BASP Frontiers Workshop 2015

Chair
Prof. Yves Wiaux, Heriot-Watt University, UK

Area Chairs
Prof. Yves Wiaux, Heriot-Watt University, UK [signal processing]
Prof. Jason McEwen, University College London, UK [astro-imaging]
Prof. Michael Lustig, UC Berkeley, USA [bio-imaging]

Session Organisers

Signal processing
Prof. Pier Luigi Dragotti, Imperial College London, UK
Prof. David Dunson, Duke University, USA
Prof. Peter Richtarik, University of Edinburgh, UK

Astro-imaging
Prof. Alan Heavens, Imperial College London, UK
Prof. Thomas Kitching, University College London, UK
Prof. Hiranya Peiris, University College London, UK
Prof. Oleg Smirnov, Rhodes University, ZA

Bio-imaging
Prof. Guang-Hong Chen, University of Wisconsin, USA
Prof. Chunlei Liu, Duke University, USA
Prof. Nicole Seiberlich, Case Western Reserve University, USA
Prof. Shreyas Vasanawala, Stanford University, USA

Local Organizing Committee
Dr. Rafael Carrillo, EPFL, Switzerland
Mrs. Rosie De Pietro, EPFL, Switzerland
Mr. Sean Farrell, Heriot-Watt University, UK
Mr. Vijay Kartik, EPFL, Switzerland
Dr. Boris Leistedt, University College London, UK
Dr. Gilles Puy, INRIA, France
Mrs. Lynn Smith, Heriot-Watt University, UK
Prof. Jean-Philippe Thiran, EPFL, Switzerland
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 the complexity of data to be acquired and processed is constantly increasing, thus challenging signal processing theories. For example, data come in larger volumes every day, can be multi-modal, multi-spectral, scalar or tensor-valued, living in high dimensional geometries, and are possibly non-Euclidean.

The international Biomedical and Astronomical Signal Processing (BASP) Frontiers workshop was created to promote synergies between selected topics in astronomy and biomedical sciences, around common challenges for signal processing.

Building on the success of the first two workshops in 2011 and 2013, the BASP Frontiers 2015 workshop will open its floor to many interesting hot topics in theoretical, astrophysical, and biomedical signal processing, with a particular focus on imaging.

Following our tradition, BASP Frontiers 2015 will take place in a very nice resort in the Swiss Alps named Villars-sur-Ollon, close to Lausanne and Lake Geneva. All participants will be accommodated in 4 star hotel in a full board regime. We believe that the most fruitful discussions often take place after the sessions themselves, on the terrace, or during breakfast, lunch, or dinner. We hope that the winter atmosphere will further promote discussion and creativity.

The Workshop Chairs
Contents

General Interest

Anthony Lasenby, The Search for Gravitational Waves in the Early Universe ......... 6

Signal Processing

Amir Beck, On the Convergence of Alternating Minimization with Applications to Iteratively Reweighted Squares and Decomposition Schemes .......................... 7
Carola-Bibiane Schoenlieb, Luca Calatroni, Juan Carlos De Los Reyes, Tuomo Valkonen, Optimising the optimisers - image reconstruction by bilevel optimisation .......... 8
Mike Davies, Gilles Puy, Pierre Vandergheynst, Yves Wiaux, Compressed Quantitative MRI using BLIP ......................................................... 9
Pier Luigi Dragotti, A ProSparse Approach to find the Sparse Representation in Fourier and Canonical Bases ..................................................... 10
Michael Elad, Yaniv Romano, SOS Boosting of Image Denoising Algorithms ........ 11
Anna Gilbert, Sparse Approximation, List Decoding, and Uncertainty Principles ...... 12
Sonja Greven, Clara Happ, Volker Schmid, A Fully Bayesian Hierarchical Framework for Scalar-on-Image Regression ........................................ 13
Vassilis Kalofolias, Xavier Bresson, Michael Bronstein, Pierre Vandergheynst, Enhanced matrix completion with manifold learning ......................... 14
Jakub Konecny, Peter Richtarik, Semi-Stochastic Gradient Descent Methods ........ 15
Jonathan Masci, Alex Bronstein, Michael Bronstein, Pablo Sprechmann, Guillermo Sapiro, Sparse similarity-preserving hashing ................................... 16
Julien Mairal, Incremental and Stochastic Majorization-Minimization Algorithms ...... 17
Jeffrey Morris, Functional Regression Methods for Biomedical Imaging Data ........ 18
Joao Mota, Nikos Deligiannis, Miguel Rodrigues, Compressed Sensing with Prior Information: Theory and Practice ........................................ 19
Jon Oñativia, Pier Luigi Dragotti, Sparsity According to Prony: From Structured to Unstructured Representations and Back ................................. 20
Frank Ong, Michael Lustig, Beyond Low Rank + Sparse: A Multi-scale Low Rank Decomposition ................................................................. 21
Hanjie Pan, Thierry Blu, Martin Vetterli, Annihilation-driven Image Edge Localization .... 22
Aurélie Pirayre, Camille Couprie, Laurent Duval, Jean-Christophe Pesquet, Discrete vs Continuous Optimization for Gene Regulatory Network Inference .......... 23
Chiara Ravazzi, Enrico Magli, Fast IRLS for sparse reconstruction based on gaussian mixtures .......................... 24
Audrey Repetti, Emilie Chouzenoux, Jean-Christophe Pesquet, Proximal Primal-Dual Optimization Methods ............................................... 25
Philip Schniter, Sundeep Rangan, Alyson Fletcher, Statistical Image Recovery: A Message-Passing Perspective ........................................ 26
Quoc Tran-Dinh, Volkan Cevher, A Primal-Dual Algorithmic Framework for Constrained Convex Optimization .............................................. 27
Susan Wei, Asymptotic Inference for Integral Curves of Noisy Vector Fields ..... 28
Dawn Woodward, Small-Feature Model-Based Image Segmentation .................. 29
Mehrdad Yaghoobi, Di Wu, Mike Davies, Non-Negative Orthogonal Matching Pursuit .... 30

Astro-imaging

Ewan Cameron, On functional regression-based emulators for faster Bayesian inference from computational simulations ...................................... 31
Rafael Carrillo, Jason McEwen, Yves Wiaux, Why CLEAN when you can PURIFY? A new approach for next-generation radio-interferometric imaging ................ 32
Arwa Dabbech, Chiara Ferrari, David Mary, Eric Slezak, Oleg Smirnov, Jonathan Kenyon, MORESANE: a sparse deconvolution algorithm for radio interferometric imaging 33

3
Hugh Garsden, Julien Girard, Jean-Luc Starck, Stéphane Corbel, Cyril Tasse, Arnaud Woiselle, John McKean, LOFAR and SKA Sparse Image Reconstruction ........................ 34
Ian Harrison, Prina Patel, Challenges in Radio Weak Gravitational Lensing .......................... 35
Alan Heavens, Raul Jimenez, Ben Panter, Astrostatistics and Brain Imaging ......................... 36
Michael Holson, Farhan Feroz, Philip Graff, Anthony Lasenby, Neural networks and accelerated Bayesian inference ................................................................. 37
Paul Hurley, Orhan Öcal, Gridding by Beamforming ............................................................ 38
Benjamin Hugo, James Gain, Oleg Smirnov, Accelerated facet-based widefield imaging ........ 39
Henrik Junklewitz, RESOLVE: A new algorithm for aperture synthesis imaging of extended emission .......................................................... 40
Vijay Kartik, Rafael Carrillo, Yves Wiaux, Dimension embedding for big data in radio interferometry ...................................................................................... 41
Jonathan Kenyon, Oleg Smirnov, Arwa Dabbech, Chiara Ferrari, PyMORESANE: Pythonic and CUDA-accelerated implementations of MORESANE ........ 42
Thomas Kitching, Weak Gravitational Lensing .......................................................... 43
Henrik Junklewitz, RESOLVE: A new algorithm for aperture synthesis imaging of extended emission .......................................................... 40
Vijay Kartik, Rafael Carrillo, Yves Wiaux, Dimension embedding for big data in radio interferometry ...................................................................................... 41
Jonathan Kenyon, Oleg Smirnov, Arwa Dabbech, Chiara Ferrari, PyMORESANE: Pythonic and CUDA-accelerated implementations of MORESANE ........ 42
Thomas Kitching, Weak Gravitational Lensing .......................................................... 43
Martin Kunz, Bruce Bassett, Michelle Lochner, Nadeem Oozeer, Iniyani Natarajan, Jonathan Zwart, Oleg Smirnov, Bayesian Inference for Radio Observations .......... 44
Boris Leistedt, Jason McEwen, Martin Büttnner, Pierre Vandalheydst, Hiranya Peiris, Yves Wiaux, Analysing the polarisation of the CMB with spin scale-discretised wavelets .......... 46
Michelle Lochner, Iniyani Natarajan, Jonathan Zwart, Oleg Smirnov, Bruce Bassett, Martin Kunz, Nadeem Oozeer, Bayesian Inference for Radio Observations: Source Separation . 47
André Offringa, Radio interferometric imaging for the SKA and its pathfinders ................. 49
Dovi Poznanski, Studies the Milky Way via stacks of low S/N spectra ................................. 50
Jean-François Roberthille, Multiscale analysis of Galactic dust emission .............................. 51
Bio-imaging ........................................................................................................... 57
Peter Basser, Opportunities and Challenges in Brain Mapping with Diffusion MRI ............. 57
Noam Ben-Eliezer, Daniel Sodickson, Timothy Shepherd, Graham Wiggins, Kai Tobias Block, Non-Analytic Model-Based Reconstruction for Accelerated Multiparametric Mapping in MRI .......................................................... 58
Stephen Cauley, Kawin Setsompop, Elfar Adalsteinsson, Mark Griswold, Lawrence Wald, Hierarchically Semiseparable Generalized Encoding Matrix Compression for Fast Inverse Imaging .................................................................................. 59
Guang-Hong Chen, More is indeed different ........................................................................ 60
Sebastien Equis, Yann Cotte, 4D in Bio-microscopy: Marker-free Live Cell Tomography .... 61
Pejman Ghanouni, MR Guided High Intensity Focused Ultrasound ...................................... 62
Charles Hennemeyer, Coming Challenges in Medical Device Development; Alternative, Counter-Current Ideas in Disease Physiology May Provide Ideas for Medical-Device-Based Solutions that will Run Countercurrent to Existing, Drug Based Therapies .......... 63
Martin Holler, Kristian Bredies, Karl Kunisch, Matthias Schögl, Rudolf Stollberger, ICTGV regularization for image sequence reconstruction of accelerated dynamic MRI ........ 64
Gregory Kicska, *Digital Chest Tomosynthesis: Clinical Applications and Technical Obstacles*. 65
Kelvin Layton, Stefan Kroboth, Feng Jia, Sebastian Littin, Huijun Yu, Maxim Zaitsev, *Spatial encoding with generalised magnetic field shapes* 66
Zhi-Pei Liang, *Multidimensional Imaging: A Path to High Resolution and High Speed through Subspaces* 67
Kathryn Nightingale, *Quantitative Elasticity Imaging With Acoustic Radiation Force: Methods and Clinical Applications* 68
Ricardo Otazo, *Low-rank plus sparse dynamic MRI: separation of background and dynamic components and self discovery of motion* 69
Jürgen Reichenbach, *MR Susceptibility Imaging and Mapping* 70
Nicole Seiberlich, *Magnetic Resonance Fingerprinting: Beyond Parameter Mapping to Clinical Application* 71
Kawin Setsompop, *Wave-CAIPI for an order of magnitude acceleration in MRI acquisition* 72
Noam Shemesh, Gonzalo Alvarez, Lucio Frydman, *Cellular microstructures revealed by Non-Uniform Oscillating-Gradient Spin-Echo (NOGSE) MRI* 73
Chantal Tax, Tom Dela Haije, Andrea Fuster, Remco Duits, Max Viergever, Evan Calabrese, G. Allan Johnson, Luc Florack, Alexander Leemans, *Towards Quantification of the Brain’s Sheet Structure in Diffusion MRI Data* 74
Shreyas Vasanawala, *Quantification in Cardiovascular MRI* 75
Ge Wang, *Imaging with X-ray Modulated Nanoparticles* 76
Jong Chul Ye, *Semi-Analytic Iterative Framework for TV Penalized Cone-beam CT Reconstruction* 77
Abstract—Considerable excitement was caused in March 2014 by the announcement of a detection by the BICEP2 experiment of gravitational waves in the early universe via their effect on the Cosmic Microwave Background (CMB). These gravitational waves imprint themselves into a particular mode of polarisation of the CMB, and measurement of their amplitude would finally reveal the energy scale at which inflation took place, as well as providing direct evidence that it actually occurred. It would also represent the farthest back in time we could ever look, and the large amplitude as discovered provides a point of contact with string cosmology and other theories of the early universe, and has stimulated much theoretical work following the announcement of the discovery. This talk will look at the background to the experimental results, the theoretical implications of the range of possible amplitudes and also the latest information from the Planck Satellite, which as well as measuring the CMB itself, provides important information on the possible foreground contamination in the results.
On the Convergence of Alternating Minimization with Applications to Iteratively Reweighted Least Squares and Decomposition Schemes

Amir Beck*

* Faculty of Industrial Engineering and Management, Technion - Israel Institute of Technology, Haifa, Israel.

Abstract — This work is concerned with the alternating minimization (AM) method for solving convex minimization problems where the decision variables vector is split into two blocks. The objective function is a sum of a differentiable convex function and a separable (possibly) nonsmooth extended-real-valued convex function, and consequently constraints can be incorporated. We analyze the convergence rate of the method and establish a nonsmooth sublinear rate of convergence where the multiplicative constant depends on the minimal block Lipschitz constant. We then analyze the iteratively reweighted least squares (IRLS) method for solving convex problems involving sums of norms. Based on the results derived for the AM method, we establish a nonsmooth sublinear rate of convergence of the IRLS method. In addition, we show an asymptotic rate of convergence whose efficiency estimate does not depend on the data of the problem. Finally, we study the convergence properties of a decomposition-based approach designed to solve a composite convex model.

In this work we consider the following minimization problem:

$$\min_{y \in \mathbb{R}^n, z \in \mathbb{R}^m} \{H(y, z) \equiv f(y, z) + g_1(y) + g_2(z)\},$$

where \( f, g_1, g_2 \) are assumed to satisfy the following two properties:

[A] The functions \( g_1 : \mathbb{R}^n \to (-\infty, \infty] \) and \( g_2 : \mathbb{R}^m \to (-\infty, \infty] \) are closed and proper convex functions.

[B] The function \( f \) is a continuously differentiable convex function over \( \text{dom } g_1 \times \text{dom } g_2 \).

[C] \( \nabla_y f \) is (uniformly) Lipschitz continuous with respect to the variables vector \( y \) over \( \text{dom } g_1 \) with constant \( L_1 \in (0, \infty) \).

[D] \( \nabla_z f \) is (uniformly) Lipschitz continuous with respect to the variables vector \( y \) over \( \text{dom } g_2 \) with constant \( L_2 \in (0, \infty] \).

Obviously, when \( L_2 = \infty \), assumption [D] is meaningless, and in this case only the gradient with respect to \( y \) is Lipschitz continuous.

We will use the convention that the variables vector \( x \in \mathbb{R}^{n+m} \) is composed of the vectors \( y \) and \( z \) as follows:

$$x = (y, z).$$

The alternating minimization (AM) is described explicitly below.

The main convergence result that we establish is a sublinear rate of convergence of the sequence of function values stating that for any \( n \geq 2 \):

$$H(x_n) - H^* \leq \max \left\{ \left( \frac{1}{2} \right)^{n+1} \left( H(x_0) - H^* \right), \frac{8 \min\{L_1, L_2\} R^2}{n - 1} \right\}$$

where \( H^* \) is the optimal value of problem \((P)\) and

$$R \equiv \max_{x \in \mathbb{R}^{n+m}} \max_{z \in \mathbb{R}^m} \{\|x - x^\ast\| : H(x) \leq H(x_0)\}.$$

In the context of the IRLS method, we consider the following general model:

$$\min_{y \in X} \{s(y) + \sum_{i=1}^m \|A_i y + b_i\|^2 \} \quad \text{s.t.} \quad y \in X.$$ 

This is a general model encompassing several important applications such as robust regression, the Fermat-Weber problem and \( l_1 \)-regularized least squares. The IRLS method is described below:

**Input:** \( \eta > 0 \) - a given parameter.

**Initialization:** \( y_0 \in X \).

**General Step \((k=0,1,\ldots)\):**

$$y_{k+1} \in \arg\min_{y \in X} \left\{s(y) + \frac{1}{2} \sum_{i=1}^m \frac{\|A_i y + b_i\|^2}{\eta} \right\}$$

The IRLS method is actually equivalent to the AM method employed on the problem

$$\min_{y \in X} \left\{s(y) + \frac{1}{2} \sum_{i=1}^m \frac{\|A_i y + b_i\|^2 + \eta^2}{\eta} \right\} : y \in X, z \geq \frac{\eta}{2}.$$ 

Using the rate of convergence result for the AM method, we establish a nonsmooth sublinear rate of convergence result for the IRLS problem. We also establish an asymptotic rate of convergence result that only depends on the constant \( R \).

The final model that we consider is the composite model

$$T^* = \min \{T(y) \equiv q(y) + r(Ay)\},$$

where \( q : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\} \) is a closed, proper convex extended real-valued convex function, \( r : \mathbb{R}^m \to \mathbb{R} \) is a real-valued convex function which is Lipschitz with constant \( L_r \) and \( A \) is an \( m \times n \) matrix. We analyze the behaviour of the AM method employed on the following approximated problem

$$T^*_n = \min_{y,z} \{T_n(y,z) = q(y) + r(z) + \frac{\rho}{2} \|z - Ay\|^2 \},$$

and establish a rate of convergence of both the approximated and exact problems is established.
Optimising the optimisers - image reconstruction by bilevel optimisation

Luca Calatroni*, Juan Carlos De Los Reyes†, Carola-Bibiane Schönlieb* and Tuomo Valkonen†

* Department of Applied Mathematics and Theoretical Physics (DAMTP), University of Cambridge, Cambridge CB3 0WA, UK.
† Centro de Modelización Matemática (MODEMAT), Escuela Politécnica Nacional, 170109 Quito, Ecuador.

Abstract—This talk is about optimal parameter learning for inverse imaging problems using a bilevel optimisation strategy.

I. INTRODUCTION

When assigned with the task of reconstructing an image from given data the first challenge one faces is the derivation of a truthful image and data model. Such a model can be determined by the a-priori knowledge about the image, the data and their relation to each other. The source of this knowledge is either our understanding of the type of images we want to reconstruct and of the physics behind the acquisition of the data or we can thrive to learn parametric models from the data itself. The common question arises: how can we optimise our model choice? Starting from the first modelling strategy this talk will lead us from the total variation as the most successful image regularisation model today to non-smooth second- and third-order regularisers, with data models for Gaussian and Poisson distributed data as well as impulse noise. Applications for image denoising, inpainting and surface reconstruction are given.

After a critical discussion of these different image and data models we will turn towards the second modelling strategy and propose to combine it with the first one using a bilevel optimisation method. In particular, we will consider optimal parameter derivation for total variation denoising with multiple noise distributions [1], [2] and optimising total generalised variation regularisation [3] for its application in photography. This is joint work with Luca Calatroni (Cambridge), Juan Carlos De Los Reyes (Quito), Tuomo Valkonen (Quito).

REFERENCES

Compressed Quantitative MRI using BLIP

Mike Davies∗, Gilles Puy†, Pierre Vandergheynst‡ and Yves Wiaux§
∗Institute for Digital Communications (IDCom), The University of Edinburgh, EH9 3JL, UK
†INRIA, Centre Ioria Rennes - Bretagne Atlantique, FR-35042 Rennes Cedex, France
‡Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, CH
§Institute of Sensors, Signals, and Systems, Heriot-Watt University, EH14 4AS, UK

Abstract—We present a compressed sensing framework for quantitative MRI based on Magnetic Resonance Fingerprinting. We show that, as long as the excitation sequence induces persistent excitation, we are able to achieve accurate recovery of the proton density, T1, T2 and off-resonance maps simultaneously from very short pulse sequences.

I. INTRODUCTION
A new type of MRI acquisition scheme called Magnetic Resonance Fingerprinting (MRF) [3] offers full quantification of multiple tissue properties simultaneously through a single acquisition process. The procedure is composed of 4 steps: (1) the material magnetization is excited through a sequence of random RF pulses; (2) each pulse response is recorded over a small portion of k-space; (3) a sequence of highly aliased magnetization response images are formed using back projection; and (4) parameter maps (proton density, ρ, T1, T2 and off-resonance, δf) are formed using a bank of matched filters, applied voxelwise.

We investigate this idea from a compressed sensing (CS) perspective and leverage recent results from [1] and develop a recovery algorithm with good theoretical guarantees.

II. THE BLOCH RESPONSE MANIFOLD
The MRF process is based upon a random pulse excitation sequence. Denote the magnetization response image sequence by X ∈ C^{N×L}, with X_{i,t} denoting the magnetization for voxel i at the t^{th} readout. The magnetization response at any voxel can be written as a parametric nonlinear mapping from \{ρ, θ\} to the sequence X_{i,:} as:

\[ X_{i,:} = \rho_i B(θ_i; α, TR) ∈ C^{1×L}, \]

where L is the sequence length, θ_i = \{T_1, T_2, δf\} ∈ M is the set of unknown parameters and B : M → C^{1×L} is a smooth mapping induced by the Bloch dynamics.

Inferring \{ρ, θ\} from an estimate for X_{i,:} (assuming identifiability) can be done by projecting onto a discretization of the cone of the response manifold, which we denote as \( R, B \).

Let \( θ^{(k)} = (T^{(k)}_1, T^{(k)}_2, δf^{(k)}) \) be a discrete sampling of M and define the MRF “dictionary” \( D ∈ C^{P×N} \) of the magnetization responses as: \( D_k = B(θ^{(k)}; α, TR), k = 1, \ldots, P \). The projection is given by the maximum matched filter of the voxel response sequence with the elements of D. After which the Bloch parameters can be retrieved using a look up table.

III. K-SPACE SAMPLING
Unfortunately, it is impractical to observe the full spatial magnetization X_{i,:} at each readout within the necessary time window and we must resort to some form of undersampling in k-space, which we denote by the mapping: \( Y = h(X) \).

In order to ensure parameter map recovery, we now exploit tools from compressed sensing. In particular we would like h to induce a low distortion embedding of the cone of the product response manifold, \((R, B)^N\), or equivalently satisfy a suitable RIP. In order to achieve this it is useful to characterize the persistent discrimination of the different magnetization responses. We quantify this persistence through the flatness of the chords of \( R, B \).

Definition 1: Let U be a collection of vectors \{u\} in \( C^L \). We denote the flatness of these vectors by

\[ \lambda := \max_{u \in U} ||u||_∞ / ||u||_2. \]

Note that from standard norm inequalities \( L^{-1/2} ≤ λ ≤ 1 \).

For our sampling operator we consider regularly subsampling k-space by a factor of p in one direction with random shifts at each readout time. This can be achieved using a randomized version of multishot Echo-Planar Imaging (EPI). The following theorem shows that random EPI can provide the desired RIP.

Theorem 1 (RIP for random EPI): Given an excitation response cone \( R, B \) of dimension \( d_b \), whose chords have a flatness \( λ \), and a random EPI operator \( h : (R, B)^N → C^{M×L} \). With probability at least \( 1 − η \), h is a restricted isometry on \((R, B)^N − (R, B)^N\) with constant \( δ \), as long as

\[ λ^{-2} ≥ Cδ^{-2}p^2d_b\log(N/δη), \]

for some constant \( C \) independent of \( p, N, d_b, δ \) and \( η \).


IV. COMPRRESSED QUANTITATIVE MRI
Assuming that \( Y = h(X) \) has a suitable RIP we can retrieve \{ρ, θ\} from Y using an efficient iterated projection algorithm [1] along with our discretized Bloch response model.

\[ X^{(n+1)} = P_{(R, B)^N}[X^{(n)} + μh^H(Y − h(X^{(n)}))], \]

where n is the recursion index, \( P_{(R, B)^N} \) is the projection onto the signal model \((R, B)^N\) approximated using D, and \( μ \) is a stepsize, which we select adaptively. We call the resulting algorithm BLIP (Bloch response recovery via Iterated Projection).

V. SIMULATIONS
In simulations on an anatomical brain phantom the BLIP procedure was able to achieve near oracle performance with a pulse sequence length of ~ 200, substantially shorter than the already impressive MRF performance. For full details see [2].

REFERENCES
A ProSparse Approach to find the Sparse Representation in Fourier and Canonical Bases

Pier Luigi Dragotti,
Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ London, UK.

Abstract—We revisit the classical problem of finding the sparse representation of a signal in a pair of bases. When both bases are orthogonal, it is known that there is a gap between the unicity condition and the one required to use the polynomial-complexity Basis Pursuit (BP) algorithm.

In this paper, we focus on the case of Fourier and canonical bases and introduce a new polynomial complexity algorithm, ProSparse, that finds the K-sparse representation of a signal under only the unicity constraint.

I. INTRODUCTION

We consider the classical problem of finding the sparse representation of a signal in the pair of Fourier and canonical bases. Specifically, let \( y \) be an \( N \)-dimensional vector given by the linear combination of \( K \) atoms of the dictionary \( D = [F_N \ I_N] \), where \( F_N \) denotes the Fourier basis and the identity matrix \( I_N \) denotes the canonical basis. Given the synthesis model \( y = Dx \), we study the problem of finding the \( K \) nonzero entries of \( x \) from \( y \).

The exact sparse representation problem can be stated as follows:

\[
(P_0) : \quad \min_{x} \|x\|_0 \quad \text{s.t.} \quad y = Dx,
\]

where \( \|x\|_0 \) is the \( \ell_0 \) “norm” which computes the number of nonzero entries of \( x \). Since \( (P_0) \) is not convex, it is convenient to consider the following convex relaxation:

\[
(P_1) : \quad \min_{x} \|x\|_1 \quad \text{s.t.} \quad y = Dx,
\]

where \( \|x\|_1 = \sum_{i=1}^{M} |x_i| \) is the \( \ell_1 \) norm.

The sparse approximation problem was first posed in these forms by Donoho and Huo in [1] and then extended to generic orthogonal pairs of bases by Elad and Bruckstein in [2]. The remarkable finding of these papers is that \( (P_0) \) has a unique solution when

\[
K < \frac{1}{\mu} = \sqrt{N}, \tag{1}
\]

where \( \mu \) is the mutual coherence of the dictionary \( D \) and, in the case of Fourier and canonical bases, \( \mu = 1/\sqrt{N} \). Moreover, if the signal \( y \) is made of \( K_F \) atoms of \( F_N \) and \( K_I \) atoms of \( I_N \) with \( K = K_F + K_I \), then \( (P_1) \) is equivalent to \( (P_0) \) when

\[
2\mu^2 K_F K_I \mu + \mu K_I - 1 < 0. \tag{2}
\]

This implies that, when the signal is sparse enough, the solution to the original nonconvex problem \( (P_0) \) can be obtained by addressing the simpler convex problem \( (P_1) \). Since \( (P_1) \) is equivalent to Linear Programming, it can be solved by using polynomial complexity algorithms such as Basis Pursuit (BP).

We note that \( (P_1) \) poses a weaker condition than \( (P_0) \) and that there is a gap between these two bounds. While we know that \( (P_0) \) has a unique solution under (1), we cannot conclude from previous results whether \( (P_0) \) can be solved with polynomial complexity algorithms unless the sparsity level is further reduced to satisfy (2). The main contribution of this work is to show that for this specific structured dictionary (with Fourier and canonical bases), a polynomial complexity algorithm based around Prony’s method can solve \( (P_0) \) under the only constraint (1). Note that Prony’s method is commonly used in spectral estimation theory [4] to estimate the frequencies of a finite sum of exponentials from its uniform samples.

For this reason we call this algorithm ProSparse in honour of Baron de Prony who first invented the method that goes under his name. The basic insight behind ProSparse is that we observe \( y = Dx \) and that if the signal is made only of Fourier atoms then \( y \) is the sum of \( K \) exponentials. In particular, its \( n \)-th entry is of the form \( y_n = \sum_{k=0}^{K-1} \alpha_k u_k^n \), where \( u_k = e^{2\pi i n k/N} \) and \( t_k \) is the integer corresponding to the location of the \( k \)-th non-zero entry of \( x \). Moreover, \( \alpha_k \) is the amplitude of that non-zero entry and the sparse representation can be found using Prony’s method directly, which requires only 2\( K \) consecutive entries of \( y \) to retrieve \( x \). In reality \( y \) contains both Fourier atoms and spikes. However, spikes are localised therefore there is a good chance to find a clean interval of \( y \) which is only due to the Fourier atoms.

Let us assume that \( y \) is made of \( 0 < K_F < K \) Fourier atoms and \( K_S = K - K_F \) spikes. We need to find an interval of 2\( K_F \) consecutive entries of \( y \) which are only due to the Fourier atoms, in order to be able to apply Prony’s on these entries and retrieve the \( K_F \) Fourier atoms. Once this has been achieved, retrieving the spike is straightforward. It is possible to show that such an interval exist when

\[
K < \sqrt{2N}. \tag{3}
\]

This new constraint is clearly weaker than (1) and thus guarantees that an interval with 2\( K_F \) entries containing only Fourier atoms always exists.

ProSparse operates by finding this window, estimating the Fourier contribution using Prony’s method, removing the contribution due to the Fourier atoms from \( y \) and then retrieving the spikes from the residual. Since Prony’s method has polynomial complexity, it is possible to show that ProSparse has also polynomial complexity. We refer to [5] for more details.

REFERENCES

SOS Boosting of Image Denoising Algorithms

Yaniv Romano∗ and Michael Elad†

∗ Electrical Engineering Department, Israel Institute of Technology, Haifa, Israel 32000
† Computer Science Department, Israel Institute of Technology, Haifa, Israel 32000

Abstract—In this work we propose a generic recursive algorithm for improving image denoising methods. Given the initial denoised image, we suggest repeating the following “SOS” procedure: (i) Strengthen the signal by adding the previous denoised image to the degraded input image, (ii) Operate the denoising method on the strengthened image, and (iii) Subtract the previous denoised image from the restored signal-strengthened outcome. The convergence of this process is studied for the K-SVD image denoising and related algorithms. Furthermore, still in the context of K-SVD image denoising, we introduce an interesting interpretation of the SOS algorithm as a technique for closing the gap between the local patch-modeling and the global restoration task, thereby leading to improved performance. We demonstrate the SOS boosting algorithm for several leading denoising methods (K-SVD, NLM, BM3D, and EPLL), showing tendency to further improve denoising performance.

I. INTRODUCTION

Image denoising is a fundamental restoration problem, where the given measurement image $y \in \mathbb{R}^{m \times n}$ is obtained from the clean signal $x \in \mathbb{R}^{m \times n}$ by a contamination of the form

$$y = x + v,$$  

(1)

where $v \in \mathbb{R}^{m \times n}$ is an iid zero-mean additive noise. Our goal is an approximation of the unknown clean image $x$. Plenty of sophisticated algorithms have been developed in order to estimate the original image, e.g., NLM, K-SVD, BM3D, LSSC [1], [2], [3], [4], and others. These algorithms rely on powerful image priors, where sparse representations and processing of local patches have become two prominent ingredients.

Despite the effectiveness of the above denoising algorithms, improved results can be obtained by applying a boosting technique. There are several such techniques that were proposed over the years, e.g., “Twicing” [5], Bregman iterations [6], SAIF [7], diffusion [8], and more.

In this work we propose a generic recursive function that treats the denoising method as a “black-box” and has the ability to push it forward to improve its performance. Differently from the above methods, instead of adding the residual (which mostly contains noise) back to the noisy image, or filtering the previous estimate over and over again (which could lead to over-smoothing), we suggest strengthening the signal by leveraging on the availability of the denoised image. More specifically, given an initial estimation of the cleaned image, improved results can be achieved by repeating iteratively the following SOS procedure:

1) Strengthen the signal by adding the previous denoised image to the noisy input image,
2) Operate the denoising method on the strengthened image, and
3) Subtract the previous denoised image from the restored signal-strengthened outcome.

The core equation that describes this procedure can be written in the following form:

$$\tilde{x}^{k+1} = f(y + \tilde{x}^k) - \tilde{x}^k,$$  

(2)

where $\tilde{x}^0 = 0$. We show that a performance improvement is achieved for various algorithms (NLM, K-SVD, BM3D, and EPLL) by up to 1dB, as the signal-strengthened image can be denoised more effectively compared to the noisy input image, due to the improved Signal to Noise Ratio (SNR).

The convergence of the proposed algorithm is studied in this paper by formulating the linear part of the denoising method and assessing the iterative systems matrix properties. In our work we put special emphasis on the K-SVD and describe the resulting denoising matrix and the corresponding convergence properties related to it. The work in [7] shows that most existing denoising algorithms can be approximated as a row-stochastic positive definite matrices. In this context, our analysis suggests that for most denoising algorithms, our proposed boosting method is guaranteed to converge.

In addition, we introduce an interesting interpretation of the SOS boosting algorithm, related to a major shortcoming of patch-based methods: the gap between the local patch processing and the global need for a whole restored image. In general, patch-based methods aggregate overlapping cleaned patches by averaging. This often leads to low-frequency artifacts in the final outcome image. Defining the difference between the local (intermediate) denoised patch and the corresponding patch from the global outcome as a disagreement, we show that, in the context of the K-SVD image denoising, the SOS algorithm is equivalent to repeating the following steps: (i) compute the disagreement per patch, (ii) subtract the disagreement from the degraded input patches, (iii) apply the restoration algorithm on these patches, and (iv) reconstruct the image. Therefore, the proposed algorithm encourages the overlapping patches to share their local information, thus reducing the gap between the local patch-processing and the global restoration task.

REFERENCES

Sparse Approximation, List Decoding, and Uncertainty Principles

Anna C. Gilbert*

* Department of Mathematics, The University of Michigan, Ann Arbor, Michigan 48109–1043.

Abstract—We consider list versions of sparse approximation problems, where unlike the existing results in sparse approximation that consider situations with unique solutions, we are interested in multiple solutions. We introduce these problems and present the first combinatorial results on the output list size. These generalize and enhance some of the existing results on threshold phenomenon and uncertainty principles in sparse approximations. Our definitions and results are inspired by similar results in list decoding. We also present lower bound examples that bolster our results and show they are of the appropriate size.
A Fully Bayesian Hierarchical Framework for Scalar-on-Image Regression

Sonja Greven*, Clara Happ*, Volker Schmid*
*Department of Statistics, Ludwig-Maximilians-Universität München, Munich, Germany

Abstract—We propose a method for scalar-on-image regression for spatially structured image covariates, as typically arise in neuroimaging studies. The model contains two main components: a Gaussian Markov random field imposes spatial smoothness of coefficient images, whereas a latent binary field accounts for the typical structure of the images, marked e.g. by hard transitions from grey matter to white matter. Samples from the posterior distribution are generated using a single-site Gibbs sampler, allowing a fast and efficient implementation. The approach is applied to the ADNI neuroimaging study for modeling the relationship between neuroimaging data and a neuropsychological test score for patients with mild cognitive impairment (MCI) and healthy controls.

I. INTRODUCTION

Alzheimer’s disease (AD) is the most common age-related form of dementia, making up more than 65% of all dementia cases. Structural and functional brain abnormalities occur already in subjects with mild cognitive impairment (MCI), who show an increased risk of AD dementia. Neuroimaging methods are promising clinically relevant tools to predict the cognitive decline and the development of dementia at an early stage of AD. The characteristic structure of neuroimages requires appropriate statistical models tailored to this kind of data.

II. A SCALAR-ON-IMAGE REGRESSION MODEL

We propose a framework for regression models of the form

$$Y = z^T \alpha + \sum_{l=1}^{L} \beta_l x_l + \varepsilon, \quad \varepsilon \overset{i.i.d.}{\sim} N(0, \sigma^2)$$

(1)

with a spatially structured image covariate $X$ with associated coefficient image $\beta$, a vector of scalar variables $z$ with coefficients $\alpha$ and a scalar response $Y$. A fully Bayesian hierarchical approach is natural here in order to account for the spatial nature of the images and the possibility to include expert prior knowledge e.g. in the form of regions of interest (ROI). The model consists of two main components: An intrinsic Gaussian Markov random field, that imposes spatial smoothness in the coefficient image, and a latent binary field $\gamma$ with

$$\beta_l = 0 \text{ if } \gamma_l = 0 \text{ and } \beta_l \neq 0 \text{ if } \gamma_l = 1, \quad l = 1, \ldots, L,$$

(2)

that accounts for the typical structure of the images, marked e.g. by hard transitions from grey matter to white matter. Its second purpose is variable selection to identify regions in the image covariate that have a significant relationship with the scalar response. In the literature, there have been different suggestions for the choice of the prior of this latent field. Smith and Fahrmeir [1] and Goldsmith et al. [2] propose a random field Ising model

$$p(\gamma | a, b) \propto \exp \left( a \sum_{l=1}^{L} \gamma_l + b \sum_{l=1}^{L} \sum_{j \in \delta(l)} I(\gamma_l = \gamma_j) \right),$$

(3)

which inherently makes use of the spatial neighbourhood structure $\delta(l)$. A drawback of this approach is that computing the normalizing constant of the Ising prior is numerically infeasible, making the update of the hyperparameters $a$ and $b$ a challenging task. An alternative suggested by Kalus et al. [3] in the setting of fMRI data is to use a latent Probit model

$$\gamma_l | \eta_l \sim \text{Bin}(1, \Phi(\eta_l)), \quad l = 1, \ldots, L$$

(4)

and include the spatial information in the linear predictor $\eta_l$, e.g. via predefined ROIs.

We extend the model of Goldsmith et al. [2] to a fully Bayesian model by an auxiliary variable approach [4] and adapt the latent Probit model by Kalus et al. [3] to the situation in (1) in order to include both in a general framework and compare their performance in a simulation study. In both cases, samples from the posterior distribution are generated using a single-site Gibbs sampler, allowing a fast and efficient implementation.

III. THE ADNI STUDY DATA

We apply our approach to the Alzheimers Disease Neuroimaging Initiative (ADNI, [5]) study including patients with mild cognitive impairment MCI (in many cases an early stage of Alzheimer’s disease) and healthy controls. Having available structural (MRI) as well as functional (FDG-PET) neuroimaging data, our model can incorporate this complementary information to model the relationship between the image covariates and a neurocognitive test score (ADAS 11) in an early stage of Alzheimer’s disease.

IV. OUTLOOK

A major future objective lies in extending the model to longitudinal image data and appropriate modeling of the cognitive decay in Alzheimer’s disease.

REFERENCES

Enhanced Matrix Completion with Manifold Learning

Vassilis Kalofolias∗, Xavier Bresson ∗, Michael Bronstein † and Pierre Vandergheynst∗

∗ Signal Processing Laboratory LTS2, Ecole Polytechnique Fédérale de Lausanne (EPFL)
† Institute of Computational Science, Faculty of Informatics SI-109, Università della Svizzera Italiana

Abstract—We study the problem of matrix completion when information about row or column proximities is available, in the form of weighted graphs. The problem can be formulated as the optimization of a convex function that can be solved efficiently using the alternating direction multipliers method. Experiments show that our model offers better reconstruction than the standard method that only uses a low rank assumption, especially when few observations are available.

I. INTRODUCTION

In matrix completion we have a set of signals in matrix form \( M \in \mathbb{R}^{m \times n} \). They are only sparsely observed, i.e. a small set \( \Omega \) of the elements of \( M \) are known, and the goal is to recover the rest. In order for the problem to be solvable, additional assumptions are needed, the most common one being that the matrix is low rank. The problem of finding the matrix of minimum rank that agrees to the given observed values is NP-hard. However, replacing \( \text{rank}(X) \) with its complex surrogate nuclear norm \( \|X\|_* \) yields the convex problem

\[
\min_{X \in \mathbb{R}^{m \times n}} \|X\|_*, \quad \text{s.t.} \quad A_\Omega(X) = A_\Omega(M),
\]

where \( A_\Omega \) denotes the operator that keeps only the observed values of a matrix. Under the assumption that \( M \) is sufficiently incoherent, if the indices \( \Omega \) are uniformly distributed and \( |\Omega| \) is sufficiently large, solving the last problem gives an exact solution [1].

II. MATRIX COMPLETION ON GRAPHS

Low rank implies the linear dependence of rows/columns. However, this dependence is unstructured. In many situations, the rows/columns of matrix \( M \) possess additional structure that can be incorporated into the completion problem in the form of regularization. In this work, we assume that the row and column signals reside on manifolds approximated by two given graphs. The weight \( w_{ij} \) of an edge of the columns’ graph \( G_c = (V_c, E_c, W_c) \) represents how close are the two adjacent columns \( x_i \) and \( x_j \). Similarly for the rows, that are nodes of the corresponding graph \( G_r = (V_r, E_r, W_r) \). More formally, we want

\[
\sum_{j \neq j'} w_{j,j'}^c \|x_j - x_{j'}\|^2 = \text{tr}(XL_cX^T) = \|X\|_{D,c}
\]

where \( L_c = D_c - W_c \) is the Laplacian of the column graph \( G_c \), \( D_c = \text{Diag}(\sum_{j'=1}^n w_{jj'}) \), and \( \| \cdot \|_{D,c} \) is the graph Dirichlet semi-norm for columns. Similarly, for the rows we get a corresponding expression \( \text{tr}(X^T L_r X) = \|X\|_{D,r} \) with the Laplacian \( L_r \) of the row graph \( G_r \).

These smoothness terms are added to the matrix completion problem as regularization terms, and the final optimization problem we solve is [3]

\[
\min_X \gamma_n \|X\|_* + \frac{1}{2} \|A_\Omega(X - M)\|_F^2 + \gamma_r \|X\|_{D,r} + \frac{\gamma_c}{2} \|X\|_{D,c},
\]

where \( \gamma_n, \gamma_r, \gamma_c \) are parameters that can be chosen with model selection methods like cross validation.

References

Semi-Stochastic Gradient Descent Methods

Jakub Konečný, Peter Richtárik
School of Mathematics, University of Edinburgh, United Kingdom.

Abstract—In this paper we study the problem of minimizing the average of a large number \( n \) of smooth convex loss functions. We propose a new method, S2GD (Semi-Stochastic Gradient Descent), which runs for one or several epochs in each of which a single full gradient and a random number of stochastic gradients is computed, following a geometric law. The total work needed for the method to output an \( \varepsilon \)-accurate solution, measured in the number of passes over data, or equivalently, in units equivalent to the computation of a single gradient of the loss, is \( O([n/\kappa] \log(1/\varepsilon)) \), where \( \kappa \) is the condition number. This is achieved by running the method for \( O((1/\varepsilon)) \) epochs, with a single gradient evaluation and \( O(\kappa) \) stochastic gradient evaluations in each. The SVRG method of Johnson and Zhang [1] arises as a special case.

I. INTRODUCTION

Many problems in data science (e.g., machine learning, optimization and statistics) can be cast as loss minimization problems of the form

\[
\min_{x \in \mathbb{R}^d} f(x), \tag{1}
\]

where

\[
f(x) \overset{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} f_i(x). \tag{2}
\]

Here \( d \) typically denotes the number of features / coordinates, \( n \) the number of examples, and \( f_i(x) \) is the loss incurred on example \( i \). That is, we are seeking to find a predictor \( x \in \mathbb{R}^d \) minimizing the average loss \( f(x) \). In big data applications, \( n \) is typically very large; in particular, \( n \gg d \).

Let us now briefly review two basic approaches to solving problem (1).

Gradient Descent. Given \( x_k \in \mathbb{R}^d \), the gradient descent (GD) method sets

\[
x_{k+1} = x_k - h f'(x_k),
\]

where \( h \) is a stepsize parameter and \( f'(x_k) \) is the gradient of \( f \) at \( x_k \). We will refer to \( f'(x_k) \) by the name full gradient. In order to compute \( f'(x_k) \), we need to compute the gradients of \( n \) functions. Since \( n \) is big, it is prohibitive to do this at every iteration.

Stochastic Gradient Descent (SGD). Unlike gradient descent, stochastic gradient descent [2], [3] instead picks a random \( i \) (uniformly) and updates

\[
x_{k+1} = x_k - h f'_i(x_k).
\]

Note that this strategy drastically reduces the amount of work that needs to be done in each iteration (by the factor of \( n \)). Since

\[
\mathbb{E}[f'_i(x_k)] = f'(x_k),
\]

we have an unbiased estimator of the full gradient. Hence, the gradients of the component functions \( f_1, \ldots, f_n \) will be referred to as stochastic gradients. A practical issue with SGD is that consecutive stochastic gradients may vary a lot or even point in opposite directions. This slows down the performance of SGD. On balance, however, SGD is preferable to GD in applications where low accuracy solutions are sufficient. In such cases usually only a small number of passes through the data (i.e., work equivalent to a small number of full gradient evaluations) are needed to find an acceptable \( x \). For this reason, SGD is extremely popular in fields such as machine learning.

In order to improve upon GD, one needs to reduce the cost of computing a gradient. In order to improve upon SGD, one has to reduce the variance of the stochastic gradients. In this paper we propose and analyze a Semi-Stochastic Gradient Descent (S2GD) method [4]. Our method combines GD and SGD steps and reaps the benefits of both algorithms: it inherits the stability and speed of GD and at the same time retains the work-efficiency of SGD.

In this work, we assume that each \( f_i \) is \( L \)-smooth, and that whole function \( f \) is \( \mu \)-strongly convex. Let \( \kappa = L/\mu \) be the condition number.

II. RESULTS

We prove that the S2GD algorithm attains linear convergence rate in expectation, and needs \( O((n + \kappa) \log(1/\varepsilon)) \) evaluations of \( \nabla f_i(x) \) to achieve \( \varepsilon \)-accuracy. This is a strong contrast with Gradient Descent, which would need \( O((n\kappa) \log(1/\varepsilon)) \) gradient evaluations.

We extend the algorithm into noisy setting, where we cannot evaluate \( \nabla f_i(x) \) exactly, but are only given an estimate. We show that the algorithm converges linearly to a neighborhood of the optimal solution, dependent on the level of inexactness.

We further prove that the algorithm can achieve super-linear speedup by utilizing parallel processors, and the possibility of running it in distributed setting.

Efficient implementation has been made available at [5].

REFERENCES

Sparse similarity-preserving hashing

Jonathan Masci∗, Alex M. Bronstein†, Michael M. Bronstein∗, Pablo Sprechmann‡ and Guillermo Sapiro‡

∗ University of Lugano, Switzerland; † Tel Aviv University, Israel; ‡ Duke University, USA.

Abstract—In recent years, a lot of attention has been devoted to efficient nearest neighbor search by means of similarity-preserving hashing. One of the plights of existing hashing techniques is the intrinsic trade-off between performance and computational complexity: while longer hash codes allow for lower false positive rates, it is very difficult to increase the embedding dimensionality without incurring in very high false negatives rates or prohibiting computational costs. We propose a way to overcome this limitation by enforcing the hash codes to be sparse. Sparse high-dimensional codes enjoy from the low false positive rates typical of long hashes, while keeping the false negative rates similar to those of a shorter dense hashing scheme with equal number of degrees of freedom.

I. SPARSEHASH

Given the data set \( X \in \mathbb{R}^n \) similarity-preserving hashing aims at representing it to the space \( \mathbb{R}^m = \{\pm 1\}^m \) of \( m \)-dimensional binary vectors with the Hamming metric \( d_{Ham}(a, b) = \frac{1}{2} - \frac{1}{2} \sum_{i=1}^{m} a_i b_i \) by means of an embedding \( \xi : X \rightarrow \mathbb{R}^m \) that preserves the original similarity relation. Masci et al. [1] realized \( \xi \) as a neural network and achieved state-of-the-art performance for single- and multi-modal retrieval. Similarly, in SparseHash we employ an \( \ell_1 \)-norm regularization, extensively used in the compressed sensing literature to promote sparsity, and minimize the average of

\[
L(\xi(x), \xi(x')) = s_X(x, x')\|\xi(x) - \xi(x')\|_1 + \frac{\lambda}{2}(1 - s_X(x, x'))\max(0, M - \|\xi(x) - \xi(x')\|_1)^2 + \alpha(\|\xi(x)\|_1 + \|\xi(x')\|_1),
\]

over the training set, where \( s_X \) is the groundtruth similarity function (1: similar, 0: dissimilar), \( \alpha \) controls the level of sparsity, \( \lambda \) governs the false positive and negative rate tradeoff, and \( M \) is a margin.

![Fig. 1. Left: Five nearest neighbors retrieved by different hashing methods. Right: P/R versus retrieval time for various Hamming radii.](image_url)

In modern retrieval applications, an object is often represented by more than one data modality; i.e. \( X \) and \( Y \). SparseHash generalizes to the multi-modal case by constructing two networks, \( \xi \) and \( \eta \), and minimizing the coupling loss

\[
L_{MM} = \mu_1 L(\xi(x), \xi(x'), s_X(x, x')) + \mu_2 L(\eta(y), \eta(y'), s_Y(y, y')) + L(\xi(x), \eta(y), s_{XY}(x, y)).
\]

\( \mu_1 \) and \( \mu_2 \) control the relative importance of the intra-modality similarity (0 in the cross-modal regime).

We implement SparseHash by coupling two ISTA-type networks, sharing the same set of parameters as the ones described in [2]. The architecture of an ISTA-type network can also be seen as a recurrent network with a soft threshold activation function.

II. EXPERIMENTAL RESULTS

We compare to several state-of-the-art hashing methods [3], [4], [5], [1]. All dense hash methods achieve an average sparsity of about 50% per sample whereas SparseHash achieves much sparser and structured codes; i.e. 6% sparsity on CIFAR10 with hash length of 128. This results in a drastic reduction of unique codes in the database which leads to higher recall for collision retrieval.

**CIFAR10** [6]: 60K labeled images belonging to 10 different classes represented with 384-dimensional GIST descriptors. Retrieval performance is shown in Fig. 2 whereas Fig. 1-right compares timing vs precision-recall.

![Fig. 2. Left: P/R curves at various Hamming radii: \( r = r_0 = 48 \) (dash), \( r = 2 \) (dot), \( r = 0 \) (line). Right: Recall versus Hamming radius for 48 (line) and 128 (dotted) hash lengths.](image_url)

**NUS** [7]: 270K images from Flickr associated with one or more of the different 81 concepts. We use a 500-dim bag-of-features representation for the image modality and a 1K-dim Tags vector for the text modality. Fig. 1-left shows image retrieval and Fig. 3 multi-modal retrieval and image annotation examples.

![Fig. 3. Left: Image annotation (in green matching Tags in the groundtruth). Right: Retrieved images for the given Tags query on the left.](image_url)

REFERENCES


Abstract—Majorization-minimization algorithms consist of iteratively minimizing a majorizing surrogate of an objective function. Because of its simplicity and its wide applicability, this principle has been very popular in statistics and in signal processing. In this work, we intend to make this principle scalable. We introduce an incremental and stochastic majorization-minimization schemes which are able to deal with large scale data sets. We compute convergence rates of our methods for convex and strongly-convex problems, and show convergence to stationary points for non-convex ones. We develop several efficient algorithms based on our framework. First, we propose new incremental and stochastic proximal gradient methods, which experimentally matches state-of-the-art solvers for large-scale $l_1$ and $l_2$ logistic regression. Second, we develop an online DC programming algorithm for non-convex sparse estimation.

The principle of successively minimizing upper bounds of the objective function is often called majorization-minimization [1]. Each upper bound is locally tight at the current estimate, and each minimization step decreases the value of the objective function. Even though this principle does not provide any theoretical guarantee about the quality of the returned solution, it has been very popular and widely used because of its simplicity. Many existing approaches can indeed be interpreted from the majorization-minimization point of view. For instance, this is the case of gradient-based or proximal methods, expectation-maximization (EM) algorithms in statistics, difference-of-convex (DC) programming, boosting. Majorizing surrogates have also been used successfully in the signal processing literature about sparse optimization, linear inverse problems in image processing, and matrix factorization.

In this work, we are interested in making the majorization-minimization principle scalable for minimizing a large sum of functions:

$$\min_{\theta \in \Theta} \left[ \frac{1}{T} \sum_{t=1}^{T} f_t(\theta) \right],$$

where the functions $f_t : \mathbb{R}^p \rightarrow \mathbb{R}$ are continuous, and $\Theta$ is a convex subset of $\mathbb{R}^p$. When $f$ is non-convex, exactly solving (1) is intractable in general, and when $f$ is also non-smooth, finding a stationary point of (1) can be difficult. The problem above when $T$ is large can be motivated by machine learning applications, where $\theta$ represents some model parameters and each function $f_t$ measures the adequacy of the parameters $\theta$ to an observed data point indexed by $t$. In this context, minimizing $f$ amounts to finding parameters $\theta$ that explain well some observed data. In the last few years, stochastic optimization techniques have become very popular in machine learning for their empirical ability to deal with a large number $T$ of training points. Even though these methods have inherent sublinear convergence rates for convex and strongly convex problems, they typically have a cheap computational cost per iteration, enabling them to efficiently find an approximate solution. Recently, incremental algorithms have also been proposed for minimizing finite sums of functions [2], [3]. At the price of a higher memory cost than stochastic algorithms, these incremental methods enjoy faster convergence rates, while also having a cheap per-iteration computational cost.

We follow this line of work: in order to exploit the particular structure of problem (1), we propose an incremental scheme whose cost per iteration is independent of $T$, as soon as the upper bounds of the objective are appropriately chosen. We call the resulting scheme “MISO” (Minimization by Incremental Surrogate Optimization). We present convergence results when the upper bounds are chosen among the class of “first-order surrogate functions”, which approximate the objective function up to a smooth error—that is, differentiable with a Lipschitz gradient. For non-convex problems, we obtain almost sure convergence and asymptotic stationary point guarantees. In addition, when assuming the surrogates to be strongly convex, we provide convergence rates for the expected value of the objective function. Remarkably, the convergence rate of MISO is linear for minimizing strongly convex composite objective functions, a property shared with two other incremental algorithms for smooth and composite convex optimization: the stochastic average gradient method (SAG) of Schmidt, Le Roux and Bach [2], and the stochastic dual coordinate ascent method (SDCA) of Shalev-Schwartz and Zhang [3].

In a second step, we assume that $f$ has the form of an expectation:

$$\min_{\theta \in \Theta} \mathbb{E}_x \left[ f(\theta) \right] := \mathbb{E}_x \left[ f(x, \theta) \right],$$

where $x$ from some set $\mathcal{X}$ represents a data point, which is drawn according to some unknown distribution, and $\ell$ is a continuous loss function. We propose a scheme consisting of iteratively building a surrogate of the expected cost when only a single data point is observed at each iteration; this data point is used to update the surrogate, which in turn is minimized to obtain a new estimate.

We study the convergence properties of our algorithm when the surrogates are strongly convex and chosen among the class of first-order surrogate functions. When the objective is convex, we obtain expected convergence rates that are asymptotically optimal, or close to optimal. More precisely, the convergence rate is of order $O(1/\sqrt{T})$ in a finite horizon setting, and $O(1/n)$ for a strongly convex objective in an infinite horizon setting. Our second analysis shows that for non-convex problems, our method almost surely converges to a set of stationary points under suitable assumptions.

REFERENCES


Abstract—In this talk, I will describe functional regression methods that can be applied to large-scale biomedical image data, and illustrate their application to several different biomedical imaging data sets.

In biomedical research, numerous assays produce complex, high dimensional image data. These data are characterized by complex structure both within and between images that should be taken into account for efficient analysis and accurate inference. Many commonly used approaches for these settings either model pixels/voxels independently, ignoring this structure, or limit analyses to simple summaries from the images that do not use all of the information in the data. Functional data analysis (FDA) methods can serve as a middle ground between these two extremes, modeling the entire image while accounting for its internal structure through basis function modeling and regularization approaches.

A common analysis goal in image analysis is regression to either identify which image regions differ across groups or to classify subjects based on their image data. As reviewed by Morris (2015), the FDA literature contains a large number of functional regression approaches that in principle can be applied to these tasks, but unfortunately most existing methods are not suitable for the complex, high-dimensional data characterizing biomedical imaging applications. Much of the development has focused on simple, smooth, independently sampled functions on 1d Euclidean domains, while biomedical image data are typically complex, irregular, high-dimensional data sampled on higher-dimensional and/or non-Euclidean domains, and frequently with sampling designs that induce spatial or longitudinal correlation across the images.

This talk will describe Bayesian functional regression approaches we have developed that are equipped to handle the complexity and size of biomedical imaging data sets. Our approach is based on the functional mixed model, a flexible framework that can model the simultaneous effects of multiple predictors and through random effect functions capture correlation between the images induced by the design. The method can be used with a variety of different basis functions and regularization/smoothing approaches, which can be chosen to best capture the internal structure of the images. The computational approach involves a full MCMC, but scales up to very large data sets and yields rigorous Bayesian inference to perform global testing as well as flag significant regions of the images while adjusting for multiple testing.

We will describe these methods along with recent adaptations to handle spatially and longitudinally correlated image data, and apply them to several applications, including event-related potentials from a smoking cessation study, spatially correlated genomics data from a study of bladder cancer natural history, and ocular image data from a study of the etiology of glaucoma.

REFERENCES

Compressed Sensing with Prior Information: Theory and Practice
João F. C. Mota, Nikos Deligiannis, and Miguel R. D. Rodrigues
Electronic & Electrical Engineering Department, University College London, UK

Abstract—We study two schemes for reconstructing Compressed Sensing (CS) signals in the presence of prior information. Prior information here means a signal similar to the signal we want to reconstruct. One scheme, which we call \(\ell_1-\ell_1\) minimization, adds the \(\ell_1\)-norm of the difference between the optimization variable and the prior information to the objective of basis pursuit. The other scheme, called \(\ell_1-\ell_2\) minimization, considers the \(\ell_2\)-norm instead. We establish bounds on the number of measurements that these schemes require for successful signal reconstruction with high probability. Our bounds reveal that, for a reasonably accurate prior information, \(\ell_1-\ell_1\) minimization performs much better than both \(\ell_1-\ell_2\) minimization and standard CS. We also consider representative applications.

I. COMPRESSED SENSING WITH PRIOR INFORMATION

Suppose that we want to reconstruct a sparse signal \(x^* \in \mathbb{R}^n\) from a set of \(m\) measurements \(y = Ax^*\), where \(A \in \mathbb{R}^{m \times n}\) is known. Suppose, in addition, that we know a signal \(w \in \mathbb{R}^n\) similar to \(x^*\), which we call prior information. We study the performance of the following schemes for reconstructing \(x^*\):

\[
\begin{align*}
\text{minimize} \quad & \|x\|_1 + \beta \|x - w\|_1 \\
\text{subject to} \quad & Ax = y
\end{align*}
\]

(1)

\[
\begin{align*}
\text{minimize} \quad & \|x\|_1 + \frac{\beta}{2} \|x - w\|_2^2 \\
\text{subject to} \quad & Ax = y
\end{align*}
\]

(2)

where \(\beta \geq 0\), and \(\|x\|_1 := \sum_{i=1}^{n} |x_i|\) and \(\|x\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2}\) are the \(\ell_1\)-norm and the \(\ell_2\)-norm of \(x\), respectively. We refer to (1) as \(\ell_1-\ell_1\) minimization and to (2) as \(\ell_1-\ell_2\) minimization. Setting \(\beta = 0\) in (1) or (2) yields the so-called basis pursuit problem, used in Compressed Sensing (CS) [1], [2] to reconstruct \(x^*\) from the compressive measurements \(y\). Problems (1) and (2) are natural generalizations of basis pursuit when prior information is available. It is known that if each entry of \(A \in \mathbb{R}^{m \times n}\) is i.i.d. Gaussian and \(m \geq 2\log(n/s) + (7/5)s\), where \(s\) is the number of nonzero components of \(x^*\), then basis pursuit recovers \(x^*\) with high probability [3]. Our bounds, which are a function both the prior information “quality” and the weight \(\beta\), establish that [4]:

- For a reasonable prior information \(w\), \(\ell_1-\ell_1\) minimization performs much better than both standard CS and \(\ell_1-\ell_2\) minimization. This is confirmed experimentally both with synthetic and real data.
- The best \(\beta\) for \(\ell_1-\ell_1\) minimization is independent of any problem parameter and is always equal to 1. The best \(\beta\) for \(\ell_1-\ell_2\) minimization, in contrast, depends on several parameters, including the unknown \(x^*\).

The following theorem provides the bound for (1) with \(\beta = 1\):

**Theorem 1** ([4]). Let \(H := \{i : x_{i}^* > 0, x_{i} > w_i\}\) and \(\xi := \{i : w_i \neq x_{i}^* = 0\}\). Assume \(H > 0\) and \(i : x_{i}^* = w_i = 0\) \(\neq 0\), and that each entry of \(A \in \mathbb{R}^{m \times n}\) is i.i.d. Gaussian. If

\[
m \geq 2n \log \left(\frac{n}{s + \xi/2}\right) + \frac{7}{5} \left(s + \frac{\xi}{2}\right) + 1,
\]

(3)

then, with high probability, \(x^*\) is the unique solution of (1) for \(\beta = 1\).

II. EXPERIMENTS

Fig. 1 illustrates the performance of classical CS (basis pursuit) and problems (1) and (2) using synthetic data. It shows the success rate of each problem over 50 different realizations of \(A\) versus the number of measurements \(m\). Here, \(n = 1000\), \(s = 70\), and \(w\) yielded \(H = 11\) and \(\xi = -42\); see [4] for details. Note that \(w\) differed significantly from \(x^*: \|w - x^*\|_2/\|x^*\|_2 \approx 0.45\). \(\ell_1-\ell_1\) minimization required much less measurements than both \(\ell_1-\ell_2\) minimization and classical CS for perfect reconstruction. Theoretical bounds are also shown.

Fig. 2 shows the reconstruction of an endoscopy image with standard CS (Fig. 2(c)), and with \(\ell_1-\ell_1\) minimization (Fig. 2(d)) using Fig. 2(b), from the previous video frame, as prior information. All images are \(128 \times 128\) and we used \(m = 9000\) measurements (55%). Since \(A\) was the product of a partial Fourier with a Wavelet matrix, our bounds are not applicable. Yet, \(\ell_1-\ell_1\) minimization still outperforms standard CS: the PSNR of the reconstructed images was 34.4 dB for standard CS and 36.0 dB for \(\ell_1-\ell_1\) minimization.

REFERENCES

Abstract—The problem of finding the sparse representation of a signal has attracted a lot of attention over the last decade due to recent advancements that have reshaped the way data is processed. Recently, the compressed sensing (CS) community has established uniqueness conditions and reconstruction algorithms by relaxing a non-convex optimisation problem [1, 2].

Finite rate of innovation (FRI) theory [3] is an alternative approach that solves the sparsity problem using algebraic methods based on Prony’s algorithm. Recent extensions to this framework [4] have shown that it is possible to recover sparse representations beyond the uniqueness limits, that is, finding all the possible sparse representations that fit the observation for the case of signals represented in a pair of bases. Here, we show the application of such methods to the extended cases of unions of frames and average performances of such algorithm.

I. INTRODUCTION

Let \( y \in \mathbb{C}^N \) be a complex finite dimensional signal that has a \( K \)-sparse representation in an overcomplete dictionary, that is,

\[
y = D x,
\]

(1)

where \( D \in \mathbb{C}^{N \times L} \) is the overcomplete dictionary with \( L > N \) atoms and \( x = (x_0, \ldots, x_{L-1}) \in \mathbb{C}^L \) satisfies \( \|x\|_0 \triangleq \# \{ \ell : |x_\ell| \neq 0 \} = K \). One way of recovering the sparse vector \( x \) from the observation \( y \) is by solving the following problem:

\[
(P_1): \quad \arg\min_{\bar{x}} \|\bar{x}\|_0 \quad \text{s.t.} \quad y = D \bar{x},
\]

(2)

which is intractable since the \( \ell_0 \) “norm” is nonconvex. In practice, this problem can be solved via a convex relaxation, which replaces the \( \ell_0 \)-norm by the \( \ell_1 \)-norm. This technique is known as Basis Pursuit. Under some conditions, the \( \ell_1 \) approach is guaranteed to find the sparsest solution.

When the dictionary is built from the union of a Fourier and a local basis the atoms can be recovered applying a variation of Prony’s method [4]. This new algorithm is called ProSparse.

II. SPARSE RECONSTRUCTION BASED ON PRONY’S METHOD

Let \( D = [\Phi, \Psi] \), where \( \Phi \) is a Fourier matrix and \( \Psi \) is the identity. Vector \( x \) can be expressed in terms of the elements that correspond to each basis as follows: \( x = [x_\Phi, x_\Psi]^T \). It follows that \( y = \Phi x_\Phi + \Psi x_\Psi \). We assume that vector \( y \) is composed of \( K_f \) Fourier atoms and \( K_s \) spikes, that is, \( x_\Phi \) is \( K_f \)-sparse and \( x_\Psi \) is \( K_s \)-sparse. The overall sparsity of the vector \( x \) is \( K = K_f + K_s \).

If we sequentially analyse sections of consecutive samples of the vector \( y \), there will be windows that are not corrupted by spikes and are only influenced by the Fourier atoms (subject to \( K_f \) not being too large). That is, for a clean window, samples \( y_n \) are given by

\[
y_n = \sum_{k=1}^{K_f} a_k e^{j2\pi m_k n/N},
\]

(3)

where \( m_k \) are the indices of the non-zero elements in \( x_\Phi \). If the clean window contains at least \( 2K_f \) samples, parameters \( a_k \) and \( m_k \) can be perfectly recovered from samples \( y_n \) by applying Prony’s method. The spikes locations and amplitudes are then obtained from the residual \( y - \Phi x_\Phi \). Prony’s method is based on finding the coefficients \( h_n \), that annihilate the sequence \( y_n \), that is,

\[
y_n + h_1 y_{n+K_f} + \ldots + h_K y_{n+K-1} = 0.
\]

(4)

Coefficients \( h_n \) form a polynomial and its roots are directly related to the locations \( m_k \).

This approach is extensible to a variety of scenarios. For instance, it can also be applied to the case where \( \Phi \) is a DCT matrix, since the atoms can be expressed as the sum of two complex exponentials, and \( \Psi \) is a local basis, such as a Haar matrix with one decomposition level. It can also be applied to unions of frames, where for example \( \Phi \) is given by a fat Fourier matrix.

III. EXPERIMENTAL RESULTS

Figure 1 shows simulation results for the case of the overcomplete dictionary constructed from a Fourier frame and the identity. The Figure compares the probability of success in recovering the sparse vector \( x \) from vector \( y \) using ProSparse and Basis Pursuit algorithms. The Fourier frame is obtained from the first \( N \) rows of an \( M \times M \) Fourier matrix. For each pair \( (K_f, K_s) \), 100 realisations of sparse vectors with \( K_f \) spikes and \( K_f \) Fourier atoms. The locations of the non-zero elements are uniformly distributed at random and the amplitudes are drawn from \( \mathcal{N}(0, 1) \). From the plots, it can be seen that there are regions where ProSparse outperforms the traditional \( \ell_1 \) optimisation technique.

REFERENCES


Beyond Low Rank + Sparse:
A Multi-scale Low Rank Decomposition

Frank Ong, Michael Lustig
Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA.

Abstract—We present a multi-scale version of low rank matrix decomposition. Our motivation comes from imaging applications, in which image sequences are correlated locally on several scales in space and time rather than globally. We model our data matrix as a sum of matrices, where each matrix has increasing scales of locally low-rank matrices. Using this multi-scale modeling, we can capture different scales of correlation in our data matrix and provide a more compact representation than conventional low rank methods. We solve the proposed decomposition via a convex formulation.

INTRODUCTION

In many image processing applications, the data matrix of interest exhibits low rank structures [1]. Often, we know in addition that entries in a neighborhood are more likely to be correlated. If the size of the locality is not known, these locally low rank structures cannot be exploited using traditional low rank methods and the matrix itself may even be considered as full rank. Inspired by low rank + sparse modeling [2], [3], we present a way of representing these matrices compactly by decomposing the matrix into low rank blocks across multiple scales. Using this multi-scale modeling, we can capture different scales of correlation in our data matrix and provide a more compact representation than conventional low rank methods.

Our motivation comes from Dynamic Contrast Enhanced (DCE) Imaging in Magnetic Resonance Imaging. In DCE imaging, different tissue contrasts evolve differently over time. When stacking each image frame as a column of the matrix, the resulting matrix is low rank with various block sizes [1], [3]. A small block size captures blood vessel dynamics better while a large block size captures background tissues better. Hence, a multi-scale low rank approach is desired to exploit all scales of correlations.

SIGNAL MODEL AND PROBLEM FORMULATION

We model our $M \times N$ data matrix $Y$ as a sum of matrices $Y = \sum_{i=0}^{L-1} X_i$, where each $X_i$ has an increasing scale of locality. That is, each $X_i$ is block low-rank with a different block size, with the smallest one being a sparse matrix and the largest one being a low rank matrix. Fig. 1 provides an illustration of the modeling.

![Fig. 1: Illustration of multiscale low rank modeling](image)

We approach the problem via a convex formulation, which falls under the class of convex demixing problems [4]. We use the block-wise nuclear norm, which is the sum of singular values, as the convex surrogate for block matrix rank. Formally, for each $i$th scale we define $\| \cdot \|_{(m_i,n_i)}$ to be the block-wise nuclear norm with block size $(m_i,n_i)$. We choose $(m_0,n_0) = (1,1)$ and $(m_{i+1},n_{i+1}) = (2m_i,2n_i)$. The number of scales is $L = \log_2 \{ \max(M,N) \}$. To obtain the decomposition, we solve the following convex problem:

$$\min_{X_i; Y = \sum_{i=0}^{L-1} X_i} \sum_{i=0}^{L-1} \lambda_i \|X_i\|_{(m_i,n_i)}$$

To select the regularization parameters $\lambda_i$, we follow the suggestion from [5], and set each regularization parameter $\lambda_i = \sqrt{m_i} + \sqrt{n_i} + \log \{ \| \cdot \|_{(m_i,n_i)} \}$, the Gaussian width of each norm $\| \cdot \|_{(m_i,n_i)}$ [6]. We use the Alternating Direction of Multiple Multipliers (ADMM) to solve the proposed optimization problem. The computational cost is dominated by SVD’s. Since we decimate our matrix size for each scale, the overall complexity per iteration is twice the SVD complexity of the entire data matrix.

RESULTS

Figure 2 shows the proposed decomposition on a DCE dataset with image size $(152 \times 112)$ over 20 time frames, in which each scale captures different scales of dynamics. In Fig. 3, we also show that the multi scale decomposition can be applied on other types of datasets, e.g. faces, where different scales of occlusions are separated.

![Fig. 2: Two instances of multiscale low rank decomposition of DCE](image)

![Fig. 3: Two instances of multiscale low rank decomposition of faces](image)

REFERENCES

Annihilation-driven Image Edge Localization

Hanjie Pan∗, Thierry Blu †, and Martin Vetterli ∗

∗School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.
†Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong SAR.

Abstract—We propose an edge localization algorithm based on the spatial domain interpretation of the curve annihilations [3]. By adopting a simple linear edge parametrization, we have achieved edge localizations with sub-pixel accuracies.

I. INTRODUCTION

Edge detection is one of the most intensively studied subproblems in low-level computer visions, where the goal is to detect sudden changes (or “discontinuities”) present in an image. A number of approaches have been proposed to detect edges since 1960’s. We refer the readers to [1], [2] for the extensive reviews on various edge detection algorithms.

We propose an efficient sub-pixel accurate edge localization technique motivated by the work on annihilations for curves with finite rate of innovation (FRI) [3]. We will show that the edge models (including the edge tangential direction and the edge positions) can be obtained with simple linear filterings.

II. MOTIVATION: FRI CURVE ANNihilation

A specific class of (periodic) curves, which are defined implicitly as the roots of a certain mask function, is considered in [3]:

\[
C : \sum_{k = -K}^{K} \sum_{l = -L}^{L} c_{k,l} e^{j \frac{2\pi k}{\tau_x} x + \frac{2\pi l}{\tau_y} y} = 0, \quad \text{where} \quad 0 \leq x < \tau_x, \quad 0 \leq y < \tau_y.
\] (1)

Here \(\tau_x\) and \(\tau_y\) are some positive real numbers that specify the periods along \(x\)-\(y\) directions respectively. An associated edge image \(I_C(x, y)\) for each curve \(C\) is an image that is analytic almost everywhere except on the curve \(C\) where it is discontinuous. It is proved that the Fourier transform of the associated edge image satisfies the annihilation equation:

\[
\sum_{k = -K}^{K} \sum_{l = -L}^{L} c_{k,l} \mathcal{F}_C \left( \omega_x - \frac{2\pi k}{\tau_x}, \omega_y - \frac{2\pi l}{\tau_y} \right) = 0, \quad \forall \omega_x, \omega_y.
\] (2)

We may interpret the annihilation equation (2), which is a Fourier domain convolution, as a simple (continuous) domain multiplication:

\[
\sum_{k,l} c_{k,l} \mathcal{F}_C \left( \omega_x - \frac{2\pi k}{\tau_x}, \omega_y - \frac{2\pi l}{\tau_y} \right) \leftrightarrow \mu(x, y) \cdot I_C(x, y).
\]

Thus, the function \(\mu(x, y)\) serves as a “mask” that automatically annihilates whatever is different from zero in the derivative image \(I'_C(x, y)\). From this perspective, the curve model may be obtained by minimizing the annihilation errors e.g., in a least square sense: \(\min \| \mu \cdot I'_C \|^2\). Here we are no longer restricted to the mask parametrization (1) in general.

III. SUBPIXEL EDGE LOCALIZATION

If we zoom-in an image large enough, then we may well approximate the image edges with a straight line locally. In the annihilation framework, we can treat the local edge model as the roots (i.e., the zero-crossings) of a mask function, which is parametrized as:

\[
\mu(x, y) = c_1 x + c_2 y + c_3.
\]

With the spatial domain interpretation in the previous section, the best linear edge model within a block centred at \((x_0, y_0)\) should minimize the annihilation error:

\[
\min_{c_1, c_2, c_3} \int \int |\mu(x - x_0, y - y_0) \cdot I'(x, y)|^2 w(x - x_0, y - y_0) dx dy \quad \text{subject to} \quad c_1^2 + c_2^2 = 1.\]

Here \(w(\cdot, \cdot)\) is a localized window, e.g., a Gaussian with certain standard deviation: \(w(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)\); and the constraint is to avoid the trivial solution where all three coefficients are zero. In a matrix/vector form, (3) is

\[
\min_{c} e^{\mathbf{c}^T \mathbf{M} \mathbf{c}} \quad \text{subject to} \quad e^{\mathbf{c}^T \mathbf{B} \mathbf{c}} = 1,
\]

where \(\mathbf{c} = \begin{bmatrix} c_1 & c_2 & c_3 \end{bmatrix}\) and \(\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}\).

Here the entries of the \(3 \times 3\) Hermitian symmetric matrix \(\mathbf{M}\) are obtained directly by filtering the square of the derivative image \(|I'(x, y)|^2\). We can express the closed form solution of the quadratic minimization in terms of \(\mathbf{M}\)‘s entries explicitly. Subsequently, the edge point \((x_e, y_e)\) is defined as the one that is located on the estimated line model and yet has the shortest distance to the block centre \((x_0, y_0)\):

\[
(x_e, y_e) = \arg \min_{x,y} (x - x_0)^2 + (y - y_0)^2 \quad \text{subject to} \quad c_1(x - x_0) + c_2(y - y_0) + c_3 = 0.
\]

Figure 1 illustrates the edge localization results with coco test image.

REFERENCES

Discrete vs Continuous Optimization for Gene Regulatory
Network Inference

Aurélie Pirayre∗§, Camille Couprie ∗, Laurent Duval† and Jean-Christophe Pesquet§
∗ IFP Energies nouvelles, Direction Mécatronique et Numérique, Rueil-Malmaison, France
§ Univ. Paris-Est, LIGM, UMR CNRS 8049, France

Abstract—With the advent of high-throughput biological techniques, arose the need to handle gene expression data. Inferring Gene Regulatory Networks (GRNs) from this kind of data is especially useful for sketching transcriptional regulatory pathways and helps to understand phenotype variations. Given all pairwise gene similarity information, we formulate GRN inference as an energy minimization problem to determine the presence of edges in the final graph. Taking into account expected patterns in the graph structure, biological a priori are incorporated into the variational formulation. Different priors lead to different mathematical properties of the cost function, for which various optimization strategies can be applied. Experimental results show a performance improvement (in terms of Area Under the Precision-Recall curve) and/or computation time compared with state-of-the-art methods.

I. INTRODUCTION

One way of improving biological knowledge is to handle and analyze “omics” data generated by high-throughput techniques. Focusing on the context of transcriptomic data, the identification of genes involved in phenotypic variations is currently performed thanks to Gene Regulatory Network (GRN) analysis such as gene clustering [1]. A GRN is a graph containing gene regulatory pathways for a given living organism. It is obtained from gene expression signals: for each gene, the signal corresponds to the gene expression level in different conditions (physico-chemical or temporal conditions, culture medium or mutated strains). Then, inferring a GRN aims at selecting, among all plausible links, a subset of regulatory links reflecting actual regulatory relationships between genes. Unfortunately, recovering useful information from this collection of signals remains a difficult task due to the small number of observations (number of conditions) compared with the number of genes. In this work, we develop a novel variational approach for taking into account expected graph patterns according to some biological a priori. Translating such biological assumptions into an appropriate cost function, we thus formulate the GRN inference problem as an optimization one.

II. MODELS

A complete gene network may be viewed as a graph $G(V, E)$, where $V = \{v_1, \ldots, v_n\}$ is a set of vertices (corresponding to the genes), $g$ is the number of genes, and $E = \{e_1, \ldots, e_\mathbf{N}\}$ a set of edges (corresponding to plausible gene interactions), the number of edges being $n = g(g - 1)$. Inferring a GRN $G^*$ from $G$ aims at selecting an optimal subset of edges $E^* \subset E$ reflecting actual regulatory relationships between genes. This selection problem may be formulated by defining a cost function to minimize where the variables $x_{i,j}$ correspond to edge labels for $e_{i,j}$ such that $x_{i,j} = 1$ if the edge $e_{i,j}$ is in the final graph and 0 otherwise.

Weighting all possible pairwise gene relationships by the similarity $s_{i,j}$ between gene expression profiles for gene $i$ and $j$ and assuming that a reliable list of putative transcription factors (i.e. regulator genes), denoted by $\mathcal{T}$, is available, we define some biological and structural a priori which may be incorporated into our cost function based on two rationales: i) the larger the edge weights $s_{i,j}$, the more favorable the edge selection, ii) links involving a regulator gene are favored. Two variational priors were used according to the choice of function $\Phi$, leading to the following general criterion form to minimize:

$$\sum_{(i,j) \in E} s_{i,j}(1 - x_{i,j}) + \sum_{(i,j) \in E} \lambda_{i,j} x_{i,j} + \mu \Phi((x_{i,j})(i',j') \in \mathcal{N}_{i,j}),$$

where $\mu$ is a regularization parameter and, for every $(i,j)$, $\mathcal{N}_{i,j}$ denotes some local neighborhood of edge $e_{i,j}$. Depending on the prior used, mathematical properties of cost function (1) are changed and suitable optimization strategies have to be devised.

- Keeping the degree of regulated genes close to a constant number $d$ is enforced by choosing $\Phi$ as a composition of a linear averaging operator with a norm. A relaxation of the binary constraint on the vector $x$ of edge labels is then necessary to minimize (1) efficiently by using recent convex optimization methods [2].

- Enforcing a co-regulation property (i.e. favoring a similar label for $x_{i,j}$ and $x_{j,i}$ when genes $j$ and $j'$ are likely to act together) $\text{via}$ a total variation like function $\Phi$ makes the criterion submodular. Thus, a discrete optimization process can be carried out, such as a maximum flow algorithm, in order to obtain an optimal labeling [3].

III. RESULTS

We performed comparisons of our approach with two state-of-the-art methods: Context Likelihood of Relatedness (CLR) [4] and GENIE3 [5]. The performance was evaluated in terms of Area Under the Precision-Recall curves (AUPR), where the precision reflects the proportion of correctly inferred edges compared to the total number of inferred edges, while the recall indicates the proportion of correctly inferred edges with respect to the edges corresponding to the gold standard. Results obtained on the synthetic data from the DREAM4 multifactorial challenge are quite promising. In term of AUPR, our method outperforms both CLR and GENIE3 approaches while having a low computational complexity.

REFERENCES


Fast IRLS for sparse reconstruction based on gaussian mixtures

Chiara Ravazzi and Enrico Magli
Dipartimento di elettronica e delle telecomunicazioni, Politecnico di Torino, Italia.

Abstract—The theory of compressed sensing has demonstrated that sparse signals can be reconstructed from few linear measurements. In this work, we propose a new class of iteratively reweighted least squares (IRLS) for sparse recovery. The proposed methods use a two state Gaussian scale mixture as a proxy for the signal model and can be interpreted as an Expectation Maximization algorithm that attempts to perform the constrained maximization of the log-likelihood function. Under some conditions, standard in the compressed sensing theory, the sequences generated by these algorithms converge to the fixed points of the maps that rule their dynamics. A condition for exact sparse recovery, that is verifiable a posteriori, is derived and the convergence is proved to be quadratically fast in a neighborhood of the desired solution. Numerical experiments show that these new reconstructions schemes outperform classical IRLS for  \( \ell_1 \)-minimization with \( \tau \in (0, 1) \) in terms of rate of convergence and accuracy.

I. SPARSE RECOVERY VIA IRLS FOR  \( \ell_\tau \)-MINIMIZATION

The basic principle of the compressed sensing theory is that a \( k \)-sparse signal \( x^* \in \mathbb{R}^n \) (i.e., it has at most \( k \) nonzero entries) can be recovered from a smaller number \( m < n \) of linear measurements \( y = Ax^* \in \mathbb{R}^m \) than traditional sampling theory believed necessary \([2]\). The estimation of the sparsest solution, consistent with the observations, is an NP-hard problem. However, the constrained  \( \ell_\tau \)-minimization with \( \tau \in (0, 1) \), which is the convex or nonconvex surrogate problem, has been proposed in \([3]\) as an appealing alternative for sparse recovery. It consists in selecting the element which is compatible with the observations which has minimal  \( \ell_\tau \)-norm with \( \tau \in (0, 1) \):

\[
\min_{x \in \mathbb{R}^n} \|x\|_{\ell_\tau}, \quad \text{s.t. } y = Ax. \tag{1}
\]

Under certain assumptions on the sensing matrix \( A \), it is known that \( (1) \) has a unique solution and it provides the desired solution \( x^* \).

The minimization in \((1)\) can be carried out by an iteratively reweighted least squares method (IRLS, \([4]\)). More precisely, given an initial guess \( x^{(0)} \), at each iteration the algorithm requires to solve a constrained weighted least-squares problem:

\[
x^{(t+1)} = \arg \min_{y = Ax} \sum_{i=1}^n w_i^{(t)} x_i^2 \text{ s.t. } y = Ax
\]

with \( w_i^{(t+1)} = (\epsilon_i^{(t)})^2 + (x_i^{(t)})^2)^{\tau/2 - 1} \) and a suitable non-increasing sequence \( \epsilon_i^{(t)} \). In particular, under certain assumptions, these methods have been proved to converge to an \( x^* \) globally linearly fast when \( \tau = 1 \) and locally superlinearly fast with rate \( 2 - \tau \) for \( \tau \in (0, 1) \).

Although classical IRLS algorithms appear very attractive for their simplicity, theoretical results guarantee the superlinear convergence only in a neighborhood of the desired solution. In fact, numerical results point out that exact recovery is achieved when \( \tau \) is not too small (i.e. \( \tau > 1/2 \)) and tends to be trapped in local minima when \( \tau < 1/2 \) \([5]\). Heuristic techniques to avoid local minima are currently object of study.

II. GSM BASED IRLS

We derive a new class of IRLS procedures for sparse recovery which outperform the classical procedures. More precisely, we model the elements of the signal as a two state gaussian mixture (GSM, \([1]\))

\[ x^*_i = z_i \sqrt{\alpha u_i} + (1 - z_i) \sqrt{\beta u_i} \]

where \( u_i \) are identically and independently distributed (i.i.d.) zero mean Gaussians and \( z_i \) are i.i.d. Bernoulli variables with probability mass function \( \mathbb{P}(z_i = 1) = 1 - p, p = k/n, \alpha \approx 0, \) and \( \beta \gg 0 \). The combination of the considered model, used as a proxy for the sparsity assumption, with the maximum log-likelihood estimation provides a new alternative to select the sparsest vector consistent to the data. More precisely, we want to minimize

\[
L(x, z, \alpha, \beta, \epsilon) = \sum_{i=1}^n \left[ \frac{z_i x_i^2 + \epsilon_i^2/n}{2\alpha} + \frac{z_i}{2} \log \frac{\alpha}{1-p} \right] + \frac{(1-z_i) x_i^2 + \epsilon_i^2/n}{2\beta} + \frac{(1-z_i)}{2} \log \frac{\beta}{p}.
\]

subject to the constraint \( y = Ax \). We design three iterative techniques: ML-based IRLS, EM-based IRLS, and K-EM based IRLS. These strategies can be interpreted as instances of the Expectation Maximization algorithm. After choosing some initial values for the mixture parameters, two updates are alternated: in the E-step, we use the current values for the parameters to estimate the signal \( x^* \) and to evaluate the posterior distribution \( \mathbb{P}(z_i = 1) \) of the signal coefficients; in M-step we use these probabilities to re-estimate the mixture parameters \( \alpha \) and \( \beta \).

Besides the design of the algorithms, we prove that, under suitable conditions, the sequence of provided estimations converges to a fixed point of the map that rules their dynamics. Moreover, we derive conditions for exact recovery that are verifiable a posteriori. Finally, the algorithm turn out to be quadratically fast in a neighborhood of \( x^* \). Numerical simulations validate our claims and show that these new procedures avoid local minima, outperforming classical IRLS for sparse recovery in terms of rate of convergence and sparsity-undersampling tradeoff.

REFERENCES

Abstract—In the field of inverse problems, one of the main benefits which can be drawn from primal-dual optimization approaches is that they do not require any linear operator inversion. In addition, they allow to split a convex objective function in a sum of simpler terms which can be dealt with individually either through their proximity operator or through their gradient if they correspond to smooth functions. Proximity operators constitute powerful tools in nonsmooth functional analysis which have been at the core of many advances in sparsity aware data processing. Using monotone operator theory, the convergence of the resulting algorithms can be shown to be theoretically guaranteed.

In this paper, we provide a survey of the existing proximal primal-dual approaches which have been proposed in the recent literature. We will also present new developments based on a randomization of these methods, which allow them to be applied block-coordinatewise or in a distributed fashion.

I. INTRODUCTION

Recently, there has been a growing interest in primal-dual methods for optimizing functions of a high number of variables. A great majority of such optimization problems arising in computer vision, inverse problems, or machine learning can be formulated as follows:

\[
\begin{align*}
\text{minimize} \quad & \sum_{j=1}^{p} \left( f_j(x_j) + h_j(x_j) \right) + \sum_{k=1}^{q} (g_k \square l_k) \left( \sum_{j=1}^{p} L_{k,j} x_j \right) \\
\text{subject to} \quad & x_j \in H_j, \quad j = 1, \ldots, p.
\end{align*}
\]

(1)

where, for every \( j \in \{1, \ldots, p\} \), \( H_j \) is a separable real Hilbert space, \( f_j \in \Gamma_0(H_j) \), \( h_j \in \Gamma_0(H_j) \) is Lipschitz-differentiable, and, for every \( k \in \{1, \ldots, q\} \), \( G_k \) is a separable real Hilbert space, \( g_k \in \Gamma_0(G_k) \), \( l_k \in \Gamma_0(G_k) \) is strongly convex, and \( L_{k,j} : H_j \to G_k \) is a bounded linear operator.\(^1\) The symbol \( \square \) here denotes the inf-convolution operation, that is

\[
(\forall k \in \{1, \ldots, q\})(\forall v_k \in G_k)(g_k \square l_k)(v_k) = \inf_{v_k^* \in C_{G_k}} g_k(v_k^*) + l_k(v_k - v_k^*).
\]

(2)

In particular, if \( l_k \) reduces to the indicator function of \( \{0\} \), \( g_k \square l_k = g_k \), which corresponds to a case frequently encountered in practical applications. One of the main interests in considering the general formulation in (1) is that the dual problem has a fully symmetric form:

\[
\begin{align*}
\text{minimize} \quad & \sum_{j=1}^{p} \left( f_j^* (h_j^*)_j \right) \left( - \sum_{k=1}^{q} L_{k,j}^* v_k \right) + \sum_{k=1}^{q} (g_k^*(v_k^*) + l_k^*(v_k^*)), \\
\text{subject to} \quad & v_k \in H_k, \quad k = 1, \ldots, q.
\end{align*}
\]

(3)

where \( \varphi^* \) designates the Fenchel-Legendre conjugate of a function \( \varphi \). Solving (1) raises a number of difficulties due to the facts that (i) the space \( H = H_1 \oplus \cdots \oplus H_p \) in which the unknown solution lies is often very high-dimensional, (ii) the functions \( (f_j)_{1 \leq j \leq p} \) and \( (g_k)_{1 \leq k \leq q} \) may be nonsmooth so as to promote the sparsity of the solution or to impose some hard constraints on it,

\(^1\)\( \Gamma_0(H) \) denotes the class of lower-semicontinuous convex functions with nonempty domain defined on \( H \) and taking their values in \([-\infty, +\infty]\).

(iii) the number of functions may be large, and (iv) the involved linear operators may be difficult to (pseudo-)inverted because of their size and lack of structure. To overcome these difficulties, primal-dual algorithms have been recently developed which aim at jointly solving the primal and the dual problems instead of dealing with the primal problem only.

II. AN OVERVIEW OF RECENT ALGORITHMS

Proximal primal-dual methods for solving (1) and (3) (or specific instances of them) can be decomposed in three main families (see [1] and the references therein for more details):

- forward-backward based methods: these methods are based on the forward-backward iteration, which means that they combine proximal steps and gradient steps. Two main variants of these methods have been proposed in the literature.
- forward-backward-forward based methods: compared with the previous approaches, these techniques include an extra-gradient step. They were the first proposed methods allowing to address Problem (1) under its general split form.
- projection-based methods: these quite recent approaches present the potential advantage of requiring no knowledge on the norms of the linear operators (or an upper bound of it). They have been however little used in practice, up to now.

Parallel versions of all these algorithms are available, but they require to activate all the primal and dual variables at each iteration which may be intensive in terms of computation and memory requirements. In the spirit of the work performed for designing stochastic versions of forward-backward-like variational methods [2], efforts have been made in order to develop block-coordinate versions of the primal-dual algorithms belonging to the first family. These novel algorithms are grounded on recent theoretical results concerning the randomization of proximal algorithms, and more generally fixed point algorithms for solving monotone inclusion problems [3].

Interestingly, by using consensus techniques, it is possible to deduce from these block-coordinate primal-dual algorithms, distributed algorithms involving linear operators where a splitting is performed on the functions instead of the variables [4].

REFERENCES


Abstract—We review MAP and MMSE-based approaches to image recovery and their implementation via generalized approximate message-passing (GAMP), highlighting recent results on GAMP convergence for general measurement operators.

We consider the recovery of image $\mathbf{x} \in \mathbb{C}^N$ from noisy outputs $\mathbf{y} \in \mathbb{C}^M$ of known linear measurement operator $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$. The “statistical” approach to image recovery models the image $\mathbf{x}$ as a realization of random $\mathbf{X} \sim p_{\mathbf{x}}$ and the measurements as a realization of random $\mathbf{Y}$ whose statistics are given by a likelihood function of the form $p_{\mathbf{Y}}(\mathbf{y} | \mathbf{\Phi} \mathbf{x})$. Here, $p_{\mathbf{Y}}$ is the pdf of $\mathbf{y}$ conditioned on the (hidden) transform outputs $\mathbf{z} = \mathbf{\Phi} \mathbf{x}$ and $\mathbf{x}$ is a hypothesis of the image. For clarity, we denote random quantities in san-serif font. Which can be interpreted (from a non-statistical viewpoint) as regularized loss minimization, i.e.,

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\max_{\mathbf{x}} p_{\Phi \mathbf{Y}}(\mathbf{y} | \mathbf{\Phi} \mathbf{x}) p_{\mathbf{x}}(\mathbf{x})$$

using the loss $L(\mathbf{z}) = -\log p_{\mathbf{Y}}(\mathbf{y} | \mathbf{\Phi} \mathbf{x})$ and regularization $R(\mathbf{x}) = -\log p_{\mathbf{x}}(\mathbf{x})$. By choosing $p_{\mathbf{Y}}$ and $p_{\mathbf{x}}$ so that both $L(\cdot)$ and $R(\cdot)$ are convex, one can readily apply convex optimization algorithms to the image recovery problem. In image recovery, it is popular to use regularizations of the form $R(\mathbf{x}) = \|\mathbf{D}\mathbf{x}\|$ for a given matrix $\mathbf{D}$.

In the minimum mean-squared error (MMSE) approach to statistical image recovery, the objective is to compute

$$\hat{\mathbf{x}}_{\text{MMSE}} = \mathbb{E}[\mathbf{x} | \mathbf{y}] = \int_{\mathbb{C}^N} p_{\mathbf{Y} | \mathbf{x}}(\mathbf{y} | \mathbf{x}) d\mathbf{x},$$

with the hope of mean-square optimal performance. Unfortunately, the high-dimensional integral in (3) is computable in closed-form for only a very narrow class of priors and likelihoods (e.g., Gaussian) and even then may require the inversion of a very large matrix.

For problems with separable loss and regularization, i.e.,

$$L(\mathbf{z}) = \sum_{i=1}^M L_i(\tilde{z}_i)$$

and

$$R(\mathbf{x}) = \sum_{j=1}^N R_j(\tilde{x}_j),$$

a computationally efficient inference methodology that supports either MAP or MMSE recovery was recently proposed under the name of “generalized approximate message passing” (GAMP) [1]. GAMP is an extension of the AMP algorithm [2] from quadratic loss (i.e., $L_i(\tilde{z}_i) = \frac{1}{2} (\tilde{z}_i - \mu)^2$ for some $\mu > 0$) to generic loss $L_i(\cdot)$, as needed for phase retrieval, Poisson noise, or quantized measurements. Interestingly, the behavior of GAMP for large i.i.d. $\mathbf{\Phi}$ is rigorously characterized by a state evolution whose fixed points, when unique, are MAP or MMSE optimal [3]. Still, important questions remain about the convergence of GAMP for generic $\mathbf{\Phi}$, and whether GAMP can be applied to non-separable regularizers like $\|\mathbf{D}\mathbf{z}\|_1$, which are commonly used in image recovery.

In this talk, we review recent results on the convergence of GAMP for generic $\mathbf{\Phi}$. First, for any $\mathbf{\Phi}$, we recall that the fixed points of MAP-GAMP are known to coincide with the critical points of the optimization (2) [4]. Meanwhile, the fixed points of MMSE-GAMP are known to coincide with the critical points of the optimization [4]

$$(f_z, f_{\mathbf{z}}) = \arg\min_{\mathbf{z}, \mathbf{b}} J(b_z, b_{\mathbf{z}}) \text{ s.t. } \mathbb{E}[\mathbf{z} | b_{\mathbf{z}}] = \mathbf{\Phi} \mathbb{E}[\mathbf{X} | b_{\mathbf{x}}]$$

$$J(b_z, b_{\mathbf{z}}) \triangleq D(b_z \| p_{\mathbf{z}}) + D(b_z \| p_{\mathbf{y} | \mathbf{\Phi} \mathbf{x}} Z^{-1}) + H(b_z, \nu^0),$$

where $J$ is a high-dimensional approximation of the Bethe free energy [5]. In (6), $b_z(\mathbf{z}) = \prod_i b_z(z_i)$ and $b_{\mathbf{z}} = \prod_i b_{\mathbf{z}}(z_i)$ are pdfs, $D(\cdot \| \cdot)$ denotes KL divergence, and $H(b_z, \nu^0) \triangleq \sum_i \text{var}(z_i | \nu_i) = \text{var}(z_i | \nu_i)$. But these fixed points don’t tell the whole story, because GAMP may diverge.

For quadratic $L_i(\cdot)$ and $R_j(\cdot)$, however, the convergence of GAMP has been fully characterized, and global convergence can be ensured by “damping” [6]. Damping can also be used to ensure local convergence under strictly convex $L_i(\cdot)$ and $R_j(\cdot)$ [6].

We also review recent results on connections between GAMP and convex optimization algorithms. For example, with MAP-GAMP, the variable updates coincide with those of the primal-dual hybrid gradient (PDHG) approach to (2) while the stepsizes are adapted in accordance with the local cost [6]. Meanwhile, with MMSE-GAMP, the mean updates coincide with an application of PDHG to (5) under a local convexification of the augmented Lagrangian, while the variance updates adapt that local convexification. Finally, we describe a recent variant on MMSE-GAMP that guarantees global convergence with generic $\Phi$ for strictly convex $F, G$ with bounded derivatives.

Finally, we describe how GAMP can be configured to use “analysis” non-separable regularizers $R(\mathbf{x}) = \sum_{d=1}^D G_d(\|\mathbf{D}_d \mathbf{x}\|)$ [7].

REFERENCES


A Primal-Dual Algorithmic Framework for Constrained Convex Optimization

Q. Tran-Dinh∗, and V. Cevher∗
∗ Laboratory for Information and Inference Systems (LIONS), EPFL, CH1015-Lausanne, Switzerland.

Abstract—In this work, we develop a class of primal-dual algorithms for solving constrained convex optimization problems arising from applications. The main idea is to exploit state-of-the-art techniques in convex optimization such as smoothing, model-based excessive gap functions and augmented Lagrangian, and combine them in a smart fashion to develop scalable algorithms for solving such constrained problems. We present a unified primal-dual optimization framework and investigate its global convergence and convergence rate. Then, we customize our framework to obtain different variants including alternating direction methods of multipliers (ADMM) and preconditioned alternating direction methods of multipliers. Many numerical examples are presented and comparison with existing methods is also made.

I. INTRODUCTION

Many practical optimization problems in different disciplines such as machine learning, signal/image processing, statistics and engineering can conveniently be formulated into the following constrained convex optimization problem:

\[
f^* := \min_{x \in X} \{ f(x) \mid Ax = b, \}
\]

where \( f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) is proper, closed and convex, \( X \subseteq \mathbb{R}^n \) is a nonempty, closed and convex set, \( A \in \mathbb{R}^{m \times n} \) and \( b \in \mathbb{R}^m \).

The aim of this work is to develop scalable primal-dual algorithms for solving (1) by exploiting the underlying structures of this problem, especially of \( f \) and \( X \). We mainly rely on two key structures that stand out among many others:

1. Separability: We say that problem (1) is separable if both \( f \) and \( X \) are separable, i.e., they can be represented as:

\[
f(x) = \sum_{i=1}^{p} f_i(x_i), \quad X = X_1 \times \cdots \times X_p,
\]

where \( p \) is the number of components, \( X_i \subseteq \mathbb{R}^{n_i} \) and \( \sum_{i=1}^{p} n_i = n \). The separable structure (also called the decomposable structure) is very useful to develop scalable algorithms by decomposing large-scale problems into smaller components so that they can be solved in a parallel and distributed manner. This structure plays crucial role to design decomposition methods in large-scale convex optimization.

2. Tractable proximity: Proximal operator is a key concept in convex optimization. Formally, for a proper, closed and convex function \( g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \), the proximal operator of \( g \) is defined as:

\[
\text{prox}_g(x) := \arg \min_{z \in \mathbb{R}^n} \{ g(z) + \frac{1}{2} \| z - x \|^2 \}.
\]

We say that \( g \) is tractably proximal if its proximal operator can be computed efficiently (e.g., closed form or by polynomial time algorithms). Examples of tractable proximity functions can be found in the literature. Proximal operators play a central role to develop proximal-type algorithms and splitting techniques for solving composite minimization as well as nonsmooth structural convex problems.

Our aim in this work is to exploit these two structures to develop a primal-dual algorithmic framework and customize it to solve a broad problem class of the form (1). We also aim at investigating its global convergence rate and its connections to existing methods.

II. OUR APPROACH AND THEORETICAL CONTRIBUTIONS

Our approach is based on three techniques: smoothing techniques, model-based excessive gap functions [1] in variational inequalities and primal-dual frameworks. First, we propose two smoothing techniques: one via Bregman distances and the other via augmented Lagrangian functions. Next, we exploit a model-based excessive gap function using in variational inequality to characterize the optimality condition of (1) and show how to smooth this gap function in order to measure the primal-dual objective residual and the primal-feasibility gap. Then, we present two update schemes using either one primal step and two dual steps (1P2D) or two primal steps and one dual step (2P1D) to update the primal-dual iterative sequence that maintains the model-based excessive gap condition. Finally, we prove an analytical update rule for parameters so that we obtain a complete primal-dual algorithm with full convergence theory.

The main contributions of this work are as follows. We propose a novel primal-dual framework for solving (1) that covers both smoothing techniques: Bregman smoothing and augmented Lagrangian smoothing. We prove the global convergence of the algorithm for two cases, which guarantees the convergence separately for the primal objective residual \( |f(x^k) - f^*| \) and the primal feasibility gap \( \|Ax^k - b\| \). More precisely, we show that:

- For the Bregman smoother: The global convergence rate is \( O(1/k) \) both on \( |f(x^k) - f^*| \) and on \( \|Ax^k - b\| \), where \( k \) is the iteration counter.

- For the augmented Lagrangian smoother: The global convergence rate is \( O(1/k^2) \) both on \( |f(x^k) - f^*| \) and on \( \|Ax^k - b\| \).

We also customize this algorithmic framework to obtain different variants. The first variant is an ADMM and the second one is a preconditioned ADMM. All algorithms are tuning-free so that all parameters are updated automatically. We also propose several implementation enhancements to improve the actual performance of our algorithms. Details of the algorithmic framework, convergence theory and its variants can be found in our manuscript [2].

III. APPLICATIONS AND NUMERICAL SIMULATIONS

Our algorithmic framework is applicable for a broad class of convex problems including composite convex minimization and non-smooth separable constrained convex problems. Concrete examples include sparse and group sparse recovery, sparse logistic regression, image de-blurring, image denoising, robust principle component analysis and Poisson image reconstruction. Details of numerical examples can be found in our manuscript [2].

Acknowledgements: This work is supported in part by ERC Future Proof, MIRG-268398, SNF 200021-132548, SNF 200021-146750 and SNF CRSII2-147633.

REFERENCES


Asymptotic Inference for Integral Curves of Noisy Vector Fields

Susan Wei and Victor Panaretos
Mathematics Section, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

Abstract—Noisy vector fields arise in many disciplines of study. One of the most fundamental operations on a vector field is its integration. We derive conditions under which the integral curve of a noisy vector field is asymptotically normal. Results in this vein have previously been obtained in the specific contexts of diffusion tensor imaging in neuroimaging and filament estimation in astrophysics. The theory we build decouples the analysis from the specific vector field estimator used. Numerous illustrative examples will be given to demonstrate the wide-ranging applicability of our results.

I. INTRODUCTION

Let $U$ be an open subset of $\mathbb{R}^d$, $v_0$ a vector field in $\mathbb{R}^d$, and $p \in \mathbb{R}^d$ a fixed initial value. Consider the initial-value problem of finding a differentiable function $\gamma(t)$ from $[a, b] \subset \mathbb{R}$ to $\mathbb{R}^d$ that satisfies

$$
\dot{\gamma}(t) = (v_0 \circ \gamma)(t) \quad \forall t \in [a, b]
$$

$$
\gamma(a) = p
$$

A solution $\gamma_0(t)$ of (1) is a curve in $U$ starting at $p$, with the property that the tangent vector of the curve at the point $\gamma_0(t)$ coincides with the value of the vector field $v_0$ at $\gamma_0(t)$. The curve $\gamma_0$ is called an integral curve of the vector field $v_0$.

We consider the challenge of estimating the integral curve of $v_0$ from a noisy vector field $\hat{v}_n$. A natural estimate is $\hat{\gamma}_n(t)$, the solution to the plug-in version of (1), where $v_0$ is replaced by $\hat{v}_n$. We shall demonstrate sufficient conditions that guarantee the asymptotic normality of $\hat{\gamma}_n$.

II. MAIN RESULT

We present a summary of the assumptions needed for asymptotic normality of the integral curve: (A1) conditions to guarantee existence and uniqueness of solution to (1); (A2) condition to guarantee that the Frechet derivative of the objective function defining $\gamma$ is continuously invertible; (A3) the Jacobian matrix of $v_0$ is strictly diagonally dominant; (A4) metric entropy condition for continuity of the limiting Gaussian process. Let $J$ be the Jacobian of $v_0$ evaluated along $\gamma_0$ and $\Gamma = G \circ \gamma_0$ where $G$ is the limiting Gaussian random field of $r_n(\hat{v}_n - v_0)$.

Theorem 1. Suppose A1-A4 hold. First we have that there exists $\beta \in (a, b)$ such that the initial-value problem (1) has a unique solution $\gamma_0(t)$ defined on the interval $a \leq t \leq \beta$. Furthermore, we have

$$
r_n(\gamma_n - \gamma_0) \overset{d}{\rightarrow} \xi
$$

in $C([a, \beta], \mathbb{R}^d)$ where $\xi$ is the mean zero Gaussian process that satisfies

$$
J(t)\xi(t) - \dot{\xi}(t) = \Gamma(t), \quad \forall t \in [a, \beta]
$$

$$
\xi(a) = 0
$$

Notice the limiting Gaussian process $\xi$ depends on unknown parameters, namely $v_0$ and $\gamma_0$. As a workaround, we will use the solution to the empirical version of Equation (2) as a proxy for $\xi$.

Consider the solution $\hat{\xi}_n(t)$ to the following system

$$
J(t)\hat{\xi}_n(t) - \dot{\hat{\xi}}_n(t) = \hat{\Gamma}(t), \quad \forall t \in [a, \beta]
$$

$$
\hat{\xi}_n(a) = p
$$

for some appropriately defined $\hat{J}$ and $\hat{\Gamma}$.

Theorem 2. Suppose the conditions of Theorem 1 hold, then $\hat{\xi}_n \overset{d}{\rightarrow} \xi$ in $C([a, \beta], \mathbb{R}^d)$.

III. EXAMPLES

The asymptotic results above can be applied to a wide-range of applications. Here we describe the setup of a few example applications.

a) Diffusion Tensor Imaging: In this example, we study the setting considered in [1]. We observe $x_i \in \mathbb{R}^d$, $i = 1, \ldots, n$ uniformly and $v_i$ according to

$$
v_i = v_0(x_i) + \epsilon_i
$$

where $E\epsilon = 0$ and $Cov(\epsilon, \epsilon) = \Sigma$. Let $\hat{v}_n$ be the kernel estimator of $v_0$ based on $v_1, \ldots, v_n$. The integral curve in this model is used in [1] for white matter fiber tractography in diffusion tensor imaging.

b) Filament Estimation: We can use our results to study the setting considered in [2]. Let $x_1, \ldots, x_n \in \mathbb{R}^2$ be a sample from a distribution with density $f$. The vector field of interest, $v_0$, is the gradient of $f$. The noisy vector field $\hat{v}_n$ is the gradient of $f_n$, the kernel density estimator of $f$. The motivation for this model arises from the problem of filament estimation in astrophysics.

c) Local Covariance Field: In this example, our initial starting point of analysis is a tensor field rather than a vector field. We construct a tensor field by associating a local covariance matrix to each point in an $\mathbb{R}^2$ point cloud. Take the vector field of interest, $v_0$, to be the first eigenvector of the covariance matrix. We study the asymptotics of the integral curve as a precursor to the study of the asymptotics of principal flows, introduced in [4]. The principal flow is a non-greedy type of integral curve and the analysis here could lend insight to this more complex object.

REFERENCES


Small-Feature Model-Based Image Segmentation

Dawn Woodard*, Roseline Bilina Falafala†, and Ciprian Crainiceanu‡

* School of Operations Research & Information Engineering, Cornell University, Ithaca, NY
† Department of Statistics, Columbia University, New York, NY
‡ Department of Biostatistics, Johns Hopkins University, Baltimore, MD

Abstract—We present an improved method for spatial model-based clustering, and apply it to segment three-dimensional Dynamic Contrast Enhanced Magnetic Resonance (DCE-MR) images. The proposed method is shown to capture small features in the image, such as small lesions or tumors, more effectively than existing approaches. We apply our method to segment DCE-MR images of a subject with multiple sclerosis. In DCE-MR imaging the concentration of a contrast agent in the brain is monitored over time; we segment the brain according to this concentration trajectory. Unlike existing methods, ours yields medically informative segmentations for this application, for instance accurately identifying a particular type of multiple sclerosis lesion.

I. INTRODUCTION

Image segmentation consists of partitioning an image into possibly non-contiguous regions, within which the measurement values are relatively homogeneous (Figure 1). It can facilitate interpretation or further analysis of the image, and has numerous applications in medicine, video and image compression, and remote sensing. Model-based clustering provides a formal statistical approach to image segmentation, allowing natural characterization of the regions and extension to other contexts such as time-course imaging, multiple subjects, non-Euclidean measurements, and longitudinal data.

The images are typically two- or three-dimensional, and consist of measurement(s) at each of a grid of voxels (the general term for a pixel in more than two dimensions). A popular approach to segmentation uses a mixture model based on a Markov random field; this captures the spatial association of the voxels, but is computationally challenging, due in part to an unknown normalizing constant in the likelihood.

Figure 1: (a)-(d): A two-dimensional slice of a DCE-MRI study at four time points. The black boundary delineates white matter; the bright spot inside the white matter in (c)-(d) corresponds to a multiple sclerosis lesion. (e): The same slice of a three-dimensional segmentation of the white matter, obtained using our method.

We extend an existing Monte Carlo Expectation-Maximization (MCEM) method for image segmentation based on Markov random fields, and apply it to three-dimensional Dynamic Contrast Enhanced Magnetic Resonance (DCE-MR) images of a subject with multiple sclerosis (MS). First, we show how to incorporate cluster weight parameters in a computationally tractable way. Second, we incorporate a covariance decomposition into the model for the mixture components. Third, we give a consistent approximation to the observed-data likelihood (an approximation that converges to the true value in the limit of the number of Monte Carlo samples). Our method, unlike existing approaches, accurately distinguishes small features like MS brain lesions, and yields medically informative segmentations in our application.

The computational and statistical challenges in the context of DCE-MRI include: 1. handling three-dimensional images, instead of a two-dimensional slice as in most previous work using Markov random fields; 2. the number of voxels, which is an order of magnitude larger than in most previous work; and 3. the time-series nature of the data.

We show in a simulation study that our method, unlike the existing MCEM method, is able to accurately capture regions of unequal sizes. We further illustrate this advantage in an application to DCE-MR images of a MS subject. In DCE-MR imaging the subject is injected with a contrast agent, and the concentration of the agent in the tissue is monitored over time using a series of MR images. This contrast enhancement improves viewing of some features of interest. For instance, MS subjects have brain lesions, which are associated with increased permeability of the blood-brain barrier; a primary method for characterizing these lesions is by observing the diffusion of a contrast agent into the brain tissue.

A two-dimensional slice of a DCE-MRI study at four time points is shown in Figure 1 (a)-(d); the region of primary interest is the white matter, delineated by the black boundary. A lesion is visible in (c)-(d) as a bright spot inside the white matter, where the contrast agent concentration is high; such lesions are called "enhancing lesions." A two-dimensional slice of an example segmentation using our approach is shown in Figure 1(e). The segmentation automatically distinguishes the enhancing lesions, and identifies other distinct regions within the white matter, which may have relevance in terms of the disease process.
Abstract—One of important classes of sparse signals is the non-negative sparse signals. Canonical greedy techniques have been modified to incorporate the non-negativity of the representations. One such modification has been proposed for Orthogonal Matching Pursuit (OMP), which chooses positive coefficients first and uses a non-negative least square as a replacement for the orthogonal projection onto the selected support, at each iteration of the algorithm, which is computationally expensive. We here present a very different modification to the canonical OMP implementation, which truly incorporate the non-negativity of the coefficients. We also present a novel fast implementation of the Non-Negative OMP (NNOMP) which is based on the QR decomposition. As a result we show that we may receive an acceleration of a factor of up to ten in a reasonable size problem with the new method\(^1\).

I. INTRODUCTION

Let \( y \in \mathbb{R}^m \) be a signal having the sparse representation using a normalised linear generative model \( \Phi \in \mathbb{R}^{m \times n} \), i.e. dictionary, where \( y = \Phi x \) and \( x \) is a sparse vector. Orthogonal Matching Pursuit gradually adds one element of the dictionary, called an atom, to the selected non-zero set and find the best possible representation of \( y \) using selected atoms, i.e. orthogonal projection. When the signal of interest has a sparse and positive representation, i.e. \( \|x\|_0 \leq k, x \in \mathbb{R}^n \), we like to incorporate the extra information and adapt OMP to the new setting. This adaptation has been proposed in [1] with two modifications: a) only selecting the atoms with positive correlation with the residual of the signal in that iteration, i.e. \( i^* = \text{argmax}_i \langle \psi_i, x \rangle \) where \( \psi_i \) is the orthogonal component of \( y \) to the span of currently selected atoms, and \( \phi_i \) is the \( i \)-th atom, b) using Non-Negative Least Square (NNLS) representation within selected atoms. The latter step is indeed computationally very expensive for large \( k \)'s. As we have a non-negativity constraint on the representation, we may find that the selected atom in each iteration forces the coefficients of some already selected atoms to be zero, which provides less energy reduction in each iteration of algorithm and reduces the efficiency.

II. NON-NEGATIVE OMP

Let \( \Phi_k = \Psi_k R_k \) be the QR factorisation of selected \( k \) atoms of dictionary \( \Phi \), where \( \Psi_k \) is a column orthonormal and \( R_k \) is an upper-triangular matrix. With some abuse of notation, we assume that in iteration \( k \), \( \Phi_k \) is sorted based on the iteration number and \( \phi_{i,k} \) for \( 1 \leq i \leq k \), is the selected atom in \( i \)-th iteration. In each iteration of NNOMP, we choose the atom which is positive and maximises \( \Phi^T_i y \). In the first iteration, we do not need any orthogonalisation and we have \( \phi_1 = \psi_1 \) and \( R = [1] \). In the \( 1 \leq k \)th iteration, let the best approximation of \( y \), using \( \Phi_k \), be \( \sum_{i=1}^{k} \alpha_i \phi_i = \sum_{i=1}^{k} z_i \psi_i \). In the \( k+1 \) iteration, we have,

\[
\sum_{i=1}^{k+1} z_i \psi_i = \sum_{i=1}^{k} z_i \psi_i + z_{k+1} \psi_{k+1} = \sum_{i=1}^{k} x_i \phi_i + z_{k+1} \psi_{k+1}.
\]

As \( z_{k+1} \psi_{k+1} \) lives in the span of non-redundant set \( \{ \phi_j \}_{j \in [1,k+1]} \), \( z_{k+1} = \sum_{j=1}^{k+1} \gamma_j \phi_j \) for some unique \( \gamma_j \)'s. We can then have,

\[
\sum_{i=1}^{k+1} z_i \psi_i = \sum_{i=1}^{k} x_i \phi_i + \sum_{j=1}^{k+1} z_{k+1} \gamma_j \phi_j = \sum_{i=1}^{k} (x_i + z_{k+1} \gamma_i) \phi_i + z_{k+1} \phi_{k+1}.
\]

As \( z_{k+1} \phi_{k+1} \) is always positive, we only need to assure that \( x_i + z_{k+1} \gamma_i \geq 0 \) or

\[
z_k \leq \min_{1 \leq i \leq n, \gamma_i > 0} \frac{x_i}{\gamma_i}.
\]

In the fast implementation of OMP using QR factorisation [2], we only need to update \( z_i \) at each iteration. To assure that \( x_i \)'s are all non-negative, \( x_i \)'s should comply the condition of (1). We then choose the atom that the corresponding \( z_{k+1} \), or shrunk by upper-bound of (1), has the largest value. It is worth mentioning that \( R_k, R_{k+1} \) and \( \gamma \) can be calculated very efficiently, using iterative reconstruction techniques.

III. SIMULATIONS

To demonstrate the performance, we randomly generated \( \Phi \) with 256 atoms and 64 or 128 rows using \( i.i.d. \) Gaussian noise, followed by column normalisation. With different sparsity and using proposed NNOMP and canonical NNOMP, we repeated the simulations 100 times. The exact recovery and computational time is shown in Figure 1. While the exact recovery is very similar, proposed method is significantly faster for large \( k \)'s.

\begin{thebibliography}{10}
  \bibitem{1} A. Bruckstein, M. Elad, and M. Zibulevsky, \textquotedblleft Sparse non-negative solution of a linear system of equations is unique,	extquotedblright in ISCCSP, 2008, pp. 762–767.
  \bibitem{2} T. Blumensath and M. Davies, \textquotedblleft On the difference between orthogonal matching pursuit and orthogonal least squares,	extquotedblright Tech. Rep., 2007.
\end{thebibliography}
On functional regression-based emulators for faster Bayesian inference from computational simulations

Ewan Cameron
Department of Zoology, University of Oxford, Oxford, United Kingdom OX1 3PS.

Abstract—Gaussian process-based emulators have lately become a popular tool for facilitating Bayesian inference over the parameters of computationally-expensive simulations when both input and output are finite-dimensional vectors. However, in the case that the input and/or output are functional data an alternative framework is required. To this end I will describe recent work towards the use of functional regression from a pre-compiled library of possibly noisy simulations, motivated by the challenge of fitting “individual simulation” transmission models of *Plasmodium falciparum* malaria infections to age-structured incidence counts across multiple sites. A novel solution to the problem of metric distance optimisation (in the input space) is presented, and its effectiveness illustrated through numerical examples, using ideas from the literature on functional uniform priors.

I. INTRODUCTION

The runtimes required for simulation from advanced computational models, such as arise frequently in cosmology, weather forecasting, & epidemiology, are often incompatible with the numbers of likelihood evaluations demanded by Monte Carlo methods for posterior exploration. One approach to facilitate inference under such conditions is to train an emulator of the simulation code as an approximate, but fast, substitute. To this end, Gaussian process-based methods, which smooth and/or interpolate over a library of reference simulations generated under a range of input parameters, are a popular choice provided both inputs and outputs are vector-valued data (cf. [1], [2], [3]). When the latter take the form of functional data, however, alternative methods must be devised (e.g. [4]).

In this talk I will describe one such method for model emulation of functional data based on the technique of functional regression [5], [6]. This work was motivated by the challenge of inferring the prevalence–incidence relationship for *Plasmodium falciparum* malaria using age-structured incidence data from multiple sites and a suite of computationally-expensive “individual simulation” codes [7]. Our novel solution to the problem of metric distance optimisation in the input space of this problem—the product of an infinite-dimensional function space and a finite-dimensional vector space—is based on ideas from the clinical trial design literature on functional uniform priors [8]. The potential of this emulation approach for astronomical/signal processing applications outside the epidemiological context will also be discussed.

REFERENCES
Why CLEAN when you can PURIFY? A new approach for next-generation radio-interferometric imaging

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

† Mullard Space Science Laboratory, University College London (UCL), Holmbury St Mary, Surrey RH5 6NT, UK.
‡ Institute of Sensors, Signals, and Systems, Heriot-Watt University, Edinburgh EH14 4AS, UK.

Abstract—In recent works, sparse models and convex optimization techniques have been applied to radio-interferometric (RI) imaging showing the potential to outperform state-of-the-art imaging algorithms in the field. In this talk, I will review our latest contributions in RI imaging, which leverage the versatility of convex optimization to both handle realistic continuous visibilities and offer a highly parallelizable structure paving the way to high-dimensional data scalability. Firstly, I will review our recently proposed average sparsity approach, SARA, which relies on the observation that natural images exhibit strong average sparsity over multiple coherent bases. Secondly, I will discuss efficient implementations of SARA, and sparse regularization problems in general, for large-scale imaging problems in a new toolbox dubbed PURIFY.

The advent of next-generation radio telescopes, such as the new LOw Frequency ARray (LOFAR), the recently upgraded Karl G. Jansky Very Large Array (VLA) and the future Square Kilometer Array (SKA), has posed several challenges for image reconstruction and the design of data processing systems [1]. The new telescopes will achieve much higher dynamic range than current instruments at a higher angular resolution. Also, these telescopes will acquire a massive amount of data, thus posing large-scale inverse problems in the perspective of image reconstruction. These challenges have triggered an intense research in the community to reformulate imaging and calibration techniques for radio interferometry (RI).

The RI measurement equation can be discretized as
\[ y = \Phi x + n, \]
where \( y \in \mathbb{C}^M \) denotes the vector of measured visibilities, \( \Phi \in \mathbb{C}^{M \times N} \) is a discretization of the measurement operator and \( n \in \mathbb{C}^M \) represents the observation noise. In [2] we propose an imaging algorithm dubbed sparsity averaging reweighted analysis (SARA) based on average sparsity over multiple bases, showing superior reconstruction qualities relative to state-of-the-art imaging methods in the field. A sparsity dictionary composed of a concatenation of \( q \) coherent bases, \( \Psi = [\Psi_1, \Psi_2, \ldots, \Psi_q] \), is used and average sparsity is promoted through the minimization of an analysis \( \ell_0 \) prior, \( \| \Psi^\dagger x \|_0 \), where \( \Psi^\dagger \) denotes the adjoint operator of \( \Psi \) [3].

SARA adopts a reweighted \( \ell_1 \) minimization scheme to promote average sparsity through the prior \( \| \Psi^\dagger x \|_0 \). The algorithm replaces the \( \ell_0 \) norm by a weighted \( \ell_1 \) norm and solves a sequence of weighted \( \ell_1 \) problems where the weights are essentially the inverse of the values of the solution of the previous problem [2], [3]. The weighted \( \ell_1 \) problem is defined as:
\[
\min_{x \in \mathbb{R}^N} \| \Psi^\dagger x \|_1 \quad \text{subject to} \quad \| y - \Phi x \|_2 \leq \epsilon, \tag{1}
\]
where \( W \in \mathbb{R}^{D \times D} \) denotes the diagonal matrix with positive weights, \( \mathbb{R}_+^N \) denotes the positive orthant in \( \mathbb{R}^N \) and \( \epsilon \) is an upper bound on the \( \ell_2 \) norm of the noise, which can be accurately estimated. Hence, we focus our attention on solving problem (1) efficiently, especially for large-scale data problems, i.e. when the number of visibilities is very large (\( M \gg N \)). In this case, we propose to split the data vector \( y \) and the measurement operator into \( R \) blocks in the following manner:
\[
y = [y_1^T, \ldots, y_R^T]^T \quad \text{and} \quad \Phi = [\Phi_1^T, \ldots, \Phi_R^T]^T, \tag{2}
\]
where each \( y_i \) is modelled as \( y_i = \Phi_i x + n_i \) and \( n_i \) denotes the noise vector. With this partition the optimization problem in (1) can be reformulated as
\[
\min_{x \in \mathbb{R}^N} \| \Psi^\dagger x \|_1 \quad \text{subject to} \quad \| y_i - \Phi_i x \|_2 \leq \epsilon_i, \quad i = 1, \ldots, R, \quad (3)
\]
where each \( \epsilon_i \) is an appropriate bound for the \( \ell_2 \) norm of \( n_i \).

In [4] we propose a general algorithmic framework based on the simultaneous-direction method of multipliers (SDMM) to solve (3). The proposed framework offers a parallel implementation structure that decomposes the original problem into several small simple problems, hence allowing implementation in multicore architectures or in computer clusters, or on graphics processing units. These implementations provide both flexibility in memory requirements and a significant gain in terms of speed, thus enabling scalability to large-scale problems. A beta version of an SDMM-based imaging software written in C and dubbed PURIFY was released that handles various sparsity priors, including SARA, thus providing a new powerful framework for RI imaging (toolbox available at http://basp-group.github.io/purify/). Even though this beta version of PURIFY is not parallelized yet, we discuss in detail the extraordinary parallel and distributed optimization potential of SDMM to be exploited in future versions. We also discuss other possible research avenues for big data scalability. One possibility is to incorporate ideas from stochastic gradient methods into proximal splitting and augmented Lagrangian methods. The key idea is to use only one data block \( y_i \), or a subset of blocks, at each iteration of the reconstruction. By doing so, the computational complexity per iteration will be reduced. Thus, the total processing time of the algorithm will also be reduced if the convergence rate of the original problem is preserved. See for example [5] and references therein for first theoretical results.

REFERENCES

MORESANE: a sparse deconvolution algorithm for radio interferometric imaging

Arwa Dabbech*, Chiara Ferrari*, David Mary*, Eric Slezk*, Oleg Smirnov†‡, and Jonathan Simon Kenyon†

* Laboratoire Lagrange, UMR7293, Université Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d’Azur, Nice, France.
† Centre for Radio Astronomy Techniques & Technologies (RATT), Department of Physics and Electronics, Rhodes University, Grahamstown, South Africa.
‡ SKA South Africa, Pinelands, South Africa.

Abstract—In this paper, we present a new deconvolution algorithm named MORESANE for radio astronomical imaging. MORESANE is a greedy algorithm that combines complementary types of sparse recovery methods in order to reconstruct the most appropriate sky model from the observed radio data. A synthesis approach is used for the reconstruction of the model where the synthesis atoms representing the unknown sources are learned using analysis priors. Application of this new deconvolution method to fully realistic simulations of radio data along with comparisons with standard algorithms indicate that MORESANE is very promising in the reconstruction of radio images.

The deconvolution of the radio interferometric images is a research field where recent innovations have built extensively on sparse representations. Sparse image models are formulated in the literature through redundant dictionaries (such as wavelets, curvelets, etc.), with respect to which sparsity can be promoted either in analysis, or in synthesis. We propose a new approach that uses sparsity promoting restoration models based on highly redundant, shift invariant dictionaries (The Isotropic Undecimated Transform dictionaries). On one hand, using sparse synthesis priors, the image to be reconstructed is modeled as a sum of few P elementary objects which, as opposed to classical sparsity-based priors, are unknown. On the other hand, these objects are iteratively estimated and deconvolved through analysis-based priors using the Isotropic Undecimated Wavelet Transform (IUWT) dictionaries (MORESANE: Dabbech et al., submitted; [1]).

We applied MORESANE to realistic simulations of observations using the measurements set of MeerKAT, a precursor to the Square Kilometre Array. We compared the results with those obtained by standard deconvolution algorithms such as MULTISCALE CLEAN [2] and a compressed sensing approach IUWT-based CS [3]. MORESANE provides a better approximation of the input model where the morphologies of the different objects are estimated in a more accurate way. It is also more robust to false detections and most of the objects in the model correspond to genuine sources when checked against the true image.

A complete self-contained implementation of MORESANE in Python is now available (https://github.com/ratt-ru/PyMORESANE) which includes GPU acceleration and can be used on large datasets. Further developments are planned, including taking into account the variations of the PSF across the field-of-view of the instrument, reconstructing spectral images and testing performances on poorly calibrated data.

REFERENCES

LOFAR and SKA Sparse Image Reconstruction

H. Garsden∗, J. N. Girard ∗, J. L. Starck ∗, S. Corbel ∗, C. Tasse†, A. Woiselle§, J. McKeen¶

∗ Laboratoire AIM, Université Paris Diderot, Paris 7/CNRS/CEA-Saclay, DSM/IRFU/SAp, 91191 Gif-sur-Yvette, France
† Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa
‡ SKA South Africa, 3rd Floor, The Park, Park Road, Pinelands, 7405, South Africa
§ Sagem (Safran), 27 rue Leblanc, 75512 Paris Cedex 15, France
¶ Netherlands Institute for Radio Astronomy (ASTRON), Postbus 2, 7990 AA Dwingeloo, The Netherlands

Abstract—Recent papers have established a clear link between the discrete nature of radio interferometry measurement and “compressed sensing” theory, which supports sparse recovery methods to reconstruct an image from the measured visibilities. We have implemented a sparse recovery method for LOFAR (and extending it to SKA) allowing us to compare the reconstructed images from both SKA simulations and LOFAR simulated and real data to results obtained with classical methods such as (Cotton-Schwab) CLEAN or MS-CLEAN. While being compatible with the Radio Interferometer Measurement Equation (RI ME) framework, we show that i) sparse recovery performs as well as CLEAN in recovering the flux of point sources, ii) performs much better on extended objects, and iii) provides a solution with an effective angular resolution at least twice times better than the CLEAN map.

I. INTRODUCTION

Recent years have seen the development and planning of very large radio interferometers such as the LOw Frequency ARray (LOFAR) [1] in Europe, and the Square Kilometre Array (SKA) [2] in Australia and South Africa (through its various precursors and pathfinders). These instruments come after a long time of preliminary developments and incarnate the new “golden” age for radio astronomy in terms of science, methods and instrumentation. Since the beginning of radio interferometry, various imaging methods have been designed to fit the requirements of different kinds of (extended) radio objects. The availability of high-performance computing, and the need for efficient, fast and accurate imaging for new wide-field interferometers, has motivated the implementation of new imaging algorithms. Given the recording time/frequency resolutions, the integration time, and the diversity of baselines of wide-field interferometers, large amounts of data storage are required to save the telescope data. These data must then be transformed into a scientifically exploitable form (typically into image cubes). Because of the nature and the dimensions of the LOFAR and SKA arrays, Direction-Dependent Effects (DDE) [3] occur during the span of an observation, and add up to the usual other effects intervening in classical radio interferometers. These effects require a “direction-dependent” calibration before imaging. In particular, the classical compact planar array & small FoV assumptions are no longer valid, especially for wide-field instruments.

The problem can be generalized and expressed in the Radio Interferometry Measurement Equation framework (e.g. [4], [5], [6] and following papers) [7]. The calibration problem therefore manifests as an inverse problem that should be solved in order to determine independently all the parameters and coefficients that describe each observation dataset.

For many years, deconvolution has been achieved through the CLEAN algorithm [7] and its variants [8], [9]. But with the advent of wide field of view (FOV) digital instruments, the need for advanced imaging methods, taking the DDE into account, was more and more manifest.

Among the recent developments of data processing/reconstruction algorithms, the “discovery” of Compressed Sensing (CS) [10], [17] (a new non-linear sampling/compression theory) has led to new approaches to solve these problems. It has been proposed for radio interferometry (e.g. [11], [12], [13], [14], [15]) as the latter constitutes a relevant practical case due to the sparse nature of the interferometric sampling. The implementation of sparse radio image reconstruction methods is expected to produce better results on large extended objects with high angular resolution, compared to other classical deconvolution methods. In our work [16], we presented the first implementation of sparse reconstruction in the standard LOFAR imager $\text{AI}w\text{m}a\text{g}e\text{r}$ and its application to radio interferometric data compared to CLEAN-based deconvolution methods.

II. SPARSE IMAGE RECONSTRUCTION

A. LOFAR sparse imaging

We will present the scientific validation of our code in terms of photometry (by inspecting the reconstructed flux density of known point sources) and angular resolution (between two point sources separated by variable angular distance). We reconstructed extended radio emissions (along with CLEAN and the multi-scale CLEAN [18]) of a simulated dataset of W50 and of a real LOFAR observation of Cygnus A.

B. SKA weak lensing studies

We performed a comparative study of CLEAN and CS in the scope of radio weak lensing measurements by SKA. We based it on the simulations of [19] to compare the impact of the sparse reconstruction (compared to that of CLEAN) for different interferometer distribution and different time/baselines integrations. We performed a statistical analysis of the quality of the reconstruction by using an input set of simulated fields containing radio galaxies.

REFERENCES

Challenges in Radio Weak Gravitational Lensing

Ian Harrison*, Prina Patel†,

* Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL UK
† Department of Physics, University of Western Cape, Cape Town 7535, South Africa

Abstract—Weak gravitational lensing experiments with the Square Kilometre Array (SKA) have the potential to overcome systematics which limit similar optical surveys and be a powerful probe of the physical nature of Dark Energy. However, a number of significant challenges need to be met in extracting from radio interferometer data the exquisitely accurate measurement of galaxy shapes necessary for weak lensing. Here we describe the organisation of radioGREAT, a community-wide effort to tackle this challenge.

I. INTRODUCTION

As light from distant galaxies traverses the Universe it is lensed by the gravitational potentials of all the intervening matter, both luminous and dark, causing distortions in the images of those galaxies we view here on Earth. By using this lensing effect to track the evolution of structures in matter over cosmic time we can also track the evolution of Dark Energy and learn about its physical nature. However, weak lensing experiments are difficult to perform. ‘Weak’ lensing refers to the regime in which the intrinsic shapes of galaxies are dominant, with only around 1% of a given objects ellipticity being caused by the lensing distortion. It is therefore necessary to cross-correlate the shapes of a large number density of galaxies in order to measure the coherent distortions in shape caused by the intervening mass along a particular line of sight in the sky. In order to obtain the precision to be able to select between different physical models of Dark Energy, this galaxy shape measurement needs to be performed with extraordinary accuracy (down to an error of less the 0.01% in ellipticity). In the past 10 years the weak lensing scientific community have made large strides in being able to achieve these levels of accuracy, developing shape measurement algorithms of the necessary sophistication to ensure current and near future optical weak lensing experiments will have errors dominated by statistical rather than systematic uncertainties.

II. WEAK LENSING WITH THE SKA

In addition to these optical and Near Infra-Red (NIR) experiments, the Square Kilometre Array (SKA) will be able to achieve the high number densities of galaxies necessary for weak lensing, allowing us to perform experiments at radio frequencies which will be competitive with premier optical/NIR ones [1]. In addition, the radio represents a number of unique advantages over the optical/NIR for weak lensing:

- Radio interferometers have precisely known, stable and deterministic Point Spread Functions (PSFs), minimising systematic uncertainties on ellipticities induced by the telescope.
- Populations of radio star-forming galaxies are expected to have higher median redshifts, providing a longer lever-arm for measuring the equation of state of Dark Energy.
- Radio observations allow extra information to be included such as polarisation [2] and rotational velocities [3], overcoming intrinsic alignment and shape noise systematics which can limit the power of optical/NIR experiments.

It is also expected that many systematics will be uncorrelated between radio and optical/NIR experiments, allowing combination of data from both wavebands to outperform what can be done with both alone. Weak lensing science has been given a high priority by the SKA cosmology Science Working Group.

In the period before the SKA, a number of pathfinding experiments such as superCLASS, CHILES and VLASS will probe the faint radio source populations we expect to use for weak lensing and provide real data on which to test shape measurement techniques of the high accuracy necessary.

III. RADIOGREAT

In developing shape measurement techniques for optical/NIR data, which arrives in the image plane onto a CCD detector, the Shear Testing Programme (STEP) and Gravitational Lensing Accuracy Testing (GREAT) initiatives challenged participants to blindly measure shapes of galaxies in increasingly realistic simulated images, with the true answer known only to the challenge organisers. These competitions have provided invaluable information in identifying important questions for shape measurements and driving development of a variety of novel and powerful algorithms. However, more needs to be learned in order to achieve the full potential of weak lensing with the SKA, where the data arrives not in the image plane but in the near-Fourier ‘UV’ or ‘visibility’ plane. Existing imaging algorithms for turning these visibility data into images are highly non-linear and were not designed with precision morphology studies in mind. Measurement of galaxy shapes directly from visibilities has been demonstrated before [4], but the accuracy of this and its scalability to the extremely large data volumes which will emanate from the SKA is unclear.

In light of these issues, we propose the radio Gravitational Lensing Accuracy Test (radioGREAT) challenge. The challenge will produce a large number of realisations of simulated radio observations of sky models representing galaxy populations relevant to weak lensing. These simulated observations will then be made available to the community to test and compare imaging and shape measurement algorithms. We aim to make this challenge accessible to as many people as possible, with the awareness previous GREAT challenges have benefited from the involvement of participants from outside the fields of weak lensing and even physics.

REFERENCES

Astrostatistics and Brain Imaging

Alan Heavens*, Raul Jimenez†, and Ben Panter‡

* ICIC, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ
† ICC, University of Barcelona, Martí i Franques 1, Barcelona 08028, Spain
‡ Blackford Analysis, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ

Abstract—Techniques in astrostatistics developed for analysing the spectra of a million galaxies very fast, employing a novel data compression algorithm (MOPED), have been used to align 3D images of brains. MOPED stands for Massive Optimised Parameter Estimation and Data compression, and works by finding linear combinations of the data that maximise the information of each parameter in turn. The algorithm can give astonishing levels of data compression (by a factor $N/M$, where $M$ is the number of parameters and $N$ is the number of data points), without increasing error bars. The application to medical images has led to the formation of Blackford Analysis, a software company that provides very rapid alignment of medical images, saving radiologist time and allowing them to concentrate on diagnosis.

I. INTRODUCTION

Bayesian statistical analysis is very well-adapted for inference of parameters when data are analysed within the framework of a theoretical model. Astronomical datasets can be large in size, and the time for analysis can represent a significant constraint. As a result there is strong motivation in astrostatistics to find rapid ways to analyse large datasets to obtain posterior probabilities for parameter values. One possible route is to use data compression, so the size of the dataset to be analysed is reduced, giving speed advantages. If the compressed data can be made uncorrelated, then the computation of the posterior may be very fast indeed, allowing very efficient searching of the parameter space. Data compression itself is of course not difficult to achieve, but the key is to do it in such a way to avoid compromising the size of the Bayesian confidence intervals. The MOPED algorithm does this, by employing a linear compression of the data. It achieves a potentially massive reduction in data size to the minimum size possible - one datum per parameter, with little or no increase (in ideal cases) in the size of the confidence regions. Use is made of the Fisher matrix formalism, to quantify the expected size of the confidence intervals under the assumption that the likelihood surface is a multivariate gaussian. Parameters are treated in turn, with a set of weights for the data, which are optimised to maximise the size of the relevant Fisher matrix element. With an additional constraint that the compressed datum is uncorrelated with the previous compressed data, the solution is found by solving a generalised eigenvalue problem. Remarkably, if the covariance matrix of the full dataset is independent of the parameters in the model, the Fisher matrix of the compressed data can be identical to that of the original dataset. To the extent that the Fisher matrix determines the errors (the likelihood surface may not be gaussian), this represents no loss of information on the model parameters. The weights are somewhat informative - they are high where the data are sensitive to the parameter, or where they have low noise, but the requirement for the compressed data to be uncorrelated complicates the issue to some degree. Since the sensitivity of the data to changes in the parameters depends in general on the value of the parameter, a fiducial parameter set needs to be chosen to compute the weights, and an additional requirement for the Fisher matrix to be unchanged is that the fiducial parameters must be the true parameters. Since these are what we are trying to find, it is not generally satisfied, but it turns out in practice that a poor choice of fiducial model makes almost no difference to the size of the confidence regions. In any case, a second analysis iterating the fiducial model could always be done if needed. The original application [1] was to galaxy spectra, where the parameters in the model were principally the formation history of the stars in the galaxy (parameterised in $\sim 10$ time periods), and the data consisted of $\sim 2000$ flux points per galaxy. Compression was a factor of 100 or so. $\sim 10^6$ galaxies from the Sloan Digital Sky Survey were analysed and the star formation history of the Universe determined [2]. The problem of the alignment of medical images can be formulated as a parameter inference problem as follows. The data are a set of (typically) $\sim 10^6$ voxels from 3D scans, and in the simplest case of brain scans, the model is that two sets of images taken at different times are identical apart from rigid-body transformations from one to the other. This is a 6-parameter problem (3 rotations and 3 translations), and these are the numbers that need to be determined to align the images to detect any changes over time. The changes are, one hopes, small, and do not affect the registration. These ideas led to the formation of Blackford Analysis as a spin-out company that has developed software in a number of areas including very rapid image registration. The head is relatively straightforward, as it is intrinsically rigid (although instrumental effects may introduce other distortions), but the software can deal with the non-rigid transformations that apply to other parts of the body, such as the lungs seen being aligned in the figure. A further complication which the software can handle is that the images may not even be taken in the same modality (PET, CT, MRI etc).

REFERENCES

Neural networks and accelerated Bayesian inference

Michael Hobson∗, Philip Graff†, Farhan Feroz∗ and Anthony Lasenby∗

∗ Astrophysics Group, Cavendish Laboratory, J.J. Thomson Avenue, Cambridge CB3 0HE, UK.
† NASA Goddard Space Flight Centre, Gravitational Astrophysics Laboratory, 8800 Greenbelt Rd., Greenbelt, MD 20771, USA

Abstract—We present a generic neural network training algorithm, called SKYNET, and describe its use in the accelerated Bayesian inference algorithm, BAMBI. SKYNET combines multiple techniques to create an efficient and robust machine-learning tool that is able to train large and deep feed-forward neural networks for use in a wide range of learning applications, such as regression, classification, density estimation, clustering and dimensionality reduction. The blind accelerated multimodal Bayesian inference (BAMBI) algorithm uses SKYNET to learn the likelihood function in a Bayesian inference problem and combines it with the nested sampler MULTINEST to explore the posterior distribution and evaluate the evidence, even for multimodal problems. In this way, such analyses are vastly accelerated, often by several orders of magnitude.

In astrophysics and cosmology, one is faced with analysing large, complicated and multidimensional data sets. Such analyses typically include tasks such as: data description and interpretation, pattern recognition, prediction, classification, compression, inference and many more. One way of performing such tasks is through the use of machine-learning methods (see, e.g., [1], [2] and [3]).

Machine-learning can be divided into two broad categories: supervised learning and unsupervised learning. In supervised learning, the goal is to infer a function from labeled training data, which consist of a set of training examples. Each example has known ‘input’ quantities whose values are to be used to predict the values of the ‘outputs’. Thus, the function to be inferred is the mapping from input to outputs. Once learned, this mapping can be applied to datasets for which the values of the outputs are not known. Supervised learning includes both classification and regression. In unsupervised learning, the data have no labels. More precisely, the quantities associated with each data item are not divided into inputs and outputs.

An intuitive and well-established approach to machine learning, both supervised and unsupervised, is based on artificial neural networks (NNs). These consist of a group of interconnected nodes, each of which processes information that it receives and then passes this product on to other nodes via weighted connections. In this way, NNs constitute a non-linear statistical data modeling tool, which may be used to model complex relationships between a set of inputs and outputs. Many machine-learning applications can be performed using only feed-forward NNs: an input layer of nodes passes information to an output layer via zero, one, or many ‘hidden’ layers in between. Moreover, a universal approximation theorem [4] assures us that we can accurately and precisely approximate any reasonable mapping with a NN of a given form. A useful introduction to NNs can be found in [1].

In astronomy, feed-forward NNs have been applied to various machine-learning problems for over 20 years (see, e.g., [3], [5]). Nonetheless, their more widespread use in astronomy has been limited by the difficulty associated with standard techniques, such as backpropagation, in training networks having many nodes and/or numerous hidden layers (i.e. ‘large’ and/or ‘deep’ networks), which are necessary to model the complicated mappings between the numerous inputs and outputs. In this presentation, we therefore introduce SKYNET [6], an efficient and robust neural network training algorithm that is capable of training such networks.

An important recent application of regression supervised learning in astrophysics and cosmology is the acceleration of the Bayesian analysis (both parameter estimation and model selection) of large data sets in the context of complicated models. At each point in parameter space, Bayesian methods require the evaluation of a ‘likelihood’ function describing the probability of obtaining the data for a given set of model parameters. For some problems each such function evaluation may be computationally expensive. Substantial gains in performance can thus be achieved if one is able to speed up the evaluation of the likelihood, and a NN is ideally suited for this task. The blind accelerated multimodal Bayesian inference (BAMBI) algorithm [7] uses SKYNET for training such NNs, and combines them with a nested sampling [8] approach that efficiently calculates the Bayesian evidence (also referred to as the marginal likelihood) for model selection and produces samples from the posterior distribution for parameter estimation. It particular, BAMBI employs the MULTINEST algorithm [9], [10], [11], which is a generic implementation of nested sampling, extended to handle multimodal and degenerate distributions, and is fully parallelised.

This presentation will demonstrate the application of SKYNET to a number of toy problems, and to astronomical problems focusing on the recovery of structure from blurred and noisy images, the identification of gamma-ray bursters, and the compression and denoising of galaxy images. I will also discuss the use of SKYNET within the BAMBI algorithm to accelerate Bayesian inference in cosmology.

REFERENCES
Abstract—The speed of the FFT renders it a standard tool in imaging for radio telescope arrays. The nuisance is all Fourier samples need to be gridded: a form of interpolation onto a discrete grid and an enormous bottleneck. This will be a particularly acute problem for the Square Kilometre Array (SKA), given its unprecedented data generation. On the other hand, standard practise reduces antenna data through beamforming at station level (collections of antennas) prior to central processing. We propose and test a beamforming technique which, at stations, strives to position virtually the beams on the discrete grid, so as to attenuate the necessity for subsequent gridding. By transforming the problem of gridding into one of approximating complex exponentials as the product of beam-shapes, we are able to reduce imaging complexity. We present initial results showing its effectiveness.


text

Extended Abstract

Under certain assumptions, the baselines between radio astronomy antennas approximate samples of the Fourier transform of the sky image. The number of visibility measurements is very large, and the FFT is used to speed up calculations. Because baselines do not necessarily lie on a grid, the Fourier samples also do not. Thus visibility data is first massaged to lie on one by the gridding process [1]. This involves convolving the visibility function with a Gaussian function. Then the visibility domain signal is sampled densely. This amounts to replication of the sky in the image domain. Because the region of interest is limited, the sampling density can be chosen so as to avoid aliasing [2].

On the other hand, data is reduced at stations by beamforming before being sent for central image creation. We propose to tailor this beamforming to grid the visibility data as measurements are acquired. This would reduce the imaging complexity enormously, as gridding is currently the biggest bottleneck in the imaging chain [3].

The samples of the sky image Fourier transform with spacing \( \Delta u \) and \( \Delta v \) in the frequency domain can be expressed as

\[
V[k, n] = \int \int I(l, m)e^{-j2\pi(k\Delta u+l\Delta v,m)}\,dl\,dm,
\]

for integers \( k \) and \( n \). The visibility measurement obtained from two beams, say \( p \) and \( q \), is

\[
V(u_{p,q}, v_{p,q}) = \int \int I(l, m)c_p(l, m)c_q^\ast(l, m)e^{-j2\pi(u_{p,q}l+v_{p,q}m)}\,dl\,dm,
\]

where \( c_p(l, m) \) and \( c_q(l, m) \) are the beam shapes of stations \( p \) and \( q \) respectively. If the product of the beam shapes and the complex exponential resulting from the baseline between the stations approximates the complex exponential in (1), that is, if

\[
e^{-j2\pi(k\Delta u+l\Delta v,m)} \approx c_p(l, m)c_q^\ast(l, m)e^{-j2\pi(u_{p,q}l+v_{p,q}m)},
\]

we have the samples of the Fourier transform on the grid. Thus, we can transform the problem of gridding the visibilities to one of approximating complex exponentials as the product of beam-shapes. Expanding the expression for the baseline between the antennas, the right hand side of (2) can be written as

\[
\left(c_p(l, m)e^{-j2\pi(xpl+ylm)}\right)\left(c_q(l, m)e^{-j2\pi(xql+yqm)}\right)\ast,
\]

where \((x_p, y_p)\) and \((x_q, y_q)\) are the coordinates of stations \( p \) and \( q \) respectively. It can be seen that, if the individual parts in the product are on a grid, then the visibility measurement between any two stations are again on a grid. This strategy corresponds to virtually aligning the stations on the grid.

Beamforming weights can be chosen to approximate the desired complex exponential beam shape. We can express the desired condition (2) as finding \( w \) such that \( w^\ast A \approx c \), where \( A \) is a matrix of the antenna samples and \( c \) is a vector of the complex exponentials from the desired grid positions.

To prevent amplifying noise in signal output, we constrain the norm of the beamforming weights, resulting in optimisation:

\[
\arg\min_w \|A^\ast w - c\|^2.
\]

This is convex with a globally optimal solution efficiently attainable. In general, how well the complex exponentials can be approximated depends on the number of antennas per station.

Fig. 1 illustrates the scheme. The reference is the dirty image obtained from a set of stations positioned on the grid points that we approximate. As can be seen, the dirty images align almost perfectly.

References

Accelerated facet-based widefield imaging

Benjamin Hugo∗, Associate Professor James Gain ∗and Professor Oleg Smirnov†‡
∗ Department of Computer Science, University of Cape Town, South Africa
† Square Kilometre Array Africa, Cape Town, South Africa
‡ Radio Astronomy Techniques & Technologies group, Department of Physics & Electronics, Rhodes University, South Africa

Abstract—Imaging in radio astronomy entails the Fourier inversion of the relation between the sampled spatial coherence of an electromagnetic field and the intensity of its emitting source. Due to significant computational savings this inversion is normally performed by performing an inverse Fast Fourier Transform. Unfortunately the resulting planar approximation to the sky is only valid over small regions. When imaging over wider fields of view, and in particular using non-East-West antenna arrays, significant distortions are introduced on the computed image. Targeted faceting removes these distortions by computing small portions of the sky-image at a time. Our work focuses on accelerating this faceted imaging approach using today’s highly parallel General Purpose Graphics Processing Unit co-processors.

I. INTRODUCTION

Over recent years the trend in radio astronomy has predominantly been towards the construction of telescopes with ever-increasing sensitivity and resolving capacity, and have taken the form of many-element arrays. Through the process of radio interferometry a telescope with comparable resolution to that of a single aperture telescope which completely encompasses the entire array can be constructed. This avoids encountering the cost (and infeasibility) of building and maintaining a very large structure.

A radio interferometer measures the coherence of an electrical field at two distinct points, by taking short-time averages between the output of pairs of antennae. Under some simplifying assumptions the following well known equation relates the measured coherence and the projected intensities of the sources on a (unit) celestial sphere (a brief mathematical treatment is given by Clark [1]):

\[ V(u, v, w) = \int_S I(l, m, n)e^{2\pi i \lambda^{-1} \hat{b} \cdot \hat{s}} dS \]  

(1)

Here \( V \) is the measured coherence, \( I \) is the sky intensity, \( \hat{b} \) has components \( u, v, w \) and is the vector between antennae, \( \hat{s} \) is in the direction of the pointing centre of the telescope and \( \hat{s} \) is the vector between some source and the observing telescope. By observing over an extended period of time it is possible to obtain samples which cover most of the local coordinate frame (at least at the plane \( w = 0 \)), since the rotation of the Earth will rotate the interferometer elements in an elliptical path through the local frame. After resampling onto a regular grid this relation may be inverted using a two dimensional inverse Fast Fourier Transform. This approach is valid only when \( \hat{s} \) is relatively close to \( \hat{s}_0 \) (ie. the planar image is a close approximation to the celestial sphere) or when the interferometer remains coplanar (ie. \( w = 0 \)) throughout the observation. This is generally true of arrays with non-East-West components [2], [3].

When larger images are produced the error associated with taking a planar approximation will show up as smearing effects around sources away from the centre of the produced image. These effects are significantly worsened by the decorrelation introduced by non-coplanar interferometer components [4], [5].

II. WIDEFIELD CORRECTIONS

Over the past two decades widefield imaging have been the topic of much discussion. Generally speaking there are three primary algorithms for correcting widefield distortions: faceting [4], snapshot imaging [6] and w-projection [5].

Our work is only concerned with performing targeted faceting, since recent estimations have projected that images up to \( 10^{10} \) pixels will be produced using the Square Kilometre Array [6]. In a faceting approach (of any form) smaller subsections of the final image is produced by introducing an artificial delay to the coherence samples. In essence this “electrically steers” the telescope to the centre of each subimage [4], [7].

The approaches of snapshotting (short time observations over which all interferometer elements remain roughly coplanar) and w-projection/w-stacking (applying a w-dependent corrective term in either the Fourier or image space respectively) computes the entire corrected image at a time, and previous widefield acceleration attempts are only useful for creating lower resolution images because of this limitation [8].

An additional advantage of faceting over the other approaches is that one may directly apply corrections for slow-varying directional dependent effects to the coherence samples before resampling and imaging [9].

III. ACCELERATION

Previous literature have shown that GPGPU-accelerated image synthesis using convolutional resampling and employing the Fast Fourier Transform can be more than 100 times faster compared to current non-parallel imaging software [8]. It is planned to expand on this work, by avoiding using the large convolution kernels employed by accelerated w-projection and adding the transformations necessary for faceting as described earlier. Additionally we’re hoping to add support for parallelization across multiple GPUs in order to best exploit the upcoming trend towards multi-GPU co-processors.

REFERENCES

RESOLVE: A new algorithm for aperture synthesis imaging of extended emission

Henrik Junklewitz∗†, Michael R. Bell †, Marco Selig † and Torsten A. Enßlin †

∗Argelander-Institut für Astronomie, Rheinische Friedrich-Wilhelms-Universität Bonn, Regina-Paci-Weg 3, 53113 Bonn, Germany
†Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany

Abstract—We present RESOLVE, a new algorithm for radio aperture synthesis imaging of extended emission in total intensity. The algorithm is derived using Bayesian statistical inference techniques, estimating the surface brightness in the sky assuming a priori log-normal statistics. RESOLVE not only estimates the measured sky brightness in total intensity, but also its spatial correlation structure, which is used to guide the algorithm to an optimal reconstruction of extended and diffuse sources. For a radio interferometer, it succeeds in deconvolving the effects of the instrumental point spread function during this process. RESOLVE also provides a map with an uncertainty estimate of the reconstructed surface brightness. Furthermore, with RESOLVE we introduce a new, optimal visibility weighting scheme that converges to robust weighting. In tests using simulated observations, the algorithm shows improved performance against two standard imaging approaches for extended sources, Multiscale-CLEAN and the Maximum Entropy Method.

I. INTRODUCTION

Aperture synthesis techniques using large interferometers have a long and successful history in radio astronomy. While enabling observers to achieve very high resolutions, data processing is considerably more complicated than with a single dish instrument. A radio interferometer effectively measures the Fourier transformation of the sky brightness [1]. Unfortunately, inverting this relationship to achieve an estimate of the desired source brightness is a non-trivial task since an interferometer only samples a fraction of the Fourier plane, effectively convolving the true image brightness with an observation-dependent point-spread function. A crucial part in data reduction is therefore the imaging, i.e. estimating the true sky brightness distribution from the observed data, and the development of new imaging techniques is still a field of ongoing research.

An important reason is that all widely used imaging algorithms in radio astronomy have a number of drawbacks. The most successful method CLEAN [2] assumes the image to be comprised of uncorrelated point sources and, therefore, is naturally non-optimal for highly resolved, extended and diffuse sources. Some of the newest enhancements of CLEAN try to address this problem using a multiscale approach [3], but it is not clear in general how to properly choose the scales, although the new Adaptive Scale Pixel method [4] is supposed to overcome these problems. MEM, the Maximum Entropy Method [5], is by design prone to oversmoothing the image in cases. Furthermore, for all of these methods, including very recent approaches using wavelets within the framework of Compressed Sensing ([6], [7]) no reliable uncertainty estimates for the image reconstruction are available to date.

A second incentive for new developments of imaging techniques are recent advances in radio astronomical instrumentation. The new generation of radio telescopes, such as the upgraded VLA, LOFAR, the SKA pathfinder missions or ultimately the SKA itself, are opening new horizons in radio astronomy. Their unprecedented capabilities of simultaneous, broadband frequency coverage including previously unexplored wavelength regimes, sensitivity, and wide fields of view will almost certainly advance astrophysical and cosmological sciences. At the same time, new developments in signal processing and data analysis are required to exploit these new capabilities.

In this article, we introduce RESOLVE (Radio Extended SOurces deCONvolution EStimator) ([8], [9]), a novel algorithm for the imaging of diffuse and extended radio sources in total intensity. A new approach to the problem is taken, using Bayesian statistics in the framework of Information Field Theory [10] and based on clearly formulated mathematical principles to overcome some of the mentioned imaging problems.

In short, RESOLVE is designed to fulfill three main requirements: 1) It should be optimal for extended and diffuse radio sources in the sense of minimizing a statistical error norm, 2) It should include reliable uncertainty propagation and provide an error estimate together with an image reconstruction, 3) It should be capable of exploiting and processing modern wideband-data.

The main scientific focus of RESOLVE is by definition on the imaging of astrophysical extended and diffuse radio sources. Typically, among those are galaxy clusters with their weak diffuse halos and strong extended relic structures, lobes of radio galaxies, radio galaxies, supernova remnants, galactic radio halos, and the diffuse radio emission from the Milky Way.

REFERENCES

Abstract—As radio telescope data is now increasingly high-dimensional, data reduction has become essential to reduce computational load while preserving accurate signal reconstruction. Gridding the continuous Fourier visibilities represents the standard approach to dimension reduction in radio interferometric imaging. This abstract describes a novel dimension embedding technique relying on the multiplication of the data vector by a fat matrix whose entries are drawn from an i.i.d. random Gaussian distribution. Preliminary results suggest that this approach may provide significant improvement in the reduction of data size with respect to standard gridding, for a given target imaging quality.

High-dimensional data acquisition from next-generation radio interferometers can be modelled through the measurement equation \( y = \Phi x + n \), where \( y \in \mathbb{C}^M \) represents the vector of continuous Fourier visibilities (measurements) corrupted by additive noise \( n \in \mathbb{C}^M \), \( \Phi \in \mathbb{C}^{M \times N} \) is the measurement operator, and \( x \in \mathbb{C}^N \) is the underlying signal, with \( M \gg N \).

The aim of dimension embedding is reducing data size to minimise computing load, in terms of memory usage and running time of the reconstruction algorithm. Dimension embedding is performed through an embedding matrix \( R \in \mathbb{C}^{M_L \times M} \), \( M_L \ll N \ll M \), leading to an embedded forward model \( y' = \Phi' x + n' \), with \( y' = R y \), \( n' = R n \), and \( \Phi' = R \Phi \). The gist of dimensional embedding is that the full embedded measurement operator \( \Phi' \) is precomputed once and then stored for further use, which avoids calculations involving the large dimension \( M \) during reconstruction, only dealing with the embedded measurement vector of dimension \( M_L \). Once the data size is reduced below the image dimension, reconstruction fits into a Compressed Sensing (CS) problem, where the main issues to address are (i) conditions that the measurement operator needs to satisfy, e.g. the Restricted Isometry Property (RIP), and (ii) reconstruction algorithms including adequate prior signal information.

The following dimension embedding techniques are being explored: (i) Holographic embedding [1] involves a ‘gridding down’ of the measurements from the continuous visibilities. In this case, \( R \) identifies with a gridding operator and \( M_L = N \). Gridding is the standard approach to dimension embedding in radio interferometry; (ii) Random Gaussian embedding involves a matrix \( R \) with zero mean i.i.d. Gaussian random entries, and \( M_L \) can take any arbitrary value, typically \( M_L \ll N \). For random Gaussian embedding, \( \Phi' \) becomes a Gaussian operator, which might actually approach the characteristics of the optimal sensing matrices promoted by the CS theory [2]. This calls for further theoretical study of the RIP satisfied by \( \Phi' \), in comparison to \( \Phi \). Another important advantage of this embedding is that the entries of the embedded noise vector \( n' \) can be shown, on average over realisations of \( R \), to be an i.i.d. Gaussian distribution, even when the original noise vector \( n \) is not. In this context, the log-likelihood for the data takes the standard form of an \( \ell_2 \)-norm term following a \( \chi^2 \) distribution with \( M_L \) degrees of freedom. The noise behaviour is therefore very well controlled analytically after embedding, and natural weighting becomes superfluous.

We provide a preliminary comparison between random Gaussian and holographic embeddings on simulated data using the PURIFY toolkit [3], which resorts to convex optimisation for image reconstruction, more specifically to the recently proposed algorithm SARA [4]. From the above considerations, a bound on the \( \ell_2 \)-norm data term used by SARA can be set analytically for the random Gaussian embedding, while empirical evaluations are necessary for the holographic embedding.

We use an \( N = 128 \times 128 \) model image of the M31 Galaxy (see figure, left plot—shown in log scale), from which \( M = 10N \) continuous visibilities are sampled following a variable density profile with Gaussian shape in the Fourier plane. The input SNR, defined as ISNR \( = 10 \log_{10}(\|\hat{x}\|^2/\|n\|^2) \) with \( \hat{x} = \Phi \hat{x} \) being the clean measurement vector, is set to 30dB. The reconstruction quality is measured in terms of the output SNR defined as OSNR \( = 10 \log_{10}(\|\hat{x}\|^2/\|\hat{x} - \hat{x}\|^2) \), \( \hat{x} \) being the reconstructed signal. Our simulations show that an OSNR of more than 32dB is reached in the absence of embedding. Holographic embedding achieves an OSNR of around 27dB with, by construction, \( M_L = N \), while random Gaussian embedding achieves the same imaging quality from an (approximately) three times lower embedding dimension \( M_L = 0.3N \). The corresponding imaging quality is illustrated in the figure through the reconstructed image (centre, log scale) and the error image (right, linear scale).

If confirmed and extended to truly high-dimensional data, such conclusions may significantly impact the field of radio interferometric imaging by providing significant reduction of memory requirements and computing time. Note that fast implementations of \( \Phi' \) as an operator are critical as it is used at each iterative step during reconstruction. Gaussian embedding—although possibly optimal in terms of imaging quality—remains computationally prohibitive, and in this context, dimension embedding with other random matrices will be further explored, beginning with Bernoulli ensembles (which are expected be computationally ‘lighter’ as compared to Gaussian ensembles, owing to their sparser and binary structure), and subsampled Hadamard transforms (for which fast implementations exist), as outlined in [5].

REFERENCES

PyMORESANE: Pythonic and CUDA-accelerated implementations of MORESANE

Jonathan Kenyon*, Oleg Smirnov*, Arwa Dabbech†, Chiara Ferrari†

* Department of Physics and Electronics, Rhodes University, Grahamstown, South Africa
† Laboratoire Lagrange, UMR7293, Université Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur, F-06300 Nice, France

Abstract—Developments in compressed sensing techniques have led to a whole new class of deconvolution algorithms. One such algorithm, MORESANE, has recently been shown to be effective at recovering faint, diffuse emission at or below the noise level. However, MORESANE and its ilk are computationally expensive. This fact, in conjunction with a desire to make MORESANE available without the use of proprietary software, inspired the development of PyMORESANE - an accelerated version of MORESANE written in Python. It is freely available and, with the addition of CUDA acceleration, it boasts execution times comparable with that of its legacy competitors such as CLEAN.

Deconvolution has long plagued the field of radio interferometry - it is vital in recovering the science embedded in radio images, but it is by no means a trivial problem. As a result, methods of effective and efficient deconvolution are essential, particularly as the next generation of radio telescopes will be more sensitive and have higher data-rates than ever before.

Currently, the tried and tested algorithm known as CLEAN [1] still produces remarkably impressive results. However, CLEAN and its derivatives are not without their flaws and CLEAN in particular struggles to model and recover diffuse emission. The aforementioned increase in sensitivity will likely yield more detectable diffuse emission in interferometric data. This, in turn, will necessitate a class of deconvolution algorithms capable of performing at least as well as CLEAN on compact sources, but far better on diffuse emission.

It is to this class of algorithm that MORESANE (MOdel REconstruction by Synthesis-ANalysis Estimators), developed by Dabbech et al., belongs. A full description of MORESANE is available in [2], which also presents positive results.

MORESANE was originally implemented in MATLAB. However, as MATLAB is proprietary software and the intention was for the algorithm to be used by the community, a Python implementation has been developed - PyMORESANE. This implementation, while heavily based on the original, boasts additional features which allow users to adapt the algorithm to suit their needs. These include access to a variety of tuning parameters, several options for improving convolution, and far superior scalability. Hardware permitting, PyMORESANE can run on virtually any problem size.

The development of PyMORESANE was a multi-stage process. The original implementation of PyMORESANE relied solely on optimisation by code re-factoring. However, even though this did improve the MATLAB version, the problem of dealing with large images rapidly became apparent. Since MORESANE operates in the image domain, execution time scales with the number of pixels in the input image. Next-generation radio interferometers will make large images all the more necessary. As a result, implementations of new algorithms need to be able to deal with large datasets in a realistic time-frame.

Fortunately, the MORESANE algorithm is conducive to GPU acceleration. This acceleration was implemented using PyCUDA - Python’s interface to NVIDIA’s CUDA (Compute Unified Device Architecture) for GPUs. Due to the massive parallel computing power of GPUs, they become invaluable when problem sizes become sufficiently large and the CPU is no longer capable of performing the sheer number of computations required. The FFT (Fast Fourier Transform), which is integral in most deconvolution procedures is well optimised for GPUs and it alone massively improved one of the most time-consuming steps of the algorithm.

The results of these optimisations is PyMORESANE as it now stands. It produces results which are identical to the original MATLAB implementation, but, with appropriate GPU hardware, does so in a fraction of the time. Its sensitivity to diffuse emission, coupled with its realistic run-times show PyMORESANE to be an effective deconvolution tool.

References

Weak Gravitational Lensing

Thomas D Kitching

University College London, Mullard Space Science Laboratory, Dorking, Surrey, UK, RH5 6NT

Abstract—Weak gravitational lensing is the effect caused by photons being influenced by the gravitational field caused by massive objects along the line of sight. Light from galaxies is lensed in this manner causing a slight change in a galaxy's observed third eccentricity (ellipticity). Light from the cosmic microwave background (CMB) is also lensed, causing a detectable change in the temperature and polarisation correlation functions. I will review the current state-of-the-art in the study of weak gravitational lensing from the act of observation to the inference of cosmological parameters. Observationally the measurement of galaxy ellipticities is particularly difficult due to the presence of noise in images, and the pre-lensed distribution of ellipticities (a prior) is unknown. The interpretation of the weak lensing observables is also challenging because a fully three dimensional (on the ball) analysis is required, and the inter-datum covariance between the galaxy weak lensing and the CMB weak lensing signal must be included, if both data sets are to be combined correctly. Weak lensing is also dependent on several systematic effects, for example that galaxies can tidally align in a manner that can mimic the lensing effect, and that on small scales feedback effects between dark matter and baryonic effects are difficult to model. I will present the challenges, current solutions, and open problems in this field.
Bayesian Inference for Radio Observations: 
Joint Calibration and Science Inference

Michelle Lochner†‡§, Iniyan Natarajan‡, Jonathan Zwart¶, Oleg Smirnov∥, Bruce Bassett†‡§, Martin Kunz¶ and Nadeem Oozeer‡†

† African Institute for Mathematical Sciences, 6 Melrose Road, Muizenberg, 7945, South Africa.
department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch, Cape Town, 7700, South Africa.
§ South African Astronomical Observatory, Observatory, Cape Town, 7935, South Africa.
× Department of Physics & Astronomy, University of the Western Cape, Private Bag X17, Bellville 7535, South Africa.
∥ SKA South Africa, 3rd Floor, The Park, Park Road, Pinelands, 7405, South Africa.
‡ Centre for Space Research, North-West University, Potchefstroom 2520, South Africa.
§ Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch, Cape Town, 7700, South Africa.
¶ South African Astronomical Observatory, Observatory, Cape Town, 7935, South Africa.
× Department of Physics & Astronomy, University of the Western Cape, Private Bag X17, Bellville 7535, South Africa.
∥ SKA South Africa, 3rd Floor, The Park, Park Road, Pinelands, 7405, South Africa.
†‡§ African Institute for Mathematical Sciences, 6 Melrose Road, Muizenberg, 7945, South Africa.
‡ Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch, Cape Town, 7700, South Africa.
§ South African Astronomical Observatory, Observatory, Cape Town, 7935, South Africa.
× Department of Physics & Astronomy, University of the Western Cape, Private Bag X17, Bellville 7535, South Africa.
∥ SKA South Africa, 3rd Floor, The Park, Park Road, Pinelands, 7405, South Africa.
† Centre for Space Research, North-West University, Potchefstroom 2520, South Africa.

Abstract—Current deconvolution methods for analysing radio interferometric data produce a single image of the sky that one must assume is its true representation. In fact, this is just one realisation of an infinite ensemble of images compatible with the noise in the raw data. Upcoming telescopes like the Square Kilometre Array (SKA) [1] will push the sensitivity into a new regime and systematics like direction-dependent effects will become more significant. We present a probabilistic approach, Bayesian Inference for Radio Observations (BIRO), that simultaneously estimates both science and instrumental parameters from the raw data providing better estimates and uncertainties than a standard deconvolution procedure. We demonstrate BIRO with simulated Westerbork Synthesis Radio Telescope (WSRT) [2] datasets.

I. INTRODUCTION

The SKA, when fully constructed, will be up to 50 times more sensitive than current instruments. This, combined with a relatively cheap antenna design, means that a rigorous treatment of instrumental effects will be necessary. The fact that the number of data points scales as the number of baselines, $N^2$, while the number of instrumental parameters scales only as the number of antennas, $N$, means that the sky and instrumental parameters can be estimated simultaneously, even for large $N$ [3]. The standard approach to imaging alternates the use of deconvolution methods like CLEAN [4] and calibration models to determine the best-fitting, calibrated image. This can only provide point estimates for model parameters that may be biased due to the presence of random and systematic noise.

II. BAYESIAN INFERENCE

In our Bayesian approach, we infer the science and instrumental parameters simultaneously and derive the correlations between them along with accurate error estimates. This requires sampling the full posterior probability distribution of the parameters which we do using an MCMC [5], [6] algorithm and the software MeqTrees [7] for modelling.

III. SIMULATIONS

We demonstrate our technique with two key simulated datasets: one where we jointly estimate science (source flux density) and instrumental (pointing errors, primary beam, receiver noise) parameters and the other where we focus on the problem of being able to reliably discriminate between different source models (between one or more point sources and an extended source). In this poster, we focus on the joint inference problem, comparing the source catalogue of fluxes retrieved by BIRO with the ones retrieved by the traditional CLEAN+SE (source extraction) method. As example instrumental errors, we include polynomial, time-varying pointing errors and telescope beam parameters, which results in a total of 103 instrumental and scientific parameters estimated. We find that BIRO extracts accurate fluxes, including reliable uncertainties, while CLEAN+SE (without calibration) fails to find several sources and produces biased fluxes for the sources it does find.

REFERENCES

Abstract—We apply MCMC sampling in Radio Interferometry (RI) imaging. Most imaging techniques give a point estimate of the observed sky. This is insufficient for capturing uncertainty about the reconstruction. Estimation of the pixel variances (as a measure of uncertainty) cannot be derived solely with noise estimates because it ignores missing data that was estimated by the reconstruction method. The Bayesian method provides a posterior distribution that can be used to study uncertainty in imaging based on a given prior. We use a Laplacian prior for the image pixels that is reasonable in RI imaging and allows the sampling to be feasible for demonstrational purposes.

I. INTRODUCTION

The most popular algorithms for image reconstruction in RI are CLEAN ([1]) and the Maximum-Entropy Method (MEM) (e.g. [2]). Recently, Compressed Sensing (CS) techniques were also proposed (e.g. [3]). However, those algorithms do not give an estimate of the certainty of the image features, e.g. in the form of posterior variances of the pixels. The MEM would allow this if the entropy is interpreted as a (log-) prior, but this doesn’t seem to be a reasonable prior for RI. There exist some recent developments in this area. In [4], the uncertainty or variance is approximated by the inverse curvature of the posterior at its maximum. In [5], Gibbs sampling provided the variance. In [6], sampling is used for a MEM-like approach.

II. HMC SAMPLING OF THE POSTERIOR

Although the $l_1$-norm is used as a regularization in CS, it also makes sense as a negative log prior (see equation 1) because it gives most weight to small pixel values around 0 and has almost heavy tails while still being log-concave, which is preferable when optimizing or sampling from the posterior. The Laplacian distribution is parametrized by $\lambda$ and given by

$$p(x) \propto \exp(-\lambda ||x||_1).$$

We assume observations of the following kind

$$v = F(a \odot x) + n$$

with the image $x$, the primary beam $a$, and the Fourier-type matrix $F$ with entries depending on the image pixel locations and the measured uv-points. Here, we account for a noise term $n$ which has a normal distribution with variance $\sigma^2$. Equations 1 and 2 yield a posterior probability distribution. We sample from this distribution using the Hybrid Monte Carlo (HMC) method because computing the pixels means and variances analytically is not possible. HMC uses the differentiability of the posterior and allows to sample it more efficiently by keeping (latent) momentum variables.

III. EXPERIMENTAL RESULTS

The test images consisted of six images of the CASA online guide one of which is shown in figure 1. We use toy array settings in which the visibilities lie on a regular grid to avoid costly gridding steps. The primary beam $a$ drops to $10^{-3}$ at the edges of the image. All images had a size of $64 \times 64$ pixels. For each image 250 samples were drawn (25 HMC chains with 10 samples each). We used Estimated Potential Scale Reduction (EPSR) to ensure that the samples were almost independent of each other.

Fig. 1: Imaging results. From left to right: the true image, the MAP estimate ($SNR = 12.51$), the posterior mean ($SNR = 20.69$) and the standard deviation of the pixel marginals.

Fig. 2: Some of the samples for the image shown in figure 1.

We compared the mean of the samples to the $l_1$-regularized reconstructions (the maximum of the posterior distribution). The mean was better in terms of SNR for all six images of the CASA guide. For the computation of the SNR the pixel values were weighted by the primary beam.

Figure 2 shows a few of the samples for the image shown in figure 1. It illustrates that there are different possible explanations for the measured data and how large the variations are.

IV. CONCLUSION

For all images the variance estimate of the pixels looked similar to the image itself: pixels with a high value also have a high posterior variance. Sampling the posterior assuming a Laplacian prior provides not only new insights about the uncertainty in RI imaging but the mean of the samples is also a good estimate of the image. We are currently working on experiments with a non-negativity constraint yielding better estimates. Experiments with real data sets are also in progress.

REFERENCES


Analysing the polarisation of the CMB with spin scale-discretised wavelets

Boris Leistedt∗, Jason D. McEwen†, Martin Böttner∗, Hiranya V. Peiris∗, Pierre Vanderghynst‡, and Yves Wiaux§.

∗ Department of Physics and Astronomy, University College London (UCL), London WC1E 6BT, UK
† Mullard Space Science Laboratory (MSSL), University College London (UCL), Surrey RH5 6NT, UK
‡ Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
§ Institute of Sensors, Signals & Systems, Heriot-Watt University, Edinburgh EH14 4AS, UK

Abstract—We discuss a new scale-discretised directional wavelet transform to analyse spin signals defined on the sphere, in particular the polarisation of the cosmic microwave background (CMB).

I. SPIN SCALE-DISCRETISED WAVELETS ON THE SPHERE

We design a directional scale-discretised wavelet transform to analyse the directional features of signals of arbitrary spin on the sphere. Following Refs. [1]–[4], the spin-discretised wavelets ψjℓm on the sphere are defined in harmonic space as ψjℓm(ξ) = κℓ(ξ)ξjℓm, where κℓ(ξ) characterises the angular localisation of the wavelets (for the jth scale), while ξjℓm controls their directionality. Fig. 1 shows examples of wavelets obtained with this construction.

Fig. 1. Spin 0 and 2 wavelets constructed with angular band-limit L = 512, azimuthal band-limit N = 15, and tiling parameters B = 2 and dmin = 2.

The wavelet transform of a spin signal sj on the sphere is

\[ f \rightarrow \mathcal{W} = \{ \mathcal{W}_j \} \quad j \geq 0 \]

where \( \mathcal{W}_j \) are the wavelet coefficients of the observable. The wavelet coefficients are synthesised exactly from its wavelet coefficients. More details about this transform can be found in Ref. [4].

II. APPLICATION TO THE CMB

The polarisation of the CMB is currently an intense avenue of research, since it may reveal signatures of primordial gravitational waves and a glimpse on the initial conditions of the universe. However, the physical quantities necessary for these investigations – namely the E- and B-modes of the polarisation, or gradient and curl modes – are not directly available. They are obtained by reducing and transforming frame-dependent observables: the Q and U “local” polarisation measured on the sky. This Q-U to E-B transformation is unambiguous when dealing with data covering the entire sky. However, when CMB observations cover fractions of the sky only, E-B reconstruction is imperfect near the boundaries of the observation mask, causing leakage and potential biases in the recovered E- and B-mode maps. This issue is traditionally addressed by smoothing and extending the mask to remove boundary regions where the leakage is important. We develop a method exploiting the novel directional spin wavelet transform to decrease the leakage near the mask boundaries and obtain a more accurate E-B reconstruction.

First, we compute the wavelet coefficients of the observable Q+iU using our spin wavelet transform. Second, we deal with the partial sky coverage by masking the data in spin wavelet space. Third, we apply to the real and imaginary parts of these wavelet coefficients a scalar inverse wavelet transform. It can be shown that this yields estimates of the E and B signals provided the wavelets of this inverse transform are spin-lowered versions of those used in the first transform Ref. [4]. This new method to estimate the E- and B-mode contributions from Q and U observations can significantly reduce the E-B leakage by exploiting improved masking in spin wavelet space.

We anticipate our novel spin wavelet transform will be of general use in analysing CMB polarisation data, beyond this first application to E-B separation. A more detailed description of spin scale-discretised wavelets, fast algorithms, and a rigorous evaluation of their performance will be given in a series of forthcoming articles.

REFERENCES

Bayesian Inference for Radio Observations: Source Separation

Martin Kunz∗†, Michelle Lochner†‡§, Iniyan Natarajan≡, Jonathan Zwart¶, Oleg Smirnov∥
Bruce Bassett†‡§, and Nadeem Oozeer†‡

Abstract—When producing an image, there is often more information than is visible to the eye which can be revealed using statistics. This is particularly relevant to radio interferometry, where there is an infinite number of possible images consistent with the data. We propose a Bayesian approach to interferometry, which determines scientific parameters, such as source position, shape and brightness, directly from the data without the need for imaging. With this technique, we can use the Bayesian evidence to perform source separation on scales smaller than the beam of the telescope. There are clear links to the biomedical imaging field with such probabilistic techniques.

I. INTRODUCTION

In radio interferometry, the gaps between antennas result in missing information in the data. The current approach to producing an image is to apply a deconvolution algorithm, under certain assumptions. However, this image is just one possible realisation and may be subject to biases from calibration errors etc. Additionally, images only contain a fraction of the information stored in the original dataset. To achieve maximum sensitivity from the upcoming SKA [1] telescope and other interferometers, we propose instead a Bayesian technique to extract science parameters directly from the data.

II. SOURCE SEPARATION WITH BIRO

Our technique, Bayesian Inference for Radio Observations (BIRO), combines the software MeqTrees [2], for modeling, with the Bayesian sampler MultiNest [3], [4], [5]. We test BIRO on simulated Westerbork Synthesis Radio Telescope (WSRT) [6] data with sources on sub-synthesised beam scales. The images produced from our datasets using a standard deconvolution algorithm reveal no structure on such small scales. However, by using the Bayesian evidence and applying various models, we find we are able to clearly distinguish between a single point source, an extended source or two point sources, showing how information can be extracted by going beyond imaging.

REFERENCES

Abstract—A new high-performance imaging tool for Radio Interferometry has been developed that performs nearly 100× faster than another general CPU-based imaging tool.

A radio interferometer indirectly measures the intensity distribution of the sky over the celestial sphere. Measurements are made over an irregularly sampled Fourier plane and various distortions have to be accounted for. Synthesising an intensity image from interferometric measurements requires substantial processing.

A new high-performance image synthesis tool, called the Malta-imager[1] has been developed. Implemented in C++ and CUDA, the imaging tool achieves unprecedented performance by means of Graphics Processing Units (GPUs). The imaging tool is divided into several components, and the back-end handling numerical calculations is generalised in a new framework (The General Array FrameWork). A new feature termed compression arbitrarily increases the performance of an already highly efficient GPU-based gridding implementation. Compression takes advantage of the behaviour of oversampled convolution functions and the baseline trajectories. A CPU-based component prepares data for the GPU which is multi-threaded to ensure maximum use of modern multi-core CPUs. Best performance can only be achieved if all hardware components in a system do work in parallel. The imaging tool is designed such that disk I/O and work on CPU and GPUs is done concurrently.

Test cases show that the imaging tool performs nearly 100× faster than another general CPU-based imaging tool. Current work will implement deconvolution and A-projection, whilst finding ways of imaging efficiently very large images made up of 1 billion pixels.

The Malta-imager is available under the GPLv3 license and can be downloaded from the Ska-sa GitHub repository1.

REFERENCES

1Author can be contacted via email at daniel.muscat@ieee.org.
Radio interferometric imaging for the SKA and its pathfinders

André R. Offringa ∗

∗ Research School of Astronomy & Astrophysics, Australian National University, Weston Creek, 2611 ACT Australia.

Abstract—To make maps of the sky with a radio interferometer, the radio data has to be processed with an imaging algorithm. The SKA pathfinders show that this imaging step can be a considerable challenge. The imaging complexity scales with the number of elements, maximum baseline length, angular field of view, data volume and dynamic range. The SKA will be larger by orders of magnitude on almost all of these aspects, and will therefore require new advances in solving the imaging problem. I will give an overview of the challenges involved, and summarize the currently available approaches.

I. INTRODUCTION

Many algorithms are available for imaging radio data, that each have their own benefits and performance dependencies. In practice, many of these algorithms are combined, making a full imaging program very complex. In the following sections, I will describe the commonly used algorithms and their use for the SKA.

II. w-TERM CORRECTION

The first step in imaging is the mathematical inversion of the function that maps the sky brightness to the correlation products. The performance of this step often dominates the total imaging time. Arrays such as the SKA that will have a wide field of view and non-coplanar elements require expensive correction for the ‘w-terms’, which describe the distance of an element to the plane. The fastest algorithms to correct for the w-terms are w-projection [1] and w-stacking [2]. These algorithms have a squared and linear dependency on the size of the w-term respectively, although the faster w-stacking algorithm requires more memory [3]. Further improvements can be made by combining either of these techniques with snapshot imaging. This decreases the size of the w-term, but requires regidding of the snapshot images before they can be integrated or deconvolved [4]. This algorithm has recently been implemented in WSCLEAN, and is useful for MWA imaging of very off-zenith or very large fields of view [3].

III. CORRECTION OF DIRECTIONAL EFFECTS

Direction-dependent effects (DDEs) such as the ionosphere or a changing or heterogeneous primary beam – as will be the case for the SKA – have to be corrected before images from different epochs are added together. It has been shown this can be corrected during the inversion step using the a-projection technique [5] or, when the changes are uniform over the elements, by correcting snapshots in image space [3]. For certain effects, a hybrid of these two can be used to reduce cost.

IV. DECONVOLUTION

An image created by inversion is convolved with the point-spread function (PSF) of the instrument, and therefore needs to be deconvolved to improve the dynamic range. For a wide-field array such as the SKA, the PSF varies spatially, which complicates this process. The Cotton-Schwab algorithm [6] corrects this: Minor iterations assume a constant PSF and produce a model for the image, while major iterations calculate and subtract the full effect of the changing PSF. The simplest algorithm for implementing minor iterations is the Högbom algorithm [7]. In more complex fields, multiscale (MS) cleaning improves the deconvolution accuracy [8], but the application of CASA’s MS algorithm on MWA data shows the algorithm is too slow for practical use with a wide-field array with so many elements. A fast multi-resolution multi-scale algorithm has recently been implemented in WSCLEAN (Offringa et al., in prep.). Recently, deconvolution using compressed sensing (CS) have shown promising results [9]. CS improves the deconvolution accuracy over Högbom cleaning, but its performance and robustness have not been extensively validated using real data.

Deconvolving multi-frequency synthesis (MFS) images is complicated further when a large bandwidth is imaged, making it necessary to account for spectral variation of sources. The Sault-Wieringa algorithm [10] takes spectral variation into account during deconvolution, and produces the MFS map as well as a spectral-index map. A different approach is taken by WSCLEAN, which splits the bandwidth in subbands and deconvolves them joinedly. Further variation can be corrected during the major iterations.

V. PARALLELISM & DATA FORMAT

In practical situations, a hybrid of these algorithms is often used, although different science goals will need different subsets of algorithms. Software packages such as CASA and WSCLEAN already implement many of these algorithms. Adaptions will be necessary for SKA’s data volumes, such as parallelization over subband, short snapshot or image subset (mosaicing). The data will need to be available in such a way that each node can work on a subset of the data.

REFERENCES

Studying the Milky Way via stacks of low S/N spectra

Dovi Poznanski∗

∗ School of Physics and Astronomy, Tel Aviv University, Tel Aviv, Israel.

Abstract—The gas and dust constitute the bulk of the baryonic mass of the Galaxy, and yet are poorly understood in many respects. Their main traces are very weak and narrow absorption lines that are typically studied using stars as background sources and costly high quality spectra. By using robust, yet simple, computational techniques: paired with novel computational tools, we have recently shown that we can extract the imprint of these very weak features from low resolution, low signal-to-noise (S/N) spectra of background quasars and galaxies. This opens a new window to study these crucial components of our Universe. Can sodium lines be used to trace dust extinction? What are the carriers of the mysterious diffuse interstellar bands (DIBs)? How do DIBs correlate with other species in the interstellar medium? These are some of the questions that we are addressing.

I. SCIENCE QUESTIONS

The interstellar matter (ISM) forms stars, is formed by stars, and through this cycle carries information about the past and future of a galaxy. Yet many of its properties remain elusive. For example, what are the carriers of the Diffuse Interstellar Bands (DIBs)? The DIBs are the oldest unidentified spectral features. Since the 1920s these hundreds of absorption bands of various width, present in the optical spectra of virtually every star, have confused the community. Space is a special environment that is difficult to reproduce in the laboratory. DIBs may also be the spectral signature of molecules never discovered before. Since hundreds of lines are known they are unlikely to all come from a single species. But if one can securely divide the DIBs into families, the list of candidate molecules is substantially reduced. Current divisions only apply to a few lines and are anecdotal, in the sense that they are determined for only some of the strongest lines, they are often based on observations of no more than a few stars, or they are based on comparisons between a small number of sight lines. Useful knowledge could be gleaned from the correlations of DIBs with other species and properties of the ISM. For example, it is strongly established that most DIBs correlate well with dust (though not all of them do). But the relations at low column densities are typically hard to measure, as color excesses and equivalent widths are small. Also, because of both intrinsic and measurement scatter, the exact functional relations between DIB strength and dust column are not well known, making groupings difficult.

II. METHODS

In a pilot project [1] we took about a million extragalactic spectra from SDSS, and used them in a novel way. Every object in SDSS is unavoidably observed through the gas and dust in the Milky Way (MW). A single such spectrum does not offer the S/N ratio needed to detect this imprint, but stacking many of them to produce a high-fidelity composite spectrum recovers the signature in absorption. We stacked thousands of spectra with a similar expected extinction and measured the dependence of the line strengths on the color excess due to dust. Our approach allowed us to beat down the ample noise and intrinsic signal from every source (i.e., we removed the quasar/galaxy lines), recovering just the effect of the Galaxys ISM with high signal-to-noise. This can be clearly seen in the figure where we recover the weak DIBs (from [2]).

The methods we first used could be qualified as brute force, yet pull a surprising amount of information from the noisy SDSS spectra. We plan to optimize them in terms of efficiency and resolution. We are using a new array-based database paradigm (www.scidb.org). This project is not trivial computationally: 2.5 million spectra, each of length ~ 30 thousand resolution elements (assuming 0.2Åfter interpolating), can effectively be represented as a 600 GB matrix. operating on so much data from disk is too slow, and holding it in memory requires dedicated hardware. Soon, surveys like DESI will generate more than 10 times more spectra. SciDB is a scientifically oriented database, that is being developed precisely with such projects in mind, allowing not only the division of the data to be transparent, but mostly enabling parallel access from multiple nodes to the database and expensive on-query sort operations (needed for calculating medians) to be performed quickly.

Our first results using SciDB, looking at the distribution of DIBs were recently submitted for publication [2]. We plan to apply super resolution algorithms to our data. These are a family of algorithms that, allow one to stack multiple low-resolution images (e.g., from security video) to obtain a high resolution single image (e.g., of a person or license plate). SDSS spectra could essentially be viewed as noisy low-resolution 1D images of the ISM. The spectrum of the source (here the ISM at similar sight lines) was sampled many times, dithered on the detector. Our current method is analogous to a convolution of the source with a window function, effectively the resolution of the instrument, about 2.5Å. Super-resolution techniques, however, use the existing data in the multiple exposures to perform a smarter interpolation, thus increasing the resolution of the target image. Depending on the resolution we achieve we might be able to study gas dynamics, looking for inflows and outflows on large scales.

REFERENCES

Multiscale analysis of Galactic dust emission

Jean-Franc ·ois Robitaille∗

∗ School of Physics & Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK.

Abstract—Anisotropic complex wavelet transforms are used to make a multiscale analysis of the distribution of fluctuations of dust emission at 250 µm from the Hi-GAL survey of the Herschel Space Observatory. By using an iterative algorithm, I will show that it is possible to separate the Gaussian and non-Gaussian part of the wavelet coefficient distribution for each scale and azimuthal direction. The reconstructed map of non-Gaussian fluctuations is well correlated with small-scale structures in the 13CO emission map.

I. INTRODUCTION

Mechanisms of molecular cloud formation in the interstellar medium (ISM) are not clear yet. Atomic gas dynamics are driven by multiple physical processes which take place on different scales (injection of energy by spiral arm shocks, explosion of supernovae, turbulence energy transfer to smaller scales, etc). In this context, a clear picture of the formation of molecular clouds is hard to establish. To describe the complex density structures seen in the ISM, we need sophisticated statistical tools. The simultaneous spatial and spectral properties of wavelets make them an excellent tool to characterise local fluctuations and the interconnection between different components present in a medium. Dust grains play an important role in the catalytic reaction of molecule formation in the ISM. For this reason, it is assumed that the dust is homogeneously mixed with the atomic and molecular gas and its emission can be used to trace the total amount of matter along the line of sight.

II. COMPLEX WAVELET TRANSFORMS

The Fourier power spectrum assumes Gaussian distributions at every scale so that the averaged power provides relevant information about distributions and can be compared from one scale to another. These assumptions are rarely true for real physical media. Complex wavelet coefficients allow direct access to the spatial distribution of power on different scales as well as the local phase information of fluctuations in the image. In this work, the continuous complex Morlet wavelet has been chosen for multiscale analysis of thermal dust emission in a representative Herschel infrared Galactic Plane Survey (Hi-GAL) subfield. In two dimensions, the Morlet wavelet is anisotropic and its Fourier transform can be written as a localised Gaussian distribution in the u-v plane:

\[
\hat{\psi}(k) = e^{-|k-k_0|^2/2} = e^{-((u-u_0)\cos \theta)^2+(v-v_0)\sin \theta)^2}/2. \tag{1}
\]

By rotating the function over the angle \( \theta \), it has been shown that the Morlet wavelet is also the best wavelet to reproduce the Fourier power spectrum using the following relation [1]:

\[
P(l, \chi) = \frac{\delta \theta}{N_\theta} \sum_{j=0}^{N_\theta-1} \hat{f}^* (l, \chi, \theta_j) \cdot \hat{f} (l, \chi, \theta_j), \tag{2}
\]

where \( \delta \theta \) is the azimuthal increment between \( N_\theta \) successive wavelets and \( \hat{f} (l, \chi, \theta) \) represents the wavelet transform of \( f(\chi) \) at a given spatial scale \( l \) and a given azimuthal direction \( \theta \).

III. NON-GAUSSIANITIES SEPARATION

Using an iterative algorithm, it is possible to isolate the two parts of a turbulent flow, non-Gaussian fluctuations embedded in homogeneous and random flows, by analyzing their wavelet coefficient distributions as a function of scale and direction [2]. The reconstruction of both parts, Gaussian and non-Gaussian fluctuations of the Hi-GAL subfield, are shown in Figure 1. Power spectra of both parts possess different power laws (\( \gamma = -2.3 \) and \(-3.1 \) for the non-Gaussian and the Gaussian part respectively; where \( P(k) \propto k^\gamma \)). We found that non-Gaussian fluctuations are well correlated spatially to the 13CO emission which also shows a similar power law \((-2.5)\).

IV. CONCLUSION

The isolated Gaussian part, with its steeper power spectrum and smooth structures, presents similar characteristics to a turbulent medium. The extraction of another independent embedded power law may suggest that one or multiple other physical processes are the origin of these isolated structures, e.g. gravitational and/or thermal instabilities. This technique is very general and can find various fields of application, from cosmology to medical image analysis.

REFERENCES

Optimal deblending and stacking across multi-band surveys using LAMBDAR

Aaron Robotham*, Angus Wright*, and Simon Driver*

* ICRAR, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

Abstract—A common procedure in astronomy has been to detect sources in a deep image and extract photometry by placing apertures on position matched images in other bands. Entering an era of truly multi-band surveys with highly discordant imaging Point Spread Functions (GAMA has 20 bands over 250 sq deg covering the far-UV to the far-IR and PSFs ranging from 0.5" to 37") means new approaches to optimally extracting photometry are required. Here I present a new astro-imaging photometry package written in R called LAMBDAR. It is designed to fully exploit diverse imaging datasets to allow stacking of shallow data and deblending of deep imaging.

I. INTRODUCTION

To extract the maximum amount of science out of astronomical surveys observations are made in as many bands as possible. By covering the full spectrum of typical galaxies we can best understand the physical processes that drive galaxy evolution in representative volumes of the Universe. GAMA is such a survey, covering the energy output of galaxies from the far-UV through to the far-IR with 20 broad bands of photometry.

A serious issue with such data is the mismatch in the depth of observations due to a mixture of different imaging facilities being used at different wavelengths and variations in the energy output of galaxies. On top of this the point-spread function describing the sharpness of the imaging can vary by multiple orders of magnitude between facilities.

With this situation in mind we have developed the Lambda Adaptive Multi-Band Deblending Algorithm in R, or LAMBDAR. Using optical priors and knowledge of the PSF in other bands the code is able to extract photometry in shallower data whilst also correctly deblending the flux between sources that merge as the PSF inflates.

Figure 1 is an example of the type of deep high resolution we have in GAMA. In comparison the top-left panel of Figure 2 is an example of the much shallower and lower quality imaging we have in some of our longer wavebands where the source galaxy is badly confused by a nearby bright star. By combining knowledge of this star’s location and the exact shape of the PSF we are able to meaningfully redistribute the flux amongst the image sources and extract good photometry for our target galaxy. Ultimately this software will be used to stack photometry where we have no significant signal in a target image.

Fig. 1. A multi-colour image of a GAMA galaxy.

Fig. 2. The top two panels are a pictorial representation of the deblending and flux extraction executed by LAMBDAR. The bottom panel is an example SED where we have constructed the energy output over all 20 bands of GAMA.

This talk will focus on the concepts behind the software and its first application to the GAMA dataset. The LAMBDAR R package will be publicly released, and we envisage many potential uses beyond astronomy.
Fast Phase Transition Estimation

Ludwig Schwardt*

* SKA South Africa, Third Floor, The Park, Park Road, Pinelands, 7405, South Africa

Abstract—Phase diagrams are useful for understanding the performance of compressed sensing reconstruction algorithms. They can be expensive to produce if these algorithms are computationally intensive. I present an active learning technique that estimates the location and width of a phase transition which is at least an order of magnitude faster than the traditional brute-force approach and apply it to the benchmarking of several sparse reconstruction algorithms applicable to radio astronomy.

I. INTRODUCTION

Phase transitions occur in many problems involving high-dimensional geometry and play an important role in the field of compressed sensing [1]. Compressed sensing is concerned with the recovery of a sparse $N$-dimensional signal with $S$ non-zero components from $M$ linear measurements, where typically $S < M < N$. The probability of recovery as a function of parameters $S$, $M$ and $N$ changes dramatically from near-guaranteed failure to near-guaranteed success within a small range of parameter values and therefore exhibits a phase transition.

This is useful for gauging the performance of compressed sensing reconstruction algorithms as the phase transition observed for each algorithm on a given problem set can be compared to the ideal phase transition curve predicted by theory [2]. The comparison is usually done in terms of a phase diagram which shows the probability of successful signal recovery as a function of some experimental parameters. The most popular version is a two-dimensional plot of recovery rate versus undersampling ($M/N$) and sparsity ($S/M$) [2].

The traditional way to create a phase diagram is to perform a number of trials (typically about 100) on a regular grid of points in parameter space. For each trial the reconstruction algorithm is evaluated on a random set of data with the appropriate parameter setting. The outcome of the trial is either success or failure. The plot of the relative number of successes as a function of the parameters then highlights any phase transitions that are present. Examples of this procedure can be found in SparseLab [3] and elsewhere [4].

This can be very inefficient, especially if the reconstruction algorithm is computationally intensive. The brute-force approach typically spends most of its time performing trials in regions far from the transition where the probability of recovery is already well known.

I present a technique that estimates the location and width of a phase transition by selectively performing trials based on the outcomes of previous trials. It is a form of active learning [5] that explores the parameter space based on the assumption that it contains a single smooth transition boundary. The quality of the estimate improves incrementally with each new trial.

I show that the algorithm can find transitions in a time that is at least an order of magnitude faster than the brute-force approach. The algorithm is then used to speed up the benchmarking of several compressed sensing algorithms that are considered for use in radio astronomy.

REFERENCES


Surveying the Universe - the past, present and future of galaxy redshift surveys.

Rita Tojeiro*

* School of Physics & Astronomy, University of St Andrews, St Andrews, Fife, KY16 9SS, UK.

Abstract—The large-scale structure of the Universe holds a vast amount of information about the content, dynamics and geometry of the Universe. To map the Universe on these large scales we measure accurate three-dimensional positions of millions of galaxies, which in turn trace the underlying and dominating Dark Matter component. We compress these three-dimensional maps into quantities that trace the expansion of the Universe, or the rate at which structures are growing. The last 15 years have seen us going from the first (marginally) statistically useful maps to one of the cleanest and most powerful probes of the Universe. I will review our journey to date and the challenges we have overcome. I will then present the state-of-the art results from present-day galaxy surveys, and discuss the challenging path that lays ahead.

I. INTRODUCTION

Current observational evidence increasingly points towards a scenario where the Universe is undergoing an accelerated expansion (see e.g. [1], [2], [3], [4], [5], [6], [7]). The physical reason behind such an acceleration remains a mystery, and potential explanations range from a simple cosmological constant or vacuum density, to modified gravity models or an inhomogeneous Universe creating the illusion of an acceleration. A key goal of modern cosmology, therefore, is to measure the expansion rate of the Universe with increasing precision, with the clear intent of glimpsing the physics behind the Universe’s acceleration by direct comparison of this measurement with predictions arising from different physical models. This increase in precision must be matched by an increase in accuracy, and both aspects are a challenge for modern-day galaxy redshift surveys.

The last 30 years have seen a phenomenal increase in our ability to measure the detailed large-scale structure of the Universe via galaxy redshift surveys. Such surveys provide a wealth of information on cosmological models and on the evolution of galaxies, and are a remarkably versatile tool of modern Astronomy. With enough volume surveyed, the expansion history of the Universe can be measured via the baryon acoustic oscillation (BAO) scale - an imprint of the comoving sound-horizon size at the time of recombination on the distribution of galaxies. The BAO scale has now been convincingly measured using a variety of datasets and methodologies and, as it can be found at large scales ($\approx 100 h^{-1}$ Mpc), it remains free of many astrophysical systematics, providing one of the most robust probes of the expansion history of the Universe (see [8] for a review). Nonetheless, as distance measurements approach a precision of 1%, significant pressure is put on our understanding of the data and all aspects of the survey.

A great advantage of modern galaxy redshift surveys is that they provide, essentially for free, a measurement of redshift-space distortions (RSD): an anisotropy in the clustering statistics of galaxies due to the coherent flow of galaxies as they infall into large-scale gravitational potential wells. RSD are therefore sensitive to the theory of gravity and a very powerful tool of General Relativity on large scales. Together, BAO and RSD, stand as one of our most powerful tools in investigating the expansion and the growth of structure of our Universe.

We now stand at a critical point in the design and exploitation of redshift surveys. A point at, whereby constructing bigger experiments that probe ever larger volumes of the Universe no longer yields immediate benefits. We are limited by systematic errors from our data-analysis and signal-processing techniques, our observation techniques, our modelling techniques and our understanding of the galaxies that we use as probes. This talk will review the past, present and future of redshift surveys, with an emphasis on the challenges that lay ahead.

REFERENCES

Bayesian hierarchical models for supernova cosmology

Roberto Trotta*, Xiyun Jao†, Hikmatali Shariff* and David van Dyk†

* Imperial College London, Astrophysics Group, Blackett Laboratory, Prince Consort Rd, London SW7 2AZ, UK
† Imperial College London, Statistics Section, Maths Department, Prince Consort Rd, London SW7 2AZ, UK

Abstract—I review Bayesian hierarchical methods, and their application to cosmological parameters inference, in particular dark energy, in the context of Supernova type Ia cosmological observations.

I. INTRODUCTION

Bayesian hierarchical models are becoming increasingly popular in astrophysics and cosmology. They offer a statistically principled way of incorporating various sources of variability: observational noise, selection effects, population variability all contribute to the observed scatter, and are often challenging to disentangle using traditional, likelihood-based methods.

Thanks to their superior ability to incorporate naturally the different physical nature of such variability, the Bayesian hierarchical models framework is gaining popularity despite its relative complexity.

In this talk I will review hierarchical modelling and the underlying inferential principles. The power of this approach will be demonstrated with an application to the illustrative case of the Gaussian linear model.

II. APPLICATION TO COSMOLOGICAL DATA

I will then show how this framework can be adopted in the context of supernovae Type Ia cosmology.

Supernovae Type Ia (SNIa) are a type of stellar explosion that can be used as a “standardisable” candle to measure distances in the cosmos. Such a measurement is necessary to reconstruct the expansion history of the Universe, and hence infer its composition. It is SNIa observations that have led, in the late 90’s, to the discovery that the matter-energy density of the Universe is currently dominated by a repulsive vacuum energy, called “dark energy”, responsible for approximately 70% of the contents of the Universe.

One of the biggest mysteries in modern-day physics, the nature of dark energy remains currently unknown. Present and future SNIa observations are one of the most important tools that will help understanding dark energy better, in particular in order to determine whether or not its properties change with time.

The number of SNIa observations is now in the several hundreds, and it is set to increase by over a factor of 10 in the next few years. Already today, our inferences about the nature of dark energy are being limited by poorly-understood systematic effects (such as the reddening and dimming introduced by dust, which can be confounded for the dimming due to the expansion of the Universe).

Bayesian hierarchical models can be used to accurately model various sources of uncertainty and systematic effects, including galactic dust reddening, population variability, residual scatter in the intrinsic SNIa magnitudes after the standardisation procedure, and possible other effects coming from correlations with galactic host galaxy mass and morphology.

I will review some recent results obtained in this framework, and describe how this can be exploited to improve our knowledge of dark energy.
3-D, physical image reconstruction in cosmology

Benjamin D. Wandelt†, Florent Leclercq‡ and Jens Jasche∗
† Institut d'Astrophysique de Paris, UPMC/CNRS, UMR 7095, 98 bis boulevard Arago, 75014 Paris, France
‡ Lagrange Institute (ILP), 98 bis boulevard Arago, 75014 Paris, France

Abstract—Astronomical surveys map the large-scale Universe and find structure everywhere and on all scales. We confront these data sets both with morphological analysis techniques and with detailed models of cosmic content and dynamics. Using advances in Bayesian analysis and sampling techniques we can fuse our data with physical model assumptions to produce quantitative, testable predictions of as yet unobserved phenomena—including the dark sector.

I. THE PREPOSTEROUS COSMOLOGICAL MODEL

To some it seems that we could not live in a more boring universe. All observed phenomena in cosmology are well-described by a 6 parameter model built on an extension of the standard canon of modern physics. And yet, our standard model of cosmology is preposterous: it posits that over 90% of the energy density of the universe has not yet been observed and may indeed be unobservable, and that the early universe emerged from the (presumed) chaos of quantum gravity in an incredibly low-entropy state. The least one can expect from such a preposterous model is that it be predictive. But as cosmologists we have only one universe to study. Since we cannot re-run the cosmic “experiment,” how can we test the predictiveness of the cosmological standard model?

II. PREDICTIONS FROM DATA AND THEORY

Models do not just link initial conditions to observed outcomes, but generally one set of observables (at some time) to another (possibly at another time). The key to confronting the cosmological model with only one observed universe is therefore to have access to many different observables. Fortunately the universe is large and offers many very different observables. All these observables derive from the presence of structure in the universe. Gathering information about the cosmological model means exploiting the presence of structure.

The theoretical description linking these structures to the model parameters ranges from very simple (on large scales where the initial, very nearly Gaussian seeds of structure have experienced minimal gravitational processing) to highly complex (on small scales where gravity and gas physics have created structures that are non-linear, with non-Gaussian and highly correlated statistics). While the large-scale modes are easy to model, the small scales modes are many and therefore potentially very powerful.

III. NON-LINEAR COSMOLOGICAL ANALYSIS

In recent papers (Jasche & Wandelt 2012, Jasche, Leclercq and Wandelt 2014) we have described a fully Bayesian analysis approach of galaxy surveys, including modeling the smaller scales where non-linear effects leading to the appearance of the cosmic web. While we coached the description of this methodology in terms of Bayesian inference of the cosmic initial conditions, another way to think about the results is in terms of “data fusion:” the resulting predictive densities can be used to guide future observations (i.e. to detect new clusters that are outside the current footprint of the survey but are predicted to be there due to the three-dimensional correlations) or to run gas dynamical simulations that produce templates that can be used to do detect new observables.

Our detailed reconstruction of the dynamical evolution and state of our universe allow finding morphological tracers voids, sheets, and filaments with much greater ease than the raw data themselves. This fact was exploited in Leclercq, Jasche, Wandelt (2014) to predict the properties of voids in dark matter using the Sloan data and LCDM.

Some of the resulting images of the cosmic web naturally evoke biological analogies. Maybe the analogies run deeper than the merely visual: the cosmic web arises from caustics due to flow along potential gradients. Is this a valid mathematical analogy for the formation of biological structures? Might the model that allows us to reconstruct the cosmic web with high fidelity be useful for enhancing images of biological structures? Can our tools to explore morphology in cosmic large-scale structure be helpful to extract features from medical data sets or vice versa?

REFERENCES

Opportunities and Challenges in Brain Mapping with Diffusion MRI

Peter J. Basser*

Section on Tissue Biophysics, Program on Pediatric Imaging and Tissue Sciences
Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, 20892-5772, USA

Abstract—Diffusion MRI provides new contrasts with which to image and map features of human brain microstructure and architecture in vivo. As powerful as this method currently is, it still has great potential to provide high-dimensional quantitative data with which to segment and cluster different brain regions, probe tissue microstructure and compartmental features, and provide clinicians with new quantitative imaging biomarkers with greater specificity and sensitivity.

I. INTRODUCTION

It has long been a dream of neuroscientists to discover the basis of thoughts and feelings, and of clinicians to understand and cure complex disorders of the brain. A common presumption in both communities is that if one elucidates the “wiring diagram” and the purpose of each cell type, one can understand brain function in its entirety. Until 1990, the primary ex vivo techniques to assess brain wiring via white matter pathways were post-mortem cell, histological and whole-brain slice analysis. In vitro methods primarily consisted of analyzing electrical activity of groups of cells, synecytia in cell cultures, and recordings from brain slices.

The advent and development of macroscopic-scale in vivo MRI methods created new opportunities to study structure, architectural organization, and function in the living brain. Functional MRI (fMRI) [1] was proposed to follow brain activation in gray matter through changes in blood oxygenation. Various structural MRI methods, like T₁-imaging showed white matter regions with new definition. Diffusion Tensor MRI (DTI) [2] revealed new microstructural features, particularly diffusion anisotropy in white matter, which provided a means to reconstruct brain white matter pathways in vivo.

II. DISCUSSION

One largely unrealized application of diffusion MRI data is to be able to study second and higher-order field properties inferred from a dense set of 10³x10³ mm³ isotropic voxels. These data could be used to segment and cluster the brain into distinct anatomical regions, e.g., identifying Brodmann-like areas in the cortex, and virtually dissecting and elucidating all of the white matter pathways connecting them. This is inherently a statistical task [3], but to date, no comprehensive, robust and reliable imaging sciences toolbox has been developed for hypothesis testing of second and higher-order tensor data.

Another challenge is to be able to “drill down into the voxel”—to be able to estimate or measure microstructural cell and tissue-scale features from macroscopic, voxel-averaged MRI measurements. Our approach is to use single [4] and multiple [5, 6] Pulsed-Field Gradient (PFG) MRI methods to be able to probe water displacements and their correlations in different intra- and extra-cellular compartments owing to the different MR signature diffusing water molecules produce in different restricted domains.

Imaging and mapping the entire net displacement distribution of water molecules, to which the diffusion tensor can be shown to be the lowest order Gaussian approximation, allows one to identify and resolve different populations of white matter fibers in complex brain regions. However, no comprehensive imaging sciences framework exists for statistically testing features of these net displacement distributions within the imaging volume.

An additional complexity is the multi-scale organization of the brain. White matter, gray matter and glia are all hierarchically organized, having a fractal-like character. Diffusion MRI captures net water displacements over a micron-length scale and a millisecond time scale. Needed are new ways to probe a larger range of length and time scales in a systematic and integrated way to be able to achieve a more comprehensive description of the structure and organization of the brain.

New approaches to assess exchange among diffusing molecules in different compartments will help us estimate their proximity and the permeability of boundaries separating them.

III. SUMMARY

Diffusion MRI is a powerful method that allows us to probe tissue microstructure by being able to measure the net displacement distribution of water molecules in different tissue compartments. While widely used clinically and in a myriad of neuroscience studies, Diffusion MRI continues to have potential to extract new and important tissue features. Opportunities exist to develop new imaging sciences methods and pipelines to analyze Diffusion MRI data for these ends.

REFERENCES

Non-Analytic Model-Based Reconstruction for Accelerated Multiparametric Mapping in MRI

Noam Ben-Eliezer¹, Daniel K Sodickson¹, Timothy Shepherd¹, Graham C Wiggins¹ and Kai Tobias Block¹
¹ The Bernard and Irene Schwartz Center for Biomedical Imaging, Department of Radiology, New York University School of Medicine, New York, NY 10016, USA

Abstract— This work presents a new methodology for quantitative multiparametric mapping that operates at clinically feasible timescales by employing advanced model-based and compressed-sensing (CS) reconstruction techniques. Accurate quantitative mapping of the T2 MRI relaxation time is achieved in vivo, through fusion of a novel signal-model calculation, with non-Cartesian radial sampling of the MR signal domain.

I. INTRODUCTION

Quantitative assessment of MRI transverse relaxation time (T2) can be used to detect pathological tissue changes in many clinical applications. Reliable mapping, however, is challenging as traditional single Spin-Echo (SE) protocols entail long scan times (10s of minutes), while faster, multi-SE (MSE) strategies, are inherently biased by stimulated echoes, non-rectangular slice profiles, and transmit field (B1+) inhomogeneities. A novel T2 mapping technique – the echo-modulation-curve (EMC) algorithm – is able to overcome these limitations using precise Bloch simulations of the pulse-sequence scheme. This work fuses the EMC algorithm with radial sampling schemes, thereby leveraging benefits such as improved scan-time efficiency, lower motion sensitivity, and acceleration using CS, or via mathematical modeling of the actual acquisition scheme. The resulting framework offers accurate quantification of both proton density (PD), and T2 information within clinically feasible scan times.

II. METHODS

The EMC algorithm is based on precise simulations of the prospective MSE protocol, repeated for a range of T2 (5...1000ms) and B1+ inhomogeneity (60...130 %) values. Reconstruction comprises of pixel-by-pixel matching of the acquired images’ time-series to the simulated database of EMCs, producing a unique [B1+,T2] value-pair for each pixel. Data were acquired on a whole-body 3T scanner for human brain and spinal cord in vivo, using golden-angle radial MSE, and reference Cartesian MSE and SSE protocols. Radially sampled data was processed using two approaches: (1) joint multi-coil CS reconstruction employing principle component analysis (PCA) as sparsifying transform, followed by matching the resulting image-series to the EMC database, and calculation of the PD map by extrapolating the first TE image to time t=0, based on the calculated T2 map. (2) model-based reconstruction, embedding the numerical EMC database into a forward signal-acquisition operator of the form:

\[ F_{TE,C,j}(\rho,T_2) = \sum_{\rho} \rho(\mathbf{r}) \cdot \text{EMC} \cdot B_1^j(\mathbf{r},T_2,TE) \cdot C_e(\mathbf{r}) \cdot e^{-k r^2} \]

This operator is integrated into a non-linear conjugate-gradient descent algorithm, implementing a three-parameter [T2, PD, B1+] optimization via iterative minimization of the cost function:

\[ \Phi = \frac{1}{2} \sum_{TE,C} \| (\hat{F}_{TE,C} - \text{Sig}_{TE,C}) \|_2^2 + \lambda_1 \| PD \|_2^2 + \lambda_2 \| B_1^j \|_2^2 \]

Cartesian data were processed using straightforward EMC fitting of the set of images. PD maps were calculated by extrapolating the first echo-time image to the point t=0 using the fitted T2 map.

III. RESULTS AND DISCUSSION

The figure compares T2 and PD maps of a human spinal cord in vivo, obtained with Cartesian MSE and with model-based radial MSE. As directly visible, the radial MSE protocol offers superior T2 and PD maps with better spatial definition and significantly lower blurring due to through-plane and in-plane CSF pulsations and flow. Extraction of the true tissue T2 values is furthermore achieved by employing the EMC-based relaxation model.

Sequential processing using CS followed by EMC matching (not shown) produced accurate morphological information, yet proved unstable with respect to regularization along the temporal domain, resulting in less deterministic T2 values.

The radial EMC reconstruction framework can be extended to model other contrasts such as T1, diffusion and T2*, multi-compartment T2 distributions, and various pulse-sequence schemes.

REFERENCES

Hierarchically Semiseparable Generalized Encoding Matrix Compression for Fast Inverse Imaging

Stephen Cauley,1,2 Kavin Setsompop,1,2 Elfar Adalsteinsson,1,4 Mark Griswold,3,5 and Lawrence Wald1,2,4

1Athinoula A. Martinos Center for Biomedical Imaging, Dept. of Radiology, Mass. General Hospital, Charlestown, Massachusetts, USA.
2Dept. of Radiology, Harvard Medical School, Boston, Massachusetts, USA.
3Dept. of Biomedical Engineering, Case Western Reserve University, Cleveland, Ohio, USA.
4Dept. of Electrical Engineering and Computer Science; Harvard-MIT Div. of Health Sci. and Tech., Cambridge, Massachusetts, USA.
5Dept. of Radiology, Case Western Reserve University and University Hospitals of Cleveland, Cleveland, Ohio, USA.

Abstract—Reconstruction of images from non-Cartesian data can be a computationally demanding problem. Iterative numerical solutions often involve repeated evaluation of Discrete Fourier operators, receive coil sensitivity profiles, and other physical MR parameters. In this work, we introduce a direct method for computing an approximate inverse of a generalized encoding matrix. The scalable compact representation can be used to quickly map acquired data directly to an image. This is accomplished through Hierarchically Semiseparable (HSS) modeling of the encoding matrix. Algorithms that use HSS modeling have been successfully employed to solve many large-scale problems in applied physics and mathematics. HSS modeling can be computed prior to data collection and is ideal for time series reconstruction, e.g. fMRI, cardiac imaging, and MR fingerprinting. We demonstrate the efficiency of our approach for the reconstruction of spiral data incorporating B0 distortion correction. Our method allows for 40x speed-up when compared to state-of-the-art iterative solvers.

I. MODEL-BASED IMAGE RECONSTRUCTION

Algorithms for image reconstruction of Cartesian data have been successfully incorporated into clinic settings. However, model-based reconstructions of non-Cartesian trajectories that include non-linear effects can be computationally prohibitive. A typical parallel imaging (PI) forward model for model-based image reconstruction [1] is as follows:

\[ s_l(t) = \int c_l(\vec{f}) e^{-j\omega_0(\vec{f})t} e^{-2\pi i k(\vec{f})t} f(\vec{f}) d\vec{f} \] (1)

The signal at channel l and time t is a function of the receive coil profiles \( c_l(\vec{f}) \), the off-resonance phase accrual \( \Delta \omega_0(\vec{f})t \), and the Fourier encoding \( 2\pi i k(\vec{f})t \). Iterative solvers [2-4] are conditioned Conjugate Gradient (PCG) [3], both tuned for 10\(^{-4}\) tolerance. Thus, HSS modeling can effectively capture generalized encoding of non-Cartesian data and should allow for near-real time reconstruction across many imaging applications.

REFERENCES

More is Indeed Different

Guang-Hong Chen
Department of Medical Physics and Department of Radiology, University of Wisconsin-Madison, Madison, Wisconsin 53705–2275

Abstract— X-ray medical imaging has evolved from 2D x-ray projection imaging, into 3D time-resolved (2+1 D) or volumetric (3D) imaging, and eventually 4D temporally or spectrally-resolved imaging, with the potential for even higher dimensions when other parametric maps are involved for specific applications. A natural question to ask is whether the data acquisition, image processing, and data storage methods in higher dimensions must be the same as that in the 2D case or whether something different can be done to acquire data, reconstruct images, and store/retrieve image data more efficiently without sacrificing diagnostic performance of a given imaging task? In this talk, examples of x-ray and CT imaging acquisitions will be discussed that demonstrate how intrinsic low dimensional structures in high dimensional imaging acquisitions can be exploited to reduce radiation dose, improve temporal resolution, and enable new clinical applications.

Since Roentgen discovered x-rays in 1895, x-ray projection images have been widely used in the diagnosis of musculoskeletal, pulmonary, and breast maladies among others. With the addition of a temporal dimension, 2+1 dimensional digital subtraction angiography (DSA) and x-ray fluoroscopic imaging methods were introduced in clinics in the 1980s. Another development was the addition of the x-ray spectral dimension to 2D projection imaging to generate another 2+1 dimensional technique: dual energy subtraction x-ray imaging. Since its introduction in the 1970s, x-ray computed tomography (CT) has also evolved from a 2D sliced imaging method to a 3D volumetric method with cone-beam acquisition geometry, generating CT image volumes with isotropic spatial resolution. In recent years, the incorporation of temporal dynamics has brought 4D cone-beam CT imaging to the forefront of basic science and clinical research. By incorporating knowledge of the x-ray spectra, 3D volumetric cone-beam CT or 4D time-resolved cone-beam CT (CBCT) can be brought into the domain of 4D or 5D imaging, respectively. This evolution in biomedical imaging, progressing from a problem with low dimensionality to problems with much higher dimensionality, resonates with mankind’s other efforts to understand the fundamental organizing principles of the physical world, which is an intrinsically higher dimensional research subject. In physics, there are two fundamentally different approaches that have been taken to understand the rich structures of the complex universe. One method, reductionism, is the belief that a thorough understanding of individual (low dimensional) components is sufficient to understanding an arbitrarily complex system. The alternative purports that Mother Nature is organized into hierarchical structures, and there are different organizing principles at each level of the hierarchy. While debates on the pros and cons of the two methods remain, it has become clear that lower dimensional structures do exist in physics, biology, and many other biomedical phenomena. Actually, the discovery and exploitation of low dimensional structures in a higher dimensional data set often facilitates understanding of complex natural phenomena; biomedical x-ray imaging is no exception.

In this talk, we revisit the image formation process in x-ray and CT imaging to demonstrate that a successful exploitation of low dimensional structures in higher dimensional x-ray imaging can be used to dramatically lower radiation dose and/or improve temporal and spatial resolution in specific applications. We will cover three applications: Ultra-low radiation dose x-ray DSA and high temporal resolution x-ray fluoroscopy imaging, high spectral resolution x-ray and CT imaging, and 4D time-resolved CBCT imaging.

A comparison between standard and ultra-low dose (1/20th dose level) DSA images using the conventional technique and the novel technique exploiting the low dimensionality of the dataset are shown here.

A comparison of CBCT myocardial perfusion data sets at extremely low radiation dose (25 mA exposure level) reconstructed with both FBP and the novel technique and a reference acquisition (500 mA exposure level) is shown here. In both this example and the previous DSA example, the ultra-low dose images generated with the novel technique are at least as good as the higher dose reference image.

In conclusion, these examples demonstrate that low dimensional structures do exist in x-ray high dimensional imaging, and appropriate exploit of these low dimensional structures may indeed be beneficial in application.
4D in Bio-microscopy: Marker-free Live Cell Tomography
Sebastien Equis*, Andreas Kern*, Yann Cotte* and Christian Depeursinge‡
* Nanolive SA, EPFL Innovation Park, 1024 Ecublens, Switzerland,
‡ King Abdullah University of Science and Technology (KAUST), 23955-6900 Thuwal, Saudi Arabia

Marker-free nanoscopy allows to “see” the activity inside of a living cell in 3D. Using only harmless light, it enables observing the living cells label-free, non-invasively, and interference-free. The combination of holography with rotational scanning opens new dimensions for live cell monitoring. The cell’s tomographic reconstruction allows measuring cellular processes with real-time kinetics down to sub-cellular scale. Since the cells are not manipulated to introduce a label, it gives the possibility to measure the response of authentic cell’s physiological activity.

1. INTRODUCTION

We present recent applications of non-invasive live cell tomography, which enables the acquisition of label-free quantitative phase images of biological samples at high frame rate through interferometric detection. As Quantitative Phase Nanoscopy (QPN) is a wide field imaging technique, it differs from most recently developed microscopy techniques, which are based on pointwise laser-scanning systems.

QPN is based on the combination of holography and rotational scanning, which allows accessing the world of live cell tomography.

Holography offers a unique means to probe cells in their native environment: Label-free, non-invasive, manipulation-free, and interference-free. Rotational scanning allows real 3D reconstructions, noise robustness, and a resolution far beyond the accepted limit for light.

Based on this coherent light detection scheme, it is possible to construct a three-dimensional super-resolved QPN image, which opens the world of seeing living cells with nanometric details. An additive feature is based on a synthetic aperture approach where several images with different illumination angles are acquired and combined. The resolution can then be further improved by employing deconvolution based on the specific coherent transfer function of QPN [1], leading to a lateral resolution below 100 nm [2]. This approach yields three-dimensional refractive index maps of unlabeled samples, which can be acquired at frame rates in the second range, by taking advantage of the high acquisition speed of QPN (in the millisecond range).

In the context of cell microscopy, the signal is proportional to the intracellular optical density, hence a physical marker. Since the cells are not manipulated in any way to introduce a label, it gives the user the possibility to measure the response of authentic cell’s physiological activity. It is therefore possible to monitor at high speed the dynamics of cells under stimulation.

An example is to use the signal as an indicator of the transmembrane water fluxes occurring during cellular neuronal activity. The signal can therefore be used as a label-free biological indicator linked to the activity of specific membrane receptors, where the dynamics of cortical neuron cells can be monitored at high-speed. It is also possible to couple such measurement with fluorescence signals [3], such as intracellular ionic indicators, making it possible to deepen the understanding of underlying cellular processes linked with transmembrane water fluxes.

It allows measuring cellular processes with real-time kinetics enabling multi-parameter analysis at single cell and sub-cellular scale.

We would like to sincerely thank the Laboratory of Lymphatic & Cancer Bioengineering (LLCB) in EPFL for their support.

REFERENCES
MR guided focused ultrasound (MRgFUS) is a minimally invasive method of targeted tissue thermal ablation

Pejman Ghanouni, MD, PhD*

*Assistant Professor of Radiology, Stanford University Medical Center, 1201 Welch Road, Stanford, CA  94305

I. INTRODUCTION

In this talk, I will cover the physical principles of MRgFUS. I will summarize the treatment protocol, and review applications in gynecology (1), oncology (2-4), and neurosurgery (5-11). Finally, future directions and limitation of the technology will be discussed.

A. Figures

Figure 1: MRgFUS devices include A) an in-table system used for fibroid and other body applications, B) a conformal system, and C) a transcranial brain system.

Figure 2: MR images from MRgFUS ablation of A) a uterine fibroid, B) a pelvic osseous metastasis, and C) a nucleus in the halamus.

REFERENCES

Coming Challenges in Medical Device Development; Alternative, Counter-Current Ideas in Disease Physiology May Provide Ideas for Medical-Device-Based Solutions that will Run Countercurrent to Existing, Drug Based Therapies

Charles Hennemeyer MD,
University of Arizona, Department of Medical Imaging, Arizona 85658

Abstract— Medical Device development has long been a source of novel devices that lead to new, minimally invasive surgical procedures. Many of the devices are mechanical, and address clearly visible anatomic problems. However, future disruptive medical device innovations are more likely to arise from an understanding of physiology rather than anatomy, and as such the challenge for existing innovators will be to better understand the physiologic basis for disease rather than simple anatomic sequela of visible disease.

I. INTRODUCTION

Disruptive innovation has long pervaded the culture of Interventional radiology. Finding a less invasive, simpler means for solving a medical problem has always been in vogue; for example use a needle probe to deliver energy and burn or destroy a tumor through the skin, or use specially designed catheters to reopen a vessel slowing blood which caused a stroke. As medical imaging gave better visualization of disease, so did the response of medical device innovators, giving us the tools to alter the course of disease. Diagnostic imaging technology gave way to improved procedural guidance, and with it a new crop of therapeutic medical devices. This interplay between diagnostic imaging technology and procedure guidance has become the basis for our profession, Ultrasound, CT and MRI were all developed for diagnosis, but are now often used for therapeutic guidance as well. However, the ability to detect pathology, and the ability to effect a change in that pathology are entirely dependent on task specific, medical devices, such as a microwave ablation probe for tumors, a suction catheter for heart and brain clots, an embolic implantable coil for bleeding, or radioactive micro-particles to radiate tumors from within.

The development of task specific mechanical devices has been the driver for much medical innovation, minimally invasive image guided procedural solutions. Up until now many medical problems had simple mechanical solutions, a narrow vessel needs an inflatable balloon to stretch open a stenosis and then a metal stent to keep it held open. Unfortunately, many of these structural, anatomically visible problems have been solved with a multitude of device options. Disruptive medical devices of the future that truly change human disease, will come from an understanding of the physiologic basis of disease, rather than the anatomic sequelae of disease. Successful device-based therapies will in one, or a few sessions replace the need for expensive, chronic drug therapies. Hypertension may become treatable with a neuro-modulation device rather than a lifetime of drug treatment. The economics of a one-time large expense versus a lifetime of chronic drug expenses will drive these changes. Existing medical device development is scheduled for a change. Reactive procedures that are designed for symptomatic relief are likely to evolve into device based solutions for permanent cures. Successful medical device innovators of the near future will be those who are willing to rethink apparently flawed models used as our physiologic understanding of diseases. New ideas and experiments based on these ideas are likely to run counter current to existing understanding of common diseases such as hypertension and diabetes which serves as the basis for current drug therapy. In this session I will attempt to give examples of these trends, and offer guidance to those who seek truly innovative solutions.

II. KEY POINTS FOR SUCCESS

- Know the Pain Point or Problem to be Solved
- Know the current state of medical practice, what needs replacing or “disruption”
- Know the Manufacturing technology available and its limitations
- How to form a small cohesive group of innovation minded experts, medical engineering and funding
- Most difficult is select fewer projects, know accurately which are likely to succeed and finally see them to completion amongst a sea of distractions
- Cannot be done in isolation, go at it alone mentality
- Support at major institutions is in the ability to associate with others, not supply funding
- Must be a passion, cannot perseue an idea only for financial gain
- IP ownership, worry and fear can be stifling
ICTGV regularization for image sequence reconstruction of accelerated dynamic MRI

Martin Holler∗, Kristian Bredies∗, Karl Kunisch†, Matthias Schlögl†, and Rudolf Stollberger†

† Institute of Medical Engineering, Graz University of Technology, Kronegasse 5/H, A-8010 Graz.

Abstract—We propose a variational method for reconstructing image sequences from accelerated dynamic MR data. Our approach uses the infimal convolution of total generalized variation functionals (ICTGV) as regularization. The resulting minimization problem is solved using a primal-dual algorithm. The method is applicable to coil-weighted MRI and, by exploiting temporal redundancy, allows a high reconstruction quality even in the case of strong undersampling.

I. INTRODUCTION

For a large number of applications in image processing, regularization is a key issue. Given a regularization functional $R$ and a forward model of the data acquisition process $F$, a generic approach for stable inversion of the data acquisition process is to solve

$$
\min_u R(u) + \lambda \| F(u) - d_0 \|^p, \tag{1}
$$

where $d_0$ is the given data, $\| \cdot \|^p$ a suitable discrepancy and $\lambda$ a parameter defining the trade off between data fidelity and regularization. In the situation of still image reconstruction, suitable choices of $R$ are well investigated [1] and allow for example the reconstruction of MR images from highly undersampled data [2].

When it comes to dynamic imaging, the situation is different as a suitable extension to spatio-temporal regularization is not obvious. Nevertheless, applications such as dynamic cardiac MRI require a significant acceleration of the acquisition process, hence a method of reconstruction from undersampled data is highly desirable. We propose such a method by incorporating the recently introduced spatio-temporal regularization functional of [3] in a model as (1). As can be seen in Figure 1, this allows to obtain high quality reconstructions from dynamic MR data for acceleration factors above 10. The success of our method was also confirmed recently in the context of the 2013 ISMRM challenge [4] for reconstruction of dynamic cardiac MRI, where the second place was achieved.

II. MATHEMATICAL BACKGROUND

For still image reconstruction, the Total Generalized Variation (TGV) functional [1] is an effective regularizer in many different situations. By incorporating higher order derivatives, it models images to be piecewise smooth while still being convex and computationally tractable.

When aiming to extend TGV regularization to the spatio-temporal situation, an important observation is that the relation between spatial and temporal regularity gives an additional degree of freedom. In [3], this has been exploited by defining a class of spatio-temporal regularization functionals that automatically balance between spatial and temporal regularity in an optimal way. A particular instance of this class of functionals is given as

$$
\text{ICTGV}_2(u) = \min_v \text{TGV}_2^2(u - v) + \text{TGV}_2^2(v), \tag{2}
$$

where the $\text{TGV}_2^2$ are spatio-temporal TGV functionals that employ a different weighting of spatial and temporal derivatives. It has been shown in [3] that ICTGV$_2^2$ is a convex functional that allows a significant improvement for image sequence reconstruction when compared to straightforward extensions of spatial regularizers.

III. APPLICATION FOR ACCELERATED DYNAMIC MRI

Given the subsampled MR data $(d_{t,c})$ for coil $c$ and frame $t$, our reconstruction method amounts to solving

$$
\min_u \text{ICTGV}_2^2(u) + \lambda \sum_{t,c} \| F_t(b_c, u_t) - d_{t,c} \|^2, \tag{3}
$$

where $F_t$ is a Fourier transform with time dependent masking and $b_c$ are the complex valued coil sensitivities. As the coil sensitivities are not given, we reconstruct them a-priori using time-averaging and spatial regularization and hence achieve a linear forward mapping.

IV. NUMERICAL SOLUTION

The minimization problem (3) can be reformulated as a convex-concave saddle-point problem having a simple structure and solved by employing the primal-dual algorithm of [5] with an adaptive stepsize strategy. The iteration steps of the resulting algorithm consist of simple, voxel-wise operations together with the evaluation of linear operators and global convergence can be assured.

REFERENCES

Digital Chest Tomosynthesis: Clinical Applications and Technical Obstacles

Gregory Kicska, M.D. Ph.D.*,
* Department of Radiology, University of Washington, Seattle, Washington, 98159, USA

Abstract—Digital chest tomosynthesis (DTS) is a relatively new imaging modality that produces cross-sectional images similar to CT but with a higher in-plane resolution and lower dose, similar to chest radiography. This unique combination of dose, resolution and cross-sectional images suggests that it may be suited to specific clinical applications. Although the outlook for DTS is promising, there are further advances that may expand its utility and increase its acceptance.

I. OVERVIEW

DTS images are produced by first obtaining multiple radiographic projections of the patient at incremental angles and then performing filtered back projection. Compared to CT were the x-ray detector and source orbit around the patient in concert, in DTS the detector is fixed and only the source moves in a straight line. A major disadvantage of DTS is that because only 30-60 of projections are used for reconstructing slices, the nominal slice thickness is large with a broad sensitivity profile, resulting in images that contain blurred out-of-plane anatomy (structural noise). DTS structural noise is anisotropic and a function of the source motion, which complicates a physicians ability to distinguish artifact from pathology. Methods that reduce the structural noise through image processing would likely increase the diagnostic performance of DTS.

Often, patients are imaged with DTS after they have been imaged with CT or radiography. It would be advantageous if prior imaging could be used to tailor the image acquisition technique so that structural noise would be reduced in the area of interest.

The decreased radiation dose of DTS is a prominent driver in its utilization. DTS delivers approximately 5% of the dose of chest CT and about twice the dose of PA & lateral chest radiography. Although this reduction is dose is substantial, the gap between CT and tomosynthesis has been narrowing with the introduction of iterative reconstruction techniques. If the same reconstruction techniques could be applied to DTS, this would maintain its dose advantage.

In conclusion, DTS is a promising modality that is gaining acceptance among clinicians. Further advances reducing structural noise and dose would likely expand its utility and increase its acceptance.

Fig. 1. Comparison of Radiography, DTS and CT for the detection of pulmonary nodules (arrow).
Spatial encoding with generalised magnetic field shapes

Kelvin Layton∗, Stefan Kroboth ∗, Feng Jia ∗, Sebastian Littin ∗, Huijun Yu∗, and Maxim Zaitsev∗
∗ Department of Radiology, Medical Physics, University Medical Centre Freiburg, Freiburg, Germany

Abstract—The possibilities for spatial encoding in magnetic resonance imaging (MRI) are greatly expanded with the use of multiple nonlinear magnetic fields. This work presents a method to optimise the encoding in this situation based on the variance of the reconstructed pixels. Experimental results suggest that highly accelerated imaging is possible with reduced artifacts compared to traditional techniques.

I. INTRODUCTION

Spatial information can be encoded in a magnetic resonance signal through RF pulses, receive coils and magnetic field gradients. Traditionally, linear magnetic field gradients are used to encode spatial information with functions derived from the Fourier basis. Magnetic field gradients that vary non-linearly across the field-of-view have recently been explored for increased imaging speed [1], targeted resolution enhancement [2] and reduced field-of-view imaging [3]. Multiple nonlinear gradients together with receive arrays provide unprecedented flexibility for spatial encoding but also make trajectory design difficult. In this work, we aim to optimise the trajectory based on the covariance of the reconstructed pixels.

II. THEORY

The signal of the $i$th receive coil with sensitivity $c_i$ is

$$s_i(t_i) = \int m(x)c_i(x)e^{i k(t)(x)} dx,$$

(1)

where $k(t)$ is a vector of gradient moments (the trajectory) and $\psi(x)$ is a vector of gradient fields at position $x$. A matrix equation can be written from (1) by defining a signal vector of all times and coils, $s$, and magnetisation vector, $m$, such that $s = Em$.

We adopt a Bayesian formulation where the pixel covariance is recursively computed after each measurement. We consider a generalised projection, parameterised by $\theta_n$, consisting of a prephasing moment and readout direction for each gradient channel. The covariance matrix is given by the Kalman filter equations,

$$P_{n+1}(\theta_n) = P_n - P_n E(\theta_n)^*(E(\theta_n)P_n E(\theta_n)^* + \Sigma)^{-1} E(\theta_n)P_n.$$

(2)

The projections are optimised based on knowledge of the previous measurements by minimising $\text{trace}(P_{n+1}(\theta_n))$.

III. METHODS

Experiments were performed on a modified 3T scanner (Siemens, Erlangen, Germany) with a nonlinear gradient insert (Resonance Research Inc, Billerica, USA) to provide simultaneous control of the five encoding fields (after slice selection). Fig. 1 illustrates the fields measured using the phase difference between gradient echo images with and without a gradient pulse. Data from a tube phantom were acquired with an eight channel receive array, calibrated with another gradient echo image. The optimisation was solved for each new projection using an evolutionary algorithm that starts with 15 randomly chosen ‘parent’ parameters, each of which produces a ‘child’ parameter vector by random mutation. The parameters producing the lowest variance are selected as the parents for the next generation. The algorithm was run for 10 generations. A total of 32 projections were optimised with 256 time points per projection. Samples along projection lines were acquired with pure phase encoding to mitigate gradient delay and eddy-current effects. Cartesian Fourier encoding with 32 phase encodes and 256 readout points (8-fold undersampling) was acquired for comparison. The field-of-view was 224 mm. Images were reconstructed with a conjugate gradient algorithm with 30 iterations.

IV. RESULTS & DISCUSSION

Fig. 2 displays the reconstructed images for undersampled Fourier encoding and optimised higher-dimensional encoding. The Fourier encoding generates strong artifacts since aliasing cannot be resolved by the receive coil array (equivalent to large g-factor). Conversely, the optimised encoding produces a clearer image with incoherent noise-like artifact.

Experimental results suggest that multiple nonlinear encoding fields can reduce artifacts for highly accelerated imaging. The variance metric used for trajectory design includes both the gradient fields and receive coils. Trajectory optimisation will be crucial as the number gradient channels increase (e.g. in a matrix coil) to properly exploit the additional flexibility.

REFERENCES


ACKNOWLEDGEMENT

This work was in part supported by European Research Council (ERC) grant 282345 ‘RANGEmri’.
Multidimensional Imaging: A Path to High Resolution and High Speed through Subspaces

Zhi-Pei Liang
Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801

Abstract—A major obstacle hindering the practical applications of high-dimensional imaging is due to the well-known “curse of dimensionality” (i.e., the amount of data required increases exponentially as the imaging dimension increases). This talk will discuss a new approach to high-dimensional imaging based on low-rank tensors (partially separable functions). These models significantly reduce the imaging complexity, making high-resolution, high-speed multidimensional imaging possible.

I. INTRODUCTION

High-dimensional imaging is desirable for a range of biomedical applications. For example, in magnetic resonance imaging (MRI), one can perform 4-dimensional dynamic imaging (for cardiovascular imaging), 5-dimensional spectroscopic imaging (for biochemical imaging of cancer), and 6-dimensional diffusion spectral imaging (for brain mapping). A major obstacle hindering the practical applications of these potentially powerful imaging techniques is due to the well-known “curse of dimensionality” (i.e., the amount of data required increases exponentially as the imaging dimension increases). This talk will discuss a new approach to high-dimensional imaging, exploiting the sparsity and partial separability (or low-rank tensor structure) of the underlying multivariate image function. Low-rank tensors significantly reduce the imaging complexity, making high-resolution, high-speed multidimensional imaging possible.

II. OVERCOMING THE CURSE OF DIMENSIONALITY WITH LOW-RANK TENSORS

Let \( f(x_1, x_2, \ldots, x_d) \) represent a general multivariate image function, where \( d \) is the imaging dimension and \( x_n \), for \( n = 1, \ldots, d \), represents any variable (e.g., space, time, chemical-shift frequency, diffusion spectral frequency, relaxation parameters, etc.). In conventional imaging methods based on the Shannon-Nyquist sampling theory, we often need \( O(N^d) \) measurements to recover such a function with \( N \) being the number of encodings needed for one dimension. The exponential growth is known as the curse of dimensionality. To reduce the imaging complexity, we express \( f \) as a sum of \( L \) separable functions [1]:

\[
f(x_1, x_2, \ldots, x_d) = \sum_{\ell=1}^{L} \phi_{\ell,1}(x_1) \phi_{\ell,2}(x_2) \cdots \phi_{\ell,d}(x_d).  \tag{1}
\]

This model can also be viewed as a rank-\( L \), order-\( d \) tensor [2], which assumes that \( f \) resides in a \( L \)-dimensional tensor-product subspace, effectively reducing the imaging complexity from \( O(N^d) \) to \( O(dLN) \). Sparsity can also be imposed on \( \phi_{\ell,n}(x_n) \) to further reduce the number of measurements needed to recover \( f \). Various extensions of (1) exist for more efficient representation. For example, we can put \( x_1, \ldots, x_d \) into \( d \) separable groups \( x_1, \ldots, x_d \) with \( x_i = (x_{i1}, \ldots, x_{im}) \) such that \( \sum_{i=1}^{d} 1_{mi} = d \), and

\[
f(x_1, x_2, \ldots, x_d) = \sum_{\ell=1}^{L} \phi_{\ell,1}(x_{1}) \phi_{\ell,2}(x_{2}) \cdots \phi_{\ell,d}(x_{d}).  \tag{2}
\]

The validity and effectiveness of these models for high-dimensional imaging will be discussed in detail during the talk.

III. EXAMPLES OF ACCELERATED MULTIDIMENSIONAL IMAGING

Low-rank tensors (partially separable functions) have been successfully used to accelerate high-dimensional MRI experiments, producing impressive results. Two examples are shown below, illustrating the ability of the models for high-speed cardiac imaging (Fig. 1) and for high-resolution spectroscopic imaging (Fig. 2), by highly sparsely sampling \((k, \ell)\)-space (see [3] and [4] for a more detailed discussion of the imaging techniques and the results).

![Fig. 1. High-speed cardiac imaging with sparse sampling: (a) high-resolution anatomical image, and (b) temporal changes along the marked line in (a).](image1)

![Fig. 2. High-resolution spectroscopic imaging with sparse sampling: (a) high-resolution metabolic map (NAA) from a phantom, and (b) localized spectrum from a selected pixel.](image2)

IV. CONCLUSION

Multidimensional imaging functions can often be efficiently modeled by low-rank tensors (partially separable functions). These models significantly reduce the imaging complexity, making high-resolution, high-speed multidimensional imaging possible. Tensor model-based imaging methods may prove useful for a number of practical applications ranging from high-dimensional dynamic imaging to spectroscopic imaging.

REFERENCES

Quantitative Elasticity Imaging With Acoustic Radiation Force: Methods and Clinical Applications

Kathryn R Nightingale  
Biomedical Engineering Department, Duke University, Durham, North Carolina, 27708.

Abstract— the goal of elasticity imaging is to non-invasively visualize the inherent mechanical contrast of tissues with higher resolution and depth penetration than is possible with manual palpation. As such, these methods introduce a mechanical perturbation of tissue, and then use conventional imaging modalities to monitor the tissue response. An overview of ultrasonic, acoustic radiation force impulse (ARFI) shear wave imaging methods is provided herein. This class of imaging methods has recently become commercially available and is finding widespread clinical use in the context of hepatic fibrosis staging based upon changes in group shearwave speed. The use of higher order material models (i.e. viscoelastic, anisotropic, and nonlinear) is an active area of investigation with the goal of identifying additional image parameters with the potential to serve as biomarkers for various disease states. Several approaches to ARFI and shear wave imaging and their application to different clinical scenarios will be discussed.

Introduction: All elasticity imaging methods introduce a mechanical excitation into the body, and monitor the tissue response to that excitation using an imaging method, typically either ultrasound or magnetic resonance imaging [1]. One of the fundamental challenges with elasticity imaging methods that use external mechanical excitation (static or dynamic) is coupling the excitation into the organ/structure of interest. The use of focused acoustic energy (i.e. acoustic radiation force impulse, ARFI) eliminates this challenge by delivering a spatially localized mechanical excitation directly within the organ of interest [2]. ARFI elasticity imaging methods use a single ultrasound transducer to both generate the focused ARFI excitation, and to monitor the resulting tissue dynamic displacement response. Data is processed to portray either relative differences in displacement amplitude or by quantifying the propagation speed of the radiation force induced shear waves. Commercial realizations of these technologies employ simple material models to quantify group shearwave speed; an active area of investigation involves the use of higher order material models to identify additional image parameters with the potential to serve as biomarkers for various diseases.

Qualitative ARFI Imaging: ARFI imaging creates images of tissue displacement within the region of excitation (ROE). These images provide information about relative differences in tissue stiffness, and have spatial resolution that is comparable to that of B-mode, often with greater contrast, providing matched, adjunctive information to that obtained with B-mode imaging. ARFI imaging involves multiple ARFI interrogations throughout the region of interest. The displacement data from the different radiation force excitations are synthesized into a single dataset representing the displacement response throughout the tissue volume to all the individual excitations at a given time after excitation, with darker regions representing smaller displacements (i.e., stiffer tissues). The contrast in ARFI images is related to the underlying tissue mechanical contrast, in addition to the relative size and position of the excitation ROE and the structures being imaged. Fig. 1 presents matched B-mode, ARFI, and quantitative SWEI (as explained below) prostate coronal images obtained in vivo under an IRB approved protocol.

Quantitative ARFI Imaging: Shear wave elasticity imaging (SWEI) [3] involves monitoring the propagation speed of the shear waves generated by ARFI excitations. In ultrasound, time-of-flight methods are employed to evaluate the propagation speed at laterally offset locations from the ROE. Under various material model assumptions, the measured shearwave speeds can be used to reconstruct the tissue elastic properties. Tradeoffs exist between estimation precision and resolution, and optimizing both remains an active area of research. Given the broadband nature of ARFI induced shearwaves, dispersion analyses (Figure 2) are possible using a range of spectral methods and a range of material models (i.e. linear elastic, viscoelastic, anisotropic, nonlinear, etc). An overview of current research in these areas will be presented.

References:

Low-rank plus sparse dynamic MRI: separation of background and dynamic components and self discovery of motion

Ricardo Otazo
Department of Radiology, New York University School of Medicine, New York, NY, USA

Abstract — New uses of sparsity in addition to increasing imaging speed are presented for dynamic MRI. Compressed sensing and low-rank matrix completion are combined using a low-rank plus sparse (L+S) model to reconstruct dynamic MR images with automatic separation of background (L) and dynamic (S) components. An extension of the L+S model named motion-guided L+S with self discovery of motion is also presented for free-breathing imaging.

Compressed sensing (CS) exploits the natural fact that image pixels are correlated to reconstruct data sampled below the Nyquist-Shannon rate [1-2]. The application of CS to Magnetic Resonance Imaging (MRI) [3] is a major breakthrough in terms of imaging speed and demonstrated great potential to overcome some of the major limitations in terms of spatial and temporal resolution, volumetric coverage and sensitivity to motion. MRI presents favorable conditions for the application of CS, since (a) medical images are naturally compressible by using appropriate sparsifying transforms, such as wavelets, finite differences, and many others, and (b) MRI data are acquired in the spatial frequency domain (k-space) rather than in the image domain, which facilitates the generation of incoherent aliasing artifacts via random undersampling of Cartesian k-space or the use of non-Cartesian k-space trajectories. The application of CS is particularly significant for high-dimensional imaging, such as 3D imaging and dynamic imaging, which present increased compressibility and call for increased imaging speed.

A closely related idea to compressed sensing is matrix completion [4], where the missing or corrupted entries can be recovered if the matrix presents low-rank conditions. In a similar fashion to sparse images, which only have few large coefficients, low-rank matrices, which only have a few large singular values, depend upon a smaller number of degrees of freedom and undersampling becomes possible. Low-rank matrix completion can be applied to MRI by sorting the images as a column in the matrix. For example, each frame of a dynamic image can represent a column in the matrix.

Most of the work on compressed sensing and low-rank matrix completion is focused to improve imaging speed. However, they can be potentially used for other applications, such as separation of sources of information and compensation for data corruption (motion, relaxation, etc.). For example, the low-rank plus sparse (L+S) reconstruction method can reconstruct undersampled dynamic MRI data with automatic separation of background (L) and dynamic (S) components [5] (Fig. 1). This method is based on the low-rank plus sparse matrix decomposition or robust principal component analysis (RPCA) and represents a combination of compressed sensing and parallel imaging.

This talk will cover the basics of L+S matrix decomposition and its application to reconstruct undersampled dynamic MRI data with separation of background and dynamic components. An extension of the L+S approach that self discovers motion fields to improve the performance in the presence of organ motion will be also discussed. Besides motion compensation, the information contained in the motion fields can be of potential clinical use. Reconstruction of highly-accelerated dynamic MRI data corresponding to cardiac perfusion, cardiac cine, time-resolved peripheral angiography, and abdominal perfusion using Cartesian and golden-angle radial sampling will be presented to show feasibility and general applicability of the L+S method.

REFERENCES

MR Susceptibility Imaging and Mapping

Jürgen R. Reichenbach*

* Medical Physics Group, Department of Diagnostic and Interventional Radiology, Jena University Hospital – Friedrich-Schiller University, D-07743 Jena, Germany

Abstract— Magnetic susceptibility is a fundamental physical property, which can significantly affect MR image contrast. Susceptibility weighted imaging (SWI) has made it possible to highlight tissue structures and compounds that can be difficult to be detected by conventional MRI by exploiting the susceptibility differences between tissues via phase effects at long gradient echo times. In recent years, impressive progress has been made in the quantification of tissue susceptibility in vivo by solving the inverse problem (QSM) to recover the susceptibility distribution of the human body from the measured local field that is expressed by the convolution of the susceptibility distribution with the magnetic field generated by a unit dipole. Clinical susceptibility mapping studies are now available and the application of QSM is growing rapidly.

I. INTRODUCTION

Magnetic susceptibility is a fundamental physical property of all materials that can significantly affect MR image contrast. Using T₂*-weighted gradient echo imaging, variations of tissue magnetic susceptibility typically lead to local signal cancellations in magnitude images and cause frequency shifts. In SWI, the magnitude and the phase are combined into a single composite image, called susceptibility weighted image to enhance the contrast in a qualitative non-local manner between tissues of differing magnetic susceptibility [1,2]. More recently, impressive progress has been made in even quantifying tissue susceptibility in vivo by solving the inverse problem (QSM) to retrieve information about the underlying magnetic susceptibility distribution from MRI phase measurements, which makes it possible to calculate iron content or oxygen saturation in the human brain [3-5].

II. FROM MR PHASE TO QUANTITATIVE SUSCEPTIBILITY MAPPING

MR phase images contain essential information about local magnetic susceptibility sources in the tissue. When processed appropriately, phase images can be used to reconstruct the local susceptibility sources creating a new type of contrast in the form of susceptibility maps. This is achieved by, first, estimating the magnetic field distribution from raw MRI phase data, second, eliminating so-called background field contributions that result from magnetization outside of the imaging field-of-view in MR-invisible areas and, third, solving the inverse problem from field perturbation to magnetic susceptibility. Each of the first two processing steps needs to be carried out with utmost rigor because the final inversion step is highly sensitive to noise and errors in the input field pattern. Starting from the forward modeling approach the field variation can be expressed by

\[ \varphi(r) = \gamma \cdot B_0 \cdot TE \cdot \left[ \chi(r) \otimes \frac{1}{4\pi} \cdot \frac{3 \cos^2 \theta - 1}{r^3} \right] \] (1)

where \( \theta \) is the azimuthal angle in the spherical coordinate system, \( r' = x^2 + y^2 + z^2 \), and \( \otimes \) represents the convolution operator. Susceptibility maps may then be derived by inverting the equation in k-space

\[ \chi(r) = \frac{1}{\gamma \cdot B_0 \cdot TE} \cdot \text{FT}^{-1} \left[ \text{FT} \left[ \varphi(r) \right] \left[ \frac{1}{3} \frac{k_2}{k_0} \right] \right] \] (2)

where FT represents the Fourier transform and the denominator represents the dipole kernel in the Fourier domain. As seen from Eq. (2) the inverse problem is ill-posed and prone to severe noise amplification and streaking artifacts due to the presence of zeroes at a cone in the Fourier representation of the unit dipole kernel. One solution for avoiding division by zero is to threshold the dipole kernel, other solutions include multiple angle acquisition or regularization. Overall, reconstruction methods have been greatly improved to give better recovery of tissue susceptibility data for QSM, and various clinical applications have been pursued.

REFERENCES


Magnetic Resonance Fingerprinting: Beyond Parameter Mapping to Clinical Application

Nicole Seiberlich
Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, USA

Abstract—Magnetic resonance fingerprinting (MRF) has been demonstrated for the simultaneous collection of maps of various MRI contrast-generating parameters. MRF has been shown to provide high quality relaxation parameter maps in the brain and body, and a database of normative values is currently being established. However, MRF can be used to solve technical challenges beyond parameter mapping. MRF has been demonstrated for the reduction of SAR in high field imaging, and can be employed to separate multiple components in a single voxel. Because the encoding gradients can be selected arbitrarily in MRF, they can be used to play music while imaging, thereby improving patient comfort. These applications will lead to clinical utility far beyond parameter mapping.

Magnetic Resonance Fingerprinting (MRF) [1] is a novel technique which has been used to generate quantitative maps of several MRI contrast mechanisms simultaneously, including proton density, T₁ and T₂ relaxation times, off-resonance [1], diffusion [2], and perfusion [3]. MRF works by using MRI pulse sequences with pseudo-random acquisition parameters to make the MRI signal sensitive to several of the parameters listed above. With a knowledge of the specific pulse sequence used and the range of physiologically relevant values for the sensitive parameters, all possible signal evolutions for the experiment can be calculated using the Bloch equations. Pattern matching algorithms are then used to compare this dictionary of signal evolutions to those seen in the acquired data, enabling the determination of the appropriate parameter values for each voxel. Often only a small subset of the MRI data is collected after each contrast block, and the pattern matching algorithm is tasked with recognizing the appropriate signal evolution even in the presence of aliasing artifacts.

MRF was first demonstrated for parameter mapping for normal brain tissue [1], and its application in clinical settings is currently under evaluation. Initial results in brain tumor patients have shown that it is possible to differentiate tumors from normal tissue and each other using MRF data [4]. Early studies are underway to generate quantitative parameter maps in other areas of the body [5], where motion and off-resonance effects play a bigger role and require additional technical developments. While the ability to quantify multiple parameters simultaneously is highly beneficial, MRF has potential for applications beyond parameter mapping. For instance, because the signal evolution “fingerprint” in each voxel is a weighted sum of the fingerprints of different tissue components, it is possible to resolve these multiple material components from single voxels [6]. By using an MRF scheme based on the QUEST sequence, which generates the maximum possible imaging signals with minimal RF energy, SAR deposition can be reduced for high-field parameter mapping [7]. Intriguingly, because the trajectory used for MRF can be freely selected, it is possible to design imaging gradients that play music while efficiently collecting quantitative data [8]. These novel features of MRF make this method powerful for areas outside of parameter mapping, including subvoxel tissue characterization, high-field imaging, and scanning of pediatric or uncooperative patients.

Because of these novel features, MRF has the potential to change the way imaging is performed clinically. By examining large numbers of patients using MRF, databases of normative and pathological parameter values could be developed; this is currently impossible due to the variation in parameter values depending on the scanner set-up, quantification method, and technical challenges in some parts of the body. Then, instead of relying on qualitative weighted images, radiologists would have access to quantitative maps as well as normal and abnormal ranges, taking much of the subjectivity out of diagnosing disease. Whereas currently multiple image contrasts must be acquired, in MRF only one scan is required, and all image contrasts can be reconstructed retrospectively. Using the same data, tissue microstructure could be assessed by separating voxels with multiple tissue components. This single MRF scan could be performed at high fields, yielding high SNR and therefore precise measurements, and set to music, significantly reducing the discomfort felt by many patients. Thus, while MRF is generally thought of as a parameter mapping technique, its clinical utility goes well beyond quantitative imaging.

References

Wave-CAIPI for an order of magnitude acceleration in MRI acquisition

Kawin Setsompop*†

*Athinoula A. Martinos Center for Biomedical Imaging, Dept. of Radiology, MGH, Charlestown, MA, USA.
† Harvard Medical School, Boston, MA, USA.

Abstract—The “controlled aliasing” methods, or CAIPI has been proposed to improve performance of parallel imaging acquisition in MRI where data which are encoded in Fourier (k-space) is undersampled, creating image aliasing that can be resolved using spatial information from multiple receiver coils. With CAIPI, the k-space under-sampling scheme is modified to increase the distance between aliasing voxels and allow for better use of coil sensitivity information. This work presents some of the latest technology advancement in CAIPI, in Wave-CAIPI, which when used together with modern large area receiver coil enable an order of magnitude acceleration in MRI acquisition at low SNR and artifact penalties.

Introduction—Wave-CAIPI [1] synergistically combines k-space under-sampling of 2D-CAIPI and a modified version of Bunched Phase Encoding [2,3] to create a staggered cork-screw trajectory (Fig. 1, left), which maximizes the distance between aliased voxels and can be adapted for both 3D and Simultaneous Multislice (SMS) accelerated acquisitions. Fig. 1, right shows the inter-slice shift along y between aliased slices in z caused by staggered k_x-k_y sampling of 2D-CAIPI and the position-dependent voxel spreading along x resulting from G_x and G_y gradients of the cork-screw trajectory. These effects combine to create a complex aliasing pattern in an R = 3x3 (9x) accelerated 3D-GRE acquisition as shown, which maximizes the distance between aliased voxels in all spatial directions, x, y and z and enables optimal use of 3D coil sensitivity information in Parallel imaging reconstruction. Using 32-chn receiver coil information at 3 T, this complex aliasing was resolved using a SENSE-based reconstruction that incorporates information about the modified k-space sampling and trajectory to provide high quality reconstruction at 9x acceleration (Fig.1 far-right) with maximum g-factor SNR loss over the image volume of just 9%.

In addition to 3D-GRE, Wave-CAIPI was also applied to SMS RARE imaging (a.k.a. TSE or FSE). Fig. 2 shows results of 15× Wave-CAIPI for SMS-T2w-RARE imaging at 1-mm isotropic resolution at 3T, acquired using MultiPIPS refocusing with low SAR [4] and a 32-chn receiver coil. Reconstructed images of a representative MB-15 slice group are shown (top-left), along with retained-SNR maps in three orthogonal views (top-right), where high-quality images and high retained-SNR are observed. Also shown are 3-mm ‘slice’ reformatted images in three orthogonal views, demonstrating multi-orientation reformatting into standard radiological viewing planes. Good reconstructions are observed with minor streaking artifacts along the slice (A-P) direction, likely from small errors in our slice-by-slice coil sensitivity estimation, which are amplified in the reconstruction at this high acceleration. Imaging parameters were: matrix size = 256×192 with 255 slices, TR/TE = 4000 ms/89 ms, bandwidth = 130 Hz/pixel, and Echo Train Length (ETL) = 8. The data were acquired in 90 s using a prototype SMS-RARE sequence, where echo spacing and gradient timing have not been fully optimized. With additional engineering, we expect that this 1-mm isotropic acquisition can be completed in 1 minute.

References:
Shape and size distributions can have a direct impact on the underlying functional and biological aspects of cells. In the Central-Nervous-System (CNS), the axonal size distribution determines conduction velocity [1], and disease-induced aberrations in cellular dimensions often lead to detrimental consequences [2]. Noninvasive characterization of the salient features of cellular distributions via existing methods [3–5] requires very strong gradient amplitudes and multidimensional experiments, as well as extensive tissue modeling. Here, we harness Non-uniform Oscillating-Gradient Spin-Echo (NOGSE)-MRI (Fig. 1A) – a methodology probing diffusion dynamics recently shown to exhibit extraordinary sensitivity towards compartmental dimensions (l) that varies as l$^{6}$ [6,7] – for reporting on size distributions noninvasively and in a simple, one-dimensional fashion. Simulations and experiments in yeast cells corroborate NOGSE’s sensitivity; contrasts derived from NOGSE-MRI in mouse brains show promise for future in vivo applications.

Axonal calibers and conduction velocities are directly related [1] and severe functional deficits can arise from aberrations in cellular-scale dimensions [2]. NOGSE-MRI [6,7] is an emerging technique capable of robustly probing cellular-scale dimensions with excellent sensitivity. To test NOGSE’s sensitivity, NOGSE signals were simulated in lognormal size distributions incorporating H$_2$O undergoing restricted diffusion. NOGSE’s l$^{6}$ sensitivity can be highly informative (Fig. 1B): the ground-truth structural distributions are faithfully reconstructed from NOGSE responses (Fig. 1B, inset). To experimentally corroborate these important trends, NOGSE experiments were performed on yeast cells on a 9.4 T Bruker Avance scanner operating at a $^1$H frequency of 400.17 MHz using a micro5 imaging probe. Ground truth cellular size distributions obtained from ancillary light microscopy were found to be in excellent agreement with the NOGSE-derived size distribution parameters (data not shown).

To demonstrate the potential of NOGSE-MRI to provide contrast in the brain, NOGSE-EPI experiments in the perfused mouse brain were performed with $T_{\text{NOGSE}}=30$ ms, $N=8$, $G=57.6$ G/cm. The ensuing data were analyzed pixel-by-pixel and fitted to lognormal distributions, and their mean, peak, and width were mapped. The results shown in Figures 2A–C represent such size distribution maps in the white matter (corpus callosum). Notably, NOGSE clearly segments this tissue into its well-known [4] five distinct anatomical regions; furthermore, all of the corpus–callosum’s size distribution features well-known from histology [8], such as the increase in distribution width from genu to splenium, are reproduced by the NOGSE maps. These striking contrasts are not confined to white matter: when experiments were performed in a coronal orientation (Fig. 2D–F), NOGSE appears to parcellate cortical gray matter into distinct layers, faithfully depicting this major GM structural hallmark. NOGSE’s ability to resolve size distributions is likely reflecting its previously reported [6,7] l$^{6}$ sensitivity, coupled to its multi-frequency probe of diffusion-driven dynamics [6]. This enables the direct imaging of cellular size distribution parameters, with minimal tissue modeling and with robustness; however, a notable limitation is the inherent bias towards longer $T_2$ species. Still, the detailed maps derived from NOGSE-MRI are highly promising for studying cellular-scale aberrations in white matter tissues upon disease, as well as modifications in the tissue’s ultrastructure upon normal CNS processes such as plasticity.

**Fig. 1.** (A) NOGSE-EPI sequence used in this study, comprising a constant-time, constant N, variable-delay probe of diffusion played out at a constant $T_{\text{NOGSE}}$, and constant gradient (G). (B) Simulations showing the marked effects of size distributions on NOGSE signals. The inset shows extracted distributions from simulated curves (symbols) and ground truth (solid curves).

**Fig. 2.** NOGSE EPI experiments in the ex-vivo mouse brain. (A–C) Maps of distribution mean, peak, and width, respectively, in images masked for the corpus callosum. (D–F) Maps of distribution mean, peak, and width.

Towards Quantification of the Brain’s Sheet Structure in Diffusion MRI Data

Chantal M.W. Tax*,1, Tom C.J. Dela Haije†,1, Andrea Fuster†, Remco Duits†, Max A. Viergever‡, Evan Calabrese2, G. Allan Johnson1, Luc M.J. Florack1, and Alexander Leemans*†‡

* Image Sciences Institute, University Medical Center Utrecht, Utrecht, the Netherlands. † Imaging Science & Technology, Eindhoven University of Technology, Eindhoven, the Netherlands. ‡Center for In Vivo Microscopy, Duke University Medical Center, Durham, North Carolina, USA

Abstract—The recent hypothesis on the occurrence of sheet structure in the brain has posed many questions to the diffusion MRI (dMRI) community as to whether this structure actually exists and can be measured with dMRI. In this work, we exploit the capability of the discrete Lie bracket to infer information on the existence of sheet structure in real dMRI data.

I. INTRODUCTION

The question whether our brain’s structure is best reflected by a three-dimensional intrinsic sheet or not remains controversial. Over the last years, several groups including our own have worked on the development of methods for the detection of sheet structure in artificial vector fields, to vector fields derived from diffusion MRI data. We have shown that spatial resolution and the curvature influence the ability to detect sheet structures. We present preliminary but promising results of a high resolution mouse brain, which shows the presence of a sheet formed by two main fODF peaks in correlation with a diffusion tensor mapping (DTI) geometry map.

II. THEORY AND METHODS

A. Lie bracket theory

The Lie bracket $\{V, W\}_p$ is a measure of the deviation from $p$ when trying to move around in an infinitesimal loop along the integral curves of the fields $V$ and $W$ (Fig. 1). If and only if $\{V, W\}_p = 0$, the vector fields form a sheet at $p$. The Lie bracket can be approximated by various difference vectors $r_{\ell, n}$ according to

$$r_{\ell, n}(x_1, y_1) \approx h_1, h_2[V, W]_{p} + \Delta(h_1, h_2),$$

where $h_1$ and $h_2$ are walking distances and $\Delta(h_1, h_2)$ is an error term that scales with $h_1$ and $h_2$. See references [5,7] for details.

B. Implementation and experiments

Starting from point $p$ in the data, we assign two fiber orientation distribution function (fODF) peaks [4] as representative members of vector fields $V$ and $W$.

We use nearest neighbor streamline tractography using steps of size $\Delta h$ to find the difference vectors. Each difference vector is based on 4 consecutive tractography paths $\{V, W\}_p$. At each streamline step the local vectors are assigned to one of the fields based on their cosine similarity with the vectors at the previous position. Tractions passing through voxels with only one peak are ignored. Subsequently, $\{V, W\}_p$ is calculated as an indicator of sheet structure in a simulated dMRI dataset that was known to represent a sheet [5,8] and in high resolution mouse brain data.

III. RESULTS AND DISCUSSION

Complementing previous results on the method’s dependence on resolution [5], Fig. 2 shows the curvature of the presentation of the sheet structure in the data, we extend this approach to real dMRI data.

RESULTS AND DISCUSSION

In this work we extend the analysis of the Lie bracket normal component as a tool for the detection of sheet structure in artificial vector fields, to vector fields derived from diffusion MRI data. We have shown that spatial resolution and the curvature influence the ability to detect sheet structures. We present preliminary but promising results of a high resolution mouse brain, which shows the presence of a sheet formed by two main fODF peaks in correlation with a diffusion tensor mapping (DTI) geometry map.

REFERENCES


IV. CONCLUSION

In this work we extend the analysis of the Lie bracket normal component as a tool for the detection of sheet structure in artificial vector fields, to vector fields derived from diffusion MRI data. We have shown that spatial resolution and the curvature influence the ability to detect sheet structures. We present preliminary but promising results of a high resolution mouse brain, which shows the presence of a sheet formed by two main fODF peaks in correlation with a diffusion tensor mapping (DTI) geometry map.

Fig. 1 Walking loop with $\{\delta_x^p, \delta_y^p\}$ to the end point. Difference vector $r_{\ell, n}$ approximates $\{V, W\}_p$.

Fig. 2 Diffusion data generated from vector fields $V = \{0,1,0\}$ and $W = \{0,1,0\}$ defined on domain $[-10 \text{mm}, 10 \text{mm}]^2$ with 1 mm voxel size and the curvature in point $p = 0$. These have zero Lie bracket by design and are locally tangent to the surface $x(y) = 0.5(y^2 + x^2)$. $\Delta h_{\text{max}} = 1 \text{mm}$ and $\Delta h = 0.4 \text{mm}$. For $\kappa > 2$, $\{V, W\}_p$ deviates significantly from 0. The number of paths used (numbers above each graph) is lower for higher $\kappa$, partially causing the increased standard deviation.

Fig. 3 Mouse brain dMRI data with $b = 4000 s/mm^2$, measured with 120 different directions and 119 $\times 119 \times 80$ images, voxel size 0.043 mm isotropic. (a) Direction encoded fractional anisotropy map. (b) $\{V, W\}_p$ between two largest fODF peaks, with $\Delta h = 0.043 \text{mm}$ and $\Delta h_{\text{max}} = 5$. The blue location shows a region with low $\{V, W\}_p$, the yellow location one with noisy $\{V, W\}_p$. (c) The corresponding DTI geometry map.

These authors contributed equally to this work.

In this work we consider the difference vectors $\{\delta_x^p \cdot \delta_y^p \cdot \delta_y^p \cdot \delta_y^p\} (p) - p = (\delta_x^p \cdot \delta_y^p \cdot \delta_y^p \cdot \delta_y^p) (p)$ and $(\delta_x^p \cdot \delta_y^p \cdot \delta_y^p) (p) = (\delta_x^p \cdot \delta_y^p) (p)$, where the flow operator $\delta_x^p (p)$ denotes moving a distance $x$ along the integral curve of vector field $X$ starting from point $p$.
Quantification in Cardiovascular MRI

Shreyas S. Vasanawala*
* Department of Radiology, Stanford University, Stanford, CA 94305.

Abstract Obtaining parameters of cardiovascular physiology is a critical aspect of MRI exams. The parameters include flow through vessels, flow through valves, and stroke volumes. However, examinations are quite lengthy, and consistently acquiring accurate and reliable assessments can be difficult. This talk will provide an overview of the challenges of obtaining accurate quantification in the setting of congenital heart disease.

I. INTRODUCTION
Imaging plays a critical role in diagnosis and monitoring of cardiovascular disease. More so than in oncologic imaging, the goals of cardiovascular imaging entail obtaining quantitative data. Multiple imaging modalities serve to obtain complementary information. For instance, echocardiography enables quantification of left ventricular function. Nuclear scintigraphy quantifies regional myocardial perfusion and split pulmonary perfusion. Catheter angiography provides blood pressure gradients. Computed tomography more recently may provide fractional flow reserve to the myocardium.

II. QUANTIFICATION IN MRI
MRI may provide multi-faceted quantification for cardiovascular disease. Metrics provided by MRI include blood flow in great vessels, ventricular size, ventricular function, and valvular regurgitation. Other parameters potentially provided by MRI include pressure gradients, fibrosis through T1 mapping, and perfusion. Although together these parameters potentially provide a comprehensive quantitative assessment of cardiovascular disease, two barriers reduce widespread use of MRI.

First, conventionally performance of MRI is quite lengthy to perform and time-consuming to interpret. This originates from a sequential and iterative approach to data acquisition and a tedious approach to post-processing of images. Second, the precision and accuracy of quantitative methods in MRI has been hard to achieve in routine practice. For example, quantification of flow often has errors of 40%.

This presentation will focus on efforts to enhance the speed of imaging, efficiency of interpretation, and quantitative validity of cardiovascular MRI exams. In particular, fast methods of quantification of cardiovascular flow and function will be presented, and placed in the clinical context of congenital heart diseases.

Fig. 1. Example of 15 year old boy with tetralogy of Fallot after repair. (a) Systolic image show high velocities in the pulmonary artery (white arrow), indicative of pulmonic stenosis and a jet of tricuspid regurgitation (black arrow). (b) Diastolic image shows pulmonic regurgitation as well (arrow). Multi-planar reformations (c, d, e, and f) allow optimal segmentation of pulmonary artery yield a flow curve (g) for calculation of regurgitant fraction of 48%. Magnitude data can be reformatted in the long axis in both diastole (h) and systole (i), and segmentation of the right ventricle performed to determine ventricular volume and ejection fraction. Alternatively, reformats can be obtained in the traditional short axis plane (j,k). Total acquisition time for the 4D flow was 7 minutes, 29 seconds.
Imaging with X-ray Modulated Nanoparticles

Ge Wang, PhD; ge-wang@ieee.org
Biomedical Imaging Center, CBIS/BME, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

Abstract — Imaging depth of modern optical microscopies has been fundamentally limited to about a millimeter for a biological sample due to strong scattering of light. X-ray microscopy can resolve spatial details of few microns deeply inside a sample but contrast resolution is poor to reveal cellular features. To enrich biological contrast, nanoparticles become essential in basic research and molecular medicine. Nanoparticles can be functionalized as imaging probes, of which nanophosphors are an emerging category with a great potential. Various types of nanophosphors emit luminescent light upon x-ray excitation, or allow energy storage with x-ray excitation and subsequent luminescence emission triggered by visible/near-infrared light. In this talk, we present imaging methods based on x-ray modulated nanophosphors for deep penetration and fine resolution. Other physical couplings are also discussed with nanoparticles including but not limited to nanophosphors. This work is in collaboration with University of Georgia, GE GRC, University of Albany, and Purdue University.

I. BACKGROUND OF OPTICAL IMAGING

Over the past decade, optical imaging has been a rapidly evolving field, especially for microscopic and preclinical studies. Imaging depth of modern optical microscopies, such as two-photon and photoacoustic methods, has been fundamentally limited to about a millimeter for a biological sample due to strong scattering of light. In preclinical scenarios, while diffuse optical tomography (DOT) targets tissue properties, fluorescence molecular tomography (FMT) and bioluminescence tomography (BLT) focus on cellular and molecular features. Because of optical diffusion, image resolution has been fundamentally limited to ≥1mm. X-ray luminescence computed tomography (XLCT) [1] was developed to excite nanophosphors locally through x-ray collimation for improved image resolution. However, due to physical and mechanical reasons, a collimated x-ray beam is insufficiently narrow and further broadened by scattering. Hence, resolution of XLCT is only ~1mm.

Over the past year, we have been coupling x-rays and luminescent light via nanophosphors. First, we proposed x-ray focusing with a zone plate, a polycapillary lens, or a grating, instead of an x-ray collimator, for precise localization of luminescence light sources. Second, we utilized energy-storing characteristics of novel nanophosphors [2] to decouple x-ray stimulation and luminescence emission for multiplexed luminescence data collection.

In this talk, we will explain our recent results along this line, namely, x-ray micro-modulated luminescence tomography (XMLT) [3, 4] and stored luminescence CT (SLCT) [5]. We will also report our explorative efforts on nanoparticle-enabled x-ray MRI (NXMRI) and x-optogenetics. The emphasis will be on unconventional multi-physics couplings.

II. COUPLING OF X-RAYS & LUMINESCENCE

In a good analogy, the aforementioned new optical imaging methods can be distributed along the unit circle on the complex plane in Fig. 1. First, the purely optical tomography methods, such as DOT, FMT and BLT, are the starting point featured by isotropic scattering. Then, XLCT uses a pencil x-ray beam for both excitation and localization of nanophosphors, but the pencil beam diverges due to x-ray scattering. On the other hand, XMLT treats x-rays as waves instead of particles, and focuses x-ray waves onto a fine focal spot deeply into tissue, promising higher image resolution than XLCT. If we position XLCT at the location (1, 0) (1 symbolizes a pencil beam), XLCT should be put at the location (-1, 0) (the x-ray beam is converging, not diverging; i.e., convergence is opposite to divergence). More interestingly, if energy-storing nanophosphors are introduced, there will be a phase difference between x-ray excitation and luminescence emission (which can be induced, within a decent time window, by visible or infrared light after x-ray excitation). Depending on how we manipulate x-rays (either collimation or focusing), we have two modes of SLCT, which can be marked at (0, i) and (0, -i) respectively. Our numerical and experimental results will be described on these imaging modes.

Fig. 1. Four combinations of x-ray excitation and luminescence emission.

III. COUPLING OF X-RAYS & MAGNETIC RESONANCE

It is our claim/hypothesis (provisional patent protection by RPJ) that x-ray induced energy storage into nanoparticles such as nanophosphors will have an effect similar to chemical shifts, and be observable via magnetic resonance. We are working to couple x-ray excitation and magnetic resonance for high-order imaging capabilities. For example, focused x-rays will naturally introduce an energy gradient in a dual cone geometry, and high resolution and fast speed could be outcomes. Also, we are evaluating the feasibility of coupling x-rays, nanophosphors and optogenetics. Collaboration and feedback will be welcome.

REFERENCES

Semi-Analytic Iterative Framework for TV Penalized Cone-beam CT Reconstruction

Jong Chul Ye and Eunhee Kang

Abstract—This paper introduces a novel 3D iterative CT reconstruction framework for low dose imaging, whose computational complexity is comparable to analytic reconstruction. The main breakthrough comes from the introduction of the differentiated backprojection (DBP) preconditioner, originated from analytic cone-beam reconstruction theory. We also explain an optimization method that simplifies the recon flow.

I. INTRODUCTION

X-ray computed tomography (CT) is now considered as the first examination in hospital before treatment plan is made by doctors. Since the introduction of CT, radiologists have been struggling to balance image noise with radiation dose for decades. Even though iterative CT reconstruction techniques have been investigated by many researchers in academia, only recently after the FDA approval of the GE Healthcare’s Model Based Image Reconstruction (MBIR) technology, the iterative reconstruction becomes main stream technology in industry. However, one of the most critical technical huddles in its wide deployment is the excessive computational burden due to the multiple applications of the forward and backward projectors. Even with the introduction of the graphic processor unit (GPU), the current reconstruction speed for helical diagnostic CT geometry is still far from routine use.

In this paper, we introduce a novel semi-analytic iterative reconstruction method, where a novel preconditioner derived from analytic cone-beam reconstruction theory is used to eliminate the necessity of multiple applications of forward and backward projectors. The new algorithm enables high quality imaging with computational complexity comparable to the analytic methods.

II. DBP PRECONDITIONED TV FORMULATION

A noisy x-ray CT measurement from linear attenuation coefficient distribution $\mu$ is described by

$$y(a, \theta) = D_\theta(a, \theta) + w$$

where $w$ denotes the additive noise, and $D_\theta(a, \theta) = \int_{-\infty}^{\infty} \mu(a+\theta t)dt$, and $\theta \in S^2$ refers the x-ray photon propagation direction; and $a \in \mathbb{R}^3$ refers the x-ray source location. The estimation problem of $\mu$ from (1) is often solved using total variation (TV) penalized least squares [1]. Here, a preconditioner is often employed to accelerate the convergence [1]. In general, a preconditioner should not amplify the noise contribution. Toward this goal, we propose a differentiated backprojection (DBP) operator as a preconditioner. More specifically, in 3D helical CT geometry, the DBP operator on a PI line is given by [2]

$$DBP\{ \} = \int_{-\lambda}^{\lambda} d\lambda' \frac{1}{\| \mathbf{x}' - a(\lambda) \|} \frac{\partial \{ \} }{\partial \mu} |_{\mu = \lambda}$$

where $a(\lambda^{-}), a(\lambda^{+})$ denote the two source locations that connect a PI line. Note that the differentiation step in DBP only amplifies the noise twice if we use one-step differentiation filter. Furthermore, backprojection step is basically summing up the contribution from multiple views, which reduces the overall noise contribution. Moreover, it is well-known that $DBP\{ D_\theta \}$ reduces the problem into separable to the 1-D Hilbert transform on PI lines, which significantly reduces the complexity [2]. Thus, we argue that the DBP is a good pre-conditioner. This observation leads to the following formulation:

$$\min_{\mu} \sum_i ||g_i - \mathcal{H} \mu_i||^2 + \lambda TV(\mu)$$

where $g_i$ denotes the DBP data on the $i$-th PI line, and $\mu_i$ refers the linear attenuation coefficients on the corresponding PI-line, and $\mathcal{H}$ is the 1-D Hilbert transform.

III. OPTIMIZATION METHOD

Note that in (2) the 3D TV penalty makes the problem non-separable. To facilitate this optimization problem, we employ the alternating direction method of multiplier (ADMM) algorithm with variable splitting [1]. More specifically, (2) is converted into the following constrained optimization problem:

$$\min_{(p_i, \nu)} \sum_i ||g_i - \mathcal{H} p_i||^2 + \lambda TV(\mu)$$

subject to $p_i = T_i \nu$, $i = 1, \ldots, n$.

where $T_i$ denotes the cartesian to the $i$-th PI line transform. The resulting ADMM subprograms consists of alternating applications of the following steps:

$$\mu^{(k+1)}_i = \arg \min_{\mu} \lambda TV(\mu) + \frac{\rho}{2} \sum_i ||p_i^{(k)} - T_i \mu + \eta_i^{(k)}||^2$$

$$p_i^{(k+1)} = \arg \min_{p_i} ||g_i - \mathcal{H} p_i||^2 + \frac{\rho}{2} ||p_i - T_i \mu^{(k+1)} + \eta_i^{(k)}||^2$$

$$\eta_i^{(k+1)} = \eta_i^{(k)} - \rho (p_i^{(k+1)} - T_i \mu^{(k+1)})$$

where the first update is simply 3D TV denoising, and the second step can be simply reduced to a weighted average of the TV denoised image and the 1-D inverse Hilbert transform of DBP on each PI line. Hence, the non-separable 3D TV penalty is only for image denoising, which significantly reduces overall computational time.

IV. RESULTS AND CONCLUSIONS

Using GPU implementation, the reconstruction time for 512×512×256 volume from 1024×135 detectors with 800 views and 3 helical rotations took about 43 seconds for backprojection filtration method [2], whereas the proposed semi-analytic iterative method using 15 iterations took only 88 sec. Therefore, even with the sophisticated regularization and optimization algorithm, the additional complexity of the proposed algorithm is only 45 seconds, which is several order smaller compared to those of fully iterative approaches.

REFERENCES
