradient descent. (joint work with Nicolas Le Roux, Eric Moulines and Mark Schmidt)

Video Compressive Sensing

Richard Baraniuk*, Aswin Sankaranarayanan*, Christoph Studer* and Thomas Goldstein*

* Rice University

Sensing and imaging systems are under increasing pressure to accommodate ever larger and higher-dimensional data sets; ever faster capture, sampling, and processing rates; ever lower power consumption; communication over ever more difficult channels; and radically new sensing modalities. Since its discovery in 2004, compressive sensing (CS) has stimulated a re-thinking of sensor and signal processing system design. In CS, analog signals are digitized and processed not via uniform sampling but via measurements using more general, even random, test functions. In contrast with conventional wisdom, the new theory asserts that one can combine “sub-Nyquist-rate sampling” with large-scale optimization for efficient and accurate signal acquisition when the signal has a sparse structure. In this talk, we will review the progress in field over the last 8 years, with a special emphasis on the pros and cons of the technique and on sensing ephemeral signals such as videos.

For video CS, we focus on spatial-multiplexing cameras (SMCs) that sample a scene through a series of coded projections using a spatial light modulator and a few optical sensor elements. SMC architectures are particularly useful when imaging at wavelengths for which full-frame sensors are too cumbersome or expensive. While existing CS recovery algorithms for SMCs perform well for static images, they typically fail for time-varying scenes (videos). We will overview a new CS Multi-scale Video (CS-MUVI) sensing and recovery framework for SMCs. The framework features a co-designed video CS sensing matrix and recovery algorithm that provide an efficiently computable low-resolution video preview. We then estimate the scene’s optical flow from the video preview and feed it into a convex optimization algorithm to recover the high-resolution video. In Fig. 1 we illustrate the performance of CS-MUVI as compared to a more conventional sliding window-based recovery video recovery technique. Even at a 64x measurement compression ratio, CS-MUVI is able to render the fine details of the scene.



(a) One frame from a target high-speed video sequence



(b) CS-MUVI recovery using 64x fewer measurements than voxels in target video



(c) Sliding window CS recovery using 64x fewer measurements than voxels in target video

Fig. 1. Performance of CS-MUVI as compared to a more conventional sliding window-based recovery video recovery technique.

Sparse Sampling in Phase Space: Theory and Applications

Ayush Bhandari* (ayush@MIT.edu)

* Massachusetts Institute of Technology, 75 Amherst St, Cambridge, MA 02139, USA

Abstract—The question that we seek to answer is: Given a continuous time sparse signal (in primal domain), is it possible to sample and reconstruct this sparse signal based on its characterization in a dual domain—the one which generalizes a number of already known transformations?

Phase Space transformations like the Linear Canonical Transform parametrically generalize a number of interesting transformations including the Fourier, Fresnel and fractional Fourier Transform among others. In this work, we develop a sampling and reconstruction methodology for sampling a sparse signal by analyzing their properties in Phase Space which is backward compatible with several known transformations. This problem, as will be seen, can be recast as a parameter estimation problem that has its roots in Prony’s method.

Given the general scope of analysis of sparse signals in transform domain, this work has applications in areas of spread spectrum methods for imaging, phase retrieval and computational photography.

Phase Space transformations are theoretically appealing for they generalize a number of existing transformations like the Fourier transform among others. One such construction includes the case of Linear Canonical Transform (LCT). The LCT (parametrized by $\Lambda \equiv \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $ad - bc = 1$) of a signal, say $x(t)$, is given by $\hat{x}_\Lambda(\omega) = \text{LCT}_\Lambda \{x\}$ or,

$$\hat{x}_\Lambda(\omega) \stackrel{\text{def}}{=} \begin{cases} \langle x(\cdot), \phi_\Lambda(\cdot, \omega) \rangle & \text{for } b \neq 0, \\ \sqrt{d} e^{j\frac{1}{2}(cd\omega^2)} x(d\omega) & \text{for } b = 0, \end{cases} \quad (2)$$

where $\phi_\Lambda(\cdot, \omega)$ is the phase space kernel defined in (1) of Table I and $\langle g, h \rangle$ denotes the inner-product in L_2 -sense. For $\Lambda \in \mathbb{R}^2$, the LCT is a unitary transformation and generalizes many transformations (Table I). Results applicable to the LCT domain can easily be extended to any of its special cases. The Fourier transform and the FrFT are the most notable byproducts. The LCT is equipped with an interesting composition property:

$$\text{LCT}_{\Lambda_2} \{\text{LCT}_{\Lambda_1} \{x\}\} = \text{LCT}_{\Lambda_3} \{x\} \Rightarrow \Lambda_2 \cdot \Lambda_1 = \Lambda_3$$

that leads to the definition of inverse-LCT of function, which is equivalent to $x(t) = \text{LCT}_{\Lambda^{-1}} \{\hat{x}_\Lambda\}$ or,

$$x(t) = \begin{cases} \langle \hat{x}_\Lambda(\cdot), \phi_{\Lambda^{-1}}(t, \cdot) \rangle & \text{for } b \neq 0, \\ \sqrt{a} e^{-j\frac{1}{2}(cat^2)} x(at) & \text{for } b = 0, \end{cases} \quad (3)$$

where Λ^{-1} is the inverse of Λ . If $\tilde{x}(t)$ is the approximation of $x(t)$, then $\|\tilde{x}(t) - x(t)\|_{L_2}^2 = 0$ whenever $(\omega_m b) \leq \frac{\omega_s}{2}$ (Nyquist rate for LCT domain), where ω_s is the sampling frequency. When $\Lambda_{\text{FT}} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, all the aforementioned results take form of Shannon’s sampling theorem. In this paper, we are interested in sampling a stream of K Dirac impulses—a signal with structure:

$$x(t) \stackrel{\text{def}}{=} \sum_{k=0}^{K-1} c_k \delta(t - t_k), \quad (4)$$

with weights and arbitrary shifts, $\{c_k, t_k\}$. By constructing an aperiodic version of Fourier Series expansion (cf. [1]) for LCT domain, it turns out that (4) has a representation of form:

$$x(t) = \frac{1}{\tau} e^{-j\frac{a}{2b}t^2} \sum_{n=-\infty}^{n=+\infty} \left(\sum_{k=0}^{K-1} \underbrace{c_k}_{a_k} \cdot \underbrace{e^{j\frac{a}{2b}t_k^2}}_{u_k^n} e^{-jn\omega_0 t_k} \right) e^{jn\omega_0 t}. \quad (5)$$

$p[n]$ —Sum of K -complex exponentials

TABLE I
LCT AS A GENERALIZATION OF OTHER TRANSFORMATIONS

$\phi_\Lambda(t, \omega) = \frac{1}{\sqrt{-j2\pi b}} \exp \left\{ -\frac{j}{2b} ((at^2 + d\omega^2) - 2\omega t) \right\}$ (1)	
LCT Parameters (Λ)	Corresponding Transform
$\begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} = \Lambda_\theta$	Fractional Fourier Transform (FrFT)
$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \Lambda_{\text{FT}}$	Fourier Transform (FT)
$\begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix}$	Fresnel Transform
$\begin{bmatrix} 1 & jb \\ j & 1 \end{bmatrix}$	Bilateral Laplace Transform
$\begin{bmatrix} 1 & -jb \\ 0 & 1 \end{bmatrix}, b \geq 0$	Gauss–Weierstrass Transform
$\frac{1}{\sqrt{2}} \begin{bmatrix} 0 & e^{-j\pi/2} \\ -e^{-j\pi/2} & 1 \end{bmatrix}$	Bargmann Transform

Discretizing (5) in m -points results in the following system:

$$\underbrace{x(m)}_{\mathbf{x}} e^{j\frac{a}{2b}m^2} = \frac{1}{\tau} \sum_{n=-\infty}^{n=+\infty} \underbrace{\left(\sum_{k=0}^{K-1} a_k u_k^n \right)}_{\mathbf{p}} e^{jn\omega_0 m} \Leftrightarrow \boxed{\mathbf{x} = \mathbf{D}_{\text{DFT}}^{m \times n} \mathbf{p}}$$

which is reminiscent of the problem tackled by Prony. For a full rank system, $\mathbf{D}_{\text{DFT}}^{-1} \mathbf{x}$ yields \mathbf{p} and then the parameters are estimated. In what remains, we will recast this as a sampling problem [1] where we use a pre-filter to sample (4) and reconstruct it using (5).

The *finite-rate-of-innovation theory* [2] is a special case of our result (set $\Lambda = \Lambda_\theta$). Similarly, our work has close connections with Phase Retrieval problem of P. Jaming [3]. These ideas have applications in computational imaging [4]. Given that the sparse signal spectrum can be uniquely characterized in (5), the problem of source separation of of bandlimited + impulsive signal (4) can be efficiently tackled. This has applications in computational imaging [5] as well as in mitigating impulsive noise [6]. The chirp-kernel of LCT might explain the success of spread spectrum imaging method of Wiaux et al. in [7]. In context of the recent results in line of *Mathematical Theory of Super-Resolution* of Candes et al. [8], our problem can be reformulated as an optimization problem (noisy case).

REFERENCES

- [1] A. Bhandari and A. I. Zayed, “Shift-Invariant and Sampling Spaces Associated with the Fractional Fourier Domain,” *IEEE Trans. Signal Proc.*, vol. 60, no. 4, pp. 1627–1637, April 2012.
- [2] “Special issue on compressive sensing,” *IEEE Journ. on Sel. Topics in Signal Proc.*, vol. 2, no. 4, pp. 608–608, 2008.
- [3] P. Jaming, “Uniqueness results for the phase retrieval problem of fractional Fourier transforms of variable order,” *arXiv:1009.3418 [math.CA]*.
- [4] C. Barsi and R. Raskar, “TOF Range Imaging,” *Private Communication*.
- [5] S. Nayar et al., “Fast separation of direct and global components of a scene using high frequency illumination,” *ACM Trans. on Graphics*, vol. 25, no. 3, pp. 935–944, 2006.
- [6] J. Wolf, “Redundancy, the discrete fourier transform, and impulse noise cancellation,” *IEEE Trans. Commun.*, vol. 31, no. 3, pp. 458–461, Mar 1983.
- [7] Y. Wiaux et al., “Spread spectrum for imaging techniques in radio interferometry,” *Monthly Notices of the Royal Astronomical Society*, vol. 400, no. 2, pp. 1029–1038, 2009.
- [8] E. Candes and C. Granda, “Super-Resolution from Noisy Data,” *arXiv:1211.0290v2 [cs.IT]*, 2012.

Blind Sensor Calibration in Sparse Recovery

Çağdaş Bilen*, Gilles Puy†, Rémi Gribonval* and Laurent Daudet‡

* INRIA, Centre Inria Rennes - Bretagne Atlantique, 35042 Rennes Cedex, France.

† Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

‡ Institut Langevin, CNRS UMR 7587, UPMC, Univ. Paris Diderot, ESPCI, 75005 Paris, France

Abstract—We consider the problem of calibrating a compressed sensing measurement system under the assumption that the decalibration consists of unknown complex gains on each measure. We focus on *blind* calibration, using measures performed on a few unknown (but sparse) signals. In the considered context, we study several sub-problems and show that they can be formulated as convex optimization problems, which can be solved easily using off-the-shelf algorithms. Numerical simulations demonstrate the effectiveness of the approach even for highly uncalibrated measures, when a sufficient number of (unknown, but sparse) calibrating signals is provided.

I. INTRODUCTION

We consider the blind calibration problem in a system with sensors having effective unknown complex valued gain and a number of unknown sparse training signals, $\mathbf{x}_l \in \mathbb{C}^N$, $l = 1 \dots L$. The measured signal, $y_{i,l} \in \mathbb{C}$, in this system is modeled as

$$y_{i,l} = d_i e^{j\theta_i} \mathbf{m}_i^* \mathbf{x}_l \quad i = 1 \dots M \quad * : \text{Conj. Transpose} \quad (1)$$

where $\mathbf{m}_i \in \mathbb{C}^N$ are known sensor projection vectors, $d_i \in \mathbb{R}^+$ are unknown gain magnitude and $\theta_i \in [-\pi, \pi]$ are the unknown phase shifts for each sensor. This problem can be simplified to 2 sub-problems for easier analysis.

II. GAIN CALIBRATION

For known phases, the calibration problem can be formulated as a convex optimization problem such that

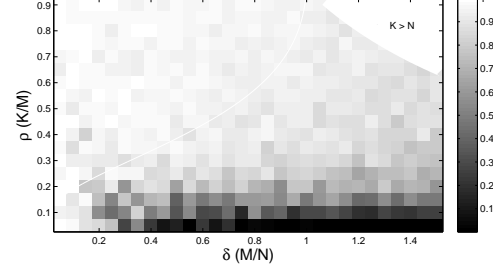
$$\begin{aligned} \mathbf{x}_{l,*}, \Delta_{i,*} = \arg \min_{\substack{\mathbf{x}_1, \dots, \mathbf{x}_L \\ \Delta_1, \dots, \Delta_M}} \sum_{l=1}^L \|\mathbf{x}_l\|_1 \quad (2) \\ \text{subject to} \quad \sum_{i=1}^M \Delta_i = c, \quad \Delta_i y_{i,l} = \mathbf{m}_i^* \mathbf{x}_l, \quad l = 1 \dots L \\ i = 1 \dots M \end{aligned}$$

where $c > 0$ is an arbitrary constant and the resulting estimated gains are $d_{i,*} = 1/\Delta_{i,*}$. This optimization problem has been investigated in [1]. The presented results show that, if there are sufficient number of training signals, the calibration approach provides significantly better performance than traditional recovery (by solving (2) with $\delta_i = 1, i = 1, \dots, M$) when the gain magnitudes have high variance.

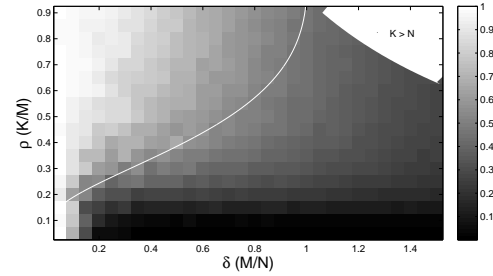
III. PHASE CALIBRATION

For known gains, the calibration problem is reduced to estimating the unknown phases. In case of single sparse training signal, this problem is equivalent to the phase retrieval problem investigated in [2]. When dealing with multiple sparse signals ($\mathbf{x}_l \in \Sigma_K$), we propose to perform the calibration and signal estimation with the semidefinite programming

$$\begin{aligned} \mathbf{X}_* = \arg \min_{\mathbf{X}} \quad \text{Tr}(\mathbf{X}) + \lambda \|\mathbf{X}\|_1 \quad (3) \\ \text{subject to} \quad \mathbf{X} \succeq 0, \quad y_{i,k} y_{i,l}^* = \mathbf{m}_i^* \mathbf{X}_{k,l} \mathbf{m}_i, \quad k, l = 1 \dots L \\ i = 1 \dots M \\ \mathbf{X}_{k,l} \triangleq \mathbf{x}_k \mathbf{x}_l^* \in \mathbb{C}^{N \times N}, \quad \mathbf{X} \triangleq \begin{bmatrix} \mathbf{X}_{1,1} & \dots & \mathbf{X}_{1,L} \\ \vdots & & \vdots \\ \mathbf{X}_{L,1} & \dots & \mathbf{X}_{L,L} \end{bmatrix} \in \mathbb{C}^{LN \times LN} \end{aligned}$$



(a) $L = 1$



(b) $L = 10$

Fig. 1: The decorrelation, σ_I , between the source signal and the estimated signal averaged over 10 randomly generated simulations for various ρ and δ ($\sigma_I(\mathbf{x}_1, \mathbf{x}_2) \triangleq 1 - \frac{|\mathbf{x}_1^* \mathbf{x}_2|^2}{\|\mathbf{x}_1\|_2^2 \|\mathbf{x}_2\|_2^2}$). The Donoho-Tanner phase transition curve is indicated with the white line.

which minimizes the rank and sparsity of the joint signal matrix \mathbf{X} . The resulting estimated signal, $\mathbf{x}_* = [\mathbf{x}_1^* \dots \mathbf{x}_L^*]^*$ is the eigenvector of \mathbf{X} that corresponds to the largest eigenvalue and the estimated phase shifts, $\theta_{i,*}$, are easily computed given \mathbf{x}_* and $y_{i,l}$. Sample simulation results comparing the joint optimization in (3) to the independent optimization described in [2] can be seen in Figure 1, which shows much higher correlation with the reconstructed signal for $L = 10$.

The talk will present further performance analysis of the proposed method for phase calibration, and discuss methods combining the gain and phase calibration approaches for calibration of complex valued gains.

REFERENCES

- [1] Gribonval, R.; Chardon, G.; Daudet, L.; , "Blind calibration for compressed sensing by convex optimization," Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on , vol., no., pp.2713-2716, 25-30 March 2012
- [2] H. Ohlsson, A. Y. Yang, R. Dong, and S. S. Sastry, "Compressive Phase Retrieval From Squared Output Measurements Via Semidefinite Programming", Accepted to the 16th IFAC Symposium on System Identification, SYSID 2012.

Optimization for inverse problems with total generalized variation regularization and applications in medical imaging

Kristian Bredies*

* Institute for Mathematics and Scientific Computing, University of Graz, Graz, Austria.

Abstract—We present and study the concept of total generalized variation for general symmetric tensor fields, in particular for the regularization of ill-posed inverse imaging problems in terms of optimization of Tikhonov functionals. Simple and efficient computational minimization algorithms are introduced, applications and numerical experiments with focus on medical imaging problems are reported.

I. INTRODUCTION

Since its introduction in [1], the total variation (TV) functional $TV(u) = \int_{\Omega} d|\nabla u|$ has become a well-established model for images and is often used as penalty functional for variational problems. One reason for its attractivity lies in the fact the the underlying space, the functions of bounded variation $BV(\Omega)$ allows for functions which admit a certain smoothness on the one hand, but also may contain jump discontinuities on the other hand. However, TV-regularization tends to deliver solutions which suffer from the so-called “staircasing effect”.

Several approaches have been proposed to overcome this problem. Here, we focus on the concept of total generalized variation introduced in [2] which realizes a functional which is able to detect higher-order smoothness while still accounting for discontinuities. We study the natural generalization to symmetric tensor fields of order l : For $u \in L^1_{loc}(\Omega, \text{Sym}^l(\mathbf{R}^d))$, the *total generalized variation* (TGV) of order k reads as

$$\text{TGV}_{\alpha}^{k,l}(u) = \sup \left\{ \int_{\Omega} u \cdot \text{div}^k v \, dx \mid v \in \mathcal{C}_c^k(\Omega, \text{Sym}^{k+l}(\mathbf{R}^d)), \right. \\ \left. \|\text{div}^i v\|_{\infty} \leq \alpha_i, \, i = 0, \dots, k-1 \right\} \quad (1)$$

where $\alpha = (\alpha_0, \dots, \alpha_{k-1})$ is a vector of positive weights.

II. REGULARIZATION WITH TGV

In order to use $\text{TGV}_{\alpha}^{k,l}$ as a regularization functional, its functional-analytic properties have to be examined. Denoting by \mathcal{E} the symmetrized derivative, we establish the equivalence of the norms

$$\|u\|_1 + \text{TGV}_{\alpha}^{k,l}(u) \sim \|u\|_1 + \|\mathcal{E}u\|_{\mathcal{M}}$$

implying that the underlying space for $\text{TGV}_{\alpha}^{k,l}$ is the space of *symmetric tensor fields of bounded deformation* $\text{BD}(\Omega, \text{Sym}^l(\mathbf{R}^d))$, studied in [3]. This will also give the coercivity estimate

$$\|u - Ru\|_{d/(d-1)} \leq C \text{TGV}_{\alpha}^{k,l}(u) \quad \forall u \in \text{BD}(\Omega, \text{Sym}^l(\mathbf{R}^d))$$

where R is a linear projection in $L^{d/(d-1)}(\Omega, \text{Sym}^l(\mathbf{R}^d))$ onto the finite-dimensional space $\ker(\mathcal{E}^k)$. This allows to regularize the linear inverse problem $Ku = f$ with TGV, i.e., to prove existence of minimizers of the Tikhonov functional

$$\min_{u \in L^p(\Omega, \text{Sym}^l(\mathbf{R}^d))} \frac{1}{2} \|Ku - f\|^2 + \text{TGV}_{\alpha}^{k,l}(u) \quad (2)$$

for $1 < p \leq d/(d-1)$ and $K : L^p(\Omega, \text{Sym}^l(\mathbf{R}^d)) \rightarrow H$ a linear and continuous mapping with some Hilbert space data $f \in H$.

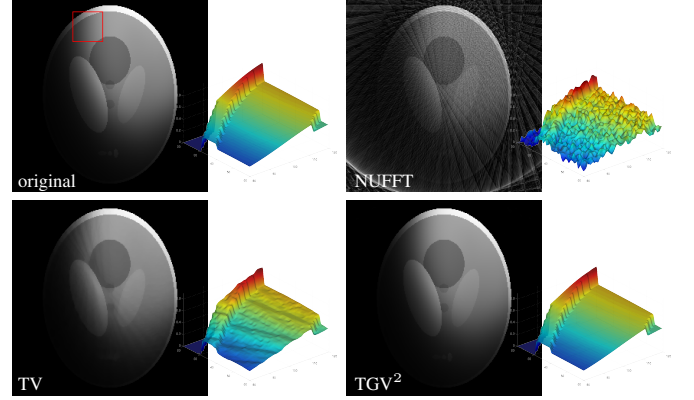


Fig. 1. Simulation results: Direct inversion (NUFFT), TV-regularization, TGV-regularization for undersampled MRI (24 radial projections).

The optimization problem (2) can easily be discretized and reformulated as an abstract saddle point problem

$$\min_x \max_y \langle \mathcal{A}x, y \rangle + \mathcal{F}(x) - \mathcal{G}(y)$$

where \mathcal{A} is a linear operator and \mathcal{F}, \mathcal{G} are proper, convex and lower semi-continuous functionals. Such problems can efficiently be solved with the iterative primal-dual problem introduced in [4]. In each update step, the resulting algorithms perform only simple computations such as linear operations and pointwise projections [5].

III. APPLICATIONS IN MEDICAL IMAGING

The computational optimization framework is then applied to three problems in medical imaging: undersampling magnetic resonance imaging (MRI) [6], reconstruction of diffusion-tensor imaging (DTI) data [7] and denoising of dual-energy computed tomography (CT) reconstructions. The numerical experiments, see, e.g. Fig. 1, confirm the favorable effects of TGV-regularization as the absence of the staircasing effect as well as the awareness of higher-order smoothness.

REFERENCES

- [1] L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D*, vol. 60, pp. 259–268, 1992.
- [2] K. Bredies, K. Kunisch, and T. Pock, “Total generalized variation,” *SIAM J. Imaging Sciences*, vol. 3, pp. 492–526, 2011.
- [3] K. Bredies, “Symmetric tensor fields of bounded deformation,” *Annali di Matematica Pura ed Applicata*, 2012, DOI 10.1007/s10231-011-0248-4.
- [4] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *J. Math. Imaging Vis.*, vol. 40, pp. 120–145, 2011.
- [5] K. Bredies, “Recovering piecewise smooth multichannel images by minimization of convex functionals with total generalized variation penalty,” University of Graz, Tech. Rep., 2012.
- [6] F. Knoll, K. Bredies, T. Pock, and R. Stollberger, “Second order total generalized variation (TGV) for MRI,” *Magnetic Resonance in Medicine*, vol. 65, no. 2, pp. 480–491, 2011.
- [7] T. Valkonen, K. Bredies, and F. Knoll, “Total generalised variation in diffusion tensor imaging,” University of Graz, Tech. Rep., 2012.

Learning Sparsifying Transforms

Yoram Bresler*,

* Coordinated Science Laboratory and Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

Abstract—Analytical sparsifying transforms or dictionaries, such as DCT, wavelets, curvelets, and finite differences have been used in many applications in signal processing. Recently, attention has turned to learning sparse signal representation that are directly adapted to data. However, while there has been extensive research on learning synthesis dictionaries and some recent work on learning analysis dictionaries, the idea of learning sparsifying transforms has received no attention. We describe a new formulation for data-driven learning of sparsifying transforms, and illustrate its advantages.

The sparsity of signals and images in a certain transform domain or dictionary has been exploited in many applications in signal and image processing, including compression, denoising, and compressed sensing. The following three alternative formulations have been considered for modeling sparsity.

- **Synthesis Dictionary:** the data vector y is modeled as $y \approx Dx$, with matrix D being the synthesis dictionary, and x a sparse vector.
- **Analysis Dictionary:** the data vector y is modeled as $y \approx q$, where $\Omega q = x$, with matrix Ω being the analysis dictionary, and x a sparse vector.
- **Transform Model:** the data vector y is modeled to satisfy $Wy \approx x$, with matrix W being the *sparsifying transform*, and x a sparse vector.

Historically, the transform model appears to have been the first to be used in signal and image processing, most notably in compression, and has been prominent in past and present image compression standards such as JPEG.

The various applications used sparsifying transforms or dictionaries such as DCT, wavelets, curvelets, and finite differences, all of which had a fixed, analytical form. Recently, sparse representations that are directly adapted to the data have become popular, especially in applications such as image denoising, inpainting, and medical image

reconstruction. However, while there has been extensive research on learning synthesis dictionaries and some recent work on learning analysis dictionaries, surprisingly the idea of learning sparsifying transforms has received no attention.

In this talk, we describe novel problem formulations and algorithms for learning sparsifying transforms from data. The formulation provides full control over the conditioning of the learned transforms, which are designed to be well-conditioned. The algorithms are shown to provide monotonic convergence of the cost function, and are insensitive to initialization. Moreover, on practical examples, their computational cost is nearly two orders of magnitude lower than that of synthesis dictionary learning algorithms such as K-SVD. Extending the idea of transform learning, we consider doubly sparse transforms, which are a product of a fixed, fast analytic transform such as the DCT, and an adaptive matrix constrained to be sparse. Such transforms can be learnt, stored, and implemented efficiently, providing further acceleration of transform learning.

The learned transforms provide much lower sparsification errors than analytical transforms. Results with natural images demonstrate that well-conditioning (but not necessarily unit conditioning) of the transforms is compatible with good sparsification and good performance in applications. Even for piecewise constant images, for which a difference operator provides optimal sparsification, but at high condition number, our well-conditioned learnt transforms provide essentially identical, or even better sparsification.

We show the superior promise of our approach as compared to analytical sparsifying transforms such as the DCT for image representation. We also show promising performance in image denoising using the learnt transforms, which compares favorably with approaches involving learnt synthesis dictionaries such as the K-SVD algorithm, but at orders of magnitude lower computational cost.

Sample Allocation in Compressive Imaging

Chunli Guo, Mike E. Davies

Institute for Digital Communications (IDCom), The University of Edinburgh, EH9 3JL, UK.

Abstract—We model the compressible signal with the two states Gaussian mixture (GM) distribution and consider the sample distortion function for the recently proposed Bayesian optimal AMP decoder. By leveraging the rigorous analysis of the AMP algorithm, we are able to derive the theoretical SD function and a sample allocation scheme for multi-resolution statistical image model. We then adopt the “turbo” message passing method to integrate the bandwise sample allocation with the exploitation of the hidden Markov tree (HMT) structure of wavelet coefficients. Experiments on natural image show that the combination outperforms either of them working alone.

I. INTRODUCTION

Suppose $X \in \mathbb{R}^n$, $i.i.d \sim p(x)$ is a realization of a random vector. In compressed sensing, we observe the linear combination $Y = \Phi X$ through the encoder $\Phi \in \mathbb{R}^{m \times n}$, $m < n$ and estimate X using the decoder Δ . Define the Sample Distortion (SD) function for X as the minimum achievable squared error distortion over all encoder-decoder pairs for a fixed undersampling ratio δ . We can show that the convex combination of two achievable SD points is also achievable by applying the two encoder-decoders to different portions of the source. Assuming $p(x) := \sum_{i=0}^1 p(s=i) \mathcal{N}(x; 0, \sigma_i^2)$, [1] has provided the theoretical basis for the SD function of Gaussian encoder-Bayesian optimal AMP (BAMP) decoder pair through the state evolution formalism.

We then extend the SD notation to the statistical multiresolution model for natural images. We restrict ourself to a block diagonal encoder which samples different wavelet band separately and consider the optimized bandwise sampling strategy. The convexified SD function enables us to perform a greedy sample allocation to achieve the least distortion. By leveraging the TurboAMP algorithm [2], We further incorporate HMT structure with the optimized bandwise sampling to maximize the SD performance.

II. SAMPLE ALLOCATION WITH INDEPENDENT MODEL

We group the wavelet coefficients of natural images according to the wavelet scale and impose the mutually independent GM distribution for each wavelet band. To derive the SD function for multi-resolution images, we need to optimize the sample allocation under the sample budget constraint $m = \delta n = \sum_i m_i$, with the aim of minimizing the total reconstruction distortion. We follow the ideas presented in [3] and use a distortion reduction (DR) function for each wavelet band: $d^{(i)}(m_i) := n_i[D(m_i/n_i) - D((m_i - 1)/n_i)]$. Assuming $m_i - 1$ samples have been allocated to the i th band, $d^{(i)}(m_i)$ is the amount of distortion decreased by adding one more sample to that band. With the convexified SD function, the optimal bandwise sampling is achieved by progressively allocating samples to the band which provides the greatest distortion reduction.

III. SAMPLE ALLOCATION WITH TREE STRUCTURE

To exploit the wavelet dependency across bands, we model the hidden states with the HMT structure and visualize the posterior of the source given the observation using a factor graph in Fig. 1. The turbo decoding [2] procedure is to exchange the local belief of $s_{j,k}$ between AMP decoding and HMT decoding alternately, by treating the likelihood on $s_{j,k}$ from one subgraph as prior for the the

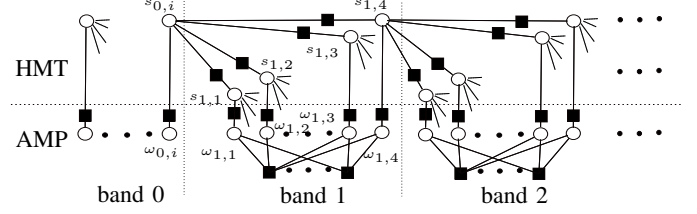


Fig. 1. Factor graph for bandwise sampling with HMT decoding. $\omega_{j,k}$: wavelet coefficient at scale j . $s_{j,k}$: hidden state for $\omega_{j,k}$.

other subgraph. The key feature of our factor graph is for the AMP decoding part, the sensing procedure is bandwise independent rather than a mixture of all the wavelet coefficients.

IV. NATURAL IMAGE EXAMPLE

We take the Haar wavelet decomposition of the 256×256 cameraman image as a prototypical example. Fig. 2 illustrated SDR for 5 encoder-decoder pairs under 4 undersampling ratios. The theoretical SD prediction is reinforced by the simulation results. The combination of turbo scheme and sample allocation generally delivers better reconstruction results than either of the two strategies working alone. Although the sample allocation based on independent model does not consider the HMT decoding, it is, in practice, a plausible choice for turbo decoding.

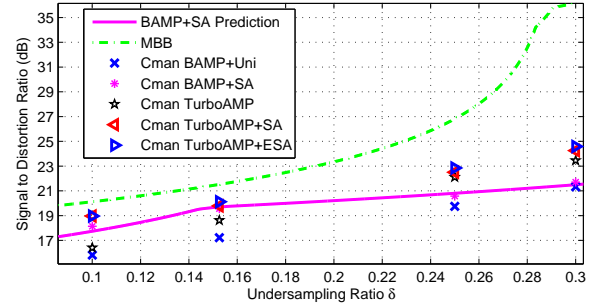


Fig. 2. SDR plot for Haar wavelet representation of cameraman. SA: sample allocation. Uni: uniformly distribute samples. ESA: empirically best sample allocation for turbo decoding. MBB: model based bound.

REFERENCES

- [1] D. Donoho, A. Maleki, and A. Montanari, “Message passing algorithms for compressed sensing: I. motivation and construction,” in *Information Theory Workshop (ITW), 2010 IEEE*, jan. 2010, pp. 1–5.
- [2] P. Som, S.; Schniter, “Compressive imaging using approximate message passing and a markov-tree prior,” *Signal Processing, IEEE Transactions on*, vol. 60, pp. 3439–3448, 2012.
- [3] M. Davies and C. Guo, “Sample-distortion functions for compressed sensing,” in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, sept. 2011, pp. 902–908.

Learning Big (Image) Data via Coresets for Dictionaries

Micha Feigin*, Dan Feldman† and Nir Sochen‡

* Media Arts and Sciences Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

† Computer Science and Artificial Intelligence lab, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

‡ Department of Mathematics, Tel-Aviv University, Tel-Aviv, Israel

Abstract—Signal and image processing have seen in the last few years an explosion of interest in a new form of signal/image characterization via the concept of sparsity with respect to a dictionary. An active field of research is dictionary learning: the representation of a given large set of vectors (e.g. signals or images) as linear combinations of only few vectors (patterns). To further reduce the size of the representation, the combination are usually sparse, i.e, each signal is a linear combination of only small number of patterns.

This work suggests a new computational approach to the problem of dictionary learning, known in computational geometry as *coresets*. Coreset is a small smart non-uniform sample from the input signals, such that the optimal dictionary for the input can be approximated by learning the coreset. Since the coreset is small, the learning is much faster. Moreover, using merge-and-reduce, the coreset can be constructed for streaming signals that do not fit in memory in logarithmic space, and can also be computed in parallel.

I. INTRODUCTION

One of the major problems in image processing is image characterization. By this we mean a system that gets a two-dimensional function and provides a probability measure as to whether this function is an image. We are still far from achieving this ultimate goal, although a few breakthroughs have been recorded. In the past, many characterizations used the decay rate of the coefficients of certain transformations. That led to a characterization in a linear space of functions. In the last decade, a new approach that involves redundant representations and sparsity has shown promise. In this framework, a signal is represented again as a superposition of signals, but unlike the representation with a basis of a linear space, the number of atoms exceeds the dimension of the signal such that a given signal may have many different representations. Uniqueness is achieved only for a subset of signals which can be represented with a limited number of atoms, called sparse signals. For this class of signals the sparsest representation is unique. This approach shifts the focus of attention from the general law of decay of coefficients to the outliers of such behavior, namely the large coefficients of such an expansion. The class of sparse signals does not form a linear space, which reflects the non-linearity of the set of images. At the same time, we still use linear techniques which helps a lot in practice.

Similar sparsity approaches have been used as well for problems such as image and texture classification, image compression and image denoising.

The sparsity approach has appealing features for image processing, but it suffers from a few problems. First, sparsity is a notion which is attached to a given dictionary. There is no one universal dictionary that can represent all image in a sparse way. This calls upon the need to construct dictionaries for different classes of images or for each application. Constructing a dictionary for a set of images from the same class/application goes under the name dictionary learning and is an active field of research.

Because of the prohibitive computational time and space complexity for computing dictionaries for large images or a large number of images, dictionary methods are generally limited to small patches of a single image or a small number of images.

This work brings the spell of coresets to cure the curse of space and time limitations. Informally, a coreset C for a set of elements Y is a compressed representation of Y that well approximates the original data in some problem-dependent sense. The given problem is then solved on the much smaller coreset and the resulting solution is applicable to the original set Y . This is done by using C to give approximate answers for queries about Y .

Coreset techniques were first introduced in the computational geometry field, and in recent years have been used to solve some well known open problems in computer science and machine learning. Coresets present a new approach to optimization and have huge success, especially in tasks which use large computation time and/or memory space. In particular, coresets suggest ways to use existing serial algorithms for distributed (parallel) computing, and provide solutions under the streaming model, where the space for solving the problem at hand is significantly smaller than its input size.

This work deals with coresets for learning dictionaries, and we demonstrate our ideas on the K-SVD method introduced by Aharon et al [1]. The K-SVD is a greedy algorithm designed to solve the following optimization problem: given positive values T_0, K and a matrix $Y \in R^{d \times n}$ of atoms, we want to find a dictionary $D \in R^{d \times K}$ and a sparse coefficient matrix $X \in R^{d \times k}$ that minimizes

$$\arg \min_{D, X} \|Y - DX\|_F \text{ s.t. } \forall i, \|x_i\|_0 \leq T_0. \quad (1)$$

where the atoms y are the columns of the matrix Y , the vector x_i is the i th column of X , $\|\cdot\|_F$ is the Frobenius norm (the sum of squared entries in the matrix) and $\|x_i\|_0 \leq T_0$ is the sparsity of x_i , i.e, x_i contains at most T_0 non-zeros.

The algorithm solves alternate optimization problem, alternating between finding a better D (using the SVD algorithm) while preserving the required sparsity of X and find the best sparse X given a dictionary D using a matching pursuit approach (where the original work uses orthogonal matching pursuit).

Formally, a coreset for the problem defined in (1) is a matrix C such that

$$\|Y - DX\|_F \sim \|C - D\bar{X}\|_F$$

for every dictionary $D \in R^{d \times k}$. Here, the symbol \sim denotes a multiplicative factor of $1 \pm \epsilon$. An ϵ -coreset C is efficient if its number of columns is $c \ll n$, and the optimization problem can be solved on the much smaller coreset without sacrificing too much accuracy.

Our coreset size depends solely on the required accuracy and problem parameters and not on the size of the input data. In fact, due to issues of stability and sensitivity to initial conditions of greedy algorithms, our method provides a better solution at a lower running time. Note though that we only provide a method to find the optimal D and X is later computed normally.

REFERENCES

- [1] 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, 2006.

Reconciling “priors” and “priors” without prejudice ?

Remi Gribonval*

* Inria Rennes – Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, France.

Abstract—We discuss a long-lasting *qui pro quo* between regularization-based and Bayesian-based approaches to inverse problems, and review some recent results that try to reconcile both viewpoints. This sheds light on some tradeoff between computational efficiency and estimation accuracy in sparse regularization.

I. INTRODUCTION

A central problem in modern signal processing is to solve inverse problems of the type $y = \mathcal{A}(x) + n$ where $\mathcal{A} : \mathbb{R}^N \rightarrow \mathbb{R}^m$ ($m \leq N$) models a linear or nonlinear observation process, and n accounts for additive noise in this process. Addressing such problems amounts to designing estimators $\Delta : \mathbb{R}^m \rightarrow \mathbb{R}^N$, also called decoders.

The last decade has witnessed a particularly impressive amount of work dedicated to linear dimensionality reducing observations processes ($m \ll N$), where $\mathcal{A}(x) = \mathbf{A}x$, with $\mathbf{A} \in \mathbb{R}^{m \times N}$. Many *sparse* decoders (greedy algorithms, iterative reweighted or thresholding schemes) have been carefully designed and their performance guarantees have been scrutinized on various types of signals.

Regularization: A particular family of decoders is associated to *regularization* through global optimization of a cost function

$$\Delta_\phi(y) := \arg \min_x \frac{1}{2} \|y - \mathcal{A}(x)\|_2^2 + \phi(x) \quad (1)$$

where $\phi : \mathbb{R}^N \rightarrow \mathbb{R}$ is a penalty function. The ℓ^1 decoder associated to $\phi(x) = \lambda \|x\|_1$ has attracted a particular attention, for the associated optimization problem is convex, and generalizations to other “mixed” norms are being intensively studied. Several facts explain the popularity of such approaches: a) these penalties have well-understood geometric interpretations; b) they are known to be sparsity promoting (the minimizer has many zeroes); c) this can be exploited in active set methods for computational efficiency; d) convexity offers a comfortable framework to ensure both a unique minimum and a rich toolbox of efficient and provably convergent optimization algorithms.

Bayesian modeling: While the convex and deterministic viewpoint on inverse problems has gained a strong momentum, there is another major route: the Bayesian one. Assuming prior distributions $x \sim P_X$, $n \sim P_N$ on the unknown and the noise, and measuring the risk with the squared loss $\|\Delta(\mathcal{A}(x)) - x\|_2^2$, the optimum decoder, in the sense of the minimum expected risk is the conditional mean, also known as posterior mean or minimum mean squared error (MMSE),

$$\Delta^*(y) := \mathbb{E}(x|y) = \int_{\mathbb{R}^N} x p(x|y) dx. \quad (2)$$

Its computation involves high-dimensional integration, which raises substantial issues typically addressed through sampling (MCMC, etc.).

II. RECONCILING TWO WORLDS?

Regularization and Bayesian estimation seemingly yield radically different viewpoints on inverse problems. In fact, they are underpinned by distinct ways of defining signal models or “priors”. The “regularization prior” is embodied by the penalty function $\phi(x)$ which promotes certain solutions, somehow carving an implicit signal

model. In the Bayesian framework, the “Bayesian prior” is embodied by where the mass of the signal distribution P_X lies.

The MAP qui pro quo: A *qui pro quo* been these distinct notions of priors has crystalized around the notion of *maximum a posteriori* (MAP) estimation, leading to a long lasting incomprehension between two worlds. In fact, a simple application of Bayes rule shows that under a Gaussian noise model $n \sim \mathcal{N}(0, \mathbf{I}_m)$ and *Bayesian prior* $P_X(x \in E) = \int_E p_X(x) dx$, $E \subset \mathbb{R}^N$, MAP estimation yields the optimization problem (1) with *regularization prior* $\phi_X(x) := -\log p_X(x)$. As an unfortunate consequence of an erroneous “reverse reading” of this fact, the optimization problem (1) with regularization prior $\phi(x)$ is now routinely called “MAP with prior $\exp(-\phi(x))$ ”. With the ℓ^1 penalty, it is often called “MAP with a Laplacian prior”. This *qui pro quo* has given rise to the erroneous but common myth that the optimization approach is particularly well adapted when the unknown is distributed as $\exp(-\phi(x))$.

A myth disproved: As a striking counter-example to this last myth, it has recently been proved [1] that when x is drawn i.i.d. Laplacian and $\mathbf{A} \in \mathbb{R}^{m \times N}$ is drawn from the Gaussian ensemble, the ℓ^1 decoder – and indeed any sparse decoder – will be outperformed by the least squares decoder $\Delta(y) = \mathbf{A}^+ y$, unless $m \gtrsim 0.15N$.

Reconciliation?: Can these routes be reconciled ? In the context of additive white Gaussian noise denoising ($m = N$, $\mathbf{A} = \mathbf{I}_m$, $n \sim \mathcal{N}(0, \mathbf{I}_m)$) it has been shown [2] that the truly Bayesian estimator $\Delta^*(y)$ is in fact the solution of an optimization problem (1), for some *regularization prior* ϕ^* fully determined by the *Bayesian prior* P_X :

$$\Delta^*(y) = \arg \min_x \frac{1}{2} \|y - x\|_2^2 + \phi^*(x). \quad (3)$$

Moreover, for any $y \in \mathbb{R}^m$, the global minimum $\Delta^*(y)$ is indeed the *unique stationary point* of the resulting optimization problem (3). In other words, for AWGN denoising, Bayesian estimation with any postulated Bayesian prior P_X can be expressed as a regularization with a certain regularization prior ϕ^* .

Is the reverse true ? The results in [2] show that the resulting regularization prior ϕ^* is necessarily smooth everywhere. Hence, many popular sparsity-promoting regularization priors cannot correspond to any Bayesian prior. In particular, the ℓ^1 penalty cannot be the MMSE estimator for any Bayesian prior P_X . In other words, the performance of any sparse-regularization scheme is necessarily sub-optimal. The talk will discuss consequences of these results in terms of tradeoffs between computational complexity and estimation performance, as well as possible extensions to under-determined linear or nonlinear problems.

REFERENCES

- [1] R. Gribonval, V. Cevher, and M. Davies, E., “Compressible Distributions for High-dimensional Statistics,” *IEEE Transactions on Information Theory*, 2012, was previously entitled “Compressible priors for high-dimensional statistics”. [Online]. Available: <http://hal.inria.fr/inria-00563207>
- [2] R. Gribonval, “Should penalized least squares regression be interpreted as Maximum A Posteriori estimation?” *Signal Processing, IEEE Transactions on*, vol. 59, no. 5, pp. 2405–2410, 2011. [Online]. Available: <http://hal.archives-ouvertes.fr/inria-00486840/PDF/MApvsOptim.pdf>

Optimal signal processing on the 2-sphere: A general operator approach to signal recovery

Rodney A. Kennedy*, Parastoo Sadeghi*, and Zubair Khalid*

* Research School Engineering,

The Australian National University, Canberra ACT 0200 Australia.

Abstract—The optimal recovery of signals whose domain is the 2-sphere is considered from an operator theoretical viewpoint. Under measurement, signals on the 2-sphere are subject to distortions and noise. These distortions may arise for example from the physical phenomena of transmission or from limitations of the measurement process. This work considers two standard engineering criteria, minimum mean square error (MMSE) and zero-forcing (ZF), for signal recovery given the distortion is linear and noise is isotropic and additive.

The class of bounded operators which we use to model the signal distortion process and the signal recovery process subsumes convolution operators defined on the 2-sphere and related elementary operators such as SO(3)-rotation operators. The theory is formulated in terms of infinite dimensional operator matrices which exist for any complete orthonormal set of functions on the 2-sphere. The signal recovery problem is solved in generality for the MMSE and ZF criteria. We show that unlike classical signal processing (where the distortion is usually taken as a convolution) the zero-forcing formulation does not result as the zero noise limit of the MMSE solution.

I. INTRODUCTION

The processing of signals whose domain is the 2-sphere, $\mathbb{S}^2 \triangleq \{\mathbf{x} \in \mathbb{R}^3: \|\mathbf{x}\| = 1\}$ is an active area of research with applications in geodesy, cosmology, and 3D beamforming/sensing [1]–[5].

Consider the complex Hilbert space $L^2(\mathbb{S}^2)$ of finite energy functions on the 2-sphere where the inner product is given by

$$\langle f, g \rangle = \int_0^{2\pi} \int_0^\pi f(\theta, \phi) \overline{g(\theta, \phi)} \sin \theta \, d\theta \, d\phi, \quad (1)$$

where θ is the co-latitude and ϕ the longitude. This inner product induces the norm $\|f\| \triangleq \langle f, f \rangle^{1/2}$. Such an $f \in L^2(\mathbb{S}^2)$ we call a signal on the 2-sphere (or simply “signal”).

II. PROBLEM FORMULATION

The goal of the paper is to estimate a desired signal $f \in L^2(\mathbb{S}^2)$ under the MMSE and ZF criteria when we observe it through a known distortion and additive noise of known statistics. The distortion is modeled via an infinite dimensional matrix representation of a bounded linear operator (see next section). The noise model is zero-mean, and isotropic [6]. The estimate of the desired signal is obtained by passing the noisy measurement through an operator also described by an infinite dimensional operator matrix. This second operator may be regarded as an equalizer in engineering terms. In essence the design is to determine the coefficients of the equalizer operator matrix. The problem is a significant generalization of that given in [6] which considered a restricted class of distortion models.

A linear operator \mathcal{B} on Hilbert space $L^2(\mathbb{S}^2)$ is *bounded* if there exists a constant $B \geq 0$ such that

$$\|\mathcal{B}f\| \leq B\|f\|, \quad \forall f \in L^2(\mathbb{S}^2). \quad (2)$$

Let $Y_\ell^m \equiv Y_\ell^m(\theta, \phi)$ denote the complex spherical harmonic of degree ℓ and order m then the following coefficients define the operator matrix \mathbf{B} corresponding to \mathcal{B} in the spherical harmonic basis [7]

$$b_{\ell,p}^{m,q} \triangleq \langle \mathcal{B}Y_p^q, Y_\ell^m \rangle. \quad (3)$$

If $f \in L^2(\mathbb{S}^2)$ is the desired signal we write \mathbf{f} to represent its Fourier coefficient sequence in the spherical harmonic basis. Similarly the Fourier coefficient sequence of the measured signal is

$$\mathbf{y} = \mathbf{B}\mathbf{f} + \mathbf{n} \quad (4)$$

where \mathbf{n} is the Fourier noise sequence and \mathbf{B} is the distortion operator matrix. To estimate the desired signal we compose an equalizer operator matrix \mathbf{Z} to form

$$\mathbf{Z}\mathbf{y} = \mathbf{Z}\mathbf{B}\mathbf{f} + \mathbf{Z}\mathbf{n}. \quad (5)$$

where criteria such as the MMSE and ZF can be used to design \mathbf{Z} . For example, the MMSE criterion seeks to find the \mathbf{Z} that minimizes the ℓ^2 complex sequence energy $\|\mathbf{Z}\mathbf{y} - \mathbf{f}\|^2$.

III. APPLICATION DOMAINS

The various 2-sphere convolution operators given in [2], [8], [9] can be used for the \mathbf{B} in the above formulation. These special cases have known applications [1], [2].

Deconvolution problems in cosmology fit within this framework including the case where the beam-pattern is axially symmetric. Spatial truncation operators which are outside the class of convolution operators fit in the framework, so that signal recovery is possible when regularizing conditions such as finite spectral degree are known [10].

REFERENCES

- [1] C. Armitage and B. D. Wandelt, “Deconvolution map-making for cosmic microwave background observations,” *Phys. Rev. D*, vol. 70, no. 12, p. 123007, Dec. 2004.
- [2] K.-W. Ng, “Full-sky correlation functions for CMB experiments with asymmetric window functions,” *Phys. Rev. D*, vol. 71, no. 8, p. 083009, Apr. 2005.
- [3] B. D. Wandelt and K. M. Górski, “Fast convolution on the sphere,” *Phys. Rev. D*, vol. 63, no. 12, p. 123002, May 2001.
- [4] B. Rafaely, “Plane-wave decomposition of the sound field on a sphere by spherical convolution,” *The Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2149–2157, Oct. 2004.
- [5] Y. Wiaux, L. Jacques, P. Vielva, and P. Vanderghelynst, “Fast directional correlation on the sphere with steerable filters,” *Astrophys. J.*, vol. 652, no. 1, pp. 820–832, Nov. 2006.
- [6] L. Wei and R. A. Kennedy, “Zero-forcing and MMSE filters design on the 2-sphere,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., ICASSP 2011*, Prague, Czech Republic, May 2011, pp. 4360–4363.
- [7] R. A. Kennedy and P. Sadeghi, *Hilbert Space Methods in Signal Processing*. New York, NY: Cambridge University Press, 2013.
- [8] J. R. Driscoll and D. M. Healy, Jr., “Computing Fourier transforms and convolutions on the 2-sphere,” *Adv. Appl. Math.*, vol. 15, no. 2, pp. 202–250, Jun. 1994.
- [9] P. Sadeghi, R. A. Kennedy, and Z. Khalid, “Commutative anisotropic convolution on the 2-sphere,” *IEEE Trans. Signal Process.*, 2013, (available as an early access article 10.1109/TSP.2012.2217337).
- [10] Z. Khalid, R. A. Kennedy, S. Durrani, and P. Sadeghi, “Conjugate gradient algorithm for extrapolation of sampled bandlimited signals on the 2-sphere,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., ICASSP 2012*, Kyoto, Japan, Mar. 2012, pp. 3825–3828.

Solving mixed-norm estimate for the M/EEG inverse problem using an active-set strategy and proximal algorithm

Matthieu Kowalski*, Alexandre Gramfort[†], Pierre Weiss[‡], Sandrine Anthoine[§]

* Univ Paris-Sud, L2S, Gif-sur-Yvette, France.

[†] Institut Mines-Telecom, Telecom ParisTech, CNRS LTCI, Paris, France

[‡] CNRS, Institut des Technologies Avancées du Vivant, Toulouse, France

[§] CNRS, LATP, Marseille, France

Abstract—We present an accelerated version of the ISTA algorithm, using a simple active-set strategy. The algorithm is shown to converge, and its effectiveness is demonstrated on M/EEG inverse problem.

I. INTRODUCTION

Magneto- and electroencephalography (M/EEG) measure the electromagnetic fields produced by the neural electrical currents. Given a conductor model for the head, and the distribution of source currents in the brain, Maxwell's equations allow one to compute the ensuing M/EEG signals. Given the actual M/EEG measurements and the solution of this forward problem, one can localize, in space and in time, the brain regions that have produced the recorded data. However, due to the physics of the problem, the limited number of sensors compared to the number of possible source locations, and measurement noise, this inverse problem is ill-posed. We consider a general class of priors based on mixed-norms. Such norms have the ability to structure the prior in order to incorporate some additional assumptions about the sources. We refer to such solvers as Mixed-Norm Estimates (MxNE). In order to be used in practice, we must be able to provide algorithms as fast as possible: an estimate should be obtained in few seconds. Regarding the size of the problem, classical solvers such as the Fast Iterative Shrinkage/Thresholding Algorithm [1] cannot be used directly.

II. MATHEMATICAL MODELING

The measurements $M \in \mathbb{R}^{N \times T}$ (N number of sensors and T number of time instants) are obtained by multiplying the source amplitudes $X \in \mathbb{R}^{S \times T}$ (S number of dipoles) by a forward operator $G \in \mathbb{R}^{N \times S}$, i.e., $M = GX$. In addition, the measurements are corrupted by an additive noise E : $M = GX + E$.

In the context of M/EEG, N lies between 50 for EEG only and 400 for M/EEG combined measurements, while S lies between 5000 and 50000 depending on the precision of the source model considered.

The functional considered here is based on the so-called group-lasso ℓ_{21} mixed norm. As a consequence, an estimation of X given by the minimization of Eq. (1) is sparse through the lines, i.e., all the coefficients of a line of X are either jointly nonzero, or all set to zero. This approach, proposed earlier for M/EEG [2], avoids the irregular time series obtained with a simple ℓ_1 norm.

$$X^* = \arg \min_X \frac{1}{2} \|M - GX\|^2 + \lambda \|X\|_{21}. \quad (1)$$

A detailed study is available in [3].

III. ALGORITHM

Starting from a practical implementation of Roth and Fisher's algorithm [4] to solve a Lasso-type problem, we propose and study the Active Set Iterative Shrinkage/Thresholding Algorithm (AS-ISTA). The main idea of this strategy, is to run (F)ISTA on subproblems

which have a small dimension. The subproblems are created by sticking on the variables which violate the most the KKT constraints. The algorithm, presented in [5], is summarized in Alg. 1

Algorithm 1: AS-ISTA

- 1) Initialization: $\mathcal{A}^{(0)} \subseteq \{k \text{ such that } |\langle g_k, M \rangle| > \lambda\}$,
 $X^{(0)} = \mathbf{0}_{\mathbb{R}^N}$.
 - 2) Let $\tilde{X}_{\mathcal{A}^{(t)}}^{(0)} = X_{\mathcal{A}^{(t)}}^{(t)}$. Iterate $J^{(t)}$ times

$$\tilde{X}_{\mathcal{A}^{(t)}}^{(j+1)} = \text{prox}_{\lambda \|\cdot\|_{21}} \left(\tilde{X}_{\mathcal{A}^{(t)}}^{(j)} + G_{\mathcal{A}^{(t)}}^T \left(M - G_{\mathcal{A}^{(t)}} \tilde{X}_{\mathcal{A}^{(t)}}^{(j)} \right) \right)$$
and define $X^{(t+1)}$ by $X_{\mathcal{A}^{(t)}}^{(t+1)} = \tilde{X}_{\mathcal{A}^{(t)}}^{(j)}$.
 - 3) Compute the dual variable $Z^{(t+1)} = M - GX^{(t+1)}$.
 - 4) Let $a^{(t+1)} \subseteq \{k, \text{ such that } |\langle g_k, Z^{(t+1)} \rangle| > \lambda\}$.
 - 5) Update the active set:

$$\mathcal{A}^{(t+1)} = \text{supp}(X^{(t+1)}) \cup \left\{ \arg \max_{k \in \{1, \dots, N\}} |\langle g_k, Z^{(t+1)} \rangle| \right\} \cup a^{(t+1)}.$$
 - 6) $t \leftarrow t + 1$ and go to 2.
-

The convergence is proven by observing that the algorithm can be seen as a particular case of a coordinate gradient descent algorithm with a *Gauss-Southwell-r* rule [6]. We will provide experimental evidence that the proposed method can outperform FISTA and significantly speed-up the resolution of very undetermined inverse problems when using sparse convex priors. The proposed algorithm significantly speeds up brain mapping with magneto- and electroencephalography (M/EEG) when promoting spatially sparse and temporally smooth solutions using an ℓ_{21} mixed-norm, which is crucial for analyzing real data.

REFERENCES

- [1] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [2] W. Ou, M. Hämaläinen, and P. Golland, "A distributed spatio-temporal EEG/MEG inverse solver," *NeuroImage*, vol. 44, no. 3, pp. 932–946, Feb 2009.
- [3] A. Gramfort, M. Kowalski, and M. Hämaläinen, "Mixed-norm estimates for the m/eeeg inverse problem using accelerated gradient methods," *Physics in Medicine and Biology*, vol. 50, no. 7, pp. 1931–1961, 2012.
- [4] V. Roth and B. Fischer, "The group-lasso for generalized linear models: uniqueness of solutions and efficient algorithms," in *ICML'08*, 2008, pp. 848–855.
- [5] M. Kowalski, P. Weiss, A. Gramfort, and S. Anthoine, "Accelerating ista with an active set strategy," in *Opt2011 (Nips workshop)*, 2011.
- [6] P. Tseng and S. Yun, "A coordinate gradient descent method for nonsmooth separable minimization," *Math. Prog. B*, vol. 117, pp. 387–423, 2009.

Beyond Incoherence: Stable and Robust Image Recovery from Variable Density Frequency Samples

Felix Krahmer*, Rachel Ward†,

* Institute for Numerical and Applied Mathematics, University of Göttingen, Germany

† Mathematics Department, University of Texas at Austin, TX, USA

Abstract—In many applications of compressed sensing, one wishes to acquire images that are sparse in transform domains such as spatial finite differences or wavelets using Fourier measurements. For these applications, overwhelming empirical evidence suggests that superior image reconstruction can be obtained through certain *variable density* sampling strategies which concentrate on lower frequencies. However, as the wavelet and Fourier domains are not incoherent, the compressed sensing theory does not directly imply the optimal sampling strategy. Here we use a more refined notion of coherence – the so-called *local coherence* – to derive guarantees that sampling frequencies from suitable power-law densities enables image reconstructions that are *stable* to sparsity defects and *robust* to measurement noise.

Inverse source or scattering problems play an important role in a wide range of imaging applications such as radar, sonar, astronomy, and computer tomography. After appropriate approximation and discretization, the measurements in such settings can be reformulated as samples from weighted discrete Fourier transforms [1]. Similarly, it is well known in the medical imaging literature that the measurements taken in Magnetic Resonance Imaging (MRI) are well modeled as Fourier coefficients of the desired image. The theory of *compressed sensing* [2], [3] seeks to reduce the number of measurements without degrading the quality of image reconstruction. The underlying key observation is that signals which allow for an approximately sparse representation in a suitable basis or dictionary – a central feature of many classes of natural images – can be recovered from relatively few linear measurements via convex optimization, provided these measurements are sufficiently *incoherent* with the basis in which the signal is sparse. However, incoherence-based theory falls short in many imaging applications, as natural sparsity domains for images such as wavelet bases are *not* incoherent to the Fourier basis.

A number of empirical studies, including the very first papers on compressed sensing MRI [4], [5], observed that image reconstructions from compressive frequency measurements could be significantly improved by subsampling frequencies according to variable densities, preferring low frequencies to high frequencies. In [6], the authors propose to use convex optimization to find a distribution such that the more coherent the basis element with respect to the sparsity basis, the higher the density at which they are sampled. While their approach is, in many regards, parallel to ours, they do not derive rigorous recovery guarantees for the resulting systems.

Note that lower frequencies are more coherent with wavelets and step functions than higher frequencies. Taking this reweighting into account, we find that a finer measure to determine the possibility of sparse recovery is the *local coherence* as given by the following definition.

Definition 1 (Local coherence): The *local coherence* of an orthonormal basis $\{\varphi_j\}_{j=1}^N$ of \mathbb{C}^N with respect to another orthonormal basis $\{\psi_k\}_{k=1}^N$ of \mathbb{C}^N is the function $\mu_{loc}(j) = \sup_{1 \leq k \leq N} |\langle \varphi_j, \psi_k \rangle|$.

Indeed, in our presentation we will argue that if the local coherence of $\{\varphi_j\}_{j=1}^N$ with respect to $\{\psi_k\}_{k=1}^N$ is pointwise bounded by a function κ , $\mu_{loc}(j) \leq \kappa(j)$, then a matrix Φ consisting of the

$m \geq C\delta^{-2}\|\kappa\|_2^2 s \log^3(s) \log(N)$ rows $\{\varphi_{L_k}\}_{k=1}^m$ will allow for the recovery of vectors with s -sparse expansions in the basis $\{\psi_j\}_{j=1}^N$, if the indices L_k are independent random variables drawn according to the local coherence,

$$\nu(j) := \mathbb{P}[L_k = j] \propto \mu_{loc}^2(j), \quad j \in \{1, 2, \dots, N\}$$

By preconditioning the sensing matrix Φ from the left by $D = \text{diag}(d_{j,j})$, where $d_{j,j} = \|\kappa\|_2/\kappa_j$, the resulting matrix can be interpreted as arising from a *bounded orthonormal system* – see [7]. This approach was implicitly introduced in the setting of function approximation in [8]. In the following proposition, we show the local incoherence of the Fourier basis with respect to the Haar wavelet system, a necessary ingredient to obtain image recovery guarantees for the case of frequency samples.

Proposition 2: Let $N = 2^p$ for integral $p \geq 8$. Then the local coherence μ_{loc} of the two-dimensional Fourier basis $\{\varphi_{\omega_1, \omega_2}\}$ with respect to the bivariate Haar wavelet basis $\{h_{n,k}^e\}$ in $\mathbb{C}^{N \times N}$ is bounded by

$$\mu_{loc}(\omega_1, \omega_2) \leq \kappa(\omega_1, \omega_2) := \min \left(1, \frac{18\pi\sqrt{2}}{(|\omega_1|^2 + |\omega_2|^2)^{1/2}} \right),$$

and one has $\|\kappa\|_2 \leq 52\sqrt{p} = 52\sqrt{\log_2(N)}$.

An immediate consequence of this result is that after increasing the number of measurements only by a single logarithmic factor in N , the well-known recovery results for frequency measurements under incoherence conditions [9] will carry over to signals which are sparse with respect to Haar wavelets.

Our main result given in the following uses these bounds combined with the methods introduced in [10] to establish recovery guarantees for total variation minimization.

Theorem 3: Fix $\delta < 1/3$ as well as integers

$$N = 2^p, \quad s \gtrsim \log(N), \quad \text{and} \quad m \gtrsim s \log^3 s \log^5 N.$$

Select m discrete frequencies (Ω_1^j, Ω_2^j) independently from the distribution $\nu \propto \mu_{loc}^2$ as defined in Proposition 2. Let $\mathcal{F}_\Omega : \mathbb{C}^{N^2} \rightarrow \mathbb{C}^m$ be the DFT matrix restricted to $\{(\Omega_1^j, \Omega_2^j)\}$.

Assume the noise ξ satisfies $\sum_{\omega_1, \omega_2} \xi_{\omega_1, \omega_2}^2 / \mu(\omega_1, \omega_2) \leq \varepsilon^2$. Then, with high probability, the following holds for all $f \in \mathbb{C}^{N \times N}$.

Given noisy measurements $y = \mathcal{F}_\Omega f + \xi$, the TV-minimizer

$$f^\# = \underset{g \in \mathbb{C}^{N \times N}}{\operatorname{argmin}} \|g\|_{TV} \quad \text{such that} \quad \sum_{\omega_1, \omega_2} \frac{(\mathcal{F}_\Omega g - y)_{\omega_1, \omega_2}^2}{\nu(\omega_1, \omega_2)} \leq \varepsilon^2,$$

approximates f to within the best s -term approximation error of the gradient ∇f and the noise level:

$$\|f - f^\#\|_2 \lesssim \frac{\|\nabla f - (\nabla f)_s\|_1}{\sqrt{s}} + \varepsilon.$$

Flaglets: Exact Wavelets on the Ball

Boris Leistedt* and Jason D. McEwen*

* Department of Physics and Astronomy, University College London, London WC1E 6BT, U.K.
 {boris.leistedt.11, jason.mcewen}@ucl.ac.uk

Abstract—We summarise the construction of exact axisymmetric scale-discretised wavelets on the sphere and on the ball. The wavelet transform on the ball relies on a novel 3D harmonic transform called the Fourier-Laguerre transform which combines the spherical harmonic transform with damped Laguerre polynomials on the radial half-line. The resulting wavelets, called flaglets, extract scale-dependent, spatially localised features in three-dimensions while treating the tangential and radial structures separately. Both the Fourier-Laguerre and the flaglet transforms are theoretically exact thanks to a novel sampling theorem on the ball. Our implementation of these methods is publicly available [1], [2] and achieves floating-point accuracy when applied to band-limited signals.

I. INTRODUCTION

Spherical wavelets have been applied successfully to numerous problems in astrophysics and geophysics to extract features of interest from signals on the sphere. But in these disciplines, signals on the sphere are often completed with radial information such as depth, redshift, or distance, in which case a full 3D analysis is required.

II. SCALE-DISCRETISED WAVELETS ON THE SPHERE

Scale-discretised wavelets [2], [3] allow one to probe and extract scale-dependent, spatially localised features in signals defined on the sphere. In the axisymmetric case (i.e. azimuthally symmetric when centered on the poles) scale-discretised wavelets reduce to needlets [4] and are constructed through a tiling of the harmonic line, thus defining an exact continuous transform on the sphere. Also, both the spherical harmonic and the scale-discretised wavelet transforms are exact in the discrete setting thanks to the use of a sampling theorem on the sphere [5]. In other words a band-limited signal, i.e. described by a finite number of spherical harmonic coefficients, is represented by a finite number of samples on the sphere without any loss of information. Since the wavelets are band-limited by construction a multiresolution algorithm is used to speed up the transform by capturing each wavelet scale in the minimal number of samples on the sphere. Our implementation of the scale-discretised wavelet transform is publicly available in the S2LET package [2] which supports the C, Matlab, IDL and Java programming languages. At present the code is restricted to axisymmetric wavelets but will be extended to directional, steerable wavelets and spin functions in a future release.

III. FLAGLETS ON THE BALL

The starting point to construct scale-discretised wavelets on the ball is a novel 3D transform, the Fourier-Laguerre transform, combining the spherical harmonics with damped Laguerre polynomials on the radial half-line [1]. We construct axisymmetric wavelets, which we call flaglets, by separately tiling the tangential and radial harmonic spaces of the Fourier-Laguerre transform. Both the Fourier-Laguerre and flaglet transforms are exact continuous transforms, which are also exact in the discrete setting thanks to a 3D sampling theorem on the ball. Flaglets extract scale-dependent, spatially localised angular and radial features in signals defined on the ball. Since the wavelets are band-limited in Fourier-Laguerre space by construction, a multiresolution algorithm is again introduced. Our implementations of these transforms on the ball are publicly available in the FLAG and FLAGLET packages [1].

IV. APPLICATIONS AND PERSPECTIVES

The flaglet transform probes tangential and radial structures at scales of interest while capturing all the information of a band-limited signal in the minimal number of samples on the ball. It is suitable for high precision analysis of 3D data that requires the separate treatment of angular and radial components. In future application we intend to exploit flaglets to study galaxy surveys, which are used in cosmology to study the large-scale structure of the Universe, specifically by confronting observations (e.g. clustering properties) with predictions of physical models. Galaxy surveys are contaminated with intrinsic uncertainties and systematics affecting the radial and angular dimensions differently. For example photometric redshifts of galaxies are estimated from colour information with much higher uncertainty than the estimate of the angular position of galaxies. Hence a separate treatment of the angular and radial information is needed to efficiently extract cosmological information from galaxy surveys and to constrain relevant physical theories. Also, the flaglet transform takes advantage of the sparsity of these surveys: gravity tends to generate a filamentary structure that is captured in a small number of flaglet scales, as shown in figure 1 for an N-body simulation.

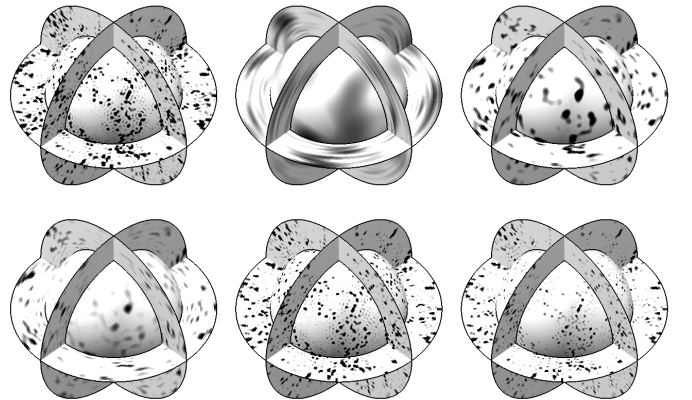


Fig. 1. N-body simulation (top left panel) and its flaglet coefficients for decreasing flaglet scales (subsequent panels from left-to-right, top-to-bottom).

REFERENCES

- [1] B. Leistedt and J. D. McEwen, “Exact wavelets on the ball,” *IEEE Trans. Sig. Proc.*, vol. 60, no. 12, 2012.
- [2] B. Leistedt, Y. Wiaux, P. Vanderghenst, and J. D. McEwen, “S2LET: A code to perform fast wavelet analysis on the sphere,” *Astron. & Astrophys.*, 2012, submitted.
- [3] Y. Wiaux, J. D. McEwen, P. Vanderghenst, and O. Blanc, “Exact reconstruction with directional wavelets on the sphere,” *Mon. Not. Roy. Astron. Soc.*, vol. 388, no. 2, pp. 770–788, 2008.
- [4] D. Marinucci, D. Pietrobon, A. Balbi, P. Baldi, P. Cabella, G. Kerkycharian, P. Natoli, D. Picard, and N. Vittorio, “Spherical needlets for cosmic microwave background data analysis,” *Mon. Not. Roy. Astron. Soc.*, vol. 383, pp. 539–545, Jan. 2008.
- [5] J. D. McEwen and Y. Wiaux, “A novel sampling theorem on the sphere,” *IEEE Trans. Sig. Proc.*, vol. 59, no. 12, pp. 5876–5887, 2011.

Classification methods applied to astrophysical and MRI data

Guillaume Madelin*, Antonios Makrimalis† and Frédérick Poidevin†

* NYU Medical Center, Department of Radiology, Center for Biomedical Imaging, New York, NY 10016, USA.

† UCL, KLB, Department of Physics & Astronomy, Gower Place, London WC1E 6BT, United Kingdom.

Abstract—We compared support vector machine (SVM), k-nearest neighbor (k-NN) and naive Bayes (NB) algorithms for classifying magnetic resonance imaging (MRI) and astrophysical data. An accuracy of $\sim 80\%$ can be achieved with k-NN on astrophysical data for classifying star forming regions (SFR) and non-SFR, and with SVM on sodium MRI data for classifying healthy subjects and subjects with osteoarthritis (OA).

I. INTRODUCTION

In this preliminary study, we compare three algorithms based on machine/statistical learning theory for classifying two different kinds of data: one from magnetic resonance imaging (MRI) and one from astrophysics. The algorithms were: support vector machine (SVM) [1], k-nearest neighbor (k-NN) [2] and naive Bayes (NB) [3]. On the astrophysical side, we use the classifiers for analyzing the SCUPOL catalog [4], in which most of the observed stellar regions are structures of the interstellar medium where star formation occurs at different stages. On the medical imaging side, we used machine learning to classify healthy subjects and patients with osteoarthritis (OA), based on tissue sodium concentration (TSC) measurements obtained with sodium MRI of cartilage in the knee.

II. DATA PROCESSING AND RESULTS

Astrophysical data - The SCUPOL catalog is a compilation of 83 regions observed at the James Clerk Maxwell Telescope between 1997 and 2005. Polarization maps at 850 microns are available on 48 star forming regions (SFR), 11 Young stellar Objects (YSO), 6 Starless Prestellar Cores (SPC), 9 Bok Globules (BG) and on 1 region toward the galactic center. Those regions have been first classified 'by eye' on a morphological basis, reflecting different structures of the interstellar medium. Each map provides a matrix of Stokes parameters containing information about the magneto-turbulent properties of the regions after integration of the signal. From this matrix, we derive the mean polarization degree $\langle p \rangle$, the standard deviation (std) of the polarization σ_p and the std of the polarization angles σ_θ . Additional parameters are the distance to each observed target d , and the size $npix$ of the sample for each map. Classification was performed between SFR and non-SFR data.

MRI data - Sodium MRI was acquired on the knee cartilage of 19 healthy volunteers and 28 OA subjects at 7 Tesla, with 2 acquisition sequences: one without fluid suppression - radial 3D (R3D) - and one with fluid suppression by inversion recovery (IR). Fluid suppression is expected to increase the sensitivity of sodium imaging to the characteristic loss of sodium content in cartilage with OA, by eliminating partial volume effect due to the presence of fluid signal within the large voxels (3D pixels) of the sodium images [5]. Sodium quantification was performed by linear regression from the signal of gel phantoms with known sodium concentrations placed on top of the knee. The mean TSC ($\langle TSC \rangle$) and its standard deviation σ_{TSC} was then measured in 3 regions of cartilage over 4 consecutive slices for each subject and each sequence (12 values/subject/sequence for $\langle TSC \rangle$ and σ_{TSC}). OA patients were previously diagnosed by the standard method of joint space narrowing from radiography.

TABLE I
BEST CLASSIFICATION RESULTS FOR ASTROPHYSICAL AND MRI DATA.

Machine Learning Algorithm	k-NN	NB	SVM
Astrophysical data: $\sigma_p + npix + d$			
Accuracy	79.5%	75.2%	77.0%
Sensitivity	80.8%	63.3%	73.0%
Specificity	78.3%	86.0%	80.0%
Astrophysical data: All 5 parameters			
Accuracy	77.0%	75.1%	73.0%
Sensitivity	81.9%	65.0%	70.0%
Specificity	72.2%	85.0%	77.0%
MRI data: $\langle TSC \rangle + \sigma_{TSC}$ from IR only			
Accuracy	62.9%	76.3%	79.4%
Sensitivity	64.8%	72.6%	75.8%
Specificity	61.7%	78.7%	82.0%
MRI data: $\langle TSC \rangle + \sigma_{TSC}$ from IR+R3D			
Accuracy	71.1%	72.7%	78.2%
Sensitivity	78.2%	60.0%	79.3%
Specificity	66.1%	81.0%	77.5%

Machine learning - All processing was performed in Matlab (Mathworks, USA). Z-score normalization was applied on the data. The train dataset was 25% of the data, with the rest of data used as test dataset. Each algorithm was applied 100 times for each set of parameters, and the mean sensitivity, specificity and accuracy (correct rate) of each method was calculated, where 'true positive' was defined as 'SFR' for astrophysical data or 'OA' for MRI data.

Results - The two best results from the accuracy point of view for both data are presented in Table I.

III. CONCLUSION

Machine learning algorithms were applied to astrophysical and MRI data and gave good accuracy ($\sim 80\%$) for binary classification, with both high sensitivity and specificity (75-80%). K-NN provided the best results for classifying SFR from non-SFR regions with only 3 parameters, while SVM proved to be superior for classifying OA from healthy subjects using fluid suppressed MRI data. Further studies will include optimization of the classification parameters (kernel functions, train/test sets) and acquisition of more data for extending the classification to other types of stellar regions (SFR, YSO, SPC, BG) or degrees of OA severity (Kellgren-Lawrence grades 1-4).

REFERENCES

- [1] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [2] T. Cover and P. Hart, "Nearest neighbor pattern classification," *Information Theory, IEEE Transactions on*, vol. 13, no. 1, pp. 21–27, 1967.
- [3] D. Lewis, "Naive (bayes) at forty: The independence assumption in information retrieval," *Machine Learning: ECML-98*, pp. 4–15, 1998.
- [4] B. Matthews, C. McPhee, L. Fissel, and R. Curran, "The legacy of scupol: 850 μm imaging polarimetry from 1997 to 2005," *The Astrophysical Journal Supplement Series*, vol. 182, no. 1, p. 143, 2009.
- [5] G. Madelin, J. Lee, S. Inati, A. Jerschow, and R. Regatte, "Sodium inversion recovery mri of the knee joint in vivo at 7t," *Journal of Magnetic Resonance*, vol. 207, no. 1, pp. 42–52, 2010.

Convex relaxation for Combinatorial Penalties

Guillaume Obozinski*, Francis Bach[†]

* ENPC, Guillaume.Obozinski@ens.fr

[†] INRIA/ENS, Francis.Bach@ens.fr

The last years have seen the emergence of the field of structured sparsity, which aims at identifying a model of small complexity given some a priori knowledge on its possible structure. Specifically, models with structured sparsity are models in which the set of non-zero parameters — often corresponding to a set of selected variables — is not only assumed to be small, but also to display structured patterns. Two important examples are group sparsity, where groups of parameters are simultaneously zero or non-zero, and hierarchical sparsity, where variables can only be selected following a prescribed partial order encoded by a directed acyclic graph. A common approach to the problem is to add to the empirical risk an implicit or explicit penalization of the structure of the non-zero patterns. In this talk, I will consider a generic formulation in which allowed structures are encoded by a combinatorial penalty, and show that when combined with continuous regularizer such as an L_p norm, a tightest convex relaxation can be constructed and used as a regularizer. The formulation considered allows to treat in a unified framework several a priori disconnected approaches such as norms based on overlapping groups, norms based on latent representations such as block-coding and submodular functions, and to obtain generic consistency and support recovery results for the corresponding estimators obtained as minimizers of the regularized empirical risk.

SURE-based Parameter Selection for Sparse Regularization of Inverse Problems

Gabriel Peyré*, Samuel Vaïter*, Charles Deledalle*, Jalal Fadili†

* Ceremade, Université Paris-Dauphine, France

† GREYC, ENSICAen, France

Abstract—This work reviews both theoretical and numerical aspects of parameter selection for inverse problems regularization. We focus our attention to a set of methods built on top of the Generalized Stein Unbiased Risk Estimator (GSURE) [1]. GSURE allows one to unbiasedly estimate the ℓ^2 -risk of an estimator on the orthogonal of the kernel of the forward operator. One can thus automatically and objectively select the value of some parameters of the estimator by minimizing the GSURE. Computing the GSURE necessitates the estimation of the generalized degrees of freedom (df) of the estimator. We prove in [2] a formula that gives an unbiased estimator of the df for sparse ℓ^1 analysis regularization. This includes analysis-type translation invariant wavelet sparsity and total variation. This theoretical analysis provides a better understanding of the sensitivity behavior of ℓ^1 -analysis regularization, but turns out to be difficult to compute numerically for large scale imaging problems. Indeed, convex optimization solvers only provide an approximate solution, which does not lead to a stable estimation of the df, hence of the GSURE. We addressed this issue in [3] by proposing a novel algorithm that recursively computes an unbiased and stable estimator of the ℓ^2 -risk from the sequence of iterates generated by a large class of convex optimization methods.

We consider the inverse problem of estimating a high resolution signal $x_0 \in \mathbb{R}^N$ from noisy observations $y = \Phi x_0 + w \in \mathbb{R}^P$ where $\Phi \in \mathbb{R}^{P \times N}$ is the imaging operator and $w \in \mathbb{R}^P$ is an additive noise.

An estimator $x_\lambda(y)$ of x_0 is a function that depends on the observation y only and is parameterized by some λ . A popular class of estimators is defined as the solution of a variational optimization problem

$$x_\lambda(y) \in \operatorname{argmin}_{x \in \mathbb{R}^N} \frac{1}{2} \|y - \Phi x\|_2^2 + \lambda J(x). \quad (1)$$

The J functional reflects a prior information on the signal, and $\lambda > 0$ can be adapted to the noise level and the signal regularity. Computing the optimal λ is a crucial but difficult issue, and this work details theoretical and numerical contributions to this problem.

I. GENERALIZED SURE

We now assume that $w \sim \mathcal{N}(0, \sigma^2 \text{Id}_P)$ is a zero-mean white Gaussian noise. We assume that $\mu_\lambda(y) = \Phi x_\lambda(y)$ is a single-valued mapping of y . In this setting, one seeks to optimally select λ that minimizes the expected risk $\mathbb{E}_w(\|x_\lambda(y) - x_0\|_2^2)$.

Denoting $\Pi = \Phi^*(\Phi\Phi^*)^+ \Phi$ the orthogonal projector on $\ker(\Phi)^\perp$, the GSURE is defined following [1] as

$$\text{GSURE}_\lambda(y) = \|\Phi^*(\Phi\Phi^*)^+ y - \Pi x_\lambda(y)\|_2^2 - \sigma^2 \operatorname{tr}((\Phi\Phi^*)^+) + 2\sigma^2 \operatorname{df}(y), \quad (2)$$

where $\operatorname{df}(y) = \operatorname{tr}\left((\Phi\Phi^*)^+ \frac{\partial \mu_\lambda(y)}{\partial y}\right)$ is the so-called generalized degrees of freedom (df). If $y \mapsto \mu_\lambda(y)$ is weakly differentiable, the GSURE gives an unbiased estimate of the risk on $\ker(\Phi)^\perp$, i.e. $\mathbb{E}_w(\|\Pi(x_0 - x_\lambda(y))\|^2) = \mathbb{E}_w(\text{GSURE}_\lambda(y))$. The optimal parameter λ is computed in practice by minimizing $\text{GSURE}_\lambda(y)$ from a single observation y .

II. ANALYTICAL EVALUATION FOR ANALYSIS SPARSITY

A popular class of regularizations is the so-called analysis-type ℓ^1 -norm prior where $J(x) = \|D^* x\|_1$ see [2]. It measures the sparsity of the correlations of the signal with the atoms of an analysis dictionary $D = (d_j)_{j=1}^Q \in \mathbb{R}^{N \times Q}$. This includes many popular regularizations, such as classical synthesis sparsity when $D = \text{Id}_N$ and anisotropic total variation when D^* is a finite difference approximation of the gradient operator.

We have shown in [2] that the mapping $y \mapsto \mu_\lambda(y)$ is C^∞ outside a set \mathcal{H} of Lebesgue measure 0, and that

$$\forall y \notin \mathcal{H}, \operatorname{df}(y) = \operatorname{tr}(\Pi \Gamma^{[J]}) \quad \text{where} \quad \Gamma^{[J]} = U(U^* \Phi^* \Phi U)^{-1} U^*$$

and U is a matrix whose columns form a basis of $\mathcal{G}_J = \ker(D_J^*)$ and $J = \{j \setminus \langle d_j, x_\lambda(y) \rangle\}$ indexes the co-support of a solution $x_\lambda(y)$ such that Φ is injective on \mathcal{G}_J (such a solution always exists). This expression of df, plugged into the GSURE definition (2), thus defines an unbiased estimate of the projected risk.

III. NUMERICAL COMPUTATION USING ITERATIVE SCHEMES

We consider an iterative algorithm of the form $x^{(\ell+1)} = \varphi(x^{(\ell)}, y)$ that converges to the estimator $x_\lambda(y)$. A popular family of such iterative schemes are the so-called proximal splitting methods, [4], designed to solve large-scale non-smooth convex optimization problems encompassing (1).

The iterate $x^{(\ell)} = x^{(\ell)}(y)$ is a function of the observation. Let

$$\forall \delta \in \mathbb{R}^P, \quad \eta^{(\ell)}[\delta] = \partial x^{(\ell)}(y)[\delta] \in \mathbb{R}^N$$

denotes the application to δ of the differential of $x^{(\ell)}$ at y . Following [3], $\eta^{(\ell)}$ are computed iteratively

$$\eta^{(\ell+1)}[\delta] = \partial_1 \varphi(x^{(\ell)}(y), y) \circ \eta^{(\ell)}[\delta] + \partial_2 \varphi(x^{(\ell)}(y), y)[\delta]$$

where $\partial_i \varphi$ denote the differential with respect to the i^{th} variable.

The expression of df can then be recursively estimated using Monte-Carlo integration of the trace from K random realizations of a zero-mean white Gaussian noise $\delta_k \sim \mathcal{N}(0, \text{Id}_P)$ for $k = 1, \dots, K$

$$\operatorname{df}^{(\ell)}(y) \approx \frac{1}{K} \sum_{i=1}^K \langle (\Phi\Phi^*)^+ \eta^{(\ell)}[\delta_k], \delta_k \rangle.$$

This can be plugged in (2) to recursively estimate the GSURE.

REFERENCES

- [1] Y. C. Eldar, "Generalized sure for exponential families: Applications to regularization," *IEEE Transactions on Signal Processing*, vol. 57, pp. 471–481, 2009.
- [2] S. Vaïter, C. Deledalle, G. Peyré, J. Fadili, and C. Dossal, "Local behavior of sparse analysis regularization: Applications to risk estimation," *Technical report, Preprint Hal-00687751*, 2012.
- [3] C. Deledalle, S. Vaïter, G. Peyré, J. Fadili, and C. Dossal, "Proximal splitting derivatives for risk estimation," *Proc. NCMIP'12*, 2012.
- [4] P. L. Combettes and J.-C. Pesquet, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer-Verlag, 2011, ch. Proximal Splitting Methods in Signal Processing, pp. 185–212.

On a first-order primal-dual algorithm for convex optimization

T. Pock* and A. Chambolle†

* Institute for Computer Graphics and Vision, Graz University of Technology, 8010 Graz, Austria

† Center of Applied Mathematics, Ecole Polytechnique, CNRS, 91128 Palaiseau, France

I. INTRODUCTION

We consider a particular class of convex optimization problems with known saddle-point structure that can be written as:

$$\min_{x \in X} \max_{y \in Y} \langle Kx, y \rangle + G(x) - F^*(y) \quad (1)$$

where X and Y are finite-dimensional vector spaces equipped with standard inner products $\langle \cdot, \cdot \rangle$, $K : X \rightarrow Y$ is a bounded linear operator and G and F^* are lower semicontinuous proper convex functions with known structure. We further assume that G and F^* are simple in the sense that their proximal maps are easy to compute.

The optimality condition for (1) is written as

$$\begin{pmatrix} K^T y^{n+1} + \partial G(x^{n+1}) \\ -Kx^{n+1} + \partial F^*(y^{n+1}) \end{pmatrix} \ni 0, \quad (2)$$

One of the most-general and best-studied iterative algorithms to solve problems such as (2) is the so-called the proximal-point algorithm [1], which for the above problem is written as follows:

$$\begin{pmatrix} \partial G(x^{n+1}) + K^T y^{n+1} \\ -Kx^{n+1} + \partial F^*(y^{n+1}) \end{pmatrix} + M \begin{pmatrix} x^{n+1} - x^n \\ y^{n+1} - y^n \end{pmatrix} \ni 0. \quad (3)$$

The algorithm is known to converge to a solution of (1), as long as M is symmetric and positive definite. It has been observed in [2], that by choosing

$$M = \begin{pmatrix} \frac{1}{\tau} I & -K^T \\ -K & \frac{1}{\sigma} I \end{pmatrix},$$

with $\tau, \sigma > 0$ and $\tau\sigma\|K\|^2 < 1$, M , the iterates of the proximal point algorithm can be written in the explicit form:

$$\begin{cases} x^{n+1} = (I + \tau\partial G)^{-1}(x^n - \tau K^T y^n) \\ y^{n+1} = (I + \sigma\partial F^*)^{-1}(y^n + \sigma K(2x^{n+1} - x^n)) \end{cases} \quad (4)$$

This is exactly the first-order primal-dual algorithm proposed in [3].

II. CONVERGENCE

It has been shown in [3] that algorithm (4) is optimal in the sense that it comes along with convergence rates that are known to be optimal for first-order methods. In particular the algorithm offers the following convergence rates:

- G and F^* not uniformly convex: In this case, convergence of the algorithm is difficult to quantify, but one can show that the partial primal-dual gap converges with rate $\mathcal{O}(1/N)$ for the averages of the iterates, which coincides with the best known rate.
- G or F^* uniformly convex: If one of the both functions are uniformly convex, then the algorithm can be accelerated by using iteration dependent steps, to yield $\mathcal{O}(1/N^2)$ which is also known to be optimal for first-order algorithms.
- G and F^* uniformly convex: In this case, convergence of the algorithm is linear, i.e. $\mathcal{O}(\omega^N)$ which equals the rate of other optimal first order algorithms including the heavy-ball method, TWIST and Nesterov's algorithm.

III. EXTENSIONS

Algorithm (4) can be extended in various ways in order to make it applicable for more general problems and also to speed up its convergence.

- **Preconditioning:** An obvious modification of the algorithm is to replace the scalar valued step sizes τ and σ in (4) by symmetric, positive definite matrices T and Σ . It is shown in [4] that algorithm (4) converges as long as $\|\Sigma^{-\frac{1}{2}} K T^{-\frac{1}{2}}\| < 1$. It is important that the choice of T , Σ still allows to efficiently compute the proximal mappings with respect to G and F^* and hence diagonal matrices are considered in [4]. For ill-conditioned linear operators K , the preconditioning can speed up the convergence of the algorithm.
- **Overrelaxation:** Since (4) is a proximal point algorithm, it is known that it can be overrelaxed. The idea is simply to perform the iterates as in (4) to obtain the point $(x^{n+1\frac{1}{2}}, y^{n+1\frac{1}{2}})$ and then to perform for $\gamma \in [0, 1)$ an overrelaxation of the form $(x^{n+1}, y^{n+1}) = (x^{n+1\frac{1}{2}}, y^{n+1\frac{1}{2}}) + ((x^{n+1\frac{1}{2}}, y^{n+1\frac{1}{2}}) - (x^n, y^n))$. It turns out that this additional overrelaxation can speed up the convergence on a number of problems while keeping the computation complexity basically unchanged.
- **Nonlinear proximal-point:** This extension refers to the generalization of the usual quadratic proximal distance function utilized in the proximal point algorithm to more general distances, say the Bregman distance on a smooth, strictly convex function $h(u)$. It turns out that such a choice can significantly simplify the computations of the proximal mappings.
- **Explicit steps:** If G and/or F^* can be written as the sum of a smooth plus a simple non-smooth function, it is natural to incorporate explicit gradient steps in (4) (see [5]). It can be shown that this setting can also be written as a primal-dual proximal point algorithm with the help of a special non-linear proximal distance function.

REFERENCES

- [1] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM Journal on Control and Optimization*, vol. 14, no. 5, pp. 877–898, 1976.
- [2] B. He and X. Yuan, "Convergence analysis of primal-dual algorithms for total variation image restoration," Nanjing University, China, Tech. Rep., 2010.
- [3] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2010.
- [4] T. Pock and A. Chambolle, "Diagonal preconditioning for first order primal-dual algorithms," in *International Conference of Computer Vision (ICCV 2011)*, 2011.
- [5] L. Condat, "A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," University of Caen, Tech. Rep., 2011, preprint hal-00609728.

Sparse Modeling of Intrinsic Correspondences

Jonathan Pokrass*, Alexander M. Bronstein*, Michael M. Bronstein†, Pablo Sprechmann‡, and Guillermo Sapiro‡

* School of Electrical Engineering, Tel Aviv University, Israel

† Faculty of Informatics, Università della Svizzera Italiana, Switzerland

‡ Department of Electrical and Computer Engineering, Duke University, USA

Abstract—We present a novel sparse modeling approach to non-rigid shape matching using only the ability to detect repeatable regions. As input, we are given only two sets of regions in two shapes; no descriptors are provided so the correspondence between the regions is unknown, nor we know how many regions correspond in the two shapes. We show that even with such scarce information, it is possible to establish dense correspondence between the shapes by using methods from the field of sparse modeling. We formulate the problem of *permuted sparse coding*, in which we solve simultaneously for an unknown permutation ordering the regions on two shapes and for an unknown correspondence in functional representation. Numerically, the problem is solved efficiently by alternating the solution of a linear assignment and a sparse coding problem.

I. INTRODUCTION

Ovsjanikov et al. [1] proposed an elegant way to avoid direct representation of correspondences as maps between shapes using a functional representation. The authors noted that when two shapes X and Y are related by a bijective correspondence $t : X \rightarrow Y$, then for any real function $f : X \rightarrow \mathbb{R}$, one can construct a corresponding function $g : Y \rightarrow \mathbb{R}$ as $g = f \circ t^{-1}$. t uniquely defines a linear mapping between two function spaces $T : L^2(X) \rightarrow L^2(Y)$, where $\mathcal{F}(X, \mathbb{R})$ denotes the space of real functions on X . Equipping X and Y with harmonic bases, $\{\phi_i\}_{i \geq 1}$ and $\{\psi_j\}_{j \geq 1}$, respectively, one can represent a function $f : X \rightarrow \mathbb{R}$ using the set of Fourier coefficients $\{a_i\}_{i \geq 1}$ as $f = \sum_{i \geq 1} a_i \phi_i$. Then, translating the representation into the other harmonic basis, one obtains a simple representation of the correspondence between the shapes

$$T(f) = \sum_{i,j \geq 1} a_i c_{ij} \psi_j, \quad (1)$$

where c_{ij} are Fourier coefficients of the basis functions of X expressed in the basis of Y , defined as $T(\phi_i) = \sum_{j \geq 1} c_{ij} \psi_j$. The correspondence can be thus by approximated using k basis functions and encoded by a $k \times k$ matrix $\mathbf{C} = (c_{ij})$ of these coefficients. In this representation, the computation of the shape correspondence $t : X \rightarrow Y$ is translated into a simpler task of finding \mathbf{C} from a set of correspondence constraints.

II. PERMUTED SPARSE CODING

Let us be given a set of q regions on X represented by indicator functions $\{f_i\}_{i=1}^q$ and similarly, g_i on Y , assuming that $g_j \approx f_i \circ t^{-1}$. We stress that the ordering of the f_i 's and g_j 's is *unknown*, i.e., we do not know to which g_j in Y a f_i in X correspond. This ordering can be expressed by an unknown $q \times q$ permutation matrix $\mathbf{\Pi}$. From (1) it follows that $\mathbf{\Pi B} = \mathbf{AC}$, where $\mathbf{A} = (f_1, \dots, f_q)^T \Phi$ and $\mathbf{B} = (g_1, \dots, g_q)^T \Psi$. Note that both $\mathbf{\Pi}$ and \mathbf{C} are unknown, and solving for them is a highly ill-posed problem. Recalling that the correspondence should be represented by a nearly-diagonal \mathbf{C} , we formulate the following problem

$$\min_{\mathbf{C}, \mathbf{\Pi}} \frac{1}{2} \|\mathbf{\Pi B} - \mathbf{AC}\|_F^2 + \lambda \|\mathbf{W} \odot \mathbf{C}\|_1. \quad (2)$$



Fig. 1. Correspondence as a sparse modeling problem: Indicator functions of repeatable regions on two shapes are represented as matrices of coefficients \mathbf{A} and \mathbf{B} in the corresponding bases Φ and Ψ . Correspondence between the shapes is encoded by the matrix \mathbf{C} . We solve $\mathbf{\Pi B} = \mathbf{AC} + \mathbf{O}$ simultaneously for an approximately diagonal \mathbf{C} and the permutation $\mathbf{\Pi}$ bringing the indicator functions into correspondence. To cope with imperfectly matching regions, we relax the surjectivity of the permutation and absorb the mismatches into a row-wise sparse outlier matrix \mathbf{O} .

where \odot denotes element-wise multiplication, and the non-negative parameter λ determines the relative importance of the penalty. Small weights w_{ij} in \mathbf{W} are assigned close to the diagonal, while larger weights are selected for the off-diagonal elements.

The solution of (2) can be obtained using alternating minimization over \mathbf{C} with fixed $\mathbf{\Pi}$ boiling down to a sparse coding problem,

$$\min_{\mathbf{C}} \frac{1}{2} \|\mathbf{B\Pi} - \mathbf{AC}\|_F^2 + \lambda \|\mathbf{W} \odot \mathbf{C}\|_1, \quad (3)$$

and minimization over $\mathbf{\Pi}$ with fixed \mathbf{C} boiling down to a linear assignment problem with a relaxation of $\mathbf{\Pi}$ as a double-stochastic matrix,

$$\max_{\mathbf{\Pi} \geq 0} \text{vec}(\mathbf{E})^T \text{vec}(\mathbf{\Pi}) \quad \text{s.t.} \quad \begin{cases} \mathbf{\Pi} \mathbf{1} = \mathbf{1} \\ \mathbf{\Pi}^T \mathbf{1} = \mathbf{1} \end{cases} \quad (4)$$

We refer to problem (2) as to *permuted sparse coding*, and propose to solve it by alternating the solution of the standard sparse coding problem (3) and the solution of the linear assignment problem (4).

Proposition. The process alternating (3) and (4) converges to a global minimizer of the permuted sparse coding problem (2).

III. ROBUST PERMUTED SPARSE CODING

Let us assume that q and s regions are detected on X and Y and only $s \leq q$ f_i 's have corresponding g_j 's. This means that there is no correspondence between $r - s$ rows of \mathbf{B} and $q - s$ rows of \mathbf{A} , and the relation $\mathbf{\Pi B} \approx \mathbf{AC}$ holds only for an unknown subset of its rows. We absorb the $r - s$ mismatches in a row-sparse $q \times n$ outlier matrix \mathbf{O} that we add to the data term of (2). This results in the following problem

$$\min_{\mathbf{C}, \mathbf{O}, \mathbf{\Pi}} \frac{1}{2} \|\mathbf{\Pi B} - \mathbf{AC} - \mathbf{O}\|_F^2 + \lambda \|\mathbf{W} \odot \mathbf{C}\|_1 + \mu \|\mathbf{O}\|_{2,1}, \quad (5)$$

which we refer to as *robust permuted sparse coding*.

REFERENCES

- [1] M. Ovsjanikov, M. Ben-Chen, J. Solomon, A. Butscher, and L. Guibas, "Functional maps: A flexible representation of maps between shapes," *TOG*, vol. 31, no. 4, 2012.

Non-convex optimization for robust multi-view imaging

Gilles Puy* and Pierre Vanderghyest*

* Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

Abstract—We study the multi-view imaging problem where one has to reconstruct a set of l images, representing a single scene, from a few measurements made at different viewpoints. We first express the solution of the problem as the minimizer of a non-convex objective function where one needs to estimate one reference image, l foreground images modeling possible occlusions, and a set of l transformation parameters modeling the inter-correlation between the observations. Then, we propose an alternating descent method that attempts to minimize this objective function and produces a sequence converging to one of its critical points. Finally, experiments show that the method accurately recovers the original images and is robust to occlusions.

I. PROBLEM FORMULATION

In multi-view imaging, we have in hand l observations $\mathbf{y}_1, \dots, \mathbf{y}_l \in \mathbb{R}^m$ of a reference image $\mathbf{x}_0 \in \mathbb{R}^n$. As these observations are done from different viewpoints, the image \mathbf{x}_0 undergoes geometric transformations. We consider here transformations represented by few parameters (e.g., homography) and denote $\boldsymbol{\theta}_j \in \mathbb{R}^q$ the parameters associated to the j^{th} observations. The reference image transformed according to $\boldsymbol{\theta}_j$ is estimated using, e.g., a cubic spline interpolation and is equal to $S(\boldsymbol{\theta}_j)\mathbf{x}_0$, with $S(\boldsymbol{\theta}_j) \in \mathbb{R}^{n \times n}$.

To handle realistic applications, we also assume that parts of the reference image might sometimes be occluded. We model these occlusions using l foreground images $\mathbf{x}_1, \dots, \mathbf{x}_l \in \mathbb{R}^n$, and assume that the image “viewed” by the j^{th} observer is $S(\boldsymbol{\theta}_j)\mathbf{x}_0 + \mathbf{x}_j$.

Finally, we model the acquisition device using a linear operator $\mathbf{A} \in \mathbb{R}^{m \times n}$, and the observation model satisfy

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_l \end{bmatrix} = \begin{bmatrix} \mathbf{A}S(\boldsymbol{\theta}_1) & \mathbf{A} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}S(\boldsymbol{\theta}_l) & \mathbf{0} & \dots & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 \\ \vdots \\ \mathbf{x}_l \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_l \end{bmatrix}, \quad (1)$$

where $\mathbf{n}_1, \dots, \mathbf{n}_l \in \mathbb{R}^m$ represent additive measurement noise.

To reconstruct the images $\mathbf{x}^\top = (\mathbf{x}_0^\top, \dots, \mathbf{x}_l^\top)$ and the transformation parameters $\boldsymbol{\theta}^\top = (\boldsymbol{\theta}_1^\top, \dots, \boldsymbol{\theta}_l^\top)$ from the observations $\mathbf{y}^\top = (\mathbf{y}_1^\top, \dots, \mathbf{y}_l^\top)$, we wish to solve the following problem

$$\min_{(\mathbf{x}, \boldsymbol{\theta})} \{L(\mathbf{x}, \boldsymbol{\theta}) = \|\Psi^\top \mathbf{x}\|_1 + \kappa \|\mathbf{A}(\boldsymbol{\theta}) \mathbf{x} - \mathbf{y}\|_2^2 + \sum_{1 \leq j \leq l} i_{\Theta_j}(\boldsymbol{\theta}_j)\}, \quad (2)$$

which is non-convex. The matrix $\Psi \in \mathbb{R}^{(l+1)n \times (l+1)n}$ is block-diagonal and built by repeating $l+1$ times, e.g., the Haar wavelet basis on the diagonal, $\kappa > 0$ is a regularizing parameter, $\mathbf{A}(\boldsymbol{\theta}) \in \mathbb{R}^{lm \times (l+1)n}$ is the matrix appearing in (1), $(\Theta_j)_{1 \leq j \leq l}$ are closed convex subsets of \mathbb{R}^q , and i_{Θ_j} is the indicator function of Θ_j .

II. NON-CONVEX OPTIMIZATION

To solve problem (2), we propose an alternating descent method producing a sequence of estimates $(\mathbf{x}^k, \boldsymbol{\theta}^k)_{k \in \mathbb{N}}$, which converges to a critical point of L . The algorithm is inspired by recent results in non-convex optimization [1], [2], and consists of two main steps that sequentially decrease the value of the objective function.

First, we update the images while keeping the parameters fixed. Let $(\mathbf{x}^k, \boldsymbol{\theta}^k)$ be the estimates obtained after k iterations, and $(\lambda_x^k)_{k \in \mathbb{N}} > 0$ be a decreasing sequence. The next estimate satisfies

$$\mathbf{x}^{k+1} \in \operatorname{argmin}_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\theta}^k) + \frac{\lambda_x^k}{2} h(\Psi^\top (\mathbf{x} - \mathbf{x}^k)),$$

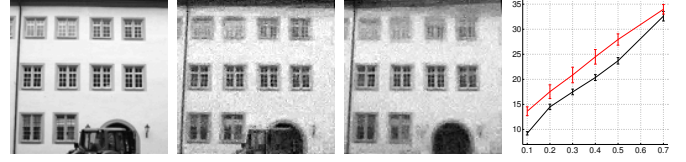


Fig. 1. From left to right: 3rd initial image; 3rd reconstructed image; estimated reference image \mathbf{x}_0 ; SNR vs. m/n for the proposed method (red) and the BP problem (black). The curves represent the mean SNR over 30 simulations, and the vertical lines represent the error at 1 standard deviation.

where h is a smooth approximation of the ℓ_1 -norm. We noticed that the addition of the cost term $\lambda_x^k h(\Psi^\top (\mathbf{x} - \mathbf{x}^k))/2$ produces a reconstruction of the images in coarse-to-fine scales fashion and improves the accuracy of the estimated transformation parameters.

Then, we update the transformation parameters by minimizing a quadratic approximation of $\|\mathbf{A}(\boldsymbol{\theta}) \mathbf{x} - \mathbf{y}\|_2^2$. To simplify notations, we introduce l new functions $Q_j(\boldsymbol{\theta}_j) = \|\mathbf{A}_j S(\boldsymbol{\theta}_j) \mathbf{x}_0^{k+1} + \mathbf{A}_j \mathbf{x}_j^{k+1} - \mathbf{y}\|_2^2$, with $j = 1, \dots, l$. Let $\mathbf{I} \in \mathbb{R}^{q \times q}$ be the identity matrix, and $\lambda_\theta > 0$. Assuming that the entries of $S(\boldsymbol{\theta}_j)$ are differentiable with respect to $\boldsymbol{\theta}_j$, the next estimates $\boldsymbol{\theta}_j^{k+1}$ is chosen as the minimizer of

$$P_j(\boldsymbol{\theta}_j) = \nabla Q_j(\boldsymbol{\theta}_j^k)^\top (\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^k) + (\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^k)^\top \frac{\mathbf{H}_j^k + 2^i \lambda_\theta \mathbf{I}}{2} (\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^k),$$

where i is the smallest positive integer such that $Q_j(\boldsymbol{\theta}_j^{k+1}) + \lambda_\theta \|\boldsymbol{\theta}_j^{k+1} - \boldsymbol{\theta}_j^k\|_2^2/2 \leq Q_j(\boldsymbol{\theta}_j^k) + P_j(\boldsymbol{\theta}_j^{k+1})$. In the above equations, the matrix $\mathbf{H}_j^k = 2(\mathbf{A}_j^k)^\top (\mathbf{A}_j^k)$ with $\mathbf{J}_j^k = (\partial_{\theta_{1j}} S(\boldsymbol{\theta}_j^k), \dots, \partial_{\theta_{qj}} S(\boldsymbol{\theta}_j^k)) \mathbf{x}_0^{k+1} \in \mathbb{R}^{n \times q}$.

III. EXPERIMENTS AND CONCLUSION

We test the proposed method using 5 images of the same scene, taken from different viewpoints, and containing occlusions. We generate 5 measurement vectors using the compressed sensing technique of [3]. Fig. 1 shows the 3rd initial image next to the corresponding reconstructed image from $m = 0.3n$ measurements. The estimated reference image, free of occlusions, is also presented. We also show the curves of the reconstruction SNR as a function of m/n obtained with our method and by solving the Basis Pursuit (BP) problem, which does not benefit from the correlation between measurements. Our method exhibits better reconstruction qualities.

We have presented a method for the joint reconstruction of a set of misaligned images. Our algorithm is an alternating descent method that produces a sequence converging to a critical point of L . Experiments show that the method correctly estimates the underlying reference image \mathbf{x}_0 , is robust to occlusions, and benefits from the inter-correlation between measurements.

REFERENCES

- [1] Attouch et al., “Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-lojasiewicz inequality,” *Math. Oper. Res.*, vol. 35(2), pp. 438–457, 2010.
- [2] Attouch et al., “Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward backward splitting, and regularized gauss-seidel methods,” *Math. Program.*, 2011.
- [3] Puy et al., “Universal and efficient compressed sensing by spread spectrum and application to realistic fourier imaging techniques,” *EURASIP J. Adv. Signal Process.*, vol. 2012(6), 2012.

Femto-Photography and Looking Around the Corners

Ramesh Raskar*

* Media Lab, Massachusetts Institute of Technology, 75 Amherst St, Cambridge, MA 02139, USA

Abstract—Can we look around corners beyond the line of sight? Our goal is to exploit the finite speed of light to improve image capture and scene understanding. New theoretical analysis coupled with emerging ultra-high-speed imaging techniques can lead to a new source of computational visual perception. We are developing the theoretical foundation for sensing and reasoning using Femto-photography and transient light transport, and experimenting with scenarios in which transient reasoning exposes scene properties that are beyond the reach of traditional computer vision. (Joint work with a large team, see <http://raskar.info/femto>)

Compressive Sensing with Structured Random Matrices

Holger Rauhut*

* Hausdorff Center for Mathematics & Institute for Numerical Simulation, University of Bonn, Endenicher Allee 60, 53115 Bonn, Germany.

Abstract—Compressive sensing predicts that sparse signals can be recovered efficiently from what was previously considered incomplete linear information using efficient algorithms. While optimal theoretical guarantees are achieved for (unstructured) Gaussian random measurement matrices, practical applications demand for more structure both due to modeling and computational reasons. We discuss several types of structured random matrices motivated by various signal processing applications as well as the corresponding available theoretical recovery guarantees.

I. INTRODUCTION

Compressive sensing considers the recovery of a signal $x \in \mathbb{C}^N$ from measured linear data $y = Ax \in \mathbb{C}^m$, where $A \in \mathbb{C}^{m \times N}$ and $m \ll N$. Without additional information recovery is clearly impossible in this scenario. Compressive sensing exploits that many real-world signals can be approximated well by sparse vectors (possibly after a basis transformation). In mathematical terms, x is called s -sparse $\|x\|_0 = \#\{\ell : x_\ell \neq 0\} \leq s$ with $s \ll N$.

The naive approach for recovering x from $y = Ax$, namely the ℓ_0 -minimization problem $\min_z \|z\|_0$ subject to $Az = y$ is unfortunately NP-hard, and therefore several alternatives have been introduced including the ℓ_1 -minimization problem

$$\min_z \|z\|_1 \quad \text{subject to } Az = y,$$

where $\|z\|_1 = \sum_\ell |z_\ell|$.

A by-now classical concept for providing recovery guarantees for a measurement matrix A is the restricted isometry property (RIP). We define the restricted isometry constants δ_s as the smallest number such that

$$(1 - \delta_s)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s)\|x\|_2^2 \quad \text{for all } s\text{-sparse } x.$$

If $\delta_{2s} < \delta_0$ for an appropriate $\delta_0 < 1$ then ℓ_1 -minimization and various other algorithms stably reconstruct x from $y = Ax$.

So far optimal recovery guarantees are only known for random matrices. A suitably scaled $m \times N$ Gaussian random matrix (having i.i.d. $\mathcal{N}(0, 1)$ entries) satisfies $\delta_s \leq \delta$ if

$$m \geq C\delta s \log(N/s)$$

and therefore allows stable s -sparse recovery via ℓ_1 -minimization (and further algorithms) in the parameter regime.

While Gaussian random matrices provide optimal recovery guarantees (concerning optimality see e.g. [1] and references therein) they are less appealing from a practical point of view. Often the matrix models a physical measurement device which allows to inject only a limited amount of randomness. This motivates to study random matrices with more structure than Gaussian random matrices.

An important example arises from random sampling. Let $\psi_1, \dots, \psi_N : \mathcal{D} \rightarrow \mathbb{C}$ be a set of functions that are orthogonal with respect to some probability measure ν , i.e., $\int_{\mathcal{D}} \psi_j(t) \overline{\psi_k(t)} d\nu(t) = \delta_{j,k}$, and which possess a uniform L^∞ -bound,

$$\sup_{j=1, \dots, N} \|\psi_j\|_\infty \leq K$$

for some suitable constant $K \geq 1$ (ideally independent of N). We consider functions of the form

$$f(t) = \sum_{j=1}^N x_j \psi_j(t)$$

which we call s -sparse if $\|x\|_0 \leq s$. Given sampling points t_1, \dots, t_m the task is to reconstruct f from the sample values $y_j = f(t_j)$, $j = 1, \dots, m$. Introducing the sampling matrix $A \in \mathbb{C}^{m \times N}$ with entries $A_{j,k} = \psi_k(t_j)$, we can write $y = Ax$, so that we arrive at a compressive sensing problem with this particular A . Choosing the sampling points t_1, \dots, t_m independently at random according to the probability measure ν makes A a structured random matrix. It is known, see e.g. [2] and references therein, that A satisfies the RIP in the sense that $\delta_s \leq \delta$ with high probability provided the number of samples satisfies $m \geq C\delta s \log^4(N)$.

An important example is the trigonometric system $\psi_j(t) = e^{2\pi i j \cdot t}$, arising for instance in MRI. With a reweighting trick also orthogonal polynomial systems [3] and spherical harmonic expansions [4] are covered, the latter being important in various astronomical signal processing applications.

Another type of structured random matrix arises from subsampled random convolutions, that is,

$$Ax = R_\Theta(b * x), \quad \Theta \subset \{1, \dots, N\}, \# \Theta = m,$$

where b is a subgaussian random vector (for instance, the $b_j = \pm 1$ being independent Rademacher variables), R_Θ is the restriction of a vector in \mathbb{C}^N to its entries in Θ , and $b * x$ is the convolution of b with x . The set Θ may be arbitrary, for instance, of the form $\Theta = \{L, 2L, 3L, \dots\}$ for some integer L . Such a partial random circulant matrix arises for instance in radar applications or compressive coded aperture imaging. A recent result [5] states that A satisfies the RIP with high probability in the parameter regime $m \geq Cs \log^2(s) \log^2(N)$.

Further structured random matrices of interest include random Gabor synthesis matrices [6], [5] and random scattering matrices arising from remote sensing [7].

REFERENCES

- [1] S. Foucart, A. Pajor, H. Rauhut, and T. Ullrich, “The Gelfand widths of ℓ_p -balls for $0 < p \leq 1$,” *J. Complexity*, vol. 26, no. 6, pp. 629–640, 2010.
- [2] H. Rauhut, “Compressive Sensing and Structured Random Matrices,” in *Theoretical Foundations and Numerical Methods for Sparse Recovery*, ser. Radon Series Comp. Appl. Math., M. Fornasier, Ed. deGruyter, 2010, vol. 9, pp. 1–92.
- [3] H. Rauhut and R. Ward, “Sparse Legendre expansions via ℓ_1 -minimization,” *J. Approx. Theory*, vol. 164, no. 5, p. 517533, 2012.
- [4] —, “Sparse recovery for spherical harmonic expansions,” in *Proc. SampTA 2011, Singapore*, 2011.
- [5] F. Krahmer, S. Mendelson, and H. Rauhut, “Suprema of chaos processes and the restricted isometry property,” *preprint*, 2012.
- [6] H. Rauhut and G. E. Pfander, “Sparsity in time-frequency representations,” *J. Fourier Anal. Appl.*, vol. 16, no. 2, pp. 233–260, 2010.
- [7] M. Hügel, H. Rauhut, and T. Strohmer, “Remote sensing via ℓ_1 -minimization,” *Preprint*, 2012.

Refined support and entropic uncertainty inequalities

Benjamin Ricaud*, Bruno Torr sani†

* Signal Processing Laboratory 2, Ecole Polytechnique F d rale de Lausanne (EPFL), Station 11, 1015 Lausanne, Switzerland

† LATP, Aix-Marseille Univ/CNRS/Centrale Marseille, UMR7353, 39 rue Joliot-Curie, 13453 Marseille cedex 13, France

Abstract—The uncertainty principle expresses the limitation for a function to be simultaneously concentrated or sparse in two different representations. We generalize some uncertainty inequalities (support and entropy based) from bases to frames and we refine them.

I. INTRODUCTION

The uncertainty principle expresses the limitation for a function to be simultaneously concentrated or sparse in two different representations. The ingredients of this principle are: a signal in a Hilbert space, different representations without loss of information (projection on bases or frames), a measure of concentration (variance, entropy, ℓ^p -norm...). We generalize the Elad-Bruckstein ℓ^0 inequality and entropic inequalities from bases to frames and suggest refined bounds by introducing the r -coherence (see below). Let us first introduce some notations and key quantities.

Coherence: Let $r \in [1, 2]$, let r' be such that $1/r + 1/r' = 1$. The mutual coherence of order r of two frames $\{u_k\}_k$ and $\{v_l\}_l$ is defined by

$$\mu_r(\mathcal{U}, \mathcal{V}) = \sup_{\ell} \left(\sum_k |\langle u_k, v_{\ell} \rangle|^{r'} \right)^{r/r'}, \quad (1)$$

$r = 1$ is the standard mutual coherence $\mu(\mathcal{U}, \mathcal{V})$.

Frame: $\mathcal{U} = \{u_k, k \in \Lambda\}$ and $\mathcal{V} = \{v_{\ell}, \ell \in \Lambda\}$ two countable frames for the Hilbert space \mathcal{H} . We denote by $A_{\mathcal{U}}, B_{\mathcal{U}}$ and $A_{\mathcal{V}}, B_{\mathcal{V}}$ the corresponding frame bounds, i.e. we have for all $x \in \mathcal{H}$, $A_{\mathcal{U}}\|x\|^2 \leq \sum_k |\langle x, u_k \rangle|^2 \leq B_{\mathcal{U}}\|x\|^2$ and $A_{\mathcal{V}}\|x\|^2 \leq \sum_{\ell} |\langle x, v_{\ell} \rangle|^2 \leq B_{\mathcal{V}}\|x\|^2$.

R nyi entropies: In physics and information theory, entropy is a measure of disorder, or information content. For $x \in \mathcal{H}$, let $a_k = \langle x, u_k \rangle$ and $\tilde{a} = a/\|a\|_2$. Given $\alpha \in [0, \infty]$, the R nyi entropy reads

$$R_{\alpha}(a) = \frac{1}{1-\alpha} \ln(\|\tilde{a}\|_2^{2\alpha}). \quad (2)$$

The limit $\alpha \rightarrow 1$ gives $S(a)$ the Shannon entropy. Also, Note: $R_0(a) = \ln \|a\|_0$.

II. REFINED ELAD-BRUCKSTEIN ℓ^0 INEQUALITIES

The classical Elad-Bruckstein ℓ^0 inequality [1] gives a lower bound for the product of support sizes of two orthonormal basis representations a and b of a single vector:

$$\|a\|_0 \cdot \|b\|_0 \geq 1/\mu_1^2. \quad (3)$$

The inequality is extended to the frame case and generalized as follows.

Theorem 1: Let \mathcal{U} and \mathcal{V} be two frames of the Hilbert space \mathcal{H} . Let $\tilde{\mathcal{U}}, \tilde{\mathcal{V}}$ be their respective dual frames. For any $x \in \mathcal{H}$, $x \neq 0$, denote by $a = Ux$ and $b = Vx$ the analysis coefficients of x with respect to these two frames. For all $r \in [1, 2]$, coefficients a and b satisfy the uncertainty inequality

$$\|a\|_0 \cdot \|b\|_0 \geq 1/\mu_*(\mathcal{U}, \tilde{\mathcal{U}}, \mathcal{V}, \tilde{\mathcal{V}})^2 \quad (4)$$

where $\mu_*(\mathcal{U}, \tilde{\mathcal{U}}, \mathcal{V}, \tilde{\mathcal{V}}) = \inf_{r \in [1, 2]} \sqrt{\mu_r(\tilde{\mathcal{U}}, \mathcal{V}) \mu_r(\mathcal{V}, \tilde{\mathcal{U}})}$.

Sketch of proof:

$$\begin{aligned} \|b\|_{\infty} &= \sup_{\ell} |\langle x, v_{\ell} \rangle| = \sup_{\ell} \left| \left\langle \sum_k a_k \tilde{u}_k, v_{\ell} \right\rangle \right| \leq \\ &\leq \sup_{\ell} \sum_k |a_k| |\langle \tilde{u}_k, v_{\ell} \rangle| \leq \|a\|_r \mu_r(\tilde{\mathcal{U}}, \mathcal{V})^{1/r}, \end{aligned}$$

and

$$\|a\|_r^r \leq \|a\|_0 \|a\|_{\infty}^r \leq \|a\|_0 \|b\|_r^r \mu_r(\tilde{\mathcal{V}}, \mathcal{U}).$$

Use the same estimate on $\|b\|_r$. ■

In addition: 1) For all $r \in [1, 2]$, the inequality can only be sharp if $|a|$ and $|b|$ are constant on their support. 2) In some cases $\mu_r \leq \mu_1$.

III. ENTROPIC INEQUALITIES

Uncertainty inequalities involving entropy measures have been derived in several different contexts (see [2], [3], [4] for example). Let \mathcal{U} and \mathcal{V} be two frames of the Hilbert space \mathcal{H} . Let $\tilde{\mathcal{U}}, \tilde{\mathcal{V}}$ be their respective dual frames. Let us introduce: $\rho(\mathcal{U}, \mathcal{V}) = \sqrt{\frac{B_{\mathcal{V}}}{A_{\mathcal{U}}}}$,

$$\sigma(\mathcal{U}, \mathcal{V}) = \sqrt{\frac{B_{\mathcal{U}} B_{\mathcal{V}}}{A_{\mathcal{U}} A_{\mathcal{V}}}} \geq 1 \text{ and } \nu_r(\mathcal{U}, \tilde{\mathcal{U}}, \mathcal{V}) = \frac{\mu_r(\tilde{\mathcal{U}}, \mathcal{V})}{\rho(\mathcal{U}, \mathcal{V})^r}.$$

We then have the following theorem, which can be seen as a frame generalization of the Maassen-Uffink uncertainty inequality [4], [3]:

Theorem 2: Let \mathcal{H} be a separable Hilbert space, let \mathcal{U} and \mathcal{V} be two frames of \mathcal{H} , and let $\tilde{\mathcal{U}}$ and $\tilde{\mathcal{V}}$ denote corresponding dual frames. Let $r \in [1, 2]$. For all $\alpha \in [r/2, 1]$, let $\beta = \alpha(r-2)/(r-2\alpha) \in [1, \infty]$. For $x \in \mathcal{H}$, denote by a and b the sequences of analysis coefficient of x with respect to \mathcal{U} and \mathcal{V} . Then the R nyi entropies satisfy the following bound:

$$(2-r)R_{\alpha}(a) + rR_{\beta}(b) \geq -2 \ln(\nu_r(\mathcal{U}, \tilde{\mathcal{U}}, \mathcal{V})) - \frac{2r\beta}{\beta-1} \ln(\sigma(\mathcal{U}, \mathcal{V})) \quad (5)$$

Sketch of proof: Use the Riesz-Thorin interpolation theorem between $L^2 \rightarrow L^2$ and $L^1 \rightarrow L^{\infty}$ and apply the log. to the L^p -norm. ■

Corollary 1: Assume \mathcal{U} and \mathcal{V} are tight frames, and let $r \in [1, 2]$:

1) For all $\alpha \in [r/2, 1]$, with $\beta = \alpha(r-2)/(2\alpha-r) \in [1, \infty]$

$$(2-r)R_{\alpha}(a) + rR_{\beta}(b) \geq -2 \ln(\nu_r(\mathcal{U}, \tilde{\mathcal{U}}, \mathcal{V})). \quad (6)$$

2) $S(a) + S(b) \geq -2 \ln(\mu_*(\mathcal{U}, \tilde{\mathcal{U}}, \mathcal{V}, \tilde{\mathcal{V}}))$.

REFERENCES

- [1] M. Elad and A. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Transactions On Information Theory*, vol. 48, pp. 2558–2567, 2002.
- [2] W. Beckner, "Inequalities in Fourier analysis on \mathbb{R}^n ," *Proc. Nat. Acad. Sci. USA*, vol. 72, no. 2, pp. 638–641, February 1975.
- [3] A. Dembo, T. Cover, and J. Thomas, "Information theoretic inequalities," *Information Theory, IEEE Transactions on*, vol. 37, no. 6, pp. 1501–1518, nov 1991.
- [4] H. Maassen and J. Uffink, "Generalized entropic uncertainty relations," *Physical Review Letters*, vol. 60, no. 12, p. 11031106, 1988.

Adaptive compressive noncoherent change detection: An AMP-based approach

Philip Schniter

Dept. of ECE, The Ohio State University, Columbus, OH 43210, Email: schniter@ece.osu.edu

Abstract—We propose a turbo approximate message passing (AMP) algorithm to detect spatially clustered changes in signal magnitude, relative to a reference signal, from compressive linear measurements. We then show how the Gaussian posterior approximations generated by this scheme can be used for mutual-information based measurement kernel adaptation. Numerical simulations show excellent performance.

I. SUMMARY

A. Compressive noncoherent change detection

In *change detection*, one observes noisy linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \in \mathbb{C}^M$ of a signal $\mathbf{x} \in \mathbb{C}^N$ and aims to detect changes in \mathbf{x} relative to a known reference signal $\mathbf{r} \in \mathbb{C}^N$. Here, \mathbf{A} represents a known measurement kernel and \mathbf{w} represents white Gaussian noise.

Our focus is *noncoherent* change detection, where the phase difference between \mathbf{r} and \mathbf{x} may be significant even in the absence of a material change. In this case, the goal is to detect *changes in magnitude* between \mathbf{x} and \mathbf{r} . An example application arises in radar, where small (e.g., wind-induced) movements in foliage can result in a large independent phase differences in each pair (x_n, r_n) even when the material present in pixel n has not changed.

We are particularly interested in the *compressive* case, where the number of measurements, M , is less than the signal length, N . Although we assume that the magnitude changes $|\mathbf{x}| - |\mathbf{r}|$ are sparse, and possibly even structured-sparse, we do not assume that the signals \mathbf{x} and \mathbf{r} themselves are sparse in a known basis, nor is their difference $\mathbf{x} - \mathbf{r}$. Note that, if (an estimate of) \mathbf{x} was available, then standard techniques [1] could be applied to detect changes between $|\mathbf{x}|$ and $|\mathbf{r}|$. However, we do not observe \mathbf{x} , and the lack of sparsity in \mathbf{x} (and $\mathbf{x} - \mathbf{r}$) prevents the use of standard compressed sensing techniques to recover \mathbf{x} from \mathbf{y} . Thus, the problem is somewhat challenging.

Our approach exploits that fact that, under the sparse magnitude-change assumption, $|\mathbf{r}|$ does provide information about $|\mathbf{x}|$ that can aid in compressive recovery of \mathbf{x} and—more importantly—*joint change detection and signal recovery*. For this, we model

$$x_n = s_n c_n + (1 - s_n)(r_n e^{j\theta_n} + d_n), \quad (1)$$

where $s_n \in \{0, 1\}$ indicates the presence of a change, $c_n \in \mathbb{C}$ represents the changed pixel value, $\theta_n \in [0, 2\pi)$ represents an unknown phase rotation, and $d_n \in \mathbb{C}$ represents a small deviation allowed in an “unchanged” pixel. We then assign the priors

$$\begin{aligned} c_n &\sim \mathcal{CN}(0, \nu^r) \text{ i.i.d with } \nu^r = \frac{1}{N} \sum_{n=1}^N |r_n|^2 \\ \theta_n &\sim \mathcal{U}[0, 2\pi) \text{ i.i.d} \\ d_n &\sim \mathcal{CN}(0, \nu^d) \text{ i.i.d with } \nu^d \ll \nu^r \\ s_n &\sim \text{Markov}, \end{aligned} \quad (2)$$

where the Markov property on $\{s_n\}$ captures the fact that changes are often spatially clustered. Finally, we jointly infer the change pattern \mathbf{s} and the signal \mathbf{x} using the turbo extension [2] of the Bayesian approximate message passing (AMP) algorithm [3]. To our knowledge, the use of AMP with a signal prior of this form is novel.

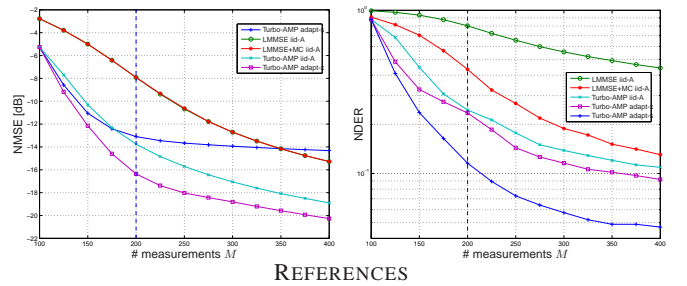
With support from NSF grant CCF-1018368 and DARPA/ONR grant N66001-10-1-4090.

B. Measurement adaptation

We now allow the aforementioned approach multiple *adaptive* measurement steps, building on the work in [4]. In step $t = 1, \dots, T$, the detector collects measurements $\mathbf{y}_t = \mathbf{A}_t \mathbf{x} + \mathbf{w}_t \in \mathbb{C}^{M_t}$ using a kernel \mathbf{A}_t optimized around the uncertainty of \mathbf{x} (or \mathbf{s}) that remains from inference based on the cumulative previous measurements $\underline{\mathbf{y}}_{t-1} \triangleq [\mathbf{y}_1^T, \dots, \mathbf{y}_{t-1}^T]^T$. When optimizing \mathbf{A}_t for the recovery of \mathbf{x} , [4] suggested to maximize the mutual information (MI) between *Gaussian approximations* of the random vectors $\mathbf{x} \sim p(\mathbf{x}|\underline{\mathbf{y}}_{t-1})$ and $\mathbf{y}_t \sim p(\mathbf{y}_t|\underline{\mathbf{y}}_{t-1}; \mathbf{A}_t)$. Indeed, when \mathbf{x} and \mathbf{y}_t are jointly Gaussian, [4] established that the MI-maximizing \mathbf{A}_t is computable using eigendecomposition and waterfilling. Conveniently, the necessary Gaussian approximation on \mathbf{x} is an output of turbo AMP. For *s*-adaptive kernel design, we now propose a similar approach based on a Gaussian approximation of $\mathbf{s} \sim p(\mathbf{s}|\underline{\mathbf{y}}_{t-1})$.

C. Numerical results

The left plot shows the normalized mean-squared error (NMSE) in recovering $\mathbf{x} \in \mathbb{C}^{200}$ versus cumulative number of measurements M , under 15 dB SNR and $\nu^d = 0.001$, averaged over 1000 realizations. All quantities were drawn according to (2), with the binary Markov chain for \mathbf{s} activating 10% changes on average, clustered with an average run-length of 10. There, turbo-AMP with MI- \mathbf{x} kernel adaptation performed best, approximately 2dB better than turbo-AMP with i.i.d-Gaussian \mathbf{A} , while LMMSE estimation of \mathbf{x} with i.i.d-Gaussian \mathbf{A} performed significantly worse. The right plot shows the corresponding normalized detection error rate (NDER), where turbo-AMP with MI- \mathbf{s} kernel adaptation performed best, and significantly better than Bayes-optimal change detection using LMMSE- \mathbf{x} , even when change clustering was exploited. Although turbo-AMP with MI- \mathbf{s} kernel adaptation did not work well for \mathbf{x} -recovery, we did not expect it to, since it was optimized for change detection.



REFERENCES

- [1] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer, 2nd ed., 1994.
- [2] P. Schniter, “Turbo reconstruction of structured sparse signals,” in *Proc. Conf. Inform. Science & Syst.*, (Princeton, NJ), pp. 1–6, Mar. 2010.
- [3] D. L. Donoho, A. Maleki, and A. Montanari, “Message passing algorithms for compressed sensing: I. Motivation and construction,” in *Proc. Inform. Theory Workshop*, (Cairo, Egypt), pp. 1–5, Jan. 2010.
- [4] P. Schniter, “Exploiting structured sparsity in Bayesian experimental design,” in *Proc. IEEE Workshop Comp. Adv. Multi-Sensor Adaptive Process.*, (San Juan, Puerto Rico), pp. 357–360, Dec. 2011.

Convergent and Scalable Algorithms for Expectation Propagation Approximate Bayesian Inference

Matthias W. Seeger*

* Probabilistic Machine Learning Laboratory Ecole Polytechnique Fédérale de Lausanne
INR 112, Station 14, CH-1015 Lausanne
matthias.seeger@epfl.ch

Abstract—We detail novel, provably convergent algorithms for variational Bayesian inference in large scale applications, such as image deconvolution (non-blind or blind) and magnetic resonance imaging.

Popular sparse reconstruction algorithms can be seen as maximum a posteriori estimation in sparse generalized linear graphical models. The move from point estimation to Bayesian inference, namely quantifying the posterior (Boltzmann) distribution through its low order moments (mean, (co)variances) and log partition function, has many potential advantages, such as robust reconstruction by minimum mean square estimation (posterior mean instead of mode), automatic calibration of free parameters such as noise variance, convolution kernel or prior parameters without costly cross-validation, scoring predictive reliability by posterior variances, and advanced decision making (Bayesian experimental design, Bayesian active learning). However, inference beyond MAP is intractable and has to be approximated. Even though inference is a harder problem than estimation, applications demand scalability of methods comparable to MAP.

In order to meet such demands, variational approximations relax Bayesian inference to tractable optimization problems. On most large scale models, they can be substantially faster than Markov chain Monte Carlo methods, the mainstream approach to Bayesian inference in statistics today. Expectation propagation [1] is maybe the most versatile and accurate variational inference relaxation known today, a direct generalization of loopy belief propagation to models with continuous and discrete variables and potentially densely coupled graphs. It is rooted in the TAP approximation from statistical physics and can be seen as (more costly) improvement of approximate message passing (AMP) methods, even though it predates the latter. However, the basic coordinate update “message passing” EP algorithm used in most applications today does not scale to large models used in imaging. Parallel message passing variants frequently fail to converge on models with hard sparsity priors (such as spike and slab). In general, coordinate-wise message passing does not make use of modern convex optimization methods such as Newton-Raphson or first order methods.

In this talk, we describe a novel EP algorithm [2], which is both provably convergent and can be scaled up to large densely connected models, drawing a connection between the double loop algorithm of Oppé and Winther [3] and earlier work by the author on scalable algorithms for simpler relaxations. Even for problems of moderate size (such as Gaussian process classification with a few thousand training points), the new algorithm converges at least an order of magnitude faster than the standard (sequential) EP algorithm. Time permitting, we will show recent results on models with hard sparsity prior potentials (spike and slab).

REFERENCES

- [1] T. Minka, “Expectation propagation for approximate Bayesian inference,” in *Uncertainty in Artificial Intelligence 17*, J. Breese and D. Koller, Eds. Morgan Kaufmann, 2001.
- [2] M. Seeger and H. Nickisch, “Fast convergent algorithms for expectation propagation approximate Bayesian inference,” in *Workshop on Artificial Intelligence and Statistics 14*, G. Gordon and D. Dunson, Eds., 2011.
- [3] M. Oppé and O. Winther, “Expectation consistent approximate inference,” *Journal of Machine Learning Research*, vol. 6, pp. 2177–2204, 2005.

Recovery of sparse translation-invariant signals with continuous basis pursuit

Eero Simoncelli*, Chaitu Ekanadham* and Daniel Tranchina*

* New York University

We consider the problem of decomposing a signal into a linear combination of features, each a continuously translated version of one of a small set of elementary features. Although these constituents are drawn from a continuous family, most current signal decomposition methods rely on a finite dictionary of discrete examples selected from this family (e.g., a set of shifted copies of a set of basic waveforms), and apply sparse optimization methods to select and solve for the relevant coefficients. Here, we generate a dictionary that includes auxiliary interpolation functions that approximate translates of features via adjustment of their coefficients. We formulate a constrained convex optimization problem, in which the full set of dictionary coefficients represent a linear approximation of the signal, the auxiliary coefficients are constrained so as to only represent translated features, and sparsity is imposed on the non-auxiliary coefficients using an L1 penalty. The well-known basis pursuit denoising (BP) method may be seen as a special case, in which the auxiliary interpolation functions are omitted, and we thus refer to our methodology as continuous basis pursuit (CBP). We develop two implementations of CBP for a one-dimensional translationinvariant source, one using a first-order Taylor approximation, and another using a form of trigonometric spline. We examine the tradeoff between sparsity and signal reconstruction accuracy in these methods, demonstrating empirically that trigonometric CBP substantially outperforms Taylor CBP, which in turn offers substantial gains over ordinary BP. In addition, the CBP bases can generally achieve equally good or better approximations with much coarser sampling than BP, leading to a reduction in dictionary dimensionality.

On localization and uncertainty in graph based representations

Pierre Vandergheynst*

* Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

Graph theoretical modelling of high dimensional datasets or signals is slowly emerging as a versatile tool, merging together elements of machine learning, signal processing but also geometrical insights. Much work remains to be done for understanding the fundamental limits of these models, though. In this talk, we will discuss the interplay between localization and uncertainty in some graph based representations with a particular emphasis on their role in graph based harmonic analysis and in the emergence of a body of methodological methods that hint at new ways of processing signal on graphs.

Overcomplete Joint Sparsity Model for Dictionary Selection

Mehrdad Yaghoobi*, Laurent Daudet† and Mike E. Davies*

* Institute for Digital Communications (IDCom), The University of Edinburgh, EH9 3JL, UK.

† Paris Diderot University / IUF, Institut Langevin, 1, rue Jussieu 75005 Paris, France.

Abstract—The problem of dictionary selection for linear sparse approximation will be revisited in this paper. A dictionary for sparsifying a class of signals is often selected based upon the domain knowledge or using some exemplar signals. We present a new exemplar based approach for the dictionary selection, which combines the two approaches. In this framework, a large set of atoms is also given as the mother dictionary and the task is to choose a subset of the atoms, which suits the given exemplars. The new dictionary learning problem is initially formulated as a new type of *joint sparsity model*, which differs from the standard joint sparsity model. A simple gradient based algorithm will then be presented here to practically solve the optimisation problem. An important advantage of the new formulation is the scalability of the learning algorithm. The new dictionary selection paradigm is here examined with some synthetic experiments¹.

I. INTRODUCTION

Let $\mathbf{Y} = [\mathbf{y}_l]_{l \in [1, L]}$ be a matrix made by training samples $\mathbf{y}_l \in \mathbb{R}^m$ and $\Phi = [\phi_i]_{i \in [1, n]}$ be a mother dictionary of normalised atoms $\phi_i \in \mathbb{R}^m$. We assume that the generative dictionary $\mathbf{D} \in \mathbb{R}^{m \times p}$, $m \leq p$ is made by a subset selection of the atoms in Φ , i.e. $\mathbf{D} = [\phi_i]_{i \in \mathcal{J}}$ where $\mathcal{J} \subset [1, n]$ and $|\mathcal{J}| = p < n$. We assume that each \mathbf{y}_l is *approximately* generated by a k -sparse coefficient vector γ_l , by $\mathbf{y}_l \approx \mathbf{D}\gamma_l$. The problem of *optimal dictionary selection*, which has a close relationship with the framework introduced in [1], is to find index set \mathcal{J} , given \mathbf{Y} , Φ , p and k . Let $\mathbf{X} \in \mathbb{R}^{m \times L}$ be a coefficient matrix and $f_{\mathcal{J}}(i) : [1, p] \mapsto [1, n]$ be the mapping that assigns the corresponding atom index of Φ to the i^{th} component of γ_l . By assigning $\{\mathbf{x}_l\}_{f_{\mathcal{J}}(i)} \leftarrow \{\gamma_l\}_i$, $\forall i \in [1, p]$, $\forall l \in [1, L]$, while the other elements of \mathbf{X} are set to zero, the generative model can be reformulated as, $\mathbf{Y} \approx \Phi\mathbf{X}$. \mathbf{X} is k -sparse in each column and p -row-sparse, i.e. only p rows of \mathbf{X} have non-zero components. It thus lies in the intersection of $\mathcal{K} := \{\mathbf{X} \in \mathbb{C}^{m \times L} : \|\mathbf{x}_l\|_0 \leq k, \forall l \in [1, L]\}$ and $\mathcal{P} := \{\mathbf{X} \in \mathbb{C}^{m \times L} : \|\mathbf{X}\|_{0, \infty} \leq p\}$, where $\|\mathbf{X}\|_{0, \infty} = \|\nu\|_0$, with $\{\nu\}_i := \|\mathbf{x}^{(i)}\|_{\infty}$ and $\mathbf{x}^{(i)}$ is the i^{th} row of \mathbf{X} .

Let the representation error be defined as $\psi(\mathbf{X}) := \|\mathbf{Y} - \Phi\mathbf{X}\|_F^2$. The optimal dictionary \mathbf{D} , which can alternatively be indicated by \mathcal{J} , is defined as the sub-dictionary used by the solution of following problem,

$$\min_{\mathbf{X}} \psi(\mathbf{X}), \text{ s. t. } \mathbf{X} \in \mathcal{K} \cap \mathcal{P}. \quad (1)$$

This optimisation problem can generally have unbounded solutions. However, if the null-space of Φ does not have any non-trivial element in $\mathcal{K} \cap \mathcal{P}$, the solutions would be bounded [2].

II. DICTIONARY SELECTION ALGORITHM

We use a gradient based method to approximately solve (1), which iteratively updates the current solution $\mathbf{X}^{[n]}$, in the negative gradient direction of $\psi(\mathbf{X})$ at $\mathbf{X}^{[n]}$ and map onto $\mathcal{K} \cap \mathcal{P}$. The gradient of ψ , which is noted by \mathbf{G} , can be found by, $\mathbf{G} = 2\Phi^T(\Phi\mathbf{X} - \mathbf{Y})$.

An important part of the gradient based methods is how to select the step size. An efficient step size selection technique was used in

¹This work was supported by EU FP7, FET-Open grant number 225913 and EPSRC grant EP/J015180/1.

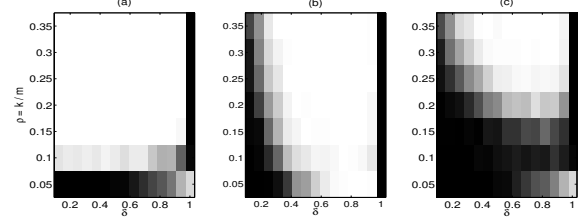


Fig. 1. Phase transitions using, (a) \mathcal{K} , (b) \mathcal{P} and (c) $\mathcal{K} \cap \mathcal{P}$ as the admissible sets. The black area indicates successful dictionary recovery.

[3] for the sparse approximations of k -sparse signals. Similarly, we calculate the initial step size using the gradient matrix \mathbf{G} , constrained to the support, by $\mu = \frac{1}{2} \frac{\mathbf{G}_S^H \Phi^H \Phi \mathbf{G}_S}{\mathbf{G}_S^H \mathbf{G}_S}$ where $\mathbf{G}_S \in \mathbb{R}^{m \times L}$ is \mathbf{G} masked by the support of \mathbf{X} , S .

Mapping the updated solution onto $\mathcal{K} \cap \mathcal{P}$ is done by, first projecting onto \mathcal{P} , then projecting onto \mathcal{K} . Although this mapping is **not** a projection onto the intersection of \mathcal{K} and \mathcal{P} , we can show that alternating projections onto \mathcal{P} and \mathcal{K} , find a point on their intersection.

The introduced gradient mapping algorithm may increase the objective with the initial step size. We should thus reduce the stepsize to guarantee monotonic decrease of the objective in each update. We have shown the convergence property of such gradient mapping algorithm in [2].

III. SIMULATION RESULT OF A CASE STUDY

A dictionary $\Phi \in \mathbb{R}^{20 \times 80}$ was randomly generated by normalising a zero mean, unit variance normal distribution. The mother dictionary \mathbf{D} was randomly selected from Φ . A set of training matrix \mathbf{Y} was generated using the generative model and k -sparse coefficient vectors, which were generated with random support and uniformly random magnitudes between 0.2 and 1. This experiment was repeated for various $\delta = p/n$ and $\rho = k/m$, while keeping $m = 20$ and $n = 80$ fixed. To recover the optimal generative dictionary \mathbf{D} , given \mathbf{Y} , p and k , we used the gradient mapping/projection algorithm with three different admissible sets, i.e. \mathcal{K} , \mathcal{P} and $\mathcal{K} \cap \mathcal{P}$, to demonstrate the superiority of the proposed framework.

The phase transitions of the average correct dictionary recovery of 100 trials, have been plotted in Figure 1, with the constraint \mathcal{K} in (a), \mathcal{P} in (b) and the proposed constraint in (c). While the black colour means high exact dictionary recovery, it is clear that the exact recovery in (c) is most probable, with respect to (a) and (b).

REFERENCES

- [1] A. Krause and V. Cevher, "Submodular dictionary selection for sparse representation," in *International Conference on Machine Learning (ICML)*, 2010.
- [2] M. Yaghoobi, L. Daudet, and M. Davies, "Optimal dictionary selection using an overcomplete joint sparsity model," in preparation.
- [3] T. Blumensath and M. Davies, "Normalized iterative hard thresholding: Guaranteed stability and performance," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 298 – 309, April 2010.

Signal processing techniques applied to future radio data.

Filipe B. Abdalla,

University College London, Department of Physics and Astronomy, Gower Place, London, WC1E 6BT.

Abstract—For many years now, several signal processing techniques have been of extensive use by the astronomical community. Several missions analysing the radiation from the Cosmic Microwave Background, including the upcoming Planck experiment, have used signal processing techniques to extract signals which would be otherwise buried deep within the noise and the systematic effects from the experiment. There are many further applications and data sets which will benefit a huge amount from signal processing techniques in the future, most notably in the radio domain. I will discuss in this talk the prospects of using signal processing techniques in simulations of future and upcoming radio data which will be able to significantly improve our capability of measuring the epoch of reionisation as well as measure cosmology and the dark energy component in our Universe.

I. INTRODUCTION

Most of Cosmology today is the study of statistical random fields. This is how we are able to test theories of the formation of structure in our Universe, measure distances via the so called Baryonic Acoustic Oscillations imprinted in the distribution of matter in our Universe and test modifications to gravity theories by looking at how fluctuations in the matter density field grow. To date, we have a huge knowledge about our Universe. We have a working theory of inflation which is yet to be fully tested by has solved already several conceptual problems in cosmology. We have measured very accurately the processes which occur in the early Universe when nucleosynthesis takes place and during recombination. Finally we have mapped extensively the distribution of galaxies in the late time Universe. However serious questions remain: for instance we have not yet seen the signal from the first sources in our Universe and we have no clue what these first sources are. We also have no working theory which explains the current accelerated expansion of the Universe.

II. THE EPOCH OF REIONISATION, COSMOLOGY AND DARK ENERGY VIA 21 CM RADIATION.

Radio telescopes are able to probe the epoch of Reionisation [1] and also measure enough statistical fluctuations to pose a very strong constraint on the properties of dark energy via the 21cm radiation [2], [3]. This radiation which originates from forbidden hyperfine transition of neutral hydrogen can be seen at around 100-200MHz at redshifts of interest for Reionisation studies and around 500-1400MHz at redshifts of interest for dark energy studies. This has been some of the main science motivations for the future Square Kilometre Array which will be one of the most sensitive radio telescopes and will share a site between South Africa and Australia. There are other telescopes which aim to measure signals in these frequency regimes which will also teach us about these two epochs in the history of our Universe.

The Signal from the 21cm radiation is a very faint line signal which will map the anisotropies of density as well as ionisation throughout space and back in time. When it comes to measure the reionisation signal this is likely to be a diffuse signal which will be modulated by the anisotropies of the density fluctuations but also by the ionisation field. For measuring dark energy we are mainly interested on the emission from the neutral hydrogen present in galaxies. Measuring such galaxies with the SKA has been shown to not only be able to

measure dark energy below a percent in accuracy [4], [5] it can also measure the masses of cosmological neutrinos [6] and test general relativity [7]. It is however possible to obtain much faster experiments by degrading the resolution of the radio telescope and mapping the intensity of the line radiation with the so called intensity mapping experiments [8].

III. FOREGROUND SEPARATION

The above science requires good foreground subtraction as there will be a strong continuum emission from our own Galaxy from sources such as synchrotron emission, free-free emission, emission from supernovae remnants as well as extragalactic emission. This has to be separated from the 21cm maps so that an analysis of the clustering as well as the anisotropies during the epoch of reionisation can be interpreted. There has been some attempts of doing this with techniques from signal processing used successfully in CMB experiments in [9], [10] for EoR data and in [11] for intensity mapping data. These are more sophisticated techniques which will gain compared to the simple types of fitting that were being used previously in this area.

I will review in this talk the current techniques which are being used and other potential problems in studying the Epoch of Reionisation, Cosmology and dark energy with future radio telescopes. This should create a good platform where we can see where the signal processing techniques can help extracting these weak signals from the data and also where further work is still required.

REFERENCES

- [1] G. Mellema and et al., “Reionization and the Cosmic Dawn with the Square Kilometre Array,” *ArXiv e-prints*, Sep. 2012.
- [2] C. A. Blake, F. B. Abdalla, S. L. Bridle, and S. Rawlings, “Cosmology with the SKA,” *New Astronomy Reviews*, vol. 48, pp. 1063–1077, Dec. 2004.
- [3] F. B. Abdalla and S. Rawlings, “Probing dark energy with baryonic oscillations and future radio surveys of neutral hydrogen,” *MNRAS*, vol. 360, pp. 27–40, Jun. 2005.
- [4] F. B. Abdalla, C. Blake, and S. Rawlings, “Forecasts for dark energy measurements with future HI surveys,” *MNRAS*, vol. 401, pp. 743–758, Jan. 2010.
- [5] J. Tang, F. B. Abdalla, and J. Weller, “Complementarity of future dark energy probes,” *MNRAS*, vol. 416, pp. 2212–2232, Sep. 2011.
- [6] F. B. Abdalla and S. Rawlings, “Determining neutrino properties using future galaxy redshift surveys,” *MNRAS*, vol. 381, pp. 1313–1328, Nov. 2007.
- [7] A. Raccanelli and et al., “Cosmological measurements with forthcoming radio continuum surveys,” *MNRAS*, vol. 424, pp. 801–819, Aug. 2012.
- [8] T.-C. Chang, U.-L. Pen, K. Bandura, and J. B. Peterson, “An intensity map of hydrogen 21-cm emission at redshift $z \sim 0.8$,” *Nature*, vol. 466, pp. 463–465, Jul. 2010.
- [9] E. Chapman and et al., “Foreground removal using FASTICA: a showcase of LOFAR-EoR,” *MNRAS*, vol. 423, pp. 2518–2532, Jul. 2012.
- [10] —, “The Scale of the Problem : Recovering Images of Reionization with GMCA,” *ArXiv e-prints*, Sep. 2012.
- [11] L. Wolz and et al., “Systematic effects from intensity mapping in cosmology,” *to be submitted*, 2012.

5D image reconstruction of stellar systems

Fabien Baron*, Brian Kloppenborg†, and John Monnier*

* University of Michigan, Department of Astronomy, 500 Church Street, Ann Arbor, MI 48109-1090, USA.

† Max Planck Institute for Radioastronomy, Auf dem Hügel 89 53121 Bonn, Germany.

Abstract—5D interferometric imaging is the capability to image directly astronomical targets in three spatial dimensions, with both the time and wavelength dependencies taken into account during the reconstruction process. We present here recent reconstruction results obtained using wavelet-based compressed sensing on the surface of spheroidal stars modeled in real time using GPU computing. Our optimization approach exploits an hybrid gradient-based and MCMC engine, and makes use of the Bayesian model selection framework to select optimal wavelet bases.

I. INTRODUCTION

The current paradigm of image reconstruction in optical interferometry is overwhelmingly that of monochromatic snapshot imaging on a two-dimensional plane (“2D imaging”). However many time dependent effects can prevent the imaging of the most interesting stellar surfaces (rotating stars, contact binaries, ...). In addition many objects present strong wavelength dependencies, and until recently [1] little work had been done on multi-spectral image reconstruction. Our work intends on solving both these issues for the problem of imaging stellar surfaces.

II. MODELING A 3D SURFACE

The surface of a star may considerably depart from spherical symmetry for various reasons. A fast-rotating star close to its breakup speed will e.g. cause a very visible elongation along its rotation axis[2]. In the case of interacting binaries such as Algol[3] some components may fill their Roche lobe. Our code uses a custom modification of Healpix[4] to describe stellar surfaces. Though Healpix pixels only have equal surfaces on a perfect sphere, this framework remains practical for any spheroid. To each Healpix pixel is associated a spherical coordinate vector, which radius is dynamically computed using the general Roche equation of the gravitational potential [5]. To each Healpix pixel is also associated a temperature: it is here the quantity to reconstruct, in contrast with classic 2D imaging where brightness/fluxes are used.

In addition to this description of static surfaces, any three-dimensional elements may be added to our models.

III. ADDING THE TIME DEPENDENCY

For a single star in solid-body rotation, the rotation parameters (axis orientation and rotation speed) are sufficient to describe the movement. Implementation of differential rotation is under study. For multiple stars, our current code incorporates orbital calculations, including the possibilities of defining hierarchal binary systems. Graphics Processing Units (GPUs) are used to render any 3D star into a 2D surface, bypassing potential numerical complexities due to eclipsing stars.

IV. ADDING WAVELENGTH DEPENDENCIES

Our implementation of the wavelength dependencies is currently rudimentary, with the geometry of the surface currently assumed not to vary with wavelength. SED information and differential quantities dependent on wavelength are taken into account in our χ^2 metric. Wavelength-dependent limb-darkening is added in post-processing on top of the brightness map derived from the intrinsic temperatures.

V. OPENCL AND OPENGL

Both the OpenGL and OpenCL GPU frameworks are used: OpenCL accelerates our vector operations (NFFT, χ^2 , stellar geometry, gradients) and OpenGL dynamically updates the stellar surfaces (rotation, limb-darkening, final rendering).

VI. RECONSTRUCTION AND REGULARIZATION

Interferometric imaging is a regularized maximum likelihood problem. Our code offers a vast choice of regularizations based on sparsity in gradient or wavelet bases. We use the total variation regularizer modified to work on spheroids, as well as a derived $\ell_{0.5}$ regularizer for spot detection. Because most of the regularizers are non-convex, we rely currently on MCMC optimization (simulated annealing and parallel tempering). In multi-wavelength mode, a stack of images is reconstructed, using several types of trans-spectral regularizations to impose smooth image continuity along spectral channels. Optimization is slow with a large number of spectral channels, and thus ADMM techniques [1] are being considered.

VII. BAYESIAN MODEL SELECTION

Our MCMC scheme has been modified to compute the marginal likelihood using the Nested Sampling algorithm [6]. This allows model selection on all hyperparameters and regularizers.

VIII. FUTURE PLANS

Our current code is focused on stellar surface imaging and thus will be extended beyond interferometric imaging to complementary inverse problems such as light-curve inversion [7] and Doppler imaging [8].

REFERENCES

- [1] E. Thiébaud and F. Soulez, “Multi-wavelength imaging algorithm for optical interferometry,” *Proc. SPIE*, vol. 8845, p. 84451C, 2012.
- [2] X. Che, J. D. Monnier, M. Zhao, E. Pedretti, N. Thureau, A. Mérand, T. ten Brummelaar, H. McAlister, S. T. Ridgway, N. Turner, J. Sturmann, and L. Sturmann, “Colder and Hotter: Interferometric Imaging of β Cassiopeiae and α Leonis,” *Astrophysical Journal*, vol. 732, no. 2, p. 68, May 2011.
- [3] F. Baron, J. D. Monnier, E. Pedretti, M. Zhao, G. Schaefer, R. Parks, X. Che, N. Thureau, T. a. ten Brummelaar, H. a. McAlister, S. T. Ridgway, C. Farrington, J. Sturmann, L. Sturmann, and N. Turner, “Imaging the Algol Triple System in the H band with the CHARA Interferometer,” *Astrophysical Journal*, vol. 752, no. 1, p. 20, Jun. 2012.
- [4] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, “HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere,” *Astrophysical Journal*, vol. 622, pp. 759–771, Apr. 2005.
- [5] J. F. Sepinsky, B. Willems, and V. Kalogera, “Equipotential Surfaces and Lagrangian Points in Nonsynchronous, Eccentric Binary and Planetary Systems,” *Astrophysical Journal*, vol. 660, pp. 1624–1635, May 2007.
- [6] J. Skilling, “Nested sampling for general Bayesian computation,” *Bayesian Analysis*, 2006.
- [7] R. O. Harmon and L. J. Crews, “Imaging Stellar Surfaces via Matrix Light-Curve Inversion,” *Astronomical Journal*, vol. 120, pp. 3274–3294, Dec. 2000.
- [8] S. S. Vogt, G. D. Penrod, and A. P. Hatzes, “Doppler images of rotating stars using maximum entropy image reconstruction,” *Astrophysical Journal*, vol. 321, pp. 496–515, Oct. 1987.

2-Channels Finite Resolution Deconvolution

Nicolas Cantale*, Frédéric Courbin* and Georges Meylan*

* Laboratoire d'Astrophysique, École Polytechnique Fédérale de Lausanne (EPFL)

Abstract—We present a new general purpose deconvolution algorithm and its implementation. As an evolution of the MCS algorithm [1], it is implementing the key ideas of the mainstream solutions, like correct sampling and a separate channel for the deconvolution of point sources. But we improved it with major evolutions in the areas of point sources characterization, regularization and simultaneous deconvolution, making it one of the most powerful tool for precise photometry and morphology analysis.

I. INTRODUCTION

During the last decades, image deconvolution was widely studied and successfully applied to astronomical data. It allows the reconstruction of an underlying scene after its alteration by the atmosphere and by the instrumental response of the instrument. However, deconvolution is by construction a mathematically ill-posed problem accepting an infinite number of solutions. For this reason, multiple algorithms have emerged (see [2], [3], [4]), using different priors to constrain the problem and build a likely model. Despite the number of proposed solutions, none are really acceptable in terms of accuracy and fidelity, in particular due to the artifacts they generate. We remedy to this by presenting a recently developed solution, derived from the MCS [1] algorithm.

II. CORRECT SAMPLING

Deconvolution by a classical Point Spread Function (PSF) leads automatically to an infinite resolution. Indeed, a point source in the data should be rendered as a point source in the reconstructed model. However, this is impossible to represent as the model itself is sampled. To bypass this problem, our algorithm implements a key idea, imagined by [1], consisting in deconvolving by a narrower PSF, leading to a correctly sampled model.

III. SOURCE SEPARATION AND CHARACTERIZATION

Our algorithm being a refinement of the MCS algorithm, it also implements the two channels philosophy [5]: we deconvolve simultaneously a background model for the extended sources and an analytical model for the point sources. The adopted solution uses some priors on the continuity of the background model to separate the sources in a very efficient way, allowing a precise separation between the sources and the background, or between two or more sources. It becomes really useful when sources are dominating and close together, like in strong lensing or with a supernova and its host galaxy.

IV. REGULARIZATION

One of the key challenges in deconvolution is to correctly regularize the problem. Our approach is to adopt a new regularization type, based on wavelet denoising. This technique preserves the shape of the underlying model but removes the high frequencies caused by the noise in the data.

V. SIMULTANEOUS DECONVOLUTION

A major contribution of the MCS algorithm was the simultaneous deconvolution: the model is built using several images in parallel. This allows the use of super-resolution, but also the making of light curves in case of variable point sources. The new algorithm also implements this feature, but extends it to a higher level. Instead of using several images from a single instrument (implying the same sampling, aligned data, etc.), we are now able to combine images from different instruments, taken at different resolutions and from different point of view. There is no need to align the images anymore, and thus to interpolate the data. For instance, combining ground-based data with HST data is now possible.

VI. APPLICATIONS

Deconvolution is applicable to almost any kind of astronomical data. We applied this algorithm for the extraction of light curves in strong lensing as well as on FORS2 data of galaxy clusters. Also, the Hubble Ultra Deep Field was entirely deconvolved to illustrate the power of the code with large data, and various applications, like direct imaging of planets, are in progress.

VII. PACKAGING

Our goal is to propose the software as a free, easy to use PYTHON package. It implements only basic PYTHON modules but supports GPU computing for the FFTs, via the CUDA libraries.

VIII. CONCLUSION

We presented a new deconvolution algorithm, implementing all the key features necessary for an accurate image reconstruction: correct sampling, efficient regularization, very good point sources handling and data combination from multiple instruments.

REFERENCES

- [1] P. Magain, F. Courbin, and S. Sohy, "Deconvolution with correct sampling," p. 22, Apr. 1997. [Online]. Available: <http://arxiv.org/abs/astro-ph/9704059>
- [2] Richardson, William Hadley, "Bayesian-Based Iterative Method of Image Restoration," *Journal of the Optical Society of America*, vol. 62, no. 1, 1972. [Online]. Available: <http://adsabs.harvard.edu/abs/1972JOSA...62...55R>
- [3] J. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astronomy and Astrophysics Supplement Series*, 1974. [Online]. Available: <http://adsabs.harvard.edu/full/1974A%26AS...15..417H>
- [4] L. Lucy, "An iterative technique for the rectification of observed distributions," *The astronomical journal*, 1974. [Online]. Available: <http://adsabs.harvard.edu/full/1974AJ.....79..745L>
- [5] R. Hook, L. Lucy, A. Stockton, and S. Ridgway, "Two channel photometric image restoration," *Space Telescope European Coordinating Facility Newsletter*, vol. 21, p. 16, 1994. [Online]. Available: <http://adsabs.harvard.edu/abs/1994STECF..21...16H>

Information Capacity of Radio Interferometers

Tobia D. Carozzi*,

* Onsala Space Observatory, Chalmers University of Technology, Observatorievägen 90, Onsala 439 92 Sweden.

Abstract—I present a theory for the spatial channel capacity of an radio interferometer. Modern radio interferometers must handle a large volume of information, indeed it has been suggested that the SKA will have a network traffic comparable to the entire Internet. But how much of this information is actually useful imaging data? I will show how an array's configuration and primary element gain determines the amount of sky information can be extracted by the interferometer. This approach is roughly analogous to looking at how many megapixels are photographic camera has, and can be used to assess imaging performance in interferometers. The results show that minimum-redundancy arrays have the largest imaging information capacity. They also suggest that 3D arrays collect more information from the true image than flat arrays (with all other design parameters identical), so in other words non-negligable w -terms in the radio interferometric measurement equation contribute more information than pure uv -terms.

Sparsity averaging for radio-interferometric imaging

Rafael E. Carrillo*, Jason D. McEwen†, and Yves Wiaux*‡§

* Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

† Department of Physics and Astronomy, University College London, London WC1E 6BT, UK.

‡ Department of Radiology and Medical Informatics, University of Geneva (UniGE), CH-1211 Geneva, Switzerland.

§ Department of Radiology, Lausanne University Hospital (CHUV), CH-1011 Lausanne, Switzerland.

We propose a novel regularization method for compressive imaging in the context of the compressed sensing (CS) theory with coherent and redundant dictionaries [1]. Natural images are often complicated and several types of structures can be present at once. It is well known that piecewise smooth images exhibit gradient sparsity, and that images with extended structures are better encapsulated in wavelet frames. Therefore, we here conjecture that promoting average sparsity or compressibility over multiple frames rather than single frames is an extremely powerful regularization prior. Define $\mathbf{x} \in \mathbb{R}^N$ to be the image of interest. We propose using a dictionary composed of a concatenation of q frames, i.e.

$$\Psi = \frac{1}{\sqrt{q}}[\Psi_1, \Psi_2, \dots, \Psi_q], \quad (1)$$

and an analysis ℓ_0 prior, $\|\Psi^\dagger \mathbf{x}\|_0$, to promote this average sparsity.

Note on a theoretical level that a single signal cannot be arbitrarily sparse simultaneously in any pair of frames. For example, a signal extremely sparse in the Dirac basis is completely spread in the Fourier basis. As discussed in [2], each frame, Ψ_i , should be highly coherent with the other frames in order to have a sparse representation for the signal. The concatenation of the Dirac basis and the first eight orthonormal Daubechies wavelet bases (Db1-Db8) represents a natural choice in the imaging context. The first Daubechies wavelet basis, Db1, is the Haar wavelet basis and, in particular, can be used as an alternative to gradient sparsity (usually imposed by a total variation (TV) prior) to promote piecewise smooth signals. The Db2-Db8 bases are coherent with Haar and Dirac while providing smoother decompositions.

The proposed approach is defined on the basis of the following problem:

$$\min_{\mathbf{x} \in \mathbb{R}_+^N} \|\Psi^\dagger \mathbf{x}\|_0 \text{ subject to } \|\mathbf{y} - \Phi \mathbf{x}\|_2 \leq \epsilon, \quad (2)$$

where the matrix $\Phi \in \mathbb{C}^{M \times N}$ identifies the measurement operator, $\mathbf{y} \in \mathbb{C}^M$ identifies the measurement vector and ϵ is an upper bound on the ℓ_2 norm of the residual noise. The constraint $\mathbf{x} \in \mathbb{R}_+^N$ represents the positivity prior on \mathbf{x} . Practically, the algorithm uses a reweighted approach to approximate ℓ_0 minimization by solving a sequence of weighted ℓ_1 problems. The associated reconstruction algorithm is dubbed Sparsity Averaging Reweighted Analysis (SARA). See [2] for a detailed description of the algorithm. This problem offers great versatility since one only needs to model correctly the measurement operator to allow reconstruction from different sensing modalities. Results comparing SARA to state-of-the-art reconstruction methods for random Gaussian, spread spectrum, and random discrete Fourier sampling are reported in [2], [3]. The results demonstrate that SARA outperforms benchmark methods in all cases.

In the case of radio-interferometric (RI) imaging, under common assumptions, the measurement equation for aperture synthesis provides incomplete Fourier sampling of the image of interest. Since

the measured visibilities provide continuous samples of the frequency plane, an interpolation operator needs to be included in Φ to model the map from a discrete frequency grid onto the continuous plane so that the FFT can be used. Direction dependent effects can also be included in the model as additional convolution kernels in the frequency plane. We here illustrate the performance of SARA in this field by recovering the well known test image of the M31 galaxy from simulated continuous visibilities affected by 30 dB of input noise, and using the Dirac-Db1-Db8 concatenation highlighted for Ψ . For comparison, we also study a variety of sparsity-based image reconstruction algorithms, some of which were identified as providing similar performance as CLEAN and its multi-scale versions, which are state of the art in RI imaging. SARA is compared with the following minimization problems: (i) BP, constrained ℓ_1 -minimization in the Dirac basis (similar to CLEAN), (ii) BPDb8, constrained analysis-based ℓ_1 -minimization in the Db8 basis (similar to multi-scale CLEAN), and (iii) TV, constrained TV-minimization. Figure 1 shows preliminary results of a reconstruction from a realistic radio telescope sampling pattern. SARA provides not only a drastic SNR increase but also a significant reduction of visual artifacts relative to all other methods. Note that all these algorithms are currently being implemented in a new package written in C named PURIFY.

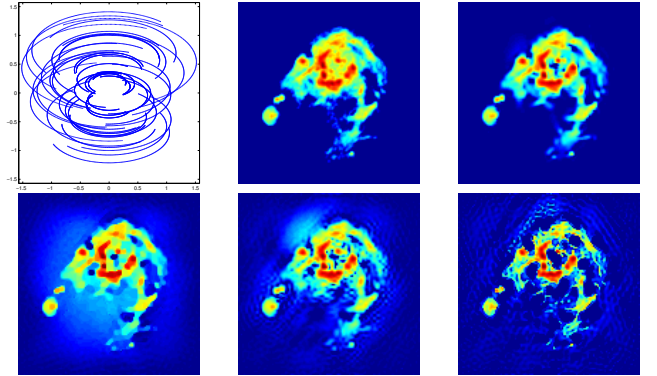


Figure 1. From left to right, top to bottom: Fourier sampling profile, original test image and reconstructions for SARA (13.35 dB), TV (11.44 dB), BPDb8 (11.13 dB) and BP (8.19 dB) in \log_{10} scale.

REFERENCES

- [1] E. J. Candès, Y. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2010.
- [2] R. E. Carrillo, J. D. McEwen, J.-P. T. D. V. De Ville, and Y. Wiaux, "Sparsity averaging for compressive imaging," *IEEE Transactions on Image Processing*, 2012, preprint, arXiv:1208.2330v1.
- [3] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "Sparsity averaging reweighted analysis (SARA): a novel algorithm for radio-interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 426, no. 2, pp. 1223–1234, 2012.

Accelerated Bayesian inference using nests and nets

Michael Hobson,

Astrophysics Group, Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, U.K.

Abstract—Bayesian inference methods are widely used to analyse observations in astrophysics and cosmology, but they can be extremely computationally demanding. New methods have been developed, however, for greatly accelerating such analyses, by up to a factor a million, using nested sampling methods and neural network training algorithms, such as the MULTINEST and SKYNET packages, respectively. These have recently been combined into the BAMBI algorithm, which fully automates the process and accelerates Bayesian inference still further. These generic approaches are illustrated in a cosmological case study.

Bayesian inference methods are widely used in astronomy and cosmology, and are gaining popularity in other fields, such as particle physics phenomenology. In general, they are used to perform two main tasks: parameter estimation and model selection.

In estimating a set of parameters Θ in a model (or hypothesis) H for the data \mathbf{D} , inferences are usually obtained by taking samples from the (unnormalised) posterior distribution of the parameters, using standard Markov Chain Monte-Carlo (MCMC) sampling methods, usually based on the Metropolis–Hastings algorithm or one of its variants, where at equilibrium the chain contains a set of samples from the parameter space distributed according to the posterior. The posterior constitutes the complete Bayesian inference of the parameter values, and can be marginalised over each parameter to obtain individual parameter constraints. The factor required to normalize the posterior over Θ , known as the evidence, is usually ignored, since it is independent of the parameters Θ .

By contrast, in selecting which of a set of competing models $\{H_i\}$ is the most probable given the data, the evidence takes the central role. As the average of the likelihood over the prior, the evidence is larger for a model if more of its parameter space is likely and smaller for a model with large areas in its parameter space having low likelihood values, even if the likelihood function is very highly peaked. Thus, the evidence automatically implements Occam’s razor. The question of model selection between two models H_0 and H_1 can then be decided by comparing their respective posterior probabilities given the observed data set \mathbf{D} .

Both the exploration of the multi-dimensional (unnormalised) posterior distribution in the parameters Θ , and the evaluation of the evidence, i.e. the integral of this distribution over the parameter space, are challenging numerical tasks. Standard MCMC techniques for exploring the posterior can be extremely computationally intensive and often need to be finely tuned in order to produce accurate results. Additionally, sampling efficiency can be seriously affected by multimodal distributions and long (curving) degeneracies in the parameter space. Furthermore, in model selection, the task of evaluating the Bayesian evidence, using the most common method of thermodynamic integration, typically requires a order-of-magnitude more computation than merely exploring the posterior. Fast methods of evidence calculation, such as assuming a Gaussian approximation to the posterior, clearly fail in multimodal and degenerate situations.

Nested sampling, introduced by [1], is an alternative Monte Carlo method, which is targeted at the efficient calculation of the evidence, but also produces posterior inferences as a by-product. [2] and [3] have built on this nested sampling framework and have introduced

the MULTINEST algorithm, which is very efficient at sampling from potentially multimodal and/or degenerate posteriors, and also calculates the evidence. The method has reduced the computational cost of Bayesian parameter estimation and model selection typically by a factor of ~ 100 in a wide range of problems in astrophysics and particle physics phenomenology.

For any sampling method, however, at each sampled point in the parameter space, one must evaluate the posterior distribution. The prior is usually simple to evaluate, but the likelihood, which embodies the entire prediction and measurement process that maps parameters into observed quantities, can be computationally very expensive to calculate. In some cosmological and particle physics applications, evaluation of the likelihood requires up to tens of seconds. Considerable gains can be achieved if one can speed up the evaluation of the likelihood itself. An artificial neural network (NN) is ideally suited to learn this mapping, since a universal approximation theorem assures us that one can always approximate the likelihood to any required accuracy. The SKYNET algorithm uses a variant of conjugate-gradient descent to train NNs using regularisation of the likelihood and a Hessian-free second-order approximation to improve convergence. Replacing the original likelihood with a trained NN can speed up its evaluation by a factor of up to $\sim 10^4$.

Recently, [4] combined MULTINEST with the SKYNET neural network training algorithm to produce the blind accelerated multimodal Bayesian inference (BAMBI) algorithm. After an initial set of new samples from MULTINEST have been obtained, BAMBI uses SKYNET to train a network on the likelihood function. After convergence, the ability of the network to predict likelihood values to within a specified tolerance level is tested. If it fails, sampling continues using the original likelihood until enough new samples have been made for training to be done again. Once a network is trained that is sufficiently accurate, its predictions are used in place of the original likelihood function for future samples for MULTINEST. Using the network typically reduces the likelihood evaluation time to milliseconds, allowing MULTINEST to complete the analysis much more rapidly. As a bonus, the user also obtains a network that is trained to easily and quickly provide more likelihood evaluations near the peak if needed, or in subsequent analyses. In this way, BAMBI both automates and enhances the acceleration process, yielding an overall speed-up of up to a factor of $\sim 10^6$.

REFERENCES

- [1] J. Skilling, “Nested Sampling,” in *American Institute of Physics Conference Series*, R. Fischer, R. Preuss, and U. V. Toussaint, Eds., Nov. 2004, pp. 395–405. [Online]. Available: <http://www.inference.phy.cam.ac.uk/bayesys/>
- [2] F. Feroz and M. P. Hobson, “Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses,” *MNRAS*, vol. 384, pp. 449–463, Feb. 2008.
- [3] F. Feroz, M. P. Hobson, and M. Bridges, “MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics,” *MNRAS*, vol. 398, pp. 1601–1614, Oct. 2009.
- [4] P. Graff, F. Feroz, M. P. Hobson, and A. Lasenby, “BAMBI: blind accelerated multimodal Bayesian inference,” *MNRAS*, vol. 421, pp. 169–180, Mar. 2012.

Methods for detecting the 3D percolation of photons in the early universe.

Daniel Jacobs*,

* School of Earth and Space Exploration, Arizona State University, Tempe, Arizona 85282.

I. ABSTRACT

Fundamental astrophysical questions can be answered by a new kind of radio interferometer (cf MWA, LOFAR, PAPER) which takes advantage of advances in computing power to dramatically increase the number of correlated antennae. The Epoch of Reionization, when photons ionized the intergalactic hydrogen in a collection of bubbles which quickly percolated through the observable universe, is one such area. These new types of arrays have extraordinary promise but violate many basic assumptions in how they measure the sky. However, the signals are also different. The phenomena under study is three dimensional in nature and dominated by smooth emission with sharp edges, rather than astronomical objects of limited extent. The fundamental reionization measurement is a spatial-spectral cube which corresponds, via hubble's law, to three physical dimensions. Finally, the signal is exceedingly faint, requiring an optimized statistical approach. Despite this non-Gaussian nature of the signal, early instruments plan to focus on the power spectrum. We will describe the various methods currently being used to detect these three-dimensional signals and examine possible alternatives. Given the experimental nature of this ongoing work we will also include several examples taken from recent measurements.

Accelerating optical interferometric image reconstruction and modeling using graphical processing units (GPUs)

Brian Kloppenborg* Fabien Baron†

* Max Planck Institute for Radio Astronomy, Auf dem Hügel 69, 53121 Bonn, Germany

†

University of Michigan, Department of Astronomy, 500 Church Street, Ann Arbor, MI 48109, USA.

Abstract—Graphical processing units (GPUs) are massively parallel processors that have been used to accelerate computations in other areas of physics by 10-100 times. Many of the tasks in optical interferometric image reconstruction and model fitting are trivially parallel and therefore ideal candidates for acceleration using GPUs. Our initial experimentation with these methods in the GPU Accelerated Image Reconstruction (GPAIR [1]) program achieved a speed up of more than 100 times on the GPU compared to a CPU. In this talk, I will present the architecture of GPUs, contrast GPU and CPU programming, and discuss how our OpenCL Interferometry Library (*liboi*) addresses many of the computational demands for image reconstruction and modeling of the soon-to-be online “next generation” of beam combiners.

I. INTRODUCTION

Within the next two years all but one optical interferometric (OI hereafter) arrays will hosts beam combiners capable of joining the light from four or more telescopes, potentially at high spectral dispersion. These devices are ideally suited for imaging. Our experience with six telescope, low spectral dispersion data from CHARA-MIRC show current software takes about one minute to reconstruct a grayscale image. With an increasing number of data and the desire to reconstruct spectrally-dependent data sets, one can anticipate that run time will quickly scale to hours or days. Indeed, in the 2010 Optical Interferometry Beauty contest [2], the winning software took 24 hours to reconstruct an initial grayscale image from the medium resolution ($R \sim 1500$) data set and eventually had to resort to considering each spectral channel independently without any cross wavelength regularization. Active work in spectral regularization [3] will make spectrally dispersed OI image reconstruction more robust; however, it is unlikely that any new algorithms will dramatically reduce the required run time. Graphical Processing Units offer a potential solution to long run times by splitting computational tasks over hundreds to thousands of processors.

II. GRAPHICAL PROCESSING UNITS

Graphical Processing Units (GPUs) are massively parallel computational platforms. GPUs are designed for high-throughput computing without a need for a hierarchy of cache. They feature memory accessed through a high-bandwidth, parallel bus and an array tens to hundreds of multiprocessors. Each multiprocessor is, in turn, composed of several simple processors, a small amount (~ 16 kb) of on-chip programmer-managed memory, and one to four transcendental processing units. Because of this arrangement, GPUs are well suited to trivially parallel problems which can exploit single instruction multiple data (SIMD) architecture.

GPUs are programmed by either the Compute Unified Device Architecture (CUDA) which runs on only NVidia hardware or the Open Compute Language (OpenCL) which runs on a heterogeneous environment of GPUs, CPUs, and specialized hardware devices. Both APIs are extended version of C99 in which you write functions, called “kernels,” which execute on the GPU.

III. OPENCL INTERFEROMETRY LIBRARY (LIBOI)

Continuing our work with GPUs and interferometry, we have segmented the GPU-accelerated portion of GPAIR, and created a dedicated C/C++ library *liboi* [4]. This library implements common functions used in OI imaging and modeling; including image normalization, Fourier transforms, OI data creation, and chi-squared evaluation. The API is quite powerful as it provides both high and low-level functionality. A programmer can leverage the power of GPU computing by simply loading the library, providing it the path to a data file, and generating a floating point image. In the most extreme case, the programmer can control the order of execution of *liboi*'s kernels and even inject new kernels derived from existing classes. *liboi* is available on GitHub and distributed under LGPL.

IV. NEW SOFTWARE

To demonstrate the power of *liboi* we have created a new model fitting program called the Simulation and Modeling Tool for Optical interferometry (SIMTOI [5]). This program attempts to offload as much computation to the GPU as possible. It renders physically informed geometric models using OpenGL and utilizes *liboi* to compute and compare the model with observed interferometric data. The parallel portions of this program are nearly 100 times faster than a CPU-only implementation.

V. FUTURE PLANS

liboi presently uses a discrete Fourier transform to compute the complex visibilities used in the rest of the code. We intend to implement a non-uniform fast Fourier transform and provide additional kernels as requested by the broader interferometric community.

REFERENCES

- [1] F. Baron and B. Kloppenborg, “GPU-accelerated image reconstruction for optical and infrared interferometry,” in *Optical and Infrared Interferometry II*, W. C. Danchi, F. Delplancke, and J. K. Rajagopal, Eds., vol. 7734, no. 1. San Diego, California, USA: SPIE, Jul. 2010, pp. 77 344D–9. [Online]. Available: <http://link.aip.org/link/?PSI/7734/77344D/1>
- [2] F. Malbet, W. Cotton, G. Duvert, P. Lawson, A. Chiavassa, J. Young, F. Baron, D. Buscher, S. Rengaswamy, B. Kloppenborg, M. Vannier, and L. Mugnier, “The 2010 interferometric imaging beauty contest,” in *Optical and Infrared Interferometry II*, W. C. Danchi, F. Delplancke, and J. K. Rajagopal, Eds., vol. 7734, no. 1. San Diego, California, USA: SPIE, Jul. 2010, pp. 77 342N–12. [Online]. Available: <http://link.aip.org/link/?PSI/7734/77342N/1>
- [3] E. Thiébaud, F. Soulez, and L. Denis, “Exploiting spatial sparsity for multi-wavelength imaging in optical interferometry,” Sep. 2012. [Online]. Available: <http://adsabs.harvard.edu/abs/2012arXiv1209.2362T>
- [4] B. Kloppenborg and F. Baron, “LibOI: The OpenCL Interferometry Library,” 2012. [Online]. Available: <https://github.com/bkloppenborg/liboi>
- [5] —, “SIMTOI: Simulation and Modeling Tool for Optical Interferometry,” 2012. [Online]. Available: <https://github.com/bkloppenborg/simtoi>

Extended abstract template for BASP Frontiers 2012: Application of Sparse Inpainting to the Integrated Sachs-Wolfe Effect

Anais Rassat^{*}, Jean-Luc Starck[†], Jalal Fadili[‡] and François-Xavier Dupé[§]

^{*} Laboratoire d'Astrophysique, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland.

[†] Laboratoire AIM, UMR CEA-CNRS-Paris 7, Irfu, SAp/SEDI, Service d'Astrophysique, CEA Saclay, 91191 GIF-SUR-YVETTE CEDEX, France.

[‡] GREYC UMR CNRS 6072, Université de Caen Basse-Normandie, ENSICAEN, 6 bvd Maréchal Juin, 14050 Caen, France.

[§] LIF - Qarma, UMR CNRS 7279 Aix-Marseille Univ, France.

Abstract—One of the main challenges of modern cosmology is to understand the nature of the mysterious dark energy which causes the cosmic acceleration. The Integrated Sachs-Wolfe (ISW) effect can be considered as an independent signature of dark energy. The ISW can be detected by cross-correlating Cosmic Microwave Background (CMB) with Large Scale Structure (LSS) data. The ISW effect occurs on large scales, where there are large amounts of missing data in both CMB and LSS maps due to Galactic confusion. This severely limits signal extraction. We propose a novel method to detect and measure the ISW signal. This method requires use of full sky data for the analysis. To address the problem of sky coverage, we use a sparse inpainting method to reconstruct missing data in both the CMB and the LSS data. With our method, we can expect a $\sim 4.7 - 7\sigma$ detection of the ISW with upcoming data.

I. INTRODUCTION

The recent abundance of cosmological data in the last few decades (for an example of the most recent results see 1; 2; 3) has provided compelling evidence towards a standard concordance cosmology, in which the Universe is composed of approximately 4% baryons, 26% ‘dark’ matter and 70% ‘dark’ energy. One of the main challenges of modern cosmology is to understand the nature of the mysterious dark energy which drives the observed cosmic acceleration (4; 5).

The ISW (6) effect is a secondary anisotropy of the CMB, which arises because of the variation with time of the cosmic gravitational potential between local observers and the surface of last scattering. The potential can be traced by LSS surveys (7), and the ISW effect is therefore a probe which links the high redshift CMB with the low redshift matter distribution and can be detected by cross-correlating the two. The ISW effect occurs on large scales, where cosmic variance is high and where there are large amounts of missing data in the Cosmic Microwave Background (CMB) and Large Scale Structure (LSS) maps due to Galactic confusion from our Milky Way. Moreover, existing methods in the literature often make strong assumptions about the statistics of the underlying fields or estimators. Together these effects can severely limit signal extraction.

We aim to define an optimal statistical method for detecting the ISW, which can handle large areas of missing data and minimise the number of underlying assumptions about the data and estimators. We first review current detections (and non-detections) of the ISW effect, comparing statistical subtleties between existing methods, and identifying several limitations. We propose a novel method to detect and measure the ISW signal. This method assumes only that the primordial CMB field is Gaussian, but requires use of full sky data for the analysis. In order to address the problem of sky coverage, we use a sparse inpainting method to reconstruct missing data in both the CMB and the LSS data. We also use a bootstrap technique to avoid assumptions about the statistics of the estimator. It is a complete method, which uses three complementary statistical methods.

In light of the upcoming Euclid mission satellite, which will

provide wide-field LSS data over the redshift range ideal for ISW detection, we apply our method to Euclid-like simulations. We show we can expect a $\sim 7\sigma$ model-independent detection of the ISW signal with WMAP7-like data, even when considering missing data. Other tests return $\sim 4.7\sigma$ detection levels for a Euclid-like survey. We find detections levels are independent from whether the galaxy field is normally or lognormally distributed. We apply our method to the 2 Micron All Sky Survey (2MASS) and WMAP7 CMB data and find detections in the $1.1 - 2.0\sigma$ range, as expected from our simulations. As a by-product, we have also reconstructed the full-sky temperature ISW field due to 2MASS data. The ISW map can be used to further study statistical isotropy and reported deviations from statistical isotropy (so called ‘anomalies’) in the primordial CMB.

REFERENCES

- [1] E. Komatsu, J. Dunkley, M. R. Nolta, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Limon, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, S. S. Meyer, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, “Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation,” , vol. 180, pp. 330–376, Feb. 2009.
- [2] W. J. Percival, R. C. Nichol, D. J. Eisenstein, D. H. Weinberg, M. Fukugita, A. C. Pope, D. P. Schneider, A. S. Szalay, M. S. Vogeley, I. Zehavi, N. A. Bahcall, J. Brinkmann, A. J. Connolly, J. Loveday, and A. Meiksin, “Measuring the Matter Density Using Baryon Oscillations in the SDSS,” *APJ*, vol. 657, pp. 51–55, Mar. 2007 a.
- [3] T. Schrabback, J. Hartlap, B. Joachimi, M. Kilbinger, P. Simon, K. Benabed, M. Bradač, T. Eifler, T. Erben, C. D. Fassnacht, F. W. High, S. Hilbert, H. Hildebrandt, H. Hoekstra, K. Kuijken, P. J. Marshall, Y. Mellier, E. Morganson, P. Schneider, E. Semboloni, L. van Waerbeke, and M. Velander, “Evidence of the accelerated expansion of the Universe from weak lensing tomography with COSMOS,” , vol. 516, pp. A63+, Jun. 2010.
- [4] A. Albrecht, G. Bernstein, R. Cahn, W. L. Freedman, J. Hewitt, W. Hu, J. Huth, M. Kamionkowski, E. W. Kolb, L. Knox, J. C. Mather, S. Staggs, and N. B. Suntzeff, “Report of the Dark Energy Task Force,” *ArXiv Astrophysics e-prints*, Sep. 2006.
- [5] J. A. Peacock, P. Schneider, G. Efstathiou, J. R. Ellis, B. Leibundgut, S. J. Lilly, and Y. Mellier, “ESA-ESO Working Group on ‘Fundamental Cosmology,’” Tech. Rep., Oct. 2006.
- [6] R. K. Sachs and A. M. Wolfe, “Perturbations of a cosmological model and angular variations of the microwave background,” *Astrophys. J.*, vol. 147, pp. 73–90, 1967.
- [7] R. G. Crittenden and N. Turok, “Looking for λ with the rees-sciamia effect,” *Phys. Rev. Lett.*, vol. 76, p. 575, 1996.

Application of Extended Factor Analysis for Interference Mitigation Using a reference array

Ahmad Mour Sardarabadi, Alle-Jan van der Veen,

Faculty Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands

Abstract—In field of radio astronomy one of the challenges is removing the man-made RF interference. Subspace and model based filters have proven to be very powerful tools for RFI mitigation. However most of the time they require the antenna array to be calibrated. Here we use an extended version of Factor Analysis model that includes a more general covariance model to estimate the RFI free covariance matrix produced by sky sources. This method produces ML estimations for the interferer subspace, power and the unknown sky sources simultaneously.

I. INTRODUCTION

Most of the current radio telescopes use interferometry to produce images of the sky and use array of antennas to this end. The radio frequencies used for radio astronomy are usually occupied by man-made interference that needs to be removed. In previous works the advantage of using reference antenna especially with spatial filter is demonstrated [1]. However in order to find the spatial signature of the interferer it is assumed that the noise covariance matrix is known and then EVD is used to find the corresponding subspace. If the noise characteristics are unknown and/or the array is not calibrated, EVD or SVD will not in general find the correct subspace, to overcome this we propose an extended version of factor analysis model to estimate the subspace and power of the RFI simultaneously.

II. DATA MODEL AND ESTIMATION STRATEGY

Following [1] we assume to have a radio telescope with p_0 elements observing the sky (primary array) and a reference array of p_1 elements that only receives the interfering signals because the sky sources are generally very weak. In total we have $p = p_0 + p_1$ elements that collect data. The output of the primary and secondary array can be modeled as

$$\mathbf{x}_0(t) = \mathbf{v}(t) + \mathbf{n}(t) + \mathbf{A}_0 \mathbf{s}(t), \quad (1)$$

$$\mathbf{x}_1(t) = \mathbf{A}_1 \mathbf{s}(t) + \mathbf{n}_1(t), \quad (2)$$

where \mathbf{x}_0 is a $p_0 \times 1$ vector obtained by stacking the signals from each receiving element of the primary array, \mathbf{v} is the signal contribution of sky sources, \mathbf{n} is the noise contributions on the primary array, \mathbf{A}_0 is a $p_0 \times m$ primary array response, \mathbf{s} is a $m \times 1$ vector representing the interfering signals with \mathbf{R}_s as covariance matrix, \mathbf{x}_1 is a $p_1 \times 1$ vector of received signals, \mathbf{A}_1 is the secondary array response and \mathbf{n}_1 is the noise contribution on the secondary array and its covariance matrix is \mathbf{R}_n . The covariance matrix of the interference free signal is given by $\Phi_{00} = \mathcal{E} \{ (\mathbf{v} + \mathbf{n})(\mathbf{v} + \mathbf{n})^H \}$.

The total system can be modeled by stacking \mathbf{x}_0 and \mathbf{x}_1 and the total covariance matrix could be partitioned as

$$\begin{aligned} \mathbf{R} &= \begin{bmatrix} \Phi_{00} + \mathbf{A}_0 \mathbf{R}_s \mathbf{A}_0^H & \mathbf{A}_0 \mathbf{R}_s \mathbf{A}_1^H \\ \mathbf{A}_1 \mathbf{R}_s \mathbf{A}_0^H & \mathbf{A}_1 \mathbf{R}_s \mathbf{A}_1^H + \mathbf{R}_{n1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}_1 \end{bmatrix} \mathbf{R}_s \begin{bmatrix} \mathbf{A}_0^H & \mathbf{A}_1^H \end{bmatrix} + \begin{bmatrix} \Phi_{00} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{n1} \end{bmatrix}. \end{aligned} \quad (3)$$

Based on this model we are interested in finding the Φ_{00} which is the RFI free covariance matrix. If we assume that the noise is

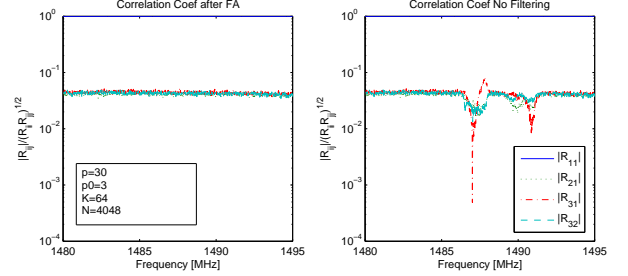


Fig. 1. EFA filtering

uncorrelated between the elements of secondary array and we let

$\mathbf{A} = \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}_1 \end{bmatrix} \mathbf{R}_s^{1/2}$ we can model the covariance matrix as

$$\mathbf{R} = \mathbf{A} \mathbf{A}^H + \mathbf{M} \odot \mathbf{R}_n, \quad (4)$$

with $\mathbf{M} = \begin{bmatrix} \mathbf{1}_{p_0} \mathbf{1}_{p_0}^H & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p_1} \end{bmatrix}$.

This is the Extended Factor Analysis model [2] which is an extension of a multivariate technique called Factor Analysis to non-diagonal noise covariances and we will use this to find Φ_{00} .

In order to make sure that the secondary array does not see the sky sources and depending on the stationarity of interferer we divide the whole integration time (~ 10 s) into small snapshots (~ 10 ms), for each snapshot we estimate $\hat{\Phi}_{00}$ which is the upper sub-matrix of $\mathbf{M} \odot \mathbf{R}_n$ at each snapshot.

Given K snapshots of length N we try to estimate $\hat{\Phi}_{00}$ using (4). The length N is chosen such that the contribution of sky sources on secondary array is negligible. As discussed in [2] there is no closed-form solution for ML FA and iterative approaches are needed. For each snapshot $\hat{\mathbf{R}}_k = 1/N \sum_{n=1}^N \mathbf{x}_n \mathbf{x}_n^H$ we need to estimate $\hat{\mathbf{A}}_k$ and $\hat{\mathbf{R}}_{nk}$. For this purpose we use various iterative methods like WALS, scoring and steep methods. The matrix $\hat{\Phi}_{00k}$ is then the upper left sub-matrix of $\hat{\mathbf{R}}_{nk}$. We take the average of these RFI free estimates to produce the long integration estimate $\hat{\Phi}_{00} = \frac{1}{K} \sum_{k=1}^K \hat{\Phi}_{00k}$. This RFI free result can now be used for further calibration or imaging.

III. EXPERIMENTAL DATA

The test setup is provided by ASTRON research institute in the Netherlands. It uses 4 parabolic dishes from the Westerbork radio telescope one of which has a receiver array in its focus and will be set off-target and used as reference antenna while the other three will track sky source 3C48. Two man made sources (satellites) are causing interference as is shown in Fig. 1 (right). The total integration time is about 13s and it is divided in 64 snapshots of about 200ms then for each snapshot the EFA model is computed. The 3C48 has an almost flat spectrum in the measured frequency band which is divided into sub-bands of about 10 kHz. Fig. 1 (left) shows the result of filtering as described above.

Challenges of radio interferometric imaging in the SKA era

Oleg Smirnov*

* Rhodes University, South Africa

Since its conception over 20 years ago, the Square Kilometre Array project has driven research into new calibration and imaging techniques. In the past decade, this has been accelerated by the new crop of “SKA pathfinder” telescopes. I will present a basic overview of radio interferometric imaging, and then focus on the specific new problems raised by the increased sensitivity and capabilities of the new telescopes, and in the future the SKA itself. Besides the obvious computational and data volume problems (simple scaling laws show that SKA will require a world-class supercomputer just to do basic imaging), there are a number of new algorithmic problems, as well as subtle instrumental effects exposed by the increased sensitivity and new observational regimes of modern instruments.

In the second part of the talk, I will focus on a specific class of problems related to primary beams (PBs). The term “PB” refers to the spatial sensitivity pattern of an individual interferometer element. This pattern serves to attenuate radiation from regions outside the field-of-interest, which simplifies the imaging problem considerably. Classic interferometry assumes a stable and restricted PB pattern that is identical across all elements; deviations from this assumption introduce subtle imaging artefacts, which we could afford to ignore at the lower sensitivities of older instruments. With new telescopes, the intrinsic sensitivity is high enough that these artefacts could in fact become the major performance limits. In this presentation, I will explore these limits and discuss how they could be addressed in practice.

Cosmic Microwave Background Sparse Recovery and Analysis

J.-L. Starck *

* Laboratoire Astrophysique, Instrumentation et Modélisation, IRFU, Service d'Astrophysique, CEA Saclay.

ABSTRACT

Cosmic Microwave Background (CMB) temperature anisotropies and polarisation measurements have been one of the key cosmological probes to establish the current cosmological model. The ESA PLANCK mission is designed to deliver full-sky coverage, low-noise level, high resolution temperature and polarisations maps. We will briefly review some of the key problem of the PLANCK data analysis, and we will present how sparsity can be used to analyze such data set..

Applying full polarization A-Projection to very-wide fields of view instruments: An imager for LOFAR

C. Tasse*, J. van Zwieten†, Ger van Diepen†, B. van der Tol‡, S. Bhatnagar§

* MeerKAT Project Office, SKA SA, 3rd Floor, The Park, Park Road, Pinelands, 7405, South Africa

† Netherlands Institute for Radio Astronomy (ASTRON), Postbus 2, 7990 AA Dwingeloo, The Netherlands

‡ Sterrewacht Leiden, PO Box 9513, 2300 RA, Leiden, The Netherlands

§ National Radio Astronomy Observatory, Socorro, NM 87801, USA

Abstract—We present a few implementations of A-Projection [1] applied to LOFAR, that can deal with non-unitary station beams and non-diagonal Mueller matrices. The algorithm is designed to correct for all the DDE, including ionospheric effects, but we focus our attention on the correction of the phased array beam patterns. They include individual antenna, projection of the dipoles on the sky, and up to a few levels of phased arrays. We describe a few important algorithmic optimizations related to LOFAR’s architecture, that allowed us to build a fast imager. We will use it for the construction of the deepest extragalactic surveys, comprising hundreds of days of integration.

I. INTRODUCTION

With the building or development of many large radio telescopes (LOFAR, EVLA, ASKAP, MeerKAT, MWA, SKA, e-Merlin), radio astronomy is undergoing a period of rapid development. New issues arise with the development of these new types of interferometer, and the approximations applicable to the older generation of instruments are not valid anymore. Specifically, they have wide fields of view and will be seriously affected by the Direction Dependent Effects (DDE). Dealing with the DDE in the framework of calibration and imaging represents an unavoidable challenge, on both the theoretical, numerical and technical aspects of the problem.

This is particularly true for the Low Frequency Array (LOFAR). It is an instrument that observes in a mostly unexplored frequency range ($\nu < 240$ MHz), and will be one of the largest radio telescopes ever built in terms of collecting area. LOFAR’s design is built on a combination of phased array and interferometer. It is made of 40 stations in the Netherlands, and 8 international stations (5 in Germany, 1 in France, England, and Sweden). The High Band Antenna stations (110-240 MHz, HBA hereafter) are made of 24 to 96 *tiles* of 4×4 antenna coherently summed, while the Low Band Antenna (10-80 MHz, LBA) are clustered in groups of 96 elements. At the station level, the signals from the individual antennas or tiles (in the cases of LBA and HBA respectively) are phased and summed by the *beam former*. This step amounts to forming a virtual antenna pointing at the targeted field location. The data is transported from the various stations of the LOFAR array to the correlator. LOFAR is affected by many complex baseline-time-frequency dependent DDE, including mainly the antenna/station beams and the ionosphere, which varies on angular scales of degrees and time scales of ~ 10 minutes and ~ 30 seconds respectively. We currently have models of both the high-band and low-band station beams (HBA and LBA respectively).

As shown in [1] A-Projection allows to estimate sky images, taking into account all the possible complicated effects associated to the DDE. However contrarily to dishes-based interferometers, where the beam shape and polarization angle are affected by pointing errors and rotated on the sky by the parallactic angle (depending on the dishes mount), LOFAR is based on phased arrays that have very wide fields of view (up to ~ 12 degrees), non-trivial and quickly varying beams, thereby driving complicated polarization effects.

Technically speaking, the very fields of view instrument that aim to reach high dynamic range have to deal with baseline-dependent non-diagonal Mueller matrices. For the VLA implementation, due to the approximate *Unitarity* of VLA beams, it was sufficient for A-Projection to take into account the diagonal terms of the Mueller matrices to demonstrate corrections for instrumental polarisation. That is not possible for LOFAR, that has heavily non-unitary beams and non-diagonal baseline-associated Mueller matrices, and all 4×4 Mueller terms have to be taken into account.

We show in this paper that the scheme described in [1] can indeed deal with the heavily non-unitary beams associated with the very-wide fields of view of phased arrays. We will describe the issues related with the usage of phased arrays in interferometers, and focus on the LOFAR-related issues *ie* the polarization aspects and baseline-dependence of the DDE. We will discuss our results based on simulations.

REFERENCES

- [1] S. Bhatnagar, T. J. Cornwell, K. Golap, and J. M. Uson, “Correcting direction-dependent gains in the deconvolution of radio interferometric images,” *A&A*, vol. 487, pp. 419–429, Aug. 2008.

Optical Interferometric Imaging

Éric Thiébaud*

* Université de Lyon, Lyon, F-69003, France;

Université Lyon 1, Observatoire de Lyon, 9 avenue Charles André, Saint-Genis Laval, F-69230, France;

CNRS, UMR 5574, Centre de Recherche Astrophysique de Lyon;

École Normale Supérieure de Lyon, Lyon, F-69007, France.

Abstract—Image reconstruction from interferometric data is an inverse problem quite similar to 2D imaging in X-ray tomography. Astronomical interferometers however give very sparse data and strict constraints such as non-negativity or the priors set by the regularization are crucial ingredients of a successful image reconstruction algorithm. At optical wavelengths, atmospheric turbulence is responsible of random optical path fluctuations which can only be avoided by integrating estimators such as the powerspectrum or the bispectrum (2 and 3 points correlations) insensitive to such errors. This however augments the amount of missing information and makes the image reconstruction more non-linear and thus more difficult. Different priors and optimization strategies to solve the resulting inverse problem have lead to the various existing algorithms. A new challenge is now to deal with multi-spectral data provided by most current interferometers and develop 3D (x, y, λ) imaging algorithms.

Ideally the coherent recombination of two telescopes, say j_1 and j_2 , of a stellar interferometer yields the complex visibility:

$$V_{j_1, j_2}(\lambda, t) = \hat{I}_\lambda(\boldsymbol{\nu}_{j_1, j_2}(\lambda, t)) \quad (1)$$

with $\hat{I}_\lambda(\boldsymbol{\nu})$ the angular Fourier transform of the specific brightness distribution $I_\lambda(\boldsymbol{\theta})$ of the observed object and:

$$\boldsymbol{\nu}_{j_1, j_2}(\lambda, t) = (\mathbf{r}_{j_2}(t) - \mathbf{r}_{j_1}(t))/\lambda \quad (2)$$

a particular spatial frequency which depends on \mathbf{r}_j the position of j th telescope projected on a plane perpendicular to the line of sight, λ the wavelength and t the time. In practice, random optical path fluctuations due to the atmospheric turbulence introduce phase distortion terms in the instantaneous complex visibility.

At long wavelengths (radio interferometry) where the effects of the turbulence evolve very slowly, the complex visibilities can be integrated and Fourier transformed to form the so-called *dirty image* of the object. Then, image reconstruction amounts to the deconvolution of the dirty image given the *dirty beam* which is the analogous of the point spread function (PSF) in more conventional imaging. When, due to the turbulence, exact calibration of the PSF is not possible, *self-calibration* has been proposed to recover it jointly with the image [1]. This is similar but, due to the more limited number of degrees of freedom, much easier than blind deconvolution.

At short wavelengths (infrared and visible), the random phase fluctuations are more important and change very quickly which forbid to directly measure the complex visibility. Hence, in the optical, other estimators have to be used, such as the powerspectrum:

$$P_{j_1, j_2}(\lambda, t) = |\hat{I}_\lambda(\boldsymbol{\nu}_{j_1, j_2}(\lambda, t))|^2, \quad (3)$$

or the bispectrum:

$$B_{j_1, j_2, j_3}(\lambda, t) = \hat{I}_\lambda(\boldsymbol{\nu}_{j_1, j_2}(\lambda, t)) \hat{I}_\lambda(\boldsymbol{\nu}_{j_2, j_3}(\lambda, t)) \hat{I}_\lambda(\boldsymbol{\nu}_{j_3, j_1}(\lambda, t)). \quad (4)$$

These quantities are insensitive to telescopewise phase errors but provide only partial Fourier phase information.

Whatever the wavelength, the interferometric data only provide a sparse sampling of the angular spatial frequencies of the object. To supplement the missing data, additional a priori information has to be

taken into account to ensure a unique and stable image reconstruction. Without loss of generality, image synthesis can be stated as seeking for the object brightness distribution which agrees the most to the priors while being compatible with the measurements:

$$\mathbf{x}^+ = \arg \min_{\mathbf{x} \in \mathbb{X}} f_{\text{prior}}(\mathbf{x}) \quad \text{s.t.} \quad f_{\text{data}}(\mathbf{x}) \leq \eta \quad (5)$$

with \mathbf{x} the image parameters (e.g., the pixel values), \mathbb{X} the set of feasible parameters, $f_{\text{data}} : \mathbb{X} \mapsto \mathbb{R}$ a likelihood term penalizing the discrepancy of the model with the data and $f_{\text{prior}} : \mathbb{X} \mapsto \mathbb{R}$ a regularization term to enforce the priors. The feasible set \mathbb{X} may be used to enforce strict constraints such as the non-negativity and the normalization of the parameters; for instance:

$$\mathbb{X} = \{\mathbf{x} \in \mathbb{R}^N \mid \mathbf{x} \geq 0, \mathbf{1}^\top \cdot \mathbf{x} = \xi\} \quad (6)$$

with N the number of parameters (e.g., pixels), ξ the given image flux, and $\mathbf{1}$ the vector of \mathbb{R}^N which has all its components equal to one. The constrained problem can be recast as:

$$\mathbf{x}_\mu = \arg \min_{\mathbf{x} \in \mathbb{X}} f_{\text{data}}(\mathbf{x}) + \mu f_{\text{prior}}(\mathbf{x}) \quad (7)$$

with $\mu > 0$ a Lagrange multiplier tuned so that $f_{\text{data}}(\mathbf{x}_\mu) = \eta$.

Many existing image reconstruction algorithms can be put in the form of problem (5) or (7). They mainly differ in the type of priors (strict constraints and regularization), the kind of data taken into account and the optimization strategy to seek for the solution [2]. Recent progress in optimization methods and compressive sensing might provide useful sources of inspiration to improve existing imaging methods or deal with larger or more complex problems such as multi-wavelength image reconstruction. For instance, the greedy algorithm of the CLEAN method [3] is equivalent to minimizing an objective function [4] which can be reformulated as a least ℓ_1 problem solved in a much more efficient way by an implementation of the alternating direction method of multipliers (ADMM). The flexibility of ADMM can then be exploited to minimize a joint ℓ_1 norm (instead of a separable one) and recover the spectra and positions of point-like sources from multi-spectral interferometric data [5].

REFERENCES

- [1] T. J. Cornwell and P. N. Wilkinson, "A new method for making maps with unstable radio interferometers," *Month. Not. Roy. Astron. Soc.*, vol. 196, pp. 1067–1086, 1981.
- [2] É. Thiébaud and J.-F. Giovannelli, "Image Reconstruction in Optical Interferometry," *IEEE Signal Process. Mag.*, vol. 27, pp. 97–109, 2010.
- [3] J. A. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astron. Astrophys. Suppl.*, vol. 15, pp. 417–426, 1974.
- [4] K. A. Marsh and J. M. Richardson, "The objective function implicit in the CLEAN algorithm," *Astron. Astrophys.*, vol. 182, pp. 174–178, 1987.
- [5] É. Thiébaud, F. Soulez, and D. Loïc, "Exploiting spatial sparsity for multi-wavelength imaging in optical interferometry," *submitted to J. Opt. Soc. Am. A*, 2012. [Online]. Available: <http://arxiv.org/abs/1209.2362>

Towards a Theory of the Measurement Set

Francois Viallefond*

* LERMA, Observatoire de Paris, 61 av. de l'Observatoire, 75014 Paris, France

Abstract—Preliminary results of a research work at the intersection between mathematics, physics and computer sciences is presented. It results from the development of a data model for the radioastronomy. It is shown that there is a structure, a complex of simplicials, which is ubiquitous in several domains. From an analysis of the grammar of the XMLSchema modeling language is derived a formalism including both the functional and syntactic facets to define what is a type. This structure is carrier of abstract concepts associated to a logical framework. A connection with the domain of compress sensing is suggested.

I. INTRODUCTION

Radio-astronomers have developed a data model to represent the data acquired by radio-telescopes, especially in the context of interferometry in aperture synthesis. A major step was achieved in 2000 with the specification of the MeasurementSet (Kemball and Wieringa)¹. Designed for offline processing it is used for data reduction by several observatories e.g. ALMA, EVLA. Since 2000 a new model was developed to structure the data during their acquisition at the telescope, the Alma Science Data Model (ASDM, Viallefond and Lucas 2006). The ASDM, currently used by ALMA and the EVLA, reuses most of the concepts specified in the MS. The most noticeable difference is the introduction of the concept of configuration, the instruments having multiple processing units operating in concurrency. Taking the opportunity of a very significant evolution of the instrumentation, in particular anticipating the context of SKA, R&D activity is devoted to investigate concepts in the perspective of a third generation model. This led to a prototype, the SDMv2 (Viallefond, 2008, 2013 in prep). The presentation will give some insights of what is this SDMv2. The most innovative aspect with the SDMv2 is that it is associated to a mathematical formalism. By definition generic, this formalism could be of interest far beyond this specific use-case in radioastronomy.

II. METHODOLOGY

A goal is to develop *efficient*, *robust* and *expressive* data models. The idea is to transform objects described using our human language into mathematical objects which can then be used efficiently in information systems. The method is based on the definition of topological spaces to model data types and their algebra; technically it is realized using parametric polymorphism. The robustness comes from coherence conditions verified at compile time. The result is high expressiveness by formulating equations and data base operations by mean of typed λ calculi.

III. THE TOPOLOGY OF THE MS

The presentation will concentrate on one aspect, the topological structure underlying the MS². I'll explain how I arrived to a 2-complex of simplicials guided by a discourse describing a state machine, any measurement instrument reflecting activities through events, eventually occurring on different time scales. Then I'll show that this structure is relevant not only to describe a physical experiment, but also to describe, in particular, *a*) the instrument itself and *b*) the data structure for the reduced data.

¹<http://casa.nrao.edu/Memos/229.html>

²acronym used for what it means, not the name of its grand father

IV. DISCUSSION

To understand this result which may potentially be useful for other domains I investigated several directions to better assess its significance: I made connections with the concepts of *a*) partial order and weak factorization by formulating the binary relation in broad sense, *b*) homotopy, the type theory and the identity type[1] and *c*) the borromean logic[2].

Transforming objects described by our human language into mathematical objects means capturing their semantics and making these functional, at least locally, a requirement to be constructible. To appreciate what can be achieved it is useful to understand the model of the grammar in modeling languages.

Familiar with the XMLSchema language I analyzed its grammar. In the modeling activity semantic is introduced by defining constraints. Therefore I have formalized two concepts of that grammar, the Component and the Uniqueness constraints. It is useful to distinguish two things, the syntactic approach based on terms defined in a lexicon, and the semantic typed approach through the classic concept of inheritance in type hierarchies. Interestingly, to formalize the Component Model of XMLSchema, I arrived to the same structure as the one found for the MS. Furthermore, to formalize the concepts of partition and key also yield to that structure. XML is famous for being described by itself (XMLSchema is written using XML and there is a schema of schema). Given the structure found for the MS we can say now that the grammar is based on a structure which is identical to the structure of models described using this grammar.

The connection with the language being done this gives a mean to express what is a type and the identity in a very general framework.

Using these results coming from different domains give insights about the meaning of this structure as a whole and the reason for its ubiquity. It is the signature of the tri-partition of a set. I'll show that its construction is the result of spatially organizing concepts following a logical structure. Inspecting the meaning of the nodes in these different domains in relation to their locations in the structure reveals systematic abstract properties.

At last, I take the view point that getting an image with an instrument is the measurement of properties of a relation the image being the carrier of semantics that we associate to natural phenomena in our concrete world and that we model on the basis of physical concepts. The formalism just described is a way to express the fact that there is a glue between the syntactic and functional approaches. This gives some insights about the compress sensing strategy, the syntactic approach used to process very large amounts of data with high efficiency assuming that there is an assimilation mechanism to build a lexicon.

REFERENCES

- [1] Awodey, S. and Warren, M.A., "Homotopy theoretic models of identity types," *Math. Proc. Camb. Phil. Soc.*, vol. 146, pp. 45–f55, 2009.
- [2] Guitart, R., "A hexagonal framework of the field f_4 and the associated borromean logic," *Log. Univers. Springer Basel, AG, DOI 10.1007/s11787-011-0033-6*, 2011.

Blind Self-Calibration of Sensor Arrays

Stefan J. Wijnholds*

* Netherlands Institute for Radio Astronomy (ASTRON), Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, The Netherlands.

Abstract—In this presentation, I discuss several techniques that could start the self-calibration process without requiring a sky model. These techniques either exploit redundancy in the array configuration or aim to optimize a measure of image quality. In this context, image quality is usually measured in terms of image contrast or Shannon entropy. After a brief introduction of these techniques, I comment on their applicability in future radio telescopes and present open questions, which is particularly relevant in the context of development of the Square Kilometre Array (SKA).

I. INTRODUCTION

The performance of many self-calibration techniques heavily depends on the quality of the initial estimate for the observed scene. This is equally true for the computational performance and the estimation performance. In radio astronomy, much effort is therefore invested in building sky models, that can be used to start the self-calibration process. In this paper, I discuss two blind self-calibration techniques, redundancy calibration and calibration based on image optimization. The key selling point of blind techniques is that they do not require prior knowledge on the observed scene. I present these methods in the radio astronomy context and comment on their potential use in future radio telescopes like the Square Kilometre Array (SKA) [1].

II. REDUNDANCY CALIBRATION

The key idea behind redundancy calibration is that the correlation of signals from pairs of antennas with identical baseline vectors measure the same spatial frequencies, regardless of the observed brightness distribution [2]. A baseline vector represents the difference between two antenna positions. If the correlations measured on two identical baselines differ, this is caused by instrumental artefacts. If we can assume that all antennas have the same directional response and that imperfections in the analog electronics have a negligible impact on the measured correlations, we can use the principle of redundancy calibration to calibrate the antenna gain and phase.

Placement accuracy is an important consideration in the context of redundancy calibration to ensure sufficient identicalness of the redundant baselines. What identicalness is “sufficient”, depends on the required calibration accuracy. This provides some room to apply redundancy calibration to baselines that are only approximately redundant. We can demonstrate that this can be exploited in initial calibration in relatively dense irregular arrays [3]. This initial calibration could be used as starting point for a normal self-calibration process instead of a sky model.

To exploit redundancy in sparse irregular configurations, redundant baselines probably have to be designed into the system. Since redundant baselines observe the same spatial frequencies, they do not provide additional information on the brightness distribution. An array configuration without redundant baselines may therefore be more attractive from an imaging perspective. Can we provide guidelines to make a trade-off between facilitating the calibration process or facilitating the imaging process?

III. IMAGE OPTIMIZATION

In the absence of noise, the vectorized measured covariance matrix is described by

$$\text{vec}(\hat{\mathbf{R}}) = \text{diag}(\bar{\mathbf{g}} \otimes \mathbf{g}) \mathbf{M} \boldsymbol{\sigma} = \mathbf{M}_{\mathbf{g}} \boldsymbol{\sigma}, \quad (1)$$

where \mathbf{g} is a vector containing the direction independent sensor gains, \mathbf{M} is the measurement matrix for a perfectly calibrated instrument, $\boldsymbol{\sigma}$ is a vector containing image parameters, e.g., wavelet coefficients or pixel values, \otimes denotes the Kronecker product and $\text{diag}(\cdot)$ forms a square matrix with its vector argument on the main diagonal. In the imaging process, (1) is inverted. Assuming that $\mathbf{M}_{\mathbf{g}}$ is invertible, this gives $\boldsymbol{\sigma} = \mathbf{M}_{\mathbf{g}}^{-1} = \text{vec}(\hat{\mathbf{R}})$. Since $\mathbf{M}_{\mathbf{g}}$ depends on the gains, the image can be optimized by finding the right gain values.

What constitutes an optimal image is still a matter of debate. Contrast optimization is a very intuitive choice that can, for example, be formulated as

$$\hat{\mathbf{g}} = \underset{\mathbf{g}}{\text{argmax}} \frac{1}{\langle \boldsymbol{\sigma} \rangle} \sqrt{\langle (\boldsymbol{\sigma} - \langle \boldsymbol{\sigma} \rangle)^2 \rangle}, \quad (2)$$

where $\langle \cdot \rangle$ denotes the averaging operator. Other measures for image contrast include maximization of the peak value in the image (with an appropriate constraint on the gain amplitudes, of course). However, the measure given in (2) seems to be more robust in scenes with multiple sources [4].

Another, but less developed idea is to maximize the Shannon entropy of the image S given by [5]

$$S = - \sum_i \frac{|\sigma_i|^2}{\|\boldsymbol{\sigma}\|^2} \log \frac{|\sigma_i|^2}{\|\boldsymbol{\sigma}\|^2}. \quad (3)$$

It has been demonstrated that maximization of the Shannon entropy can be used for phase calibration [5], but does it also allow gain amplitude calibration? Also, more analysis is required to assess the calibration quality achievable with entropy maximization. This need not be a problem as long as it provides sufficient gain calibration accuracy to make a first image that can then be improved using self-calibration cycles.

REFERENCES

- [1] P. E. Dewdney, P. J. Hall, R. T. Schilizzi, and T. J. L. W. Lazio, “The Square Kilometre Array,” *Proceedings of the IEEE*, vol. 97, no. 8, pp. 1482–1496, Aug. 2009.
- [2] J. E. Noordam and A. G. de Bruyn, “High dynamic range mapping of strong radio sources with application to 3C84,” *Nature*, vol. 299, no. 5884, pp. 597–600, Oct. 1982.
- [3] S. J. Wijnholds, “Self-Calibration of Non-Redundant Arrays Using Redundancy Calibration,” in *Astronomical Data Analysis VII*, Cargèse (Corsica), 14–18 May 2012.
- [4] F. Berzini and G. Corsini, “Autofocussing of Inverse Synthetic Aperture Radar Images Using Contrast Optimization,” *IEEE Trans. Aerospace and Electronic Systems*, vol. 32, no. 3, pp. 1185–1191, 1996.
- [5] P. Lopez Dekker, F. Gordon, and S. J. Frasier, “Entropy based phase calibration of antenna arrays for digital beamforming remote sensing radars,” in *Proceedings of the IEEE Radar Conference*, 2002, pp. 445–452.

Factor Analysis as a Tool for Signal Processing

Stefan Wijnholds[†], Ahmad Mouri Sardarabadi^{*} and Alle-Jan van der Veen^{*}

[†] ASTRON, Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, The Netherlands

^{*} TU Delft, Fac. EEMCS, Mekelweg 4, 2628 CD Delft, The Netherlands

Abstract—In array signal processing, eigenvalue decompositions are commonly used to analyze covariance matrices, e.g. for subspace estimation. An implicit assumption is that the noise power at each antenna element is the same. If the array is not calibrated and the noise powers are different, a Factor Analysis is a more appropriate alternative. It is not a well-known tool in signal processing.

Factor Analysis proves to be very useful for interference mitigation in uncalibrated radio astronomy arrays. It also plays a role in image formation and self-calibration. We will study some of these applications.

I. INTRODUCTION

Factor analysis considers covariance data models where the noise is uncorrelated but has unknown powers at each sensor, i.e., the noise covariance matrix is an arbitrary diagonal with positive real entries. In these cases the familiar eigenvalue decomposition (EVD) has to be replaced by a more general “Factor Analysis” decomposition (FAD), which then reveals all relevant information. It is a very relevant model for the early stages of data processing in radio astronomy, because at that point the instrument is not yet calibrated and the noise powers on the various antennas may be quite different.

As it turns out, this problem has been studied in the psychometrics, biometrics and statistics literature since the 1930s (but usually for real-valued matrices) [1], [2]. The problem has received much less attention in the signal processing literature. In this presentation, we briefly describe the FAD and some algorithms for computing it, as well as some applications.

II. PROBLEM FORMULATION

Assume that we have a set of Q narrow-band Gaussian signals impinging on an array of J sensors. The received signal can be described in complex envelope form by

$$\mathbf{x}(n) = \sum_{q=1}^Q \mathbf{a}_q s_q(n) + \mathbf{n}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{n}(n) \quad (1)$$

where $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_Q]$ contains the array response vectors. In this model, \mathbf{A} is unknown, and the array response vectors are unstructured. The source vector $\mathbf{s}(n)$ and noise vector $\mathbf{n}(n)$ are considered i.i.d. Gaussian, i.e., the corresponding covariance matrices are diagonal. Without loss of generality, we can scale the source signals such that the source covariance matrix is identity.

The data covariance matrix thus has the form

$$\mathbf{R} = \mathbf{A}\mathbf{A}^H + \mathbf{D} \quad (2)$$

where we assume $Q < J$ so that $\mathbf{A}\mathbf{A}^H$ is rank deficient.

Many signal processing algorithms are based on estimates of the signal subspace, i.e. the range of \mathbf{A} . If the noise is white ($\mathbf{D} = \sigma^2 \mathbf{I}$), this information is provided by the eigenvalue decomposition of \mathbf{R} . This does not work if the noise is not uniform. The objective of factor analysis is, for given \mathbf{R} , to identify \mathbf{A} and \mathbf{D} , as well as the factor dimension Q .

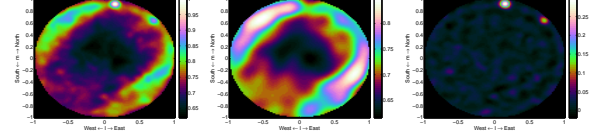


Fig. 1. (a) Calibrated all-sky map for a LOFAR prototype station; (b) image of \mathbf{R}_n , corresponding to baselines shorter than 4 wavelengths; (c) image of the residual.

III. RESEARCH ISSUES

Issues to be discussed in the presentation are:

- 1) *Identifiability*: What constraints provide unique results; what is the maximal factor rank;
- 2) *Detection*: how can the factor rank be determined;
- 3) *Estimation*: how can the factors be estimated.

Some answers are obtained by viewing the problem as a form of covariance matching (cf. [3]). An extension of the data model is

$$\mathbf{R} = \mathbf{A}\mathbf{A}^H + \mathbf{R}_n$$

where \mathbf{R}_n is a band matrix. This can be further generalized to more general (sparse) \mathbf{R}_n with known locations of the nonzero entries.

IV. APPLICATIONS TO RADIO ASTRONOMY

In the context of radio astronomy, factor analysis shows up in a number of applications, see [4] for an overview. Interference cancellation is demonstrated in the accompanying poster presentation. A rank-1 factor analysis problem occurs in the calibration of an array of telescopes pointing at a single calibrator source [5]. As application of the extended factor analysis problem, consider a field with point sources and an extended emission [6]. The extended emission has mostly effect on the short baselines (a band matrix of sorts) whereas the point sources give a low rank contribution. After extended factor analysis, the two components can be imaged separately. Figure 1 shows LOFAR station data, and the resulting image components.

REFERENCES

- [1] D. N. Lawley and A. E. Maxwell, *Factor Analysis as a Statistical Method*. Butterworth & Co, London, 1971.
- [2] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. Academic Press, 1979.
- [3] B. Ottersten, P. Stoica, and R. Roy, “Covariance matching estimation techniques for array signal processing applications,” *Digital Signal Processing, A Review Journal*, vol. 8, pp. 185–210, Jul. 1998.
- [4] A. J. van der Veen and S. J. Wijnholds, “Signal processing tools for radio astronomy,” in *Handbook of Signal Processing Systems, 2nd ed.*, S. Bhattacharyya, Ed. Springer, Feb. 2013, p. (40 pp.).
- [5] A. J. Boonstra and A. J. van der Veen, “Gain calibration methods for radio telescope arrays,” *IEEE Trans. Signal Processing*, vol. 51, no. 1, pp. 25–38, Jan. 2003.
- [6] S. J. Wijnholds and A. J. van der Veen, “Self-calibration of radio astronomical arrays with non-diagonal noise covariance matrix,” in *Proc. 17th European Signal Processing Conference (EUSIPCO 2009)*. Glasgow (UK): Eurasp, Aug. 2009, pp. 1146–1150.

The varying w spread spectrum effect for radio interferometric imaging

Laura Wolz^{*†}, Filipe B. Abdalla^{*}, Rafael E. Carrillo[‡], Yves Wiaux^{‡§¶}, Jason D. McEwen^{*}

^{*} Department of Physics and Astronomy, University College London, London WC1E 6BT, UK

[†] Sub-Dept. of Astrophysics, Dept. of Physics, University of Oxford, The Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

[‡] Institute of Electrical Engineering, Ecole Polytechnique Federale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

[§] Department of Radiology and Medical Informatics, University of Geneva (UniGE), CH-1211 Geneva, Switzerland

[¶] Department of Radiology, Lausanne University Hospital (CHUV), CH-1011 Lausanne, Switzerland

Abstract—We study the impact of the spread spectrum effect in radio interferometry on the quality of image reconstruction. This spread spectrum effect will be induced by the wide field-of-view of forthcoming radio interferometric telescopes. The resulting chirp modulation improves the quality of reconstructed interferometric images by increasing the incoherence of the measurement and sparsity dictionaries. We extend previous studies of this effect to consider the more realistic setting where the chirp modulation varies for each visibility measurement made by the telescope. In these first preliminary results, we show that for this setting the quality of reconstruction improves significantly over the case without chirp modulation and achieves almost the reconstruction quality of the case of maximal, constant chirp modulation.

I. INTRODUCTION

The next generation of radio interferometers will see a large field of view. Consequently, the planar interferometric imaging setting considered typically needs to be adapted to a wide field-of-view by incorporating the so-called w -term component, introducing a chirp modulation.

II. COMPRESSED SENSING IN RADIO INTERFEROMETRY

Previous studies have shown the power of the compressed sensing formalism in radio interferometric imaging [1]–[4]. Radio interferometers acquire incomplete Fourier measurements, so-called visibilities, of the image on the sky under observation. Recovering an image from the visibilities measured by the telescope is hence an ill-posed inverse problem, which is solved through convex optimisation methods (e.g. [1], [3]).

A crucial factor controlling the fidelity of reconstruction in this approach is the incoherence of the measurement and sparsity dictionaries. In the wide field-of-view setting, the chirp modulation that is induced acts to increase incoherence. For radio interferometry, the measurement basis can essentially be identified with the Fourier basis. In this case the coherence is given by the maximum modulus of the Fourier coefficient of the sparsity atoms. The chirp modulation corresponds to a norm-preserving convolution in Fourier space, spreading the spectrum of the sparsity atoms, thus reducing the maximum modulus of their Fourier coefficients and increasing incoherence. The increased incoherence due to this spread spectrum effect acts to improve the fidelity of image reconstruction [3].

III. FIRST RESULTS AND OUTLOOK

In this preliminary work we first confirm previous results [3] that demonstrate the effectiveness of the spread spectrum phenomenon, however here we consider more realistic interferometric images and alternative sparsity dictionaries. We then extend the constant chirp analysis of previous studies [3], [4] to the more realistic setting where every measurement in the (u, v) -space of visibilities undergoes a different w -term modulation. This is a computationally demanding setting which we address by incorporating the w -projection algorithm [5] into our framework. We consider uniform visibility sampling in (u, v) and w , with w samples ranging from zero to $2/L$ times the maximum values of u and v , where L corresponds to the size of

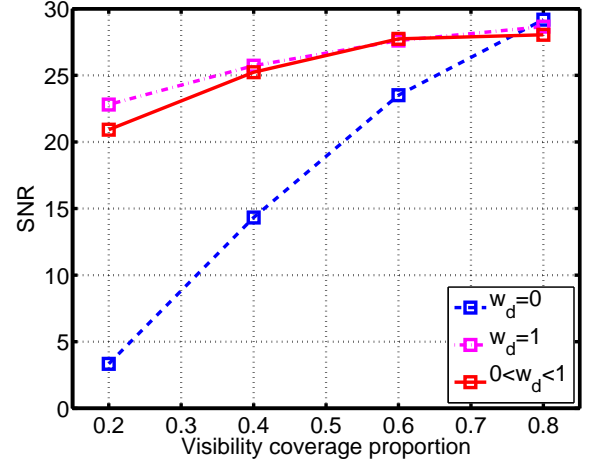


Fig. 1. Signal-to-noise ratio of the recovered image of M31 with Daubechies 8 wavelets for no chirp (blue, dashed), a constant maximal chirp (magenta, dash-dotted) and a varying w -modulation (red, solid) as a function of visibility coverage.

the field-of-view (identical to the maximum w considered previously [3]). We denote the w range by the discrete component $0 < w_d < 1$ respectively. Hence, $w_d = 0$ corresponds to no chirp modulation, $w_d = 1$ corresponds to the maximal chirp modulation studied previously [3], and the range $0 < w_d < 1$ corresponds to uniformly random sampling over the entire range.

Preliminary results on a small test image of an HII region of M31 (see e.g. [2]) of dimension 120×120 pixels show that reconstruction fidelity is significantly improved compared to the analysis without chirp modulation when using Daubechies 8 wavelets (see Fig. 1), extending previous findings to more realistic images and alternative sparsity dictionaries. Furthermore, Fig. 1 shows that reconstruction fidelity for the varying w case is almost as good as the constant, maximal chirp modulation. The study of the varying w spread spectrum effect in the context of the SARA algorithm [2] is ongoing.

REFERENCES

- [1] Y. Wiaux, L. Jacques, G. Puy, A. M. M. Scaife, and P. Vanderghenst, “Compressed sensing imaging techniques for radio interferometry,” *MNRAS*, vol. 395, pp. 1733–1742, May 2009.
- [2] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, “Sparsity Averaging Reweighted Analysis (SARA): a novel algorithm for radio-interferometric imaging,” *MNRAS*, vol. 426, pp. 1223–1234, Oct. 2012.
- [3] Y. Wiaux, G. Puy, Y. Boursier, and P. Vanderghenst, “Spread spectrum for imaging techniques in radio interferometry,” *MNRAS*, vol. 400, pp. 1029–1038, Dec. 2009.
- [4] J. D. McEwen and Y. Wiaux, “Compressed sensing for wide-field radio interferometric imaging,” *MNRAS*, vol. 413, pp. 1318–1332, May 2011.
- [5] T. J. Cornwell, K. Golap, and S. Bhatnagar, “The Noncoplanar Baselines Effect in Radio Interferometry: The W-Projection Algorithm,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, pp. 647–657, Nov. 2008.

Divergence-free phase contrast MRI

Nikolaos Arvanitopoulos*, Claudio Santelli^{†‡}, Matthias Seeger* and Sebastian Kozerke^{†‡}

* Laboratory for Probabilistic Machine Learning, Ecole Polytechnique Federale de Lausanne, Switzerland.

[†] Imaging Sciences and Biomedical Engineering, King's College London, United Kingdom.

[‡] Institute for Biomedical Engineering, University and ETH Zurich, Switzerland.

Abstract—In this work, we extend a separate magnitude-phase regularization framework for Phase Contrast MRI by incorporating the divergence-free condition.

Introduction: 3D phase-contrast (PC) MRI is a powerful tool to assess hemodynamic parameters. However, this method is hampered by long acquisition times and residual phase errors due to system imperfections. The latter can be addressed by incorporating physical priors. In [1] the reconstructed vector field was processed to approximate incompressibility of blood using normalized convolution and divergence-free basis functions. Using compressed sensing (CS) for scan acceleration, regularizers are often designed for magnitude reconstruction [2], and therefore, may not be robust for phase encoding. In [3] it was demonstrated that improved phase accuracy can be achieved by separate magnitude and phase regularization. In this work, we extend this framework for PC MRI by incorporating the divergence-free and smoothness condition of the velocity flow field.

Theory: In a 4-point PC experiment, the velocities at position \mathbf{r} along three orthogonal ($i = 1, 2, 3$) directions are given by $v_i(\mathbf{r}) = (\phi_i(\mathbf{r}) - \phi_0(\mathbf{r}))/k_{v,i}$. ϕ_i and ϕ_0 denotes the phase of the velocity encoded $\rho_i(\mathbf{r})$, and reference image $\rho_0(\mathbf{r})$, respectively. $k_{v,i}$ is the first moment of the applied bipolar gradient along i . Using incoherent undersampling and collecting all four images into $\boldsymbol{\rho} \in \mathbb{C}^{4n}$ and the acquired k-space samples into $\mathbf{d} \in \mathbb{C}^{4m}$, the separate magnitude-phase reconstruction problem is initialized with the solution minimizing the following convex baseline cost function:

$$\Psi_1(\boldsymbol{\rho}) = \frac{1}{2} \|\mathbf{d} - (\mathbf{I}_4 \otimes \mathbf{E})\boldsymbol{\rho}\|_2^2 + \lambda_1 \|(\mathbf{I}_4 \otimes \mathbf{B})\boldsymbol{\rho}\|_1 + \lambda_2 \|(\mathbf{H} \otimes \mathbf{I}_n)\boldsymbol{\rho}\|_1, \quad (1)$$

with encoding matrix \mathbf{E} relating the reconstructed images to the acquired k-space trajectory, \mathbf{B} an operator implementing several sparsifying transforms and \mathbf{H} the last 3 rows of the Hadamard matrix, producing sparse complex difference images with signal concentrated in the vessels. The image is then decomposed into its magnitude $m_j \in \mathbb{R}$, and phase component $\theta_j = e^{i\varphi_j} \in \mathbb{C}$, s.t. $\boldsymbol{\rho} = \mathbf{m} \circ \boldsymbol{\theta}$. Both components are reconstructed by minimizing the cost function,

$$\Psi_2(\mathbf{m}, \boldsymbol{\theta}) = \frac{1}{2} \|\mathbf{d} - (\mathbf{I}_4 \otimes \mathbf{E})(\mathbf{m} \circ \boldsymbol{\theta})\|_2^2 + \lambda_1 \|(\mathbf{I}_4 \otimes \mathbf{B})\mathbf{m}\|_1 + \lambda_2 \sum_{i=1}^3 \|\mathbf{m}_i - \mathbf{m}_0\|_1 + \mathcal{R}(\boldsymbol{\varphi}), \quad (2)$$

with separate phase regularization

$$\mathcal{R}(\boldsymbol{\varphi}) = \lambda_3 \left\| \boldsymbol{\omega} \circ \left(\sum_{i=1}^3 \nabla_i(\boldsymbol{\varphi}_i - \boldsymbol{\varphi}_0) \right) \right\|_2^2 + \lambda_4 \sum_{i=1}^3 \|\mathbf{C}(\boldsymbol{\varphi}_i - \boldsymbol{\varphi}_0)\|_2^2,$$

where $\boldsymbol{\omega} \in \{0, 1\}^n$ is a masking vector, ∇_i the gradient operator along i and \mathbf{C} the 3D finite difference matrix penalizing divergence and enforcing smoothness of the vector field, respectively.

Methods: Simulated PC MRI data with phase wraps was generated using a computational fluid dynamic (CFD) vector field in a U-bend. Gaussian noise was added (SNR = 30) and the resulting reference images were projected on 8-fold undersampled radial trajectories in 3D k-space. Reconstruction was then performed on the reduced data by subsequently minimizing Eq. 1 and Eq. 2.

Following [4], Eq. 1 and Eq. 2 are reformulated as constrained optimization problems using variable splitting, where parts of the objective are decoupled by introducing equality constraints. These constraints are incorporated by adding augmented Lagrangian terms with additional split variables and Lagrange multipliers. The optimization consists of iteratively updating split variables, unknown variables and Lagrange multipliers. The advantage of this algorithm is that most of the variable updates have simple form and can be efficiently computed. Furthermore, the number of FFT computations needed is largely minimized.

Results: Figure 1 shows reconstructed phase and divergence profiles along the indicated lines in the noisy phase reference images encoding two perpendicular in-plane velocity components.

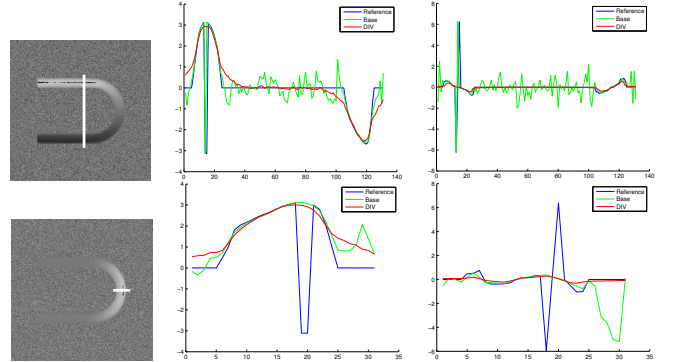


Fig. 1. Phase and divergence profiles from noise-less reference data (blue), baseline (green) and divergence regularized reconstruction (red).

Discussion: In this work, an extension of the separate magnitude-phase regularization framework for PC MRI has been developed and evaluated on CFD data. It has been shown that the proposed method is able to decrease the divergence of the reconstructed 3D velocity field. Furthermore, by using the additional smoothness regularization of the velocities, the method is also able to deal with phase wrapping.

REFERENCES

- [1] B. J. et al., “Construction of Divergence-Free Velocity Fields From Cine 3D Phase-Contrast Flow Measurements,” *MRM*, 2012.
- [2] L. M. et al., “Sparse MRI: The application of compressed sensing for rapid MR imaging,” *MRM*, 2007.
- [3] Z. F. et al., “Separate Magnitude and Phase Regularization via Compressed Sensing,” *IEEE TMI*, 2012.
- [4] R. S. et al., “Parallel MR Image Reconstruction Using Augmented Lagrangian Methods,” *IEEE TMI*, 2011.

Photo-acoustics and acousto-optics: mixing acoustics and optics for biomedical applications

Claude Boccara*

* Institut Langevin ESPCI-ParisTech, 1 rue Jussieu 75005 PARIS FRANCE
claude.boccara@espci.fr

We will summarize the main approaches that are used to image through scattering media such as biological tissue and discuss their limits in term of resolution, depth and signal to noise ratio. If for shallow explorations one can rely on Optical Coherence Tomography that uses singly backscattered photons, just like in acoustic echography, for deeper exploration of biological tissues one has to rely on diffuse photons. Diffuse Tomography is difficult, not only because it is an ill posed problem but also because the body is highly heterogeneous at various scales: resolution is thus practically limited to about one third of the depth. In this context coupling optics and acoustics using acousto-optics or photoacoustics was found useful to get acoustic resolution (typically $\approx 1\text{mm}$) at a few cm depths in order to reveal an optical contrast. We will illustrate the principles and some applications of these two techniques that are based on fairly different physical basis. We know that wavefront engineering has been very helpful to correct aberrations induced by atmosphere turbulence (in optics) or body induced aberrations in acoustics. In the same spirit we will point the progresses that have been achieved these last years in term of wavefront control in the space domain or in the time domain and the perspective that they open to image through aberrating and scattering media and we will show examples using time reversal photoacoustics. Finally we would like discussing how these wavefront controls could help to revisit the field of optical tomography.

Early detection of epileptic seizures by entropy-based methods

Davide Conigliaro*, Paolo Manganotti†, Gloria Menegaz*

* Dept. of Computer Science, University of Verona

† Department of Neurology, University of Verona

Abstract—This work presents a novel method for early detection of epileptic seizures from EEG data. Seizure detection was accomplished in three stages: multiresolution overcomplete decomposition by the *à-trous* algorithm, feature extraction by computing power spectral density and sample entropy values of sub-bands and detection by using z-test and support vector machine (SVM). Results highlight large differences between the subband sample entropy values for the epileptic and the control subjects, respectively, revealing a substantial increase of such parameter during the crisis. This enables high detection accuracy and specificity especially in beta and gamma bands (16-125 Hz). The detection performance of the proposed method was evaluated based on the ground truth provided by the expert neurophysiologist as well as by objective indexes when two crisis had been recorded.

I. INTRODUCTION

While automatic detection of epileptic crisis dates back to many years, early detection with high temporal accuracy is still largely under-investigated. The early detection of a seizure has many potential benefits: it enables early warning allowing patients to take actions to minimize their risk of injury and provides information about the onset of the crisis which would greatly contribute to understanding of the pathophysiological mechanisms that generate seizures. In this study, we propose a new detection method based on the multiresolution decomposition of the EEG signal using the stationary wavelet transform (SWT). From the subbands, the sample entropy, which provides information about the complexity of the signal, and the power spectral density were extracted. Such features were used for revealing the onset of the crisis using two different detection modalities depending on the availability of a second crisis. The proposed method was validated using clinical EEG data recorded from five epileptic patients during both ictal and interictal periods. Results show that the it was able to detect 100% of epileptic seizures with very high accuracy and specificity.

II. METHODS

The EEGs were preprocessed by using the notch filter to remove artifact caused by electrical power lines at 50 Hz and detrending was performed on time windows of one second duration.

Then EEGs were analyzed using the SWT which ensures a time-invariant representation. A five levels wavelet transform was performed in order to match the physiological subbands that are relevant in epilepsy. In particular, this representation the detail coefficients D1, D2, D3, D4, D5 represent the the from low to high frequencies bands called delta δ , theta θ , alpha α , beta β and gamma γ . Daubechies 4 (DB4) [1] was chosen.

Sample entropy (SampEn) [2] values of the approximation and detail coefficients were computed as features. SampEn is a useful tool for investigating the dynamics of time series, that quantifies the complexity of signals.

Define $B^m(r)$ and $A^m(r)$ as:

$$K^\nu(r) = \frac{1}{N-\nu} \sum_{i=1}^{N-\nu} K_i^\nu(r) \quad (1)$$

where $\nu = m, m+1$ and $K = A$ for $\nu = m+1$ and $K = B$ for $\nu = m$. To give an intuition, $B^m(r)$ and $A^{m+1}(r)$ represent the probability that two sequences will match in m and $m+1$ points, respectively. The sample entropy is defined as

$$\text{SampEn}(m, r) = \lim_{N \rightarrow \infty} \ln \frac{B^m(r)}{A^{m+1}(r)} \quad (2)$$

thus it represents the conditional probability that two sequences remain within r of each other [2]. The parameters values were as follows: window size $m = 2$, vector comparison distance $r = 0.2\sigma$ and window length $N = 250$ where σ is the standard deviation of the EEG signal in resting state. The power spectral density for each subbands was calculated by using the Welch method.

For seizure detection, two methods were used depending on the number of seizures. In case of a single seizure, a z-test was used to identify changes in the probability density function of the sample entropy with respect to the resting state. Instead, the Support Vector Machine (SVM) classification was used for repeated seizures. In the first case, the main issue consisted in choosing the significance level α due to the trade-off between the accuracy in temporal detection (high value) and the specificity in the selection (low values). The best value was estimated at 10^{-7} by inspection. In the second case, the first seizure was used for training the model and the others to test it. The two class weighted SVM was applied using radial basis functions and five fold cross-validation to find the best values of parameters C and γ . Finally the performance of the detection system was evaluated and it validated with the help of the neurophysiologist.

III. RESULTS AND DISCUSSION

The algorithm was applied to clinical data collected at Verona hospital. The dataset consisted of five EEG recordings from epileptic patients, for a total of 7 hours and 37 minutes. The duration of the seizures was varying between 8 and 120 sec. Twenty electrodes positioned according to the international 10-20 system of electrode placement were used. The EEG data were sampled at 250 Hz. Overall, sensitivity is generally low. However, our detection system is able to detect 100% of the seizures, with an average accuracy reaching the 98% for 4 out of 5 patients. These results generally improve by using the power spectral density as an additional feature and SVM as classification tool. In conclusion, the proposed method has the advantage of enabling the detection of seizures with very low latency. The changes in the value of the sample entropy with respect to seizure free conditions are subband and pathology dependent, besides being subject to the inherent inter-subject variability thus further investigation is needed for a complete validation.

REFERENCES

- [1] Adeli H., et al., "Analysis of EEG records in an epileptic patient using wavelet transform," *Journal of Neuroscience Methods*, vol. 123, pp. 69–87, 2003.
- [2] Richman J. S., Moorman J. R., "Physiological time-series analysis using approximate entropy and sample entropy," *Am J Physiol*, vol. 278, pp. 2039–2049, 2000.

ℓ_0 -deconvolution for compressive diffusion MRI

Alessandro Daducci*, Anna Auría*, Jean-Philippe Thiran*[†], Yves Wiaux*^{†‡}

* Signal Processing Lab (LTS5), École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

[†] University Hospital Center (CHUV) and University of Lausanne (UNIL), Switzerland

[‡] Department of Radiology and Medical Informatics, University of Geneva (UniGE), Geneva, Switzerland

Diffusion MRI is a well established imaging modality providing a powerful way to non-invasively probe the structure of the white matter. Despite the potential of the technique, the intrinsic long scan times of these sequences have hampered their use in clinical practice. For this reason, a wide variety of methods have been proposed to shorten acquisition times. Among them, *spherical deconvolution approaches* have gained a lot of interest for their ability to reliably recover the intra-voxel distribution of neuronal fiber orientations (FOD), represented as a function on the sphere \mathbf{x} , with a relatively small number of samples. To overcome the ill-posedness of deconvolution, these methods make use of regularization schemes generally based on the assumption that the FOD is sparse due to the small number of fiber populations present in each voxel. On one hand, the well-known Constrained Spherical Deconvolution [1] approach (herein CSD) relies on an ℓ_2 prior which presents the drawback of not promoting sparsity explicitly. On the other hand, convex optimization methods have recently been advocated in a compressed sensing perspective. A recent approach [2] (herein L2L1) relies on some ℓ_1 minimization which unfortunately conflicts with the physical constraint that the fiber compartments must sum to unity: $\|\mathbf{x}\|_1 = 1$.

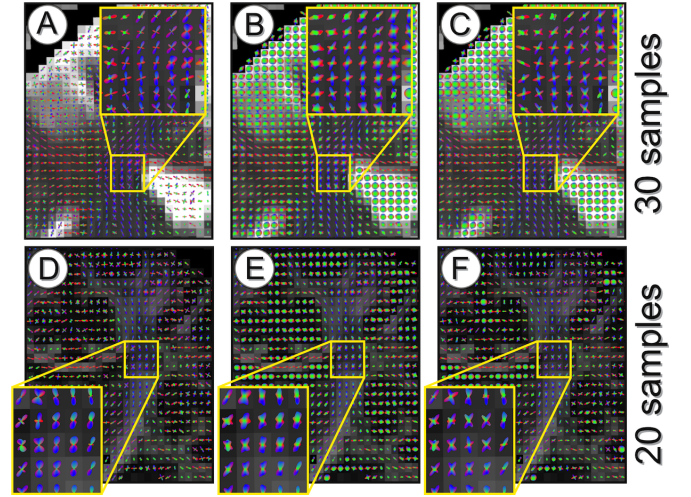
We here review a recent work [3] where we propose to further exploit the versatility of compressed sensing and convex optimization with the aim to characterize the FOD sparsity more optimally. We re-formulate the spherical deconvolution problem as a constrained ℓ_0 minimization:

$$\underset{\mathbf{x} \geq 0}{\operatorname{argmin}} \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 \quad \text{subject to} \quad \|\mathbf{x}\|_0 \leq k, \quad (1)$$

where $\|\cdot\|_0$ explicitly counts the number of nonzero coefficients and k represents the expected number of fiber populations in a voxel. We call this formulation *Reweighted Sparse Deconvolution* or RSD. Surely, ℓ_0 problems as such are intractable. However, the reweighting scheme proposed in [4] preserves the tractability of the problem by sequentially solving weighted ℓ_1 problems of the form (1), where the ℓ_0 norm is substituted by a weighted ℓ_1 norm defined as $\|\mathbf{w}\alpha\|_1 = \sum_i \mathbf{w}_i |\alpha_i|$, for positive weights \mathbf{w}_i and where i indexes vector components. At each iteration, the weights are essentially set as the inverse of the values of the solution of the previous problem, i.e. $\mathbf{w}_i^{(t)} \approx 1/\mathbf{x}_i^{(t-1)}$. At convergence, this set of weights makes the weighted ℓ_1 norm independent of the precise value of the nonzero components, thus mimicking the ℓ_0 norm behavior.

To test our reconstruction method, we compared it against CSD and L2L1 on two human brain datasets acquired using standard clinical protocols, respectively using 30 and 20 diffusion directions, with $b = 1000 \text{ s/mm}^2$. The results are reported in the Figure. Subplots A, B and C correspond to the dataset acquired using 30 samples. Even though the acquisition scheme used for this dataset is not the setting where our numerical simulations (not reported here for brevity) highlighted the most substantial differences between the three approaches, important conclusions can be drawn. First, the ability of both L2L1 (center plots) and RSD (righthand plots) to properly model

the isotropic compartment in voxels with full or partial contamination with CSF is clearly visible. Moreover, comparing B and C we can observe that RSD successfully differentiates gray matter (light gray regions) from CSF voxels with pure isotropic and fast diffusion (very bright areas), while L2L1 appears unable to distinguish them. The yellow frames in the figures highlight the corona radiata, a well-known region in the white matter containing challenging crossing fibers. We observe that RSD clearly results in sharper and more defined profiles than L2L1, whereas the improvements with respect to CSD (lefthand plots) are confined only to few voxels.



The performances of the three methods sensibly change when decreasing the acquisition samples to 20 (subplots D, E and F). Reconstructions with RSD are definitely much better resolved than both CSD and L2L1. In fact CSD clearly breaks, missing many fiber compartments probably due to limitations in the internal Spherical Harmonics representation. The same deterioration happens to L2L1, whose reconstructions appear very blurred and noisy. These results show that our proposed regularization scheme is indeed very effective and that the improvements are most remarkable in a *high q -space under-sampling regime*, thus opening the way for a further scan time reduction of high angular resolution acquisitions closer to DTI.

REFERENCES

- [1] J. Tournier, F. Calamante, D. Gadian, and A. Connelly, “Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution,” *Neuroimage*, vol. 23, pp. 1176–85, 2004.
- [2] B. Landman, J. Bogovic, H. Wan, F. Elshahaby, P. Bazin, and J. Prince, “Resolution of crossing fibers with constrained compressed sensing using diffusion tensor MRI,” *Neuroimage*, vol. 59, pp. 2175–86, 2012.
- [3] A. Daducci, D. Van De Ville, J.-P. Thiran, and Y. Wiaux, “Reweighted sparse deconvolution for high angular resolution diffusion MRI,” 2012, pre-print in arXiv:1208.2247, submitted to Neuroimage.
- [4] E. Candès, M. Wakin, and S. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *J Fourier Anal Appl*, vol. 14(5), pp. 877–905, 2008.

High angular resolution diffusion imaging (HARDI): from local reconstruction to fiber tracking

Maxime Descoteaux*

* Sherbrooke Connectivity Imaging Lab, Computer Science department, Université de Sherbrooke, Sherbrooke, Canada.

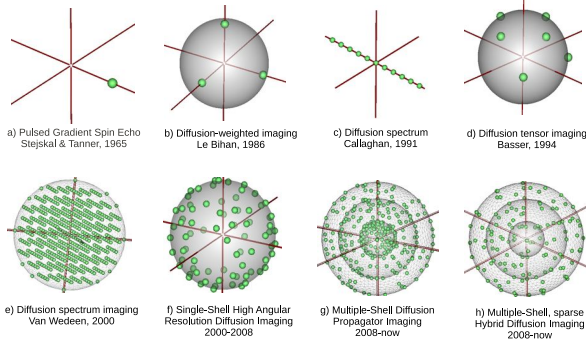


Fig. 1. History of diffusion-weighted imaging and sampling schemes that affect the local reconstruction problem [1].

Abstract—In this presentation, state-of-the-art reconstruction techniques will be presented for high angular resolution diffusion imaging (HARDI). HARDI has been extensively studied in the last ten years. I will cover the most important model-free and model-based techniques. HARDI reconstruction is often the first step towards tractography. Considering local HARDI reconstruction techniques and studying their impact on the tractography results is a challenge in terms of validation. Tons of parameters come into play and I will bring light on the open questions and remaining challenges for the diffusion MRI community.

I. OVERVIEW OF THE PRESENTATION

Fig. 1 shows a short history and evolution of the strategies adopted during the last two decades to perform diffusion-weighted imaging. In this presentation, I will mostly focus on single-shell HARDI samplings and more complex sampling schemes (seen in e) to h)). The sampling scheme directly affects the reconstruction technique.

HARDI techniques were originally developed to overcome limits of diffusion tensor imaging (DTI) in crossing configurations. DTI assumes a Gaussian diffusion process and by now, we know that this is a wrong assumption in nearly 90% of white matter voxels [2].

In this talk, we will cover the model-free techniques family, summarized in Fig. 2. These techniques generally do not assume a physical model of compartments in the imaging voxels, as the model-based techniques such as the one illustrated in Fig. 3.

All these HARDI techniques affect the tractography algorithms. I will review the four classes of tracking techniques: deterministic, probabilistic, geodesics, and global techniques. They all have advantages and inconveniences. We will address the challenge of validation of these techniques and show some results in clinical applications (Fig. 4).

REFERENCES

- [1] M. Descoteaux and C. Poupon, *Diffusion-Weighted MRI*, D. Belvic and K. Belvic, Eds. In *Comprehensive Biomedical Physics*, Elsevier, 2012.
- [2] B. Jeurissen, A. Leemans, J.-D. Tournier, D. K. Jones, and J. Sijbers, "Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging," *Human Brain Mapping*, p. in press, 2012.

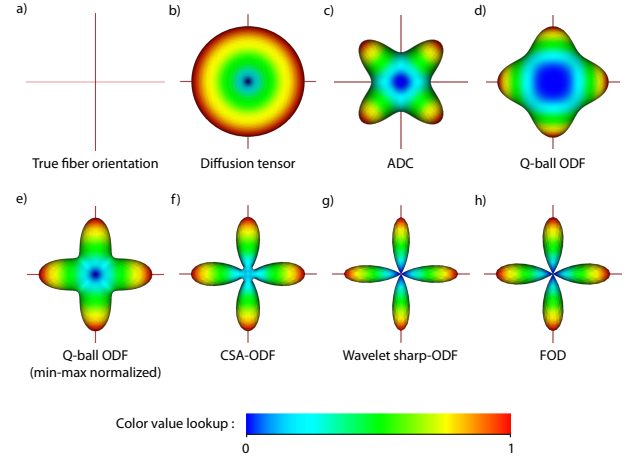


Fig. 2. Model-free techniques. From the true fiber distribution and limitations of the diffusion tensor, to different model-free techniques such as apparent diffusion coefficient (ADC) modeling, q-ball imaging and reconstruction of the diffusion orientation distribution function (ODF) with and without constant solid angle (CSA) consideration, to spherical wavelet techniques and finally, spherical deconvolution techniques reconstructing the fiber orientation distribution (FOD).

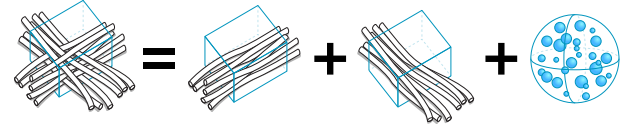


Fig. 3. Model-based techniques. Here, the figure shows a classical example of a three compartment model: two compartments modeling the crossing fibers separately and an isotropic compartment modeling free diffusion. This is also called the *ball and multi-stick* model.

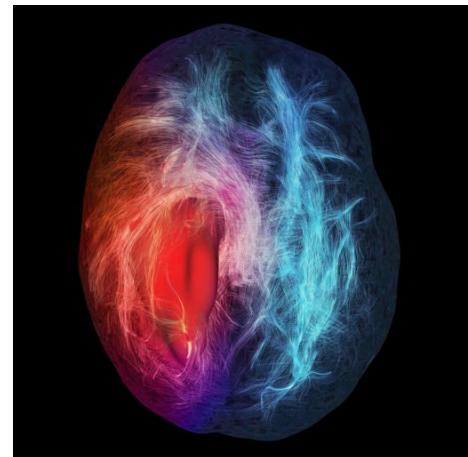


Fig. 4. Tracking techniques in a neurosurgical application. The red surface is a tumor that deviate fiber tracts.

Describing Neuronal Modulations with Bayesian Adaptive Regression Splines

Jack DiGiovanna, Gregoire Courtine, and Silvestro Micera

Abstract—Neuronal modulations are fundamentally “noisy”, especially single units in the motor cortex where the baseline firing rates are low. Standard techniques such as event-triggered histograms extract the underlying neural response by averaging out inter-trial variability. However, histograms are not a compact representation. Data fitting can provide compactness, but may induce signal loss. We describe our implementation of Bayesian Adaptive Regression Splines (BARS) for extracting modulation features from hindlimb motor cortex during over-ground locomotion.

I. INTRODUCTION

NEURAL modulation is an ambiguous term which has been used to describe a wide variety of features including: firing rate, inter-spike interval, ensemble covariance, local field potential, etc. Here we are specifically interested in single unit firing rates. However, action potential generation includes a stochastic element, i.e. given the same stimulus (desire) there is a randomness superimposed on neural response (command) [1, 2]. This inter-trial variability is greatly reduced in the Peri-Event Spike Histogram (PESH) that represents the average response from all trials [1, 2]. However, the variability only disappears as the number of trials included in the histogram goes to infinity; an impossible requirement for behavioral studies. We use fitting to further reduce inter-trial variability but preserve the underlying signals and find BARS superior to 5th order polynomials.

II. ANIMAL MODEL AND RECORDINGS

A cohort of female Lewis rats was cross-trained on multiple tasks; then implanted chronically with a 4x8 micro-wire array (TDT, Alachua FL) in layer V covering left hindlimb sensorimotor cortex. Kinematic and EMG activity was recorded (Vicon, Oxford UK) while rats walked overground on an elevated runway. Single units were identified via offline spike sorting [3] and the gait cycle defined as foot strike (0%), maxima of hindlimb angle (70%), and next foot strike (100%) for the right hindlimb. Firing times were converted to cycle percentage and a PESH was created over all gait cycles.

Abstract received October 25, 2012. This work supported in part by the European Commission’s 7th Framework Programme (CP-IP 258654).

J. DiGiovanna is with the Translational Neural Engineering Lab, Center for Neuroprosthetics, EPFL Lausanne, Switzerland (jack.digiovanna@epfl.ch).

G. Courtine is with the Center for Neuroprosthetics and Brain Mind Institute, EPFL Lausanne, Switzerland (gregoire.courtine@epfl.ch)

S. Micera is with the Translational Neural Engineering Lab, EPFL Lausanne, Switzerland and the BioRobotics Institute, Scuola Superiore Sant’Anna, Pisa, Italy (silvestro.micera@epfl.ch).

III. NEURAL MODULATION FITTING

BARS [4] is well suited for neural data analysis because it captures rapid transitions in firing rates while incorporating the concept that neurons are noisy and usually change firing rate slowly [5]. Fig 1 shows a representative PESH and both fits. BARS has been used for sensory responses [6], here it is used for motor commands. Across six rats (n=282 neurons), the mean RMS fit errors were 4.8 and 3.5 Hz for 5th order polynomials and BARS fits respectively. Additionally, BARS had 53.5% higher energy (averaged over all frequencies), preserving high frequency information as expected.

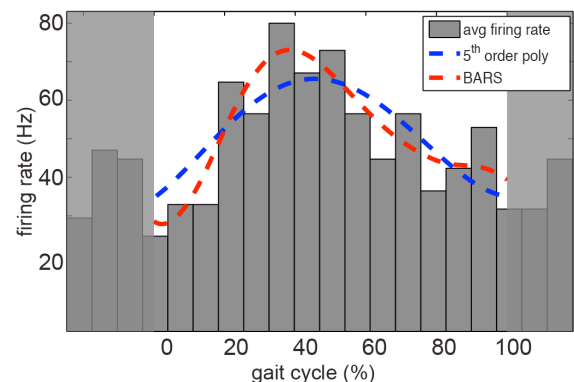


Figure 1: Fitted PESH (n = 38 gait cycles) of neuronal modulation over gait cycle. BARS fitting better captures the rapid excitation at 20% and peak at 35%. Shaded areas were included to avoid edge-effects in both fits.

In summary, BARS better captures transitions in motor cortex modulations during locomotion; this sets the foundation for analyzing the relationship between the neural ensemble and muscular activity, which also contains sharp transitions.

REFERENCES

- [1] T. P. Trappenberg, *Fundamentals of Computational Neuroscience*. New York: Oxford University Press, 2002.
- [2] D. Purves, G. J. Augustine, D. Fitzpatrick, W. C. Hall, A.-S. LaMantia, J. O. McNamara, and S. M. Williams, Eds., *Neuroscience*. Sunderland, MA: Sinauer Associates, Inc., 2004.
- [3] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering," *Neural Computation*, vol. 16, pp. 1661-87, 2004.
- [4] I. Dimatteo, C. R. Genovese, and R. E. Kass, "Bayesian curve fitting with free, knot splines," *Biometrika*, vol. 88, pp. 1055-1071, 2001.
- [5] R. E. Kass, V. Ventura, and E. N. Brown, "Statistical issues in the analysis of neuronal data," *Journal of Neurophysiology*, vol. 94, pp. 8-25, 2005.
- [6] S. Behseta and R. E. Kass, "Testing equality of two functions using BARS," *Statistics in Medicine*, vol. 24, pp. 3523-3534, 2005.

Applications of ultrasonic time-reversal in biomedical imaging

Mathias Fink*

* Langevin Institute, ESPCI ParisTech, CNRS, 1 rue Jussieu, 75005, Paris, France

This talk will present an overview of the research conducted on ultrasonic time-reversal methods applied to biomedical imaging.

We will first describe iterative time-reversal techniques that allow tracking and focusing ultrasonic waves on reflectors in tissues (kidney stones, micro-calcifications). Because time reversal is also able to correct for the strong distortions induced by the skull bone on ultrasonic propagation, this adaptive focusing technique allows non-invasive therapy of brain diseases and high resolution brain neurostimulation. We will also show that time-reversal focusing does not need the presence of bright reflectors but it can be achieved only from the speckle noise generated by random distributions of non-resolved scatterers. We will describe the applications of this concept to correct distortions and aberrations in ultrasonic imaging.

In the second part of the talk we will describe the concept of time-reversal processors to get ultrafast ultrasonic images with typical frame rates of order of 10.000 F/s. It is the field of ultrafast ultrasonic imaging that have plenty medical applications. We will focus on two of them: Quantitative Elasticity imaging of tissues by following shear wave propagation and Ultrafast Doppler imaging that allows ultrasonic functional imaging of the brain activity.

Riemann-Finsler Geometry and its Applications to Diffusion Magnetic Resonance Imaging

Luc Florack*, Andrea Fuster*, and Tom Dela Haije*

* Imaging Science & Technology Eindhoven, Eindhoven University of Technology, NL-5600 MB Eindhoven, The Netherlands.

Abstract—Riemannian geometry has become a popular mathematical framework for the analysis of diffusion tensor images (DTI) in diffusion weighted magnetic resonance imaging (DWMRI). If one declines from the a priori constraint to model local anisotropic diffusion in terms of a 6-degrees-of-freedom rank-2 DTI tensor, then Riemann-Finsler geometry appears to be the natural extension. As such it provides an interesting alternative to the Riemannian rationale in the context of the various high angular resolution diffusion imaging (HARDI) schemes proposed in the literature. The main advantages of the proposed Riemann-Finsler paradigm are its manifest incorporation of the DTI model as a limiting case via a “correspondence principle” (operationalized in terms of a vanishing Cartan tensor), and its direct connection to the physics of DWMRI expressed by the (appropriately generalized) Stejskal-Tanner equation and Bloch-Torrey equations furnished with a diffusion term.

I. INTRODUCTION

Riemann-Finsler geometry, already hinted upon by Riemann in his “Habilitation” [1], is a generalization of Riemannian geometry. The latter has found important applications in Maxwell theory and Einstein’s theory of general relativity, contributing greatly to its popularity. The general case was taken up by Finsler [2], Cartan [3] (referring to it as “Finsler geometry”), and others [4], [5], [6].

Despite its great potential, Riemann-Finsler geometry has not become nearly as popular as its Riemannian limit. To some extent this may be explained by its rather mind-boggling technicalities and heavy computational demands (due to the introduction of an additional vectorial dimension extending the base manifold). Another key factor is the still open challenge to find important “natural” application areas for it, and to show its added value in these areas. We conjecture that DWMRI could be one such application area. This imaging modality plays an important role in the unravelment of the human brain connectome, among others.

II. THEORY

The pivot of Riemann-Finsler geometry is a generalised notion of length of a spatial curve C (“Hilbert’s invariant integral” [4]):

$$\mathcal{L}(C) = \int_C F(x, dx). \quad (1)$$

The so-called *Finsler function* $F(x, \xi)$ is positive definite for $\xi \neq 0$, and homogeneous of degree one in ξ , i.e. $F(x, \lambda\xi) = |\lambda|F(x, \xi)$ for all λ . In addition, the *Riemann-Finsler metric tensor*, defined as

$$g_{ij}(x, \xi) = \frac{1}{2} \frac{\partial^2 F^2(x, \xi)}{\partial \xi^i \partial \xi^j}, \quad (2)$$

is positive definite. It is easy to see that (applying summation convention)

$$F(x, \xi) = \sqrt{g_{ij}(x, \xi) \xi^i \xi^j}. \quad (3)$$

Riemann’s “quadratic restriction” pertains to the “mildly anisotropic” case $g_{ij}(x, \xi) = g_{ij}(x)$.

The non-Riemannian nature of the Riemann-Finsler manifold is most concisely expressed in terms of the so-called *Cartan tensor*:

$$C_{ijk}(x, \xi) = \frac{1}{4} \frac{\partial^3 F^2(x, \xi)}{\partial \xi^i \partial \xi^j \partial \xi^k}. \quad (4)$$

A dual, or Hamiltonian formulation rests upon the identity

$$g^{ik}(x, y) g_{kj}(x, \xi) = \delta_j^i, \quad (5)$$

in which the first factor on the l.h.s. defines the *dual Riemann-Finsler metric tensor*, and in which it is understood that

$$y_i = g_{ij}(x, \xi) \xi^j \quad \text{or, equivalently,} \quad \xi^i = g^{ij}(x, y) y_j. \quad (6)$$

We stipulate that the *dual Finsler function*, $H(x, y) \equiv F(x, \xi)$, governs signal attenuation in DWMRI if, as with DTI, one relies on the Stejskal-Tanner (mono-exponential decay, Gaussian diffusion) and Bloch-Torrey equations with diffusion term [7], [8], [9], viz.

$$S(x, y) = S(x, 0) \exp(-\tau H^2(x, y)). \quad (7)$$

Here τ denotes a time constant related to the time Δ between a pair of balanced diffusion-sensitizing gradients G_i and pulse duration δ (in Stejskal-Tanner’s scheme we have $\tau = \Delta - \delta/3$), and

$$H(x, y) = \sqrt{g^{ij}(x, y) y_i y_j}, \quad (8)$$

with “momentum” $y_i = \gamma \delta G_i$ given in terms of δ , G_i , and hydrogen gyromagnetic ratio γ . The DTI rationale [10], [11] is based on the (strong) simplification that the (y -independent) diffusion tensor image $D^{ij}(x)$ can be identified with $g^{ij}(x, y)$.

Further details can be found in a forthcoming publication [12].

REFERENCES

- [1] B. Riemann, “Über die Hypothesen, welche der Geometrie zu Grunde liegen,” in *Gesammelte Mathematische Werke*, H. Weber, Ed. Leipzig: Teubner, 1892, pp. 272–287.
- [2] P. Finsler, “Ueber kurven und Flächen in allgemeinen Räumen,” Ph.D. dissertation, University of Göttingen, Göttingen, Germany, 1918.
- [3] E. Cartan, *Les Espaces de Finsler*. Paris: Hermann, 1934.
- [4] D. Bao, S.-S. Chern, and Z. Shen, *An Introduction to Riemann-Finsler Geometry*, ser. Graduate Texts in Mathematics. New York: Springer-Verlag, 2000, vol. 2000.
- [5] H. Rund, *The Differential Geometry of Finsler Spaces*. Berlin: Springer-Verlag, 1959.
- [6] —, *The Hamilton-Jacobi Theory in the Calculus of Variations*. Huntington, N.Y.: Robert E. Krieger Publishing Company, 1973.
- [7] F. Bloch, “Nuclear induction,” *Physical Review*, vol. 70, pp. 460–473, 1946.
- [8] E. O. Stejskal and J. E. Tanner, “Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient,” *Journal of Computational Physics*, vol. 42, no. 1, pp. 288–292, 1965.
- [9] H. C. Torrey, “Bloch equations with diffusion terms,” vol. 104, pp. 563–565, 1956.
- [10] P. J. Basser, J. Mattiello, and D. Le Bihan, “MR diffusion tensor spectroscopy and imaging,” *Biophysics Journal*, vol. 66, no. 1, pp. 259–267, 1994.
- [11] D. Le Bihan, J.-F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriet, “Diffusion tensor imaging: Concepts and applications,” *Journal of Magnetic Resonance Imaging*, vol. 13, pp. 534–546, 2001.
- [12] L. M. J. Florack and A. Fuster, “Riemann-Finsler geometry for diffusion weighted magnetic resonance imaging,” Eindhoven University of Technology, Tech. Rep. CASA-Report 12-23, cf. www.win.tue.nl/analysis/reports/rana12-23.pdf, 2012, to be published in the Lecture Notes in Computer Science.

On the Riemannian Rationale for Diffusion Tensor Imaging

Andrea Fuster, Tom Dela Haije, and Luc Florack

Imaging Science & Technology Eindhoven (IST/e), Eindhoven University of Technology, The Netherlands

Abstract—One of the approaches in the analysis of brain diffusion MRI data is to consider white matter as a Riemannian manifold, with a metric given by the inverse of the diffusion tensor. Such a metric is used for white matter tractography and connectivity analysis. Although this choice of metric is heuristically justified it has not been derived from first principles. We propose a modification of the metric tensor motivated by the underlying mathematics of diffusion.

I. INTRODUCTION

A possible approach to study white matter from diffusion MRI is to consider a geometric framework in which quantities of interest, such as connectivity measures, are derived from a Riemannian metric. In this way white matter is represented as a Riemannian manifold, and candidate neural fibres are postulated to coincide with geodesic curves. The common choice in the literature is to consider a Riemannian metric given by the inverse of the diffusion tensor D [1], [2]. The intuitive idea behind this choice is to transform anisotropic diffusion in Euclidean space to (free) Brownian diffusion in a curved Riemannian space. Brownian motion is characterized by an infinitesimal diffusion generator \mathcal{L} which can be written as $\mathcal{L} = \Delta_{\text{LB}}$, with Δ_{LB} the Laplace-Beltrami operator for the appropriate metric tensor g [3]. Such a generator is called an intrinsic Laplacian. However, the usual choice of metric from an anisotropic diffusion generator, $g = D^{-1}$, does not lead to Brownian motion in the corresponding curved space. We propose a slight modification of the Riemannian metric in order to accomplish this. The question of how to choose an appropriate metric has already been addressed in [4], [5].

II. DISCREPANCY

Inhomogeneous anisotropic diffusion is commonly described by the generator

$$\mathcal{L} = \partial_i (D^{ij} \partial_j) = D^{ij} \partial_i \partial_j + (\partial_j D^{ij}) \partial_i \quad (1)$$

where $i, j = 1, 2, 3$, D^{ij} is the diffusion tensor, $\partial_i = \partial/\partial x^i$, and in which we use Einstein's summation convention. A Riemannian metric $g_{ij} = D_{ij}$ can be introduced, where D_{ij} is the inverse diffusion tensor.

The generator (1) can then be expressed as

$$\mathcal{L} = \Delta_g - \sqrt{d} \left(\partial_j \frac{1}{\sqrt{d}} \right) D^{ij} \partial_i \quad (2)$$

where d is the determinant of the diffusion tensor D^{ij} and Δ_g is the Laplace-Beltrami operator

$$\Delta_g = \frac{1}{\sqrt{g}} \partial_j (\sqrt{g} g^{ij} \partial_i) \quad (3)$$

Here, $g = \det g_{ij}$. In our case, $g_{ij} = D_{ij}$, we have

$$\Delta_g = D^{ij} \partial_i \partial_j + \sqrt{d} \partial_j \left(\frac{1}{\sqrt{d}} D^{ij} \right) \partial_i \quad (4)$$

From Eq. (2) we see that the usual identification $g = D^{-1}$ does not lead to Brownian motion on the manifold (M, g) since the diffusion generator \mathcal{L} is not an intrinsic Laplacian. This is only the case when the second term on the right-hand side of Eq. (2) vanishes, which occurs for $d = \det D^{ij}$ constant. Clearly, this cannot be assumed in general.

III. PROPOSAL

Consider now the diffusion generator given by

$$\tilde{\mathcal{L}} = d^{-1} \mathcal{L} = d^{-1} D^{ij} \partial_i \partial_j + d^{-1} (\partial_j D^{ij}) \partial_i \quad (5)$$

where we use the same notation as in section II. Again a Riemannian metric can be introduced, namely, $\tilde{g}_{ij} = d D_{ij}$. It can be shown that

$$\tilde{\mathcal{L}} = \Delta_{\tilde{g}} \quad (6)$$

The generator (5) is therefore an intrinsic Laplacian, and the proposed choice of metric results in Brownian motion on the manifold (M, \tilde{g}) .

IV. DISCUSSION

We propose a new Riemannian metric in the context of diffusion tensor imaging, motivated by first principles. In future work experiments will be performed to assess whether our modified metric leads to improved results for tractography and connectivity analysis in comparison to the usual choice of metric. It would also be very interesting to clarify the relation to other modified Riemannian metrics, such as the one in [4].

REFERENCES

- [1] L. O'Donnell, S. Haker, and C.-F. Westin, "New approaches to estimation of white matter connectivity in diffusion tensor MRI: elliptic PDEs and geodesics in a tensor-warped space," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2002*, T. Dohi and R. Kikinis, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, vol. 2488, pp. 459–466.
- [2] C. Lenglet, R. Deriche, and O. Faugeras, "Inferring white matter geometry from diffusion tensor MRI: application to connectivity mapping," in *Lecture Notes in Computer Science, T. Pajdla and*. Springer, 2004, pp. 127–140.
- [3] M. C. de Lara, "Geometric and symmetry properties of a nondegenerate diffusion process," *The Annals of Probability*, vol. 23, no. 4, pp. 1557–1604, Oct. 1995.
- [4] X. Hao, R. Whitaker, and P. Fletcher, "Adaptive Riemannian metrics for improved geodesic tracking of white matter," in *Information Processing in Medical Imaging*, 2011, p. 1324.
- [5] S. Jbabdi, P. Bellec, R. Toro, J. Daunizeau, M. Plgrini-Issac, and H. Benali, "Accurate anisotropic fast marching for diffusion-based geodesic tractography," *International Journal of Biomedical Imaging*, vol. 2008, pp. 1–12, 2008.

Single Trial ERP detection Exploiting Sparsity in Time

Foad Ghaderi

Robotics Group, University of Bremen, Bremen, Germany

Abstract—Sparsity of event related potentials (ERP) in time domain is used to develop a single trial ERP detection method. It is assumed that ERPs are deterministic whereas the ongoing EEG does not have this property [1]. Using the recursive projected compressive sensing (ReProCS) method [2] the electroencephalogram (EEG) data is decomposed into a high dimensional, low rank component and a sparse component which represents the ERPs. The method shows promising results on EEG data which is synthetically generated using real EEG and ERP samples.

I. INTRODUCTION

Brain responses to low-probability stimuli (targets) inter-mixed with high-probability stimuli (non-targets) differ in time domain such that only target stimuli evoke a type of ERP called P300. Low signal to noise ratio of the ERP signals is a major drawback for their usability in systems which are based on single trial ERP detection.

II. METHOD

The problem is formulated as $M_t = L_t + S_t$, where the $n \times 1$ vector M_t is the measurement vector at time t , L_t is low rank and dense, and S_t is a sparse vector. In other words, it is assumed that $L_t = Ux_t$, where U and x_t are respectively an unknown orthonormal matrix and an uncorrelated sparse vector whose support is N_t . The principal components (PC) matrix (the columns of $P_t = (U)_{N_t}$) span L_t , i.e., $L_t = P_t(x_t)_{N_t}$.

The ReProCS method can separate correlated sparse data from low rank subspace [2]. It is assumed that an initial estimate of the PCs of L_t exists at time t , and the goal is to use the new measurement vectors to estimate S_t , L_t , and hence update the PC matrix, \hat{P}_t .

If \hat{P}_t is an estimate of P_t at time t , and $\hat{P}_{t,\perp}$ is an orthogonal complement of \hat{P}_t , we can rewrite M_t as $M_t = \hat{P}_t\alpha_t + \hat{P}_{t,\perp}\beta_t + S_t$, where $\alpha_t = \hat{P}_t' L_t$ and $\beta_t = (\hat{P}_{t,\perp})' L_t$. By projecting M_t into the space spanned by $(\hat{P}_{t,\perp})'$, i.e., $y_t = (\hat{P}_{t,\perp})' M_t = (\hat{P}_{t,\perp})' S_t + \beta_t$, most of the contribution of the low rank part of the measurement vector is nullified. If $\|\beta_t\|_2$ is small, the problem of estimating S_t will convert to the classical noisy sparse reconstruction problem.

ERPs are always time locked to some stimuli which are known and are captured simultaneously with the EEG signals. Windows of length l (usually $l < 1s$) starting from the onset of each stimulus are cut to represent the ERP samples. The EEG stream (n channels) is then cut into windows of length w ($w < l$), and the data in each window is vectorized to construct an $(n.w.F_s) \times 1$ vector M_t , where F_s is the sampling frequency. The objective is to decompose M_t into L_t and the sparse part S_t . Because the EEG part in the vectorized windows is low rank, and ERPs and the ongoing EEG originate from different sources, ReProCS can be utilized for separation. The support of the sparse signal in each M_t is known *a priori* which is also used in the sparse reconstruction phase. Assuming that T_t^c is the complement of the support of S_t at time t , the sparse signal can be recovered by solving the following:

$$\min_{S_t} \|(S_t)_{T_t^c}\|_1 \text{ subject to } \|y_t - (\hat{P}_{t,\perp})' S_t\|_2 \leq \epsilon. \quad (1)$$

The proposed ERP detection method is summarized as follows:

- 1) Perform principal component analysis on $[M_1, \dots, M_m]^*$ and select significant components as PC matrix, \hat{P}_{t_0} .

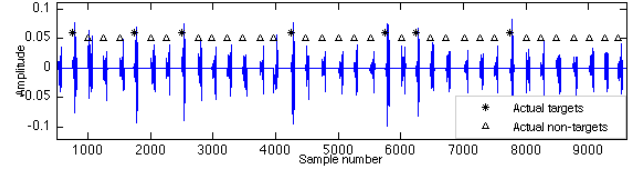


Fig. 1: The ERPs of channel PoZ recovered from applying the method to 8 channel EEG data.

TABLE I: Averages and standard deviations of balanced accuracies (BA) resulted from the proposed method on different simulated datasets. Scaling ratio = $(\text{norm}(\text{targets}) + \text{norm}(\text{non-targets})) / (2 * \text{norm}(\text{EEG}))$, where $\text{norm}(x) = \frac{\|x\|_1}{|x|}$, $|x|$ is the cardinality of x .

	Scaling ratio						
	0.09	0.11	0.14	0.16	0.18	0.21	0.23
BA (average)	0.61	0.78	0.83	0.87	0.89	0.91	0.92
BA (std)	0.03	0.02	0.02	0.03	0.02	0.02	0.02

- 2) $\hat{P}_{t,\perp} \leftarrow \hat{P}_{t-1,\perp}$, update T_t using the timing of the stimuli,
- 3) Estimate S_t using (1),
- 4) Estimate $\hat{L}_t = M_t - \hat{S}_t$,
- 5) Update \hat{P}_t using Recursive PCA algorithm [2], go to step 2.

* To initialize P_t , there should not exist any ERP samples in M_i , $\forall 1 \leq i \leq m$.

Each recovered sparse vector \hat{S}_t is restructured in the matrix form to represent the corresponding ERPs.

III. RESULTS

The proposed method was used to detect ERPs in synthetically mixed EEG signals. For this, target and non-target ERPs were extracted and averaged from EEG data recorded in a previous study [3]. The ERPs were then scaled and mixed (with the ratio of 1:5) in random sequence and added to EEG from a different dataset. Figure 1 shows the result of the algorithm on one of the datasets. To discriminate target and non-target ERPs, a threshold is defined as $\text{thr} = c * \frac{1}{K} \sum_{k=1}^K \max(\mathbf{e}_k)$, where \mathbf{e}_k is the k th ERP, K is the total number of ERPs and $1.2 \leq c \leq 1.4$.

Balanced accuracies ($0.5 * [\frac{TP}{TP+FN} + \frac{TN}{TN+FP}]$) of the experiments averaged over 20 runs are reported in Table 1. TP, FN, TN, and FP, represent the number of true positives, false negatives, true negatives, and false positives, respectively.

IV. CONCLUSION

Sparsity of the ERPs in time domain was used to develop a method for discriminating target from non-target ERPs. The method applied to synthetically mixed EEG data and the results were promising with simple thresholding. Future work includes using classifiers in the process flow and applying the method to real EEG signals.

REFERENCES

- [1] A. Ivannikov et al., "ERP denoising in multichannel EEG data using contrasts between signal and noise subspaces," *J. of Neurosci. Methods*, vol. 180, no. 2, pp. 340–351, 2009.
- [2] C. Qiu et al., "ReProCS: A missing link between recursive robust PCA and recursive sparse recovery in large but correlated noise," *CoRR*, 2011.
- [3] J. H. Metzen et al., "Minimizing calibration time for brain reading," in *Proc. 33rd int. conf. on Pattern recog.*, ser. DAGM'11, 2011, pp. 366–375.

Local vs global tractography to study brain connectivity

Alessandra Griffo^{†§} Alia Lemkaddem[†] Alessandro Daducci[†] Jean-Philippe Thiran^{†§}
[†] École Polytechnique Fédérale de Lausanne (EPFL), Signal Processing Laboratory (LTS5)
[§] University Hospital Center (CHUV) and University of Lausanne (UNIL), Department of Radiology

Since its advent, diffusion MRI has gained increasing interest for its extraordinary ability to probe the micro-structure of tissues. In the white matter this allows to infer the major axonal tracts and estimate a map of these connections (brain network). The reliability of these connectivity measures have been shown to be influenced not only by the quality of acquired data, but also by the choices of (i) the *acquisition scheme* [1], (ii) the *local reconstruction* model used to recover the intra-voxel structure [2] and (iii) the *fiber-tracking* algorithm to estimate the underlying fiber bundles [3]. The growing interest in understanding the structural connectivity of the brain has significantly boosted the development of these key aspects. In particular, in the last decade a wide gamma of tractography algorithms have been proposed. The simplest and most widely used are local approaches, which reconstruct the tracts based on information from individual or neighboring voxels, in a deterministic or stochastic vision of the problem. Recently, global approaches have been introduced. They tackle the estimation of the tracts as a global optimization problem and have been shown to outperform local methods [4]. As these techniques are potentially of high clinical value, it is very important to characterize and understand in detail their behavior. In our work we extend previous findings by investigating more systematically the reproducibility of structural brain networks as estimated by diffusion MRI, studying the influence of both the acquisition scheme and the tractography approach (local or global) adopted.

Five subjects, all female and aged between 24 and 30 years, underwent three consecutive MRI scanning sessions which included 1 high resolution T1-weighted acquisition and 3 different diffusion sequences. Diffusion was acquired using classical acquisition schemes: single shell at low b-value, or DTI (20 directions with b-value 1000 s/mm² and voxel size 2.2x2.2x3 mm), single shell at medium b-value, or HARDI (64 directions with b-value 3000 s/mm² and voxel size 2x2x3 mm) and dense cartesian sampling, or DSI (515 q-space samples with maximum b-value 8000 s/mm² and voxel size 2.2x2.2x3 mm). The T1-weighted volume was used as anatomical reference to segment the brain into white, gray and CSF using FREESURFER. The cortical and subcortical parenchyma was further subdivided into 82 regions (excluding the brain stem) according to the Desikan-Killiany atlas. The estimated regions of interest (ROI) were then linearly registered to the diffusion space using FSL.

Local tractography was performed using the standard streamline algorithm and exploiting the diffusion profiles estimated in each voxel to propagate the tracts. 32 seed points per voxel were randomly generated and an angle threshold of 35° was adopted. We used DIFFUSION TOOLKIT to reconstruct the intra-voxel fiber structure; in particular, a single tensor was fitted from DTI data, classical Q-Ball Imaging was used for HARDI scans and standard Diffusion Spectrum Imaging was used for cartesian sampling acquisitions. *Global tractography* was performed on the acquired raw signal using the algorithm described by [5]. For each acquisition scheme and tractography approach, the corresponding connectivity matrices (brain networks) were estimated using the CONNECTOME MAPPER by intersecting the reconstructed tracts with the cortical-subcortical

ROIs previously segmented. The connectivity architecture of the estimated brain networks was characterized by means of standard graph measures [6]: *network density* (D), *global efficiency* (E) and *average clustering coefficient* (CC). To assess the reproducibility of such measures we used the *intraclass correlation coefficient* (ICC).

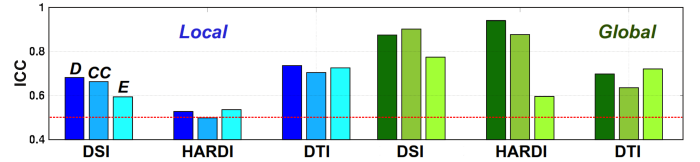


Fig. 1. Reproducibility of three well-established graph measures (D, CC and E) of the brain networks estimated by means of local and global tractography and using different acquisition schemes (DTI, HARDI and DSI).

Overall, network measures appeared clearly *most reproducible when using global tractography*. The reproducibility of DTI did not sensibly change whether local or global tractography was used. In case of local tracking, though, DTI seemed to produce more reproducible results than HARDI and DSI. This is a central observation, as local tractography combined with DTI had been previously associated with high robustness (and relatively poor sensitivity) [1]. On the other hand, however, the reproducibility of the network measures *dramatically improved* when global tractography was used in combination with HARDI or DSI, especially with respect to density D and clustering coefficient CC where the best results were reached (ICC \approx 0.9). In particular, the poor performance of local tracking with HARDI data might be caused by more sensitivity to propagate local estimation errors in the intra-voxel fiber structure due to more sensitivity to noise with only one shell at medium-high b-value. The work reported here is part of a larger study involving different institutions, where additional local models and tractography approaches are being investigated with the aim to provide a more comprehensive picture of the current processing pipelines panorama used for connectivity analyses with diffusion MRI.

REFERENCES

- [1] D. S. Bassett, J. A. Brown, V. Deshpande, J. M. Carlson, and S. T. Grafton, "Conserved and variable architecture of human white matter connectivity," *NeuroImage*, pp. 1262–1279, 2011.
- [2] T.-S. Yo, A. Anwander, M. Descoteaux, P. Fillard, C. Poupon, and T. R. Knösche, "Quantifying brain connectivity: A comparative tractography study," in *14th MICCAI*, 2009, pp. 886–893.
- [3] P. Fillard, M. Descoteaux, A. Goh, S. Gouttard, B. Jeurissen, J. Malcolm, A. Ramirez, M. Reisert, K. Sakaie, F. Tensauti, T. Yo, J.-F. Mangin, and C. Poupon, "Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom," *NeuroImage*, pp. 220–234, 2011.
- [4] L. Li, J. Rilling, T. M. Preuss, M. F. Glasser, F. W. Damen, and X. Hu, "Quantitative assessment of a framework for creating anatomical brain networks via global tractography," *NeuroImage*, pp. 1017–1030, 2012.
- [5] M. Reisert, I. Mader, C. Anastasopoulos, M. Weigel, S. Schnell, and V. Kiselev, "Global fiber reconstruction becomes practical," *NeuroImage*, pp. 955–962, 2011.
- [6] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *NeuroImage*, pp. 1059–1069, 2010.

The importance of priors for l2 regularization and total variation methods in quantitative susceptibility mapping

Diana Khabipova*, Yves Wiaux^{‡¶} † Rolf Gruetter^{*†¶} and José P. Marques^{*†}

* Laboratory for Functional and Metabolic Imaging, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

† Department of Radiology, University of Lausanne (UNIL)/CHUV, Switzerland

‡ Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

¶ Department of Radiology, University of Geneva (UniGE), Switzerland

Abstract—Quantitative susceptibility mapping in MRI is an ill-posed problem. In this abstract a systematic evaluation of an l2 method and an l1 total variation method (TV) is performed, using numerical and experimental data and different morphology priors, which is shown to make both methods more independent from their regularization parameters.

Introduction. Phase imaging has been demonstrated to offer a good contrast between and within brain tissues at 7T [1]. However, phase imaging suffers from a non-local contrast variation which can be overcome by calculating the underlying magnetic susceptibility maps χ [2]. As this problem is ill-posed, many regularization methods have been proposed over the past years [2], [3], [4], [5]. In this abstract we do a thorough comparison of some of these methods.

Subjects and Methods. Two methodologies were implemented: Firstly an l2 method [2] uses a least-square conjugated algorithm to calculate the minimum $\min_{\chi} (\|W(\mathbf{F}^{-1}\mathbf{C}\mathbf{F}\chi - \delta B)\|_2^2 + \beta\|M_{B/C}\nabla\chi\|_2)$, where $\mathbf{F}^{-1}\mathbf{C}\mathbf{F}$ represents the convolution with the dipole kernel, δB is the measured field, W is a diagonal weighting matrix inversely proportional to the noise standard deviation in the measured field, β is a regularization parameter. The diagonal matrices M_B and M_C enable local weighting based on prior information from the magnitude (M_{image}). They are respectively defined as $M_B = \{0, \text{if } \nabla M_{image} > n\sigma; 1, \text{if } \nabla M_{image} \leq n\sigma\}$, $M_C = \{1/\text{res}, \text{if } \nabla M_{image} > n\sigma; 1/\nabla M_{image}, \text{if } \nabla M_{image} \leq n\sigma\}$, where σ is the noise standard deviation and n is a threshold parameter, that indirectly defines the amount of random noise present in the prior information masks.

Secondly, an l1 TV method [4] minimizes the TV-norm of χ subject to the data constraint $\min_{\chi} \|M_{B/C}\nabla\chi\|_1$ s.t. $\|W(\mathbf{F}^{-1}\mathbf{C}\mathbf{F}\chi - \delta B)\|_2^2 < \varepsilon$, where ε can be measured from the data.

Simulations and Experiments. A synthetic field map was calculated using a numerical phantom with seven susceptibility compartments with different susceptibility values. Random noise was added to achieve an SNR of 10.

Experimental data were acquired using a 3D multi-echo gradient echo sequence with the following acquisition parameters: TR = 49 ms, 5 echoes TE = (3.35-34.71) ms, BW = 260 Hz/Px, 1 mm isotropic spacial acquisition. A 26 years old, healthy female volunteer was scanned. The phase images were unwrapped, the background removal was performed using the SHARP method [6].

Susceptibility maps were calculated with both methods while the prior information parameter n as well as the parameters ε and β were varied systematically. The quality of the reconstruction was measured as the power of the difference to the original phantom (ground truth for numerical data).

Results. Using the binary prior information, the optimum ε value (ε_{opt}) remains constant throughout different n , while the optimum β value (β_{opt}) increases with the reduction of n , 1a),c). Both methods, l2 and l1 TV, benefit from having the threshold of the binary prior

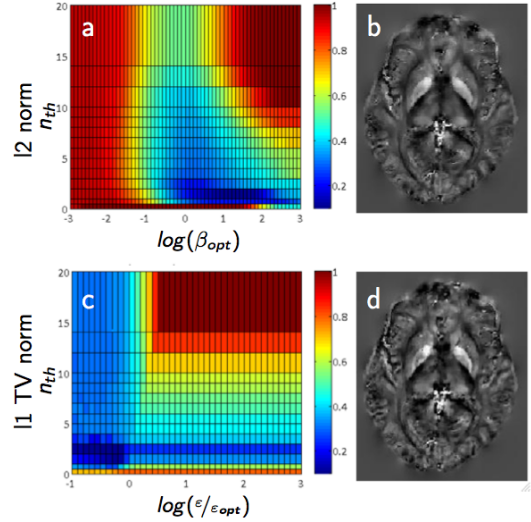


Fig. 1. first rows shows the reconstruction error dependence on regularization parameters a) β and c) ε in the x-axis and threshold value n for the Binary Priors and second row shows the quantitative susceptibility map with a),b) l2, c),d) l1 TV method information set at approximately twice the noise level as this gives both, the lowest reconstruction error and the highest independence from the regularization parameters ε and β used. The susceptibility maps reconstructed with l2, l1 TV method using optimal parameters are shown in 1b),d)

Discussion and Conclusion. The results show, that both methods are effective at calculating susceptibility maps. The binary mask outperformed the continuous mask on our simulations (data not shown). Moreover, TV reconstructions are less dependent on the prior information when noise has been correctly estimated. A selection of the optimum mask makes both methods, l1 TV and l2, more independent from their regularization parameters. The usage of low thresholds for the binary prior information allows a good compromise between: i) having as much morphological information as possible and ii) applying the spatial constraint in the smoothness (l2 method) or sparsity (l1 TV method) in enough contiguous areas in order to penalize magic angle related artifacts. Therefore, including all edge information, even if contaminated with background noise, is more important than excluding noise and edge information from the prior information. Using optimal parameters both methods show a comparable quality of the reconstructed susceptibility maps 1b),d).

REFERENCES

- [1] J. H. Duyn, *PNAS*, vol. 104, no. 28, pp. 11 796–11 801, 2007.
- [2] L. De Rochefort, *MRM*, vol. 63, no. 1, pp. 194–206, 2010.
- [3] B. Bilgic, *NeuroImage*, vol. 59, no. 3, pp. 2625–2635, 2012.
- [4] T. Liu, *MRM*, vol. 66, no. 3, pp. 777–783, 2011.
- [5] K. Shmueli, *MRM*, vol. 62, no. 6, pp. 1510–1522, 2009.
- [6] F. Schweser, *NeuroImage*, vol. 54, no. 4, pp. 2789–2807, 2011.

Optical Coherent Imaging from Tissue to Molecule

Theo Lasser*

* Laboratoire d'optique biomédicale, Faculté des Sciences et Techniques de l'Ingenieur, Ecole Polytechnique Federale de Lausanne

Imaging is key for medical diagnosis and provides new insights for the life sciences. Tissue, cell and subcellular structures can all be visualized using optical microscopy and so provide a variety of information with high spatial resolution. Structural information complemented by the functional information made possible by new optical techniques like FourierDomain Optical Coherence Tomography (FDOCT), Doppler Imaging and extended-focus Optical Coherence Microscopy (xf-OCM) Dark field Coherence Microscopy (df-OCM), and its latest extension photothermal optical lock-in Coherence Microscopy (poli-OCM) which allows extending these methods into the cellular dimensions. We will present selected examples ranging from retina blood flow, diabetes, Alzheimer's disease to brain research with an emphasis on the underlying optical concepts.

Water diffusion MRI: what are we looking at?

Denis Le Bihan*

* INRIA, 2004 Route des Lucioles, BP 93 - 06902 Sophia-Antipolis, France

In the mid 1980s it was shown that water diffusion in the brain could be imaged using MRI. During their random displacements water molecules probe tissue structure at a microscopic scale, thus providing unique information on the functional architecture of tissues. A dramatic application of diffusion MRI has been brain ischemia, following the discovery that water diffusion drops immediately after the onset of an ischemic event, when brain cells undergo swelling through cytotoxic edema. On the other hand, it was found that water diffusion is anisotropic in white matter, because axon membranes limit molecular movement perpendicularly to the fibers. This feature can be exploited to map out the orientation in space of the white matter tracks and image brain connections, as well as providing information on white track microstructure and integrity. Diffusion MRI has also the potential to give clues on the cellular organization within brain cortex on an individual basis, a step forward to segregating brain areas at mesoscale level. More recently, it has been shown that diffusion MRI could even be used to detect cortical activation. The diffusion response very closely reflects neuronal activity, as evidenced from electrophysiological recordings, and persists when the BOLD response is abolished in the presence of drugs interfering with the neurovascular coupling. This discovery represents a significant breakthrough, allowing non invasive access to a direct physiological marker of brain activation. This approach will bridge the gap between invasive optical imaging techniques in neuronal cell cultures, and current functional neuroimaging approaches in humans, which are based on indirect and remote blood flow changes. The aim of this presentation will be to review the general concepts, often overlooked or forgotten, that led to diffusion MRI as we know it, and to retrace the multidisciplinary and often complex evolution of the field since its birth circa 1985.

Imaging Tissue Magnetic Response with MRI

Chunlei Liu^{*†}

^{*} Brain Imaging and Analysis Center, Duke University, Durham, NC 27705.

[†] Department of Radiology, Duke University, Durham, NC 27710

Abstract—Magnetic response of biological tissues, though weak, becomes measurable under the strong magnetic field provided by MRI scanners. This interaction creates a unique tissue phase contrast which can be further used to quantify tissue-specific magnetic susceptibility. Susceptibility of white matter is found to be anisotropic and has to be described by a tensor. Imaging susceptibility tensor may allow visualization of brain connectivity in vivo.

I. INTRODUCTION

Magnetic response of biologic tissue is a fundamental process involved in MRI. Until very recently, however, B_0 -field induced magnetic field perturbation within tissues has largely been regarded as a source of image artifacts. This treatment was warranted as no good soft tissue contrast was readily available and large field offset creates unwanted distortions. Recent methodological developments in MRI signal processing and the availability of high-field magnets have started to reveal some unique contrast and meaningful information of tissue magnetic properties [1][2]. While higher field strength increases the contrast-to-noise ratio, methodological advances have been critical in quantifying minute field perturbations ($\sim 10^{-2} ppm$) induced by local susceptibility variations. Recent studies have significantly improved our understanding of tissue magnetic property with several interesting potential applications being reported. For example, susceptibility may be a good indicator of iron stores in certain brain regions and a good measure of myelination in brain white matter; susceptibility may also provide a new window to white matter architecture.

II. PHASE CONTRAST

Magnetic response is typically characterized by the quantity of magnetic susceptibility, the extent to which a material is magnetized by an applied magnetic field. Magnetic susceptibility, χ , is equal to the ratio of the magnetization \mathbf{M} to the applied magnetic field strength H , i.e. $\chi = \mathbf{M}/H$. This definition of susceptibility is the volume susceptibility or bulk susceptibility. In MRI images, it is the volume susceptibility representing the magnetism (a dipole moment) per voxel. Magnetic materials are classically classified as diamagnetic, paramagnetic, or ferromagnetic on the basis of their susceptibilities. Biological tissues can be either diamagnetic or paramagnetic depending on its molecular contents and microstructure. An overview of the effect of susceptibility in MRI can be found in [3]. While MRI signal originates from nuclear magnetization, the dominant magnetization that contributes to bulk susceptibility originates from orbital electrons.

The most commonly used sequence for measuring susceptibility is the spoiled-gradient-recalled-echo (SPGR or GRE) sequence. The phase of GRE images gives a measure of local frequency offset which in turn can be used to calculate susceptibility quantitatively. The magnitude of a series of multi-echo GRE images is used to estimate the $T2^*$ relaxation times.

III. QUANTITATIVE SUSCEPTIBILITY MAPPING

Although phase provides a unique contrast among a variety of tissues, phase is not an intrinsic measure of tissue property due to

the long range dipole field distribution. Quantitative susceptibility values can be calculated from background-phase removed frequency shift maps by solving a linear equation. The relationship between frequency shift $\Delta f(r)$ and the spatially distributed susceptibility $\chi(r)$ is expressed in a Fourier transform as

$$\Delta f = F^{-1}[\chi(k)/3 - k_z^2(k)\chi(k)/k^2]\gamma\mu_0 H_0 \quad (1)$$

Here $\chi(k)$ is the 3D Fourier transform of $\chi(r)$ and k is the spatial frequency vector. F^{-1} is the inverse Fourier transform. k is the spatial frequency vector; γ is the gyromagnetic ratio of water proton; μ_0 is the vacuum permeability; H_0 is the magnitude of the applied magnetic field; The equation can be inverted in the k -space. However, when $k^2 = 3k_z^2$ (magic angle), the coefficient is zero prohibiting a direct inversion. A simple strategy for avoiding dividing by zero is to threshold the coefficients. Other strategies utilize regularization, multiple angle acquisition [4] and the LSQR algorithm [5]. Compressed sensing is particularly useful for solving this ill-posed problem [6].

IV. SUSCEPTIBILITY TENSOR IMAGING

An interesting recent finding is the extensive susceptibility anisotropy in brain white matter. The anisotropy existed in susceptibility can be characterized by a second-order susceptibility tensor [5]. The resonance frequency shift Δf of each voxel is related to the spatially distributed macroscopic susceptibility tensors as [5]

$$\Delta f = F^{-1}[\hat{\mathbf{H}}^T \chi(\mathbf{k}) \hat{\mathbf{H}}/3 - \mathbf{k}^T \hat{\mathbf{H}} \frac{\mathbf{k}^T \chi(\mathbf{k}) \hat{\mathbf{H}}}{k^2}] \gamma\mu_0 H_0 \quad (2)$$

Here, $\hat{\mathbf{H}}$ is the unit directional vector of the applied magnetic field; χ is a second-order (or rank-2) susceptibility tensor. A minimum of 6 measurements of Δf at different orientations with respect to the \mathbf{H}_0 field is needed. This requires rotation of the object or the field. However, STI without rotation is also possible. STI may be used to track white matter fibers in the brain in vivo and non-invasively.

REFERENCES

- [1] E. Haacke, Y. Xu, Y. Cheng, and J. Reichenbach, "Susceptibility weighted imaging (swi)," *Magn Reson Med*, vol. 52, no. 3, pp. 612–618, 2004.
- [2] J. Duyn, P. van Gelderen, T. Li, J. de Zwart, A. Koretsky, and M. Fukunaga, "High-field mri of brain cortical substructure based on signal phase," *Proc Natl Acad Sci USA*, vol. 104, no. 28, pp. 11 796–11 801, 2007.
- [3] J. Schenck, "The role of magnetic susceptibility in magnetic resonance imaging: Mri magnetic compatibility of the first and second kinds," *Med Phys*, vol. 23, no. 6, pp. 815–850, 1996.
- [4] L. T., S. P., de Rochefort L., K. B., and W. Y., "Calculation of susceptibility through multiple orientation sampling (cosmos): a method for conditioning the inverse problem from measured magnetic field map to susceptibility source image in mri," *Magn Reson Med*, vol. 61, no. 1, pp. 196–204, 2009.
- [5] C. Liu, "Susceptibility tensor imaging," *Magn Reson Med*, vol. 63, no. 6, pp. 1471–1477, 2010.
- [6] B. Wu, W. Li, A. Guidon, and C. Liu, "Whole brain susceptibility mapping using compressed sensing," *Magn Reson Med*, vol. 67, pp. 137–147, 2011.

Rapid, Robust High Resolution MRI

Michael Lustig*, Department of Electrical Engineering and Computer Sciences, UC Berkeley, USA.

Magnetic Resonance Imaging (MRI) is a non-invasive imaging modality. Unlike Computed Tomography (CT), MRI does not use ionizing radiation. In addition, MRI provides a large number of flexible contrast parameters. These provide excellent soft tissue contrast. Since its invention more than 30 years ago, MRI has improved dramatically both in imaging quality and speed. This has revolutionized the field of diagnostic medicine. Imaging speed is a major part of this revolution as it is essential in many MRI applications. Improvements in MRI hardware and imaging techniques have enabled faster data collection, and hence faster imaging. However, we are currently at the point where fundamental physical and physiological effects limit our ability to simply encode data more quickly.

This fundamental limit has led many researchers to look for methods to reduce the amount of acquired data without degrading the image quality. These reduced sampling methods are based on the fact that MRI data is redundant, so the underlying information may be extracted from less measurements than traditionally considered necessary. One of the most significant clinical impacts of reduced sampling methods has been accomplished by parallel imaging with multiple receiver channels. Imaging with multiple channels provides more useful data per MRI acquisition, so fewer acquisitions are needed per scan. Another source of redundancy that has been gaining significant attention is the sparsity and compressibility of various MR signals. This effort has been motivated by the recent introduction of the theory of compressed sensing (CS).

Medical images, much like natural images taken by digital cameras can be compressed many folds (for example using the popular JPEG compression). The typical paradigm of compression is to first collect all the necessary data and then compress it. The question that arises is why is it necessary to collect so many measurements if most of the data is non-important? Compressed sensing provides a way to address this question. It is a new sampling theory for compressible signals that allows sampling at rates much lower than the Nyquist-rate. CS implicitly compresses data within the signal acquisition process by obtaining fewer so-called "incoherent measurements". This is accomplished through various non-uniform and pseudo-random k-space sampling schemes. Images can be accurately reconstructed from these measurements using non-linear recovery processes that enforce data consistency with the measurements and compressibility of the reconstruction. The practical result of CS in the context of MRI is that MR images require much less data for reconstruction, and hence can be scanned much faster.

Another model that has been gaining popularity is the so called low-rank model in which data is reorganized into matrices that exhibit low-rank properties. This model turns out to be particularly effective for dynamic imaging. Interestingly, parallel imaging reconstruction can also be posed as a low-rank reconstruction. This approach yields highly robust and effective algorithms for parallel imaging.

Beyond speed, sparsity and low-rank models provide additional advantages. The redundancy can not only be used to reduce the acquisition time. Instead, the redundancy can be used to provide more robust acquisitions that mitigate with many sources of artifacts in MRI such as: motion, relaxation, off-resonance and other data

corruptions. In these applications, incoherence is indeed the key to success.

The talk will cover the basics of compressed sensing theory in the context of MRI, incoherent sampling parallel imaging and low-rank models. It will also cover current status and trends in using these state of the art techniques for clinical applications. Finally, it will discuss some of the major hurdles for broad clinical adoption.

Accelerated Mouse Cardiac MR by Exploiting Retrospective Triggering Random Acquisition and CS Reconstruction

Abdallah G. Motaal*, Rui Castro[†], Luc Florack[†], Klaas Nicolay* and Gustav Strijkers*

* Biomedical NMR, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

[†] Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, Netherlands

Abstract—High temporal resolution cardiac Cine is an important diagnostic tool to assess heart diastolic function. A significant drawback is that in order to achieve high temporal resolution generally a long scan time is needed [1-3]. However, by exploiting the random nature of retrospectively triggered Cine, spatio-temporal sparsity, and by incorporating variable density k-space sampling, high acceleration factors could be achieved using compressed sensing reconstruction. In this work, mouse cardiac Cine is demonstrated with a high temporal resolution (90 frames/cardiac cycle) - i.e. time gap between two time frames = 1.2-1.5 ms \sim 4 times lower than TR - within a short acquisition time of 1-2 minutes.

I. METHODS

The scan was performed with a 9.4 T Bruker animal scanner. The undersampled 90-frames Cine was acquired with a retrospectively triggered FLASH sequence (matrix = 128x128, TR/TE = 4.7/2.35 ms). The retrospective triggering acquisition scheme satisfy randomness, i.e. high incoherence, which is important for CS reconstruction as shown in Fig.1a. Weighted k-space Cartesian sampling (Fig.1c) with scan time (1-2 min) was conducted, resulting in random and weighted sampled k - t space (Fig.1d). By exploiting the sparsity in the Cine-CMR imaging, accelerated Cine movies were reconstructed with a non-linear compressed sensing algorithm [4]. Mathematically, the following constrained optimization problem was solved:

$$\min \|\varphi(m)\|_1, \text{ subject to } \|F_s m - y\|_2 < \epsilon,$$

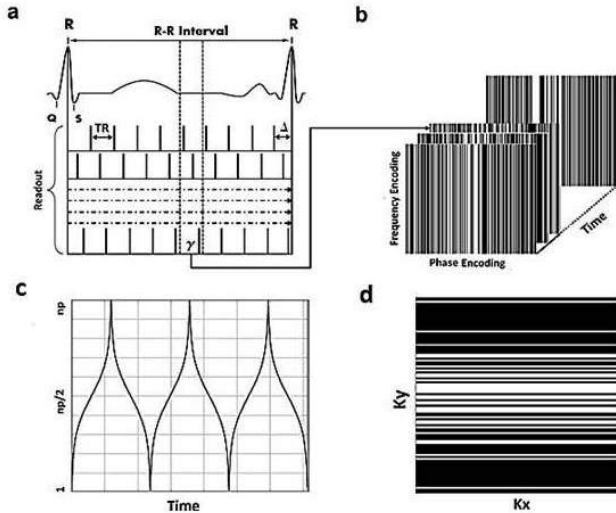


Fig. 1. The retrospectively triggered self-gated acquisition scheme and the resulting (b) k - t space after short acquisition time. (c) The weighted phase encoding sampling scheme. (d) The resultant randomly weighted Cartesian k -space.

II. RESULTS

In-vivo measurements using C57BL/6 male mice were performed. After 1-2 min acquisition, \approx 3X under-sampled k - t is acquired. Fig.2b shows the CS reconstruction after 2 min acquisition. The difference map for a time frame between the CS and gold standard reconstruction is shown in Fig.2c. Numerically, the RMSE errors were 0.0197 and 0.0206 for the 2 and 1.5 min CS reconstructions respectively. Besides, left ventricular functional parameters derived from the standard and the accelerated Cine movies were nearly identical. The E/A ratios of accelerated CS reconstructions were analyzed with respect to standard acquisition showing differences in E/A centered on a bias of just 5 %.

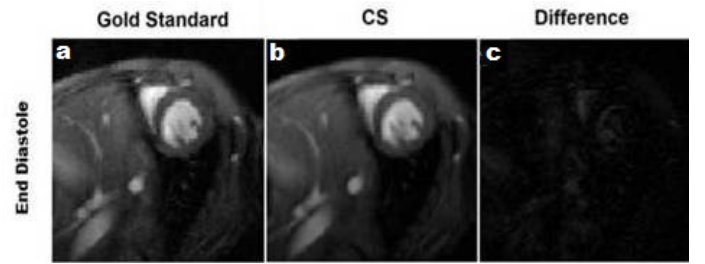


Fig. 2. (a-c) Gold standard, CS reconstructions and the difference map

III. DISCUSSION AND CONCLUSION

By using retrospectively triggered acquisition, random acquisition that satisfy incoherence is achieved. Besides, cardiac MR Cine with high temporal resolution, where the time gap between two time frames is much lower than TR, could be reconstructed. Finally, short scan time and accelerated reconstructions without significant losses in image quality and derived cardiac functional could be achieved by exploiting the spatio-temporal sparsity and using CS reconstruction. In this work, for a single slice, reconstructing 90 frames/cardiac cycle in 1-2 minutes becomes possible. Thus multi-slice imaging becomes feasible.

IV. ACKNOWLEDGMENT

This work is supported and funded by Imaging Science and Technology/Eindhoven (IST/e). We would like to thank Desiree Abdurrahim for helping in the animal experiments.

REFERENCES

- [1] Coolen et al, MRM, 2012.
- [2] Motaal et al, NMR Biomed, 2012.
- [3] Heijman et al, NMR Biomed, 2007.
- [4] Lustig et al, MRM, 2007.

Tissue microstructure recovery with dMRI: Computational challenges

E. Özarslan*

* Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02215, USA.

Abstract—Diffusion-weighted magnetic resonance imaging (DW-MRI) signals provide information about underlying tissue microstructure and architecture. Recovery of such important information requires robust and accurate computational methods. A number of practical issues (e.g., noise and limitations in acquisition time and gradient strength, etc.) specific to MRI acquisitions demand special computational tools tailored to address these issues. One promising method is the representation of the signal in a series expansion of basis functions whose properties are consistent with general DWI signal profiles.

The function of a biological tissue is determined largely by its structure. Processes such as activation, development, and aging as well as pathological perturbations lead to alterations at microscopic levels, which eventually influence the function of the tissue or organ. As a consequence, a probe that provides information about the microstructure of the tissue is expected to be of great value in many fields of biological research. The non-invasive character of MRI makes it particularly useful for diagnostic purposes, and an indispensable tool for routine clinical practice.

Perhaps the most significant shortcoming of traditional MRI acquisitions is its poor spatial resolution, which prohibits direct visualization of tissue microstructure. Diffusive-attenuation of the MR signal, e.g., achieved by applying pulsed field gradients (PFGs) [1], can be exploited to obtain the desired sensitization to the microscopic environment. This technique, whose inception precedes the invention of MRI, can be used for microstructure elucidation because distances traversed by randomly moving molecules during the course of the MR signal acquisition are in the micrometer range. Assuming that the microstructure does not vary drastically within each voxel, diffusion in different parts of the voxel is expected to be similar. Thus, there is no need to ‘localize’ or ‘isolate’ the signal for one microscopic domain from the signal in another domain as necessitated by traditional microscopy. Instead, we can collect a series of diffusion-weighted signal values for the same macroscopic voxel by varying the level of diffusion sensitization, and subsequently fit an appropriate model, descriptive of the microstructure, to the acquired signal profile. The end result yields the quantitative markers of tissue microstructure.

Extracting such quantitative measures assumes two interdependent steps: (i) a diffusion-weighted acquisition scheme that could generate a series of signal values, and (ii) a biophysical model that relates the microstructure to the diffusion sensitized signal. Neither of these steps is rigid, and a particular choice has to be made based on the characteristics of the tissue being examined. Most common approach to generate a profile of signal values involves repeating measurements by varying the PFGs’ strength and/or orientation, though variations in other parameters of the pulse sequence [2] or employing alternative sequences designed to yield the measured characteristics [3] could greatly strengthen the ability to obtain the desired information.

Many biophysical models with ever increasing complexity have been utilized in the literature. The parameters of such models are linked to the DW-MRI signal via various statistical features of the observed diffusional process. Among such features is the orientational

preference of diffusion, whose recovery has been instrumental in visualizing the neural pathways between the anatomically —hence functionally— connected regions of the central nervous system. Other applications have revealed many different microstructural characteristics of tissue such as cell dimensions, membrane permeability and exchange, apparent fractal dimension as a measure of microscopic complexity, and anisotropy in microscopic length scales.

Once the particular diffusion acquisition scheme and the biophysical model have been established, the challenge is the recovery of the microstructural features from the diffusion-attenuated signals. Among the essential characteristics of the magnitude-valued diffusion signal is that it is Rician distributed, which necessitates special signal transformation, recovery and regularization techniques [4]. Since acquisition time is limited, particularly in clinical settings, sampling is typically very sparse, and accurate interpolation methods are needed. This sampling typically occurs in the spatial frequency domain. Due to limitations on the magnitude of the diffusion gradients, the signal is always band-limited, and meaningful extrapolations of the signal would be helpful for high-resolution reconstructions. Other confounding factors include subject motion, and eddy current effects, which all demand tools tailored to the specific features of these factors. In recent years, we addressed some of these issues by expressing the signal in a series expansion of Hermite functions [5], and employed them in a number of problems to recover the microstructural features of the specimen.

Given the enormous complexity of typical biological tissues, numerous alternative measurement techniques, and availability of different theoretical approaches all suggest that there are many exciting questions that are yet to be asked. When complemented by robust and accurate computational tools, structure elucidation using diffusion-weighted MR techniques is expected to yield many novel quantitative markers of tissue microstructure, which could improve the diagnostic utility and specificity of diffusion-weighted MRI.

REFERENCES

- [1] E. O. Stejskal and J. E. Tanner, “Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient,” *J Chem Phys*, vol. 42, no. 1, pp. 288–292, 1965.
- [2] E. Özarslan, T. M. Shepherd, C. G. Koay, S. J. Blackband, and P. J. Basser, “Temporal scaling characteristics of diffusion as a new MRI contrast: findings in rat hippocampus,” *NeuroImage*, vol. 60, no. 2, pp. 1380–1393, 2012. [Online]. Available: <http://dx.doi.org/10.1016/j.neuroimage.2012.01.105>
- [3] E. Özarslan, “Compartment shape anisotropy (CSA) revealed by double pulsed field gradient MR,” *J Magn Reson*, vol. 199, no. 1, pp. 56–67, 2009. [Online]. Available: <http://dx.doi.org/10.1016/j.jmr.2009.04.002>
- [4] C. G. Koay, E. Özarslan, and P. J. Basser, “A signal transformational framework for breaking the noise floor and its applications in MRI,” *J Magn Reson*, vol. 197, no. 2, pp. 108–119, 2009. [Online]. Available: <http://dx.doi.org/10.1016/j.jmr.2008.11.015>
- [5] E. Özarslan, C. G. Koay, and P. J. Basser, “Simple harmonic oscillator based estimation and reconstruction for one-dimensional q-space MR,” in *Proc Intl Soc Mag Reson Med*, vol. 16, 2008, p. 35.

Ultra-Fast diffusion and resting-state fMRI imaging with Simultaneous Multi-Slice EPI and Q-space compressed sensing

Kawin Setsompop^{*†}

^{*} Athinoula A. Martinos Center for Biomedical Imaging, Dept. of Radiology, MGH, Charlestown, MA, USA.

[†] Harvard Medical School, Boston, MA, USA.

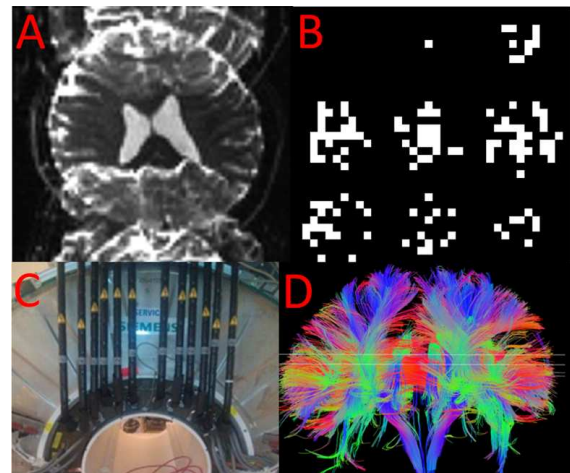
Diffusion and resting-state imaging have demonstrated the potential for non-invasive mapping of the structural and functional connectivity of the human brain in health and disease. Nonetheless, these methods face technical limitations arising from their limited sensitivity and specificity to connections as well as their clinical applicability. Specifically, the standard 2D EPI encoding used in these methods can result in lengthy and inefficient acquisitions. In this work, we utilized Simultaneous Multi-Slice (SMS) imaging and Q-space compressed sensing methodology to enable a dramatic, order of magnitude, speed up in image acquisition time while maintaining image quality and fidelity. We further derived metrics to aid the performance quantification of these acquisitions, which will help guide in parameters selection. The technology developed in this work will enable acquisition of high quality resting-state functional connectivity mapping and Diffusion Spectrum Imaging (DSI) data in a clinically relevant time frame, suitable for large Human Connectome clinical studies.

In SMS imaging [1]–[5], multiple imaging slices are excited and acquired simultaneously to accelerate the acquisition. Parallel imaging technique, which utilizes information from multiple receiver coils, is then used to untangle/unalias the simultaneously acquired imaging slices. In such acquisition, the CAIPIRINHA method [2] can be utilized to control the tangling/aliasing pattern of the imaging slices to significantly improve the conditioning of the unaliasing problem. The Blipped-CAIPI technique [6] is a variant of such technique that is applicable to EPI, where the controlled aliasing capability is achieved by playing out an extra set of magnetic encoding gradients along the slice direction during the EPI encoding. To reconstruct the Blipped-CAIPI SMS dataset, a k-space based parallel imaging algorithm, slice-GRAPPA, was created. With Blipped-CAIPI SMS technology, the acquisition speed of diffusion imaging can be reduced by 3-4x without significant noise amplification. For fMRI, this technology has been used to accelerate the acquisition up to a factor of 8x or higher.

In this work, Compressed Sensing (CS) is used to further accelerate the acquisition of Diffusion Spectrum Imaging (DSI). DSI is a particularly time consuming (1 hr) type of diffusion imaging acquisition that offers detailed information on complex distributions of intravoxel fiber orientations. It is possible to accelerate DSI by sub-Nyquist sampling of the q-space followed by nonlinear reconstruction to estimate the diffusion probability density functions (pdfs) [7]. As the performance of CS reconstruction depends strongly on the level of sparsity in the selected transform space, a dictionary specifically tailored for sparse representation of diffusion pdfs is used here to yield higher fidelity results [8]. With this technique, we reduce the scan time of whole brain DSI by 4x while retaining high image quality. Further, we demonstrate that a dictionary trained using pdfs from a single slice of a particular subject generalizes well to other slices from the same subject, as well as to slices from another subject.

SMS imaging with Blipped-CAIPI acquisition and dictionary based Q-space compressed sensing was combined to provide a 12 fold ac-

celerated DSI acquisition. The data for this acquisition was collected from a healthy volunteer using on a novel 3T system equipped with the AS302 Connectom gradient with $G_{max}=300$ mT/m and Slew=200 T/m/s, whose strong gradients are particularly suited for DSI. The figure below illustrates: A. Blipped-CAIPI acquisition, B. q-space compressed sensing, C. Connectome gradients, and D. tractography result from a 12 fold accelerated 4 minute DSI acquisition.



As part of the development of these new acquisition techniques, new metrics to quantify the performance and characterize artifacts have also been developed to aid in understanding the speed vs. data quality tradeoff of these acquisitions. These new metrics include signal leakage measures, spurious noise correlation measure, and white matter tract bundles quantification.

REFERENCES

- [1] Larkman D. J., et al., “Use of multicoil arrays for separation of signal from multiple slices simultaneously excited,” *JMRI*, vol. 13, pp. 313–317, 2001.
- [2] Breuer, F. A., et al., “Controlled aliasing in parallel imaging results in higher acceleration (caipirinha) for multi-slice imaging,” *MRM*, vol. 53, pp. 684–691, 2005.
- [3] Nunes, R. G., et al., “Simultaneous slice excitation and reconstruction for single shot epi,” *ISMRM*, vol. 14, p. 293, 2006.
- [4] Moeller, R. G., et al., “Multiband multislice geepi at 7 tesla, with 16fold acceleration using partial parallel imaging with application to high spatial and temporal wholebrain fmri,” *MRM*, vol. 63, pp. 1144–1153, 2010.
- [5] Feinberg, D. A., et al., “Multiplexed echo planar imaging for sub-second whole brain fmri and fast diffusion imaging,” *PlosOne*, vol. 5, 2010.
- [6] Setsompop, K., et al., “Blippedcontrolled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced gfactor penalty,” *MRM*, vol. 67, pp. 1210–1224, 2012.
- [7] Menzel, M., et al., “Accelerated diffusion spectrum imaging in the human brain using compressed sensing,” *MRM*, vol. 56, pp. 1226–1233, 2011.
- [8] Bilgic, B., et al., “Accelerated diffusion spectrum imaging with compressed sensing using adaptive dictionaries,” *MRM*, vol. 68, pp. 1747–1754, 2012.

Dictionary Learning on Riemannian Manifolds and Applications to EAP Reconstruction

Yuchen Xie*, Jiaqi Sun*, Jeffery Ho*, and Baba C. Vemuri*

* Department of CISE, University of Florida, Gainesville, Florida 32611-6120.

Abstract—In this abstract, we present a general framework of dictionary learning for data living on Riemannian manifolds and apply it to the problem of EAP reconstruction from diffusion MRI.

I. INTRODUCTION

Dictionary learning, which seeks to find a collection of atoms for sparse representation of the input data, has been widely used in image recognition, classification and restoration (e.g., [1]). Under this model, each data point is assumed to be generated *linearly* using only a small number of atoms, and this linear sparsity assumption is responsible for much of its generalization power and success. However, the underlying linear process requires that the data points as well as the atoms be treated as vectors in some vector space R^d , and the dictionary is learned from the input data using only the vector space structure (and its associated inner product). For many applications in medical image analysis that involve *data points belonging to some known Riemannian manifolds* such as the space of symmetric positive-definite matrices, hyperspheres for square-root densities, Stiefel and Grassmann manifolds, etc., the existing extrinsic approaches that ignore the intrinsic structure implied by the data are clearly inadequate. To remedy this deficiency and inadequacy, we present a generalization to incorporate intrinsic geometry implied by the input data.

The applicability of existing dictionary learning methods to solve medical imaging problems that have to deal with manifold-valued data can pose two thorny issues. First, as a prerequisite, the data manifold must admit an embedding into some R^d in order to be able to apply the existing dictionary learning methods. However, for most manifolds, such as Grassmann and Stiefel manifolds, there simply does not exist known canonical embedding into R^d (or such embedding is difficult to compute). Second, even in the case when the existing method can be applied, due to their extrinsic viewpoint, important intrinsic properties of the data may not be represented in the dictionary. This can be illustrated by a simple example that it is possible that two points x, y on the manifold \mathcal{M} have a large geodesic distance separating them but under the embedding $i : \mathcal{M} \rightarrow R^d$, $i(x), i(y)$ has a small distance in R^d . Therefore, sparse coding using dictionary learned in R^d is likely to code $i(x), i(y)$ (and hence x, y) using the same set of atoms with similar coefficients. Clearly, this will be undesirable if the applications require tasks such as classification and clustering, for which one would prefer the sparse coding to reflect some degree of actual similarity (i.e., geodesic distance) between the two samples x, y .

While the above example provides the motivation for seeking an extension to the existing dictionary learning framework to the more general Riemannian setting, it is by no means obvious how the extension should be correctly formulated. Let \mathcal{M} denote the Riemannian manifold on which a collection of data points x_1, \dots, x_n are given. At the minimum, the goal of dictionary learning on \mathcal{M} is

to compute a collection of atoms $\{a_1, \dots, a_m\} \subset \mathcal{M}$, also points on \mathcal{M} , such that each data point x_i can be *generated* using only a small number of atoms (sparsity). One immediate technical hurdle that any satisfactory generalization (to the Euclidean setting) needs to overcome is the lack of a global linear structure that will allow the data to be generated from the atoms. Instead, the Riemannian geometry provides only local linear structures through the Riemannian exponential and logarithm [2] maps, and by moving to the more general Riemannian setting, we essentially trade the unique global linear structure with infinitely many local linear structures, which is the main source of the various technical difficulties present in our generalization. However, this diversity of linear structures also provides us with an opportunity to formulate the dictionary learning using *data specific* approach.

Specifically, the sparse coding of a data x_i with respect to the atoms $\{a_1, \dots, a_m\} \subset \mathcal{M}$ is obtained by minimizing

$$\min_{w_i} \left\| \sum_{j=1}^m w_{ij} \log_{x_i} a_j \right\|_{x_i}^2 + \mathbf{Sp}(w_i), \quad (1)$$

with the important affine constraint that $\sum_{j=1}^m w_{ij} = 1$, where $w_i = (w_{i1}, \dots, w_{im})^T$. That is, we are using the Riemannian exponential and logarithm maps at each data point x to define the generative process, and the sparse approximation of a given data point is first computed in its tangent space $T_x \mathcal{M}$ and then realized on \mathcal{M} by applying the exponential map. We remark that this formulation is entirely coordinate-independent since each $\log_{x_i} a_j$ is coordinate-independent, and Equation 1 can be minimized using any local chart and its associated basis for $T_{x_i} \mathcal{M}$ (with a result that will be independent of these choices). Furthermore, a subspace S in a given coordinate system is represented as an affine subspace in a different coordinate system with a different origin. In short, Equation 1 is the direct generalization of linear sparsity condition with the exception that now the origin has been moved to the data point x_i . Computationally, the resulting optimization problem using Equation 1 can be effectively minimized.

We validate the proposed method by applying it to the ensemble average propagator (EAP) reconstruction problem from HARDI data sets. Preliminary results demonstrate that the dictionary learned using the proposed method can provide real improvements when compared with other direct approaches.

REFERENCES

- [1] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [2] M. Spivak, *A comprehensive introduction to differential geometry*. Publish or perish Berkeley, 1979.

Acknowledgments

This research was in part funded by the NIH grant NS066340 to BCV.

Magnetic Resonance Imaging with Non-Linear Spatial Encoding Magnetic Fields

Maxim Zaitsev*

* Department of Radiology - Medical Physics, University Medical Centre Freiburg, Freiburg, Germany

Abstract—Traditional MRI favors homogeneous image intensity, resolution and SNR. More recent approaches offer faster image acquisition, but sacrifice intensity homogeneity and simple noise statistics. Using non-linear spatial encoding magnetic fields (SEMs) allows one to gain new degrees of freedom to further accelerate imaging with additional advantages in local spatial resolution and localized signal selection. In this contribution we review recent advances in MR imaging using non-linear encoding fields and indicate potential future directions and applications.

Traditional spatial selection and encoding MRI is bound to planes and simple box-like shapes, as rectangles and parallelepipeds. Image encoding and reconstruction is usually done on a Cartesian grid with a homogeneous spatial resolution, with the measurement time typically proportional to the number of lines in the image. Quite contrary the anatomy of living beings tends to be curvilinear, ranging from round or elliptic shapes, as human head or trunk, to almost fractal surfaces as for brain or kidney cortex.

In the early days of MRI as single-channel volume receiver coils were considered state-of-the-art and images were recovered by an application of the Fourier transform it was the homogeneity of the image intensity and contrast, which were considered to be of the primary value for image quality. Also further quality parameters as signal-to-noise ratio (SNR) or spatial resolution were trivial to characterize. Noise of MR images was spatially and spectrally homogeneous with a simple Gaussian statistics. Similarly, image resolution had a global character and could be characterized with a single point spread function. The only disadvantage was the imaging speed, as the Fourier encoding process is sequential by nature.

The classical period in MRI has come to its end in late 1990th with the advent of parallel imaging [1]. It was discovered that multiple localized receiver coils with spatially different sensitivities were able to provide additional information on signal localization and thus allow skipping certain steps of the sequential Fourier encoding. Image reconstruction now requires additional steps (e.g. matrix inversion), therefore image noise in parallel imaging is no longer homogeneous. In fact, image intensity delivered by a typical receiver coil array has a tendency to increase towards the periphery of the object, therefore image SNR in the majority of MR images acquired these days (e.g. with coil arrays and parallel imaging) is rather difficult to characterize. More recent imaging concepts based on randomized sampling and compressed sensing reconstruction [2] have even more complex behavior of image intensity, contrast, features and noise.

In an effort to trade homogeneous spatial resolution for focal advantages in localized imaging in terms of performance and resolution, we have introduced in 2007 a PatLoc-concept (=parallel imaging with localized gradients) [3]. Spatial localization in PatLoc is effected by means of local, non-linear, non-bijective spatial encoding magnetic fields (SEMs). Gradients of these fields are not necessary orthogonal to each other and the number of fields used may exceed the dimensionality of the encoding problem, e.g. 3 or 4 fields may be applied to encode a 2D image [4]. Non-unique encoding, which

appears as non-linear aliasing in image domain is resolved by the receiver array processing similar to the established parallel imaging methodology [5]. In contrast to the traditional parallel imaging, through changing the encoding field geometry the spatial aliasing pattern can potentially be optimized to match the receiver array unwrapping ability [6].

PatLoc concept holds a great promise for more efficient MR imaging, which has been realized by a number of research groups, who are now entering the field. Trajectory design in multiple dimensions [4], [6] and sampling strategies taking into account a higher number of SEMs than the actual dimensionality of the encoding problem [7] are the areas of the most active research.

Noteworthy that the PatLoc concept does not necessarily require the signal readout to occur under the action of non-linear gradients. Along these lines we have demonstrated a possibility to further accelerate parallel imaging by applying nonlinear phase modulation followed by a Cartesian k-space acquisition using linear gradients. Alternatively, non-linear phase pre-modulation may be employed for signal selection during the subsequent readout with linear gradients [8]. Furthermore, non-linear gradients offer interesting options for signal excitation in curved coordinates [9].

MR imaging with non-linear spatial encoding magnetic fields allows one to gain additional degrees of freedom to accelerate data acquisition and most fully exploit both the encoding capabilities of the available receiver coil arrays and the redundancies in the image, with additional advantages in local spatial resolution and localized signal selection. These degrees of freedom may eventually make it possible to adapt MRI encoding to the underlying anatomy for more efficient imaging of the structure and function in vivo.

REFERENCES

- [1] Pruessmann, K. P., et al., "Sense: sensitivity encoding for fast mri," *Magn Reson Med*, vol. 42, pp. 952–62, 1999.
- [2] Lustig, M., et al., "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magn Reson Med*, vol. 58, pp. 1182–95, 2007.
- [3] Hennig, J., et al., "Parallel imaging in non-bijective, curvilinear magnetic field gradients: a concept study," *Magn Reson Matter Phys*, vol. 21, pp. 5–14, 2008.
- [4] Gallichan, D., et al., "Simultaneously driven linear and nonlinear spatial encoding fields in mri," *Magn Reson Med*, vol. 65, pp. 702–14, 2011.
- [5] Schultz, G., et al., "Reconstruction of mri data encoded with arbitrarily shaped, curvilinear, nonbijective magnetic fields," *Magn Reson Med*, vol. 64, pp. 1390–403, 2010.
- [6] Tam, L.K., et al., "Null space imaging: Nonlinear magnetic encoding fields designed complementary to receiver coil sensitivities for improved acceleration in parallel imaging," *Magn Reson Med*, vol. 68, pp. 1166–75, 2012.
- [7] Lin, F.-H., "Multidimensionally encoded magnetic resonance imaging," *Magn Reson Med*, vol. [Epub ahead of print], 2012.
- [8] Witschey, W.R., et al., "Localization by nonlinear phase preparation and k-space trajectory design," *Magn Reson Med*, vol. 67, pp. 1620–32, 2012.
- [9] Weber, H., et al., "Excitation and geometrically matched local encoding of curved slices," *Magn Reson Med*, vol. [Epub ahead of print], 2012.

Waveform decoding and detection in hdEEG

Mauro Zucchelli * Paolo Manganotti[†] Gloria Menegaz*

* Dept. of Computer Science, University of Verona, Italy.

[†] Dept. of Neurological, Neuropsychological, Morphological and Movement Sciences, University of Verona, Italy

Abstract—Brain oscillations are very powerful descriptors of both physiological and pathological brain states. In general, EEG signals consist of complex mixtures of components whose characterization provides reliable information about the neuronal activity. This study is inspired to the *consensus matching pursuit* (CMP) representation and proposes an effective method for the detection and modeling of interictal prototypical signal patterns in temporal lobe epilepsy. CMP allows accounting for inter-trial variability in temporal jitter, frequency and number of oscillations. In this work, we propose to generalize the approach and exploit the resulting spike representation for automatic interictal spike detection. Performance was evaluated on both synthetic and real high density EEG signals. Results show high sensitivity and specificity in spike detection as well as an accurate separation in the transient and oscillation components.

I. INTRODUCTION

The purpose of this work is to model interictal waveforms (IW) in temporal lobe epilepsy for both characterization and detection. The target signals feature a prototypical shape consisting of a spike and a slow oscillation. Such two components hold different physiological meaning calling for effective methods for their respective detection and characterization. The spectral overlap between them makes the use of Fourier and classical multi resolution representations unsuitable to the purpose. Consensus matching pursuit (CMP) was recently introduced by Benar et al.[1], [2]. While good results were obtained on simulated signals, a clear advantage was not proved on real signals. However, only single channel recordings on one patient who was diagnosed a temporal lobe epilepsy was considered, calling for a more complete characterization of the method. In this work, the CMP approach was generalized by introducing an additional parameter for the synthesis of the dictionary waveforms, namely the phase of the atoms, leading to what we call *Phase Sensitive-CMP* (PS-CMP). Additionally, the resulting representation was exploited for automatic detection based on support vector machines (SVM) enabling the accurate signal decoding into the transient and slow oscillation components as well as the effective detection of the spikes on the whole set of channels.

II. METHODS

MP is a signal approximation method where a signal is represented by linear combination of atoms chosen in a dictionary $D = \{\Psi_{\vec{p}}\}$ according to an iterative projection procedure. At each iteration, the algorithm selects the element of the dictionary, i.e. the *atom* $\Psi_{\vec{p}_i}(t)$ that minimizes the residual error, which becomes the target signal for the next iteration. In this way the atom that is selected at step i only depends on the current realization k . CMP overcomes this limitation by making such a choice a function of both the realization and the distribution of the parameters \vec{p}_i over the whole set of realization (e.g. the training set). This is accomplished by estimating the probability density function of the underlying stochastic process in a non-parametric manner, following the weighted voting procedure as in [1].

In the proposed method, the parameters were first estimated on the dataset and then used for building the dictionary. This allows shaping the dictionary to the case study without losing its generalization

capabilities. Furthermore, CMP was reinterpreted in the framework of pattern recognition. Firstly, the prototypical waveforms (PWs) are constructed relying on PS-CMP. This allows the training of the classifier, which learns the properties of the target signals. Then, the PWs are used for guiding the recognition process in the PS-CMP framework. The model provides the description of the spike and the oscillation in terms of the proximity to the *consensus atoms* that were identified during the first step in two successive iterations of the MP algorithm. The pipeline was applied to both simulated and real hdEEG data obtained from six patients affected by right temporal lobe epilepsy. Since only one category of signals was targeted and modeled, the one class SVM classifier was used for the detection. The two features chosen for the classification were the magnitude of the first projection coefficient and the goodness of fit (GOF) of the reconstruction.

III. RESULTS AND DISCUSSION

Due to space limitations here we focus on hdEEG data of one patient. The training set consisted of 28 spikes that had been manually selected by the neurophysiologist within one channel on the right temporal lobe. Two different testing sets were considered, respectively consisting of 60 trials extracted from a single channel (out of which 30 representing an IW), and 256 trials taken from all the hdEEG channels over a 500 ms time interval. Out of these, a subset of 60 trials were selected and manually labelled by an expert neurologist to be used as the ground truth. In the first case, all the IWs were correctly identified by the classifier. In the second, very good performance was reached in terms of specificity ($spec = 1$), selectivity ($sel = 1$) and sensitivity ($sens = 0.88$). Figure 1 shows the topology of the IW distribution on the scalp, highlighting the involvement of the right temporal lobe as expected. Overall, the method provided good performance in both IW modeling and detection on both synthetic and real data.

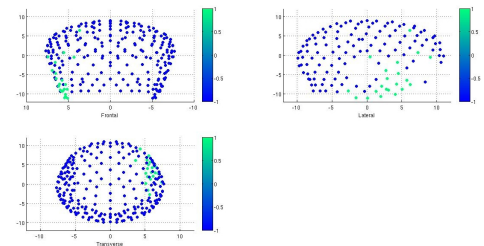


Fig. 1. Topology of IWs classification. Blue: detected IW.

REFERENCES

- [1] C. Benar, T. Papadopoulou, B. Torresani, and M. Clerc, “Consensus matching pursuit for multi-trial eeg signals,” *J. of Neuroscience Methods.*, vol. 180, pp. 161–170, 2009.
- [2] N. Jmail, M. Gavaret, F. Wendling, and C. Benar, “A comparison of methods for the separation of transient and oscillatory signals in eeg,” *J. of Neuroscience Methods.*, vol. 199, pp. 273–289, 2011.