Proceedings of the International BASP Frontiers Workshop 2017

January 29 - February 3, 2017

Villars-sur-Ollon, Switzerland

BASP Frontiers Workshop 2017

Chair

Prof. Yves Wiaux, Heriot-Watt University, UK

Area Chairs

Prof. Jason McEwen, University College London, UK [astro-imaging]
Prof. Michael Lustig, UC Berkeley, USA [bio-imaging]
Prof. Philip Schniter, Ohio State University, USA [signal processing]

Session Organisers

Astro-imaging

Dr. Tim Eifler, NASA JPL/Caltech, USA
Prof. Anna Scaife, University of Manchester, UK
Prof. Benjamin Wandelt, Lagrange Institute Paris, France

Bio-imaging

Prof. Karla Miller, Oxford Centre for Functional MRI of the Brain, UK
Prof. Ricardo Otazo, Department of Radiology, New York University, USA
Prof. Daniel K. Sodickson, Department of Radiology, New York University, USA
Prof. Martin Uecker, University Medical Center Göttingen, Germany

Signal processing

Dr. Emilie Chouzenoux, Université Paris-Est Marne-la-Vallée, France
Prof. Mário A. T. Figueiredo, Universidad de Lisboa, Portugal
Prof. Felix Krahmer, Technical University of Munich, Germany
Prof. Jean-Christophe Pesquet, CentraleSupélec, France

Local Organizing Committee

Dr. Arwa Dabbech, Heriot-Watt University, UK
Prof. Jean-Philippe Thiran, Ecole Polytechnique Fédérale de Lausanne, Switzerland
Dr. Boris Leistedt, New York University, USA
Mr. Vijay Kartik, Ecole Polytechnique Fédérale de Lausanne, Switzerland
Ms. Rosie De Pietro, Ecole Polytechnique Fédérale de Lausanne, Switzerland
Ms. Lynn Smith, Heriot-Watt University, UK
Dr. Alexandru Onose, Heriot-Watt University, UK

Foreword

Astronomy and biomedical sciences find common roots in their need to process acquired data into interpretable signals or images. In these applications of signal processing the complexity of data to be acquired and processed is constantly increasing, thus challenging signal processing theories. 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 three workshops in 2011, 2013, and 2015, the BASP Frontiers 2017 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 2017 takes 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.

On Behalf of BASP Frontiers organising committee

Prof. Y. Wiaux

Workshop Chair

Contents

A distributed algorithm for wide-band radio-interferometry

Abdullah Abdulaziz, Alexandru Onose, Arwa Dabbech and Yves Wiaux
Institute of Sensors, Signals and Systems, Heriot-Watt University, Edinburgh EH14 4AS, UK

***Abstract*—We propose a scalable, randomised algorithm to solve the inverse imaging problem in wide-band radio-interferometry. In the big-data context of the next-generation radio-telescopes, the scalability is paramount due to the large-scale of the problem to be solved. The proposed method distributes the data measured at each frequency and processes it in parallel. We showcase the algorithm capabilities through realistic simulations.**

I. Introduction

In wide-band radio-interferometry (RI), the electromagnetic signal coming from the sky is probed by an array of antennas, at multiple frequencies ν_i, and is correlated at each antenna pair, producing radio measurements $\boldsymbol{y}_i \in \mathbb{C}^M$ for each band ν_i. To recover a hyper-spectral image of the sky, an ill-posed problem has to be solved, which under simplifying assumptions, can be modelled as $\mathbf{Y} = \boldsymbol{\Phi}(\mathbf{X}) + \mathbf{N}$, where $\mathbf{Y} = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_b) \in \mathbb{C}^{M\times b}$ denotes the wide-band measured data at b bands, corrupted by additive white Gaussian noise $\mathbf{N} = (\boldsymbol{n}_1, \dots, \boldsymbol{n}_b) \in \mathbb{C}^{M\times b}$ and $\mathbf{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_b) \in \mathbb{R}_+^{N\times b}$ is the unknown hyper-spectral image. The linear operator $\boldsymbol{\Phi}(\mathbf{X}) = \left([\boldsymbol{\Phi}_i \boldsymbol{x}_i]_{i=1:b}\right)$ models the acquisition process, that is an incomplete Fourier sampling.

II. Convex minimisation problem

We assume a linear mixture model for the image cube and solve a convex minimisation problem imposing low-rankness, join-sparsity and positivity of the image cube $\mathbf{X}$ [1]. We introduce multiple data fidelity terms defined for each frequency band, to achieve a high degree of parallelism. The minimisation problem can be defined as

$$\min_{\mathbf{X}} f(\mathbf{X}) + \mu g_1(\boldsymbol{\Psi}^\dagger \mathbf{X}) + g_2(\mathbf{X}) + \sum_{i=1}^{b} h_i\left(\overline{\boldsymbol{\Phi}}_i(\mathbf{X})\right), \qquad (1)$$

with the functions involved: $f = \iota_{\mathcal{D}}, \mathcal{D} = \mathbb{R}_+^{N\times b}$ accounting for the positivity constraint; $g_1(\mathbf{Z}) = \|\mathbf{Z}\|_{\ell_{2,1}}$ imposing joint-sparsity in a concatenation of wavelet basis $\boldsymbol{\Psi}$; $g_2(\mathbf{Z}) = \|\mathbf{Z}\|_*$ imposing low-rankness onto the desired solution; $h_i = \iota_{\mathcal{B}_i}, \mathcal{B}_i = \{\mathbf{Z} \in \mathbb{C}^{M\times b} : \|\mathbf{Z} - \mathbf{Y}_i\|_F \leq \epsilon_i\}$ enforcing data fidelity by constraining the solution to belong to the ϵ_i-balls defined by the known noise statistics. We denote with $\mathbf{Y}_i = (\alpha_1 \boldsymbol{y}_1, \dots, \alpha_b \boldsymbol{y}_b) \in \mathbb{C}^{M\times b}$ the measurement matrix active only at the band ν_i such that $\alpha_j = 0, \forall j \neq i$. The associated linear operator is $\overline{\boldsymbol{\Phi}}_i(\mathbf{X}) = ([\alpha_j \boldsymbol{\Phi}_i \boldsymbol{x}_i]_{i=1:b})$ with $\alpha_j = 0, \forall j \neq i$.

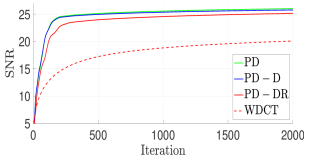
To solve (1), we use a randomised primal-dual algorithm [2] that relies on forward-backward (FB) iterations to manage the non-smooth functions. The algorithmic structure has been employed for distributed, single-band imaging [3] and for non-distributed wide-band imaging [1]. The operations are detailed in Algorithm 1. All the proximal FB steps have closed-form solutions. The proximity operator for the joint-sparsity prior is a row-wise soft-thresholding operation, for row k defined as $(\boldsymbol{\mathcal{S}}_\alpha^{\ell_{2,1}}(\mathbf{Z}))_{k,:} = \frac{\bar{\boldsymbol{z}}(\|\bar{\boldsymbol{z}}\|_{\ell_2} - \alpha)}{\|\bar{\boldsymbol{z}}\|_{\ell_2}}$ if $\|\bar{\boldsymbol{z}}\|_{\ell_2} > \alpha$ and $(\boldsymbol{\mathcal{S}}_\alpha^{\ell_{2,1}}(\mathbf{Z}))_{k,:} = 0$ otherwise. The nuclear norm produces the soft-thresholding of the eigenvalues of $\mathbf{Z}$, $\boldsymbol{\mathcal{S}}_\alpha^*(\mathbf{Z}) = \mathbf{H}_1 \boldsymbol{\mathcal{S}}_\alpha^{\ell_1}(\boldsymbol{\Sigma}) \mathbf{H}_2^\dagger$. Data fidelity is enforced by the projections $\boldsymbol{\mathcal{P}}_{\mathcal{B}_i}$ onto the ϵ_i sized ℓ_2 balls, for each band and positivity is imposed via the projection $\boldsymbol{\mathcal{P}}_{\mathcal{D}}$ onto the positive orthant $\mathcal{D}$.

Algorithm 1 Randomised PD for distributed WB RI.

given $\mathbf{X}^{(0)}, \tilde{\mathbf{X}}^{(0)}, \mathbf{V}_1^{(0)}, \mathbf{V}_2^{(0)}, \mathbf{U}_1^{(0)}, \dots, \mathbf{U}_b^{(0)}, \mu, \tau, \sigma_1, \sigma_2, \sigma_3$
repeat for $t = 1, \dots$
  generate active set $\mathcal{A} \subset \{1, \dots, b\}$
  do in parallel
    $\mathbf{V}_1^{(t)} = \boldsymbol{V}_1^{(t-1)} + \boldsymbol{\Psi}^\dagger \tilde{\mathbf{X}}^{(t-1)} - \boldsymbol{\mathcal{S}}_{\mu/\sigma_1}^{\ell_{2,1}}\left(\mathbf{V}_1^{(t-1)} + \boldsymbol{\Psi}^\dagger \tilde{\mathbf{X}}^{(t-1)}\right)$
    $\mathbf{V}_2^{(t)} = \mathbf{V}_2^{(t-1)} + \tilde{\mathbf{X}}^{(t-1)} - \boldsymbol{\mathcal{S}}_{1/\sigma_2}^{*}\left(\mathbf{V}_2^{(t-1)} + \tilde{\mathbf{X}}^{(t-1)}\right)$
    $\forall i \in \mathcal{A}$ **do in parallel**
      $\mathbf{U}_i^{(t)} = \mathbf{U}_i^{(t-1)} + \overline{\boldsymbol{\Phi}}_i\left(\tilde{\mathbf{X}}^{(t-1)}\right) - \boldsymbol{\mathcal{P}}_{\mathcal{B}_i}\left(\mathbf{U}_i^{(t-1)} + \overline{\boldsymbol{\Phi}}_i\left(\tilde{\mathbf{X}}^{(t-1)}\right)\right)$
    end
  end
  $\mathbf{X}^{(t)} = \boldsymbol{\mathcal{P}}_{\mathcal{D}}\left(\mathbf{X}^{(t-1)} - \tau\left(\sigma_1 \boldsymbol{\Psi} \mathbf{V}_1^{(t)} + \sigma_2 \mathbf{V}_2^{(t)} + \sigma_3 \sum_{i=1}^{b} \overline{\boldsymbol{\Phi}}_i^\dagger\left(\mathbf{U}_i^{(t)}\right)\right)\right)$
  $\tilde{\mathbf{X}}^{(t)} = 2\mathbf{X}^{(t)} - \mathbf{X}^{(t-1)}$
until convergence

III. Simulations and results

We simulate a wide-band image cube following the spectral curvature model $\mathbf{x}_i = \mathbf{x}_0 \left(\nu_i/\nu_0\right)^{-\boldsymbol{\gamma} + \boldsymbol{\beta}\log(\nu_i/\nu_0)}$, where $\mathbf{x}_0$ is a 256×256 sized image of a radio region in the M31 galaxy; $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are the spectral index maps of size N and modelled as correlated Gaussian random fields. The wide-band cube is generated for $b = 16$ bands in the range $[1.4, 2.8]$ GHz. The wide-band data are simulated using realistic uv-coverages from the VLA array-configuration with $M = 33120$ measurements at each band and are corrupted with zero-mean Gaussian noise with an input signal-to-noise ratio (SNR) of 30 dB.



The figure reveals the SNR evolution for the different algorithms. We can see that the non-distributed primal-dual algorithm denoted by PD [1] and the distributed version PD-D exhibit comparable behaviour, reaching a SNR = 26 dB. For the proposed distributed randomised algorithm PD-DR, we fix the probability of selecting an active subset $\mathcal{A}$ from the full data $\mathbf{Y}$ to 0.5. This has the advantage of lower infrastructure and memory requirements, at the expense of an increased number of iterations to achieve convergence. Also, when compared to the approach proposed in [4] and denoted by WDCT, our proposed algorithm presents superior performance; PD-DR reaches a SNR = 25 dB that is 5 dB higher than WDCT.

References

[1] A. Abdulaziz, A. Dabbech, A. Onose, and Y. Wiaux, in *Proceedings of the 24th EUSIPCO*, 2016, pp. 388–392.
[2] J.-C. Pesquet and A. Repetti, *J. Nonlinear Convex Anal.*, vol. 16, no. 12, pp. 2453–2490, 2015.
[3] A. Onose, R. E. Carrillo, A. Repetti, J. D. McEwen, J.-P. Thiran, J.-C. Pesquet, and Y. Wiaux, *MNRAS*, vol. 462, no. 4, pp. 4314–4335, 2016.
[4] A. Ferrari, J. Deguignet, C. Ferrari, D. Mary, A. Schutz, and O. Smirnov, *arXiv preprint arXiv:1504.06847*, 2015.

CMB delensing for detecting primordial B-mode fluctuations

Ethan Anderes*,
* Department of Statistics, University of California, Davis, California 95616.

***Abstract*—One of the major targets for next-generation cosmic microwave background (CMB) experiments is the detection of the primordial B-mode signal. Planning is under way for Stage-IV experiments that are projected to have instrumental noise small enough to make lensing and foregrounds the dominant source of uncertainty for estimating the tensor-to-scalar ratio r from polarization maps. This makes delensing a crucial part of future CMB polarization science. In this talk we present two likelihood methods for estimating the tensor-to-scalar ratio r from CMB polarization observations. These two methods combine the benefits of a full scale likelihood approach with the tractability of the quadratic delensing technique. The tractability of both methods relies on a crucial factorization of the pixel space covariance matrix of the polarization observations which allows one to compute the full likelihood profile, as a function of r, at the same computational cost of a single likelihood evaluation.**

I. Introduction

Inflation paradigm has successfully explained the origin of primordial density perturbations that grew into the Cosmic Microwave Background (CMB) anisotropies and large scale structure we observed Another key prediction of inflation is the background of primordial gravitational waves or tensor-mode perturbations which imprints a unique polarization pattern, we call primordial B-mode, on the CMB anisotropies. The strength of primordial gravitational waves or tensor-mode power is commonly quantified by the tensor-to-scalar ratio r.

The primordial B modes are contaminated by several sources: notably the foreground emission from polarized galactic dust and the B-mode polarization generated from gravitational lensing of CMB. The lensed B-mode power is nearly a constant at small multipoles ($\ell \lesssim 1000$) and therefore manifests as an effective white noise with amplitude $\sim 5\mu$K-arcmin. For CMB-S4, we expect to decrease the instrumental noise to $\sim 1\mu$K-arcmin. In this regime, the lensing B noise (and foregrounds) would become a dominant noise source and limit the primordial B-mode survey. Fortunately, the lensing B noise is well understood. Up to linear order, one can effectively delense observed B-modes by utilizing a quadratic combination of observed E-modes and an estimate of the lensing potential. However, in the regime of small instrumental noise and lensing uncertainty, higher order lensing terms, ignored by the quadratic delensing technique, can have an appreciable effect. These higher order terms not only induce quadratic delensing bias but also contain information on primordial B-modes. Moreover, experimental complexities such as nonstationary noise and sky cuts become non-trivial for spectral based methods such as the quadratic delenser. As an alternative, a full scale likelihood analysis of the tensor-to-scalar ratio r can, in principle, optimally account for all the information in the CMB observations. Unfortunately, a full likelihood analysis requires computation resources beyond what is available in the near future. In this talk, we present two likelihood based methods which are modified from the full scale likelihood one, so as to be computationally tractable.

In this talk we will present two likelihood methods for estimating the tensor-to-scalar ratio r from CMB polarization observations. These two methods combine the benefits of a full scale likelihood approach with the tractability of the quadratic delensing technique. The first method is a pixel space, all order likelihood analysis of the low frequency (high signal-to-noise) quadratic delensed B-modes. This technique essentially builds upon the quadratic delenser by taking into account all order lensing and pixel space anomalies. The second method probes high frequency primordial B-mode fluctuations via a pixel space local likelihood approximation. The tractability of both methods relies on a crucial factorization of the pixel space covariance matrix of the polarization observations which allows one to compute the full likelihood profile, as a function of r, at the same computational cost of a single likelihood evaluation.

Data Processing Challenges in the SKA era

Rosie Bolton*
* Battcock Centre for Experimental Astrophysics, University of Cambridge, CB3 0HE, UK

***Abstract*—The first phase of the Square Kilometre Array telescopes are due for completion in 2023. I will describe the data processing and analysis challenges that SKA presents, and the opportunities, challenges and limitations that are likely to come with the data in the first decade of operations.**

I. Introduction

When SKA operations begin in 2023 a new era of Radio Astronomy will commence. The mode of interacting with SKA data will be driven by the necessity to reduce the data (in the original sense of the word) in order to be able to store it for timescales longer than a few weeks, and to get it out of the remote SKA sites and into centres where astronomers can access and analyse the data products.

II. SKA data processing in context

The SKA observatory, in its first decade from 2023-2033 will comprise two telescopes working independently, one in Australia (SKA1LOW) and one in South Africa (SKA1MID). SKA1LOW will comprise around 500 groups (stations) of log- periodic antennas distributed in a centrally condensed arrangement across the Western Australian desert, on distances up to 40km from the centre, and it will be sensitive to radiation between 50 and 350MHz frequency. SKA1MID will comprise close to 200 parabolic dishes (subsuming the 64 13.5m MeerKAT antennas and adding a further 130 (or so) new 15m antennas. The frequency range for SKA1MID will be from 300MHz to 13 GHz. Both arrays of antennas are capable of functioning in a variety of different modes including: interferometric imaging mode; time- domain search mode (where a subset of antennas is phased up to produce tied-array-beams to identify time variable phenomena); VLBI mode and the ability to capture raw antenna voltage data to save in a rolling buffer in case of a time variable source detection resulting in a trigger to save these raw data. The SKA data processing will occur within the boundaries of the SKA observatory, in O(200-400) PFLOPS scale HPC centres (called Science Data Processors, SDP) located in Perth and Cape Town. These SDP centres will not support interactive data processing driven by PI astronomers, but instead will deliver a pre-agreed set of data products. The required processing pipelines and their parameters are used to predict SDP processing load and thereby contribute to schedule planning for the whole telescope (i.e. the SDP processing is a resource to be scheduled in just the same way as time-on-source is). I will describe some of the low latency functional requirements for SKA and the pipelines that will be run on the complete datasets. Data products from the SDP pipelines are passed into the archives for each telescope and then the original data are deleted. This flushing out of the SDP into these long-term preservation systems means that each Scheduling Block of SKA data (typically between a few minutes to several hours on sky) will be processed in the SDP independently. User interaction with the data products will then be via SKA Regional Centres. This is significant: SKA science for projects requiring long integration times (up to 1000 hours) will be extracted from final aggregate data products generated from a hundred or so 6 hour scheduling blocks separately; yet the noise floor in the individual products will be an order of magnitude higher than the final target noise level. How will this be achieved?; is it even feasible? I will discuss the challenges for designing the SKA Regional Centers and the AENEAS H2020 project and then go on to describe the possibilities for Exa-Byte scale astronomy and what I believe will become the normal ways for astronomers to work in the future.

Compression, Sampling, and Classification: techniques for the analysis of a new generation of Petascale surveys

Andrew Connolly*
* Department of Astronomy, University of Washington, Seattle, WA 98115

Abstract— **Astronomy has entered an era of massive data streams covering thousands of square degrees of the sky and many decades of the electromagnetic spectrum. With catalogs comprising hundreds of millions of stars and galaxies measured at thousands of separate time-steps and each with hundreds to thousands of measured attributes, we face the challenge of how to extract knowledge from these large and complex data sets. A number of statistical developments over the last decade have made some of these challenges computationally tractable including fast techniques for sampling high-dimensional likelihood spaces and techniques for the robust classification of observational data. This talk will review some of the notable successes and failures and look forward to techniques that might be needed to address the Petabyte surveys of the next decade.**

I. Introduction

It has been over 15 years since the term "precision cosmology" and the concepts of "big-data" were first introduced into the astronomical literature. In that time, the rate at which experiments collect data has increased over a thousand-fold. This increase will continue over the next decade with the commissioning of a new generation of experiments and satellites (e.g. the Dark Energy Survey [1], the Large Synoptic Survey Telescope [2], Euclid [3], the Wide Field Infrared Survey Telescope [4], and the Square Kilometer Array [5]). These missions will survey large fractions of the sky in a matter of days, and generate petabytes of data, for tens of billions of sources each with hundreds of measured attributes. Together they will address some of the most compelling questions in physics today; impacting our understanding of cosmology, the physics of the early universe, and even gravity at cosmological scales.

This shift in the way that astrophysicists collect, aggregate, and serve data has resulted in a number of statistical and computational challenges: how do we extract knowledge from large and complex data sets; how do we relate observations to the simulations of our universe (which are themselves reaching petabyte scales) in order to understand the underlying physical processes that give rise to our universe; how do we account for the noise and gaps within data streams; and how do we understand when we have detected a fundamentally new class of event or physical phenomena. This is not just a question of the size of the data (collecting and processing petabyte data sets scales well with projected technology developments) it is a fundamental question of how we discover, represent, visualize and interact with the knowledge that these data contain.

Complementary to the developments in instrumentation, are the advances in the statistical techniques that are applied to survey data. Many of these approaches can be broken into three broad categories: compression, sampling, and classification. In this talk, I will look back at the last 10 years of statistical developments in these areas and the impact they have had on survey science. This will include the introduction of techniques to expand the distribution of galaxies in terms of eigenfunctions that optimize signal-to-noise in order to estimate the underlying power spectrum or clustering of galaxies. I will discuss how techniques that enabled early science from cosmological surveys (as they were robust to missing and incomplete data) were eventually superseded by more traditional approaches as the volume of data and the availability of computational resources increased.

I will illustrate how some techniques, such as Markov Chain Monte Carlo or random decision forests, became ubiquitous throughout astronomy and how their widespread use drove the need for improvements in the robustness and efficiency of the underlying methodology. This in turn led to the development of novel sampling techniques such as importance and nested sampling, the decomposition of problems based on their computational cost, and the parallelization of the underlying algorithms.

Finally, looking forward to the next decade, I will briefly review some of the more recent developments in statistics that are gaining traction in our field, such as Approximate Bayesian Computation and hierarchical Bayes, and whether these might have a similar impact on the science of large survey astronomy.

References

[1] http://www.darkenergysurvey.org
[2] http://www.lsst.org
[3] http://www.euclid-ec.org
[4] http://wfirst.gsfc.nasa.gov
[5] https://www.skatelescope.org

Parallel image reconstruction for multi-frequency radio-interferometry

André Ferrari, David Mary and Chiara Ferrari
Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, France
Parc Valrose, F-06108 Nice cedex 02, France

***Abstract*—The perspective of the SKA telescope brings new challenges in image reconstruction. One of these is the spatio-spectral reconstruction of large (Terabytes) data cubes with high fidelity. MUFFIN (MUlti-Frequency image reconstruction For radio INterferometry) is a 3D reconstruction algorithm which combines spatial and spectral priors which are designed to achieve an efficient parallel implementation. This implementation opens the possibility of comparing efficient dictionaries, such as translation invariant wavelet transforms (IUWT), in spatio-spectral reconstruction.**

In the framework of the Square Kilometer Array, an increasing number of researchers in signal processing and radio astronomy join in common efforts. One of SKA challenges is the ability to reconstruct high-fidelity spatio-spectral data cubes (multifrequency images) of several TeraBytes (TB).

So far, existing image reconstruction algorithms are mostly monochromatic. Interestingly, sparsity was early recognized as a powerful principle for reconstruction and has lead to the most populated family of imaging algorithms. Their patriarch is the CLEAN algorithm ([1], devised in 1974), which expresses and exploits the sparsity of the sky intensity distribution in the canonical basis. Efficient monochromatic algorithms relying on more general sparse models (through redundant dictionaries) have since then proven their efficiency in radio imaging: recent examples include the works [2] (IUWT), [3], [4] (union of bases), which rely on global minimization of sparsity-regularized functionals, or [5] (IUWT), which combines complementary types of sparse recovery methods in a greedy manner.

Most of the approaches for multi-frequency reconstruction rely on a physical model for the frequency-dependent brightness distribution. In [6], a Taylor expansion of a power-law is adopted to model the flux dependence in frequency of astrophysical radio sources. More recently, reconstruction algorithms relying on parametric models for this dependence have been proposed. In [7], the authors propose to address the estimation problem using a Bayesian framework. The works [8] propose a constrained maximum entropy estimation algorithm in order to account for the frequency dependence of the intensities. These "semi-parametric" methods rely on spectral models and thus clearly offer advantages and estimation accuracy when the model is indeed appropriate. However, across the broad frequency coverage of current radio facilities, radio sources exhibiting complex spectral shapes (not simple power laws) are expected.

In [9], [10], the authors proposed to reconstruct a multi-wavelength sky image using a fully non-parametric approach. In [11] the authors also proposed a similar non-parametric approach using a low rank prior on the sky image cube, whose proximity operator requires the computation of the singular value decomposition of the data cube.

Denote as $(\boldsymbol{y}_l, \boldsymbol{H}_l, \boldsymbol{x}_l)$ the measurement set, the measurement operator and the image, at wavelength $l = 1 \dots L$. MUFFIN relies on a sparse ℓ_1 analysis prior w.r.t. a dictionary $\mathbf{W}_s$ for each image $\boldsymbol{x}_l$ and $\mathbf{W}_\lambda$ for the spectra associated to each pixel. The optimization relies on the primal-dual optimization algorithm [12], [13]. Denoting as (μ_s, μ_λ) the spatial and spectral regularization parameter, the algorithm reduces to:

1) The master node computes $\boldsymbol{T} = \mu_\lambda \boldsymbol{V}^- \boldsymbol{W}_\lambda^\dagger$ and sends the column l of $\boldsymbol{T}$, denoted as $\boldsymbol{t}_l$, to node l.
2) Each node $l = 1 \dots L$ computes sequentially:

$$\boldsymbol{\nabla}_l = \boldsymbol{H}_l^\dagger(\boldsymbol{H}_l \boldsymbol{x}_l^- - \boldsymbol{y}_l),\ \boldsymbol{s}_l = \mu_s \boldsymbol{W}_s^\dagger \boldsymbol{u}_l^- \quad (1)$$

$$\boldsymbol{x}_l^+ = \left(\boldsymbol{x}_l^- - \tau(\boldsymbol{\nabla}_l + \boldsymbol{s}_l + \boldsymbol{t}_l)\right)_+ \quad (2)$$

$$\tilde{\boldsymbol{u}}_l^+ = \text{sat}\left(\boldsymbol{u}_l^- + \sigma\mu_s \boldsymbol{W}_s(2\boldsymbol{x}_l^+ - \boldsymbol{x}_l^-)\right) \quad (3)$$

3) Each node sends $\boldsymbol{x}_l^+$ to the master and the master computes sequentially: $\boldsymbol{V}^+ = \text{sat}\left(\boldsymbol{V}^- + \sigma\mu_\lambda(2\boldsymbol{X}^+ - \boldsymbol{X}^-)\boldsymbol{W}_\lambda\right)$

Note that the particularly time consuming steps associated to: 1. the computation of the gradients, 2. the decompositions w.r.t. $\boldsymbol{W}_s$ and 3. the application of the adjunct operator $\boldsymbol{W}_s^\dagger$, see Eqs. (1,3), are all computed in parallel at each wavelength. It is worthy to note that μ_λ couples the problem w.r.t. the wavelengths. If $\mu_\lambda = 0$ the algorithm iterates Eqs. (1,2,3): each node reconstructs independtly $\boldsymbol{x}_l$.

This work was supported by MAGELLAN (ANR-14-CE23-0004-01).

References

[1] J. A. Högbom, "Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines," *AAPS*, vol. 15, pp. 417–426, Jun. 1974.
[2] H. Garsden, J. N. Girard, J.-L. Starck, S. Corbel *et al.*, "LOFAR sparse image reconstruction," *A&A*, vol. 575, p. A90, Mar. 2015.
[3] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "PURIFY: a new approach to radio-interferometric imaging," *MNRAS*, vol. 439, pp. 3591–3604, Apr. 2014.
[4] A. Onose, R. E. Carrillo, A. Repetti *et al.*, "Scalable splitting algorithms for big-data interferometric imaging in the SKA era," *arXiv:1601.04026*, Jan. 2016.
[5] A. Dabbech, C. Ferrari, D. Mary, E. Slezak, O. Smirnov, and J. S. Kenyon, "MORESANE: MOdel REconstruction by Synthesis-ANalysis Estimators. A sparse deconvolution algorithm for radio interferometric imaging," *A&A*, vol. 576, p. A7, Apr. 2015.
[6] U. Rau and T. J. Cornwell, "A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry," *A&A*, vol. 532, p. 71, Aug. 2011.
[7] H. Junklewitz, M. Bell, and T. Enßlin, "A new approach to multi-frequency synthesis in radio interferometry," *A&A*, vol. 581, Sep. 2015.
[8] A. Bajkova and A. Pushkarev, "Multifrequency synthesis algorithm based on the generalized maximum entropy method: application to 0954+658," *MNRAS*, vol. 417, no. 1, pp. 434–443, Oct. 2011.
[9] A. Ferrari, J. Deguignet, C. Ferrari, D. Mary, A. Schutz, and O. Smirnov, "Multi-frequency image reconstruction for radio interferometry. A regularized inverse problem approach," in *SPARCS*, Apr. 2015.
[10] J. Deguignet, A. Ferrari, D. Mary, and C. Ferrari, "Distributed multi-frequency image reconstruction for radio-interferometry," in *EUSIPCO*, 2016.
[11] A. Abdulaziz, A. Dabbech, A. Onose, and Y. Wiaux, "A low-rank and joint-sparsity model for hyper-spectral radio-interferometric imaging," in *EUSIPCO*, 2016.
[12] L. Condat, "A Primal-Dual Splitting Method for Convex Optimization Involving Lipschitzian, Proximable and Linear Composite Terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, Dec. 2012.
[13] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. in Comput. Math.*, vol. 38, no. 3, pp. 667–681, Nov. 2011.

Sparse Reconstruction of Radio Transients and Multichannel Images

Julien Girard*†, Ming Jiang†, Jean-Luc Starck† and Stéphane Corbel*†
* AIM Service d'Astrophysique, CEA Saclay, Orme des Merisiers, 91410 GIF-Sur-YVETTE, France
† Université Paris Diderot (Paris 7), Paris, France

***Abstract*—Imaging by aperture synthesis using interferometric data is a strong ill-posed and inverse problem. Detection of sources on deconvolved maps depends on time and frequency integration which enable to increase the set of Fourier samples of the sky. However, the improvement of the detection level assumes persistent and flat-spectrum sources whereas real radio sources can also have their own in the data (i.e. spectral and temporal behaviors made invisible due to the dilution of time/freq integration). Hopefully, the spatial, temporal and spectral reconstructions are separable problem which can be addressed separately. For transient sources, we introduce independent spatial and temporal wavelet dictionaries to sparsely represent a transient source in both spatial domain and temporal domain. For spectral sources, we designed a new deconvolution algorithm derived from GMCA to image data cubes in their spatial and spectral dimensions. We present the design and results of these new deconvolution algorithms developed in the Compressed Sensing framework. They enable the reconstruction the time profile of radio transients and the reconstruction of spectral profiles of wide-frequency sources.**

I. New challenges for radio image deconvolution

Radio interferometric Imaging via aperture synthesis has been an active field of research for ∼40 years. The incomplete knowledge of the visibility function of the sky sampled by an interferometer requires solving a strong ill-posed deconvolution problem. Tools such as CLEAN and its derivates (e.g. [1], [2], [3]) have been standards for recovering missing information in the Fourier plane. However, in the framework of Compressed Sensing, several teams have proposed new methods (e.g. [4], [5], [6] and other references therein). In addition, next generation of giant interferometers such as LOFAR, suffers from "direction-dependent" effects which distort the Fourier relationship between the measurements and the sky (such as array non-coplanarity and dipole projection). In [4], a new imager compatible with LOFAR combined both a sparse approach given by the CS theory and corrections for A and W effects [7]. It also demonstrated better angular resolution and lower residuals as compared to classical methods, on simulated and real datasets.

These giant instruments also probe the Universe in a wide temporal and spectral window that need to be properly observed and reconstructed. We present in the following, two parallel studies: sparse temporal reconstruction of transient sources and sparse multichannel image deconvolution which could bring important implication on the study of spectrally or temporally rich astrophysical sources.

II. Transient detection

Thanks to new sensitive instrumentation, the study of known class of transient sources (e.g. pulsars for general relativity tests, Active Galactic Nuclei, etc) and the recent discovery of new class of transients (e.g. Rotating Radio Transients, Fast Radio Bursts, Lorimer type bursts, see [8]) has motivated further development for transient detection and characterization.

A lot of effort has been put into the development of detection pipelines (e.g. the LOFAR TRAnsient Pipeline – TRAP [9], based on fast iterative closed-loop performing calibration / imaging / source detection / catalogue cross-matching). However, being variable and mostly point-like, the transients imaging suffers from the imaging rate. On the one hand, short time integration enables temporal monitoring of a transient, but each snapshot provides poor visibility coverage. On the other hand, long time integration images ensures a good sampling, but it will average out the temporal variation of the source. As a result, a variety of transient radio sources might be missed due to uncertainties or timescale filtering. Consequently, it is difficult to use classical imagers to detect and image transient source when the temporal variability of the transient source is unknown.

We present here a new deconvolution method, based on the CS framework, which take into account the temporal dependence of the sky. By extending previous work [4], we proposed a "2D-1D" sparse reconstruction technique based on the Condat-Vũ splitting method[10], [11] and the use of the isotropic undecimated wavelets transform [12] to reconstruction the spatial 2D space, and Haar or biorthogonal CDF 9/7 wavelets to reconstruct the 1D temporal axis. We will show the performance of the reconstruction on simulated data and on real data containing a radio transient.

III. Multi-channel deconvolution

Instrument like the SKA will provide a high number of frequency channels that contain physical information on the sources. Its instantaneous sensitivity enables channelized imaging (as opposed to frequency-blurred images) and therefore, enables source spectroscopy. By dealing with a mixture of convolved point sources and extended emission (each following its own spectral behaviour), spectral imaging requires spectral source separation in addition to spatial deconvolution.

Blind Source Separation (BSS) is a challenging matrix factorisation problem that plays a central role in multichannel imaging science. In a large number of applications, such as astrophysics, current unmixing methods are limited since real-world mixtures are generally affected by extra instrumental effects like blurring. Therefore, BSS has to be solved jointly with a deconvolution problem, which requires tackling a new inverse problem: deconvolution BSS (DBSS) . In this work, we introduce an innovative DBSS approach, called DecGMCA [13], based on sparse signal modelling and an efficient alternative projected least square algorithm. Numerical results demonstrate that the DecGMCA algorithm performs very well on simulations. It further highlights the importance of jointly solving BSS and deconvolution instead of considering these two problems independently. Furthermore, the performance of the proposed DecGMCA algorithm is demonstrated on simulated radio-interferometric data.

REFERENCES

Find all references for this abstract on
www.cosmostat.org/wp-content/uploads/2017/01/BASP2017.pdf

On Denoising Crosstalk in Radio Interferometry

Nezihe Merve Gürel[†], Paul Hurley[†], and Matthieu Simeoni[†‡]
[†] IBM Zurich Research Laboratory, Rüschlikon; [‡] École Polytechnique Fédérale de Lausanne (EPFL)

***Abstract*—Noise reduction in radio interferometers is a formidable task due to the relatively weak signals under observation. The usual strategy is to compensate with a large number of antennas and extend the observation time. We showed recently that one could de-noise directly antenna time-series, under the assumption of uncorrelated noise. This work is a first step to extend it for when crosstalk is present. We first propose a subspace based algorithm to estimate noise covariance, and then demonstrate that the noise covariance can be accurately estimated, and the image de-noised. We show sky images generated using a core station LOFAR, and that even with one tenth the observation time (and thus one tenth the data) the estimate can still be enhanced.**

Extended Abstract

To date, radio interferometers compensate for noise by throwing data and energy at the problem: observe longer, then store and process it all [1]. Furthermore, only the end sparse image is denoised, reducing flexibility substantially.

In [3], we proposed to denoise the phased-array signals directly in real-time given uncorrelated noise. The method is intimately related to low-rank approximation, removing the noise subspace explicitly at the antenna level. In real-life scenarios, however, interference can be introduced by the adjacent circuits embedded in antenna receivers. While widely applicable, the method as presented struggles to remove crosstalk from the samples due to correlation.

Denoising antenna samples in the presence of crosstalk requires a model of the underlying correlation. The present work is two-fold: infer the parameters of the noise distribution using covariance samples, and show that, already at the visibility level at stations, impressive denoising occurs. We will therefore justify our algorithm by ascertaining a sky image from the residual covariance matrix after noise is removed.

Consider L closely located antennas where the sample taken by lth antenna at time t_n is denoted by $x_l(t_n) \in \mathcal{C}$. The spatial-series governed by antenna samples can be formed as $\mathbf{x}(t_n) = (x_1(t_n), ..., x_L(t_n))$ such that $\sum_{l=2}^{L} |x_l(t_n) - x_{l-1}(t_n)|$ is minimised. The respective covariance sample is then given by

$$\mathbf{\Sigma}(t_n) = \mathbf{x}(t_n)\mathbf{x}(t_n)^\dagger.$$

Typically, an estimate of the covariance matrix over N samples is computed by

$$\hat{\mathbf{\Sigma}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{\Sigma}(t_n). \tag{1}$$

An inherent assumption follows. Source signals are uncorrelated to the noise which in turn implies that noise and the true (the absence of noise) subspaces are separable, namely that the covariance sample can be decomposed such that

$$\mathbf{\Sigma}(t_n) = \mathbf{\Sigma}_{\text{true}}(t_n) + \mathbf{\Sigma}_{\text{noise}}(t_n) \tag{2}$$

where the true and noise components are uncorrelated. In practical applications, however, separability refers to small correlation coefficient [2].

Given the above argument, here is a sketch of the method:

1. Perform SVD of $\mathbf{\Sigma}(t_n)$
2. Reconstruct rank-one bi-orthogonal elementary matrices to decompose $\mathbf{\Sigma}(t_n) = \mathbf{\Sigma}_1(t_n) + ... + \mathbf{\Sigma}_{\text{L}}(t_n)$
3. Cluster the elementary matrices by computing similarities between the elementary matrices, i.e., correlation coefficients
4. Assign the cluster with the Empirical Orthogonal Functions (EOF) at higher frequencies to the noise covariance and denote the respective set of indices by $\mathcal{I}_{\text{noise}} \subset \{1, ..., L\}$ (This labeling is due to slow-varying property of the true spatial-series.)
5. Reconstruct the noise covariance as follows

$$\mathbf{\Sigma}_{\text{noise}}(t_n) = \sum_{i \in \mathcal{I}_{\text{noise}}} \mathbf{\Sigma}_{\text{i}}(t_n).$$

Repeat the above steps for each time-instance to reconstruct $\hat{\mathbf{\Sigma}}_{\text{noise}}$ (See Equation 1). The denoised covariance matrix $\hat{\mathbf{\Sigma}}_{\text{true}}$ is thus given by $\hat{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}_{\text{noise}}$.

An application of the method to data from a LOFAR core station numerically validated our algorithm. Fig.1 demonstrates that the least-square images has been significantly cleared up by removing the noise from the covariance estimate. In particular, Fig.1(b, e) shows that the algorithm successfully split the correlated noise even when fewer number of samples are used over time. It thus suggests that we can drastically reduce the observation time to accurately generate images as well as to extract the noise parameters. This shows that heuristic arguments are in agreement with simulation.

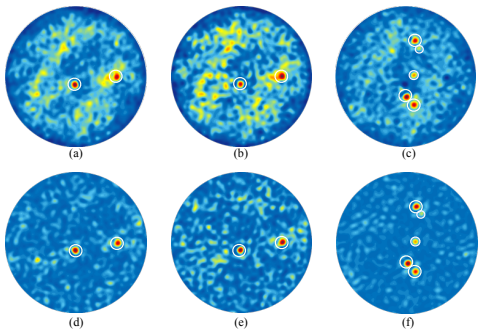


Fig. 1. The white circles denote where sources are. Noisy/denoised least-square estimates of the real sky are provided for 1000 and 100 time samples, respectively in (a)/(d) and (b)/(e). For a synthetically generated sky corrupted by real noise, we repeated the experiment. The noisy and denoised estimates over 1000 time samples are given in (c) and (f), respectively. We observe that identification of true sources is far easier and artefacts are significantly reduced when noise is split from the measurements.

The method yields accurate estimates of the noise covariance samples. The question of fitting a statistical model to the noise is thus the next direction for future research. Stochastically well-modelled noise thus permits us to extract the true behaviour of antenna signals.

References

[1] Garrett M., "Calibration and advanced radio interferometry," 2015.
[2] Golyandina N., Zhigljavsky A., "Singular spectrum analysis for time series," in *Springer Science and Business Media*, 2013.
[3] Gürel N. M., Hurley P., Simeoni M., "Phased-array data denoising at the antenna level under very low snr and its application to radio astronomy," in *submitted to International Conference on Acoustics, Speech and Signal Processing*, 2017.

Bayesian Hierarchical Modelling of data

Alan Heavens*
* Imperial Centre for Inference and Cosmology (ICIC), Imperial College London,
Blackett Laboratory, Prince Consort Road, London SW7 2AZ, U.K.

***Abstract*—A typical experiment contains many elements, each of which may be subject to error, or noise. In order to perform robust statistical inference, a complete statistical model of the data needs to be built, and this may be a complex task. Here we describe one such complex experiment, a cosmological weak lensing survey, and show how a Bayesian Hierarchical Model (BHM) may be utilised effectively. For a given set of theoretical model parameters, the power spectra of various lensing fields are specified, but the map values are not, and these, when pixelised, constitute a very large number of unknown hidden or latent variables. Sampling from such a high-dimensional space is challenging, but possible with Gibbs sampling or Hamiltonian Monte Carlo techniques. In this talk I show how this problem can be solved, with messenger fields for gaussian fields, demonstrate results from application to data, and discuss future directions.**

I. Introduction

Light from distant galaxies is continuously deflected by the gravitational potential fluctuations of large-scale structure on its way to us, resulting in a coherent distortion of observed galaxy images across the sky — weak gravitational lensing. This weak lensing effect is a function of both the geometry of the Universe (through the distance-redshift relation) and the growth of potential fluctuations along the line-of-sight, making it a tremendously rich cosmological probe; the statistics of the weak lensing fields are sensitive to the initial conditions of the potential fluctuations, the relative abundance of baryonic and dark matter (through baryon acoustic oscillations), the linear and non-linear growth of structure, the mass and hierarchy of neutrinos, dark energy and gravity on large scales. The goal of cosmic shear analyses is to extract cosmological inferences from the statistics of the observed weak lensing shear field — the distortion of observed galaxy shapes measured across the sky and in redshift.

In [1] we developed a Bayesian hierarchical modelling (BHM) approach to inferring the cosmic shear power spectrum (and thus cosmological parameters) from a catalogue of measured galaxy shapes and redshifts, building on previous work on cosmic microwave background (CMB) power spectrum inference (e.g. [2]) and large scale-structure analysis methods (e.g. [3]). The Bayesian hierarchical approach has a number of desirable features and advantages over traditional estimator-based methods: In contrast to frequentist estimators whose likelihoods need calibrating against large numbers of forward simulations (introducing assumptions and uncertainties that are often hard to propagate), the Bayesian approach explores the posterior distribution of the parameters of interest directly with clearly stated (and minimal) model assumptions, without the need for calibration. The Bayesian approach is exact and optimal, up to our ability to model the cosmic shear statistics (and systematics). Masks and complicated survey geometry are dealt with exactly and cleanly, in contrast to, e.g., pseudo-C_ℓ estimators where the mask inversion leads to mixing of E- and B-modes and physical (angular) scales that needs to be carefully accounted and corrected for. The BHM approach can be readily extended to include models of non-Gaussian fields, exploiting more of the information-content of the weak lensing fields than is possible through n-point statistic estimators [3], [4], [5]. More generally, the BHM approach can also be extended to incorporate more of the weak lensing inference pipeline (e.g., shape measurement, PSF modelling etc), formally marginalising over nuisance parameters and systematics in a principled way and ultimately leading to more robust science at the end of the day (see [1], [6] for a discussion of the global hierarchical modelling approach to weak lensing).

In this work, we develop a Bayesian approach to cosmic shear inference, whereby we jointly sample the shear maps and cosmological parameters, rather than the maps and power spectra. By going straight to cosmological parameters and bypassing the explicit power spectrum inference step, we circumvent the need to transform posterior samples into a continuous posterior density and hence avoid the prickly (high-dimensional) density estimation issues altogether. There are other advantages, too: by parametrising the power spectrum with a handful of cosmological parameters, the number of interesting parameters has been reduced from a few thousand power spectrum coefficients to typically ~ 10 cosmological parameters – this reduction in the parameter space will inevitably improve the sampling efficiency. Map-cosmology inference also extends naturally to incorporate models for non-Gaussian shear where the power spectrum no longer fully specifies the lensing statistics and is a (comfortably) computationally feasible approach for current and future surveys.

In this work we apply the map-power spectrum and map-cosmology sampling schemes to infer power spectra, shear maps and cosmological parameters from the Canada-France-Hawaii Telescope (CFHTLenS) weak lensing survey - the first application to data.[1]

References

[1] J. Alsing, A. Heavens, A. H. Jaffe, A. Kiessling, B. Wandelt, and T. Hoffmann, "Hierarchical cosmic shear power spectrum inference," *Monthly Notices of the Royal Astronomical Society*, vol. 455, no. 4, pp. 4452–4466, 2016.

[2] B. D. Wandelt, D. L. Larson, and A. Lakshminarayanan, "Global, exact cosmic microwave background data analysis using gibbs sampling," *Physical Review D*, vol. 70, no. 8, p. 083511, 2004.

[3] J. Jasche and B. D. Wandelt, "Methods for bayesian power spectrum inference with galaxy surveys," *The Astrophysical Journal*, vol. 779, no. 1, p. 15, 2013.

[4] F. Leclercq, J. Jasche, and B. Wandelt, "Bayesian analysis of the dynamic cosmic web in the sdss galaxy survey," *Journal of Cosmology and Astroparticle Physics*, vol. 2015, no. 06, p. 015, 2015.

[5] J. Carron, "Information escaping the correlation hierarchy of the convergence field in the study of cosmological parameters," *Physical review letters*, vol. 108, no. 7, p. 071301, 2012.

[6] M. D. Schneider, D. W. Hogg, P. J. Marshall, W. A. Dawson, J. Meyers, D. J. Bard, and D. Lang, "Hierarchical Probabilistic Inference of Cosmic Shear," *Ap.J.*, vol. 807, p. 87, Jul. 2015.

[7] J. Alsing, A. F. Heavens, and A. H. Jaffe, "Cosmological parameters, shear maps and power spectra from cfhtlens using bayesian hierarchical inference," *arXiv.org*, 2016.

[1] Abstract adapted from [7]

Statistical properties of nested sampling parameter estimation

Edward Higson*†, Will Handley*†, Mike Hobson* and Anthony Lasenby*†
* Astrophysics Group, Battcock Centre, Cavendish Laboratory, JJ Thomson Avenue, Cambridge CB3 0HE, UK.
† Kavli Institute for Cosmology, Madingley Road, Cambridge CB3 0HA, UK.

***Abstract*—The statistical properties of parameter estimation with nested sampling differ from those of Bayesian evidence calculation, but have been little studied in the literature. This poster explains the different sources of stochastic errors in nested sampling parameter estimation, and includes a new diagrammatic representation of the process. We find no current method can accurately measure the error in parameter inferences from a single nested sampling run, and give a new method for doing so. We introduce dynamic nested sampling — a generalisation of the nested sampling algorithm which promises increased efficiency in parameter estimation.**

I. Introduction

Nested sampling [1] is a Monte Carlo method for Bayesian analysis which simultaneously calculates both Bayesian evidences and posterior samples. The early development of the algorithm was focused on evidence calculation. However contemporary implementations such as MultiNest [2]–[4] and PolyChord [5], [6] are now also extensively used for parameter estimation from posterior samples (see for example [7]).

Nested sampling compares favourably to MCMC-based parameter estimation for degenerate, multi-modal likelihoods as it has no "thermal" transition probability and exponentially compresses the prior distribution to the posterior. Despite its increasing popularity, the stochastic errors in nested sampling parameter estimation are poorly understood.

II. Nested sampling parameter estimation

This poster explains stochastic errors in parameter estimation, and shows that any such calculation can be represented in 2 dimensions. We present a new diagram for visualising the process.

Correctly quantifying uncertainty is vital for identifying spurious results — in particular we find stochastic errors often significantly affect estimates of confidence intervals on parameters. Conversely, finding such errors are very small may imply an unnecessarily large amount of computational resource is being used for the calculation. Interestingly, we find no current method can accurately estimate errors on parameter estimation from a single nested sampling run, and so we describe a new method for doing this.

III. Dynamic nested sampling

Nested sampling was designed for calculating Bayesian evidences, and a typical calculation spends most of its computational effort iterating towards the posterior peak, producing posterior samples with negligible weights. The fraction of computational effort spent near the posterior peak is set by the likelihood and the priors; in contemporary nested sampling this cannot be adjusted when parameter estimation is the primary goal.

We propose modifying the nested sampling algorithm by dynamically varying the number of "live points" in order to maximise the information gained from posterior samples, subject to practical constraints. We term this more general approach *dynamic nested sampling*, with conventional nested sampling representing the special case where the number of live points is constant. This can increase the effective number of posterior samples and the accuracy of their relative weights — additional errors on the absolute weights of all samples do not affect parameter estimation.

References

[1] J. Skilling, "Nested sampling for general bayesian computation," *Bayesian Analysis*, vol. 1, no. 4, pp. 833–859, 2006.
[2] F. Feroz and M. P. Hobson, "Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses," *Monthly Notices of the Royal Astronomical Society*, vol. 384, pp. 449–463, Feb. 2008.
[3] F. Feroz, M. P. Hobson, and M. Bridges, "MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics," *Monthly Notices of the Royal Astronomical Society*, vol. 398, pp. 1601–1614, 2009.
[4] F. Feroz, M. Hobson, E. Cameron, and A. Pettitt, "Importance nested sampling and the multinest algorithm," *Monthly Notices of the Royal Astronomical Society*, vol. 398, no. 4, pp. 1601–1614, 2013.
[5] W. Handley, M. Hobson, and A. Lasenby, "Polychord: nested sampling for cosmology," *Monthly Notices of the Royal Astronomical Society: Letters*, vol. 450, no. 1, pp. L61–L65, 2015.
[6] ——, "Polychord: next-generation nested sampling," *Monthly Notices of the Royal Astronomical Society*, vol. 453, no. 4, pp. 4384–4398, 2015.
[7] Planck, P. A. R. Ade, N. Aghanim, M. Arnaud, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, and et al., "Planck 2015 results. XX. Constraints on inflation," *ArXiv e-prints*, Feb. 2015.

Bayesian compressed sensing

Mike Hobson *,
* Astrophysics Group, Battcock Centre, Cavendish Laboratory, JJ Thomson Avenue, Cambridge CB3 0HE, UK.

***Abstract*—Bayesian inference and compressed sensing are both well-established methods for data analysis. Typically, however, these two approaches are considered to somewhat distinct from one another, and this is often reflected in the relatively small overlap of the communities who develop and apply each technique. Nonetheless, the two methods do, in fact, have a considerable amount in common and a Bayesian interpretation of compressed sensing, and sparsity more generally, has been pursued by a number of authors. In this talk, I will develop this link more fully and discuss how Bayesian inference provides a very natural framework for sparsity and compressed sensing.**

In particular, I will outline a principled Bayesian approach for simultaneously imposing sparsity and performing dictionary learning to determine the optimal basis set for representing the signal. In this method, a signal is modelled as the superposition of a set of basis functions, whose number and form are determined by the data themselves. Sparsity is imposed directly via the prior on the number of basis functions, while simultaneous dictionary learning is performed through the estimation of parameters describing the location and shape of the basis functions. In principle, the number n of basis functions may be determined through Bayesian model selection and the evaluation of the evidence as a function of n. It is both more convenient and flexible, however, but nonetheless equivalent, to determine n using parameter estimation, whereby n is instead allowed to vary dynamically and one samples directly from the joint posterior of n and the parameters describing these n basis functions. The final inference may then be obtained by either by choosing the maximum aposteriori value of n, or better, by marginalising over n to yield an implicit multimodel solution.

I will demonstrate the practical implementation of this method in some simple test problems. To achieve an algorithm that is computationally not too burdensome, a number of issues need to be addressed. Rather than using costly transdimensional sampling technqiues, such as reversible-jump MCMC, to accommodate the fact that the dimension of the parameter space can vary, we instead use a product-space approach in which one considers a space of fixed dimensionality equal to the largest that can be encountered. Since the number n of basis functions is an (effective) integer parameter, the technique used to explore the parameter space should not rely on gradient information. Moreover, by considering the full joint space of n and the parameters describing the basis functions, one will typically need to accommodate spaces of moderate to large dimensionality, most likely possessing multiple modes and/or pronounced degeneracies. In practice, nested sampling is well suited to such problems, and therefore we adopt its latest and most efficient implementation, PolyChord, which can accommodate up to around 1000 dimensions.

Since our approach is based in parameter estimation using nested sampling, it is important to understand the statistical properties of this process, which differ from those of Bayesian evidence calculation, but have been little studied in the literature. In particular, I will outline different sources of stochastic errors in nested sampling parameter estimation, and introduce a new diagrammatic representation of the process. We find no current method can accurately measure the error in parameter inferences from a single nested sampling run, and I will describe a new method for doing so. I will also introduce dynamic nested sampling: a generalisation of the nested sampling algorithm which promises increased efficiency in parameter estimation. This is based on modifying the nested sampling algorithm by dynamically varying the number of 'live points' in order to maximise the information gained from posterior samples, subject to practical constraints.

On Flexibeam for radio interferometry

Paul Hurley*, Matthieu Simeoni*†
*IBM Zurich Research, Rüschlikon; †École Polytechnique Fédérale de Lausanne (EPFL)

***Abstract*—Beamforming in radio astronomy focuses at and around a direction using matched beamforming or a derivative, to both maximise the energy coming from this point and reduce the data rate to the central processor. Such beamformers often result in large side-lobes, with influence from undesired directions. Moreover, there is a fundamental lack of flexibility when, for example, targeting extended regions or tracking objects with uncertainty as to their location.**

We show how the analytic framework Flexibeam can be leveraged to achieve beamshapes that cover general spatial areas with substantially more energy concentration within the region-of-interest. The method is numerically stable, and scalable in the number of antennas, and does not magnify noise.

Extended Abstract

Beamforming in radio astronomy has mostly been a byword for matched beamforming: focus on and around one point in the sky by phase aligning the antenna signals. It is essentially dual-purpose: get information with high SNR around that point, while reducing the amount of data to send from a station to the central processing, so as to compress and reduce complexity.

While simple, there are quite a few drawbacks. The side-lobes induced are large, polluting substantially the data observed. Power is maximised from this one direction in the sky, but sees relatively little of the rest of it. Hence, surveying large portions requires multiple observations, steering towards various locations successively. Additionally, it is very sensitive to uncertainty in the target, and indeed only one point can be targeted at any given time.

Instead, it would be desirable to specify a general spatial sky filter, not necessarily contiguous, and determine how to beamform so as to approximate the filter. This framework is provided by Flexibeam [1]. A spatial filter is described over the sphere $\mathbb{S}^2$, from which an extended filter in $\mathbb{R}^3$ is chosen. A beamforming function is then obtained over Euclidean space $\mathbb{R}^3$. From this, the beamforming weight for an antenna is given by sampling the beamforming function at its position $\boldsymbol{p} \in \mathbb{R}^3$.

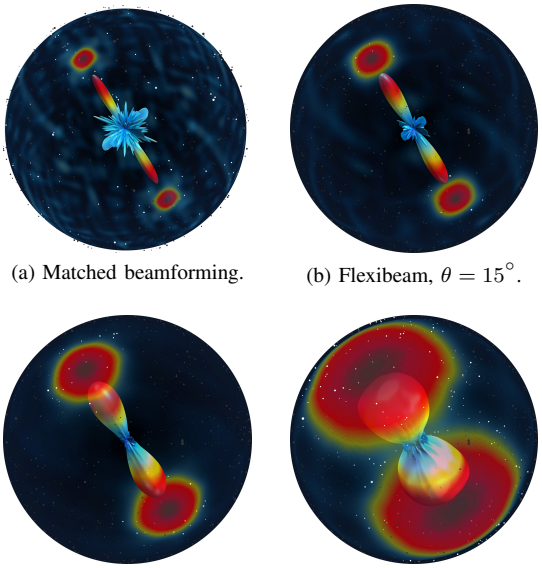
The analytical framework allows tractable, and numerically stable weight determination. It scales linearly with the number of antennas (just add additional samples for more antennas) – a key advantage in radio interferometers with thousands of antennas.

Suppose we wish to observe a specific region on the sphere. One good choice of extended filter is then the 3D ball indicator defined by $\hat{\omega}(\boldsymbol{r}) = 1$ if $\|\boldsymbol{r} - \boldsymbol{r}_0\| \leq R$ and 0 otherwise, where $\boldsymbol{r}_0 \in \mathbb{S}^2$, and $R > 0$ specifying the width of the targeted region. The resultant beamforming function, from which the beamforming weights are obtained, is given by:

$$\omega(\boldsymbol{p}) = R^{-1/2} \|\boldsymbol{p}/\lambda\|^{-3/2} J_{3/2}\left(2\pi R \|\boldsymbol{p}/\lambda\|\right) e^{-j2\pi\langle \boldsymbol{r}_0, \boldsymbol{p}/\lambda\rangle},$$

where $J_{3/2}$ is a Bessel function of the first kind, and $\boldsymbol{p} \in \mathbb{R}^2$. We can thus approximate beamshapes of various widths, as illustrated in Fig. 1 (for an LBA core LOFAR station composed of 96 antennas, with the frequency of 45MHz). Notice how Flexibeam dramatically reduces side lobes, and that we can get much more of the beam energy where we want it.

Suppose we wish to scan a region as shown in Fig. 2, which compares the use of multiple matched beams through progressive scan, versus multiple observations using a Flexibeam-determined beamshape. More signal in the area of interest is obtained with Flexibeam. Here it amplifies the signal by 26.2% in the region of interest with respect to matched beamforming. Flexibeam also has 33% less energy in the side-lobes.



(a) Matched beamforming. (b) Flexibeam, $\theta = 15^\circ$. (c) Flexibeam, $\theta = 35^\circ$. (d) Flexibeam, $\theta = 85^\circ$.

Fig. 1: Beamshapes illustrating how Flexibeam can survey large portions of the sky in radio-astronomy with reduced side-lobes.

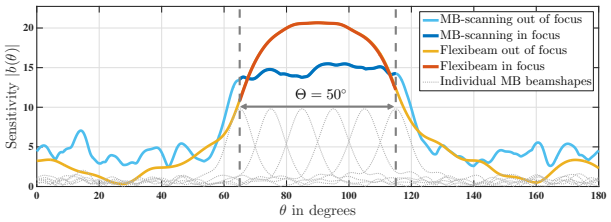


Fig. 2: Comparison, in 2D, of matched beamforming successive scan vs. a single Flexibeam beamshape with the same observation time.

By use of the computationally low-cost Flexibeam framework we are able to use one instrument for many different use cases, designing spatial filters with corresponding beamforming functions so as to search in multiple areas or track a pulse. Recent work has also shown how these beamshapes can be incorporated efficiently into the imaging pipeline [2].

References

[1] P. Hurley and M. Simeoni, "Flexibeam: analytic spatial filtering by beamforming," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE*, March 2016.

[2] M. Simeoni, "Towards more accurate and efficient beamformed radio interferometry imaging," Master's thesis, EPFL, Spring 2015.

The Murchison Widefield Array: Exploring the challenges of low-frequency radio astronomy

Natasha Hurley-Walker* Paul Hancock*, and André Offringa†
* Curtin Institute of Radio Astronomy, International Centre for Radio Astronomy Research, 1 Turner Avenue, Bentley WA 6102, Australia
† Netherlands Institute for Radio Astronomy (ASTRON), PO Box 2, 7990 AA Dwingeloo, The Netherlands

***Abstract*—The Murchison Widefield Array (MWA)[1] is a low-frequency radio telescope sited in the Western Australian outback. It is a precursor to the low-frequency Square Kilometre Array and has been fully operational since 2013, testing software and hardware solutions for the challenge of low-frequency radio astronomy. A large team of engineers and astronomers across twenty-two partner institutes and six countries have worked to deliver cutting edge science from the resulting data. This talk will explore some of the unique signal processing challenges for the MWA, in both its current state and after future upgrades.**

I. Introduction

The MWA consists of 2048 dual-polarization dipole antennas optimized for the 80–300 MHz frequency range, arranged as 128 "tiles", each a 4x4 array of dipoles. A complete technical description of the telescope is given in the journal article: The Murchison Widefield Array: The SKA Low Frequency Precursor by Tingay et al. (2013) [1]. The array has no moving parts, and all telescope functions including pointing are performed by electronic manipulation of dipole signals, each of which contains information from nearly four steradians of sky centered on the zenith. Each tile performs an analog beamforming operation, narrowing the field of view to a fully steerable 25 degrees at 150 MHz.

To best exploit these data, imaging challenges have been met with appropriate solutions, which will be detailed in this talk:

- The antennas are not co-planar, necessitating the use of advanced widefield imaging software such as WSClean [2];
- The aperture array antennas and minimum analogue beamformer delays result in an unusual primary beam, necessitating snapshot imaging, and complicating flux calibration;
- The ionosphere refracts incoming astronomical radio signals, causing apparent source position shifts over the sky; while visibility-based solutions are possible [3], cheaper image-based solutions are also useful (Fig. 1);
- Final images have resolutions differing by a factor of two over the band; the new source-finding technique of *priorised fitting* is necessary to properly describe sources (Fig. 2).

This talk will present solutions to these varied problems and how they were used to perform a large-scale sky survey, completely imaging the southern radio sky and flux-calibrating it to better than 8 %, impressive for an aperture array. A description of this GaLactic and Extragalactic All-sky MWA (GLEAM) survey is given by Wayth et al. (2015) [4] and the first extragalactic sky catalogue is presented by Hurley-Walker et al. (2016, accepted).

The telescope is now being reconfigured into a configuration with redundant baselines, in order to maximise calibratability and sensitivity on angular scales sensitive to the Epoch of Reionisation. Next year, it will be transformed again, into a long-baseline configuration, for high-resolution observations of distant astronomical objects. In future, a planned upgrade will increase the instantaneous bandwidth from 30 MHz to > 100 MHz, and improve spectral smoothness for spectral line and EoR observations. All of these configurations entail new signal processing challenges, which will explore a calibration parameter space of interest to the upcoming Square Kilometre Array.

[1] http://mwatelescope.org/

References

[1] S. J. Tingay, R. Goeke, J. D. Bowman *et al.*, "The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies," *Publications of the Astronomical Society of Australia*, vol. 30, p. 7, Jun. 2013.
[2] A. R. Offringa, B. McKinley, N. Hurley-Walker *et al.*, "WSCLEAN: an implementation of a fast, generic wide-field imager for radio astronomy," *Monthly Notices of the Royal Astronomical Society*, vol. 444, pp. 606–619, Oct. 2014.
[3] A. R. Offringa, C. M. Trott, N. Hurley-Walker *et al.*, "Parametrizing Epoch of Reionization foregrounds: a deep survey of low-frequency point-source spectra with the Murchison Widefield Array," *Monthly Notices of the Royal Astronomical Society*, vol. 458, pp. 1057–1070, May 2016.
[4] R. B. Wayth, E. Lenc, M. E. Bell *et al.*, "GLEAM: The GaLactic and Extragalactic All-Sky MWA Survey," *Publications of the Astronomical Society of Australia*, vol. 32, p. e025, Jun. 2015.

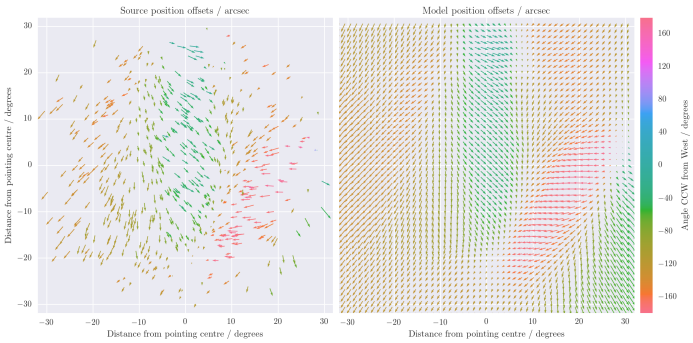


Fig. 1. Ionospheric model-fitting over the MWA's 60° field-of-view: the left panel shows how the ionosphere distorts apparent source positions, and the right panel shows a radial basis function model fit to these distortions, which can be used to correct the images during typical ionospheric conditions.

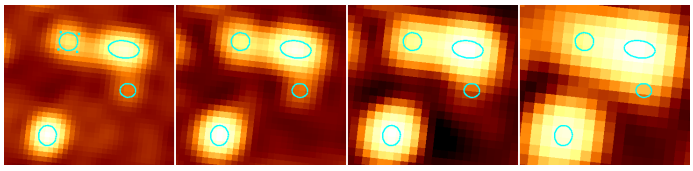


Fig. 2. Four example images from the GLEAM survey: source-finding is performed in the highest-resolution image at 200 MHz (left-most panel), while priorised fitting is performed on lower-resolution sub-bands (right panels), in order to prevent noise bias when fitting faint sources, avoid a difficult cross-matching problem, and preserve morphological information.

Bayesian data interpretation with large scale cosmological models

Jens Jasche*, Guilhem Lavaux†, Florent Leclercq‡, Benjamin Wandelt §
* Excellence Cluster Universe, Technische Universität München, Boltzmannstrasse 2, 85748 Garching, Germany.
† Sorbonne Universités, CNRS et UPMC Univ Paris 6, UMR 7095,
Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France
‡ Institute of Cosmology and Gravitation (ICG), University of Portsmouth,
Dennis Sciama Building, Burnaby Road, Portsmouth PO1 3FX, United Kingdom
§ Sorbonne Universités, CNRS et UPMC Univ Paris 6, UMR 7095,
Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France

***Abstract*—Next generation cosmological surveys will provide an avalanche of cosmological observations. This increase in scientific data needs to be accompanied with the development of novel information processing techniques to interpret such observations. Analyses of three dimensional inhomogeneous large scale structures require to jointly account for complex dynamical processes, associated to the gravitational formation history of the dark matter distribution, together with systematic as well as stochastic effects arising from observational procedures. Addressing these issues often requires to solve non-linear statistical inference problems in multi-million dimensional parameter spaces. We address this problem of high dimensional Bayesian inference from cosmological data sets via our recently developed BORG algorithm. This approach couples three dimensional numerical models of cosmological structure formation with a Hybrid Monte Carlo algorithm. By exploring plausible data constrained realizations of cosmic history this approach provides new avenues to study the full four dimensional state of our Universe in data.**

I. Introduction

Just recently the standard model of cosmology, describing the homogeneous evolution of the Universe and formation of structure, has been celebrated to successfully explain cosmic microwave background (CMB) observations provided by ESA's Planck satellite mission [1]. According to this model present dynamical evolution of our Universe is governed by dark energy and dark matter, constituting about 95 % of its total content. Although required to explain the formation of all observable structures within the picture of Einstein's gravity, so far dark matter and dark energy elude direct observations and have not yet been identified as particles within fundamental theories.

To make progress on these prevailing cosmological mysteries next generation cosmological surveys will go beyond two dimensional observations of the CMB by mapping the three dimensional cosmic large scale structure (LSS) with observed galaxies in the Universe. This promises to provide orders of magnitudes of additional information on the dynamical processes governing the evolution of our Universe.

However, connections between observations and our physical theories of the Universe cannot be established trivially. Besides typical stochastic and systematic effects associated to noisy and incomplete observations, analyses of modern deep observation also needs to account for the non-linear and dynamical evolution of the cosmic LSS throughout the observed domain. In particular, deep cosmological observations do not consist in time homogeneous measurements. Due to the finite speed of light we observe the galaxy distribution at earlier and earlier cosmic epochs with increasing distances. This light cone effect poses a challenge for cosmological data analysis as the nature and statistical properties of the subject to study vary greatly across observed volumes. In addition to such light cone effects, observations also suffer from non-linear and linear redshift space distortions associated to the motion of observed objects inside gravitationally forming structures.

We propose to address these issues with our recently developed algorithm for Bayesian Origin Reconstruction from Galaxies (**BORG**). The **BORG** algorithm is a fully probabilistic inference machinery incorporating a physical model of gravitational structure formation. More specifically the algorithm fits numerical simulations to three dimensional observations. This results in a highly non-trivial Bayesian inverse problem, requiring to explore the very high-dimensional and non-linear space of plausible solutions to the initial conditions problem from incomplete observations [2], [3]. These parameter spaces typically consist in 10^6 to 10^7 parameters, corresponding to the discretized volume elements of the observed domain. We solve this large scale inference problem with the implementation of efficient Hamiltonian Monte Carlo techniques permitting us to recover the cosmic initial conditions and the formation history of observed structure over a period of about 13.6 billion years.

As scientific results the **BORG** algorithm infers three dimensional initial conditions from which observed structures originate, non-linear density and velocity fields and provides dynamic structure formation histories including a detailed treatment of uncertainties [2], [3], [4]. This work is a clear demonstration of the technical feasibility of large scale data interpretation with complex numerical models and provides novel avenues to study the full four dimensional state of our Universe in observations.

References

[1] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al., "Planck 2013 results. XVI. Cosmological parameters," *A&A*, vol. 571, p. A16, Nov. 2014.

[2] J. Jasche and B. D. Wandelt, "Bayesian physical reconstruction of initial conditions from large-scale structure surveys," *Monthly Notices of the Royal Astronomical Society*, vol. 432, pp. 894–913, Jun. 2013. [Online]. Available: http://adsabs.harvard.edu/abs/2013MNRAS.432..894J

[3] J. Jasche, F. Leclercq, and B. D. Wandelt, "Past and present cosmic structure in the SDSS DR7 main sample," *JCAP*, vol. 1, p. 36, Jan. 2015.

[4] G. Lavaux and J. Jasche, "Unmasking the masked Universe: the 2M++ catalogue through Bayesian eyes," *MNRAS*, vol. 455, pp. 3169–3179, Jan. 2016.

Teaching computers to see colours in the Hubble Frontier Fields with Morphological Component Analysis-based method

Rémy Joseph*, Frederic Courbin*, Jean-Luc Starck† and Pascale Jablonka*
* Laboratoire d'astrophysique, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland
† Laboratoire AIM, CEA/DSM-CNRS-Universite Paris Diderot, Irfu, Service d'Astrophysique, CEA Saclay, Orme des Merisiers, 91191 Gif-sur-Yvett

***Abstract*—We present a new Morphological Component Analysis-based method designed to separate astronomical objects with different colours and show the results of its application to real Hubble images. We also show how such separation can be used to infer cosmological and astrophysical results on dark matter distribution in galaxy and cluster and on galaxy evolution and formation.**

I. Introduction

The relative crowdedness of our Universe makes images of objects overlap with one another. These objects very often present different colours, whether it is due to their age or to their evolution and star formation history. Young and star forming object will appear blue, whereas old quiescent galaxies will be seen red in visible multi-band observation. We showed that it was possible to teach our computers to distinguish between colourful objects in astronomical images and to separate them in a model independent way. We expressed the problem of colour blending as a blind source separation problem that we solve by inverting the linear problem under a sparsity constraint in wavelet space [1]. We implemented the solution algorithm, called MuSCADeT (https://github.com/herjy/MuSCADeT), and applied our method to real observations: the Hubble Frontier Fields (http://www.stsci.edu/hst/campaigns/frontier-fields/HST-Survey).

The Hubble Frontier Fields (HFF) is one of the deepest high resolution survey available to date in the visible and near infra-red. The survey targeted 6 galaxy clusters that had been selected for their lensing strength. These massive objects produce number of magnified multiple images of single background galaxies through an effect called strong gravitational lensing. The positions and magnifications of strongly lensed images can be used to constrain the mass density profile of the clusters, thus tracing the distribution of dark matter in these regions of our Universe.

Once the red and blue images of galaxies separated, our results revealed new hidden candidates for strongly lensed objects, thus allowing the prospect for more precise constraints on the mass density distribution of the clusters. More importantly, this allows for accurate photometry of the background lensed objects. Until now, blending made any attempt of measuring how gravitational lensing magnifies the multiple images of background objects impossible. Our method allows to remove, or at least diminish this bias, and we show that we can already use this new measurement to discriminate between existing mass models of the HFF clusters. With this method we should be able to provide an extra constraint on mass models that has never been used until now, thus helping refine our knowledge of dark matter properties.

Not only our method is very well adapted to study strong gravitational lenses, but also, we showed that it was possible to separate the young from the old stellar populations inside single spiral galaxies. For the first time, we are able to study the morphology of these populations beyond the limitation of model fitting of either component. We compare the properties of red and blue components of galaxies as shown on fig. 1 and relate these properties to their environment and position, relative to the cluster.

More about MuSCADeT, see http://www.cosmostat.org/research/galaxy-colour-component-separation/

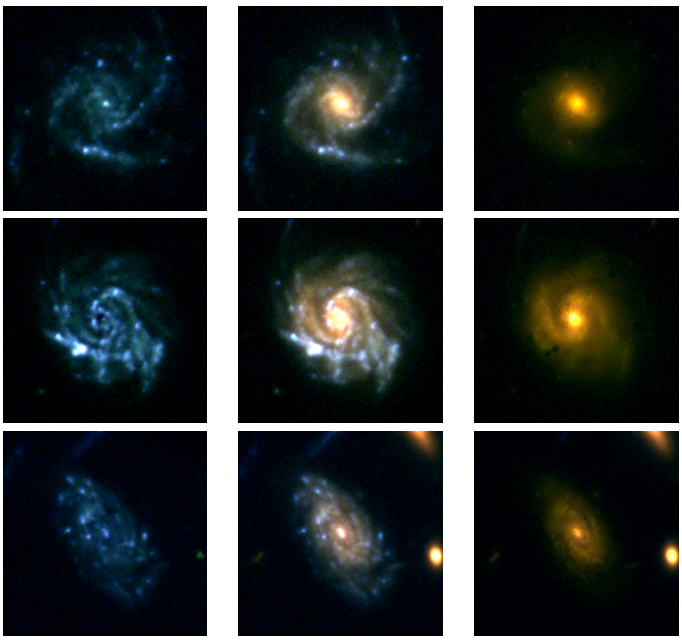


Fig. 1. Result of the separation of three Frontier Fields galaxies by `MuSCADeT`. *Left panels:* blue component, *middle panel:* Original image, *left panel:* red component.

References

[1] R. Joseph, F. Courbin, and J.-L. Starck, "Multi-band morpho-Spectral Component Analysis Deblending Tool (MuSCADeT): Deblending colourful objects," *aap*, vol. 589, p. A2, May 2016.

Fourier dimensionality reduction of radio-interferometric data

S. Vijay Kartik*, Rafael E. Carrillo*, Jean-Philippe Thiran* and Yves Wiaux†
* 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*—Next-generation radio-interferometers face a computing challenge with respect to the imaging techniques that can be applied in the big data setting in which they are designed. Dimensionality reduction can thus provide essential savings of computing resources, allowing imaging methods to scale with data. The work presented here approaches dimensionality reduction from a compressed sensing theory perspective, and links to its role in convex optimization-based imaging algorithms. We describe a novel linear dimensionality reduction technique consisting of a linear embedding to the space spanned by the left singular vectors of the measurement operator. A subsequent approximation of this embedding is shown to be practically implemented through a weighted subsampled Fourier transform of the dirty image. Preliminary results on simulated data with realistic coverages suggest that this approach provides significant reduction of data dimension to well below image size, while achieving comparable image quality to that obtained from the complete data set.**

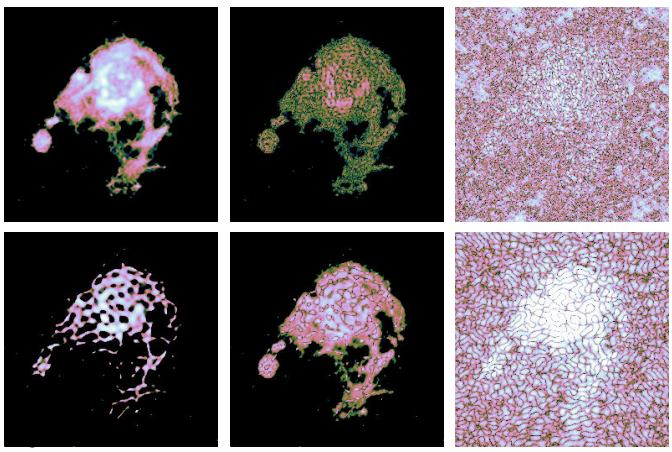
The large amount of data produced from next-generation telescopes like the Square Kilometre Array (SKA) presents a computational challenge for imaging methods, and calls for High Performance Computing (HPC)-ready solutions. Here we present our dimensionality reduction technique as a way to handle big data, and show that radio-interferometric (RI) imaging algorithms applied on significantly reduced data using the proposed method retain image reconstruction quality. The results detailed here are based on preliminary studies as presented in [1].

RI data acquisition can be modelled through the discretized form of a measurement equation, given by $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{n}$. Here $\boldsymbol{y} \in \mathbb{C}^M$ is a vector of continuous Fourier measurements (visibilities) corrupted by additive noise $\boldsymbol{n} \in \mathbb{C}^M$; we assume $\boldsymbol{n}$ to have i.i.d. Gaussian noise statistics. The visibilities $\boldsymbol{y}$ measure an underlying vectorized image $\boldsymbol{x} \in \mathbb{C}^N$, and $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$ is the measurement operator, with $M \gg N$.

Linear dimensionality reduction is performed through an embedding matrix $\mathbf{R} \in \mathbb{C}^{M_L\times M}$, $M_L \ll M$, leading to the reduced inverse problem $\boldsymbol{y'} = \boldsymbol{\Phi' x} + \boldsymbol{n'}$, with $\boldsymbol{y'} = \mathbf{R}\boldsymbol{y}$, $\boldsymbol{n'} = \mathbf{R}\boldsymbol{n}$, and $\boldsymbol{\Phi'} = \mathbf{R}\boldsymbol{\Phi}$. Consequently, imaging algorithms need only deal with the embedded measurement vector of dimension M_L, thus avoiding expensive computations involving large vectors of size M. As an embedding operator, $\mathbf{R}$ affects not only the mapping to $\boldsymbol{y'}$ but also the properties of $\boldsymbol{\Phi}$ required by CS theory to guarantee stable signal recovery. Additionally, retaining the i.i.d. Gaussian properties of the original measurement noise is important for the convex optimization-based algorithms used for image reconstruction.

The optimal dimensionality reduction $\mathbf{R}_{\text{optim}}$, with respect to CS-based image reconstruction, is a projection to the left singular vectors of the measurement operator $\boldsymbol{\Phi}$ that correspond to non-zero singular values. For a Singular Value Decomposition (SVD) of $\boldsymbol{\Phi}$ given by $\boldsymbol{\Phi} = \mathbf{U\Sigma V}^\dagger$, the optimal embedding is then given by $\mathbf{R}_{\text{optim}} = \mathbf{U}_0{}^\dagger = \boldsymbol{\Sigma}_0{}^{-1}\mathbf{V}_0{}^\dagger\boldsymbol{\Phi}^\dagger$, where the final data size $M_L = N_0 \leq N$ is the number of non-zero (or *significant*) singular values of $\boldsymbol{\Phi}$, and where $\mathbf{U}_0$, $\boldsymbol{\Sigma}_0$ and $\mathbf{V}_0$ are truncated versions of $\mathbf{U}$, $\boldsymbol{\Sigma}$, $\mathbf{V}$ by only retaining columns (rows for $\mathbf{V}$) corresponding to the N_0 non-zero (or significant) singular values of $\boldsymbol{\Phi}$. However, this requires the SVD, which is computationally expensive and hence infeasible. So, a practical implementation of this optimal embedding is constructed through the approximations $\mathbf{V}^\dagger \approx \mathbf{F}$ and $\boldsymbol{\Sigma}^2 \approx \text{Diag}(\mathbf{F}\boldsymbol{\Phi}^\dagger\boldsymbol{\Phi}\mathbf{F}^\dagger)$, leading to the embedding operator $\mathbf{R}_{\text{sing}} = \boldsymbol{\Sigma}_0{}^{-1}\mathbf{SF}\boldsymbol{\Phi}^\dagger \in \mathbb{C}^{N_0\times M}$. In words, this involves the following operations in sequence: computing the dirty image by applying $\boldsymbol{\Phi}^\dagger$, an N-sized Fourier transform $\mathbf{F}$, a subsampling through $\mathbf{S}$, retaining only those dimensions corresponding to non-zero (or significant) singular values of $\boldsymbol{\Phi}$, and finally, a weighting $\boldsymbol{\Sigma}_0{}^{-1}$. The weighting ensures that the noise covariance matrix in the embedded dimension has diagonal elements corresponding to the original variance of the measurement noise $\boldsymbol{n}$. We also note that $\mathbf{R}_{\text{sing}}$ has a fast implementation as it consists of diagonal, sparse and Fourier matrices only. Simulations were performed to compare this proposed dimensionality reduction $\mathbf{R}_{\text{sing}}$ with a weighted subsampled version of the standard 'gridding' operation $\mathbf{G}$ performed in radio interferometry, given by $\mathbf{R}_{\text{grid}} = \overline{\mathbf{WSG}} \in \mathbb{C}^{\overline{N}\times M}$ ($\overline{N} \leq 4N$ for an oversampling factor of 2 in the computation of the Fourier transform).

Here we show reconstruction results on an $N = 256 \times 256$ model image of the M31 Galaxy. $M = 50N$ continuous visibilities are sampled following a realistic SKA-like *uv* coverage. The 'input' SNR, defined as $\text{ISNR} = 20\log_{10}(\|\boldsymbol{y}_0\|_2/\|\boldsymbol{n}\|_2)$ with $\boldsymbol{y}_0 = \boldsymbol{\Phi x}$ being visibilities uncorrupted by noise, is set to 30 dB. Similarly, the 'output' SNR is defined as $\text{OSNR} = 20\log_{10}(\|\boldsymbol{x}\|_2/\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2)$, $\hat{\boldsymbol{x}}$ being the reconstructed image. Our simulations show that an OSNR of ≈25 dB is reached in the absence of embedding, although at a heavy computational cost owing to the 3.2 million visibilities. Reconstruction after $\mathbf{R}_{\text{grid}}$ achieves an OSNR of ≈25 dB with $M_L = 4N$. Crucially, much more aggressive dimensionality reduction is possible with $\mathbf{R}_{\text{sing}}$, which obtains an OSNR of ≈24.5 dB, but from a data size $M_L = 0.25N$. The robustness of $\mathbf{R}_{\text{sing}}$ compared to $\mathbf{R}_{\text{grid}}$ is illustrated in the figure below through the reconstructed, error and residual images for both methods embedding data to 5% of image size. We note that image reconstruction from data embedded to $0.05N$ using $\mathbf{R}_{\text{grid}}$ is poor compared to $\mathbf{R}_{\text{sing}}$, as is seen from the artefacts in the reconstructed image and the more prominent residual structure in the bottom row for $\mathbf{R}_{\text{grid}}$.



References

[1] S. V. Kartik, R. E. Carrillo, J.-P. Thiran, and Y. Wiaux, "A fourier dimensionality reduction model for big data interferometric imaging," *arXiv:1609.02097 [astro-ph]*, Sep. 2016.

Deep Generative Models of Galaxy Images for the Calibration of the Next Generation of Cosmological Surveys

François Lanusse* Siamak Ravanbakhsh† Rachel Mandelbaum* Peter Freeman‡ Jeff Schneider† Barnabás Póczos†
* McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15215, USA
† School of Computer Science, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15215, USA
‡ Department of Statistics, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15215, USA

***Abstract*—In order to shed some much needed light on fundamental questions such as the nature of dark energy, most next generation cosmological surveys will probe the Universe by accurately measuring the apparent shapes of billions of distant galaxies. As current shape measurement methods suffer from various biases, a precise calibration will be necessary in order to meet the requirements of the science analysis. In this work, we investigate the application of recent advances in deep conditional generative models as a way to synthesize large calibration sets of high-quality realistic galaxy images. We implement a conditional variant of the Variational Auto-Encoder, a state-of-the-art deep generative model based on variational Bayesian methods, and fit the model on data from the COSMOS survey. We assess the quality of the generated images with an extensive set of morphological statistics and find excellent agreement with real galaxy populations.**

I. Introduction

Weak gravitational lensing, i.e. the minute deflection of the light from distant objects by the intervening massive large scale structures of the Universe, has long been identified as one of the most powerful probes to investigate the nature of dark energy. As such, weak lensing is at the heart of the next generation of cosmological surveys such as LSST [1] or Euclid [2] . One particularly crucial source of systematic errors in these surveys comes from the shape measurement algorithms tasked with estimating galaxy shapes (from which the weak lensing signal is extracted). The last community challenge [3] to assess the quality of state-of-the-art shape measurement algorithms has in particular demonstrated that all current methods are biased to various degrees and, more importantly, that these biases depend on the details of the galaxy morphologies. These biases can be measured and calibrated by generating mock observations where a known lensing signal has been introduced and comparing the resulting measurements to the ground-truth. Producing these mock observations however requires input galaxy images of higher resolution and S/N than the simulated survey which typically implies extremely expensive and scarce space-based observations. The goal of this work is to train a deep generative model on already available Hubble Space Telescope (HST) data which can then be used to sample new galaxy images conditioned on parameters such as magnitude, size or redshift and exhibiting complex morphologies. Such model can allow us to inexpensively produce large set of realistic calibration images.

II. Conditional Variational-AutoEncoder

We implement a conditional generative model based on the Variational Auto-Encoder (VAE) framework introduced in [4]. The observations $\boldsymbol{x}$ are assumed to be generated by a random process conditioned on some observed properties $\boldsymbol{y}$ and involving an unobserved latent variable $\boldsymbol{z}$ according to some parametric distribution $p_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{y}) = p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{y})p_{\boldsymbol{\theta}}(\boldsymbol{z}|\boldsymbol{y})$, where multi-layered neural networks can be used to provide an expressive form of the prior $p_{\boldsymbol{\theta}}(\boldsymbol{z}|\boldsymbol{y})$ and likelihood $p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{y})$ distributions. For instance, in this work we assume a Gaussian observation model so that $p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{y}) = \mathcal{N}(\mu_{\boldsymbol{\theta}}(\boldsymbol{z}, \boldsymbol{y}), \Sigma)$ where $\mu_{\boldsymbol{\theta}}$ is a deep convolutional neural network. Assuming a given set of parameters $\boldsymbol{\theta}$, sampling from the model simply involves sampling $\boldsymbol{z}$ from the prior, and then sample from $p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{y})$. Fitting the model to the data is more complex however as one needs to adjust its parameters $\boldsymbol{\theta}$ as to maximize the marginal likelihood of the model $p_\theta(\boldsymbol{x}|\boldsymbol{y}) = \int p_\theta(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{y})p_\theta(\boldsymbol{z}|\boldsymbol{y})d\boldsymbol{z}$, which is in practice intractable, as is the posterior density $p_{\boldsymbol{\theta}}(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y})$. The VAE approach is to introduce an additional recognition model $q_\phi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y})$, also encoded as a multi-layer neural network, to approximate the true posterior $p_{\boldsymbol{\theta}}(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y})$. The parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are then fitted jointly by optimizing the following variational lower bound on the marginal log likelihood of the model:

$$\log p_\theta(\boldsymbol{x}|\boldsymbol{y}) \geq -D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y}) \parallel p_\theta(\boldsymbol{z}|\boldsymbol{y})) + \mathbb{E}_{z\sim q_\phi(z|x,y)}\left[\log p_\theta(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{y})\right]$$

Contrary to the original likelihood, this variational bound becomes easily tractable and efficient optimization of the large number of parameters of the model is made possible by stochastic gradient descent algorithms.

III. Results on COSMOS galaxies

We train our model on deep space-based observations of galaxy images from the COSMOS survey [5], conditioning on redshift, size or apparent magnitude. The quality of the model is assessed by computing an extensive set of galaxy morphology statistics on the generated images. Beyond simple second moment statistics such as size and ellipticity, we apply more complex statistics specifically designed to be sensitive to disturbed galaxy morphologies [6]. We find excellent agreement between the morphologies of real and model generated galaxies.

References

[1] LSST Science Collaboration *et al.*, "LSST Science Book, Version 2.0," *ArXiv e-prints*, Dec. 2009.
[2] R. Laureijs, J. Amiaux, S. Arduini, J. . Augueres *et al.*, "Euclid Definition Study Report," *ArXiv e-prints*, Oct. 2011.
[3] R. Mandelbaum, B. Rowe, J. Bosch, C. Chang, F. Courbin *et al.*, "The Third Gravitational Lensing Accuracy Testing (GREAT3) Challenge Handbook," *The Astrophysical Journal Supplement*, vol. 212, p. 5, May 2014.
[4] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
[5] N. Scoville, H. Aussel, M. Brusa, P. Capak *et al.*, "The Cosmic Evolution Survey (COSMOS): Overview," *The Astrophysical Journal Supplement Series*, vol. 172, pp. 1–8, Sep. 2007.
[6] P. E. Freeman, R. Izbicki, A. B. Lee, J. A. Newman *et al.*, "New image statistics for detecting disturbed galaxy morphologies at high redshift," *Monthly Notices of the Royal Astronomical Society*, vol. 434, pp. 282–295, Sep. 2013.

Data-driven, interpretable photometric redshifts for deep galaxy surveys with unrepresentative training data

Boris Leistedt*§ and David Hogg*
* Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA
§ Einstein Fellow

***Abstract*—We briefly summarize the novel method presented in [1] for estimating photometric redshifts when only unrepresentative spectroscopic training data are available.**

Testing cosmological models with deep imaging surveys such as the Dark Energy Survey and LSST requires accurate photometric redshifts of millions of galaxies in extended redshift ranges. However, very few spectroscopic observations of faint, high redshift galaxies are available. As a result, photometric redshifts rise as a major source of uncertainty in the exploitation of current and upcoming surveys.

Standard methods for obtaining point estimates and posterior distributions for the redshift of a galaxy given noisy flux measurements are based on template fitting or machine learning algorithms. Template fitting involves forward modelling the spectral energy distribution of galaxies and redshifting them. This is a well defined parameter estimation problem, but current models and templates are insufficient to describe deep imaging surveys at the required statistical accuracy. Furthermore, they fail to capture the complexity of real flux measurements. Machine learning methods resolve this issue, but require large representative training data, which are not available for ongoing or upcoming galaxy surveys. We derive a conceptually novel method which combines the advantages of template fitting and machine learning and is capable of exploiting unrepresentative training data.

The photometric flux in a band b of a galaxy (or a quasar) of rest frame luminosity density $L_\nu(\lambda_{\rm rest})$ (coined SED) at redshift z and observed wavelength λ reads

$$F_b(z) = \frac{1+z}{4\pi D_L^2(z)}\ C_b^{-1} \int_0^\infty L_\nu\left(\frac{\lambda}{1+z}\right)\ W_b(\lambda)\ {\rm d}\lambda/\lambda \qquad (1)$$

where D_L is the luminosity distance and C_b denotes a normalization constant which depends on the filter response $W_b(\lambda)$ and on the photometric system of interest.

As in template fitting methods, we introduce a variable t labelling galaxy types or classes, described by a (continuous or discrete) ensemble of SEDs $L_\nu(\lambda, t)$, so that the the flux becomes $F_b(z, t)$.

For a **target galaxy** of interest, the posterior distribution on its redshift given a set of noisy photometric bands $\{\hat{F}_b\}$ is

$$p(z|\{\hat{F}_b\}) \propto \int {\rm d}t\ p(\{\hat{F}_b\}|z,t)\ p(z,t) \approx \sum_i w_i\ p(\{F_b\}|z,t_i) \qquad (2)$$

The last equation assumes that we model a finite number of types from a training set, with the weights capturing prior information. Each type t_i is constructed from a galaxy from a **training set**, which consists of noisy photometric fluxes $\{\hat{F}_b\}_i$ and its redshift z_i (*e.g.,* spectroscopic). Hence, for each pair of target and training galaxies, we aim to compute

$$\begin{aligned} p(\underbrace{\{\hat{F}_b\}}_{\rm target}|z,t_i) &= p(\{\hat{F}_b\}|z,\underbrace{z_i,\{\hat{F}_b\}_i}_{\rm training}) \qquad (3)\\ &= p\left(\{\hat{F}_b\}|\{F_b(z,t_i)\}\right)\ p\left(\{F_b(z,t_i)\}|z_i,\{\hat{F}_b\}_i\right). \end{aligned}$$

The first term is the likelihood function, comparing the predicted and measured fluxes, and is usually a multivariate Gaussian. For the second term, we will use a Gaussian Process,

$$F(b,z) \sim \mathcal{GP}\Big(\mu^F(b,z),\ k^F(b,b',z,z')\Big) \qquad (4)$$

which we will fit to the fluxes of the training galaxy $\{\hat{F}_b\}_i$ at its redshift z_i. Thus, $p\left(F_b(z,t_i)|z_i,\{\hat{F}_b\}_i\right)$ becomes a standard prediction for a Gaussian process with 2 input dimensions, redshift and photometric band (described by filter responses).

We want the mean function μ^F and the kernel k^F to capture the expected correlations across redshift and bands resulting from the known setup and physics of the problem: the bands have known responses $\{W_b(\lambda)\}$, and galaxy SEDs are redshifted according to Equation (1). We model the latent, underlying SED of each training galaxy as a linear sum of templates $T_\nu^k(\lambda)$ (*e.g.,* taken from a standard template fitting method) and residuals that take the form of a zero-mean Gaussian Process with kernel $k(\lambda,\lambda')$. Therefore,

$$L_\nu(\lambda) = \underbrace{\sum_k \alpha_k T_\nu^k(\lambda)}_{\rm templates} + \underbrace{R_\nu(\lambda)}_{\rm residuals} \sim \mathcal{GP}\Big(\sum_k \alpha_k T_\nu^k(\lambda),\ k(\lambda,\lambda')\Big)$$

Since Equation (1) is a linear operation on L_ν, the fluxes $F(b,z)$ are indeed a Gaussian Process. As described in [1], closed analytical forms can be derived for the mean function μ^F and the kernel k^F for specific descriptions of the filter responses and the kernel k. In this case, they capture correlations allowed by redshifted SEDs.

The method presented above delivers interpretable redshift posterior distributions based on a data-driven model trained on measured fluxes. It is conceptually different from existing machine learning and template fitting photo-z methods but combines their main advantages. Importantly, the method does not require the training set to be representative of the target data, *i.e.,* for them to have similar redshift or flux distributions. Furthermore, the photometric bands in the training and target set need not to be identical. Instead, the training galaxy sample needs only to be as diverse as the target galaxies in terms of SEDs; better redshift, band and flux coverage will only improve the predictability of the model. This approach is the first capable of exploiting heterogeneous training sets, consisting of shallow and deep spectroscopic galaxy samples with different redshift and wavelength coverages.

As detailed in [1], this novel approach benefits from various other computational and methodological advantages, and performs well on real data even in the presence of shallow training data.

References

[1] B. Leistedt and D. Hogg, "Data-driven, interpretable photometric redshifts for galaxy and quasar surveys with heterogeneous, unrepresentative training data." *In prep*.

PURIFYing real radio interferometric observations

Luke Pratley*, Jason D. McEwen*, Mayeul d'Avezac† Rafael E. Carrillo‡, Alexandru Onose§, Yves Wiaux§
* Research Software Development Group, Research IT Services University College London (UCL), London WC1E 6BT UK
† Mullard Space Science Laboratory (MSSL), University College London (UCL), Holmbury St Mary, Surrey RH5 6NT, UK
‡ Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne CH-1015, Switzerland
§ Institute of Sensors, Signals, and Systems, Heriot-Watt University, Edinburgh EH14 4AS, UK

***Abstract*—Next-generation radio interferometers, such as the Square Kilometre Array (SKA), will revolutionise our understanding of the universe through their unprecedented sensitivity and resolution. However, standard methods in radio interferometry produce reconstructed interferometric images that are limited in quality and they are not scalable for big data. In this work we apply and evaluate alternative interferometric reconstruction methods that make use of state-of-the-art sparse image reconstruction algorithms motivated by compressive sensing, which have been implemented in the PURIFY software package. In particular, we implement and apply the proximal alternating direction method of multipliers (P-ADMM) algorithm presented in a recent article. We apply PURIFY to real interferometric observations. For all observations PURIFY outperforms the standard CLEAN, where in some cases PURIFY provides an improvement in dynamic range by over an order of magnitude. The latest version of PURIFY, which includes the developments presented in this work, is made publicly available.**

I. Introduction

Radio interferometry allows imaging of the radio universe at higher resolution and sensitivity than possible with a single radio telescope. Image reconstruction methods are needed to reconstruct the true sky brightness distribution from the raw data acquired by the telescope, which amounts to solving an ill-posed inverse problem. Traditional methods, which are mostly variations of the Högbom CLEAN algorithm [1], do not exploit modern state-of-the-art image reconstruction techniques.

Next-generation radio interferometers, such as the Square Kilometer Array (SKA; [2]), must meet the challenge of processing and imaging extremely large volumes of data. These experiments have ambitious, high-profile science goals, including detecting the Epoch of Re-ionisation (EoR) [3]. If these science goals are to be realised, state of the art methods in image reconstruction are needed to process big data and to reconstruct images with high fidelity.

In [5] we implement the P-ADMM algorithm developed by [4] in the PURIFY software package, which has been entirely redesigned and re-implemented in C++, and apply it to observational data from the VLA and the ATCA. The previous version of PURIFY supported only simple models of the measurement operator modelling the telescope. PURIFY now supports a wider range of more accurate convolutional interpolation kernels (for gridding and degridding). We found that the Kaiser-Bessle kernel performs as well as the prolate spheroidal wave funtions. Additionally, PURIFY provides higher dynamic range images than CLEAN on real observations, sometimes by an order of magnitude improvement. Figure 1 shows an example of a CLEAN and PURIFY reconstruction of the radio galaxy 3C129. The PURIFY reconstructions produce higher dynamic range without the need for post processing to create a restored image.

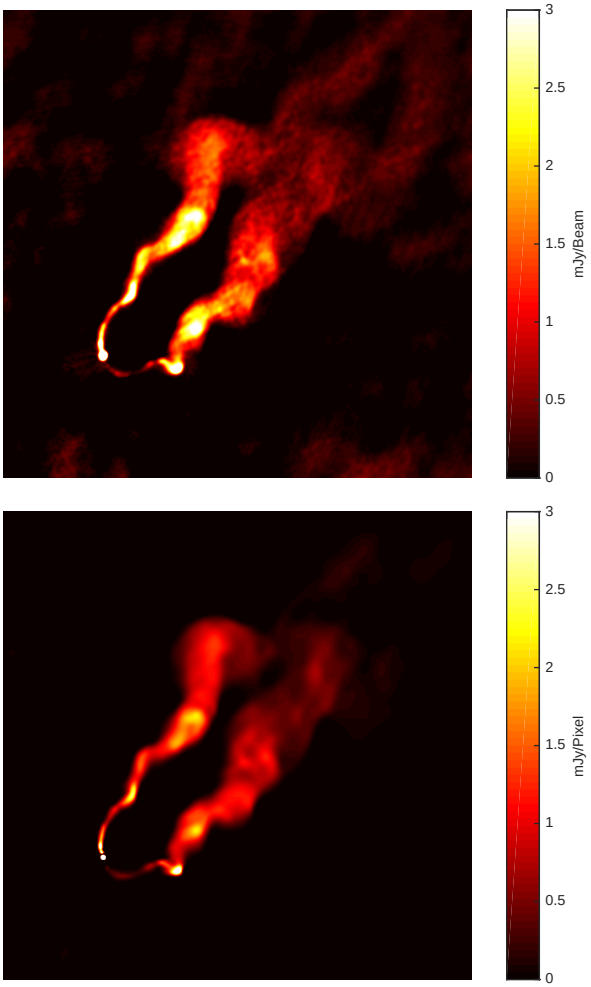


Fig. 1. The top and bottom show CLEAN and PURIFY reconstructions of the radio galaxy 3C129 respectively. The PURIFY recosntruction shows less contamination with higher dynamic range than the CLEAN reconstruction. Additionally, the PURIFY reconstruction does not require post processing to create a restored image. The details of these reconstructions can be found in [5].

References

[1] J. A. Högbom, "Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines," A&AS, vol. 15, p. 417, Jun. 1974.

[2] P. Dewdney, W. Turner, R. Millenaar, R. McCool, J. Lazio, and T. Cornwell, "Ska1 system baseline design," *Document number SKA-TEL-SKO-DD-001 Revision*, vol. 1, no. 1, 2013.

[3] L. Koopmans and et al, "The Cosmic Dawn and Epoch of Reionisation with SKA," *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, p. 1, Apr. 2015.

[4] A. Onose, R. E. Carrillo, A. Repetti, J. D. McEwen, J.-P. Thiran, J.-C. Pesquet, and Y. Wiaux, "Scalable splitting algorithms for big-data interferometric imaging in the SKA era," MNRAS, vol. 462, pp. 4314–4335, Nov. 2016.

[5] L. Pratley, J. D. McEwen, M. d'Avezac, R. E. Carrillo, A. Onose, and Y. Wiaux, "Robust sparse image reconstruction of radio interferometric observations with PURIFY," MNRAS, *(submitted)*, Oct. 2016.

Joint imaging and DDEs calibration for radio interferometry

Audrey Repetti, Jasleen Birdi, and Yves Wiaux
Institute of Sensors, Signals and Systems, Heriot-Watt University, Edinburgh EH14 4AS, UK.

***Abstract*—In the context of radio interferometry, the objective is to find an estimate of an unknown image of the sky from degraded observations, acquired through an antenna array. In the theoretical case of a perfectly calibrated array, it has been shown that solving the corresponding imaging problem by iterative algorithms based on convex optimization and Compressive Sensing (CS) theory can be competitive with classical algorithms such as CLEAN. However, in practice, antenna-based gains are unknown and have to be modelled during the calibration process. The latter aims to estimate the DIEs and DDEs related to each antenna of the interferometer. In this work, we propose an alternated minimization algorithm to estimate jointly the DIEs/DDEs and the image while promoting its sparsity in a dictionary, leveraging the CS theory.**

In radio interferometry, the imaging problem of finding an estimation of an original unknown image $\overline{\mathbf{x}} = (\overline{\mathsf{x}}(n))_n \in \mathbb{R}_+^N$ from complex visibilities $\mathbf{y} \in \mathbb{C}^M$ can be formulated as an inverse problem. More precisely, for an interferometer with n_a antennas, we have $M = n_a(n_a-1)/2$ measurements acquired in the Fourier domain of the image of interest, from antenna pairs indexed by $(\alpha,\beta) \in \{1,\dots,n_a\}^2$, with $\alpha < \beta$. Therefore, each degraded complex measurement $\mathsf{y}_{(\alpha,\beta)} \in \mathbb{C}$ acquired by the antenna pair (α,β) at the spatial frequency $\mathsf{u}_{\alpha,\beta} = \mathsf{u}_\alpha - \mathsf{u}_\beta$ can be modelled as

$$\mathsf{y}_{(\alpha,\beta)} = \sum_{n=-N/2}^{N/2-1} \mathsf{d}_\alpha(n)\mathsf{d}_\beta^*(n)\overline{\mathsf{x}}(n)\mathsf{e}^{-2\mathrm{i}\pi(\mathsf{u}_\alpha-\mathsf{u}_\beta)\frac{n}{N}} + \mathsf{b}_{(\alpha,\beta)}, \quad (1)$$

where $\mathbf{d}_\alpha = (\mathsf{d}_\alpha(n))_n \in \mathbb{C}^N$ represents a direction dependent effect (DDE) related to antenna α, and $\mathsf{b}_{(\alpha,\beta)}$ is a realization of a Gaussian additive noise. Note that direction independent effects (DIEs) can be seen as a special case of DDEs where $\mathbf{d}_\alpha = \delta_\alpha \mathbf{1}_N$, with $\delta_\alpha \in \mathbb{C}$ and $\mathbf{1}_N$ the unitary vector of dimension N.

When the antenna array is perfectly calibrated, i.e. when the DDEs are known, new methods based on both convex optimization and CS theory have been developed recently to find an estimation of $\overline{\mathbf{x}}$ from the observations (1) [1]. In particular, the estimated image can be defined as a solution to:

$$\underset{\mathbf{x}\in\mathbb{R}_+^N}{\text{minimize}} \ \frac{1}{2}\|\mathbf{G}\mathbf{F}\mathbf{x}-\mathbf{y}\|^2 + \eta\|\mathbf{\Psi}^\top\mathbf{x}\|_1, \quad (2)$$

where $\eta > 0$, $\mathbf{\Psi}^\top \in \mathbb{R}^{D\times N}$ is a given dictionary, $\mathbf{F} \in \mathbb{C}^{N\times N}$ denotes the Fourier matrix, and $\mathbf{G} \in \mathbb{C}^{M\times N}$ is a matrix containing on each line the antenna-based gain for the pair (α,β). Then, each line of $\mathbf{G}$ corresponds to the convolution of the Fourier transforms $\widehat{\mathbf{d}}_\alpha$ and $\widehat{\mathbf{d}}_\beta$ of $\mathbf{d}_\alpha$ and $\mathbf{d}_\beta$ respectively, centered at the frequency $\mathsf{u}_{\alpha,\beta}$.

However, in practice, antenna-based gains $(\mathbf{d}_\alpha)_\alpha$ have to be calibrated. The last years, several methods have been developed to estimate DIEs and/or DDEs, when the image is assumed to be known. In particular, in the StEFCal method [2] only DIEs are considered and the complex visibilities are rewritten as a data matrix $\mathbf{Y} \in \mathbb{C}^{n_a\times n_a}$ where, for every (α,β), $\mathsf{Y}_{\alpha,\beta} = \mathsf{y}_{(\alpha,\beta)}$. Note that due to the symmetry of measurements in (1) and reality of $\overline{\mathbf{x}}$, we have $\mathsf{Y}_{\beta,\alpha} = \mathsf{y}^*_{(\alpha,\beta)}$. The corresponding least squares minimization problem can be recasted as follows:

$$\underset{\mathbf{D}_1\in\mathbb{C}^{n_a\times N},\mathbf{D}_2\in\mathbb{C}^{n_a\times N}}{\text{minimize}} \ \frac{1}{2}\|\mathbf{D}_1\widehat{\mathbf{X}}\mathbf{D}_2^\top - \mathbf{Y}\|^2, \quad (3)$$

where $\mathbf{D}_1$ (resp. $\mathbf{D}_2$) is the matrix such that, each line α contains $\widehat{\mathbf{d}}_\alpha$ (resp. $\widehat{\mathbf{d}}^*_\alpha$) centered in u_α (resp. $-\mathsf{u}_\alpha$), and $\widehat{\mathbf{X}} \in \mathbb{C}^{N\times N}$ is the matrix containing on each line/column a shifted version of the Fourier transform of the image, to model the convolution operation. Note that for DIEs, each vector $\widehat{\mathbf{d}}_\alpha$ has one non-zero value. Thus, to solve (3), the StEFCal method needs to estimate only $2n_a$ values.

In this work, we propose a new method for the joint estimation of the original image, and the DDEs when both are unknown, based on a block-coordinate forward-backward algorithm [3]. It involves alternating between the estimation of the image and the DDEs, which are given as a solution to (2) and (3) respectively. In our approach, we assume that the DDEs have a bounded support in the Fourier domain, thus for the sparse matrices $\mathbf{D}_1$ and $\mathbf{D}_2$, we estimate only the coefficient values within this support (of size 1 when DDEs reduce to DIEs). Moreover, we consider constraints on the coefficient amplitudes of the DDEs, assuming that the amplitude of the central coefficient of $\widehat{\mathbf{d}}_\alpha$ is larger than the others. Finally, we make use of the prior information on the bright sources of the image to be estimated.

We analyze the performance of our method on simulated sky images of size 128×128, consisting of point sources, generated randomly on three intensity levels (Fig.1). While the first level is assumed to be known, the aim is to estimate the other two levels. We consider $n_a = 200$ randomly distributed antennas, and $\widehat{\mathbf{d}}_\alpha$ of support size 5×5. We performed simulations to reconstruct (i) the DIEs and the image (a) by combining StEFCal with an imaging algorithm, (b) using our method, and (ii) jointly the DDEs and the image with our method. The results show that the second level given in Fig. 1(center) is recovered only in case (ii). Finally, reconstructions have been compared in terms of SNR, on an average of 10 simulations, varying the random images, antennas distributions, DDE values and noise realizations. The SNR of the prior image (Fig.1 (left)) is equal to 38.8 dB, while for the reconstructed images, we have SNR= 37.8 dB for cases (i)(a-b) and SNR=53.2 dB for case (ii).

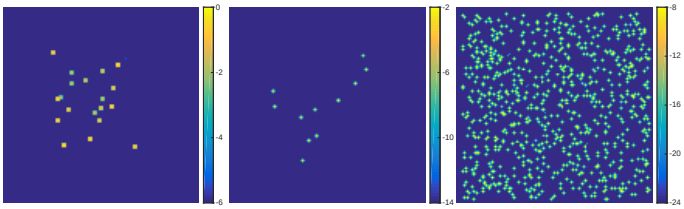


Figure 1. Images, in log scale, corresponding to the first level with the known bright sources $\overline{\mathbf{x}}_1$ (left), and to the fainter sources belonging to the second $\overline{\mathbf{x}}_2$ (center) and third $\overline{\mathbf{x}}_3$ (right) levels. We have $\overline{\mathbf{x}} = \overline{\mathbf{x}}_1 + \overline{\mathbf{x}}_2 + \overline{\mathbf{x}}_3$, with energy of $\overline{\mathbf{x}}_1$, $\overline{\mathbf{x}}_2$ and $\overline{\mathbf{x}}_3$ of the order of 1, 10^{-2} and 10^{-6}, respectively.

References

[1] Y. Wiaux, L. Jacques, G. Puy, A. M. M. Scaife and P. Vandergheynst, "Compressed sensing imaging techniques for radio interferometry," *Mon. Not. R. Astron. Soc.*, vol. 395, pp. 1733–1742, 2009.

[2] S. Salvini and S. J. Wijnholds, "Fast gain calibration in radio astronomy using alternating direction implicit methods: Analysis and applications," *A&A*, vol. 571, p. A97, 2014.

[3] E. Chouzenoux, J.-C. Pesquet and A. Repetti, "A block coordinate variable metric forward-backward algorithm," *Accepted in J. Global Optim.*, 2016.

A new perspective on turbulent Galactic magnetic fields

J.-F. Robitaille*, A. M. M. Scaife*, E. Carretti†‡, B. M. Gaensler§, J. D. McEwen¶, B. Leistedt‖** & S-PASS consortium
* Jodrell Bank Centre for Astrophysics, The University of Manchester, Oxford Road, Manchester M13 9PL, UK
†INAF Osservatorio Astronomico di Cagliari, Via della Scienza 5, 09047 Selargius (CA), Italy
‡CSIRO Astronomy and Space Science, PO Box 76, Epping, NSW 1710, Australia
§Dunlap Institute for Astronomy and Astrophysics, The University of Toronto, Toronto, ON M5S 3H4, Canada
¶Mullard Space Science Laboratory (MSSL), University College London (UCL), Surrey RH5 6NT, UK
‖Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK
**Department of Physics, New York University, 4 Washington Place, New York, NY 10003, USA

***Abstract*—We propose the comparison of two rotationally invariant decomposition techniques on linear polarisation data: the spin-2 spherical harmonic decomposition in two opposite parities, the E- and B-mode, commonly used for the cosmic microwave background analysis, and the multi-scale analysis of the gradient of linear polarisation, $|\nabla \mathbf{P}|$, used for the turbulence analysis of polarised synchrotron emission. We demonstrate that both decompositions have similar properties in the image domain and the spatial-frequency domain.**

I. Introduction

Surveys of the diffuse synchrotron emission are giving a new view of the Galactic magnetic field structure and are revealing its complexity at all spacial scales. However, to interpret the structures in such polarised data, we need robust analysis techniques. In this project, the spin-2 spherical harmonic decomposition and the gradient of Stokes parameters Q and U are used as complementary tools in order to develop a better understanding of the origin of energy sources in the turbulent magnetic field, the origin of peculiar magnetic field structures and their underlying physics.

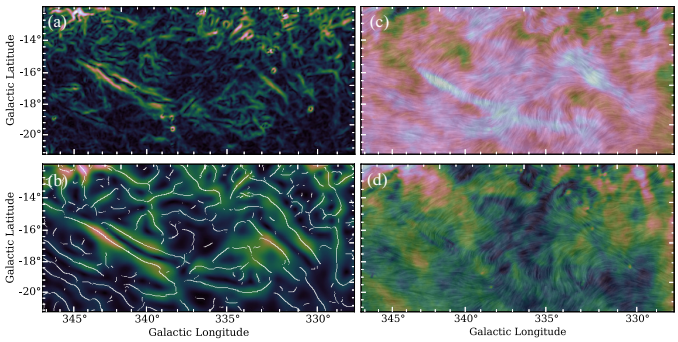


Fig. 1. (a) shows the $|\nabla \mathbf{P}|$ for a subregion of the synchrotron S-PASS survey [1] at its original resolution ($\sim$ 10 arcmin); (b) shows the $|\nabla \mathbf{P}|$ at a larger angular scale of $\sim$ 161 arcmin; (c) presents the B-mode decomposition overlaid with drapery structures showing the orientation of the magnetic field; (d) shows the polarised intensity, $|\mathbf{P}| = \sqrt{Q^2 + U^2}$, where intensity fluctuations are not sensitive to the polarisation angle. Colour scales are linear, where dark green is the minimum value and bright pink is the maximum value.

II. The multi-scale analysis of polarisation gradient

The gradient of linear polarisation measures the rate at which the polarisation vector traces out a trajectory in the Q–U plane as a function of position on the sky:

$$|\nabla \mathbf{P}| = \sqrt{\left(\frac{\partial Q}{\partial x}\right)^2 + \left(\frac{\partial U}{\partial x}\right)^2 + \left(\frac{\partial Q}{\partial y}\right)^2 + \left(\frac{\partial U}{\partial y}\right)^2}. \quad (1)$$

For the multi-scale analysis, multiple continuous wavelet transforms are calculated with different scales l on Q and U before the calculation of $|\nabla \tilde{P}(l, \mathbf{x})|$ in order to evaluate the scaling behaviour of the interstellar magnetic field [2]. The continuous wavelet transform is defined as

$$\tilde{f}(l, \mathbf{x}) = \begin{cases} \tilde{f}_1 = l^{-1} \int \psi_1[l^{-1}(\mathbf{x}' - \mathbf{x})] f(\mathbf{x}') d^2\mathbf{x}' \\ \tilde{f}_2 = l^{-1} \int \psi_2[l^{-1}(\mathbf{x}' - \mathbf{x})] f(\mathbf{x}') d^2\mathbf{x}'. \end{cases} \quad (2)$$

The two wavelet functions are

$$\begin{aligned} \psi_1(x, y) &= \partial\phi(x, y)/\partial x, \\ \psi_2(x, y) &= \partial\phi(x, y)/\partial y, \end{aligned} \quad (3)$$

where the function ϕ is a Gaussian distribution and l is the wavelet scaling factor. The multi-scale polarisation gradient is defined as

$$|\nabla \tilde{P}(l, \mathbf{x})| = \sqrt{|\tilde{Q}_1|^2 + |\tilde{U}_1|^2 + |\tilde{Q}_2|^2 + |\tilde{U}_2|^2}, \quad (4)$$

III. The spin-2 decomposition

The E- and B-mode are scalar representations of the pseudo-vectors Q and U. Their spherical harmonic coefficients are defined as $a_{E,\ell m} = -(a_{+2,\ell m} + a_{-2,\ell m})/2$ and $a_{B,\ell m} = i(a_{+2,\ell m} - a_{-2,\ell m})/2$, where $a_{\pm 2,\ell m}$ are the spherical harmonic coefficients of the spin-2 signal ${}_{\pm 2}(Q \pm iU)$ [3]. The spin-2 decomposition of the polarised synchrotron emission is a good strategy to highlight coherent features where a particular alignment of the magnetic field lines occurs (see Fig. 1).

IV. Power Spectra

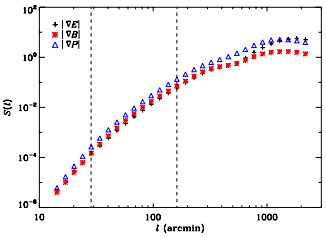


Fig. 2. The wavelet power spectra of $|\nabla \tilde{P}|$, $|\nabla \tilde{E}|$ and $|\nabla \tilde{B}|$ for the region presented in Fig. 1. It can be noticed that $|\nabla \tilde{\mathbf{P}}(\mathbf{l}, \mathbf{x})| \simeq \sqrt{|\nabla \tilde{E}(l, \mathbf{x})|^2 + |\nabla \tilde{B}(l, \mathbf{x})|^2}$. For this region, an excess in B-mode is present at intermediate scales and in E-mode at large scales.

Wavelet coefficients can be used to measure the energy transfer from large to smaller scales. The wavelet power spectrum is defined as $S_X(l) = \langle |\nabla \tilde{X}(l, \mathbf{x})|^2 \rangle_{\mathbf{x}}$, where $\tilde{X}$ can be $\tilde{P}$, $\tilde{E}$ or $\tilde{B}$. Wavelet coefficients can later be used to locate, in the map, the features responsible for the excess in E- or B-mode.

V. Conclusion

The two linear polarisation decomposition techniques can be used as complementary tools to describe and quantify fluctuations in the magneto-ionic medium (MIM). Local EB asymmetries correlated with high intensity $|\nabla \tilde{\mathbf{P}}|$ fluctuations can be inspected as local sources of energy injection and/or perturbation in the MIM.

References

[1] E. Carretti et al., "Giant magnetized outflows from the centre of the Milky Way," *Nature*, vol. 493, pp. 66–69, Jan. 2013.
[2] J.-F. Robitaille and A. M. M. Scaife, "Multiscale analysis of the gradient of linear polarization," *MNRAS*, vol. 451, pp. 372–382, Jul. 2015.
[3] M. Zaldarriaga and U. Seljak, "All-sky analysis of polarization in the microwave background," *PhysRevD*, vol. 55, pp. 1830–1840, Feb. 1997.

Spin-SILC: CMB polarisation component separation for next-generation experiments

Keir K. Rogers*, Hiranya V. Peiris†, Boris Leistedt‡, Jason D. McEwen§ and Andrew Pontzen*
* Department of Physics & Astronomy, University College London, London WC1E 6BT, UK
† Oskar Klein Centre for Cosmoparticle Physics, Stockholm University, Stockholm SE-106 91, Sweden
‡ Department of Physics, New York University, New York, NY 10003, USA
§ Mullard Space Science Laboratory, University College London, Surrey RH5 6NT, UK

***Abstract*—Spin-SILC is a component separation method that accurately extracts the cosmic microwave background (CMB) polarisation E and B modes from raw multi-frequency Stokes Q and U measurements of the microwave sky. It is an internal linear combination (ILC) method that uses spin wavelets to analyse the spin-2 polarisation signal $P = Q + iU$. The wavelets are additionally directional (non-axisymmetric). This allows different morphologies of signals to be separated and used in the analysis. The advantage of spin wavelets over standard scalar wavelets is to simultaneously and self-consistently probe scales and directions in the polarisation signal P and in the underlying E and B modes. Spin-SILC can be combined with pseudo- and pure E-B spin wavelet estimators to reliably extract the cosmological signal in the presence of complicated sky cuts and noise. These proceedings review the work in [1], [2].**

I. Motivations

Cosmology can constrain the tensor-to-scalar ratio of primordial fluctuations and hence the energy scale of inflationary expansion in the early universe, as well as properties of neutrinos, in maps of the polarisation of the CMB. These constraints are now limited by the contamination from Galactic dust and synchrotron emission in those maps. These polarised foreground signals demonstrate complex morphology, with filamentary structures that trace the Galactic magnetic field (see Fig. 1). They are poorly physically modelled. It is certain that blind source separation methods will form an essential part of disentangling the cosmological signal from polarised foregrounds. In particular, ILC methods are the most blind, only assuming knowledge of the CMB electromagnetic spectrum. ILC methods can be improved by the choice of wavelet frame in which they are localised.

II. Method

In the Spin-SILC method, we use a frame of spin, directional, scale-discretised wavelets [3]. When they are convolved with signals, they localise structure by spatial scale. Our wavelets are additionally directional, i.e. they are non-axisymmetric. This also localises by orientation on the sphere; in particular, the filamentary structures of polarised foregrounds can be separated. Moreover, our wavelets are built on the basis of spin spherical harmonics. This allows the spin-2 signal $P = Q+iU$ to be represented. E and B modes are respectively reconstructed by inverse scalar (spin-0) wavelet transforms of the real and imaginary parts of the spin wavelet coefficients. This is achieved thanks to the particular construction of our wavelets.

Spin-SILC is an ILC method executed in spin, directional wavelet space. The ILC reconstructs the CMB by co-adding multi-frequency maps of the microwave sky. It calculates the frequency weights by minimising the variance of the output map, while conserving the CMB component by inputting its electromagnetic spectrum. Spin-SILC localises these weights spatially, harmonically and directionally in the wavelet space. In this way, we use all the information in the polarisation signal $P = Q + iU$.

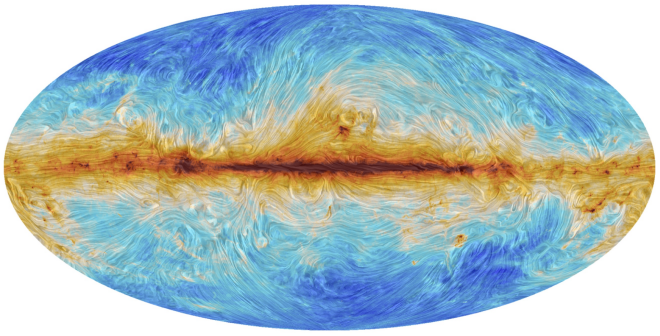


Fig. 1. A line integral convolution map of the *Planck* polarisation channel at 353 GHz. This maps the polarisation pattern of Galactic dust emission. It is dominated by complex filamentary structures. This figure is taken from [4].

III. Testing and future applications

Ref. [2] shows the results of testing Spin-SILC on full-mission *Planck* FFP8 simulations of the CMB, foregrounds and noise. The residuals between the reconstructed and input CMB are small in magnitude demonstrating the efficacy of the method. We tested the method on the biggest public dataset, *Planck*, where the polarisation data is dominated by noise. This limits all component separation methods. Nonetheless, with minimal tuning to the *Planck* dataset, our method matches the performance of existing methods; and even has smaller power spectrum residuals in higher signal-to-noise regimes [1]. It will be interesting to see how Spin-SILC performs for the next generation of high-sensitivity experiments, where, in particular, the use of directionality can be further optimised.

Future CMB polarisation experiments will normally only map a part of the sky. E-B mode decomposition is not well-defined in this setting, where, in particular, E modes can erroneously be counted as B. This E-B leakage problem can be resolved by combining Spin-SILC with the spin wavelet pure mode estimators of [5].

References

[1] K. K. Rogers, H. V. Peiris, B. Leistedt, J. D. McEwen, and A. Pontzen, "SILC: a new Planck Internal Linear Combination CMB temperature map using directional wavelets," *MNRAS*, vol. 460, pp. 3014–3028, Jan. 2016.
[2] ——, "Spin-SILC: CMB polarisation component separation with spin wavelets," *MNRAS*, vol. 463, pp. 2310–2322, Aug. 2016.
[3] J. D. McEwen, B. Leistedt, M. Büttner, H. V. Peiris, and Y. Wiaux, "Directional spin wavelets on the sphere," *IEEE TSP, submitted*, 2015.
[4] Planck Collaboration, R. Adam, P. A. R. Ade, N. Aghanim, Y. Akrami, M. I. R. Alves, F. Argüeso, M. Arnaud, F. Arroja, M. Ashdown, and et al., "Planck 2015 results. I. Overview of products and scientific results," *A&A*, vol. 594, p. A1, Sep. 2016.
[5] B. Leistedt, J. D. McEwen, M. Büttner, and H. V. Peiris, "Wavelet reconstruction of E and B modes for CMB polarisation and cosmic shear analyses," *ArXiv e-prints*, May 2016.

Estimated covariance matrices in large-scale structure observations

Elena Sellentin*,
* Imperial Centre for Inference and Cosmology (ICIC), Department of Physics, Imperial College, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, U.K.

***Abstract*—When computing the likelihood for most large-scale structure observations, a covariance matrix is needed that describes the data errors and their correlations. Usually, this covariance matrix is not known a priori, and may be estimated from simulations only. It thereby becomes a random object with some intrinsic uncertainty itself. We show how to infer parameters in the presence of such an estimated covariance matrix, by marginalising over the unknown true covariance matrix, conditioned on its estimated value. We then quantify the loss of precision of parameter constraints and describe how far away a given sky surveys are from the ideal case of a known covariance matrix. We point out that it is insufficient to estimate this loss by debiasing a Fisher matrix as previously done, due to a fundamental inequality that describes how biases arise in non-linear functions. We apply our results to DES Science Verification weak lensing data, detecting a 10% loss of information that increases their credibility contours. No significant loss of information is found for KiDS. For a Euclid-like survey, with about 10 nuisance parameters we find that 2900 simulations are sufficient to limit the systematically lost information to 1%, with an additional uncertainty of about 2%. Without any nuisance parameters 1900 simulations are sufficient to only lose 1% of information.**

I. Introduction

Cosmological parameter inference frequently assumes a Gaussian likelihood of the data, which is fully specified by a mean, which depends on cosmological parameters, and a covariance matrix which describes the measurement uncertainties. Due to the complexity of cosmological observations, such a covariance matrix is often not calculated from first principles. Approximate solutions are sometimes used, but is still more common to estimate the covariance matrix from numerical simulations of the experiment instead. Such simulations create N random samples of synthetic data sets. An unbiased estimator for the covariance matrix is then

$$\mathsf{S} = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T, \tag{1}$$

where the $\boldsymbol{X}_i$ are simulated data sets, and $\bar{\boldsymbol{X}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{X}_i$ is the ensemble mean over all simulations. The decisive difference between an approximately derived covariance matrix, and one estimated via Eq. (1), is that although both may contain systematic errors, the latter is additionally a function of the random data sets $\boldsymbol{X}_i$ and therefore a random variable with statistical uncertainties itself. Similarly, all functions of such an estimated covariance matrix will again be random. This impacts cosmological parameter inference in the following ways.

II. Deformation of the likelihood

The shape of the likelihood is being deformed. Instead of a Gaussian likelihood with a true covariance matrix, the likelihood of a sample-estimated covariance matrix can be shown to be an adapted t-distribution. The proof utilizes that the estimator Eq. (1) follows a Wishart distribution, centered on the unknown true covariance matrix. This allows a construction of a prior P for the unknown true covariance matrix $\boldsymbol{\Sigma}$, over which the originally Gaussian distribution of the data can then be marginalized. In other words, the likelihood of observed data $\boldsymbol{X}_\mathrm{o}$, given a covariance matrix S from N simulations, $P(\boldsymbol{X}_\mathrm{o}|\boldsymbol{\mu}, \mathsf{S}, N)$, is the integral

$$\begin{aligned} P(\boldsymbol{X}_\mathrm{o}|\boldsymbol{\mu}, \mathsf{S}, N) &= \int \mathrm{d}\boldsymbol{\Sigma}\, G(\boldsymbol{X}_\mathrm{o}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) P(\boldsymbol{\Sigma}|\mathsf{S}, N) \\ &= \frac{\bar{c}_p |\mathsf{S}|^{-1/2}}{\left[1 + \frac{(\boldsymbol{X}_\mathrm{o}-\boldsymbol{\mu})^T \mathsf{S}^{-1}(\boldsymbol{X}_\mathrm{o}-\boldsymbol{\mu})}{N-1}\right]^{\frac{N}{2}}}, \end{aligned} \tag{2}$$

where the last line is a modified t-distribution and $\bar{c}_p$ is a normalization constant.

III. Information loss

Information about the parameters in the mean $\boldsymbol{\mu}$ is lost, due to the uncertainty of the covariance matrix. This information can be restored if the number of simulations is increased. Accordingly, simulation numbers in the range of multiple ten-thousands are currently often being called for in cosmological discussions. These numbers were derived from Fisher-matrix calculations of an approximate likelihood, before the more accurate t-distribution discussed above became available. Going beyond the level of Fisher matrices, and utilizing the t-distribution, these numbers can be shown to be lower than expected so far, see e.g. the numbers here mentioned in the abstract, or Fig. 6 of [1] for other surveys. In [1], it was also found that although the information loss depends on the number of simulations, it depends often more strongly on the number of nuisance parameters. We therefore advocate reserving some computational time for running slow but sophisticated simulations which help to understand better the role of nuisance parameters, rather than running as many speedy simulations as possible.

This workshop contribution is based on the results presented in [1] and [2] .

References

[1] E. Sellentin and A. F. Heavens, "Quantifying lost information due to covariance matrix estimation in parameter inference," *ArXiv e-prints*, Sep. 2016.

[2] ——, "Parameter inference with estimated covariance matrices," *MNRAS*, vol. 456, pp. L132–L136, Feb. 2016.

Laplace Beamshapes for Phased-Array Imaging

Matthieu Simeoni[*,†], Paul Hurley[*]
[*]IBM Zurich Research Laboratory, Rüschlikon; [†]École Polytechnique Fédérale de Lausanne (EPFL), Lausanne.

***Abstract*—The imaging capabilities of phased-array systems are governed by the properties of their array beamshape, directly linked to the instrument impulse response. To ensure good spatial resolution, beamshapes are designed with a very narrow main lobe, at the cost of a complex sidelobe structure, potentially leading to severe image artifacts. We propose the use of a new beamshape, called the Laplace beamshape, built with the Flexibeam framework. This beamshape trades spatial resolution for smoother sidelobes, resulting in an artifact-free image that is much easier to process. This tradeoff can be optimally assessed through a single parameter of the beamshape, allowing the analyst to perform a multi-scale analysis.**

Extended Abstract

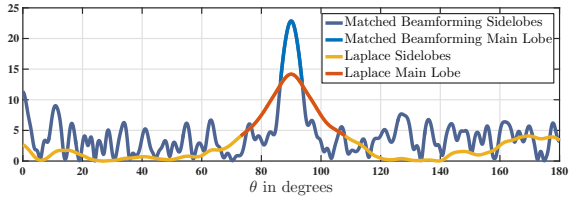
Beamforming combines networks of antennas coherently so as to achieve specific radiation patterns with desirable properties. For simplicity, we restrict our attention to 2D-beamforming in all that follows. Assuming hence an array of L antennas with positions $\boldsymbol{p}_1, \dots, \boldsymbol{p}_L \in \mathbb{R}^2$, the *beamformed signal* $y(t)$ is obtained by combining linearly the antenna signals $x_i(t)$:

$$y(t) = \boldsymbol{w}^H \boldsymbol{x}(t) = \sum_{i=1}^{L} w_i^* x_i(t), \quad \forall t \in \mathbb{R},$$

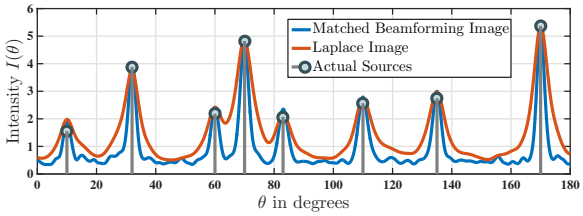
where $\boldsymbol{x}(t) := [x_1(t), \cdots, x_L(t)]$ and $\boldsymbol{w} := [w_1, \cdots, w_L] \in \mathbb{C}^L$. By properly choosing the beamforming weights w_i, it is possible to steer the antenna array towards specific directions in the sky. This is typically done via the popular *matched beamforming* method, which sets the weights to $w_i(\theta) = \left(e^{-j\frac{2\pi}{\lambda}\|\boldsymbol{p}_i\|\cos(\theta)}\right)/\sqrt{L}$, for a signal of wavelength $\lambda \in \mathbb{R}$ and a direction $\theta \in [0, 2\pi]$. By computing the variance of the beamformed signal $y(t)$, one can then obtain an estimate of the sky intensity at this location

$$I(\theta) = \mathbb{E}[y(t)y(t)^*] = \boldsymbol{w}^H(\theta)\Sigma\boldsymbol{w}(\theta), \tag{1}$$

where $\Sigma := \mathbb{E}[\boldsymbol{x}(t)\boldsymbol{x}^H(t)]$. Ranging across directions $\theta \in [0, 2\pi]$ produces an estimate $I(\theta)$ of the sky intensity field. This procedure is known as *imaging by beamforming*, or *B-scan imaging*, and is commonly used in radio-astronomy, sonar/radar and ultrasound imaging. The imaging capabilities of the instrument can then be assessed through its *point spread function*, response of the tool to an idealised point-source. We can show that this function is directly proportional to the squared magnitude of the array far-field radiation pattern, also called *beamshape*. Properties of this beamshape hence completely determine the quality of the image in Eq. (1). For optimal performance, two competing features must be optimised: the main lobe width, which controls the achievable angular resolution, and the sidelobes structure, which can translate into severe artifacts within the image. As described in [1], matched beamforming is attempting to achieve a beamshape as close as possible to a Dirac $\delta(\theta - \theta_0)$. As such, it typically performs quite well in terms of angular resolution, with a very narrow main lobe around the direction of focus θ_0, but often demonstrates strong sidelobes (see Fig. 1a). These prominent sidelobes are a consequence of the ill-defined nature of the Dirac function, which makes it a very difficult object to approximate. To avoid such complications, we hence propose to target a much better



(a) Matched beamforming versus Laplace beamshape (squared magnitude in logarithmic scale).



(b) B-scan Images.

Fig. 1: Imaging with Laplace and matched beamforming.

behaved function, namely a circular Laplace function:

$$\mathcal{L}(\theta) = \exp\left[-\frac{\sqrt{(\cos\theta - \cos\theta_0)^2 + (\sin\theta - \sin\theta_0)^2}}{\Theta}\right], \tag{2}$$

where $\Theta > 0$ controls the width of the main lobe. Being continuous and quickly decaying from the direction of focus θ_0, this function has a well-behaved Fourier spectrum, and can hence be easily approximated with a finite number of complex plane-waves. Moreover, its sharp central peak permits the very accurate estimation of source locations. To construct the *Laplace beamshape*, we used the very general Flexibeam framework introduced in [1]. Beamforming weights were obtained by sampling the so-called *beamforming function*, which, for the circular Laplace function is given by:

$$w_i = \omega(\boldsymbol{p}_i) = \frac{2\pi\Theta^2}{\left(1 + 4\pi^2\Theta^2\|\boldsymbol{p}_i\|^2\lambda^{-2}\right)^{3/2}} e^{-j\frac{2\pi}{\lambda}\|\boldsymbol{p}_i\|\cos(\theta_0)}.$$

As expected, the resulting beamshape exhibits much smoother sidelobes (see Fig. 1a), with most of its energy contained in the main lobe (∼86% against 74% for matched beamforming). As a result, the image obtained with the Laplace beamshape appears much smoother, facilitating enormously the recovery of the actual sources within the field. Observe that this smoother behaviour was obtained at the cost of lower angular resolution. This fundamental tradeoff can be formally assessed by varying the parameter Θ, leading to a multi-scale analysis of the image. The Laplace beamforming strategy above described readily extends to 3D beamforming, and this will be the subject of a future publication.

References

[1] P. Hurley and M. Simeoni, "Flexibeam: analytic spatial filtering by beamforming," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE*, March 2016.

Challenges of Extreme Dynamic Range Imaging: The Cygnus Files

Oleg Smirnov* †
* 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

***Abstract*—The new generation of radio interferometers, including the SKA precursors (such as MeerKAT, LOFAR, ASKAP, the upgraded JVLA, etc.) represent a large upgrade in terms of theoretical sensitivity and/or extremely large fields of view. The SKA itself will push these trends even further. Dynamic ranges (DR) in excess of 1 million should, in theory, become routine. However, at present only a few carefully hand-crafted data reductions have broken the million-DR barrier. I will present some case studies to illustrate why high DR remains a problem, and discuss the relevance of compressive sensing-based approaches in this context.**

The first case study is a 20-hour observation of the field around the source 3C147 between 1-2 GHz, using three configurations of the JVLA. This exhibits a world-record DR of 8 million to one, however the field is rather "simple" in terms of its spatial structure, and is therefore quite amenable to traditional deconvolution algorithms such as CLEAN. It does, however, represent significant challenges in the calibration of direction-dependent effects (DDEs), due to the rotating primary beam pattern of the JVLA. The high DR means even subtle instrumental effects are exacerbated, and must be carefully accounted for.

The second study is a 61-hour JVLA observation of the radio source Cyg A, one of the most iconic sources in radio astronomy. It is an morphologically complicated FR2-type radio galaxy, exhibiting a central source, relativistic jets, extremely bright "hotspots" where the jets create a shock front as they hit the intergalactic medium, and extended "lobes" resulting from this interaction. Even without calibration effects, Cyg A represents the ultimate imaging problem, and this observation is currently limited by deconvolution algorithms rather than sensitivity. I will show some results of applying CS-based approaches to this data, and discuss the way forward.

Space variant deconvolution of galaxy survey images

Jean-Luc Starck,*, Samuel Farrens,*, Fred Ngole,*
* CEA Saclay, Service dAstrophysique, Gif-sur-Yvette, 91191, France

***Abstract*—Removing the aberrations introduced by the Point Spread Function (PSF) is a fundamental aspect of astronomical image processing. The presence of noise in observed images makes deconvolution a nontrivial task that necessitates the use of regularisation. This task is particularly difficult when the PSF varies spatially as is the case for big surveys such as LSST or Euclid surveys. The first step is therefore to estimate accurately the PSF field. In practice, isolated stars provide a measurement of the PSF at a given location in the telescope field of view. Thus we propose an algorithm to recover the PSF field, using the measurements available at few these locations. This amounts to solving an inverse problem that we regularize using both matrix factorization and a sparsity. Then we show that, for these new surveys providing images containing thousand of galaxies, the deconvolution regularisation problem can be considered from a completely new perspective. In fact, one can assume that galaxies belong to a low-rank dimensional space. This work introduces the use of the low-rank matrix approximation as a regularisation prior for galaxy image deconvolution and compares its performance with a standard sparse regularisation technique. This new approach leads to a natural way to handle a space variant PSF. Deconvolution is performed using a Python code that implements a primal-dual splitting algorithm. The data set considered is a sample of 10 000 space-based galaxy images convolved with a known spatially varying Euclid-like PSF and including various levels of Gaussian additive noise. Performance is assessed by examining the deconvolved galaxy image pixels and shapes. The results demonstrate that the low-rank method performs as a well as sparsity for small samples of galaxies and outperforms sparsity for larger samples.**

Sampling methods and pipeline design in modern cosmology

Joe Zuntz*
* Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

***Abstract*—Cosmologists mostly follow a well-established approach to constraining physical models of the Universe and its contents with astronomical data: Bayesian exploration of model parameter posterior distributions using MCMC or similar approaches. Over the last decade the community has gradually become aware of newer and more powerful samplers than the traditional Metropolis-Hastings; I will present an overview of the many samplers included in our CosmoSIS framework and some notes on using different samplers effectively. I will also advocate a particular modular approach to structuring theory calculations in cases where model predictions themselves are difficult to make.**

I. Introduction

In many branches of science the answer to the complaint "Statistics is difficult!" is the retort "Just get more data!". In cosmology, as in many of the more interesting parts of science, getting new data is very, very expensive, and we must use all the statistical tools at our disposal to claw as much information from our existing data as possible. As well as getting the best return on a given investment, this lets us plan future experiments with more care.

Mathematically we cast this information extraction in a Bayesian framework as parameter estimation. We construct a likelihood function that models the probability of the data that we observed, given some theory and values of that theory's parameters. We vary those parameters using some exploration scheme, simple or complex, to obtain a map of their probability.

II. Sampling methods

We will generically refer to these exploration schemes as "sampling methods", since many (though not all) of them provide representative samples from the probability distribution whose histogram tends to the underlying distribution.

In this talk we will survey a number of these methods that have proved useful for general cosmological problems.We may crudely classify them as follows.

Classic methods include Metropolis-Hastings, in which a single serial chain uses likelihood ratios to decide whether to take some proposed jump, and Importance sampling, in which a previously obtained distribution is re-weighted to match a new one.

Grid-based methods explore a regular grid in parameter space, either naively, or more cleverly, as in the Snake sampler.

Ensemble methods, like Emcee, PMC, Kombine, and Multinest, use a large number of points that swarm across the space with some collective behavior.

A host of miscellaneously useful methods can also be usefully incorporated in the same framework, such as maximum likelihood methods, which climb to the peak of the likelihood, and the Fisher matrix, which assumes a known maximum likelihood and approximates the whole system as Gaussian.

I will briefly discuss our work evaluating these sampling methods, including how to choose one and configure it for a given problem

III. A typical LSST parameter problem

A typical problem of the type LSST will face in cosmological parameter estimation is to constrain cosmological parameters given a measurement of the correlations between shape measurements in two-dimensional radial slices of a sample of galaxies. In the ideal case for such a problem, to calculate the theoretical predictions of a theory, we must: calculate the background evolution of the universe; calculate the 3D spectrum of density fluctuations in the cosmic matter distribution; integrate this radially to obtain 2D spectra; and Hankel transform the spectrum into a correlation function.

In practice this relatively simple sequence becomes more and more complex as we introduce the systematic errors we will find in real data like that from LSST. We must account for redshift errors, which skew the kernel of our radial integration; galaxy shape measurement errors, which add multiplicative and additive distortions to our final signal; and theoretical uncertainties like intrinsic correlations in galaxy shapes, and the behavior of matter on small scales.

It is occasionally claimed that there is no general way to deal with systematic errors in data. This is false. The general methodology is: study the error; model it; parametrize it; and sample over its parameters at the same time as those of scientific interest.

This is the approach we I will advocate for in this talk.

IV. Structuring theory

Coupled to a generic set of sampling methods it has proved immensely useful to build a general framework for making theoretical predictions from cosmological models. I will advocate for the structure presented in our code CosmoSIS: a modular framework in which steps in cosmological calculations are split into their constituents, each one encoded as a shared library or python framework, and connected together via a strict data passing system that prevents tight coupling.

This framework provides natural splits where one can model systematic errors, eases debugging, comparison, and data serialization, and can help with sharing and verifying codes.

Our CosmoSIS code guides the scientist towards this way of thinking about physical problems and thus improves the quality of parameter estimation results.

V. Writing samplers

I will also discuss what those in the statistical community can do to make their sampling or optimization methods more useful to practicing scientists.

The theoretical predictions of cosmological theories rarely if ever take simple analytic forms; they are rather complicated numerical computations. This severely limits the utility of whole classes of method which rely on analytic derivatives of theory observables or similar.

Bespoke algorithms that need a great deal of fine tuning are also unlikely to be widely useful except in special circumstances - minimal human supervision is a major plus for a method.

Finally, we are in the era of parallel computing - a method that can make best use of large numbers of parallel cores is hugely more useful than a serial one like Metropolis-Hastings.

Medical imaging across spatial scales: Microscopic to macroscopic

Daniel C. Alexander
Centre for medical image computing, Department of Computer Science, University College London, Gower Street, London, WC1E 6BT, UK.

Abstract— **My talk will give an overview of microstructure imaging with diffusion MRI. The microstructure-imaging paradigm aims to estimate and map microscopic properties of biological tissue using a model that links those properties to the voxel scale MR signal. I'll go through the basic principles, key challenges, summarise the state of the art, and highlight current hot topics and future directions.**

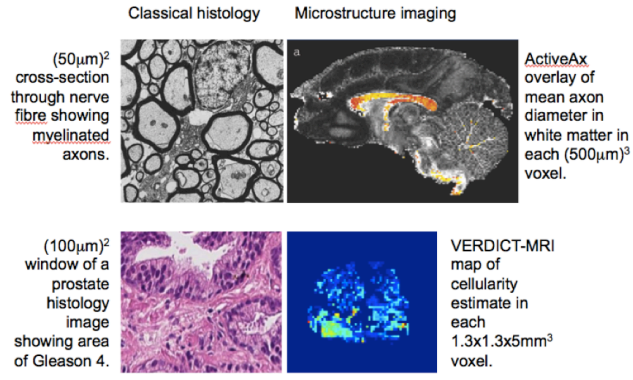


Fig. 2. Comparison of typical outputs from classical histology (left) and virtual histology using diffusion-MRI microstructure imaging (right) in two example applications: white matter imaging in the brain (top – from [1]) on estimating axonal geometry and the second on cancer imaging in the prostate i.e. cell density and organisation (bottom – from [2]).

Microstructure imaging ultimately aims to achieve virtual histology: estimating and mapping histological features of tissue using non-invasive imaging techniques, such as MRI. This virtual histology has several advantages over classical histology: i) it is non-invasive and avoids, for example, biopsy and its potential side-effects; ii) it views intact in-situ tissue avoiding disruptions that arise from tissue extraction and preparation; iii) it is non-destructive so enables repeat measurements for monitoring; iv) it provides a wide field of view, typically showing a whole organ or body, rather than the small samples typically obtained by biopsy or slices of fixated brain tissue for traditional histology. Figure 1 compares typical images from classical histology and microstructure imaging using diffusion MRI in two different scenarios. The clear advantage of classical histology is its level of anatomical detail; its submicron image resolution provides vivid insight into the cellular architecture of tissue, whereas microstructure imaging provides only statistical descriptions of the tissue over the extent of millimetre-sized image voxels. In some applications, the visualisation of individual cells is important; for example, a cancer histopathologist may need to identify the presence of one in a million mitotic cells. However, many histopathological tasks seek less specific information that reflects broad statistical changes over a wide extent of tissue: e.g. different density, shape, and configuration of cells discriminate different grades of prostate cancer. In such applications, precise detail is less important and the benefits of microstructure imaging can significantly outweigh those of traditional histology.

Microstructure imaging relies on a model that relates microscopic features of tissue architecture to MR signals. In general, the approach acquires a set of images with different sensitivities and fits a model in each voxel to the set of signals obtained from the corresponding voxel in each image. The process yields a set of model parameters in each image voxel, which constitute parameter maps of microscopic tissue features. Figure 2 illustrates with an example from diffusion MRI.

Diffusion MRI is a key modality for microstructure imaging, because of its unique sensitivity to cellular architecture. The technique sensitizes the MR signal to the dispersion of signal-bearing particles, typically water molecules, over diffusion times in the order of 1-100 milliseconds. The mean free-path over this time at room or body temperature is in the micrometer range, i.e. the cellular scale, so that the cellular architecture of the tissue strongly influences the dispersion pattern of the molecules. Thus diffusion MR measurements support inferences on tissue microstructure.

REFERENCES

[1] Alexander, D.C. et al, Neuroimage, vol. 52, pp.1374-89, 2010.
[2] Panagiotaki, E. et al, Cancer Research, vol. 74, pp. 1902-12, 2014.

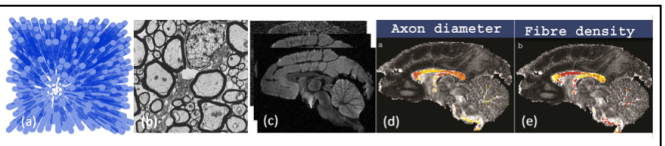


Fig. 1. ActiveAx [1] uses a simple geometric model (a), consisting of parallel non-abutting impermeable cylinders, of white matter microstructure (b). It fits the model in each voxel of a set of images (c) with varying diffusion contrast to recover estimates of cylinder size and packing density, which provide maps of indices of axon diameter (d) and axon density (e), respectively.

Multi-shell Sampling Scheme with Accurate and Efficient Transforms for Diffusion MRI

Alice P. Bates*, Zubair Khalid*, Rodney A. Kennedy* and Jason D. McEwen†
* Research School of Engineering, Australian National University, ACT 2601, Australia.
† Mullard Space Science Laboratory, University College London, Surrey RH5 6NT, UK.

***Abstract*—We propose a multi-shell sampling grid and develop corresponding transforms for the accurate reconstruction of the diffusion signal in diffusion MRI by expansion in the spherical polar Fourier (SPF) basis. The transform is exact in the radial direction and accurate, on the order of machine precision, in the angular direction. The sampling scheme uses an optimal number of samples equal to the degrees of freedom of the diffusion signal in the SPF domain.**

I. Introduction

The diffusion signal in diffusion MRI can be reconstructed from a finite number of measurements in q-space, where $\mathbf{q}$ is the diffusion wavevector, from which the brain's connectivity and microstructure properties can be determined. In diffusion MRI, the number of measurements that can be acquired is highly restricted due to the need for scan times to be practical in a clinical setting. For this reason, multi-shell sampling schemes, where samples are collected on multiple concentric spheres with different q-space radii, are commonly used rather than large Cartesian sampling grids. Existing multi-shell sampling schemes require more than the optimal number of samples, defined as the degrees of freedom in the basis used to reconstruct the diffusion signal, in order to allow for the accurate reconstruction of the diffusion signal and use least-squares to calculate coefficients, which is computationally intensive (e.g. [1]).

The spherical polar Fourier (SPF) basis [1] is a 3D complete orthonormal basis commonly used for reconstructing the diffusion signal. The normalised MR signal attenuation, $E(\mathbf{q})$ can be expanded in the SPF basis, as

$$E(\mathbf{q}) = \sum_{n=0}^{N-1} \sum_{\ell=0}^{L-1} \sum_{m=-\ell}^{\ell} E_{n\ell m} R_n(q) Y_\ell^m(\hat{\mathbf{q}}), \qquad (1)$$

where $\hat{\mathbf{q}} = \frac{\mathbf{q}}{|\mathbf{q}|}$, $q = |\mathbf{q}|$, $Y_\ell^m(\hat{\mathbf{q}})$ are spherical harmonic coefficients of degree ℓ and order m, and the expansion coefficients are given by

$$E_{n\ell m} = \langle E(\mathbf{q}), R_n(q) Y_\ell^m(\hat{\mathbf{q}}) \rangle. \qquad (2)$$

The radial functions are Gaussian-Laguerre polynomials R_n with[1]

$$R_n(q) = \left[\frac{2}{\zeta^{0.5}} \frac{n!}{\Gamma(n+1.5)} \right]^{0.5} \exp\left(\frac{-q^2}{2\zeta} \right) L_n^{1/2}\left(\frac{q^2}{\zeta} \right), \qquad (3)$$

where ζ denotes the scale factor and $L_n^{1/2}$ are the n-th generalised Laguerre polynomials of order half. The expansion Eq. (1) assumes that $E(\mathbf{q})$ is band-limited at radial order N and angular order L.

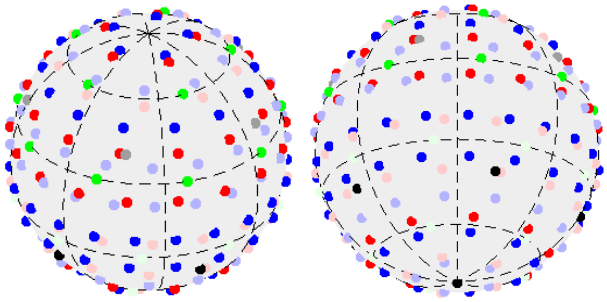
II. Multi-shell Sampling Scheme and SPF Transform

The 3D transform for calculating the diffusion signal coefficient (Eq.2) can be separated into transforms in the radial and angular directions due to the separability of the SPF basis. For the radial transformation, Gauss-Laguerre quadrature can be used, where N sampling nodes is sufficient for exact quadrature. The N shells of our proposed multi-shell sampling scheme are placed at $q_i = \sqrt{\zeta x_i}$ where x_i are the roots of the N-th generalised Laguerre polynomial of order a half. We determine the corresponding weights to be

$$w_i = \frac{0.5\zeta^{0.5}\Gamma(N+1.5) x_i e^{x_i}}{N!(N+1)^2 [L_{N+1}^{0.5}(x_i)]^2}. \qquad (4)$$

The number of shells required for accurate reconstruction was recommended as $N = 4$ in [2]. We set the scaling factor ζ so that shells are located at b-values within an interval of interest. In this work, we use a maximum b-value of 8000 $\mathrm{s/mm^2}$, resulting in shells at $b = 411.3, 1694.4, 4036.3$ and 8000 $\mathrm{s/mm^2}$.

For sampling within each shell, we use the recently proposed single-shell sampling scheme [3] which allows accurate reconstruction on the order of machine precision accuracy, has an efficient forward and inverse spherical harmonic transforms, and uses an optimal number of samples for the band-limited diffusion signal on the sphere, $L(L+1)/2$. The spherical harmonic band-limit, and therefore the number of samples within each shell, is determined using [4], where the authors determined the spherical harmonic band-limit L required to accurately reconstruct the diffusion signal at different b-values, the shells have $L = 3$, 5, 9 and 11 respectively. The proposed sampling scheme has a total of 132 samples. Fig. 1 shows the sampling scheme projected onto a single sphere, samples are shown in black, green, red and blue for each shell respectively. Locations where antipodal symmetry is used to infer the value of the signal are lighter in color.



(a) (b)

Fig. 1: (a) North pole view and (b) South pole view.

[1]We use a slightly different constant to [1] so that R_n are orthonormal with respect to a radial inner product.

References

[1] H. Assemlal, D. Tschumperlé, and L. Brun, "Efficient and robust computation of PDF features from diffusion MR signal," *Med. Image Anal.*, vol. 13, pp. 715–729, Jun. 2009.

[2] H. Assemlal, D. Tschumperlé, and L. Brun, "Evaluation of q-space sampling strategies for the diffusion magnetic resonance imaging," in *Med. Image Comput. Comput. Assist. Interv., MICCAI'2009*, London, UK, 2009, vol. 12, no. Pt 2, pp. 406–414.

[3] A. P. Bates, Z. Khalid, and R. A. Kennedy, "An optimal dimensionality sampling scheme on the sphere for antipodal signals in diffusion magnetic resonance imaging," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., ICASSP'2015*, Brisbane, Australia, Apr. 2015, pp. 872–876.

[4] A. Daducci, J. D. McEwen, D. V. D. Ville, J. P. Thiran, and Y. Wiaux, "Harmonic analysis of spherical sampling in diffusion MRI," in *Proc. 19th Ann. Meet. Int. Soc. Magn. Reson. Med.*, Jun. 2011.

Higher-order variational regularization approaches for imaging

Kristian Bredies*, Martin Holler*, Karl Kunisch*, Thomas Pock† and Benedikt Wirth‡
* Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, A-8010 Graz, Austria
† Institute for Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16, A-8010 Graz, Austria
‡ Institute for Computational and Applied Mathematics, University of Münster, Einsteinstraße 62, D-48149 Münster, Germany

***Abstract*—We present an overview and recent developments of convex regularization functionals involving higher-order derivatives and their applications in imaging. In particular, dedicated functionals for the reconstruction of multichannel images and spatio-temporal data as well as curvature-based regularization are discussed.**

I. Introduction

In variational imaging, smoothing-based image priors that lead to convex functionals are widely used due to their well-studied structure, well-posedness properties, resolution-independence and efficient numerical realizability. While functionals depending on first-order derivatives like the total variation (TV) are a popular choice, reconstruction quality can significantly benefit from higher-order smoothness information obtained with measure-based functionals.

II. Total generalized variation and extensions

In order to allow for jump discontinuities (as TV) as well as for incorporation of higher-order derivative information, the total generalized variation (TGV) has been proposed [1]. Its second-order version may also be written as [2]

$$\mathrm{TGV}^2_\alpha(u) = \min_{w\in \mathrm{BD}(\Omega)} \alpha_1\|\nabla u - w\|_{\mathcal{M}} + \alpha_0\|\mathcal{E}w\|_{\mathcal{M}}$$

where $\|\cdot\|_{\mathcal{M}}$ denotes the Radon norm and $\mathcal{E}$ is the symmetrized derivative. This can be interpreted as an optimal balancing between the first and second derivative of u where an optimal w represents the smooth part of the gradient ∇u. This way, both edge information as well as smooth regions can accurately be recovered, see Figure 1.

For the regularization of multichannel images, we have additional choices regarding the respective matrix and tensor norms for the (higher-order) derivatives [3]. This additional degree of freedom can be used to couple edge and smoothness information. In particular, an utilization of the nuclear norm for the derivative may enforce aligned edges (which correspond to rank-1 gradients). A second-order multichannel total generalized variation then reads as

$$\mathrm{TGV}^2_\alpha(u) = \min_{w\in \mathrm{BD}(\Omega)^k} \alpha_1\||\nabla u - w|_{\mathrm{nucl}}\|_{\mathcal{M}} + \alpha_0\||\mathcal{E}w|_{\mathrm{frob}}\|_{\mathcal{M}}$$

where $|\cdot|_{\mathrm{nucl}}$ and $|\cdot|_{\mathrm{frob}}$ denote the nuclear matrix norm and Frobenius tensor norm, respectively. This concept was successfully applied for joint MR-PET reconstruction, see Figure 2(a).

These notions can also be extended to dynamic data, where one has to balance spatial and temporal derivatives by a weighting factor. This factor then determines the temporal scale of regularization. Often, however, dynamic data admits several temporal scales. In order to account for the latter, an infimal convolution of functionals with different spatio-temporal weighting can be used:

$$\mathrm{ICTGV}^2_{\alpha,\beta_1,\beta_2}(u) = \inf_{u=u_1+u_2} \mathrm{TGV}^2_{\alpha,\beta_1}(u_1) + \mathrm{TGV}^2_{\alpha,\beta_2}(u_2),$$

where $\beta_1, \beta_2 > 0$ represent different spatio-temporal weighting parameters. Such an approach is, for instance, beneficial for cardiac MR imaging, see Figure 2(b).

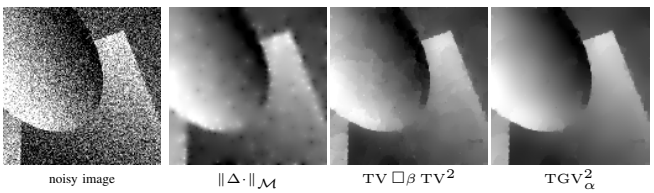
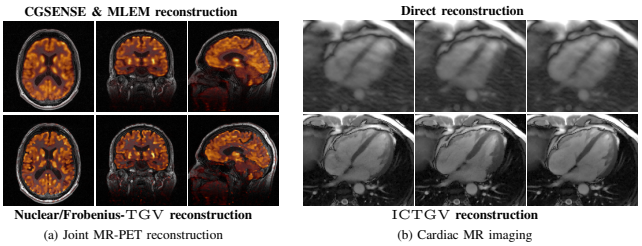


Fig. 1. Denoising with some second-order measure-based penalties.



(a) Joint MR-PET reconstruction (b) Cardiac MR imaging

Fig. 2. Application of TGV and ICTGV to biomedical imaging problems.

III. Curvature-based regularization

Besides incorporating higher-order smoothness information from the function-representation of an image, it is also possible to penalize higher-order features of their level sets. In order to obtain a convex framework, the gradient of an 2D-image can be lifted. This allows to introduce convex functionals that respect curvature information, for instance, by counting vertices or measuring the total curvature [4]:

$$\mathrm{TVX}^{\alpha,\beta}_q(u) = \inf_{\substack{\mu\in\mathcal{M}(\Omega\times S^1)\\(u,\mu)\in M_\nabla}} \alpha\|\nabla u\|_{\mathcal{M}} + \beta \sup_{\substack{\psi\in\mathcal{C}^\infty_c(\Omega\times S^1)\\ \psi\in M_q(\Omega)}} \langle \nabla_\vartheta \mu,\, \psi\rangle$$

where M_∇ corresponds to the lifting, $M_q(\Omega)$, $q\in\{0,1\}$ is a predual norm ball and ∇_ϑ realizes the tangential derivative. Employing such regularization functionals then might be advantageous for recovering thin, elongated structures, see Figure 3.

References

[1] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 492–526, 2010.
[2] K. Bredies and M. Holler, "Regularization of linear inverse problems with total generalized variation," *Journal of Inverse and Ill-posed Problems*, vol. 22, no. 6, pp. 871–913, 2014.
[3] K. Bredies, "Recovering piecewise smooth multichannel images by minimization of convex functionals with total generalized variation penalty," *Lecture Notes in Computer Science*, vol. 8293, pp. 44–77, 2014.
[4] K. Bredies, T. Pock, and B. Wirth, "Convex relaxation of a class of vertex penalizing functionals," *Journal of Mathematical Imaging and Vision*, vol. 47, no. 3, pp. 278–302, 2013.

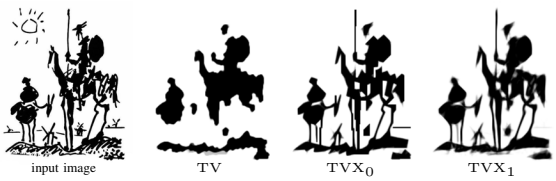


Fig. 3. Segmentation with length- vs. curvature-based regularization.

Temporal Scales: From Cellular Activity to Network Dynamics

Joana Cabral[*], Gustavo Deco[†] and Morten L. Kringelbach[†]
[*] Department of Psychiatry, University of Oxford, UK.
[†] Computational Neuroscience Group, Center of Brain and Cognition, Barcelona, Spain

***Abstract*— To explore the brain's network dynamics at the macroscale, it is useful to go beyond the microscopic activity of individual neurons and consider instead the behavior of mesoscopic ensembles of neurons when coupled together in the neuroanatomical network of white matter fibers. In whole-brain computational models, each brain area is represented by a neural-mass model with oscillatory activity and receives input from anatomically connected areas. Simulations show that their interaction at the whole-brain level gives rise to slow fluctuations (<0.1Hz) similar to the ones observed in brain activity during rest. In this work, we provide an overview of the current mechanisms linking the fast local oscillatory activity (8-100Hz) and global slow fluctuations (<0.1Hz) observed resting-state activity.**

I. Introduction

Understanding the genesis of spatially and temporally structured brain rhythms is a crucial matter in neuroscience [1]. In vitro studies have shown that the cortical tissue is excitable, displaying the emergence of coherent oscillations under specific medium conditions while the single neurons fire only intermittently [2]. Detailed computational models of spiking neurons have helped to investigate how neurons (and interneurons) connected in specific network topologies can generate firing patterns replicating electrophysiological measurements [3]. The frequency of such oscillations is determined by time constants such as the feedback delay, the synaptic time constants and the axonal transmission times. Furthermore, the ratio of time scales of excitatory and inhibitory currents and the balance between excitation and inhibition also affect the properties of the rhythms.

To investigate how these locally generated oscillations interact at the macroscopic level of the whole brain network, it is useful to use neural-mass models in order to reduce the complexity of spiking neuron models to a small set of differential equations describing the population activity [4-7]. This approach is motivated by neuroimaging observations showing that neurons within a densely connected neural ensemble tend to share the same physiological properties, exhibit dense reciprocal interconnectivity and show strong dynamical correlations. Each neural-mass receives input from anatomically connected areas. A realistic connectivity matrix can be non-invasively obtained from tractography algorithms applied to Diffusion-MRI (Figure 1) and subsequently down-sampled into a reduced number of brain areas using a parcellation template.

Over the last decade, a number of computational studies [4-8] have used whole-brain network models of coupled neural masses to investigate the relationship between the fast dynamics generated locally at the level of brain areas and the ultra-slow fluctuations of BOLD signal that appear correlated across distant brain areas forming functionally relevant resting-state networks. More recently, these BOLD signal correlations have been associated to correlated fluctuations in the power of fast oscillatory activity, as obtained with LFP, EEG or MEG.

Results show that resting-state BOLD correlations emerge spontaneously from the interaction between brain areas in the neuroanatomical network. However, the biophysical mechanisms linking the fast oscillatory activity with the slow network dynamics remain under debate [7-8].

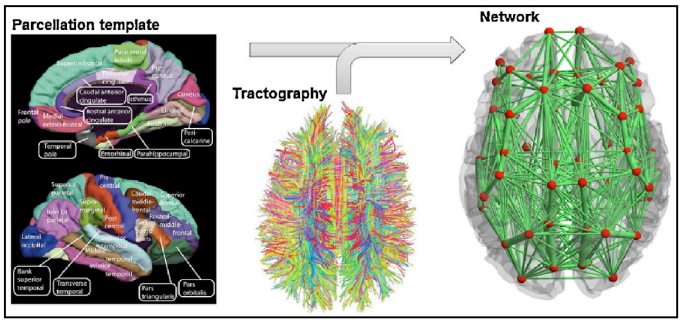


Fig. 1. Whole-brain anatomical network model obtained from DTI-based tractography. Adapted from [8}.

References

[1] Buzsáki G. "Rhythms of the brain", *Oxford University Press*, 2006

[2] Buhl EH, Tamas G, Fisahn A "Cholinergic activation and tonic excitation induce persistent gamma oscillations in mouse somatosensory cortex in vitro", *The Journal of physiology*, 513:117-126, 1998

[3] Brunel N, Wang XJ "What determines the frequency of fast network oscillations with irregular neural discharges?" *Journal of neurophysiology*, 90(1):415-430, 2003

[4] Deco G, Jirsa V, McIntosh AR, Sporns O, Kotter R "Key role of coupling, delay, and noise in resting brain fluctuations. *PNAS USA*, 106(25):10302-10307, 2009

[5] Honey CJ, Kotter R, Breakspear M, Sporns O "Network structure of cerebral cortex shapes functional connectivity on multiple time scales" *PNAS USA*, 104(24):10240-10245, 2007

[6] Cabral J, Hugues E, Sporns O, Deco G: "Role of local network oscillations in resting-state functional connectivity" *NeuroImage*, 57(1):130-139, 2011

[7] Cabral J, Kringelbach ML, Deco G, Exploring the network dynamics underlying brain activity during rest. *Progress in neurobiology* 2014, 114:102-131.

[8] Cabral J, Luckhoo H, Woolrich M, et al. "Exploring mechanisms of spontaneous functional connectivity in MEG" *NeuroImage*, 90:423-435, 2014

Wavelet-Based Segmentation Method for Spherical Images

Xiaohao Cai*, Christopher G. R. Wallis*, Jennifer Y. H. Chan* and Jason D. McEwen*
*Mullard Space Science Laboratory (MSSL), University College London, UK

***Abstract*—In this work, we review a new wavelet-based method [1] proposed to segment images on the sphere, accounting for the underlying geometry of spherical data. The method is compatible with any arbitrary type of wavelet frame defined on the sphere, such as axisymmetric wavelets, directional wavelets, curvelets, and hybrid wavelet constructions. Numerical experiments on projected spherical retina images demonstrate the superior performance of the proposed method.**

I. Introduction

Segmentation is the process of identifying object outlines within images. There are a number of efficient algorithms for segmentation in Euclidean space that depend on the variational approach and partial differential equation modelling, e.g. [2], [3]. Wavelets have been used successfully in various problems in image processing, including segmentation, inpainting, noise removal, and many others. Wavelets on the sphere have been developed to solve such problems for data defined on the sphere, which arise in numerous fields such as cosmology and geophysics, e.g. [4], [5], [6].

We review the wavelet-based spherical image segmentation method proposed in [1], which is a direct extension of the tight-frame based segmentation method [3] used to automatically identify tube-like structures such as blood vessels in medical imaging. It is compatible with any arbitrary type of wavelet frame defined on the sphere. Such an approach allows the desirable properties of wavelets to be naturally inherited in the segmentation process. Moreover, the algorithm is efficient with convergence usually within a few iterations.

The segmentation method devised in [1] provides, for the first time, a segmentation framework for spherical images. The framework used an iterative strategy with the flexibility to tailor the iterative procedure according to data types and features.

II. Algorithm

Let $f \in \mathrm{L}^2(\mathbb{S}^2)$ be the given image defined on the sphere $\mathbb{S}^2$. Without loss of generality, we assume f in [0, 1]. Let $\mathcal{A}$ and $\mathcal{A}^{-1}$ be the forward and backward spherical wavelet transforms, respectively.

The idea behind the method is to detect the candidates of possible pixels on (near) the boundary first, then gradually purify these boundary-like pixels via an iterative procedure until all pixels on the sphere are classified as inside or outside of a boundary. When a binary result is obtained the algorithm stops. In the following, we discuss each of the iterative steps of the method in more detail.

Preprocessing. Suppress the noise in f by using one iteration step of the tight-frame algorithm in [3], then represent it by $\bar{f}$.

Initialisation. Let $\Lambda^{(0)}$ be the initial set of potential boundary pixels, which is identified by using the gradient of $\bar{f}$, i.e. pixels with gradient larger than a given threshold ϵ are in $\Lambda^{(0)}$.

The i-th iteration of our algorithm can be described by: 1) find a range $[a_i,\ b_i]$ from image $f^{(i)}$ and threshold it into three parts – those below, inside and above the range (represented by $f^{(i+\frac{1}{2})}$), and obtain $\Lambda^{(i+1)}$ which contains fewer potential boundary pixels; 2) compute a new image using the following formula

$$f^{(i+1)} \equiv (\mathcal{I} - \mathcal{P}^{(i+1)})f^{(i+\frac{1}{2})} + \mathcal{P}^{(i+1)}\mathcal{A}^{-1}\mathcal{T}_\lambda(\mathcal{A}f^{(i+\frac{1}{2})}),$$

where $\mathcal{T}_\lambda$ represents the soft-thresholding with threshold λ; $\mathcal{I}$ is the identity operator and $\mathcal{P}^{(i+1)}$ is the operator generated from $\Lambda^{(i+1)}$.

Stopping criterion. As soon as all the pixels of $f^{(i+\frac{1}{2})}$ are either of value 0 or 1, or equivalently when $\Lambda^{(i)} = \emptyset$, the iteration is terminated, then all the pixels with value 1 constitute the objects of interest otherwise they are considered as background.

III. Experimental results

Experimental results of a spherical retina image are given in Fig. 1, to demonstrate the superiority of the method and show its capability of segmenting spherical images, including those with prominent directional features. The test image is generated by projecting a 2D retina image in the DRIVE data-set[1] on the sphere.

The K-means method is applied to data on the sphere according to the pixels intensities for comparison purpose, using the MATLAB built-in function `kmeans`. The result by the method [1] equipping hybrid wavelets constructed by combining the directional wavelets and curvelets is obtained with $\epsilon = 0.04$. Code to compute these wavelet transforms is public and available in the existing S2LET[2] package.

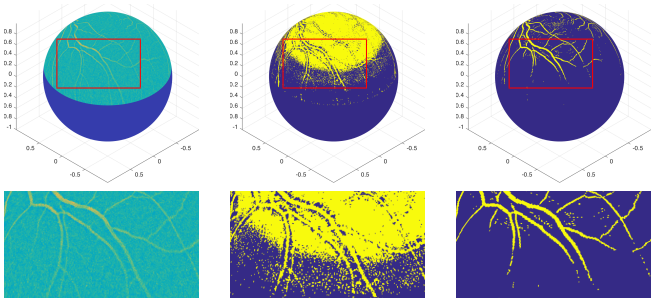


Fig. 1. Results of spherical retina image. First row from left to right gives the spherical retina image, the segmentation result of K-means method and that of the method in [1] (takes 8 iterations); with the zoomed-in details of the red rectangle areas on them shown in the second row, respectively.

References

[1] Cai, X., Chan, J., Wallis, C. and McEwen, J. D., "Wavelet-based segmentation on the sphere," *arXiv:1609.06500*, 2016.
[2] Cai, X., Chan, R. and Zeng, T., "A two-stage image segmentation method using a convex variant of the mumford-shah model and thresholding," *SIAM Journal on Imaging Sciences*, vol. 6, pp. 368–390, 2013.
[3] Cai, X., Chan, R., Morigi, S. and Sgallari, F., "Vessel segmentation in medical imaging using a tight-frame based algorithm," *SIAM Journal on Imaging Sciences*, vol. 6, pp. 464–486, 2013.
[4] Chan, J., Leistedt, B., Kitching, T. and McEwen, J. D., "Second-generation curvelets on the sphere," *IEEE Trans. Sig. Proc.*, p. to appear, 2016.
[5] Leistedt, B., McEwen, J. D., Vandergheynst, P. and Wiaux, Y., "S2LET: A code to perform fast wavelet analysis on the sphere," *Astronomy & Astrophysics*, vol. 558, pp. 1–9, 2013.
[6] McEwen, J. D., Leistedt, B., Büttner, M., Peiris, H. V., and Wiaux, Y., "Directional spin wavelets on the sphere," *arXiv:1509.06749*, 2015.

[1] http://www.isi.uu.nl/Research/Databases/DRIVE/
[2] http://www.s2let.org

Beamforming-deconvolution: A novel concept of deconvolution for ultrasound imaging

Zhouye Chen*, Adrien Besson†, Jean-Philippe Thiran†‡ and Yves Wiaux*
*Institute of Sensors, Signals and Systems, Heriot-Watt University, Edinburgh, United-Kingdom
†Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
‡Department of Radiology, University Hospital Center (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland

***Abstract*—In ultrasound (US) imaging, beamforming is usually separated from the deconvolution or some other post-processing techniques. The former processes raw data to build radio-frequency (RF) images while the latter restore high-resolution images, denoted as tissue reflectivity function (TRF), from RF images. This work is the very first trial to perform deconvolution directly with raw data, bridging the gap between beamforming and deconvolution, and thus reducing the estimation errors from two separate steps. The proposed approach retrieves both high quality RF and TRF images and exhibits better RF image quality than a classical beamforming approach.**

The deconvolution problem for ultrasound (US) imaging has been intensively considered to enhance the image quality after Jensen *et al.* [1] introduced a convolution model from the standard wave equation. According to such a model, the *radio-frequency* (RF) image, obtained from the *raw data* after the beamforming operation, can be represented as a convolution between the *point spread function* (PSF) of the US system and the *tissue reflectivity function* (TRF). The TRF image can thus be recovered from the RF image using deconvolution algorithms. The quality of the RF image, which has an impact on the recovered TRF, is linked to the beamforming technique. In classical US systems, the delay-and-sum (DAS) method is used which results in a relatively poor quality RF image.

Recently, we have expressed a linear forward model which relates the raw data to the RF image [2]. Formally, if we denote by $\boldsymbol{y} \in \mathbb{R}^M$ the raw data and by $\boldsymbol{r} \in \mathbb{R}^N$ the RF image, we have formulated a linear operator $\mathsf{G} \in \mathbb{R}^{N \times M}$ such that $\boldsymbol{y} = \mathsf{G}\boldsymbol{r} + \boldsymbol{n}$ [2].

In this study, we propose a new method, recalled as beamforming-deconvolution framework, which bridges the gap between the two techniques described above and aims at obtaining both higher quality RF and TRF images. The direct model of the beamforming-deconvolution framework is expressed as $\boldsymbol{y} = \mathsf{G}\mathsf{H}\boldsymbol{x} + \boldsymbol{n}$, where $\boldsymbol{x} \in \mathbb{R}^N$ stands for the TRF, $\mathsf{H} \in \mathbb{R}^{N \times N}$ represents the PSF and $\boldsymbol{n} \in \mathbb{R}^M$ is the additive Gaussian noise.

Instead of estimating the RF image and TRF sequentially, we hereby propose to recover the TRF and RF images altogether. With the US adapted assumption that TRF is general Gaussian Distributed, the corresponding ℓ_p - minimization ($p > 0$) problem is formulated as:

$$\min_{\boldsymbol{x} \in \mathbb{R}^N} \quad \alpha \parallel \boldsymbol{x} \parallel_p^p + \parallel \boldsymbol{y} - \mathsf{G}\mathsf{H}\boldsymbol{x} \parallel_2^2 \tag{1}$$

where α is a hyperparameter. In order to avoid the computation of the inverse of G, the forward-backward splitting (FBS) algorithm with a proximal operator of the ℓ_p-norm is adopted to solve Problem (1).

We provide a basic comparison between the proposed algorithm and a sequential method, which performs beamforming and deconvolution sequentially and separately. The method of DAS is used for beamforming and the deconvolution step was processed with FBS by minimizing $\alpha \parallel \boldsymbol{x} \parallel_p^p + \parallel \boldsymbol{r} - \mathsf{H}\boldsymbol{x} \parallel_2^2$. As a preliminary investigation, we should note that the PSF for both methods was estimated in a preprocessing step. A 128-elements linear probe, with a central frequency of 5 MHz, has been simulated with Field II, a state-of-the art ultrasound simulator. The anechoic cyst shown in figure below is composed of a 8-mm diameter anechoic occlusion at 4 cm depth embedded in a medium with high density of scatterers (30 per resolution cell) and insonified with one plane wave (PW) with normal incidence. No apodization is used neither at transmission nor at reception.

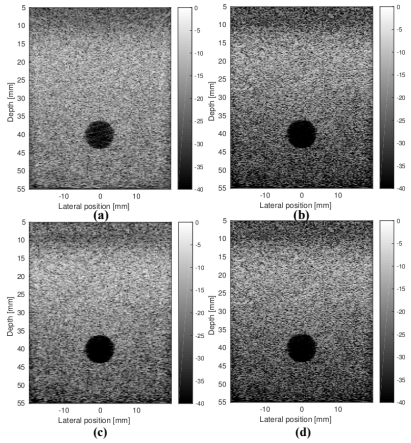


Figure 1 Comparison with a sequential method. (a) RF image with DAS (CNR=7.50 dB), (b) TRF image with sequential method (CNR=3.93 dB), (c) RF image with proposed method (CNR=7.71 dB), (d) TRF image with proposed method (CNR = 5.01 dB).

Figure 1 confirm that the proposed method is capable of recovering both high quality RF and TRF. The door from raw data to TRF is thus opened, bringing us many possibilities in the near future. On the one hand, we can perform some other post-processing techniques such as super-resolution directly to raw data. On the other hand, the compressive sampling with raw data can be introduced by including an undersampling operator and the reconstruction of enhanced US image from compressed measurements will thus become true [3, 4]. Our future work will also include the consideration of blind deconvolution techniques with variant PSFs.

REFERENCES

[1] J. A. Jensen, J. Mathorne, T. Gravesen, and B. Stage, "Deconvolution of in vivo ultrasound B-mode images," *Ultrasonic Imaging*, vol. 15, no. 2, pp. 122–133, 1993.

[2] A. Besson, R. Carrillo, M. Zhang, D. Friboulet, O. Bernard, Y. Wiaux, and J.-P. Thiran, "Sparse regularization methods in ultrafast ultrasound imaging," in *Eur. Signal Process. Conf. (EUSIPCO)*, 2016.

[3] A. Besson, R. E. Carrillo, O. Bernard, Y. Wiaux, and J.-P. Thiran, "Compressed delay-and-sum beamforming for ultrafast ultrasound imaging," in *2016 IEEE Int. Conf. Image Process*, 2016, pp. 2509–2513.

[4] Z. Chen, A. Basarab, and D. Kouamé, "Compressive deconvolution in medical ultrasound imaging," *IEEE Trans. Med. Imaging*, vol. 35, no. 3, pp. 728–737, 2016.

Rapid Motion-Robust MRI

Joseph Y. Cheng
School of Radiology, Stanford University, Stanford, California, USA.

***Abstract*—The lengthy scan durations required for magnetic resonance imaging (MRI) limit its effectiveness for characterizing rapid physiological dynamics and hinder its reliability due to sensitivity to image artifacts from motion. However, the scan duration must not only be shortened, but the sensitivity of the modality to motion must also be addressed. Speed and motion-robustness must be jointly considered for more effective solutions.**

I. Introduction

MRI is a powerful imaging modality with the flexibility to generate different soft-tissue contrasts. Unfortunately, MRI data acquisition is inherently slow: data are acquired point-by-point in the frequency domain. Moreover, signal decay requires data to be acquired over many time segments. The quality and reliability of MRI are hindered by lengthy scan durations (0.5–5 min for volumetric acquisitions). In order to truly design MRI to be both (a) rapid and (b) motion-robust, these two features must be jointly considered as they are synergistic in nature. A more rapid MRI scan will decrease the amount of motion corruption. Conversely, improving motion robustness will improve the performance of model-based reconstruction techniques that enable significant sub-sampling. The purpose of this paper is to discuss four areas of recent developments that are building blocks to achieve rapid and motion-robust MRI.

II. Data acquisition

The first component in rapid motion-robust MRI scans is the data acquisition process. Advanced techniques using model-based reconstruction, such as compressed-sensing, rely on pseudo-random variable-density sub-sampling to structure aliasing artifacts to appear noise-like [1]. To enable both advanced reconstruction techniques and motion robustness, an appealing strategy is to extend the compressed-sensing framework to include time. Popular approaches use the golden-ratio ordering [2] and its variants to pseudo-randomly sample data in the spatial-frequency-domain and in time.

III. Motion-artifact suppression

High sub-sampling factors have been enabled by model-based reconstruction techniques that exploit the localized sensitivity-profiles of each element in a coil-receiver array [3], [4]. Additionally, image priors such as sparsity in another transform domain can be incorporated using compressed sensing [5]. By reducing the amount of sub-sampling, these same algorithms can be leveraged to recover corrupted data samples. This framework can be relaxed through weighting (or "soft-gating") the data based on data corruption rather than binary data rejection [6]. These approaches suppress image artifacts from motion but require enough additional "motion-free" samples to be acquired.

IV. Motion correction

With significant sub-sampling, the few data points acquired must be sufficiently accurate to enable proper estimation of the missing or overly-corrupt data samples. By correcting for motion-corruption, more data samples can be used by the reconstruction methods [7], [8]. Advanced algorithms for modelling non-rigid image warping from motion can be used. Alternatively, complex motions can be approximated with simpler localized linear translations. This approximation simplifies the correction algorithm and reduces the required computation.

V. Motion-resolved reconstruction

Motion correction increases the number of accurate data samples used by the model-based reconstruction. Unfortunately, patient motion not only results in misaligned data samples, but will also result in other types of data corruption (e.g., field inhomogeneity variations and signal magnitude fluctuations). It has been recently proposed to extend the reconstruction framework to include motion as an additional dimension [9]. With the latest developments in image reconstruction, multiple motion dimensions, such as respiratory and temporal dynamics, can be included while maintaining feasible scan durations (5–10 min). This ultra-high-dimensional imaging approach increases robustness to motion and enables more clinically relevant information to be extracted from a single dataset.

VI. Conclusion

Developments in data acquisition and model-based reconstruction have significantly improved MRI. The dependency of the discussed building blocks (i.e., data acquisition enables effective motion suppression) reflects the synergistic nature of speed and motion-robustness. Future work should entail focusing on these different aspects simultaneously for greater advancements.

References

[1] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging." *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, dec 2007.

[2] S. Winkelmann, T. Schaeffter, T. Koehler, H. Eggers, and O. Doessel, "An optimal radial profile order based on the Golden Ratio for time-resolved MRI." *IEEE Trans Med Imaging*, vol. 26, no. 1, pp. 68–76, jan 2007.

[3] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, jun 2002.

[4] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: sensitivity encoding for fast MRI." *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, nov 1999.

[5] R. Otazo, D. Kim, L. Axel, and D. K. Sodickson, "Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI," *Magnetic Resonance in Medicine*, vol. 64, no. 3, pp. 767–776, 2010.

[6] K. M. Johnson, W. F. Block, S. B. Reeder, and A. Samsonov, "Improved least squares MR image reconstruction using estimates of k-space data consistency," *Magnetic Resonance in Medicine*, vol. 67, no. 6, pp. 1600–1608, jun 2012.

[7] J. Y. Cheng, T. Zhang, N. Ruangwattanapaisarn, M. T. Alley, M. Uecker, J. M. Pauly, M. Lustig, and S. S. Vasanawala, "Free-breathing pediatric MRI with nonrigid motion correction and acceleration," *Journal of Magnetic Resonance Imaging*, vol. 42, no. 2, pp. 407–420, aug 2015.

[8] X. Chen, M. Salerno, Y. Yang, and F. H. Epstein, "Motion-compensated compressed sensing for dynamic contrast-enhanced MRI using regional spatiotemporal sparsity and region tracking: Block low-rank sparsity with motion-guidance (BLOSM)," *Magnetic Resonance in Medicine*, vol. 72, no. 4, pp. 1028–1038, 2014.

[9] L. Feng, L. Axel, H. Chandarana, K. T. Block, D. K. Sodickson, and R. Otazo, "XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing," *Magnetic Resonance in Medicine*, vol. 75, no. 2, pp. 775–788, feb 2016.

Recovering Brain Network Structure from Highly Under-Sampled FMRI using Electrophysiological Constraints

Mark Chiew*, Jostein Holmgren†, Dean Fido*, Catherine E. Warnaby* and Karla L. Miller*
* FMRIB Centre, University of Oxford, Oxford, United Kingdom
† Institute of Psychology, University of Oslo, Oslo, Norway

***Abstract*—Functional MRI (FMRI) is a non-invasive imaging technology that is sensitive to brain activity, although encoding an entire brain at high resolution can take seconds. Recently, we developed a novel approach to characterising brain function by using the existence of low-rank brain network structure to constrain highly under-sampled reconstructions. We improve upon this work by incorporating constraints derived from simultaneous electroencephalography (EEG) measurements. We demonstrate for the first time that EEG-derived constraints can improve the functional sensitivity of rank-constrained FMRI data reconstruction.**

In functional magnetic resonance imaging (FMRI), brain activity is characterised through fluctuating blood oxygen levels, which generate local magnetic field susceptibility differences that affect MR signal contrast. These blood oxygenation changes have a well established link to neuronal activity and metabolic demand, so while the FMRI measurements are indirect, they are robust indicators of underlying brain function. Conventionally, FMRI data are acquired across the whole brain in seconds, which can limit achievable spatial and temporal resolution, and restrict data degrees of freedom.

Recently, we developed a new strategy for accelerating FMRI data acquisition by leveraging information about how the brain is intrinsically organised[1], [2]. Specifically, we exploited the correspondence between robust low-dimensional brain network models that are ubiquitous in brain functional analysis[3], and low-rank matrix completion or recovery approaches[4].

However, relying primarily on the low-rank constraint can result in a loss of fidelty for functionally relevant, but low-variance brain networks. Here, we propose a novel enhancement to rank-constrained FMRI using simultaneous recording of electroencephalography (EEG) data. The temporal fidelity and functional specificity of EEG provide an excellent opportunity for constraining the estimated temporal subspace of the FMRI data. To our knowledge, this is the first demonstration of using EEG to aid the reconstruction of FMRI data, rather than simply forming *post hoc* associations.

To do this, we solve the following problem:

$$\begin{aligned} \text{minimize} \quad & ||\Phi(UV^*) - y||_2 \\ \text{such that} \quad & \text{rank}(UV^*) \leq r \\ \text{and} \quad & \forall\ v \in W, v \in V \end{aligned} \tag{1}$$

where U are spatial components, V are temporal components, Φ is a linear operator containing MRI gradient and receive coil sensitivity encodings, and y are the under-sampled measurements. Additionally, W contains vectors derived from the EEG measures, which constrain the recovery of temporal characteristics by its explicit inclusion in the subspace defined by V.

Simultaneous EEG and FMRI were acquired using a 3 T MRI system (Siemens Healthcare), and a 32-channel MRI-compatible EEG system (BrainProducts). A conventional multi-slice 2D echo-planar imaging acquisition was used to acquires a 4-slice, 2x2x2 mm^3 at 200 ms temporal resolution data set, over a 3 minute duration in a single subject. The EEG data were corrected for gradient and ballistocardiographic artefacts, followed by extraction of the theta-band (4-7 Hz) power envelope using the short-time Fourier Transform. The first four principal components across all channels were used as the electrophysiological constraint, which were then transformed into haemodynamic space using a haemodynamic response function.

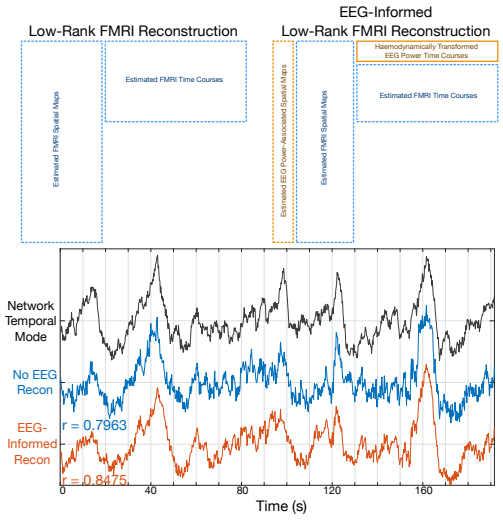


Fig. 1. Figure 1 – The role of EEG information in the rank-constrained FMRI reconstruction process. In the recovered time-courses, the EEG-informed reconstruction improves correlation with the true network temporal mode.

We demonstrate, preliminarily, that using an externally derived EEG measurement can improve the fidelity of under-sampled FMRI signal recovery. We retrospectively under-sampled the FMRI data to 10% (R=10), using a rank constraint of $r = 16$. We show that including four EEG components in the FMRI temporal subspace can improve the recovery of the temporal mode associated with the strongest resting state brain network, improving canonical correlation from 0.796 to 0.848. This work introduces a potentially exciting avenue for multi-modality brain imaging, through unique integration of the EEG and FMRI measures to exploit their mutual information.

References

[1] M. Chiew, S. M. Smith, P. J. Koopmans, N. N. Graedel, T. Blumensath, and K. L. Miller, "k-t FASTER: Acceleration of functional MRI data acquisition using low rank constraints." *Magn Reson Med*, vol. 74, no. 2, pp. 353–364, Aug. 2015.

[2] M. Chiew, N. N. Graedel, J. A. Mcnab, S. M. Smith, and K. L. Miller, "Accelerating functional MRI using fixed-rank approximations and radial-cartesian sampling." *Magn Reson Med*, p. (Early View), Jan. 2016.

[3] C. F. Beckmann and S. M. Smith, "Probabilistic independent component analysis for functional magnetic resonance imaging." *IEEE Trans Med Imaging*, vol. 23, no. 2, pp. 137–152, Feb. 2004.

[4] E. J. Candes and B. Recht, "Exact Matrix Completion via Convex Optimization," *Found Comput Math*, vol. 9, no. 6, pp. 717–772, Apr. 2009.

CT imaging from sparsely sampled data

Seungryong Cho*, Jongduk Baek†
* Department of Nuclear and Quantum Engineering, Korea Institute of Advanced Science and Technology, Daejeon, Korea.
† School of Integrated Technology, Yonsei University, Incheon, Korea

***Abstract*— Sparse sampling method is one viable option to low-dose CT, and has been actively investigated in terms of both reconstruction algorithm and physical realization of the sampling. Sparse-view sampling is a straightforward way and is considered a feasible solution particularly to various cone-beam CT applications. Recently, we have proposed and developed another type of sparse sampling scheme, which is called many-view under-sampling (MVUS). In the MVUS scheme, the x-ray beam is partially blocked by multiple radio-opaque strips thereby reducing the radiation dose to the patient. In this abstract, we summarize our work that demonstrate its feasibility in a cone-beam CT setup and also report preliminary results acquired from a fast-gantry type diagnostic CT system. For image reconstruction, we used a modified total-variation minimization (TV) algorithm that masks the blocked data in the back-projection step leaving only the measured data through the slits to be used in the computation.**

I. Introduction

THERE are increasing needs for and related researches of low-dose CT for various clinical applications. Recently, sparse sampling approaches have been also proposed in conjunction with the compressive-sensing (CS)-inspired reconstruction theories [1]. Sparse-view sampling has been particularly exploited for cone-beam CT applications. In a diagnostic CT system which is based on a fast gantry rotation, however, sparse-view sampling would be technically difficult if not impossible. As an alternative approach of sparse-view sampling, we have proposed a sparse sampling technique called many-view under-sampling (MVUS) and have experimentally demonstrated its feasibility in CBCT applications [2, 3]. In the MVUS approach, the x-ray cone-beam is partially blocked by multiple radio-opaque strips that are placed between the x-ray source and the patient.

II. Methods

A. Systems

To experimentally implement the MVUS scanning, we used an object-rotating cone-beam CT system (EBSCAN #DCT, EBTECH, Daejeon, Republic of Korea) and also used a diagnostic CT system (Bodytom, Neurologica, USA). A reciprocating beam-blocker was used for CBCT and a rotating disk-type beam-blocker was used for the diagnostic CT considering their mechanical durability as well as data sampling efficiency.

B. Reconstruction algorithm

$$\hat{f} = \arg\min \left\| \vec{f}_j \right\|_{TV} \qquad (1)$$

$$\text{, such that } \left\| THf_j - g \right\| < \varepsilon$$

, where f_j represents an image under jth iteration, and $\hat{f}$ the minimum image total-variation solution. $\| \ \|_{TV}$ represents the total-variation of an image in 3-dimension. The system matrix H was based on a ray-driven model with T representing the masking operation according to the available measured data g through the slits, and ε was determined empirically considering the data distance convergence in the POCS only iterations.

III. Results

Optimally chosen set of parameters for the beam-blocker motion with the aid of the reconstruction algorithm resulted in an outperforming image quality as shown in Fig. 1 in CBCT.

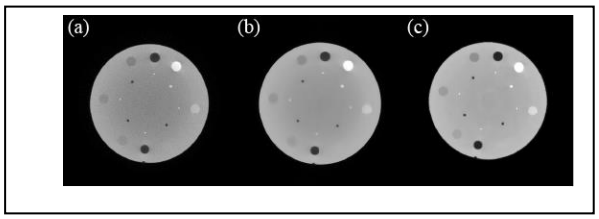


Fig. 1. Image reconstructed by (a) low mA scan + FBP, (b) low mA + TV, and (c) MVUS + TV are shown, respectively. The display window is [-1000, 100] HU.

We have experimentally demonstrated the feasibility of MVUS scanning for low-dose CBCT, and investigated various scanning configurations to seek an optimum condition. With further studies, the MVUS is believed to contribute to low-dose CT imaging in combination with low-mA scanning method.

References

[1] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, pp. 4777–807, Sep. 2008.

[2] S. Abbas, J. Min, and S. Cho, "Super-sparsely view-sampled cone-beam CT by incorporating prior data," *J. Xray. Sci. Technol.*, vol. 21, no. 1, pp. 71–83, Jan. 2013.

[3] T. Lee, C. Lee, J. Baek, and S. Cho, "Moving beam-blocker-based low-dose cone-beam CT," *IEEE Trans. Nucl. Sci.*, (in press) 2016.

Automatic Motion Signal Extraction for Fully Self-Gated 5D Whole-Heart Imaging: Preliminary Results

Lorenzo Di Sopra[*], Davide Piccini[†‡], Matthias Stuber[*†], Jessica Bastiaansen[*] and Jérôme Yerly[*†]
[*] Department of Radiology, University Hospital (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland
[†] Center for Biomedical Imaging (CIBM), Lausanne, Switzerland
[‡] Advanced Clinical Imaging Technology, Siemens Healthcare, Lausanne, Switzerland

I. Introduction

Recent advances in cardiac MR imaging, which make use of modern acquisition [1] and reconstruction schemes [2,3,4], have the potential to challenge existing paradigms by enabling simultaneous functional and anatomical 3D assessment of the whole-heart in one scan and during free-breathing. However, current solutions still present some important limitations regarding cardiac and respiratory gating. In particular, cardiac gating usually requires the use of external ECG devices [2], which need setup time, cause patient discomfort and are susceptible to magnetic perturbation. Here we propose a fully self-gated strategy to 5D reconstruction, where readouts are automatically and efficiently sorted according to the respiratory and cardiac motion.

II. Methods

This preliminary study was performed on a 1.5T clinical MRI scanner (MAGNETOM Aera, Siemens Healthcare) using a prototype free-running non-ECG-triggered 3D golden angle radial bSSFP sequence in N=3 healthy volunteers. Continuous acquisition using a novel 3D radial sampling pattern and a 1-1 180 binomial water excitation radiofrequency (RF) pulse for fat suppression allowed to minimize eddy current effects and to acquire in fully preserved steady-state magnetization. A fully automated algorithm extracted the physiological cardiac and respiratory motion signals by analyzing the modulation of the central k-space coefficient of all radial readouts (Fig 1a). Subject-dependent bandpass filters (automatically centered on the subject's specific motion frequency) were applied to the frequency spectra to isolate respiratory and cardiac self-gating signals (Fig 1b). The coils yielding the strongest motion information were automatically selected (Fig 1a). Self-gated cardiac triggers were obtained by detecting the peaks on the filtered and selected signals, and then compared to the ECG triggers. Correlation between the two sets of cardiac cycle duration was investigated with linear regression analysis and Bland-Altman plot (Fig 2). Extracted signals were employed to sort the acquired data into cardiac and respiratory motion-resolved bins. Readouts were first sorted according to the respiratory motion amplitude (6 different states) and then to the cardiac phase (50 ms window width). The resulting 5D (x-y-z-cardiac-respiratory dimensions) undersampled datasets were reconstructed using a k-t sparse SENSE algorithm [2], which exploited sparsity along both cardiac and respiratory dimensions.

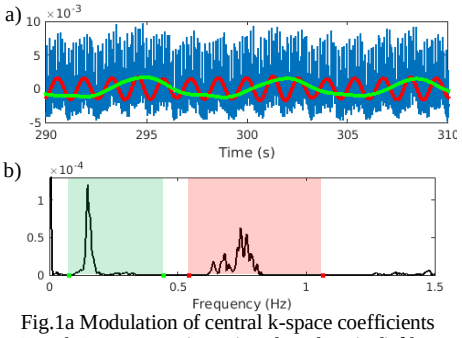


Fig.1a Modulation of central k-space coefficients
Fig.1b Respiratory (green) and cardiac (red) filters

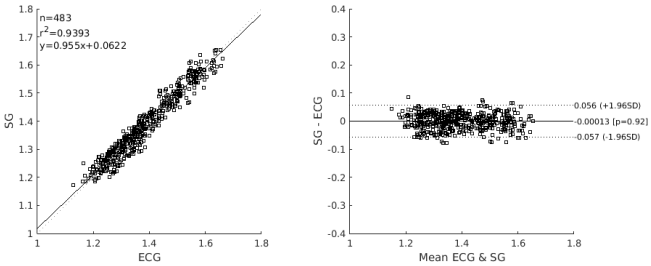
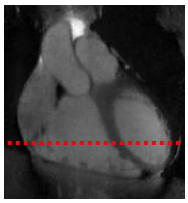


Fig.2 Linear Regression and Bland-Altman plot of SG vs ECG RR-intervals.



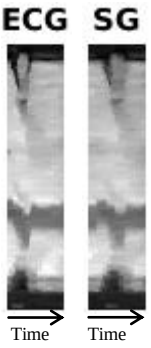


Fig.3 Temporal evolution of a transversal section profile across the ventricles.

III. Results

Both respiratory and cardiac motion signals were automatically and successfully extracted in all volunteers. The self-gated cardiac triggers were relatively accurate with respect to the reference ECG triggers: in particular, self-gated cardiac cycle duration deviated from the reference ECG RR-interval by 30.7±12.2 ms. Visual comparison between the fully self-gated and ECG-gated 5D reconstructions, as well as between temporal motion evolutions of ventricular section profiles (Fig 3), did not exhibit any significant difference.

IV. Conclusion

The proposed framework enables fully automated self-gated cardiac and respiratory motion resolved imaging of the whole-heart with isotropic spatial resolution and minimal operator-interaction. These preliminary results show promising correlation between self-gated and ECG-gated triggers. Further quantitative validation is necessary.

References

[1] Coppo S, ISMRM 2015; 28.
[2] Feng L, ISMRM 2015; 27.
[3] Lustig M, MRM 2007; 58:1182-1195.
[4] Pang J, MRM 2014; 72: 1208-1217

Adaptive-BLIP for Magnetic Resonance Fingerprinting

Roberto Duarte*, Zhouye Chen*, Mohammad Golbabaee†, Zaid Mahbub†, Ian Marshall†, Mike Davies†, Yves Wiaux*
* Institute of Signals, Sensors and Systems, Heriot-Watt University, EH14 4AS, UK
† Institute for Digital Communications (IDCom), The University of Edinburgh, EH9 3JL, UK

***Abstract*—We present an improved version of the BLoch response recovery via Iterative Projection (BLIP) algorithm for Magnetic Resonance Fingerprinting (MRF), that drastically reduces the computation time using an adaptive dictionary. At each iteration, the BLIP dictionary is updated through a clustering technique in the quantitative parameter space based on the fingerprint distribution across all voxels. Similar to a random tree, new parameter sets are selected around these clusters, making it possible to obtain a higher resolution than the original dictionary. Without loss of accuracy in reconstruction, simulations with a numerical phantom demonstrated that the computation time and the required memory to store the dictionary is significantly reduced in comparison to a dictionary with finer but fixed resolution.**

I. Introduction

Inspired by the recent growth of Compressed Sensing (CS) techniques in MRI, the Magnetic Resonance Fingerprinting (MRF) was introduced to accelerate the quantitative imaging [1]. However, the exact link to CS was not made explicit. More recently, a full CS strategy was formulated in [2] including a random pulse excitation sequence following the MRF technique, a random Echo-Planar Imaging subsampling strategy, and an iterative projection algorithm that imposes consistency with the Block equations, namely BLoch response recovery via Iterative Projection (BLIP). The algorithm is given by

$$X^{(n+1)} = \mathcal{P}_{(R_+\mathcal{B})^N}\left[X^{(n)} + \mu h^H\left(Y - h\left(X^{(n)}\right)\right)\right], \quad (1)$$

where $X \in \mathbb{C}^{N\times L}$ represents the magnetization response of the image with N voxels, $Y \in \mathbb{C}^{M\times L}$ corresponds to the measurements, L is the excitation sequence length, h is an operator that describes the undersampling in k-space, n stands for the recursion index, and μ is a stepsize, which is selected adaptively. $\mathcal{P}_{(R_+\mathcal{B})^N}$ is the voxelwise projection on to signal model $(R_+\mathcal{B})^N$ approximated by the dictionary D. The projection for the voxel i can be computed as $\hat{k}_i = \arg\max_k \text{real}\langle D_k, X_{i,:}/\rangle||D_k||_2$, where $X_{i,:}$ is the magnetization sequence of the voxel i, $D = [D_1, ..., D_d] \in \mathbb{C}^{L\times d}$, d is the size of the dictionary. It has been shown that BLIP outperforms the MRF technique proposed in [1] especially with a shorter magnetization sequence. Nevertheless, the computation time increases linearly with the size of the dictionary which needs to be big for high quality reconstructions, becoming a trade off between speed and accuracy.

II. Adaptive-BLIP

In order to address this problem, we propose to project onto an adaptive dictionary that is updated in each iteration, namely Adaptive-BLIP. The number of tissues to be imaged in MRI is usually small compared to the number of voxels in the image, we use this as prior to update the dictionary. To begin with, a coarse dictionary is first defined using a fixed grid. After the projection, quantitative parameters θ_c are clustered by K-means based on the fingerprint distribution across all voxels. The number of clusters n_c can be defined proportional to the number of expected tissues in the volume to be imaged. For each computed cluster, n_r new parameter sets are chosen randomly using a Gaussian distribution with standard deviation σ_{T_1}, σ_{T_2} and σ_{off} defined according to the T_1, T_2 and off-resonance intervals. All of the these parameter sets help to generate the new adaptive dictionary using the Bloch equation. The Bloch equation manifold is thus explored in a similar way to a random tree, allowing the algorithm to have a better resolution than the original dictionary, and resulting in a much smaller dictionary that is updated in each iteration.

III. Simulations

We provide a comparison between the proposed method and BLIP on the same numerical phantom used in [2]. BLIP is tested with two different size dictionaries and the parameters for Adaptive-BLIP is set as $n_c = 10$ and $n_r = 10$, resulting in $d = 110$, the maximum number of iterations is set to 20. Two experiments are given in Figure 1 and 2. They evaluate the performance of the algorithms in terms of the sequence length and input SNR respectively.

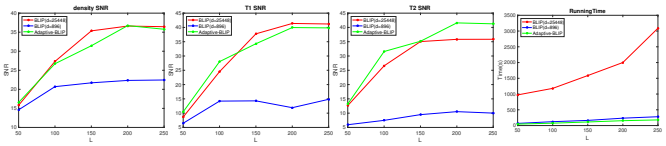


Fig. 1: Reconstruction performance as a function of L

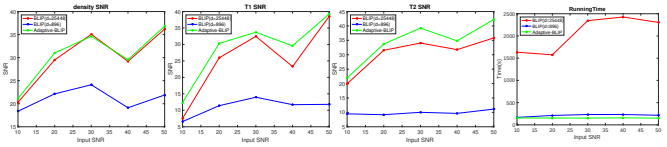
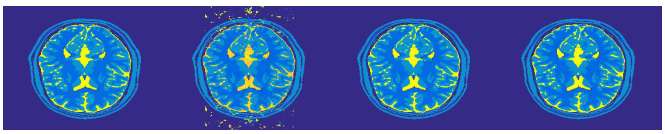


Fig. 2: Reconstruction performance as a function of the input SNR

In order to get a visual indication of the performance of algorithms, we provide also images of T_1 map for $L = 200$ with input SNR of $40dB$ in Figure 3. We may remark from the figures that the Adaptive-BLIP can achieve high quality reconstruction with significantly less processing time. Our future work will include more simulations and real data reconstructions.



(a) Original (b) $d = 896$ (c) $d = 25448$ (d) A-BLIP

Fig. 3: A visual comparison of the T_1 map estimates

References

[1] D. Ma, V. Gulani, N. Seibelich, K. Liu, J. L. Duerk, and M. A. Griswold, "Magnetic resonance fingerprinting," *Nature*, vol. 465, pp. 187–192, 2013. [Online]. Available: http://dx.doi.org/10.1038/nature11971

[2] M. Davies, G. Puy, P. Vandergheynst, and Y. Wiaux, "A compressed sensing framework for magnetic resonance fingerprinting," *Siam journal on imaging sciences*, vol. 7, no. 4, p. 2623, 10 2014.

Parametric Models of Phase-Amplitude Coupling in Neural Time Series

Tom Dupré la Tour, Yves Grenier, Alexandre Gramfort
Télécom ParisTech, LTCI, CNRS, Université Paris-Saclay, Paris, France

***Abstract*—In neuroscience, phase-amplitude coupling (PAC) refers to the interaction between the phase of a slow neural oscillation and the amplitude of high frequencies within the same signal or at a distinct brain location. To model PAC, we use new parametric generative driven auto-regressive (DAR) models. These statistical models provide a non-linear and non-stationary spectral estimation of the signal, and are able to capture the time-varying behavior of PAC. We also show that they are more robust to short signals than two state-of-the-art empirical PAC metrics.**

I. Introduction

Auto-regressive (AR) models are stochastic signal models that have proved their usefulness in many applications. One of their advantages is the existence of fast inference algorithms [1] and to provide a compact representation of the spectral content of a signal. Standard AR models are so-called stationary, meaning that the statistics of the signal are assumed to be stable over time. When working with such models, the spectrum is therefore not a function of time. In many applications, this modeling assumption is not adapted to describe the interesting dynamics of the physical system observed. This is for example the case in the field of econometrics where time-varying or non-linear AR models were first studied [2], but it is also the case for physiological signals as it will be illustrated below with a phenomena known as phase-amplitude coupling (PAC) [3].

II. Driven auto-regressive models

An AR model specifies that a signal y depends linearly on its own p past values, where p is the *order* of the model:

$$y(t) + \sum_{i=1}^{p} a_i y(t-i) = \varepsilon(t) \tag{1}$$

where ε is the *innovation*, modeled with a Gaussian white noise: $\varepsilon(t) \sim \mathcal{N}(0, \sigma(t)^2)$. To extend this AR model to a non-stationary model, one can assume [4] that the AR coefficients a_i and the innovation variance σ^2 are driven by a polynomial function of a given exogenous signal x, here called the *driver*:

$$a_i(t) = \sum_{j=0}^{m} a_{ij} x(t)^j, \qquad \log(\sigma(t)) = \sum_{j=0}^{m} b_j x(t)^j \tag{2}$$

We call this model a driven auto-regressive (DAR) model.

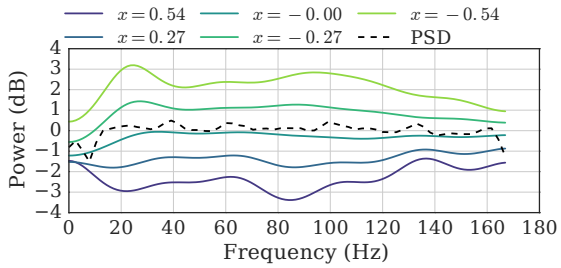


Fig. 1. Power spectral density (PSD) evaluated through a DAR model, for different driver's values x.

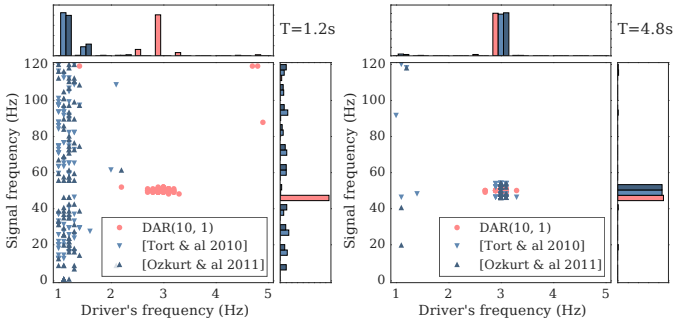


Fig. 2. Frequencies of the maximum PAC value, for three methods: a DAR model with $(p, m) = (10, 1)$ and two metrics from [5] and [6]. The simulated signals last $T = 1.2$ (left) and 4.8 (right) seconds, at a sampling frequency of 240 Hz. The signals are simulated with a PAC at 3 Hz and 50 Hz. The DAR models correctly estimate these frequencies even with a short signal length, while the two other metrics fail.

III. Phase-Amplitude Coupling

In neuroscience, phase-amplitude coupling (PAC) refers to the interaction between the phase of a slow neural oscillation and the amplitude of high frequencies. To give a proper model to PAC, we applied DAR models on a human electro-corticogram (ECoG) channel from [3], using a band-pass filter to extract the driver x from the signal y. From the estimated DAR model, we computed the power spectral density (PSD) conditionally to the driver's values x, as presented in Fig. 1. PAC can be identified in the difference of the PSD as the driver x varies: the PSD has more power for negative driver's values than for positive driver's values, and PSD shapes are also different.

We also simulated 100 signals with a PAC between $f_x = 3$ Hz and $f_y = 50$ Hz, estimated DAR models on them, and selected the frequency with the maximal PSD modulation. The results, presented in Fig. 2, show that DAR models are more robust than two state-of-the-art PAC metrics [5], [6] when the signals length is short.

References

[1] J. Durbin, "The fitting of time-series models," *Review of the Int. statistical institute*, pp. 233–244, 1960.
[2] H. Tong and K. S. Lim, "Threshold autoregression, limit cycles and cyclical data," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 245–292, 1980.
[3] R. T. Canolty, E. Edwards, S. S. Dalal, M. Soltani, S. S. Nagarajan, H. E. Kirsch, M. S. Berger, N. M. Barbaro, and R. T. Knight, "High gamma power is phase-locked to theta oscillations in human neocortex," *Science*, vol. 313, no. 5793, pp. 1626–1628, 2006.
[4] Y. Grenier, "Estimating an AR Model with Exogenous Driver," Telecom ParisTech, Tech. Rep. 2013D007, 2013.
[5] A. B. Tort, R. Komorowski, H. Eichenbaum, and N. Kopell, "Measuring phase-amplitude coupling between neuronal oscillations of different frequencies," *J. Neurophysiol.*, vol. 104, no. 2, pp. 1195–1210, 2010.
[6] T. E. Özkurt and A. Schnitzler, "A critical note on the definition of phase–amplitude cross-frequency coupling," *Journal of Neuroscience methods*, vol. 201, no. 2, pp. 438–443, 2011.

Proton-constrained iterative reconstructions of ^{17}O-MRI images for $CMRO_2$ quantification

Dmitry Kurzhunov[*†], Robert Borowiak[*†‡], Marco Reisert[*†] and Michael Bock[*†]
[*] Medical Physics, Department of Radiology, Medical Center – University of Freiburg, Germany
[†] Faculty of Medicine, University of Freiburg, Germany
[‡] German Cancer Consortium (DKTK), Heidelberg, Germany

Abstract **This study presents a comparison analysis of different reconstruction techniques for ^{17}O MR images and quantification of cerebral metabolic rate of oxygen consumption ($CMRO_2$). ^{17}O MR 3D data sets of a healthy volunteer's brain were acquired at a clinical 3 Tesla MR system with inhalation of $^{17}O_2$ gas. Conventional Kaiser-Bessel (KB) reconstruction was compared with total variation (TV) constrained reconstruction and with our recently proposed method of iterative reconstruction of ^{17}O MR images with ^{1}H constraint. Here, anisotropic diffusion (AD) of coregistered ^{1}H image was used for the penalty term and this ^{1}H MR image of high spatial resolution acted as edge-preserving constraint. AD constraint reconstruction showed higher SNR and improved quality of ^{17}O MR images. It is essential for localized $CMRO_2$ quantification in patients with heterogeneous Glioblastoma brain tumors, which is hardly possible with conventional reconstruction techniques.**

I. Introduction

Abnormalities in brain oxygen metabolism are found in tumors, cerebrovascular and neurodegenerative diseases. A useful biomarker of metabolic brain activity is the cerebral metabolic rate of oxygen consumption ($CMRO_2$). $CMRO_2$ can be quantified with ^{15}O-PET [1] or direct ^{17}O-MRI [2-5] by fitting a pharmacokinetic model to the signal dynamics seen during and after the administration of ^{17}O-enriched gas. So far, ^{17}O-MRI was applied mainly at ultra-high fields ($B_0 \geq 7T$) to overcome the low SNR at clinical field strengths ($B_0 \leq 3T$). Recently, we showed that ^{17}O-MRI is feasible at clinical field strengths [3] and the method of profile likelihood analyses showed that $CMRO_2$ can be reliably quantified [4].

The aim of this study was to compare conventional reconstruction techniques with iterative ^{1}H constraint reconstruction of ^{17}O MR images.

II. Theory

In iterative reconstruction the objective function is minimized:

$$J(\boldsymbol{x}) = \|\boldsymbol{A}\cdot\boldsymbol{x}-\boldsymbol{y}\|_2^{\,2} + \lambda\cdot\boldsymbol{R}, \qquad (1)$$

where **A** denotes the system matrix that maps the image **x** to the corresponding raw data $\boldsymbol{y}$, λ is the weighting factor of the regularization term $\boldsymbol{R}$. In our recently proposed AD reconstruction [5] regularization term contains a gradient operator $\boldsymbol{g}$ which is applied to ^{1}H MPRAGE image:

$$\boldsymbol{R}_{\mathrm{D}} = \int \boldsymbol{x}\nabla(\boldsymbol{D}\nabla\boldsymbol{x}) \qquad (2)$$

$$\boldsymbol{D} = \left(1 - \frac{\boldsymbol{g}\cdot\boldsymbol{g}^T}{|\boldsymbol{g}|^2}\Big/\sqrt{1+\boldsymbol{g}^2/\alpha^2}\right) \qquad (3)$$

III. Results and Discussion

First, realistic ^{17}O MRI phantom [5] was constructed from segmented ^{1}H MPRAGE data set with tissue-specific $H_2{}^{17}O$ concentrations, experimentally measured T_2^* and SNR values of ^{17}O images and used 3D radial acquisition. Structural similarity analysis [6] of reconstructed ^{17}O phantom images showed that AD constraint (2) reconstruction has higher precision than KB gridding and TV constraint reconstruction.

Secondly, 45 dynamic 3D ^{17}O data sets were acquired in human brain in ^{17}O MR experiment with inhalation of ^{17}O-enriched gas with 1 min temporal resolution. Figure 1 shows comparison of single ^{17}O-MR image (TA = 1 min) reconstructed with different methods. AD constraint reconstruction (d) has higher SNR compered to KB gridding (a) and TV constraint reconstruction (c). Anatomical structures (e.g., ventricles) are better seen in (d) compared to KB with Hann filter. Quantified $CMRO_2$ values of 0.67-0.83/0.86-1.07 $\mu mol/g_{tissue}/min$ in WM/GM regions are in a good agreement with results of ^{15}O-PET studies [1].

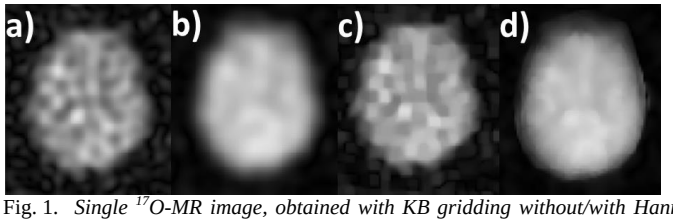


Fig. 1. *Single ^{17}O-MR image, obtained with KB gridding without/with Hann filter (a/b) and in iterative constrcution with TV (c) and AD constraint (d).*

TV constraint reconstruction did not give any significant improvement of SNR. Averaging among the neighboring pixels was beneficial: AD smoothes data among pixels with similar intensity (mostly within one brain tissue component) and preserves borders within different tissue compartments (edge-preservation), which is favorable compared to isotropic homogeneous Hanning filtering. Increased SNR of ^{17}O MR images is needed for localized $CMRO_2$ quantification and calculation of $CMRO_2$ maps. It is of high importance for investigation of oxygen metabolism of heterogeneous Glioblastoma tumor regions.

References

[1] Leenders KL et al. Brain 1990;113:27–47.
[2] Atkinson IC et al. Neuroimage 2010;51:723–733.
[3] Borowiak R et al. MAGMA 2014;27:95–99.
[4] Kurzhunov D et al. MRM 2017; doi 10.1002/mrm.26476.
[5] Kurzhunov D et al. ISMRM 2015 Toronto, p.2450.
[6] Wang Z et al. IEEE Trans. Image Process.2004;13:600-61

A Subspace Approach to Ultrahigh-Resolution MR Spectroscopic Imaging

Fan Lam* and Zhi-Pei Liang*
* Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801.

***Abstract*—Research and clinical applications of MR spectroscopic imaging (MRSI) have been limited by several fundamental technical hurdles, including low signal-to-noise ratio, limited spatial resolution, slow imaging speed, and overwhelming nuisance water/lipid signals. Recently, a subspace-based MRSI approach called SPICE (SPectroscopic Imaging by exploiting spatiospectral CorrElation) has been proposed that enables new solutions to these challenges. This work presents our latest developments in data acquisition and spatiospectral processing within the SPICE framework, and demonstrates their capability in achieving ultrahigh-resolution ^{1}H-MRSI of the brain in 5 to 10 minutes.**

I. Introduction

MR spectroscopic imaging (MRSI) has been recognized as a potentially powerful tool for in vivo, label-free molecular imaging [1], [2]. Despite significant progress has been made in data acquisition and processing over the past several decades [3], practical applications of MRSI are still hindered by the challenges of low signal-to-noise ratio (SNR), poor spatial resolution, and slow imaging speed, and for ^{1}H-MRSI in particular, the overwhelming water/lipid signals. SPICE (SPectroscopic Imaging by exploiting spatiospectral CorrElation) has been recently proposed as a new approach to address these problems. SPICE uses a low-dimensional subspace model that exploits the spatiospectral partial separability within the high-dimensional spectroscopic signals to design special acquisition and processing strategies for rapid, high-resolution MRSI. In this work, we will present some latest developments that extend the SPICE framework and lead to unprecedented combinations of SNR, resolution and speed for ^{1}H-MRSI of the brain.

II. Subspace Model

We have extended the original subspace representation in SPICE to the following union-of-subspaces model [4], [5]

$$\rho(\mathbf{r},t) = \sum_{l_m=1}^{L_m} c_{l_m}(\mathbf{r})\phi_{l_m}(t) + \sum_{l_w=1}^{L_w} c_{l_w}(\mathbf{r})\phi_{l_w}(t) + \sum_{l_f=1}^{L_f} c_{l_f}(\mathbf{r})\phi_{l_f}(t) + \sum_{l_b=1}^{L_b} c_{l_b}(\mathbf{r})\phi_{l_b}(t), \quad (1)$$

where $\rho(\mathbf{r},t)$ is the spatiotemporal function of interest in MRSI, and the partial separability representations on the right-hand side denote the metabolite, water, lipid, and macromolecule baseline components, each of which resides in a low-dimensional subspace spanned by $\{\phi_{l_m}(t)\}_{l_m=1}^{L_m}$, $\{\phi_{l_w}(t)\}_{l_w=1}^{L_w}$, $\{\phi_{l_f}(t)\}_{l_f=1}^{L_f}$ and $\{\phi_{l_b}(t)\}_{l_b=1}^{L_b}$ (with L_m, L_w, L_f and L_b typically very small numbers). This low-dimensional subspace model can be written also as a sum of low-rank matrices (generalizable to low-rank tensors), meaning that the high-dimensional spatiotemporal/spatiospectral function of interest can be represented using a significantly reduced number of degrees-of-freedom, making accelerated, ultrahigh-resolution MRSI possible. Specifically, based on this subspace model, we have developed a novel acquisition strategy for rapid, volumetric spatiospectral encoding and time-interleaving sampling strategy which allows for correcting effects of system instability. A novel processing strategy is also developed for field inhomogeneity correction, water/lipid removal, and spatiospectral reconstruction from the noisy MRSI data which integrates subspace constraints and edge-preserving regularization.

III. Ultrahigh-Resolution ^{1}H-MRSI of the Brain

The proposed acquisition and reconstruction have been successfully applied for in vivo brain ^{1}H-MRSI experiments. Figure 2 shows a set of representative results from data acquired in an approximately 6-minute scan. As can be seen, high-resolution, high-SNR spatiospectral reconstruction is obtained within such a short acquisition window. We believe this new capability can enhance the utility of MRSI in various science and clinical applications.

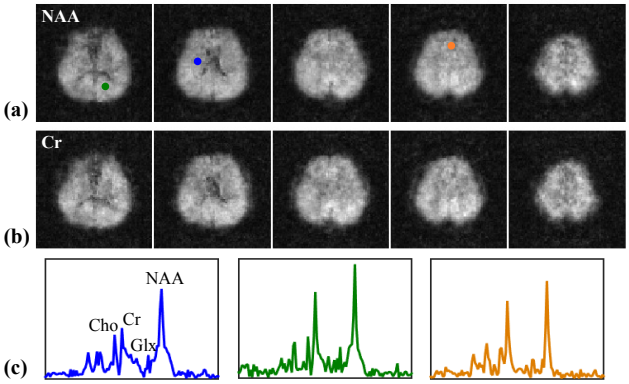


Fig. 1. Representative brain ^{1}H-MRSI results from a 6-min acquisition with 4 ms TE, 260 ms TR, a field-of-view of 230×230×72mm^3, and a matrix size of 76×76×24 (corresponding to a nominal resolution of isotropic 3mm). Images (a) and (b) show reconstructed high-resolution metabolite distributions, i.e., NAA and creatine (Cr) maps produced by the proposed method; image (c) shows spatially resolved spectra from different voxels indicated by the dots with different colors.

References

[1] P. C. Lauterbur, D. M. Kramer, W. V. House, and C.-N. Chen, "Zeugmatographic high resolution nuclear magnetic resonance spectroscopy: Images of chemical inhomogeneity within macroscopic objects," *J. Amer. Chem. Soc.*, vol. 97, pp. 6866 – 6868, 1975.
[2] R. A. de Graaf, *In Vivo NMR Spectroscopy: Principles and Techniques*. Hoboken, NJ: John Wiley and Sons, 2007.
[3] S. Posse, R. Otazo, S. R. Dager, and J. Alger, "MR spectroscopic imaging: Principles and recent advances," *J. Magn. Reson. Imag.*, vol. 37, pp. 1301 – 1325, 2013.
[4] F. Lam and Z.-P. Liang, "A subspace approach to high-resolution spectroscopic imaging," *Magn. Reson. Med.*, vol. 71, no. 4, pp. 1349–1357, 2014.
[5] C. Ma, F. Lam, C. L. Johnson, and Z.-P. Liang, "Removal of nuisance signals from limited and sparse ^{1}H MRSI data using a union-of-subspaces model," *Magn. Reson. Med.*, vol. 75, pp. 488 – 497, 2016.

The Challenges and Importance of Scale in Neuroscience

Karla Miller *,
* Oxford Centre for Functional MRI of the Brain (FMRIB), University of Oxford, Oxford, UK

***Abstract*—Modern neuroscience has an incredibly rich toolkit at its disposal, with a broad range of techniques that operate at massively different spatial and temporal scales. These techniques are crucial given that brain structure relates to function over at least eight orders of magnitude. A major challenge for neuroscience over the coming decade is to relate these measurements to each other in order to understand how microscopic features relate to whole-brain phenomena. For example, the exploding field of connectomics includes research on defining precise interconnections between local groups of neurons forming micro-circuits, as well as the long-range pattern of connections between large-scale brain regions. Relating these scales to each other is a major challenge that will require unifying models and detailed experimental work. Similarly, with the advent of population imaging, challenges exist in leveraging increasingly large-n investigations to provide insight at the level of individuals. This session will explore these challenges from the perspective of signal processing, biophysical modelling and data analytics.**

Multimodality reconstruction in PET/CT and PET/MR

Johan Nuyts*,
* Division of Nuclear Medicine, KU Leuven, Leuven, Belgium.

***Abstract*—Nowadays, PET systems are almost exclusively available as hybrid systems: PET/CT is a well established technology, PET/MR is commercially available since several years and its potential is currently being explored. The (almost) simultaneous acquisition of the anatomical and/or functional images and the molecular PET image creates new opportunities for reconstructing the tomographic images from all available data. Four such applications are discussed in this contribution. The anatomical image can be used to better regularize the resolution recovery during PET reconstruction. For MR imaging accelerated by undersampling, both modalities may benefit from joint reconstruction. Dynamic MR imaging providing motion fields can be used for motion compensated reconstruction from simultaneously acquired PET data. CT or MR can be used to estimate the attenuation image, and the time-of-flight information in TOF-PET data can be used to adjust and/or align that attenuation MAP maximizing consistency with the PET data.**

I. PET RECONSTRUCTION WITH ANATOMICAL PRIORS

By modeling the resolution in the system matrix, a sharper PET image is obtained. However, the image typically suffers from Gibbs artifacts, because deblurring is an ill-posed problem. In addition, for short acquisition times, these images can be very noisy. Edge preserving priors can be used to suppress the noise and Gibbs artifacts while preserving as much as possible the edges between regions with different tracer uptake. In some applications, in particular in brain imaging, it is reasonable to assume that the anatomy, as revealed by MR images, strongly correlates with tracer uptake. For those applications, the PET resolution recovery can be further improved by encouraging edges in the PET image to be aligned with boundaries in the MR image [1].

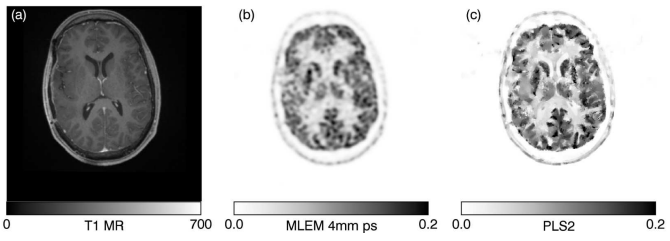


Fig. 1. PET/MR results, left: T1 weighted MR image; center: MLEM PET image (^{18}F-FDG, 5 min); right: PET image regularized with a parallel level sets prior based on the MR image.

II. JOINT PET AND MR RECONSTRUCTION

PET imaging suffers from limited spatial resolution and rather high levels of noise, but the sampling is redundant. MR imaging often has low noise and excellent resolution, but sampling may be insufficient to reduce the imaging time or improve the temporal resolution. These features seem complimentary, and therefore several groups are investigating to what extent both modalities can benefit from joint reconstruction, using priors that encourage similarity between some image features. The features should be chosen such that sampling artefact's are suppressed in the MR image and resolution recovery and noise suppression are improved in the PET image, but without exchanging features not shared by both images. Several joint priors have been proposed for this, including "parallel level sets" [3] and several flavors of joint total variation, e.g. [2]. A prior which performs well for this application will typically also do very well as an anatomical prior for PET reconstruction.

III. MR BASED MOTION CORRECTION

Because of the limited amount of administered radioactivity, tens of seconds or more are typically required to produce a PET image. Any motion occurring during the acquisition of such image frame will create motion blurring and therefore loss of resolution and degradation of quantitative accuracy. For periodic motion, a gated PET acquisition can be carried out, ensuring that images free of motion blurring are created by using for a particular image frame only data associated with a particular motion phase [4]. The gating signal can be obtained from external motion tracking devices or from MR navigator signals. Rigid and non-rigid aperiodic motion can be compensated for as well, provided that an accurate motion field is available. Again, such motion field can be acquired from external devices (e.g. optical rigid motion tracking in head imaging) or from fast MR sequences.

IV. ALIGNING THE ATTENUATION MAP

PET/CT imaging produces aligned PET and CT images which have been shown to have significantly more diagnostic value than PET or CT alone. In addition, the CT image is used for PET attenuation correction, which is required for quantitative and artifacts-free imaging. However, the images are often not perfectly aligned: the CT and PET data are acquired sequentially rather than simultaneously. In addition, the CT acquisition is very fast, enabling the acquisition of the image in a single breath hold. The PET acquisition requires more time and respiratory gating may be necessary to suppress blurring due to the motion. As a result, the CT-based attenuation map is often not well aligned to the true attenuation map that affected the PET acquisition. This poor alignment creates artifacts and adversely affects image quantification. Consequently, artifact-free PET reconstruction often calls for better alignment than obtained from the independent image reconstruction of both modalities.

Time-of-flight PET data are five dimensional and clearly redundant for the reconstruction of the three-dimensional image data. It has been shown that this additional time-of-flight information provides valuable information about the photon attenuation that occurred during the acquisition of the PET signal. That information can be used to align the CT image to the PET image by maximizing the consistency of the (attenuation corrected) PET data. The resulting optimal alignment eliminates the attenuation correction artifacts. In addition, an improved alignment may also be helpful in the interpretation of the multimodal PET/CT image.

REFERENCES

[1] K. Vunckx, et al. *IEEE Trans Med Imaging* 2012; 31.3: 599-612.
[2] F. Knoll, et al. *IEEE Trans Med Imaging*, 2016.
[3] MJ Ehrhardt et al., *Inverse Problems* 2015; 31: 015001.
[4] G. Bal, et al. *IEEE NSS/MIC*, 2014
[5] Rezaei, Ahmadreza, et al. *Phys Med Biol* 2016; 61.4: 1852.

Joint kq-space acceleration for fibre orientation estimation in diffusion MRI

Marica Pesce*, Anna Auria†, Alessandro Daducci†, Jean-Philippe Thiran†‡, Yves Wiaux*
* Institute of Sensors, Signals and Systems, Heriot-Watt University, Edinburgh, UK
† Signal Processing Lab (LTS5), École Polytechnique Fédérale de Lausanne, Switzerland
‡ University Hospital Center (CHUV) and University of Lausanne (UNIL), Switzerland

***Abstract*—We propose a method to accelerate the acquisition of High Angular Resolution Diffusion Magnetic Resonance Imaging (HARDI) in order to promote its application in a clinical setting. The method relies on a Spherical Deconvolution approach where the fibre orientation distribution (FOD) is recovered in all voxels simultaneously. The dMRI acquisition is meant to be accelerated through the partial Fourier sampling of each diffusion weighted image. Despite the further reduction of the acquired information, the FOD estimation still preserves its angular resolution thanks to the structured sparsity prior that is imposed in the problem.**

Diffusion magnetic resonance imaging is a unique non-invasive technique, enabling to extract information about the microscopic structure of white matter tissue in vivo. Spherical deconvolution approach is one of the several methods that have been developed in order to extract the fibre orientation information at high angular resolution. However, they rely on, at least, 60 diffusion images leading to considerable long acquisition times that prevent their application in the clinical setting.

The present study is based on the works of Daducci [1] and Auria [2], which exploit the recent theory of Compressive Sampling in order to recover the fibre orientations from a reduced number of diffusion images (q-space sampling). In particular, the work of Daducci. promotes the fibre orientation distribution sparsity through a voxel-wise ℓ_0-minimisation, suggesting an accurate reconstruction from no more than 30 q-space images. Auria et al. have built on this approach and introduced a spatial regularization prior promoting the smoothness of the spatial variation of fibre orientations, suggesting the FOD reconstruction from no more than 15 q-space images. We propose an extension of the above-cited methods where the acquisition of each diffusion image is accelerated through partial Fourier (k-space) sampling in order to fully exploit the spatial regularisation prior of Auria et al..

A novel linear measurement model is defined, mapping the matrix $X \in \mathbb{R}^{n\times N}$, which represents the FOD in each voxel of the imaged brain, onto the kq-space samples $\hat{Y} \in \mathbb{C}^{(N_c\times N_g)\times k}$ as follows:

$$\hat{Y}_{q,c} = M_q F P X M S_0 C^{(c)} H^{(q,c)} F M_k^{(q)} + \eta_{q,c} \qquad (1)$$

Each line of $\hat{Y}$ corresponds to the sub-sampled k-space of the DW-image acquired with gradient q by the channel with sensitivity c. F represent the Fourier matrix, the matrices $M_q \in \mathbb{R}^{1\times n}$ and $M_k \in \mathbb{R}^{N\times k}$ are binary masks representing the joint kq-space under-sampling of interest. The columns of the matrix $P \in \mathbb{R}_+^{n\times n}$ represent the ensemble average propagator for a single fibre oriented in all possible n directions. Voxels outside the brain are modelled with zero-signal through a diagonal matrix $M \in \mathbb{R}^{N\times N}$ while the diagonal matrix $S_0 \in \mathbb{R}^{N\times N}$ stores the intensities of the non-diffusion weighted image. The acquisition of the diffusion signal from multiple channels is taken into account through the diagonal matrix $C^{(c)} \in \mathbb{C}^{N\times N}$ which stores the sensitivity map of channel c. Motion and magnetic field inhomogeneities generate a phase distortion that is accounted in the matrix $H^{(q,c)} \in \mathbb{C}^{N\times N}$. The multi-channel sensitivities are assumed to be estimated from the non-diffusion weighted image while the phase distortion can be calibrated from low-resolution diffusion weighted images. Measurements $\hat{Y}_{q,c} \in \mathbb{C}^{1\times k}$ are assumed to be contaminated by Gaussian noise $\eta_{q,c} \in \mathbb{C}^{1\times k}$.

The fiber orientation distribution in each voxel is recovered solving a minimisation problem of the following form:

$$\min_{X\in\mathbb{R}_+^{n\times N}} \|\mathcal{A}(X) - \hat{Y}\|_2^2 \quad \text{subject to} \quad \|K_d \cdot X \cdot K_v\|_0 \leqslant \kappa \qquad (2)$$

where $K_d \in \mathbb{R}^{n\times n}$ and $K_v \in \mathbb{R}^{N\times N}$ represent two blurring bases imposing correlation within neighbour directions and neighbour voxels respectively, while $\mathcal{A}(\cdot)$ is the linear operator acting on X and modelling the measurements $\hat{Y}$. The parameter κ acts as a bound on the sparsity of X and it is computed as the number of voxels times the average number of fibre expected per voxel. We propose a multiple-shell approach for the q-space sampling in order to gain more accurate identification of the white matter tissue, joint with a uniform random k-space sampling with a fully sampled low frequency zone.

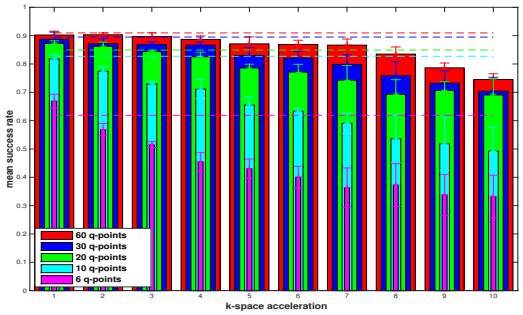


Fig. 1. Mean success rate index evaluating the performances of the FOD reconstruction in the case of different q-space and k-space undersampling settings. The results have been obtained from synthetic data with SNR=30.

The fibre orientation reconstruction has been tested through numerical simulations and real data in presence of different acceleration rates. The results suggest that the kq-space approach can significantly outperform the q-space sampling up to an acceleration of 7 with 60 diffusion images in simulated data (see Figure 1), rising the hopes to open the door of clinical applications for high angular resolution diffusion imaging methods.

References

[1] A. Daducci, D. Van De Ville, J. P. Thiran, and Y. Wiaux, "Sparse regularization for fiber odf reconstruction: From the suboptimality of ℓ_2 and ℓ_1 priors to ℓ_0," *Medical Image Analysis*, vol. 18, pp. 820–833, 2014.

[2] Auría, A. and Daducci, A. and Thiran, J.-P. and Wiaux, Y., "Structured sparsity for spatially coherent fibre orientation estimation in diffusion mri," *NeuroImage*, vol. 115, 2015.

Learning a variational model for image reconstruction

Kerstin Hammernik*, Yunjin Chen*, Florian Knoll†, Daniel Sodickson† and Thomas Pock*
* Institute of Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16, 8010 Graz, Austria.
† New York University School of Medicine, 660 First Avenue, New York, NY 10016, USA.

***Abstract*—We consider a learned variational approach for image reconstruction. The variational model contains a highly parametrized regularizer that can adapt to the structural properties of natural images. The parameters of the model are learned from data using a loss-based approach. The variational model is approximately minimized using one cycle of a block-incremental gradient descent algorithm. We show applications to image denoising, image superresolution, JPEG deblocking and MRI reconstruction from sparsely sampled data.**

I. Introduction

We consider the ill-posed linear inverse problem of finding a reconstructed image u that satisfies the following system of equations

$$Au = f,$$

where f is the input data and A is a linear forward operator. Due to noisy or sparse measurements f, the equation cannot be solved directly for u. A classical approach to overcome this problem is to consider a regularized least squares approach

$$\min_u \mathcal{R}(u) + \frac{1}{2}\left\|Au - f\right\|_2^2,$$

where $\mathcal{R}$ is a regularization term that imposes a smoothness prior onto the minimizer u. One of the most influential regularization terms in the context of image reconstruction is the total variation semi-norm [1]. However, it is well-known that the total variation is too simple to capture the complicated structure of real-world images.

The aim of this work is to consider a very flexible regularization term that is learned from data such that the solution of the variational model is as close as possible to ground truth solutions.

II. The proposed method

We consider the so-called fields of experts (FoE) model [2]. The model is written as:

$$R(u) = \sum_{j=1}^{N} r_j(u), \text{ with } r_j(u) = \phi_j(k_j * u),$$

where $N \gg 1$, is the number of "experts" and each expert r_j consists of a convolution kernel k_j together with its potential function ϕ_j. The particular choice of the kernels and functions is free and will be learned from data.

For computing an (approximate) minimizer of the variational model, we perform one cycle of a block-incremental gradient method [3]. Starting from an initial image u^0, the iterations $k = 0...K-1$ are given by

$$u^{k+1} = u^k - \alpha_k \left(\sum_{j=kb+1}^{(k+1)b} \nabla_u r_j(u^k) + A^*(Au^k - f) \right),$$

where b is the block size, α_k are step sizes and $K = N/b$ is the number of iterations. The idea of the algorithm is to perform a finite number of K steps where each step consists of a gradient step with respect to the data term and a block-gradient descent step with respect to the regularizer. The algorithm might also be interpreted as an explicit scheme of a non-linear reaction-diffusion equation or a deep convolutional neural network, see [4] for more information.

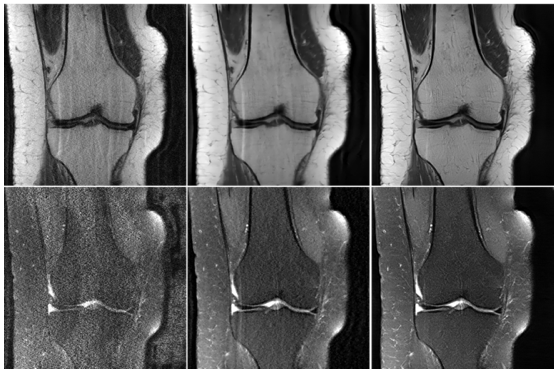


Fig. 1. MRI reconstruction using the proposed method. From left to right: Least squares solution, TGV reconstruction [5], and reconstruction using the learned method.

III. Learning

Inspired by recent deep learning activities, we propose to learn the parameters of the variational model (filters k_j, potential functions ϕ_j, and steps α_k) from data. Therefore, we make use of training data consisting of input data $(f_i)_{i=1}^M$ and its corresponding ground truth solutions $(u_i^*)_{i=1}^M$. For learning, we minimize a loss function ℓ that measures the similarity between the obtained reconstructions u_i^K and the ground truth solution u_i^*.

$$\min_{(k_j,\phi_j)_{j=1}^N,(\alpha_k)_{k=0}^{K-1}} \sum_{i=1}^{M} \ell(u_i^K - u_i^*).$$

Figure 1 shows an example of MRI reconstruction from 4-fold undersampled data. One can see that the learned variational model yields significantly better results compared to a classical least-squares solution and TGV regularized variational approach [5].

References

[1] L. Rudin, S. J. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.
[2] S. Roth and M. J. Black, "Fields of Experts," *Int'l J. Computer Vision*, vol. 82, no. 2, pp. 205–229, 2009.
[3] A. Nedic and D. Bertsekas, "Incremental subgradient methods for non-differentiable optimization," *SIAM J. on Optimization*, vol. 12, no. 1, pp. 109–138, Jan. 2001.
[4] Y. Chen, W. Yu, and T. Pock, "On learning optimized reaction diffusion processes for effective image restoration," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
[5] F. Knoll, K. Bredies, T. Pock, and R. Stollberger, "Second order total generalized variation (TGV) for MRI," *Magnetic Resonance in Medicine*, vol. 65, no. 2, pp. 480–491, 2011.

The New Age of Optical Microscopy

Eli Rothenberg*
* New York University, School of Medicine Medical Science Building, New-York, USA

***Abstract*—For the past two decades we witnessed unprecedented advances in the field of optical microscopy. These resulted in an array of novel methods which provide diverse modalities for probing biological samples, ranging from selective illumination over a large sample field to detection of individual molecules and resolving molecular features at nanometer scales, far below the classical diffraction limit of light (super-resolution microscopy). While these techniques offer unique and previously unattainable data they also present new challenges in the handling, derivation and analyses of this new data. I will provide an overview of the progress and the different modalities of recent optical microscopy methods, and discuss the technology, capabilities, applications, analysis and ongoing and future developments and challenges in the field.**

Population Imaging: MRI Meets Big Data

Stephen Smith *,
* Oxford Centre for Functional MRI of the Brain (FMRIB), University of Oxford, Oxford, UK

***Abstract*—In this talk, I will discuss how population imaging, as embodied in several very large scale projects focused on non-invasive imaging of the human brain, is aiming to remap the neuroscience landscape. These projects are producing datasets of unprecedented size and complexity, and are not only serving as important open resources for neuroscientists, but are also spurring rapid technological innovations in data acquisition and analysis.**

The brains connectivity structure is a key determining factor since this determines the ways that information can be integrated. The Human Connectome Project based in the US aims to provide the most comprehensive map of the human brains connectivity structure. The original project scanned 1200 healthy adults using bleeding-edge imaging technologies, many newly developed as part of the project. A number of major advances have already come out of this project, including rapid imaging methods, techniques for dissecting network dynamics, and the most highly detailed delineation of brain regions yet produced. This endeavour has now been expanded to include focussed disease groups and a range of ages to enable the insights provided by the HCP to inform us about disease and development.

Medical imaging has enormous potential for early disease prediction, but is impeded by the difficulty and expense of acquiring data sets before symptom onset. UK Biobank aims to address this problem directly by acquiring high-quality, consistently acquired imaging data from 100,000 predominantly healthy participants, with health outcomes then being tracked over the coming decades. The brain imaging includes structural, diffusion and functional modalities. Along with body and cardiac imaging, genetics, lifestyle measures, biological phenotyping and health records, this imaging is expected to enable discovery of imaging markers of a broad range of diseases at their earliest stages, as well as provide unique insight into disease mechanisms.

As these data resources grow and become enriched by long-term health outcomes and follow-on studies, a number of important challenges need to be addressed. While brain image analysis is highly automated for many types of questions, many of these techniques are not currently suitable for analyzing large data sets. More interestingly, large numbers of subjects create new opportunities in the domain of data-driven analysis. Finally, we will need to tackle challenges of interpretation in the face of a large number of statistically significant results that nevertheless explain relatively little variance.

Pulseq: A Rapid and Hardware-Independent Pulse Sequence Prototyping Framework

Stefan Kroboth*, Kelvin Layton*, Feng Jia*, Sebastian Littin*, Huijun Yu*,
Jochen Leupold*, Jon-Fredrik Nielsen†, Tony Stoecker‡ and Maxim Zaitsev*
* Department of Radiology, Medical Physics, University Medical Center Freiburg, Freiburg, Germany
‡ German Center for Neurodegenerative Diseases, Bonn, Germany
† Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan, USA

***Abstract*—We introduce a novel framework for sequence design and execution on arbitrary MR hardware. The proposed architecture allows for the decoupling of sequence design and execution on MR hardware via a novel file format for sequence description. This makes sequence design hardware-independent and aids researchers in sharing sequences and running the same sequences on different platforms.**

I. Introduction

Implementing new MR sequences often involves extensive programming on vendor-specific platforms, which can be time consuming and costly. Even more so if research sequences need to be implemented on several platforms simultaneously, for example, at different field strengths. This work presents an alternative programming environment that is hardware-independent, open-source, and promotes rapid sequence prototyping. We introduce a novel file format which allows for the efficient description of hardware events and timing information required for an MR pulse sequence. Platform-dependent interpreter modules convert the file to appropriate instructions to run the sequence on MR hardware. The design of sequences is done in high-level programming languages such as MATLAB (The Mathworks, Natick, MA) or with a graphical interface (i.e. JEMRIS [1]). This highly-flexible pulse sequence programming environment is called Pulseq [2] and allows for the decoupling of sequence design (hardware independent) and sequence execution (hardware dependent). Furthermore, JEMRIS can be used to simulate spin physics, allowing for comparison between real and virtual experiments.

II. Methods

The main components of the Pulseq environment are illustrated in Fig. 1. The high-level sequence can be described directly in MATLAB using functions from a custom toolbox. Alternatively, sequences can be programmed using the graphical interface of the JEMRIS simulation packages. Regardless of the choice of high-level interface, a sequence file is created containing low-level sequence instructions such as RF pulses, gradients, ADC events and delays. This sequence file can then be executed on various platforms through hardware-dependent interpreter modules.

III. Experiments

Figure 2 shows the results of a gradient echo sequence executed on three different hardware platforms: a 3T Siemens Trio equipped with a single-channel wrist RF coil (Siemens Healthcare, Erlangen, Germany); a 3 T GE Discovery MR750 with a 8 channel head coil (GE Healthcare, Waukesha, WI, USA); and a 9.4 T Bruker BioSpec MRI with a single-channel rat coil (Bruker Biospin, Ettlingen, Germany).

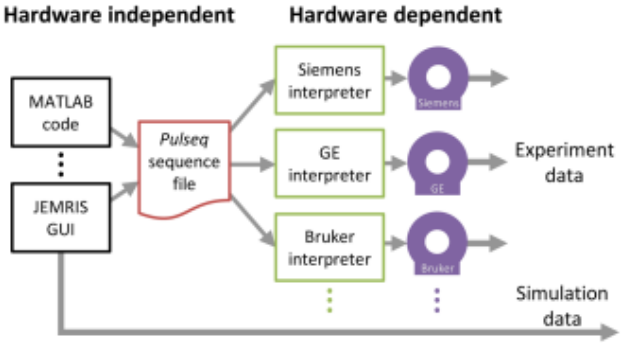


Fig. 1. Overview of the Pulseq environment. Sequences are described in a high-level design tool (left). A hardware-indpendent sequence format is output (middle) and executed using a hardware-dependent interpreter module (right). Simulation data may also be generated from the Bloch equation solver JEMRIS.

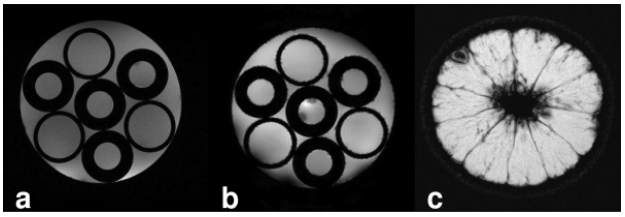


Fig. 2. Images from the same sequence file executed on Siemens (a), Bruker (b) and GE (c).

IV. Results and Discussion

The new sequence format compactly represents arbitrary sequences in an hardware-independent way. The platform independence can be advantageous for institutions with multiple scanners and also aid in the sharing of sequences between institutions. Sequences can easily be evaluated by comparing simulations and measurements based on the same sequence file.

References

[1] T. Stoecker, K. Vahedipour, D. Pflugfelder, and N. J. Shah, "High-performance computing MRI simulations," *Magnetic Resonance in Medicine*, vol. 64, no. 1, pp. 186–193, 2010. [Online]. Available: http://dx.doi.org/10.1002/mrm.22406

[2] K. J. Layton, S. Kroboth, F. Jia, S. Littin, H. Yu, J. Leupold, J.-F. Nielsen, T. Stoecker, and M. Zaitsev, "Pulseq: A rapid and hardware-independent pulse sequence prototyping framework," *Magnetic Resonance in Medicine*, pp. n/a–n/a, 2016. [Online]. Available: http://dx.doi.org/10.1002/mrm.26235

ASTRA Toolbox: a flexible, efficient and open source toolbox for tomographic reconstruction

W. Van Aarle[*], W. J. Palenstijn[†] , J. De Beenhouwer[*], K. J. Batenburg[†], J. Sijbers[*],
[*] iMinds-Visionlab, University of Antwerp, Belgium.
[†] Centrum Wiskunde & Informatica, Amsterdam, The Netherlands

Introduction
In this abstract, the ASTRA toolbox (All Scales Tomographic Reconstruction Antwerp) is introduced [1-3]. It is an open source, GPU-accelerated library for 3D and 4D image reconstruction in tomography. The basic building blocks are fast forward and backward projectors. The design of the ASTRA toolbox allows for full flexibility in specifying the geometry while still maintaining an efficient mapping onto the underlying hardware. The toolbox can be downloaded from http://www.astra-toolbox.com/

Flexible acquisition geometries
The ASTRA toolbox provides a set of highly flexible building blocks that can be used to construct advanced reconstruction algorithms for arbitrary acquisition geometries (e.g., circular cone-beam, arbitrary planar cone beam [6], laminography [9,10], tomosynthesis [11], conveyor belt setup, etc).

Multiple modalities
The ASTRA toolbox provides building blocks for Radon transform based tomography. Hence, it can be invoked not only for X-ray or Gamma tomography with synchrotron radiation or micro-CT systems [1], but as well for electron tomography [2], neutron tomography [8], etc.

N-D tomography
The ASTRA toolbox allows to build novel reconstruction methods for 2D (e.g., parallel beam), 3D (e.g. cone-beam), 4D (e.g. dynamic CT [7,13]), or even 6D (diffraction tomography).

User friendly
The ASTRA toolbox comes with a MATLAB and PYTHON interface for easy user interaction and is available for both Windows and Linux. It can be easily integrated into other tomographic processing software [4,5]. Some examples from different application fields will be touched upon. Finally, an outlook is given towards large data tomographic reconstruction with the ASTRA toolbox.

REFERENCES

[1] W. Van Aarle, et al, "Fast and Flexible X-ray Tomography Using the ASTRA Toolbox", Optics Express, In Press
[2] W. Van Aarle, et al, "The ASTRA Toolbox: a platform for advanced algorithm development in electron tomography", Ultramicroscopy, 157, 35–47, 2015
[3] W J. Palenstijn, K. J. Batenburg, and J. Sijbers, "The ASTRA Tomography Toolbox", 13th International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE 2013, vol. 4, 1139-1145, 2013
[4] D. Pelt, et al, "Integration of TomoPy and the ASTRA toolbox for advanced processing and reconstruction of tomographic synchrotron data", Journ. Synch. Radiation, 23, 842-849, 2016
[5] F. Bleichrodt, et al, "Easy implementation of advanced tomography algorithms using the ASTRA toolbox with Spot operators", Numerical Algorithms, 71(3), 673-697, 2016
[6] A. Dabravolski, et al, "Adaptive zooming in X-ray computed tomography", Journal of X-Ray Science and Technology, vol. 22, no. 1, 77-89, 2014
[7] G. Van Eyndhoven, et al, "An iterative CT reconstruction algorithm for fast fluid flow imaging", IEEE Transactions on Image Processing, vol. 24, issue 11, pp. 4446-4458, 2015
[8] H. K. Jenssen, et al., "Neutron radiography and tomography applied to fuel degradation during ramp tests and loss of coolant accident tests in a research reactor", Progress in Nuclear Energy, vol. 72, pp. 55-62, 2014
[9] J. Cant, et al, "Continuous Digital Laminography", 6th Conference on Industrial Computed Tomography, 2016
[10]K. J. Batenburg, et al, "3D imaging of semiconductor components by discrete laminography", Stress induced phenomena and reliability in 3D microelectronics: AIP Conference Proceedings, vol. 1601, pp. 168-179, 2014
[11]J. Cant, et al., "Automatic geometric calibration of chest tomosynthesis using data consistency conditions", The 4th International Conference on Image Formation in X-Ray Computed Tomography, pp. 161-164, 2016
[12] T-T, Thibault; et al, Soft Route to 4D Tomography, Physical Review Letters, 117(2), 2016
[13]V. Van Nieuwenhove, et al, "Affine deformation correction in cone beam Computed Tomography", Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, Newport, Rhode Island, USA, pp. 182-185, 2015

JEMRIS: a general-purpose MRI simulator

Tony Stöcker *†
* German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany
† Department of Physics and Astronomy, University of Bonn, Bonn, Germany

***Abstract*—MR physics computer simulations are of high educational value. Further, they serve as essential tools in basic MRI method development, sequence design and protocol optimization, and generating ground truth data for image reconstruction and post-processing algorithms. The open-source C++ software project JEMRIS provides a versatile multi-platform MRI simulation environment. General Bloch equation-based modeling of a large spin system is combined with the most important off-resonance effects, parallel receive and transmit, nonlinear gradient fields, and spatiotemporal parameter variations at different levels, e.g. to simulate object motion, molecular diffusion or tissue contrast agent uptake. The software provides a highly flexible module for the interactive design of general MR pulse sequences.**

I. Introduction

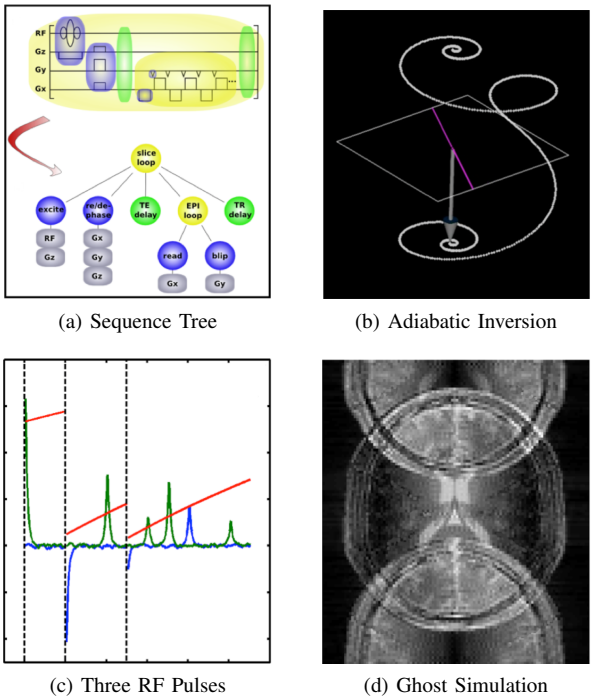
Beyond its high educational value, MR physics simulation is an important research tool in many different areas such as rapid prototyping of new ideas, validating physical models and hypotheses, RF pulse design, or the generation of ground-truth data to test novel algorithms for image reconstruction and image analysis. Moreover, MR simulations provides easy access to investigate the limits of MRI which are often harder to explore in real experiments. Therefore, the open source C++ project JEMRIS was started in 2006 to provide a versatile MR simulation environment to the community [1]. Active development over the past 10 years strongly improved the software. It has many active users world-wide and was successfully used in different research projects, e.g. to investigate the spin-echo BOLD signal at high field [2] or to generate ground-truth MRI data of the carotids for image analysis [3]. JEMRIS provides a flexible pulse sequence design interface which was recently extended to control the sequence on real MR hardware by means of the novel low-level hardware independent file format PulSeq [4].

II. Methods

JEMRIS performs numerical integration of the general Bloch equations on user-defined objects consisting of a large spin ensemble. The semi-classical approach on non-interacting spins, a valid assumption in most MRI applications, is inherently well-suited for massive parallel computing. The software provides an easy-to-use framework for MR pulse sequence design which impose little limitations on the complexity of the experiments. The simulations take various physical effects into account which have impact on real MR experiments: susceptibility-induced off-resonance, chemical shift, eddy currents, concomitant fields, spatial non-linearity of the encoding fields ("gradients") and RF reception and transmission with arbitrary number of channels in user-defined coil configurations. Moreover, a general approach was implemented to simulate spatiotemporal parameter variations of the object. This can be utilized for realistic MR simulations of rigid motion or flow and diffusion of spins.

III. Results

Figure 1 (a) shows the concept of pulse sequence serialization in JEMRIS. Figs. 1, (b), (c), and (d) show different use-cases of basic MR simulations: inversion of a single spin, the famous five echoes generated by three RF pulses, and artificial ghost images induced by eddy currents as observed in EPI. More examples and use-cases will be presented and discussed at the workshop.



(a) Sequence Tree (b) Adiabatic Inversion

(c) Three RF Pulses (d) Ghost Simulation

Fig. 1. (a) Sketch of a native EPI pulse sequence (top) and its software representation as a left-right ordered tree (bottom). (b) Spin dynamics for an adiabatic inversion pulse. (c) MR signals resulting from three RF pulses. (d) EPI ghost artifacts induced by eddy currents.

IV. Conclusions

The versatile MR simulation environment JEMRIS is a general-purpose tool for MRI physicists and other scientists interested in generating realistic MRI simulation data. Implementation details and computational costs for different applications will be discussed at the workshop.

References

[1] Stöcker, T. et al., "High performance computer MRI simulation," *MRM*, vol. 64, pp. 186–193, 2010.
[2] Pflugfelder, D. et al., "On the numerically predicted spatial BOLD fMRI specificity for spin echo sequences." *Magnetic Resonance Imaging*, vol. 29, pp. 1195–204, 2011.
[3] Nieuwstadt, H. et al., "Numerical simulations of carotid MRI quantify the accuracy in measuring atherosclerotic plaque components in vivo," *MRM*, vol. 72, pp. 188–201, 2014.
[4] Layton, K. et al., "Pulseq: A rapid and hardware-independent pulse sequence prototyping framework," *MRM*, vol. Early View, 2016. [Online]. Available: http://dx.doi.org/10.1002/mrm.26235

Photon Counting Spectral X-Ray CT: Toward Tissue-Specific Quantitative Imaging

Katsuyuki Taguchi
Division of Medical Imaging Physics, The Russell H. Morgan Department of Radiology and Radiological Science,
The Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA

***Abstract*—The development of energy-sensitive photon-counting x-ray detectors (PCD) has created great excitement in x-ray and x-ray CT systems. Such innovative new x-ray detectors count individual photons and sort them into selected energy bins. It is said that PCDs will not only improve anatomical or functional CT imaging significantly but also provide an opportunity for molecular CT imaging and low-dose CT. On the side of enthusiasm, a lot of questions are being asked. Are count rates of PCDs sufficient for intense x-ray flux of CT systems? Is the current energy resolution sufficient? What are the imaging technologies that need to be developed for PCD-CT and what are the remaining issues? When will the first commercial PCD-CT system be introduced? Aiming at providing answers to the questions listed above, we will review the current status and perspectives of the imaging technologies for PCD-CT. Methods to model, calibrate, and compensate for the non-ideal properties of PCDs will be discussed. Algorithms to reconstruct images from spectral data will be presented.**

We will present overview of two papers [1, 2] in this talk.

Since their discovery in 1895 by Wilhelm C. Röntgen, x-rays have been playing a critical role in medical imaging. They are helping radiologists and physicians to detect and characterize disease processes of the skeletal system, soft tissue, and their functionality. Transmitted and detected x-ray beams generate a snapshot projection image, a series of projection images, or cross-sectional tomographic images. Multi-slice x-ray computed tomography (CT) scanners provide three-dimensional images of the linear attenuation coefficient distribution within a patient, accurately delineating organs and tissues. However, there are four major limitations to current CT and x-ray technologies: 1) the contrast between different soft tissues is often insufficient; 2) images are not tissue-type specific (different tissue types can appear with similar pixel values); 3) "CT scanning is a relatively high-dose procedure" [3]; and 4) gray-scale pixel values of CT images, which should be linear attenuation coefficients, are not quantitative but qualitative (see Sec. V.E for more discussion). These limitations result from or are made worse by the energy integrating detectors used in CT.

Factors influencing the x-ray linear attenuation coefficients include the chemical composition and mass density of the object, and the energy of the x-ray photons. Therefore, the transmitted x-ray spectra carry information about different tissue types. The energy-integration detectors (EIDs), however, measure the energy-integrated signals of x-ray photons, thus losing all of the energy-dependent information. In addition, EIDs weight lower energy photons less, which carry larger contrast between tissues than higher energy photons. This results in increased noise and decreased contrast.

Recently, photon counting detectors (PCDs) with energy discrimination capabilities based on pulse height analysis have been developed for medical x-ray imaging. These PCDs count the number of photons of the transmitted x-ray spectrum using between two and eight energy windows. PCD-based CT systems with multiple energy windows have the potential to improve the four major limitations we discussed before. Electronic and Swank noise affect the measured energy, but do not change the output signal intensity (i.e., the counts), and the energy overlap in the spectral measurements can be smaller than that from any of the current dual-energy techniques using EIDs. In addition, more than one contrast medium can be imaged simultaneously and becomes distinguishable if the detectors have four or more energy thresholds or windows. PCDs may therefore lead to novel clinical applications as will be discussed.

The performance of PCDs is not flawless, however, especially with the large count rates in current clinical CT. Due to the stochastic nature of time intervals between photon arrivals and the limited pulse resolving time, quasi-coincident photons generate overlapping pulses which may be recorded as a single count with a wrong energy. This phenomenon is called pulse pileup and results both in a loss of counts, referred to as dead time loss, and a distortion of the recorded spectrum. It is therefore critical to develop schemes to compensate for these effects. Other phenomena may also degrade the spectral response of PCDs, including incomplete charge collection generated by x-rays due to charge sharing and charge trapping effects. We will review various performance degradation factors.

Our aim with this paper is to provide the current status and future perspective of key technologies and applications of PCDs in medical imaging. Three technologies discussed are the detector technologies, imaging technologies, and system technologies. Potential benefits and clinical applications of PCD-based CT systems are discussed.

REFERENCES

[1] K. Taguchi and J. S. Iwanczyk, "Vision 20/20: Single photon counting x-ray detectors in medical imaging," *Medical Physics,* vol. 40, p. 100901, 2013.

[2] K. Nakada, K. Taguchi, G. S. K. Fung, and K. Amaya, "Joint estimation of tissue types and linear attenuation coefficients for photon counting CT," *Medical Physics,* vol. 42, pp. 5329-5341, 2015.

[3] F. A. Mettler, Jr., P. W. Wiest, J. A. Locken, and C. A. Kelsey, "CT scanning: patterns of use and dose," *Journal of Radiological Protection,* vol. 20, pp. 353-359, 2000.

The BART Toolbox for Computational Magnetic Resonance Imaging

Martin Uecker*†, Jonathan I. Tamir‡, Frank Ong‡ and Michael Lustig‡
* Department of Diagnostic and Interventinal Radiology, University Medical Center Göttingen, Göttingen, Germany
† DZHK (German Centre for Cardiovascular Research), partner site Göttingen, Germany
‡ Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, USA

***Abstract*—We present our BART Toolbox for computational Magnetic Resonance Imaging (MRI). The main motivation for the development of this toolbox was the simultaneous need for rapid prototyping of new computational imaging methods for MRI and for highly efficient implementations. The main philosophy is the use of generic numerical algorithms and re-usable and highly configurable software components. The BART toolbox consists of programming libraries and flexible command-line tools. It contains tools for simulation, pre-processing, calibration, and image reconstruction.**

I. Compressed Sensing and Parallel Imaging

Compressed sensing is based on the idea that a sparse signal can be recovered using iterative denoising from undersampled data if the aliasing is incoherent. This idea can be applied to MRI and also combined with parallel imaging [1], [2]. In parallel imaging, the signal can be modelled as samples of the Fourier transform of the magnetization image ρ modulated by the receive-coil sensitivities c_j along a given k-space trajectory $k(t)$:

$$y_j(t) = \int \mathrm{d}\vec{r}\, \rho(\vec{r}) c_j(\vec{r}) e^{-2\pi i \vec{k}(t)\cdot\vec{r}}$$

If the coil sensitivities c_j are known, image reconstruction can be formulated as a linear inverse problem [3]. For autocalibrating parallel imaging the sensitivities must be estimated from the data. This yields a bilinear problem which is similar to blind multi-channel deconvolution. Three different reconstruction approaches are implemented in BART: non-linear inversion (NLINV) [4], structured low-rank matrix completion (SAKE) [5], and ESPIRiT [6] which is based on identification of the signal subspace.

II. Generalized Reconstruction

Many recent methods are based on high-dimensional reconstruction problems which include additional time and parametric dimensions (see Fig. 1 for examples). To facilitate experimentation, BART implements a generic framework based on the following optimization problem [7]:

$$\arg\min_x \quad \sum_j \| W (P \mathcal{F} \sum_k S_j{}^k x^k - y_j) \|_2^2 + \sum_i^N \lambda_i f_i(B_i x)$$

Here, $\mathcal{F}$ is a multi-dimensional Fourier transform, P is a sampling operator, S the multiplication with the sensitivities, W a weighting matrix, the B_i are linear operators, f_i convex functions, λ_i regularization parameters. For arbitrary combinations of certain regularization terms and transforms along arbitrary dimensions, this problem can be solved using ADMM and a library of proximity functions using the following BART command:

> bart pics -Rf:B:C:λ -R ... [-t P] -p W y S x

III. Conclusion

BART provides a flexible and efficient framework for rapid prototyping of advanced computational methods in MRI.

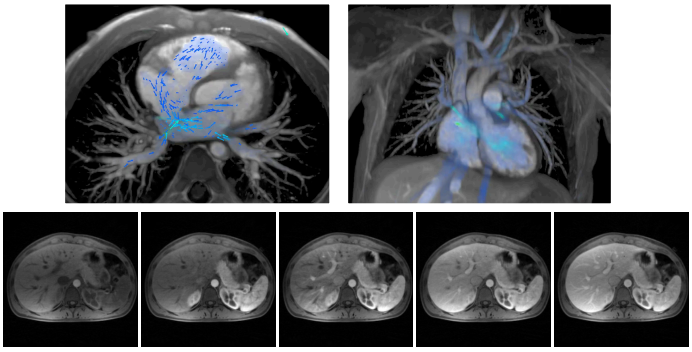


Fig. 1. Reconstruction of high-dimensional data using BART. **Top:** Highly accelerated 4D-flow [9]. **Bottom:** Dynamic contrast-enhanced MRI using GRASP [10]. Data courtesy of Joseph Y. Cheng and Tobias Block.

References

[1] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging." *Magn. Reson. Med.*, vol. 58, pp. 1182–1195, 2007.
[2] K. T. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils. iterative image reconstruction using a total variation constraint," *Magn. Reson. Med.*, vol. 57, pp. 1086–1098, 2007.
[3] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories." *Magn. Reson. Med.*, vol. 46, pp. 638–651, 2001.
[4] M. Uecker, T. Hohage, K. T. Block, and J. Frahm, "Image reconstruction by regularized nonlinear inversion - joint estimation of coil sensitivities and image content," *Magn. Reson. Med.*, vol. 60, pp. 674–682, 2008.
[5] P. J. Shin, P. E. Z. Larson, M. A. Ohliger, M. Elad, J. M. Pauly, D. B. Vigneron, and M. Lustig, "Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion." *Magn. Reson. Med.*, vol. 72, pp. 959–970, 2014.
[6] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig, "ESPIRiT - an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA." *Magn. Reson. Med.*, vol. 71, pp. 990–1001, 2014.
[7] J. I. Tamir, F. Ong, J. Y. Cheng, M. Uecker, and M. Lustig, "Generalized magnetic resonance image reconstruction using the berkeley advanced reconstruction toolbox," in *ISMRM Workshop on Data Sampling and Image Reconstruction*, 2016.
[8] M. Uecker, F. Ong, J. I. Tamir, D. Bahri, P. Virtue, J. Y. Cheng, T. Zhang, and M. Lustig, "Berkeley advanced reconstruction toolbox," in *Proc. Intl. Soc. Mag. Reson. Med*, vol. 23, 2015, p. 2486.
[9] J. Y. Cheng, K. Hanneman, T. Zhang, M. T. Alley, P. Lai, J. I. Tamir, M. Uecker, J. M. Pauly, M. Lustig, and S. S. Vasanawala, "Comprehensive motion-compensated highly-accelerated 4D flow MRI with ferumoxytol enhancement for pediatric congenital heart disease," *Journal of Magnetic Resonance Imaging*, vol. 43, pp. 1355–1368, 2016.
[10] L. Feng, R. Grimm, K. T. Block, H. Chandarana, S. Kim, J. Xu, L. Axel, D. K. Sodickson, and R. Otazo, "Golden-angle radial sparse parallel MRI: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI," *Magn. Reson. Med.*, vol. 72, pp. 707–717, 2014.

Predictive models to compare subjects from functional connectivity

Gaël Varoquaux*†‡, Alexandre Abraham*†‡, Kamalaker Dadi*†‡, Mehdi Rahim*†‡, and Bertrand Thirion*†‡,
* Parietal, Inria Saclay, Palaiseau, France
† Neurospin, I2BM/DRF/CEA, Saclay, France
‡ Université Paris-Saclay, France

***Abstract*—Brain functional connectivity, derived from resting-state fMRI acquisition can be used to study inter-subject differences, extracting neuro-phenotypes capturing that capture individual's psychological traits or psychiatric disorders. Here we study data-analysis approaches that build a connectome –a matrix of interactions– between brain regions and rely on supervised learning to discriminate subjects. Using a variety of applications to neuro-psychiatric disorders, we benchmark the choice of different methods for the various steps of the processing pipelines.**

Functional-connectivity fMRI captures neural interactions via fluctuations in the observed brain signals. Comparing functional connectivity across subjects can reveal mechanisms or biomarkers of pathologies from resting-state experiments. Resting-state acquisition are particularly interesting in a population-imaging as they are easy to run at a large scale, even on diminished population.

Standard neuroimaging analyzes are based on mapping the response of individual brain modules. However, studying interactions calls for a more complex model. It has given rise to a hoard of approaches, ranging from ICA to small-word networks.

We present a consistent inference framework bridging these various methods. The central notion is that of a "connectome", describing interaction strengths between regions of the brain. Specifically, we consider connectome-extraction pipelines composed of four steps: 1) region estimation, 2) time series extraction, 3) matrix estimation, and 4) classification –see figure 1. The corresponding description of brain activity maps easily to intuitions on brain function and connectivity as well as solid mathematical underpinnings that can guide data-processing choices.

We give benchmarks of the various aspects of the connectome extraction and comparison pipelines in predictive-modeling settings. Used on populations of subjects they capture individual behavior traits from resting-state activity. We have successfully used connectome-based prediction on various neuro-psychiatric disorders. In particular, we have shown prediction of autism spectrum disorder from resting-state acquisition of completely unknown scanning sites.

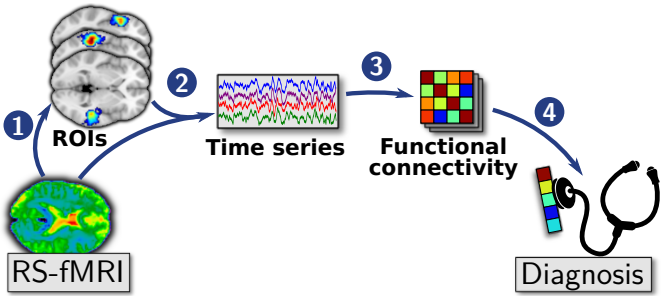


Fig. 1. The different steps of a connectome-based predictive pipeline.

I. Acknowledgments

This work is supported by the NiConnect project (ANR-11-BINF-0004 NiConnect).

References

[1] A. Abraham, M. Milham, A. Di Martino, R. C. Craddock, D. Samaras, B. Thirion, and G. Varoquaux, "Deriving robust biomarkers from multi-site resting-state data: An autism-based example," *bioRxiv*, p. 075853, 2016.

[2] K. Dadi, A. Abraham, M. Rahim, B. Thirion, and G. Varoquaux, "Comparing functional connectivity based predictive models across datasets," in *PRNI 2016: 6th International Workshop on Pattern Recognition in Neuroimaging*, 2016.

Blind Equalization of Multipath Channels using Random Coding

Ali Ahmed*
* School of Electrical and Computer Engineering, Information Technology University, Pakistan.

***Abstract*—We consider recovering vectors $\boldsymbol{h} \in \mathbb{R}^L$, and $\boldsymbol{x}_n \in \mathbb{R}^L$ from their circular convolutions $\boldsymbol{y}_n = \boldsymbol{h} * \boldsymbol{x}_n, \quad n = 1, \dots, N$. As is evident that in these multiple observed convolutions, one of the vectors is fixed while the other varies. The vector $\boldsymbol{h}$ is only assumed to be S-sparse, and each of $\boldsymbol{x}_n$ is a member of a known K-dimensional subspace.**

Using lifting, one can frame the deconvolution as a joint rank-1, and sparse matrix recovery problem [1]. The joint structure cannot be efficiently relaxed, however, we show that by allowing the quantity N of inputs to exceed a minimal number, it is possible to effectively solve the problem by relinquishing the sparsity constraint altogether. In particular, we show that if each of the $\boldsymbol{x}_n$ lies in a known random subspace then it is possible to recover all of the $\boldsymbol{x}_n$, and $\boldsymbol{h}$ using nuclear-norm minimization whenever $K + S\log^2 S \lesssim L/\log^4(LN)$, and $N \gtrsim \log^2(LN)$. Importantly, we only require $\boldsymbol{h}$ to be sparse, i.e., the support of $\boldsymbol{h}$ is not fixed a priori, which makes it a first result of this form to the best of our knowledge.

We motivate the problem by discussing its application in wireless communication over a multipath channel. Both the delays and fading coefficients of the channel are unknown. The encoder codes the messages randomly and transmits them one after the other over the channel and the decoder is able to discover all of the messages and the channel impulse response.

I. Introduction

We consider the problem of recovering length-L vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$, and $\boldsymbol{h}$ from their circular convolutions

$$\boldsymbol{y}_n = \boldsymbol{h} * \boldsymbol{x}_n, \quad n = 1, \dots, N.$$

This problem referred to as the blind deconvolution and is one of the core problems in system theory, signal processing, and communications. The problem is very ill-posed, and it is easy to see this in the Fourier domain

$$\boldsymbol{F}\boldsymbol{y}_n = \sqrt{L}(\boldsymbol{F}\boldsymbol{x}_n \odot \boldsymbol{F}\boldsymbol{h}), \quad n = 1, \dots, N,$$

where $\boldsymbol{F}$ is an $L \times L$ normalized DFT matrix. As is evident, the blind deconvolution turns into a question of recovering two vectors by observing only their Hadamard product. We show that the ill-posed nature of the problem can be averted under a very general set of structural assumptions. We assume that each of the input vector $\boldsymbol{x}_n$ resides in a known subspace, i.e.,

$$\boldsymbol{x}_n = \boldsymbol{C}_n\boldsymbol{m}_n, \ n = 1, \dots, N$$

for some known $L \times K$ matrix $\boldsymbol{C}_n$, and an $\boldsymbol{m}_n \in \mathbb{R}^K$. Furthermore, we assume that the vector $\boldsymbol{h}$ is S-sparse. This means that the support of $\boldsymbol{h}$ is not known a priori. Note that the earlier work [1] requires the support of $\boldsymbol{h}$ to be known as well. The main goal of this work is to extend the results to the case when the support of $\boldsymbol{h}$ is unknown.

Let $\boldsymbol{f}_\ell^*$ denote the ℓth row of $\boldsymbol{F}$, and $\bar{\boldsymbol{c}}_{\ell,n} = (\sqrt{L}\boldsymbol{C}_n^*\boldsymbol{f}_\ell)$. The ℓth entry of $\boldsymbol{F}\boldsymbol{y}_n$ can now be expressed as

$$(\boldsymbol{F}\boldsymbol{y}_n)_\ell = \left\langle \boldsymbol{f}_\ell\boldsymbol{c}_{\ell,n}^*, \boldsymbol{h}\boldsymbol{m}_n^* \right\rangle_F, \quad \ell = 1, \dots, L, \text{ and } n = 1, \dots, N,$$

where $\boldsymbol{c}_{\ell,n}$ is the entry-wise conjugate of the column vector $\bar{\boldsymbol{c}}_{\ell,n}$ defined above, and $\langle \cdot, \cdot \rangle_F$ is the conventional trace inner product. Thus the entries of a nth convolution are just the trace inner product of a known $L \times K$ measurement matrix $\boldsymbol{f}_\ell\boldsymbol{c}_{\ell,n}^*$, and an unknown rank-1 $L \times K$ matrix $\boldsymbol{h}\boldsymbol{m}_n^*$ that varies with n. Let $\boldsymbol{m} = [\boldsymbol{m}_1^*, \dots, \boldsymbol{m}_N^*]^*$, and $\boldsymbol{\phi}_{\ell,n} = \boldsymbol{c}_{\ell,n} \otimes \boldsymbol{e}_n$, where $\otimes$ is a standard Kronecker product. Then we can finally write

$$(\boldsymbol{F}\boldsymbol{y}_n)_\ell = \left\langle \boldsymbol{f}_\ell\boldsymbol{\phi}_{\ell,n}^*, \boldsymbol{h}\boldsymbol{m}^* \right\rangle_F, \quad \ell = 1, \dots, L, \text{ and } n = 1, \dots, N.$$

We want to find $\boldsymbol{X} = \boldsymbol{h}\boldsymbol{m}^*$ given LN linear measurements. The matrix $\boldsymbol{X}$ is row-sparse and of rank one by construction implying that the actual number of unknowns are only $\sim S\log L + KN$. The sparse and rank one constraints can be efficiently relaxed individually using the ℓ_1 norm and the nuclear norm. However, there is provably no effective convex relaxation for the joint sparse, and rank one structure [2]. Fortunately, we can drop the sparsity constraint altogether here and may only impose a rank one constrain if we are willing to have multiple inputs ($N > 1$). To see this, note that when the rank-one constraint is imposed, the number of unknowns is $L + KN$ and $LN > L + KN$ as soon as $N \geq L/L - K$.

The main theme of this study is to use multiple unknown inputs $\{\boldsymbol{x}_n\}_{n=1}^N$ with only known subspaces to resolve $\boldsymbol{h}$ completely without knowing its support a priori. Our main result shows that if we take $\boldsymbol{C}_n$ to be standard Gaussian random matrices and $\boldsymbol{h}$ is flat in teh Fourier domain then the nuclear norm minimization recovers the $\boldsymbol{X}$ exactly whenever $L/\log^4(LN) \gtrsim K + S\log^2 S$, and when $N \gtrsim \log^2(LN)$.

An natural application arises in the context multipath wireless communications, where one wants to transmit a series of messages $\boldsymbol{m}_1, \dots, \boldsymbol{m}_N$ over an unknown multipath channel, which can be modeled by a sparse $\boldsymbol{h}$. The encoder then randomly codes each of the messages by multiplying it with a tall random matrix $\boldsymbol{C}_n$ and the coded message is transmitted over the multipath channel. The decoder observes the convolutions $\boldsymbol{C}_n\boldsymbol{m}_n * \boldsymbol{h}, \quad n = 1, \dots, N$ and discovers the messages and channel response using nuclear norm minimization. This shows that the coding framework in wireless communication which has been traditionally thought to be as error protection mechanism can also be very useful in channel equalization!

References

[1] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *IEEE Trans. Inform. Theory*, vol. 60, no. 3, pp. 1711–1732, 2014.

[2] A. Aghasi, S. Bahmani, and J. Romberg, "A tightest convex envelope heuristic to row sparse and rank one matrices." in *GlobalSIP*, 2013, p. 627.

A Proximal Approach for Solving Matrix Optimization Problems Involving a Bregman Divergence

Alessandro Benfenati*, Emilie Chouzenoux*†, and Jean–Christophe Pesquet†,
* LIGM, University Paris–Est Marne–la–Vallée
† Center for Visual Computing, CentraleSupelec, University Paris-Saclay

***Abstract*—In recent years, there has been a growing interest in problems such as shape classification, gene expression inference, inverse covariance estimation. Problems of this kind have a common underlining mathematical model, which involves the minimization in a matrix space of a Bregman divergence function coupled with a linear term and a regularization term. We present an application of the Douglas-Rachford algorithm which allows to easily solve the optimization problem.**

In recent years, some applications such as shape classification models [1], gene expression [2], or inverse covariance estimation [3] have led to matrix variational formulations of the form:

$$\underset{C\in\mathcal{S}_+}{\text{minimize}}\ \ D_f(C,S)+g(C) \tag{1}$$

where $\mathcal{S}_+$ is the cone of symmetric semidefinite positive matrices of size $n\times n$, S is a given matrix in $\mathcal{S}_+$, f and g are proper lower-semicontinuous (lsc) convex functions defined on the space of $n\times n$ matrices, and D_f is the Bregman divergence associated with f. Recall that

$$D_f(C,S)=f(C)-f(S)-\operatorname{tr}\left(T(C-S)\right) \tag{2}$$

where $T\in\partial f(S)\neq\varnothing$. Note also that solving (1) amounts to computing the proximity operator of $g+\iota_{\mathcal{S}_+}$ at S,[1] with respect to the divergence D_f, which has also been found to be useful in a number of recent works [4], [5].

Very often, due to the nature of the problems, the regularization functional g has to promote the sparsity of C. A generic class of regularization is obtained by assuming that $g=g_0+g_1$ where

$$g_0(C)=\begin{cases}\psi(d) & \text{if } C\in\mathcal{S}_+\\ +\infty & \text{otherwise,}\end{cases} \tag{3}$$

where $\psi\colon\mathbb{R}^n\to]-\infty,+\infty]$ is a proper lsc function and d is the vector of eigenvalues of C, whereas g_1 is a function which cannot be expressed under this form. Typical examples are the nuclear norm $\|\cdot\|_*$ (or any Schatten norm) for g_0 and the ℓ_1 norm $\|\cdot\|_1$ (of the matrix elements) for g_1 [6].

In this paper, we will assume that function f can be expressed similarly to g_0 as $f(C)=\varphi(d)$ if $C\in\mathcal{S}_+$, $f(C)+\infty$ otherwise, where $\varphi\colon\mathbb{R}^n\to]-\infty,+\infty]$ is a proper lsc convex function. In particular, this assumption is satisfied when

$$f(C)=\begin{cases}-\log\det(C) & \text{if } C\succ 0\\ +\infty & \text{otherwise.}\end{cases} \tag{4}$$

Various algorithm have been proposed to solve Problem (1) when f is the above function and some specific choices of the function g are made: the popular GLASSO algorithm [3], a Gradient Projection method [1], and a splitting technique on the regularization term [6]. Here we propose to employ the Douglas–Rachford algorithm [7], which enables us to solve (1) in a fast manner, as soon as an efficient procedure for the eigenvalue decomposition is provided. The Douglas–Rachford approach alternates proximity steps on $D_f(\cdot,S)+g_0+\iota_{\mathcal{S}_+}$ and on g_1. For many functions g_1 of practical interest, the proximity operator of g_1 (e.g, $g_1=\|\cdot\|_1$) has a closed form solution [7] . Let us define $F(C)=f(C)+g_0(C)$. Let $\gamma\in]0,+\infty[$. It can be noted that computing the proximity operator of $\gamma\big(D_f(\cdot,S)+g_0+\iota_{\mathcal{S}_+}\big)$ w.r.t. the Frobenius metric $\|\cdot\|_{\mathrm{F}}$, at some symmetric matrix $\overline{C}$, is equivalent to find

$$\widehat{C}=\underset{C\in\mathbb{R}^{n\times n}}{\operatorname{argmin}}\ \left(F(C)-\operatorname{tr}\left(TC\right)+\frac{1}{2\gamma}\|C-\overline{C}\|_{\mathrm{F}}^2\right).$$

Classical properties of the proximity operator [7] state that

$$\widehat{C}=\operatorname{prox}_{\gamma F-\gamma\operatorname{tr}(T\cdot)}(\overline{C})=\operatorname{prox}_{\gamma F}(\overline{C}+\gamma T).$$

Moreover, if $\overline{C}+\gamma T=U\operatorname{Diag}(\sigma)U^\top$ where U is an orthogonal matrix and $\sigma\in\mathbb{R}^n$, then $\widehat{C}=UDU^\top$ with $D=\operatorname{Diag}(\operatorname{prox}_{\gamma(\varphi+\psi)}(\sigma))$. For example, if f is the log-det function (4) and $g_0=\mu\|\cdot\|_*$ where $\mu\in[0,+\infty[$, according to [8], the diagonal matrix of eigenvalues of $\widehat{C}$ is given by

$$D=\frac{1}{2}\left(\Sigma-\gamma\mu I_n+\sqrt{(\Sigma-\gamma\mu I_n)^2+4\gamma I_n}\right)$$

where $\Sigma=\operatorname{Diag}(\sigma)$. The operations to compute $\operatorname{prox}_{\gamma(\varphi+\psi)}$ are thus component–wise.

The proposed Douglas–Rachford approach is easy to implement: if an efficient procedure for the eigenvalue decomposition is available, according to our numerical experiments, it is also very fast.

Acknowledgements: This work was supported by the Agence Nationale de la Recherche under grant ANR-14-CE27-0001 GRAPHSIP.

[1] ι_E designates the indicator function of a set E.

References

[1] J. Duchi, S. Gould, and D. .Koller, "Projected subgradient methods for learning sparse gaussians," in *Proceedings of the Twenty-fourth Conference on Uncertainty in AI (UAI)*, 2008.

[2] S. Ma, L. Xue, and H. .Zou, "Alternating direction methods for latent variable gaussian graphical model selection." *Neural Computation*, vol. 25, no. 8, pp. 2172–2198, 2013.

[3] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, jul 2008.

[4] H. H. Bauschke, P. L. Combettes, and D. Noll, "Joint minimization with alternating bregman proximity operators," *Pacific Journal of Optimization*, vol. 2, no. 3, pp. 401–424, 2006.

[5] A. Benfenati and V. Ruggiero, "Inexact Bregman iteration with an application to Poisson data reconstruction," *Inverse Problems*, vol. 29, no. 6, pp. 1–32, 2013.

[6] V. Chandrasekaran, P. Parrilo, and A. S.Willsky, "Latent variable graphical model selection via convex optimization," *Ann. Statist.*, vol. 40, no. 4, pp. 1935–1967, 08 2012.

[7] P. Combettes and J.-C. Pesquet, "Proximal Splitting Methods in Signal Processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, B. B. C. E. L. W. (Eds.), Ed. Springer, 2011, pp. 185–212.

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

A Block Parallel Majorize-Minimize Memory Gradient Algorithm

Sara Cadoni*, Emilie Chouzenoux*, Jean-Christophe Pesquet* and Caroline Chaux†
* Université Paris-Est Marne-la-Vallée, LIGM and UMR-CNRS 8049, Marne-la-Vallée, France.
† Aix Marseille Université, CNRS, Centrale Marseille, I2M and UMR-CNRS 7373, Marseille, France

***Abstract*—In the field of 3D image recovery, huge amounts of data need to be processed. Parallel optimization methods are then of main interest since they allow to overcome memory limitation issues, while benefiting from the intrinsic acceleration provided by recent multicore computing architectures. In this context, we propose a Block Parallel Majorize-Minimize Memory Gradient (BP3MG) algorithm for solving large scale optimization problems. This algorithm combines a block coordinate strategy with an efficient parallel update. The proposed method is applied to a 3D microscopy image restoration problem involving a depth-variant blur, where it is shown to lead to significant computational time savings with respect to a sequential approach.**

I. Introduction

In many inverse problems encountered in image processing, one has to generate an image estimate $\hat{\boldsymbol{x}} \in \mathbb{R}^N$ by minimizing an appropriate cost function F, which has the following composite form:

$$(\forall \boldsymbol{x} \in \mathbb{R}^N) \quad F(\boldsymbol{x}) = \sum_{s=1}^{S} f_s(\boldsymbol{L}_s \boldsymbol{x})$$

where, for every $s \in \{1, \ldots, S\}$, $\boldsymbol{L}_s \in \mathbb{R}^{P_s \times N}$, $P_s \in \mathbb{N}^*$, and f_s is a function from $\mathbb{R}^{P_s}$ to $\mathbb{R}$. In the case of large scale image recovery problems, a major challenge is to design an optimization algorithm able to deliver reliable numerical solutions in a reasonable time.

When all the involved functions $(f_s)_{1 \leqslant s \leqslant S}$ are differentiable on $\mathbb{R}^N$ (but non necessarily convex), a very efficient strategy is the Majorize-Minimize Memory Gradient (3MG) algorithm [1]. It relies on a Majorize-Minimize (MM) approach, combined with a subspace acceleration technique. The 3MG algorithm enjoys nice convergence properties in both convex and non-convex cases and comparisons with state-of-the-art optimization methods on a number of image restoration problems have shown its good performance in terms of practical convergence speed [1], [2]. However, when the size of the problem becomes increasingly large, as it may happen in 3D image processing or video processing, running this kind of algorithm becomes difficult, due to memory limitation issues.

II. Proposed method

The MM approach relies on the existence of symmetric positive matrices

$$(\forall \boldsymbol{x} \in \mathbb{R}^N) \quad \boldsymbol{A}(\boldsymbol{x}) = \sum_{s=1}^{S} \boldsymbol{L}_s^\top \operatorname{Diag}\{\boldsymbol{\omega}_s(\boldsymbol{L}_s \boldsymbol{x})\} \boldsymbol{L}_s,$$

with for every $s \in \{1, \ldots, S\}$, $\boldsymbol{\omega}_s : \mathbb{R}^{P_s} \to]0, +\infty[^{P_s}$, such that, for every $(\boldsymbol{x}, \boldsymbol{x}') \in (\mathbb{R}^N)^2$, the following majoration holds:

$$F(\boldsymbol{x}) \leqslant F(\boldsymbol{x}') + \nabla F(\boldsymbol{x}')^\top (\boldsymbol{x} - \boldsymbol{x}') + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}')^\top \boldsymbol{A}(\boldsymbol{x}')(\boldsymbol{x} - \boldsymbol{x}').$$

In the 3MG algorithm, a new iterate results from the minimization of the latter quadratic majorant within a two-dimensional subspace spanned by the current gradient and the previous direction. In order to overcome difficulties related to memory requirements, we propose to combine 3MG with a parallel block alternating strategy. The target vector $\boldsymbol{x}$ is split into J non-overlapping block vectors $\boldsymbol{x}^{(j)}$ of reduced dimension $N_j \neq 0$. At each iteration, only a subset $\mathcal{S} \subset \{1, \ldots, J\}$ of them is selected, and the associated entries $\boldsymbol{x}^{(\mathcal{S})} = (x_p)_{p \in \mathbb{S}}$ of $\boldsymbol{x}$ are updated. To this end, a clever use of Jensen's inequality allows us to show that, for every $\mathcal{S} \subset \{1, \ldots, J\}$,

$$(\forall \boldsymbol{x} \in \mathbb{R}^N) \quad \boldsymbol{A}^{(\mathcal{S})}(\boldsymbol{x}) \preceq \boldsymbol{B}^{(\mathcal{S})}(\boldsymbol{x}) = \operatorname{BDiag}\left\{ \left(\boldsymbol{B}^{(j)}(\boldsymbol{x})\right)_{j \in \mathcal{S}} \right\},$$

where, for every $j \in \mathcal{S}$,

$$(\forall \boldsymbol{x} \in \mathbb{R}^N) \quad \boldsymbol{B}^{(j)}(\boldsymbol{x}) = \sum_{s=1}^{S} \left((\boldsymbol{L}_s^{(j)})^\top \operatorname{Diag}\{\boldsymbol{b}_s(\boldsymbol{L}_s \boldsymbol{x})\} \boldsymbol{L}_s^{(j)} \right),$$

with, for every $s \in \{1, \ldots, S\}$ and $p \in \{1, \ldots, P_s\}$,

$$(\forall \boldsymbol{x} \in \mathbb{R}^N) \quad [\boldsymbol{b}_s(\boldsymbol{L}_s \boldsymbol{x})]_p = [\boldsymbol{\omega}_s(\boldsymbol{L}_s \boldsymbol{x})]_p [|\boldsymbol{L}_s^{(\mathcal{S})}| \mathbf{1}_{|\mathbb{S}|}]_p / [|\boldsymbol{L}_s^{(j)}| \mathbf{1}_{N_j}]_p.$$

Thanks to the block-diagonal structure of the majorant matrix, the selected blocks with indices $j \in \mathcal{S}$ can be updated in a parallel manner according to a 3MG scheme, leading to the so-called block-parallel 3MG algorithm. The monotonic convergence of the criterion sequence $(F(\boldsymbol{x}_k))_{k \in \mathbb{N}}$ to a (locally) optimal value is established, using the same theoretical tools as in [3].

III. Application to 3D microscopy

The proposed algorithm is applied for solving a 3D image restoration problem with depth-variant blur. Figure 1 illustrates its high efficiency in terms of acceleration for multi-core architectures.

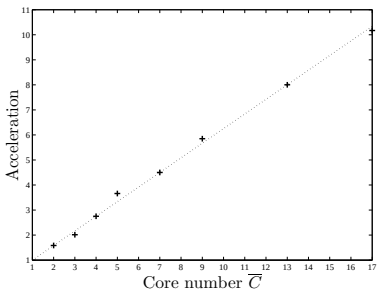


Fig. 1. Ratio between the computation time for one core and the computation time for $\overline{C}$ slave cores (crosses) with linear fitting (dotted line), for the restoration of a 3D microscopy image with size $N = 256 \times 256 \times 48$ pixels.

References

[1] E. Chouzenoux, J. Idier, and S. Moussaoui, "A Majorize-Minimize strategy for subspace optimization applied to image restoration," *IEEE Trans. Image Process.*, vol. 20, no. 18, pp. 1517–1528, Jun. 2011.

[2] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, "A majorize-minimize subspace approach for ℓ_2-ℓ_0 image regularization," *SIAM J. Imag. Sci.*, vol. 6, no. 1, pp. 563–591, 2013.

[3] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "A block coordinate variable metric forward-backward algorithm," *To appear in J. Global Optim.*, 2015, http://www.optimization-online.org/DB_HTML/2013/12/4178.html.

Fast and effective image restoration with trainable nonlinear reaction diffusion

Yunjin Chen*, Wensen Feng†, and Thomas Pock*
* Institute of Computer Graphics and Vision, Graz University of Technology, Graz, Austria
†School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China.

***Abstract*—We describe a flexible learning framework based on the concept of nonlinear reaction diffusion models for various image restoration problems. We propose a dynamic nonlinear reaction diffusion model with time-dependent parameters (i.e., linear filters and influence functions). In contrast to previous nonlinear diffusion models, all the parameters, including the filters and the influence functions, are simultaneously learned from training data. We call this approach TNRD (Trainable Nonlinear Reaction Diffusion). The TNRD approach is applicable for a variety of image restoration tasks by incorporating appropriate reaction force. We demonstrate its capabilities with several representative applications. Experiments show that our trained nonlinear diffusion models lead to state-of-the-art performance for the tested applications. Our trained models preserve the structural simplicity of diffusion models and only take a small number of diffusion steps, thus are highly efficient. Moreover, they are also well- suited for parallel computation on GPUs, which makes the inference procedure extremely fast.**

I. Introduction

Among the approaches to tackle the problem of image restoration, nonlinear diffusion [1] defines a class of efficient approaches. In this work, we start with a conventional nonlinear diffusion model (P-M model [1]), and extend it to a trainable framework with a few highly parametrized linear filters as well as influence function, which are learned from training data. Our proposed nonlinear diffusion process has several remarkable benefits as follows: *(1) It is conceptually simple as it is merely a standard nonlinear diffusion model with trained filters and influence functions; (2) It has broad applicability to a variety of image restoration problems; (3) It yields excellent results for many tasks in image restoration, including image denoising with Gaussian, Poisson or speckle noise, single image super resolution/interpolation, compression artifacts reduction and non-blind/blind image deblurring; (4) It is highly computationally efficient, and well suited for parallel computation on GPUs.*

II. Proposed nonlinear diffusion process

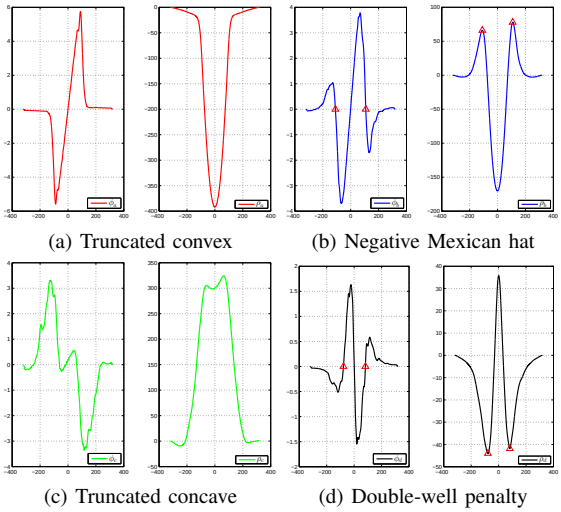
The discrete P-M model is reformulated as the following discrete PDE with an explicit finite difference scheme

$$\frac{u_{t+1}-u_t}{\Delta t} = -\sum_{i=\{x,y\}} \nabla_i^\top \Lambda(u_t)\nabla_i u_t \doteq -\sum_{i=\{x,y\}} \nabla_i^\top \phi(\nabla_i u_t)\,. \tag{1}$$

Model (1) can be improved from the aspects: (a) more filters of larger kernel size; (b) more flexible influence functions, instead of hand-crafted ones with fixed shape; (c) truncated gradient descent procedure with fixed iterations/stages to accelerate the inference phase; (d) varying parameters per stage. Therefore, we arrive at a general nonlinear reaction diffusion model given as

$$u_t = \text{Prox}_{\mathcal{G}^t}\left(u_{t-1} - \left(\sum_{i=1}^{N_k} \bar{k}_i^t * \phi_i^t(k_i^t * u_{t-1}) + \psi^t(u_{t-1}, f)\right)\right), \tag{2}$$

where $*$ is 2D convolution, filters k_i^t and influence functions ϕ_i^t vary across stages and are trained from data. $\text{Prox}_{\mathcal{G}^t}(\hat{u})$ is the proximal mapping operation related to the convex function $\mathcal{G}^t$, see [2] for more details.



(a) Truncated convex (b) Negative Mexican hat

(c) Truncated concave (d) Double-well penalty

Fig. 1. The figure shows four characteristic influence functions (left plot in each subfigure) together with their corresponding penalty functions (right plot in each subfigure), learned by our proposed method in the $\text{TNRD}_{5\times5}^{5}$ model.

III. Overall training model

As our goal is to train a diffusion network with T stages, the training task is formulated as the optimization problem:

$$\begin{cases} \min\limits_{\Theta} \mathcal{L}(\Theta) = \sum\limits_{s=1}^{S} \frac{1}{2}\|u_T^s - u_{gt}^s\|_2^2 \\ \text{s.t.} \begin{cases} u_0^s = I_0^s, \quad t = 1\cdots T \\ u_t^s = \text{Prox}_{\mathcal{G}^t}\left(u_{t-1}^s - \left(\sum\limits_{i=1}^{N_k} (K_i^t)^\top \phi_i^t(K_i^t u_{t-1}^s) + \psi^t(u_{t-1}^s, f^s)\right)\right) \end{cases} \end{cases}. \tag{3}$$

In this work, we further improve our proposed trainable diffusion model by investigating multi-scale image analysis technology and nonlocal similarity information. The resulting models can lead to significantly better restoration performance for a few applications.

IV. Important findings

A major finding in this paper is that our learned penalty functions significantly differ from the usual penalty functions adopted in partial differential equations and energy minimization methods. In contrast to their usual robust smoothing properties which is caused by a single minimum around zero, most of our learned functions have multiple minima different from zero and hence are able to enhance certain image structures. Four representative shapes are shown in Fig. 1.

References

[1] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[2] Y. Chen and T. Pock, "Trainable Nonlinear Reaction Diffusion: A flexible framework for fast and effective image restoration," *To appear in IEEE TPAMI*, 2016.

Proximal splitting methods on convex problems with a quadratic term: Relax!

Laurent Condat*,
* GIPSA-lab, Univ. Grenoble–Alpes, F-38000 Grenoble, France. Contact: see http://www.gipsa-lab.fr/∼laurent.condat/

***Abstract*—Proximal splitting algorithms are appropriate for the large-scale convex optimization problems encountered in signal and image processing. They generally contain a relaxation parameter, which is easy to tune: the larger, the better for the convergence speed. We show that when a quadratic term is present in the optimization problem to solve, this parameter can be chosen larger than what is commonly known.**

I. Introduction

A large number of problems in signal and image processing [1], [2], computer vision, data mining, and many other fields, can be formulated as large-scale convex optimization problems. They typically involve several terms, like data fidelity terms, regularization or structural penalties, and constraints. Thus, in all generality, one wants to

$$\text{Find } \hat{x} \in \arg\min_{x\in\mathcal{X}} \sum_{m=1}^{M} f_m(L_m x), \tag{1}$$

for linear operators $L_m : \mathcal{X} \to \mathcal{U}_m$, where $\mathcal{X}$ and $\mathcal{U}_m$ are real Hilbert spaces of large dimension, and convex, proper, lower semicontinuous [3] functions $f_m : \mathcal{U}_m \to \mathbb{R} \cup \{+\infty\}$, which may be smooth or not. Nonsmooth penalties are beneficial to constrain the solution to be parsimonious in some sense. We refer the reader unfamiliar with convex optimization to the book [3] or to the tutorial papers [1], [2].

To solve such large-scale convex problems, first order proximal splitting algorithms are particularly appropriate [1]–[4]. They proceed by calling individually, at every iteration, the gradient ∇f_m or the proximal operator prox_{f_m} [1]–[4] of each function, and the linear operators L_m or their adjoints. The development of splitting algorithms has become a very active topic in the last years, driven by the need to solve highly demanding problems, e.g. reconstruction of 3D volumes in microscopy, astronomy, or medical imaging; see e.g. the recent papers [5]–[7] and references therein.

In this work, we focus on the case where one of the function in (1) is quadratic: $f_1(x) = \frac{1}{2}\langle x, Qx\rangle + \langle x, b\rangle$ for some positive self-adjoint bounded linear operator Q and element $b \in \mathcal{X}$. One typical example is a least-squares term $\frac{1}{2}\|Ax - y\|^2$, which corresponds to $Q = A^*A$ and $b = -A^*y$. Many problems involve a least-squares term, like inverse imaging problems in presence of Gaussian noise, the LASSO estimation of a sparse vector, the unmixing problem in multispectral imaging. It is natural to view a quadratic term as a smooth function with a $\|Q\|$-Lipschitz-continuous gradient $\nabla f_1(x) = Qx + b$. Consequently, a splitting algorithm based of the forward-backward kind [6] will be used to solve (1). We advocate an alternative, which is to change the metric in the proximity operator.

Let us look at the simple case of minimizing a quadratic function plus another proximable function:

$$\text{Find } \hat{x} \in \arg\min_{x\in\mathcal{X}} \tfrac{1}{2}\langle x, Qx\rangle + \langle x, b\rangle + f_2(x). \tag{2}$$

The relaxed forward-backward iteration is

$$\left\lfloor \begin{array}{l} \tilde{x}^{(i+1)} := \text{prox}_{\gamma f_2}\big(x^{(i)} - \gamma(Qx^{(i)} + b)\big) \\ x^{(i+1)} := \rho\,\tilde{x}^{(i+1)} + (1-\rho)x^{(i)} \end{array} \right. .$$

The second step, which is a simple linear combination of the new iterate and the previous one, is called a **relaxation**. Convergence is guaranteed if $\gamma \in (0, 2/\|Q\|)$ and the relaxation parameter $\rho \in (0, 2 - \gamma\|Q\|/2)$. It is generally observed in practice that, for a given parameter γ, the **larger** ρ, the **faster** the convergence. Since relaxation is very **cheap**, it is a pity that this fact is largely unknown.

Now, let us consider, instead, the relaxed proximal point algorithm, which iterates the proximity operator of $f_1 + f_2$, but defined with a different inner product $\langle\cdot,\cdot\rangle_P = \langle\cdot, P\cdot\rangle$, for a strongly positive, self-adjoint, bounded linear operator P. The iteration is

$$\left\lfloor \begin{array}{l} \tilde{x}^{(i+1)} := \arg\min_{x\in\mathcal{X}} \gamma f_1(x) + \gamma f_2(x) + \frac{1}{2}\|x - x^{(i)}\|_P^2 \\ x^{(i+1)} := \rho\,\tilde{x}^{(i+1)} + (1-\rho)x^{(i)} \end{array} \right. .$$

If $\gamma \in (0, 1/\|Q\|)$ and we choose $P = \text{Id} - \gamma Q$, the subproblem becomes

$$\begin{aligned} \tilde{x}^{(i+1)} &:= \arg\min_{x\in\mathcal{X}} \gamma f_2(x) + \gamma\langle x, b\rangle + \tfrac{1}{2}\|x - x^{(i)}\|^2 + \gamma\langle x, Qx^{(i)}\rangle \\ &= \text{prox}_{\gamma f_2}(x^{(i)} - \gamma Qx^{(i)} - \gamma b). \end{aligned} \tag{3}$$

So, both algorithms are the same! But now, convergence is guaranteed for $\rho \in (0, 2)$, which is better. More generally, a continuum between these two regimes can be derived. These better bounds seem to be specific to the quadratic case. Note that the idea of preconditioning, or linearizing, to cancel the quadratic term in a minimization subproblem, is not new, see e.g. [8]. In the poster, we will extend this principle and show new parameter intervals, for well known algorithms like the Douglas–Rachford or Chambolle–Pock algorithms. We will show with practical examples that a larger relaxation parameter yields faster convergence.

References

[1] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. New York: Springer-Verlag, 2010.

[2] A. Chambolle and T. Pock, "An introduction to continuous optimization for imaging," *Acta Numerica*, vol. 25, pp. 161–319, 2016.

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

[4] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 3, no. 1, pp. 127–239, 2014.

[5] L. Condat, "A generic proximal algorithm for convex optimization—Application to total variation minimization," *IEEE Signal Process. Lett.*, vol. 21, no. 8, pp. 1054–1057, Aug. 2014.

[6] P. L. Combettes, L. Condat, J.-C. Pesquet, and B. C. Vũ, "A forward–backward view of some primal–dual optimization methods in image recovery," in *Proc. of IEEE ICIP*, Paris, France, Oct. 2014.

[7] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal–dual approaches for solving large-scale optimization problems," *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 31–54, Nov. 2015.

[8] K. Bredies and H. Sun, "Preconditioned Douglas–Rachford splitting methods for convex-concave saddle-point problems," *SIAM Journal on Numerical Analysis*, vol. 43, no. 1, pp. 421–444, 2015.

SURE-ly better reconstructions using AMP

Mike Davies* and Alessandro Perelli*
* Institute for Digital Communications (IDCOM), The University of Edinburgh, EH9 3JL, UK

***Abstract*—We use Stein's Unbiased Risk Estimate (SURE) to substantially enhance AMP methods through a denoising perspective, enabling parameter optimisation on-the-fly without the explicit need for a full generative model. We also show these methods can be very effective in an imaging setting that does not fit the rigorous compressed sensing framework, suggesting that the AMP framework is more widely applicable than was first thought.**

I. Introduction

Compressed sensing (CS) aims to retrieve a sparse or compressible signal $\mathbf{x}$ from an appropriate set of measurements $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$, with a sampling ratio below the Nyquist rate. Approximate message passing solutions stand out as not only being very computationally efficient but also having performance that is accurately quantifiable through the State Evolution (SE) equations. Intuitively at each iteration the AMP estimate, $\hat{\mathbf{x}}$ can be viewed as a noisy version of the original signal: $\hat{\mathbf{x}}^t = \mathbf{x} + \mathbf{A}^\mathbf{T}\mathbf{w} + \mathbf{w}^\mathbf{t}_{\text{eff}}$ where $\mathbf{w}^\mathbf{t}_{\text{eff}}$ is the effective noise induced by signal interference and retrieving the signal amounts to successive denoising steps. Here we leverage this idea to use different denoising functions to provide an efficiently estimate of the signal at each iteration.

II. Parametric SURE-AMP

Parametric SURE AMP [1] extends the generic AMP by using an adaptive signal denoising module, $f_t(\mathbf{r}^t, c^t|\theta^t)$. Our goal is to select from this flexible family of denoisers a θ that minimizes the error of the estimate at each iteration. The implementation of parametric SURE-AMP is summarized in Algorithm 1 where $\gamma = m/n$ is the sampling ratio. The key differences from the original AMP are highlighted in red.

Algorithm 1 Parametric SURE-AMP

]

initialization: $\hat{\mathbf{x}}^0 = \mathbf{0}, \mathbf{z}^\mathbf{0} = \mathbf{0}, \mathbf{c}^\mathbf{0} = \frac{1}{\mathbf{m}}\|\mathbf{z}\|_\mathbf{2}^\mathbf{2}$
for $t = 0, 1, 2, 3, \ldots$ **do**
$\mathbf{r}^t = \hat{\mathbf{x}}^t + \mathbf{A}^\mathbf{T}\mathbf{z}^\mathbf{t}$,
$\theta^t = H_t(\mathbf{r}^t, c^t)$, // parameter selection function
$\hat{\mathbf{x}}^{t+1} = f_t(\mathbf{r}^t, c^t|\theta^t)$, // parametric denoiser
$\mathbf{z}^{t+1} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{\mathbf{t+1}} + \frac{\mathbf{1}}{\gamma}\langle \mathbf{f}'_\mathbf{t}(\mathbf{r}^\mathbf{t}, \mathbf{c}^\mathbf{t}|\theta^\mathbf{t})\rangle \mathbf{z}^\mathbf{t}$ // Onsager term
$c^{t+1} = \frac{1}{m}\|\mathbf{z}^{t+1}\|_2^2$
end for

Although we do not have access to $\mathbf{x}_0$ we can still estimate the MSE using SURE. Furthermore, since SURE becomes more accurate as more data is available it is particularly appropriate for the large system limit setting of AMP.

Given a noisy scalar signal $r = x_0 + w$ with $\text{var}(w) = \sigma^2$ and the scalar denoising function in the form: $f(r, \sigma^2|\theta) := r + g(r, \sigma^2|\theta)$ we have:

$$\mathbb{E}_{\hat{x},x_0}\{(\hat{x} - x_0)^2\} = \sigma^2 + \mathbb{E}_r\{g^2(r, \sigma^2|\theta) + 2\sigma^2 g(r, \sigma^2|\theta)\}.$$

Thus we use the SURE to define:

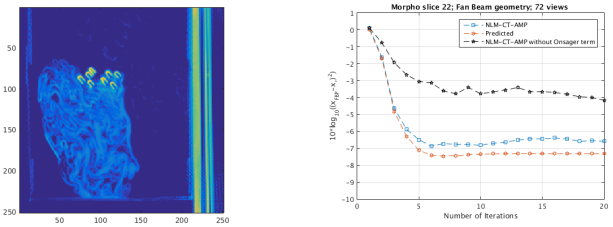
$$H_t(\mathbf{r}^t, c^t) = \underset{\theta}{\text{argmin}}\langle g^2(r, \sigma^2|\theta) + 2\sigma^2 g(r, \sigma^2|\theta)\rangle$$

Furthermore, if $f(\cdot|\theta)$ is a linear function of θ then the optimisation can to computed in closed form, making this extremely quick.

Experimentally we have found that using simple families of denoisers can achieve essentially Bayes optimal performance for Bernoulli-Gaussian data without prior information and is 20 times faster than empirical Bayesian EM-GM-GAMP approach [1].

III. Denoising Message Passing for CT

Independently in [2] a similar denoising perspective was investigated to incorporate very general denoisers such as Nonlocal Means (NLM) into the AMP framework. For such denoisers optimising parameters and calculating the Onsager correction do not have a closed form. Instead [2] proposed using a simple Monte Carlo estimate from [4]. In [3] we have subsequently applied these ideas to computational tomography (CT) to test whether such algorithms are applicable outside of the 'comfort zone' of Gaussian random measurements. This required dealing with the ill-conditioned (and deterministic) measurement operators and the Poisson noise model characteristic of CT systems. The former was tackled by incorporating cone filter preconditioning while for the latter problem we were able to directly absorb the Poisson model into the algorithm using a Generalized AMP formulation - something that cannot be done in classical Penalized Weighted Least Squares solutions - enabling us to separate the effects of system conditioning and the noise model. The results shown in figure 1 show that the NLM-CT-AMP solution provides excellent performance and better than state-of-the-art. Furthermore, as can be seen the role of the Onsager term in this improvement appears to be crucial.



(a) (b)

Fig. 1. CT reconstruction of luggage scan obtained using Morpho CTX5500 Air Cargo dual energy system at 100kVp using only 72 views: (a) NLM-CT-AMP with cone filter preconditioner, (b) State Evolution with and without Onsager correction

References

[1] C. Guo and M. E. Davies, Near Optimal Compressed Sensing without priors: Parametric SURE Approximate Message Passing. IEEE Trans. SP vol. 63 (8), pp. 2130–2141, 2015.
[2] C. A. Metzler, A. Maleki, and R. G. Baraniuk, From denoising to compressed sensing. ArXiv preprint arXiv:1406.4175, 2014.
[3] A. Perelli, M. Lexa, A. Can and M. E. Davies, Denoising Message Passing for X-ray Computed Tomography Reconstruction. ArXiv preprint: arXiv:1609.04661, 2016.
[4] S. Ramani and T. Blu and M. Unser, Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms. IEEE Trans. Im. Proc., vol. 17 (9), pp. 1540–1554, 2008.

Divide (Twice) and Conquer: Patch-based Image Restoration using Mixture Models

Mário A. T. Figueiredo
Instituto de Telecomunicações and Instituto Superior Técnico, University of Lisbon, Portugal

ABSTRACT

The use of patches in image processing is clearly an instance of the "divide and conquer" strategy: since it is admittedly too difficult to formulate a global prior/model for an entire image, patch-based approaches process patches thereof, and combine the processed patches to obtain the processed image. The early patch-based methods (namely the seminal *non-local means*–NLM–denoising method) extract patches from the noisy image, then process/denoise them independently (or maybe collaboratively, as in BM3D), and finally return them to their original locations (averaging overlapping pixel estimates).

More recent classes of approaches that build global image models/priors also adopt a divide and conquer paradigm, by forming a function computed from image patches. This class was arguably initiated with the *expected patch log-likelihood* (EPLL), and has been adopted by most of the recent work. Unlike their earlier counterparts, these methods do provide a coherent global statistical image model/prior.

A particular subclass of patch-based statistical image models uses Gaussian mixtures (GM) to model the patches, in what can be seen as a second type application of the divide and conquer principle, now in the space of patch configurations. Different components of the GM specialize in modeling different types of typical patch configurations. Although many other statistical image models exist, using a GM as a patch-prior has several relevant advantages: (i) the corresponding *minimum mean squared error* (MMSE) estimate can be obtained in closed form; (ii) the GM parameters can be estimated from a dataset of clean or noisy patches, using the *expectation-maximization* (EM) algorithm; (iii) theoretically, the class of Gaussian mixture densities can approximate arbitrarily well any probability density (under mild conditions).

In this keynote presentation, I will overview the class of GM-based patch-based approaches to image restoration and reconstruction. After reviewing the first members of this family of methods, which addressed only denoising under Gaussian noise, I will describe more recent advances, namely: denoising under non-Gaussian noise; use of class-adapted GM priors, *i.e.*, tailored to specific image classes (*e.g.*, faces, fingerprints); addressing of problems other than denoising (namely, deblurring, super-resolution, compressive image reconstruction), by plugging GM-based denoisers in the loop of an iterative algorithm (in what has recently been called the plug-and-play approach); joint restoration/segmentation of images.

Blind Deconvolution and Polynomial Factorization

Peter Jung*, Philipp Walk† and Götz E. Pfander‡
*Communications and Information Theory Group, Technische Universitaet Berlin
†Department of Electrical Engineering California Institute of Technology, Pasadena
‡Department of Mathematics and Computer Science, Philipps-Universität Marburg
{*peter.jung@tu-berlin.de, pwalk@caltech.edu, pfander@mathematik.uni-marburg.de*}

***Abstract*—We consider one-dimensional blind deconvolution in the context of sporadic communication of short messages. Blind deconvolution is known to be ill–posed in general and the required univariate case highly suffers from non–trivial ambiguities. Classical blind equalization methods therefore often fail in such applications. Hence, we investigate signal recovery again in the framework of polynomial factorization and discuss algorithmic approaches for the original and the lifted problems. In the Wiener-Hopf setting recovery up to trivial ambiguities is possible using efficient iterative algorithms.**

Recent progress in low–rank matrix recovery have put blind deconvolution as a prototypical bilinear inverse problem back into focus. In the noiseless setting the task is to infer from observations:

$$y_k = (\boldsymbol{h} * \boldsymbol{x})_k = \sum_l h_l x_{k-l}$$

the unknown vectors $\boldsymbol{h}$ and $\boldsymbol{x}$. The multivariate case has been intensively investigated, for example, in image processing. In control theory and communication engineering instead the one-dimensional case is relevant which highly suffers from combinatorial non–trivial ambiguities. The unknown vector $\boldsymbol{h} \in \mathbb{C}^K$ is called channel (impulse response) and $\boldsymbol{x} \in \mathbb{C}^L$ is the transmit signal to recover. With statistical assumptions on $\boldsymbol{x}$ the receiver can estimate from $\boldsymbol{y} \in \mathbb{C}^{L+K-1}$ an inverse operation (equalization) in the regime $L \gg K$ which is known as blind equalization. However, sporadic and short message communication means $K \approx L$.

New results are obtained in [1] for randomized cyclic convolutions $\boldsymbol{y} = \boldsymbol{w} \circledast \boldsymbol{x} \in \mathbb{R}^L$ where $\boldsymbol{w} = \boldsymbol{B}\boldsymbol{h} \in \mathbb{R}^L$ for fixed $\boldsymbol{B}$ and $\boldsymbol{x} = \boldsymbol{C}\boldsymbol{m} \in \mathbb{R}^L$ lies in a random (given by $\boldsymbol{C}$) subspace of dimension N. It has been shown that for $L = \mathcal{O}(N+K)$ the (convex) nuclear norm minimization:

$$\min \|\boldsymbol{X}\|_* \quad \text{s.t.} \quad \mathcal{A}(\boldsymbol{X}) = \boldsymbol{y} \tag{1}$$

recovers $\boldsymbol{h}\boldsymbol{m}^T \in \mathbb{R}^{K\times N}$ with high probability under incoherence assumptions on $\boldsymbol{B}$. Here, the nuclear norm $\|\boldsymbol{X}\|_*$ of a matrix $\boldsymbol{X}$ is the absolute sum of its singular values and $\mathcal{A}$ is the (random) lifted map such that $\mathcal{A}(\boldsymbol{h}\boldsymbol{m}^T) = \boldsymbol{B}\boldsymbol{h} \circledast \boldsymbol{C}\boldsymbol{m}$.

Although (1) renders blind deconvolution tractable in theoretical terms this is (i) challenging complex for communication applications since lifting considerably increases problem size, (ii) it requires common randomness which is often not feasible, (iii) recovery guarantees are probabilistic and can not be strict and (iv) cyclic extensions for $\boldsymbol{x}$ causing additional overhead of $\mathcal{O}(K)$. The computational aspect of the unlifted problem has been tackled recently in [2] with a clever initialization overcoming with high probability the nonconvex nature such that gradient based algorithms will not stuck in local minima.

We will address in this context blind (non–cyclic) deconvolution again in the classical framework of polynomial factorization. Let X and H be the z–transforms of the vectors $\boldsymbol{x}$ and $\boldsymbol{h}$. The z–transform $Y : \mathbb{C} \to \mathbb{C}$ of $\boldsymbol{y} = \boldsymbol{h} * \boldsymbol{x}$:

$$Y(z) = (H \cdot X)(z) = \sum_{k=0}^{L+K-2} y_k z^{-k}$$

is a polynomial in the variable $z^{-1} \in \mathbb{C}$ (same for H and X). Hence, given the observation $\boldsymbol{y}$ and further constraints on $\boldsymbol{h}$ and $\boldsymbol{x}$, the recovery problem is equivalent to find a suitable factorization $Y(z) = H(z)X(z)$ such that H and X belong to particularly constrained classes. Obviously, without sufficiently constraining the unknowns unique factorization is not possible. The non-trivial ambiguities can be characterized in the polynomial description and classifying locations of the zeros could yield uniqueness. Following this line, we also review for the desired application the important minimum-phase assumption where the unique factorization is known as Wiener-Hopf factorization. For this setting a low-complexity algorithm for blind deconvolution based on [3] will be investigated which exactly (up trivial ambiguities) recovers $\boldsymbol{x}$ and $\boldsymbol{h}$. The algorithm allows for large-scale problems and will be compared to existing methods.

REFERENCES

[1] A. Ahmed, J. Romberg, and B. Recht, "Blind deconvolution using convex programming," *IEEE Trans. Inf. Theory*, no. 3, pp. 1711–1732.

[2] X. Li, S. Ling, T. Strohmer, and K. Wei, "Rapid, Robust, and Reliable Blind Deconvolution via Nonconvex Optimization," pp. 1–49, 2016. [Online]. Available: http://arxiv.org/abs/1606.04933

[3] A. Böttcher and M. Halwass, "Wiener-Hopf and spectral factorization of real polynomials by Newton's method," *Linear Algebra and Its Applications*, vol. 438, no. 12, pp. 4760–4805, 2013.

Learning Bayesian Optimal FISTA with Error Backpropagation

Ulugbek S. Kamilov and Hassan Mansour
Mitsubishi Electric Research Laboratories (MERL), 201 Broadway, Cambridge, MA 02139, USA.

***Abstract*—Regularized least-squares (RLS) estimator is one of the most commonly used approaches for solving linear inverse problems. In the context of statistical signal reconstruction, RLS is typically interpreted as a maximum posteriori probability (MAP) estimator. However, recent works have showed that minimum mean squared error (MMSE) estimator can also be expressed as the solution of RLS. In this work, we present a scheme for training the popular fast iterative shrinkage/thresholding algorithm (FISTA) for computing MMSE estimator. Specifically, we show that by representing FISTA as a deep neural network (DNN), the error backpropagation algorithm can be used to learn thresholding functions that minimize the MSE for a given statistical distribution of data.**

I. Training FISTA for MMSE estimation

We consider a linear inverse problem $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}$, where the goal is to recover the unknown signal $\mathbf{x} \in \mathbb{R}^N$ from the noisy measurements $\mathbf{y} \in \mathbb{R}^M$. The matrix $\mathbf{H} \in \mathbb{R}^{M\times N}$ is known and models the response of the acquisition device, while the vector $\mathbf{e} \in \mathbb{R}^M$ represents unknown errors in the measurements. Inverse problems are often ill-posed, which means that measurements $\mathbf{y}$ cannot explain the signal $\mathbf{x}$ uniquely. One standard approach for solving such problems is the regularized least-squares (RLS) estimator

$$\widehat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 + \mathcal{R}(\mathbf{x}) \right\}, \tag{1}$$

where $\mathcal{R}$ is a regularizer that imposes prior structure in order to promote more meaningful solutions. For example, the choice of $\mathcal{R}(\cdot) = \|\cdot\|_{\ell_1}$ leads to the famous Lasso estimator.

In the context of statistical signals $\mathbf{x} \sim p_{\mathbf{x}}$ and $\mathbf{e} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$, it is possible to interpret RLS as a maximum a posteriori probability (MAP) estimator by setting $\mathcal{R}_{\text{MAP}}(\mathbf{x}) = -\sigma^2 \log(p_{\mathbf{x}}(\mathbf{x}))$. Such interpretation establishes a popular connection between Lasso and the Laplace statistical prior on the signal $\mathbf{x}$. There is, however, an alternative interpretation by Gribonval and Machart [1], establishing a connection between RLS and a minimum mean square estimator (MMSE) that can be defined as $\widehat{\mathbf{x}}_{\text{MMSE}} = \mathbb{E}[\mathbf{x}|\mathbf{y}]$. While, MMSE estimator is optimal in terms of quadratic error, its direct implementation involves tedious integral computations that cannot be practically performed. Gribonval and Machart [1] have proved that for a non-degenerate priors $p_{\mathbf{x}}$, noise levels σ^2, and full-rank matrices $\mathbf{H}$, there exists a regularizer $\mathcal{R}_{\text{MMSE}} = \mathcal{R}_{\mathbf{H},p_{\mathbf{x}},\sigma^2}$ such that the MMSE estimator is a minimizer of RLS with $\mathcal{R} = \mathcal{R}_{\text{MMSE}}$. This result highlights the fact that while one Bayesian interpretation of RLS is the MAP estimator $\widehat{\mathbf{x}}_{\text{MAP}}$ with prior $p_{\mathbf{x}}(\mathbf{x}) \propto \exp(-\mathcal{R}_{\text{MAP}}(\mathbf{x}))$, there is another admissible Bayesian interpretation as the MMSE estimator $\widehat{\mathbf{x}}_{\text{MMSE}}$, where in general $\mathcal{R}_{\text{MMSE}}(\mathbf{x}) \neq -\sigma^2 \log(p_{\mathbf{x}}(\mathbf{x}))$. Another important consequence of this result is that by identifying the regularizer $\mathcal{R}_{\text{MMSE}}$, one can reformulate MMSE estimation as RLS and rely on numerical optimization for computing $\widehat{\mathbf{x}}_{\text{MMSE}}$ in a potentially tractable fashion.

A common approaches for solving (1) is the *fast iterative shrinkage/thresholding algorithm (FISTA)* [2] that can be expressed as

$$\mathbf{z}^t \leftarrow \mu_t \mathbf{S}\mathbf{x}^{t-1} + (1-\mu_t)\mathbf{S}\mathbf{x}^{t-2} + \mathbf{b} \tag{2a}$$
$$\mathbf{x}^t \leftarrow \eta(\mathbf{z}^t). \tag{2b}$$

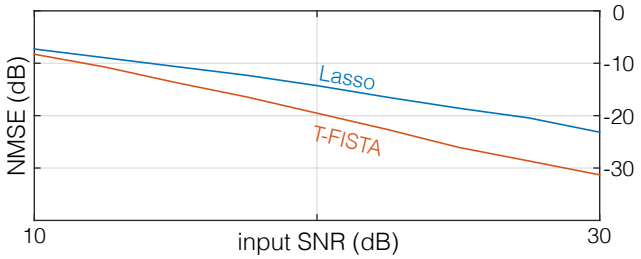


Fig. 1. Average NMSE of Lasso and trained FISTA plotted for different noise levels at the measurement rate $M/N = 0.6$ when recovering Bernoulli-Gaussian signals $\mathbf{x}$ of sparsity $\rho = 0.2$ from measurements with i.i.d. $\mathbf{H}$.

where $\gamma = 1/\lambda_{\max}(\mathbf{H}^{\mathrm{T}}\mathbf{H})$ is the step-size, $\{\mu_t\}_t$ are the relaxation parameters, $\mathbf{S} \triangleq \mathbf{I} - \gamma\mathbf{H}^{\mathrm{T}}\mathbf{H}$, $\mathbf{b} \triangleq \gamma\mathbf{H}^{\mathrm{T}}\mathbf{y}$, and $\eta(\cdot)$ is a nonlinearity that reduces to a scalar function for a separable $\mathcal{R}$.

Iterations of FISTA can be represented as a feedforward neural network whose adaptable parameters correspond to the nonlinearity $\eta(\cdot)$ [3]. We adopt the following parametric representation for the nonlinearity $\eta(z) \triangleq \sum_{k=-K}^{K} c_k \phi(z/\Delta - k)$, where $\mathbf{c} \triangleq \{c_k\}_{k\in[-K,\ldots,K]}$ are the coefficients of the representation and ϕ are basis functions. Then, $\eta(\cdot)$ can be learned by solving

$$\widehat{\mathbf{c}} = \underset{\mathbf{c}\in\mathcal{C}}{\arg\min} \left\{ \frac{1}{L}\sum_{\ell=1}^{L} \mathcal{E}_\ell(\mathbf{c}) \right\}, \tag{3}$$

where $\mathcal{C} \subseteq \mathbb{R}^{2K+1}$ represents prior constraints on the coefficients, and $\mathcal{E}$ is the empirical MSE defined as $\mathcal{E}_\ell(\mathbf{c}) \triangleq \frac{1}{2}\|\mathbf{x}_\ell - \mathbf{x}^T(\mathbf{c}, \mathbf{y}_\ell)\|_{\ell_2}^2$, where $\mathbf{x}^T$ is the solution of FISTA at iteration T. Given a large number of i.i.d. realizations of the signals $\{\mathbf{x}_\ell, \mathbf{y}_\ell\}$, the empirical MSE is expected to approach the true MSE of FISTA for the given statistical distribution of data.

Let $\mathbf{\Phi}^t$ be the matrix whose entries $\Phi^t_{mk} = \phi(z^t_m/\Delta - k)$ at row m and column k. The gradient $\nabla\mathcal{E}_\ell(\mathbf{c})$ is given by $\nabla\mathcal{E}_\ell(\mathbf{c}) = \mathbf{g}^1 + [\mathbf{\Phi}^1]^{\mathrm{T}}\mathbf{r}^1_1$, where $\mathbf{g}^1$ and $\mathbf{r}^1$ are computed using the following error backpropagation algorithm for $t = T, T-1, \ldots, 2$

$$\mathbf{g}^{t-1} \leftarrow \mathbf{g}^t + \mathbf{\Phi}^t\boldsymbol{\epsilon}^t \tag{4a}$$
$$\mathbf{v}^{t-1} \leftarrow \mathbf{S}\mathrm{diag}(\eta'(\mathbf{z}^t))\boldsymbol{\epsilon}^t \tag{4b}$$
$$\boldsymbol{\epsilon}^{t-1} \leftarrow \mu_t\mathbf{v}^{t-1} + (1-\mu_{t+1})\mathbf{v}^t, \tag{4c}$$

where $\boldsymbol{\epsilon}^T = \mathbf{x}^T - \mathbf{x}$, $\mathbf{v}^T = 0$, $\mu_{T+1} = 0$, and $\mathbf{g}^T = 0$. Given the gradient, we can optimize over the coefficients $\mathbf{c}$ in an online fashion using the projected gradient algorithm.

References

[1] R. Gribonval and P. Machart, "Reconciling "priors" & "priors" without prejudice?" in *Proc. Advances in Neural Information Processing Systems 26*, Lake Tahoe, NV, USA, December 5-10, 2013, pp. 2193–2201.

[2] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[3] U. S. Kamilov and H. Mansour, "Learning optimal nonlinearities for iterative thresholding algorithms," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 747–751, May 2016.

Optimal Injectivity Conditions for Bilinear Inverse Problems with Applications to Identifiability of Deconvolution Problems

Michael Kech* Felix Krahmer*
* Department of Mathematics, Technical University of Munich, Boltzmannstr. 3, 85747 Garching/Munich, Germany

***Abstract*—We study identifiability for bilinear inverse problems under sparsity and subspace constraints. We show that, up to a global scaling ambiguity, almost all such maps are injective on the set of pairs of sparse vectors if the number of measurements m exceeds $2(s_1+s_2)-2$, where s_1 and s_2 denote the sparsity of the two input vectors, and injective on the set of pairs of vectors lying in known subspaces of dimensions n_1 and n_2 if $m \geq 2(n_1+n_2)-4$. We also prove that both these bounds are tight in the sense that one cannot have injectivity for a smaller number of measurements. Our proof technique draws from algebraic geometry. As an application we derive optimal identifiability conditions for the deconvolution problem, thus improving on recent work of Li et al. [1].**

I. Introduction

Bilinear inverse problems are ubiquitous in signal processing and communications and have been studied extensively over several decades in both the engineering and the mathematics literature. In such problems, one observes a (potentially noisy) bilinear map on two input signals, and the goal is to recover the inputs.

More precisely, the aim is to reconstruct two vectors $u \in \mathbb{C}^{n_1} \setminus \{0\},\ v \in \mathbb{C}^{n_2} \setminus \{0\}$ from the measurement outcome $\tilde{B}(u,v)$, where $\tilde{B} : \mathbb{C}^{n_1} \times \mathbb{C}^{n_2} \to \mathbb{C}^m$ is a bilinear map. At best, this is possible up to a global multiplicative factor.

A natural first step in analyzing such problems is the quest for conditions of identifiability, that is, to ask when such problems with no added noise have a unique solution up to this ambiguity.

It is not difficult to describe bilinear maps (including certain real life scenarios), where strong identifiability fails despite very restrictive signal classes. These cases can, however, be considered as exceptional. As we will show, *generic* bilinear maps will be identifiable for m very close to the number of unknowns. The bounds we derive are tight as no bilinear inverse problem will be identifiable m lower than our bound.

II. Tight Bounds for the Injectivity Problem

Our analysis uses the fact that every bilinear map can be "lifted" to a linear map on the outer product, that is, for $\tilde{B}$ bilinear there exists a linear map B such that $\tilde{B}(u,v) = B(uv^t)$. Thus to show that a bilinear map is injective on pairs of sparse vectors (for the case of no sparsity constraint, set the sparsity to be the space dimension), it suffices to show injectivity of B on the set

$$M^1_{s_1,s_2}(n_1,n_2) := \{uv^t :\ u \in \mathbb{C}^{n_1}_{s_1},\ v \in \mathbb{C}^{n_2}_{s_2}\},$$

where $\mathbb{C}^n_s := \{x \in \mathbb{C}^n :\ x \text{ is } s\text{-sparse}\}$. We hence seek stably (s_1,s_2)-injective maps as introduced in the following definition.

Definition 1 (Stably (s_1,s_2)-injective): A linear map $B : \mathbb{C}^{n_1 \times n_2} \to \mathbb{C}^m$ is called stably (s_1,s_2)-injective iff there exists a constant $C > 0$ such that $\|B(X)\| \geq C\|X\|_{HS}$ holds for all $X \in \{\lambda(X-Y) | \, X,Y \in M^1_{s_1,s_2}(n_1,n_2),\ \lambda > 0\}$.

The first result gives a general lower bound on the number of measurement outcomes of a stably (s_1,s_2)-injective linear map.

Theorem 1 (Lower bound [2]): If the linear map $B : \mathbb{C}^{n_1 \times n_2} \to \mathbb{C}^m$ is stably (s_1,s_2)-injective, then

$$m \geq \begin{cases} 2(n_1+n_2)-4 & \text{if } s_1 = n_1, s_2 = n_2, \\ 2(s_1+s_2)-2 & \text{else.} \end{cases}$$

The following theorem shows that the lower bound given by Theorem 1 is indeed tight.

Theorem 2 (Upper bound [2]): Almost all[1] linear maps $B : \mathbb{C}^{n_1 \times n_2} \to \mathbb{C}^m$ are stably (s_1,s_2)-injective if

$$m \geq \begin{cases} 2(n_1+n_2)-4 & \text{if } s_1 = n_1, s_2 = n_2, \\ 2(s_1+s_2)-2 & \text{else.} \end{cases}$$

III. Blind Deconvolution

A very important class of bilinear inverse problems are blind deconvolution problems, where one observes the convolution of two signals. This model has applications for example in image deblurring, where the inputs represent the image and a blur kernel, or in communication, where the two inputs represent signal and channel. In this talk, we will consider the following variant.

The circular convolution of vectors $v,w \in \mathbb{C}^m$ is denoted by $v \circledast w \in \mathbb{C}^m$, i.e., for all $i \in \{0,\dots,m-1\}$ one has $(v \circledast w)_i := \sum_{j=0}^{m-1} v_j w_{(i-j) \bmod m}$, which is clearly bilinear. By design, the number of measurements agrees with the space dimension, so one cannot hope for injectivity on the set of all inputs, but one needs to consider subclasses imposing additional structure. Here we consider the class of signals that are sparse in given bases or frames. Here the term "almost all" refers to the Lebesgue measure on $\mathbb{C}^{m \times k}$.

Theorem 3 (Deconvolution with sparsity constraint [2]): Let $s_1, s_2 \in \mathbb{N}_+$ be such that $2(s_1+s_2)-2 \leq m$ and let $k,l \in \mathbb{N}$ be such that $s_1 < k \leq m$, $s_2 < l \leq m$. Then, for almost all $(D,E) \in \mathbb{C}^{m \times k} \times \mathbb{C}^{m \times l}$ with D,E both a frame, i.e., of full rank, the linear map C representing the circular convolution map, that is, $C(uv^t) = (Dv) \circledast (Ew)$ is stably (s_1,s_2)-injective.

Corresponding results are also derived for the case that the signal is known to lie in a generic subspace. For both the sparse case and the subspace case, slightly suboptimal identifiability conditions had been derived prior to our work in [3], [1]. In contrast, it follows from Theorem 1 that the bounds given in Theorem 3 as well as the ones for the subspace case are optimal.

References

[1] Li, Y., Lee, K., and Bresler, Y., "Identifiability in blind deconvolution under minimal assumptions," 2015, arXiv: 1507.01308.
[2] Kech, M. and Krahmer, F., "Optimal injectivity conditions for bilinear inverse problems with applications to identifiability of deconvolution problems," 2016, arXiv: 1603.07316.
[3] Li, Y., Lee, K.and Bresler, Y., "Identifiability in blind deconvolution with subspace or sparsity constraints," 2015, arXiv:1505.03399.

[1] By vectorizing the input matrices, the set of linear maps $B : \mathbb{C}^{n_1 \times n_2} \to \mathbb{C}^m$ can be identified with $\mathbb{C}^{n_1 n_2 \times m}$. The term "almost all" refers to the Lebesgue measure on $\mathbb{C}^{n_1 n_2 \times m}$.

Convex signal reconstruction with positivity constraints

Richard Kueng*,
* Institute for Theoretical Physics, University of Cologne

***Abstract*—Convex signal reconstruction is the art of reconstructing structured signals, e.g. vectors, or matrices, from an under-determined set of linear measurements via convex optimization techniques. Important examples include *compressed sensing* and *low rank matrix reconstruction*. Reconstruction is typically performed via a constrained norm minimization, where the norm serves as a surrogate for structure and the constraints take into account the measurement data.**

Additional structural constraints, such as positivity, can have profound impacts on the algorithmic reconstruction. These insights date back to the very early days of compressed sensing. Here, we formulate novel implications of positivity constraints that are geared towards noise robustness and take into account more recent progress in the field of convex signal reconstruction.

I. Sparse reconstruction of non-negative vectors

Compressed sensing aims at reconstructing s-sparse vectors $x \in \mathbb{R}^n$ from $m \ll n$ noisy, linear measurements. Such a measurement process may succinctly be written as

$$y = Ax + e, \tag{1}$$

where $A \in \mathbb{R}^{m\times n}$ models the measurement process and $e \in \mathbb{R}^m$ denotes additive noise. Typically, the actual reconstruction is then performed by solving a constrained ℓ_1-norm minimization:

$$z^\sharp = \text{argmin} \quad \|z\|_{\ell_1} \quad \text{subject to} \qquad \|Az - y\|_{\ell_2} \leq \eta. \tag{2}$$

Here, the constant η denotes an upper bound on the noise strength $\|e\|_{\ell_2}$ present in the measurement process.

To this date, several sufficient criteria on A have been established that assure that (2) indeed converges to the right solution, see e.g. [1] for an overview. Among these, the *null space property* (NSP) has the added benefit of being both necessary and sufficient. Loosely speaking, a matrix $A \in \mathbb{R}^{m\times n}$ obeys a null space property of order $s \leq n$, if its kernel (nullspace) does not contain any s-sparse vectors. Robust versions of this property assure that (2) stably reconstructs any s-sparse $x \in \mathbb{R}^n$ from noisy measurements of the form (1):

$$\|z^\sharp - x\|_{\ell_2} \leq \frac{D}{\sqrt{m}}\eta. \tag{3}$$

This implies that the reconstruction error scales linearly in η—the upper bound on the noise strength $\|e\|_{\ell_2}$.

The study of reconstructing sparse vectors that are in addition entry-wise nonnegative ($x \geq 0$) has a comparatively long history [2]. Here we adopt one geometric condition by Bruckstein *et al.* [3] and combine it with more recent techniques from compressed sensing:

Theorem 1 (Main result in [4])**.** *Suppose that $A \in \mathbb{R}^{m\times n}$ obeys a robust version of the NSP of order s and there exists a combination of measurement vectors $w = \sum_{k=1}^m t_k a_k$ that is entry-wise positive. Then, every nonnegative, s-sparse vector $x \in \mathbb{R}^n$ may be reconstructed from* (1) *via solving*

$$z^\sharp = \underset{z\geq 0}{\textit{argmin}} \|Az - y\|_{\ell_2}. \tag{4}$$

This reconstruction is stable, in the sense that $z^\sharp$ is guaranteed to obey

$$\|z^\sharp - x\|_{\ell_2} \leq \frac{D}{\sqrt{m}}\|e\|_{\ell_2}. \tag{5}$$

Compared to conventional compressed sensing results, this statement has two benefits: (i) the reconstruction error (5) is directly proportional to $\|e\|_{\ell_2}$—the actual noise corruption in the measurements; (ii) the algorithm (4) is considerably simpler than (2), or similar reconstructions protocols that require a ℓ_1-regularization of some sort.

A concrete example for measurement matrices A that meets the requirements of Theorem 1 are $0/1$-Bernoulli matrices with $m = Cs\log(n)$ rows [4].

II. Low rank matrix reconstruction of positive semidefinite matrices

The task of reconstructing a rank-r matrix X from noisy, linear measurements of the form $y = \mathcal{A}(X) + e \in \mathbb{R}^m$ may be viewed as a non-commutative analogue of compressed sensing. Typically, reconstruction is performed via solving a constrained norm minimization similar to (2) over the set of all matrices. However, the ℓ_1-norm is replaced by the nuclear norm which serves as a convex surrogate for rank. Positive-semidefiniteness ($X \geq 0$) is the natural non-commutative analogue of non-negativity. One of the main results in [5] corresponds to a matrix version of Theorem 1:

Theorem 2 (Theorem 8.3 in [5])**.** *Suppose that a measurement process $\mathcal{A} : H_d \to \mathbb{R}^m$ obeys a matrix version of the NSP of order r and that there exists a linear combination $W = \sum_{k=1}^m t_k W_k > 0$ that is positive definite. Then, any positive semidefinite rank-r matrix may be reconstructed from the acquired measurement data via solving*

$$Z^\sharp = \underset{Z\geq 0}{\textit{argmin}} \|\mathcal{A}(Z) - y\|_{\ell_2}.$$

This reconstruction is stable towards additive noise in a way that is completely analogous to (3)*.*

III. Applications

Positivity constraints are implicitly present in many applications. Entry-wise non-negativity of vectors occurs, for instance naturally in imaging problems and optical communication. Likewise, positive semidefinite matrices arise in kernel-based learning methods, convex relaxations of the phase retrieval problem and quantum mechanics.

References

[1] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, ser. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013.

[2] D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 27, pp. 9446–9451, 2005.

[3] A. M. Bruckstein, M. Elad, and M. Zibulevsky, "On the uniqueness of non-negative sparse & redundant representations," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, no. 796, pp. 5145–5148, 2008.

[4] R. Kueng and P. Jung, "Robust Nonnegative Sparse Recovery and the Nullspace Property of 0/1 Measurements," 2016. [Online]. Available: http://arxiv.org/abs/1603.07997

[5] M. Kabanava, R. Kueng, H. Rauhut, and U. Terstiege, "Stable low-rank matrix recovery via null space properties," *Information and Inference*, 2016. [Online]. Available: http://imaiai.oxfordjournals.org/content/early/2016/08/28/imaiai.iaw014.full.pdf+html

New Physically Plausible Compressive Sampling Schemes for High Resolution MRI

Carole Lazarus*, Nicolas Chauffert*, Pierre Weiss†, Jonas Kahn‡, Alexandre Vignaud* and Philippe Ciuciu*
* CEA/NeuroSpin Center & INRIA-CEA Parietal team, University of Paris-Saclay, F-91191 Gif-sur-Yvette, France
† PRIMO Team, Advanced Life Science Institute, CNRS (USR 3505), University of Toulouse, Toulouse, France
‡ Institute of Mathematics (UMR CNRS 5219), University of Toulouse, Toulouse, France.

***Abstract*—Magnetic resonance imaging (MRI) is one of the most successful application fields of compressed sensing (CS). Despite recent advances, there is still a large discrepancy between theories and most actual implementations. In [1], [2], we have proposed a new generic principle for designing 2D k-space sampling trajectories that are fast enough, compatible with hardware and imaging constraints and that follow a prescribed sampling density. In this communication, we illustrate the performance of this approach on a retrospective CS scenario. During the workshop, 20-fold acceleration results in prospective CS will be illustrated on *ex-vivo* high resolution T_2^* imaging at 7 Tesla.**

I. Introduction

MRI data are collected in the k-space (spatial Fourier domain) along regular trajectories which are subject to kinematic constraints. Indeed, the gradient waveforms which are responsible for this displacement in k-space are obtained by energizing gradient coils with electric currents, whose amplitude and slew rate are upper bounded. Since high resolution MR imaging requires visiting larger k-space domains (i.e., larger $k_{\max}$), collecting such data is time consuming.

On the other hand, MR image resolution improvement in standard scanning times (e.g., 200 μm in-plane in 15 min) would allow neuroscientists and doctors to push the limits of their current knowledge and to significantly improve both their diagnosis and patients' follow-up. One critical path to achieve this goal relies on the CS theory [3], [4], which has revolutionized how data can be collected in a compressed manner while ensuring conditions for optimal image recovery. This breakthrough has been accomplished by combining three key ingredients: (i) pseudo-random acquisitions, (ii) image representation using sparse decompositions (e.g., wavelets) and (iii) nonlinear image reconstruction.

Although heuristic application of CS in MRI has provided promising results [5], CS theory cannot be directly cast to the MRI setting. The reasons are: 1) the acquisition (Fourier) and representation (wavelets) bases are coherent and 2) 2D sampling schemes obtained using CS theorems are composed of isolated measurements and cannot be efficiently implemented by magnetic field gradients. In the recent literature [6], [7], variable density sampling (VDS) theory has addressed the first impediment. Moreover, in the seminal paper [5], 2D pointwise sampling was performed along parallel lines in the orthogonal readout direction to the the slices of interest, thus implementing a 2D VDS within each slice. However, in a 3D perspective, this 2D-VDS is likely suboptimal since high frequencies along the readout direction are sampled too densely.

To go beyond this approach, new 2D sampling trajectories that fulfill acquisition constraints while traversing the k-space as fast as possible according to a prescribed variable density have been proposed in [1], [2]. In brief, the proposed framework consists of projecting a probability distribution (i.e. π) onto a set of measures that are brought by admissible curves with respect to the gradient constraints. The proposed algorithm also allows to handle arbitrary affine constraints (e.g. echo time specification) and automatically generates efficient sampling patterns. So far, it has been implemented in 2D although its 3D extension is currently under study. On retrospectively undersampled ($\sim$ 5 % of full k-space) simulated data (Fig. 1), we illustrate its impact on the signal-to-noise ratio (SNR) of reconstructed MR images that were computed using a non-Cartesian implementation of the FISTA algorithm [8]. The reconstruction results using this strategy outperform existing acquisition trajectories (spiral, radial) by about 3 dB. More recently, we adapted a GRE sequence to acquire a T_2^* weighted image of an *ex-vivo* baboon brain at 7 Tesla (Siemens Magnetom) with an adapted version of these multi-shot trajectories (sense of k-space traversal, starting point selection, ...). Our preliminary results proved the practical feasibility of these sampling schemes and the 20-fold acceleration of acquisitions with shots lasting less than 40 ms.

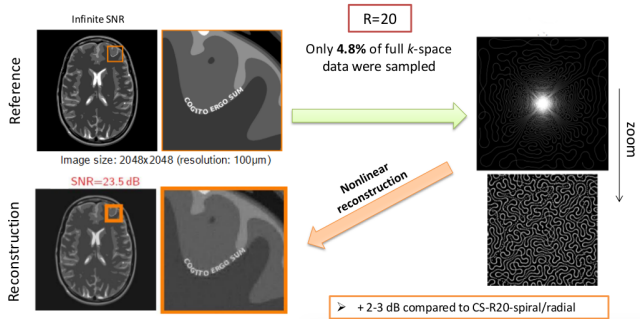


Fig. 1. Simulation results for $n = 2048 \times 2048$, $G_{\max} = 40$ mT.m^{-1}, $S_{\max} = 150$ mT.m^{-1}.ms^{-1}. The trajectory (right side) is made of 4 segments of 25,000 points each distributed according to $\pi = 1/(\|k\| + 1)^2$.

References

[1] C. Boyer, N. Chauffert, P. Ciuciu, J. Kahn, and P. Weiss, "On the generation of sampling schemes for Magnetic Resonance Imaging," *accepted to SIAM Journal on Imaging Science*, Sep. 2016.
[2] N. Chauffert, P. Weiss, J. Kahn, and P. Ciuciu, "A projection algorithm for gradient waveforms design in Magnetic Resonance Imaging," *IEEE Trans. Med. Imag.*, vol. 35, no. 9, pp. 2026–2039, Sep. 2016.
[3] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.
[4] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[5] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.
[6] G. Puy, P. Vandergheynst, and Y. Wiaux, "On variable density compressive sampling," *IEEE Trans. Sig. Proc. Let.*, vol. 18, no. 10, pp. 595–598, 2011.
[7] B. Adcock, A. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing," 2013.
[8] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

A nonconvex optimization method for multichannel blind deconvolution with sparsity channel models

Kiryung Lee* and Justin K. Romberg*
* School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332–0250.

I. Extended Abstract

Multichannel blind deconvolution resolves unknown input signal from multiple channel outputs with unknown impulse responses. It is often easier to estimate only the unknown channel impulse responses in applications in wireless communications and underwater acoustics, where a compact deterministic model with few parameters is not available for the input signal. In 1990s, various reconstruction methods based on statistics of the input signal and/or the commutativity of the convolution operation have been proposed with algebraic performance guarantees. However, the provided guarantees are restricted to the case where the observations are noise-free or the input has infinite length. Furthermore, the empirical performance of these classical methods deteriorates dramatically with a short input signal, which restricts its utility in estimating time-varying channels.

Recently, motivated by a channel estimation problem in underwater acoustics, the authors with Ning Tian introduced a bilinear system model to multichannel deconvolution and provided a fast algorithm with non-asymptotic performance guarantees in the presence of noise [1]. In this work, we consider a new channel model with sparsity priors and propose corresponding solutions that exploit the given channel model. While the previous bilinear model with its separability structure enabled natural alternating minimization approach, the same strategy is not applicable to our new channel model. Instead, we propose an iterative algorithm for a nonconvex optimization formulation of multichannel blind deconvolution under the sparsity channel model. We show performance guarantees for the proposed algorithm in terms of fast convergence and stable recovery of the channel model parameters. Furthermore, numerical results with generic data and with synthesized data in a realistic setup will be presented to support our theoretic findings.

References

[1] K. Lee, N. Tian, and J. K. Romberg, "Fast and guaranteed blind multichannel deconvolution under a bilinear channel model," in *Information Thoery Workshop*, Cambridge, UK, September 2016.

Iterated Conditional Expectations for Total Variation Image Restoration

Cécile Louchet*, Lionel Moisan†
* Université d'Orléans, MAPMO, CNRS UMR 7349, France
† Université Paris Descartes, MAP5, CNRS UMR 8145, France

***Abstract*—From a Bayesian point of view, variational models derived from the original model of Rudin, Osher and Fatemi can be interpreted as a maximum a posteriori estimate associated with a total variation Gibbs prior. While the posterior mean is known to be a better estimate in general, its practical computation with Monte-Carlo Markov chains is computationally very expensive. We here explore a recent alternative, based on the iteration of conditional expectations. We discuss several possibilities to extend this model from simple denoising to more general inverse problems, while maintaining the fixed point structure and thus the linear convergence of the algorithms.**

Let $v : \Omega \to \mathbb{R}$ be a discrete gray-level image, defined on a rectangular domain $\Omega \subset \mathbb{Z}^2$. According to the ROF (Rudin, Osher, Fatemi) model [1], a denoised version $u : \Omega \to \mathbb{R}$ of v can be estimated by minimizing the energy

$$E(u) = \|u - v\|^2 + \lambda TV(u),$$

where $TV(u)$ is the discrete total variation of u and $\lambda > 0$ is a parameter that controls the level of regularization. Various discretization schemes can be used to define $TV(u)$, but we shall here focus on the so-called *anisotropic* total variation

$$TV(u) = \frac{1}{2} \sum_{|x-y|=1} |u(y) - u(x)|.$$

Assuming that u follows the (improper) apriori distribution

$$p(u) \propto e^{-\beta TV(u)}$$

(for an appropriate choice of β) and that $v - u$ is a white Gaussian noise with variance σ^2, the minimizer $\hat{u}_{\text{ROF}} = \arg\min_u E(u)$ can be interpreted as the maximum density point of the posterior distribution

$$\pi(u) = p(u|v) \propto \exp\left(-\frac{\|u - v\|^2 + \lambda TV(u)}{2\sigma^2}\right).$$

As discussed in [2], choosing such a maximum a posteriori estimate is questionable, and the posterior mean estimate

$$\hat{u}_{\text{LSE}} = \mathbb{E}_\pi(u) = \int_{\mathbb{R}^\Omega} \pi(u)\, u\, du,$$

which achieves the least square error (in average), is often to be preferred. Such a high-dimensional (typically, 10^6 for a 1000×1000 image) integral can be estimated using a Monte-Carlo Markov Chain algorithm, but it is computationally quite intensive.

It is interesting to remark that for any $x \in \Omega$, $\hat{u}_{\text{LSE}}(x)$ is the expectation of the marginal distribution

$$\pi(u(x)) \propto \int_{\mathbb{R}^{x^c}} \pi(u)\, du(x^c),$$

where $u(x^c)$ is the restriction of u to $x^c = \Omega \setminus \{x\}$. This marginal distribution is difficult to compute, and is it tempting to use the conditional marginal distribution $\pi(u(x)|u(x^c))$ instead, whose expectation only involves the computation of a one-dimensional integral. This leads to the Iterated Conditional Expectation (ICE) estimate, obtained as the limit of the fixed point recurrence

$$u^{n+1}(x) = \mathbb{E}_\pi\left(u(x) \,\middle|\, u(x^c) = u^n(x^c)\right) = f_{v(x)}(u^n), \qquad (1)$$

where the integral $f_{v(x)}(u^n)$ can be explicitly computed using the error function erf. In [3], we showed that this iterative scheme linearly converges to a solution $\hat{u}_{\text{ICE}}$, and experimentally observed that the images delivered by $\hat{u}_{\text{ICE}}$ and $\hat{u}_{\text{LSE}}$ were visually very close.

How to extend the ICE model to more general inverse problems? If we now consider the energy

$$E(u) = \|Au - v\|^2 + \lambda TV(u)$$

(for some known linear operator A) and the associated posterior distribution $\pi(u) \propto \exp(-E(u)/(2\sigma^2))$, one can show that the natural ICE iteration can be rewritten under the form

$$\begin{cases} w^n(x) = u^n(x) - \gamma(x) A^*(Au^n - v)(x) \\ u^{n+1}(x) = f_{w^n(x)}(u^n), \end{cases}$$

where $\gamma(x) = \|A\delta_x\|^{-2}$ (δ_x being the discrete Dirac at pixel x), and f is the function appearing in the iteration of ICE denoising (1), but here considered for the parameters $\lambda(x) = \gamma(x)\lambda$ and $\sigma^2(x) = \gamma(x)\sigma^2$. One can prove the convergence of this iterative scheme, but with important restrictions on A that unfortunately discard many interesting cases (in particular, medium and high blur operators). To overcome this limitation, we study three variants of the ICE iteration that considerably extend the convergence cases. The associated algorithms exhibit a linear convergence rate (as in the denoising case), and we illustrate them with classical imaging inverse problems like deblurring, image magnification [4] and spectrum extrapolation. Although theses variants do not theoretically define the same solution, differences are visually barely noticeable in practice.

References

[1] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1, pp. 259–268, 1992.
[2] C. Louchet and L. Moisan, "Posterior expectation of the total variation model: Properties and experiments," *SIAM J. Imaging Sciences*, vol. 6, pp. 2640–2684, 2013.
[3] ——, "Total variation denoising using iterated conditional expectation," *Proc. Eur. Signal Process. Conf.*, pp. 1592–1596, 2014.
[4] F. Guichard and F. Malgouyres, "Total variation based interpolation," *Proceedings of the European signal processing conference*, vol. 3, pp. 1741–1744, 1998.

A differential-geometric derivation of MAP estimation

Marcelo Pereyra
School of Mathematics, University of Bristol, U.K.
School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, U.K.

***Abstract*—Bayesian estimators arise from Bayesian decision theory as optimal summaries of the posterior $p(x|y)$ w.r.t an expected loss. In this paper we revisit the question of the choice of this loss function in the context of convex problems. We show that under mild regularity conditions, this loss is intimately related to $p(x|y)$ by the differential geometry of the parameter space. Precisely, $p(x|y)$ induces a dually-flat Riemannian geometry on the parameter space, and taking into account this geometry naturally leads to a canonical loss function to perform Bayesian estimation. We then show that this canonical loss is given by the Bregman divergence associated with $-\log p(x|y)$, and that the Bayes estimator w.r.t. this loss is the maximum-a-posteriori estimator.**

I. Bayesian point estimation for convex problems

We consider the Bayesian estimation of $x \in \mathbb{R}^n$ from an observation y, related to x by the posterior $p(x|y) = p(y|x)p(x)/p(y)$. We assume that

$$p(x|y) = \frac{\exp\{-\phi(x)\}}{\int_{\mathbb{R}^n} \exp\{-\phi(s)\}\mathrm{d}s} \tag{1}$$

for some ϕ strongly convex and almost everywhere $\mathcal{C}^3$ over $\mathbb{R}^n$.

Because drawing conclusions directly from $p(x|y)$ is difficult we deliver summaries, namely Bayesian estimators that summarise $p(x|y)$ optimally in the following decision-theoretic sense [1]:

$$\hat{x}_L = \operatorname*{argmin}_{u\in\mathbb{R}^n} \mathrm{E}_{x|y}[L(u,x)] \triangleq \int_{\mathbb{R}^n} L(u,x)p(x|y)\mathrm{d}x.$$

where $L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is some relevant loss function to quantify the difference between two points in $\mathbb{R}^n$ and that verifies:

- $L(u,x) \geqslant 0,\ \forall\, u, x \in \mathbb{R}^n$,
- $L(u,x) = 0$ iff $u = x$,
- L strictly convex w.r.t. u (to guarantee estimator uniqueness).

The careful choice of L is extremely important, yet it has received very little attention in the imaging literature. Moreover, most modern imaging methods use the maximum-a-posteriori (MAP) estimator

$$\hat{x}_{MAP} = \operatorname*{argmax}_{x\in\mathbb{R}^n} p(x|y) = \operatorname*{argmin}_{x\in\mathbb{R}^n} \phi(x),$$

that can be calculated efficiently by convex optimisation and generally delivers very accurate results for convex models, but which is widely thought to not be a Bayes estimator in a decision theory sense [1].

II. Canonical Bayesian estimation

A. Differential geometry of the parameter space

Suppose that (1) holds, and let $\phi(x) = -\log p(x|y)$. Because ϕ is $\mathcal{C}^3$ and strongly convex it induces a Riemannian geometry on $\mathbb{R}^n$ [2]. Precisely, from differential geometry, we have a dually-flat Riemannian manifold $(\mathbb{R}^n, g, \Gamma, \Gamma^\star)$ with a metric

$$g_{i,j}(x) = \partial_i\partial_j\phi(x), \quad \forall x \in \mathbb{R}^n,\ \forall\, i,j \in \{1,\dots,n\}, \tag{2}$$

primal and dual affine connections Γ and $\Gamma^\star$ with coefficients

$$\Gamma_{ij,\,k}(x) = \partial_i\partial_j\partial_k\phi(x), \quad \Gamma^\star_{ij,\,k}(x) = \partial_i\partial_j\partial_k\phi^\star(x), \tag{3}$$

and primal and dual coordinates x and η, related by the duality $\eta = \nabla\phi(x)$ and $x = \nabla\phi^\star(\eta)$ where $\phi^\star(\eta) = \max_{x\in\mathbb{R}^n} x^\top\eta - \phi(x)$.

Similarly to Euclidean spaces, this kind of manifold supports divergence functions. In particular, dually-flat Riemannian manifolds are equipped with a *canonical* divergence that generalises the Euclidean quadratic distance $d(u,x) = \|u-x\|^2$ to these non-Euclidean spaces.

Definition II.1 (Canonical divergence [3])**.** *For any two points $u, x \in \mathbb{R}^n$, the canonical divergence on $(\mathbb{R}^n, g, \Gamma, \Gamma^\star)$ is given by*

$$D_\phi(u,x) = \int_0^1 t\dot{\gamma_t}^\top g(\gamma_t)\dot{\gamma}_t\mathrm{d}t \tag{4}$$

where $\gamma_t = u + t(x-u)$ is the Γ-geodesic connecting $u \to x$.

A dual canonical divergence $D^\star_\psi$ w.r.t. η is defined by using $\psi^\star$ and the $\Gamma^\star$-geodesic, and verifies the duality $D^\star_\phi(\eta_u, \eta_x) = D_\phi(x,u)$. Also, in the Euclidean case $D_\phi(u,x) = \|u-x\|^2/2$ as expected.

B. Differential-geometric derivation of MAP estimation

Theorem II.1 (Canonical Bayesian estimator)**.** *The canonical divergence on the Riemannian manifold $(\mathbb{R}^n, g, \Gamma, \Gamma^\star)$ induced by $\phi(x) = -\log p(x|y)$ is the Bregman divergence*

$$D_\phi(u,x) = \phi(u) - \phi(x) - \nabla\phi(x)(u-x).$$

In addition, the Bayesian estimator associated with D_ϕ is unique and is given by the maximum-a-posteriori estimator, i.e.,

$$\hat{x}_{D_\phi} \triangleq \operatorname*{argmin}_{u\in\mathbb{R}^n} \mathrm{E}_{x|y}[D_\phi(u,x)], \tag{5}$$

$$= \operatorname*{argmin}_{x\in\mathbb{R}^n} \phi(x) \tag{6}$$

$$= \hat{x}_{MAP}. \tag{7}$$

Proof. The proof of Theorem II.1 is reported in [4].

Theorem II.1 provides several valuable new insights into MAP estimation. It establishes that MAP estimation stems from Bayesian decision theory, and that the definition $\hat{x}_{MAP} = \operatorname{argmax}_{x\in\mathbb{R}^n} p(x|y)$ is mainly algorithmic. Hence, the predominant view of MAP estimation as computationally efficient pseudo-estimation is fundamentally incorrect. In fact, MAP estimation provides a general strategy to derive a loss function and a Bayes estimator that are tailored for $p(x|y)$; this is in sharp contrast with minimum mean square error (MMSE) estimation, which provides a general strategy to approximate quadratically any strongly convex loss function and its estimator. Theorem II.1 also presents an interesting interpretation of MMSE estimation as an approximation of MAP estimation where the manifold $(\mathbb{R}^n, g, \Gamma, \Gamma^\star)$ is approximated by an Euclidean space.

References

[1] C. P. Robert, The Bayesian Choice (second edition). New-York: Springer Verlag, 2001.
[2] S. Amari and H. Nagaoka, Methods of Information Geometry (Translations of Mathematical Monographs), v. 191. American Mathematical Society, 2000.
[3] A. Nihat and S. Amari, "A novel approach to canonical divergences within information geometry," Entropy, vol. 17, no. 12, pp. 8111–8129, 2015.
[4] M. Pereyra, "Canonical Bayesian estimation in convex models: a decision-theoretic and differential-geometric derivation of maximum-a-posteriori estimators," SIAM J. Imaging Sc., in preparation.

A multi–component method for high–dynamic range image deconvolution

Marco Prato*, Andrea La Camera†, Patrizia Boccacci† and Mario Bertero†
* Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio Emilia, Italy
† Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi, Università di Genova, Italy

***Abstract*—Deconvolving an image might be a particularly challenging problem when the object consists of very bright sources superimposed to diffuse structures, a common situation in astronomical imaging. In this work we propose a solution based on the representation of the image as the sum of a pointwise component and a smooth one, with different regularization for the two components, and we address the corresponding minimization problem by means of the scaled gradient projection method. Application to the deconvolution of high-dynamic range Poisson images is considered.**

I. The deconvolution model

We consider a standard Poisson image deconcolution problem and we assume that the unknown target $\boldsymbol{f}$ is the sum of a point-wise part $\boldsymbol{f}_P$ and an extended and smooth one $\boldsymbol{f}_E$. We also assume that the sources are localized inside small regions so that one can construct a mask which is 1 on these regions and 0 elsewhere. The resulting minimization problem becomes

$$\min_{(\boldsymbol{f}_E,\boldsymbol{f}_P)\in\overline{\Omega}} J_\beta(\boldsymbol{f}_E,\boldsymbol{f}_P;\boldsymbol{g}) \equiv J_0(\boldsymbol{f}_E+\boldsymbol{f}_P;\boldsymbol{g}) + \beta J_1(\boldsymbol{f}_E) \quad , \quad (1)$$

where $\overline{\Omega}$ is the set of non–negative couples $(\boldsymbol{f}_E,\boldsymbol{f}_P)$ with $\boldsymbol{f}_P$ equal to 0 outside a prefixed subregion P of the image and object domain S and such that $\boldsymbol{f}=\boldsymbol{f}_E+\boldsymbol{f}_P$ satisfies the flux constraint

$$\sum_{\boldsymbol{n}\in S} \boldsymbol{f}(\boldsymbol{n}) = c \quad , \quad c = \sum_{\boldsymbol{m}\in S} [\boldsymbol{g}(\boldsymbol{m}) - \boldsymbol{b}(\boldsymbol{m})] \quad ; \qquad (2)$$

$\boldsymbol{g}$ is the detected image; the background $\boldsymbol{b}$ is the known expected value of the sky emission; the data-fidelity function $J_0(\boldsymbol{f}_E+\boldsymbol{f}_P;\boldsymbol{g})$ is the generalized Kullback–Leibler (KL) divergence [1] between detected and computed images; the positive parameter β plays the role of a regularization parameter; and the regularization function $J_1(\boldsymbol{f}_E)$ is the Markov random field (MRF) operator defined as

$$J_1(\boldsymbol{f}) = \frac{1}{2}\sum_{\boldsymbol{n}\in S}\sum_{\boldsymbol{n}'\in\mathcal{N}(\boldsymbol{n})} \sqrt{\delta^2 + \left(\frac{\boldsymbol{f}(\boldsymbol{n})-\boldsymbol{f}(\boldsymbol{n}')}{\epsilon(\boldsymbol{n}')}\right)^2} \quad ,$$

where $\delta>0$, $\mathcal{N}(\boldsymbol{n})$ is a symmetric neighborhood made up of the eight first neighbors of $\boldsymbol{n}$ and $\epsilon(\boldsymbol{n}')$ is equal to 1 for the horizontal and vertical neighbors and equal to $\sqrt{2}$ for the diagonal ones.

II. Multi–component scaled gradient projection method

Problem (1) is addressed by means of the scaled gradient projection (SGP) method [2], [3], born as a natural way to accelerate the split gradient method (SGM) proposed by Lanteri [4] by introducing variable step-lengths and projections. In its general form, SGP can be applied to the minimization of any smooth objective function subject to a feasible set on which the projection is fast to compute, as in the case of box (possibly plus an equality) constraint. Feasibility of the iterations and stationarity of the limit points of the sequence are achieved by a projection on the constraints P_Ω and a line-search parameter λ_k automatically detected by means of a monotone Armijo backtracking rule, thus resulting in the iteration

$$\boldsymbol{f}^{(k+1)} = \boldsymbol{f}^{(k)} + \lambda_k \left(P_\Omega(\boldsymbol{f}^{(k)} - \alpha_k D_k \nabla J_\beta(\boldsymbol{f}^{(k)};\boldsymbol{g})) - \boldsymbol{f}^{(k)} \right),$$

Here the matrix D_k is defined according to the SGM strategy and the step-length parameter α_k is the one described e.g. in [2] and based on the Barzilai–Borwein rules.

III. Numerical results

We simulate a 256×256 Io-like object by generating a disc with the same diameter of Io as observed by Keck [5] and a smoothly variable brightness, including a sort of limb darkening; we superimpose to the disc very bright sources and we convolve the result with a PSF modeling the Keck PSF in M-band. It is obtained from the K-band PSF of Keck provided with the Io images. Finally the result is perturbed by Poisson noise. The positions of the bright sources (and the related mask) are obtained by a rough reconstruction obtained by running SGP on a standard KL + MRF minimization (here $\beta=10^{-4}$ and $\delta=500$). Then, SGP is used again to address problem (1) and obtain an estimate of the smooth surface f_E (here $\beta=10^{-2}$ and $\delta=10$), which is used as background in a last SGP run on the minimization of the non–regularized KL divergence in order to better reconstruct the bright spots f_P. The results are shown in Figure 1, in which the blurred and noisy image, the surface reconstructed after the second minimization and the complete reconstruction are provided.

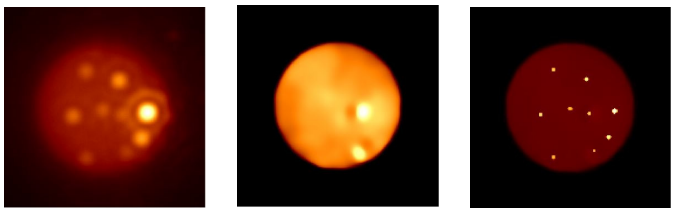


Fig. 1. Blurred and noisy image (left), surface reconstruction (middle) and final image (right).

References

[1] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *Ann. Stat.*, vol. 19, no. 4, pp. 2032–2066, 1991.
[2] S. Bonettini, R. Zanella, and L. Zanni, "A scaled gradient projection method for constrained image deblurring," *Inverse Probl.*, vol. 25, no. 1, p. 015002, 2009.
[3] S. Bonettini and M. Prato, "New convergence results for the scaled gradient projection method," *Inverse Probl.*, vol. 31, no. 9, p. 095008, 2015.
[4] H. Lantéri, M. Roche, and C. Aime, "Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms," *Inverse Probl.*, vol. 18, no. 5, pp. 1397–1419, 2002.
[5] I. de Pater, C. Laver, A. G. Davies, K. de Kleer, D. A. Williams, R. R. Howell, J. A. Rathbun, and J. R. Spencer, "Io: Eruptions at Pillan, and the time evolution of Pele and Pillan from 1996 to 2015," *Icarus*, vol. 264, pp. 198–212, 2016.

Stochastic Reformulations of Linear Systems and Fast Randomized Iterative Methods

Peter Richtárik*, Martin Takáč†
* School of Mathematics, The University of Edinburgh, United Kingdom
† Department of Industrial and Systems Engineering, Lehigh University, USA

***Abstract*—We propose a new paradigm for solving linear systems. In our paradigm, the system is reformulated into a *stochastic problem*, and then solved with a randomized algorithm. Our reformulation can be equivalently seen as a stochastic optimization problem, stochastically preconditioned linear system, stochastic fixed point problem and as a probabilistic intersection problem. We propose and analyze basic and accelerated stochastic algorithms for solving the reformulated problem, with linear convergence rates.**

I. Introduction

In this work we are concerned with the problem of solving a consistent linear system. In particular, consider the problem

$$\text{solve} \quad \mathbf{A}x = b, \tag{1}$$

where $0 \neq \mathbf{A} \in \mathbb{R}^{m\times n}$. We shall assume throughout the paper that the system is consistent, i.e., $\mathcal{L} \overset{\text{def}}{=} \{x \,:\, \mathbf{A}x = b\} \neq \emptyset$.

II. Stochastic Reformulations

We propose a fundamental and flexible way of reformulating each consistent linear system into a *stochastic problem*. To the best of our knowledge, this is the first systematic study of such reformulations. Stochasticity is introduced in a controlled way into an otherwise deterministic problem, as a decomposition tool which can be leveraged to design efficient, granular and scalable randomized algorithms. In particular, we consider a user-defined distribution $\mathcal{D}$ describing an ensemble of random matrices $\mathbf{S} \in \mathbb{R}^{m\times q}$. We make use of one more parameter: a user-defined $n \times n$ symmetric positive definite matrix $\mathbf{B}$. Our reformulation of (1) as a stochastic problem has several seemingly different, yet equivalent interpretations:

1) **Stochastic optimization problem.** Consider the problem

$$\text{minimize} \quad f(x) \overset{\text{def}}{=} \mathrm{E}_{\mathbf{S}\sim\mathcal{D}}\left[f_{\mathbf{S}}(x)\right], \tag{2}$$

where $f_{\mathbf{S}}(x) = \frac{1}{2}(\mathbf{A}x - b)^\top \mathbf{H}(\mathbf{A}x - b)$, $\mathbf{H} = \mathbf{S}(\mathbf{S}^\top\mathbf{A}\mathbf{B}^{-1}\mathbf{A}^\top\mathbf{S})^\dagger\mathbf{S}^\top$, and $\dagger$ denotes the Moore-Penrose pseudoinverse. When solving the problem, we do not have explicit access to f, its gradient or Hessian. Rather, we can repeatedly sample $\mathbf{S} \sim \mathcal{D}$ and receive unbiased samples of these quantities at points of interest.

2) **Stochastically preconditioned linear system.** We consider what we call a *stochastically preconditioned* version of (1), namely

$$\text{solve} \quad \mathbf{B}^{-1}\mathbf{A}^\top \mathrm{E}_{\mathbf{S}\sim\mathcal{D}}\left[\mathbf{H}\right]\mathbf{A}x = \mathbf{B}^{-1}\mathbf{A}^\top \mathrm{E}_{\mathbf{S}\sim\mathcal{D}}\left[\mathbf{H}\right]b. \tag{3}$$

The preconditioner, $\mathbf{P} \overset{\text{def}}{=} \mathbf{B}^{-1}\mathbf{A}^\top \mathrm{E}_{\mathbf{S}\sim\mathcal{D}}\left[\mathbf{H}\right]$, is not assumed to be known explicitly. Instead, when solving the problem, we are able to sample $\mathbf{S} \sim \mathcal{D}$, obtaining an unbiased estimate of the preconditioner (not necessarily explicitly), $\mathbf{B}^{-1}\mathbf{A}^\top\mathbf{H}$, for which we coin the name *stochastic preconditioner*. This gives us access to a random sample of system (3): $\mathbf{B}^{-1}\mathbf{A}^\top\mathbf{H}\mathbf{A}x = \mathbf{B}^{-1}\mathbf{A}^\top\mathbf{H}b$. This information can be obtained by repeatedly querying the stochastic sampling $\mathbf{S} \sim \mathcal{D}$ and utilized by an iterative algorithm.

3) **Stochastic fixed point problem.** Let $\Pi^{\mathbf{B}}_{\mathcal{L}_{\mathbf{S}}}(x)$ denote the projection of x onto $\mathcal{L}_{\mathbf{S}} \overset{\text{def}}{=} \{x \,:\, \mathbf{S}^\top\mathbf{A}x = \mathbf{S}^\top b\}$, in the norm $\|x\|_{\mathbf{B}} \overset{\text{def}}{=} \sqrt{x^\top\mathbf{B}x}$. Consider the *stochastic fixed point problem*

$$\text{solve} \quad x = \mathrm{E}_{\mathbf{S}\sim\mathcal{D}}\left[\Pi^{\mathbf{B}}_{\mathcal{L}_{\mathbf{S}}}(x)\right]. \tag{4}$$

That is, we seek to find a *fixed point* of the mapping $x \to \mathrm{E}_{\mathbf{S}\sim\mathcal{D}}\left[\Pi^{\mathbf{B}}_{\mathcal{L}_{\mathbf{S}}}(x)\right]$. When solving the problem, we do not have an explicit access to the average projection map. Instead, we are able to repeatedly sample $\mathbf{S} \sim \mathcal{D}$, and use the stochastic projection map $x \to \Pi^{\mathbf{B}}_{\mathcal{L}_{\mathbf{S}}}(x)$.

4) **Probabilistic intersection problem.** Consider the problem:

$$\text{find} \quad x \in \cap_{\mathbf{S}\sim\mathcal{D}}\mathcal{L}_{\mathbf{S}} \overset{\text{def}}{=} \{x \,:\, \mathrm{Prob}(x \in \mathcal{L}_{\mathbf{S}}) = 1\}. \tag{5}$$

As before, we typically do not have an explicit access to the probabilistic intersection when designing an algorithm. Instead, we can repeatedly sample $\mathbf{S} \sim \mathcal{D}$, and utilize the knowledge of $\mathcal{L}_{\mathbf{S}}$ to drive the iterative process. If $\mathcal{D}$ is a discrete distribution, probabilistic intersection reduces to standard intersection.

Theorem 1. *The four stochastic formulations are equivalent.*

With all of the above reformulations we associate the same *condition number*. Letting $\mathbf{W} \overset{\text{def}}{=} \mathbf{B}^{-1/2}\mathbf{A}^\top \mathrm{E}\left[\mathbf{H}\right]\mathbf{A}\mathbf{B}^{-1/2}$, we define the condition number as $\kappa = \kappa(\mathbf{A},\mathbf{B},\mathcal{D}) \overset{\text{def}}{=} \|\mathbf{W}\|\|\mathbf{W}^\dagger\| = \lambda_{\max}/\lambda^+_{\min}$, where $\|\cdot\|$ is the spectral norm, $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{W}$ and $\lambda^+_{\min}$ is the smallest nonzero eigenvalue of $\mathbf{W}$. Let $\mathcal{X}$ be the set of solutions of any of the reformulations. We now give necessary and sufficient conditions for this set to be equal to $\mathcal{L}$.

Theorem 2. $\mathcal{L} = \mathcal{X} \quad \Leftrightarrow \quad \mathrm{Null}\left(\mathrm{E}\left[\mathbf{A}^\top\mathbf{H}\mathbf{A}\right]\right) = \mathrm{Null}\left(\mathbf{A}\right).$

III. Algorithms

We propose several algorithms for solving the reformulations. Our *basic method* has the form

$$x_{k+1} = \phi_\omega(x_k, \mathbf{S}_k) \overset{\text{def}}{=} x_k - \omega\mathbf{B}^{-1}\mathbf{A}^\top\mathbf{H}_k(\mathbf{A}x_k - b), \tag{6}$$

where $\mathbf{S}_k \sim \mathcal{D}$ is sampled afresh in each iteration. This method can be interpreted as *stochastic gradient descent* and *stochastic Newton descent*, with stepsize ω, applied to the stochastic optimization problem; as a stochastic fixed point method, and as a *stochastic projection method.* Our *accelerated method* can be written as

$$x_{k+1} = \gamma\phi_\omega(x_k, \mathbf{S}_k) + (1-\gamma)\phi_\omega(x_{k-1}, \mathbf{S}_{k-1}), \tag{7}$$

where the matrices $\{\mathbf{S}_k\}$ are independent samples from $\mathcal{D}$, and $\gamma \in \mathbb{R}$ is an *acceleration parameter*. Our theory suggests that γ should be always between 1 and 2. In particular, for well conditioned problems (small κ), one should choose $\gamma \approx 1$, and for ill conditioned problems (large κ), one should choose $\gamma \approx 2$.

Theorem 3. *For suitable parameters ω and γ, the basic (resp. accelerated) method converges linearly in expectation, with rate $\mathcal{O}(\kappa)$ (resp. $\mathcal{O}(\sqrt{\kappa})$).*

Phase Retrieval meets Statistical Learning Theory

Justin Romberg*, Sohail Bahmani*
* School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA

***Abstract*—We will present a new convex relaxation for the phase retrieval problem based on linear (in the real case) or quadratic (in the complex case) programming. The number of variables in the convex relaxation is the same as the dimension of the signal, so we avoid the increase in dimensionality inherent to lifting schemes. Our method has a clear geometric interpretation: by relaxing each magnitude measurement into a pair of inequality constraints, we know that the signal of interest lies on the surface of the polytope defined by the intersection of these constraints. We show that even a rough guess of where the signal lies (a guess that can itself be formed from the measured data) is enough to define a linear functional over this polytope whose maximum is at the true signal. Our analysis uses classical results from statistical learning theory, in particular the VC dimension and generalization bound.**

Half-Quadratic Image Models: From Sampling to Discriminative Training

Uwe Schmidt*, Qi Gao†, and Stefan Roth‡
* Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
† CellNetworks, Heidelberg University, Heidelberg, Germany
‡ Department of Computer Science, TU Darmstadt, Darmstadt, Germany

***Abstract*—Since their introduction in the early 90s, half-quadratic regularization approaches have become a popular tool for modeling a-priori information about images and, more generally, visual scenes. For the most part, their popularity stems from efficient alternating optimization schemes that arise directly from their model structure. This talk will give a summary of our work on two other uses of half-quadratic schemes. First, we will show how they can be used to derive efficient sampling schemes for learning the prior and for performing posterior inference. This enables concrete applications to denoising and deblurring, and allows a Bayesian treatment of nuisance parameters such as the noise strength. Second, we will demonstrate how half-quadratic optimization can be turned into multi-stage discriminative architectures that allow for loss-based training. The resulting shrinkage fields combine ideas from random fields, shrinkage, and discriminative learning into an efficient model architecture that scales well to megapixel images.**

I. Introduction

Half-quadratic approaches to image restoration were introduced by Geman and colleagues over 20 years ago [1], [2]. Their basic idea is to augment the image prior $p(\mathbf{x})$ with auxiliary variables $\mathbf{z}$ into a joint distribution $p(\mathbf{x}, \mathbf{z})$ such that the image prior can be obtained by maximizing or marginalizing over the auxiliary variables. Maximization over $\mathbf{z}$, i.e. $p(\mathbf{x}) = \max_{\mathbf{z}} p(\mathbf{x}, \mathbf{z})$, gives rise to the popular envelope type, while marginalization, i.e. $p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{z})\, d\mathbf{z}$, yields the integral type. The name half-quadratic stems from the fact that the augmentation is chosen such that the model energy $E(\mathbf{x}, \mathbf{z})$ with $p(\mathbf{x}, \mathbf{z}) \propto \exp(-E(\mathbf{x}, \mathbf{z}))$ is quadratic in $\mathbf{x}$ (but not in $\mathbf{z}$). The crucial benefit is that this enables inference schemes that alternate between updating $\mathbf{x}$ and updating $\mathbf{z}$, which build the basis of a number of widely used efficient image restoration schemes, *e. g.* [3], [4].

II. Half-Quadratic Sampling

In the first half of this talk, we will turn to the less widely used integral type. We exploit that the half-quadratic formalism allows setting up an auxiliary variable Gibbs sampler that alternates between sampling from $p(\mathbf{x}|\mathbf{z})$ and $p(\mathbf{z}|\mathbf{x})$. Its benefit is threefold: Since $p(\mathbf{x}|\mathbf{z})$ is Gaussian (following from the quadratic model energy), sampling is rather efficient mainly involving solving linear equation systems. Second, as the individual auxiliary variables z_i are conditionally independent given $\mathbf{x}$, sampling $p(\mathbf{z}|\mathbf{x})$ is straightforward. Finally, this sampling scheme mixes much more rapidly than traditional single-site Gibbs samplers or Hamiltonian Monte Carlo.

We first explore sampling image priors based on pairwise and high-order Markov random fields, which can be used to learn faithful image models [5], [6] as well as evaluate how well the resulting models reproduce key statistical properties of (photographic) images. Next, we show how the auxiliary variable sampler can be used for sampling the posterior in image denoising [5], [6] and deblurring [7] applications, which allows computing a Bayesian mean squared error estimate. Finally, the sampler can be extended to nuisance variables, which allows for a Bayesian treatment of the noise strength or the point spread function [7].

III. Shrinkage Fields

In the second half of the talk, we will turn to applications of half-quadratic schemes for energy minimization that enable highly efficient and effective discriminatively-trained image restoration models. Their efficiency stems from using the so-called additive envelope type, which when combined with suitable boundary conditions enables diagonalizing the system matrix of the quadratic part of the energy using the FFT. The effectiveness on the other hand stems from unrolling the half-quadratic optimization into a fixed set of stages, replacing the optimization over certain potential functions using a more powerful shrinkage operation, and training individual shrinkage functions and image filters for each stage in a discriminative end-to-end fashion.

The resulting *shrinkage fields* [8] combine ideas from classical shrinkage with iterative inference in random fields and discriminative training. One benefit over "black-box" discriminative learning approaches is that a connection to the original generative image model is retained, which for example allows adapting the approach to different image formation parameters (*e. g.*, point spread functions) at test time without retraining. Moreover, the approach scales well to large image sizes and yields highly competitive restoration quality.

Acknowledgements

This work was done while Uwe Schmidt and Qi Gao were with TU Darmstadt. The authors would like to thank Kevin Schelten for his contributions to sampling-based deblurring. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013) / ERC Grant Agreement No. 307942.

References

[1] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE T. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.
[2] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE T. Image Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.
[3] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE T. Image Process.*, vol. 6, no. 2, pp. 298–311, Feb. 1997.
[4] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for Total Variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.
[5] U. Schmidt, Q. Gao, and S. Roth, "A generative perspective on MRFs in low-level vision," in *CVPR*, 2010, pp. 1751–1758.
[6] Q. Gao and S. Roth, "How well do filter-based MRFs model natural images?" in *DAGM*, 2012, pp. 62–72.
[7] U. Schmidt, K. Schelten, and S. Roth, "Bayesian deblurring with integrated noise estimation," in *CVPR*, 2011, pp. 2625–2632.
[8] U. Schmidt and S. Roth, "Shrinkage fields for effective image restoration," in *CVPR*, 2014, pp. 2774–2781.

Denoising-based Vector AMP

Philip Schniter*, Sundeep Rangan†, and Alyson Fletcher‡
* Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH.
† Department of Electrical and Computer Engineering, New York University, Brooklyn, NY.
‡ Departments of Statistics, Mathematics, and Electrical Engineering, University of California, Los Angeles, CA.

***Abstract*—The D-AMP methodology, recently proposed by Metzler, Maleki, and Baraniuk, allows one to plug in sophisticated denoisers like BM3D into the AMP algorithm to achieve state-of-the-art compressive image recovery. But AMP diverges with small deviations from the i.i.d.-Gaussian assumption on the measurement matrix. Recently, the VAMP algorithm has been proposed to fix this problem. In this work, we show that the benefits of VAMP extend to D-VAMP.**

Consider the problem of recovering a (vectorized) image $\boldsymbol{x}_0 \in \mathbb{R}^N$ from compressive (i.e., $M \ll N$) noisy linear measurements

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}_0 + \boldsymbol{w} \in \mathbb{R}^M, \tag{1}$$

known as "compressive imaging." The "sparse" approach to this problem exploits sparsity in the coefficients $\boldsymbol{v}_0 \triangleq \boldsymbol{\Psi}\boldsymbol{x}_0 \in \mathbb{R}^N$ of an orthonormal wavelet transform $\boldsymbol{\Psi}$. The idea is to rewrite (1) as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{v}_0 + \boldsymbol{w} \quad \text{for} \quad \boldsymbol{A} \triangleq \boldsymbol{\Phi}\boldsymbol{\Psi}^{\mathsf{T}}, \tag{2}$$

recover an estimate $\widehat{\boldsymbol{v}}$ of $\boldsymbol{v}_0$ from $\boldsymbol{y}$, and then construct the image estimate as $\widehat{\boldsymbol{x}} = \boldsymbol{\Psi}^{\mathsf{T}}\widehat{\boldsymbol{v}}$.

Although many algorithms have been proposed for sparse recovery of $\boldsymbol{v}_0$, a notable one is the approximate message passing (AMP) algorithm from [1]. It is computationally efficient (i.e., one multiplication by $\boldsymbol{A}$ and $\boldsymbol{A}^{\mathsf{T}}$ per iteration and relatively few iterations) and its performance, when M and N are large and $\boldsymbol{\Phi}$ is zero-mean i.i.d. Gaussian, is rigorously characterized by a scalar state evolution.

A variant called "denoising-based AMP" (D-AMP) was recently proposed [2] for *direct* recovery of $\boldsymbol{x}_0$ from (1). It exploits the fact that, at iteration t, AMP constructs a pseudo-measurement of the form $\boldsymbol{v}_0 + \mathcal{N}(\boldsymbol{0}, \sigma_t^2\boldsymbol{I})$ with known σ_t^2, which is amenable to any image denoising algorithm. By plugging in a state-of-the-art image denoiser like BM3D [3], D-AMP yields state-of-the-art compressive imaging.

AMP and D-AMP, however, have a serious weakness: they diverge under small deviations from the zero-mean i.i.d. Gaussian assumption on $\boldsymbol{\Phi}$, such as non-zero mean or mild ill-conditioning. A robust alternative called "vector AMP" (VAMP) was recently proposed [4]. VAMP has similar complexity to AMP and a rigorous state evolution that holds under right-rotationally invariant $\boldsymbol{\Phi}$—a much larger class of matrices. Although VAMP needs to know the variance of the measurement noise $\boldsymbol{w}$, an auto-tuning method was proposed in [5].

In this work, we integrate the D-AMP methodology from [2] into auto-tuned VAMP from [5], leading to "D-VAMP." (For a matlab implementation, see http://dsp.rice.edu/software/DAMP-toolbox.)

To test D-VAMP, we recovered the 128×128 *lena*, *barbara*, *boat*, *fingerprint*, *house*, and *peppers* images using 10 realizations of $\boldsymbol{\Phi}$. Table I shows that, for i.i.d. Gaussian $\boldsymbol{\Phi}$, the average PSNR and runtime of D-VAMP is similar to D-AMP at medium sampling ratios. The PSNRs for $\boldsymbol{v}$-based indirect recovery, using Lasso (i.e., "ℓ_1")-based AMP and VAMP, are significantly worse. At small sampling ratios, D-VAMP behaves better than D-AMP, as shown in Fig. 1.

To test robustness to ill-conditioning in $\boldsymbol{\Phi}$, we constructed $\boldsymbol{\Phi} = \boldsymbol{JSPFD}$, with $\boldsymbol{D}$ a diagonal matrix of random ± 1, $\boldsymbol{F}$ a (fast) Hadamard matrix, $\boldsymbol{P}$ a random permutation matrix, and $\boldsymbol{S} \in \mathbb{R}^{M\times N}$ a diagonal matrix of singular values. The sampling rate was fixed at $M/N = 0.1$, the noise variance chosen to achieve SNR=32 dB, and the singular values were geometric, i.e., $s_i/s_{i-1} = \rho \ \forall i > 1$, with ρ chosen to yield a desired condition number. Table II shows that (D-)AMP breaks when the condition number is ≥ 10, whereas (D-)VAMP shows only mild degradation in PSNR (but not runtime).

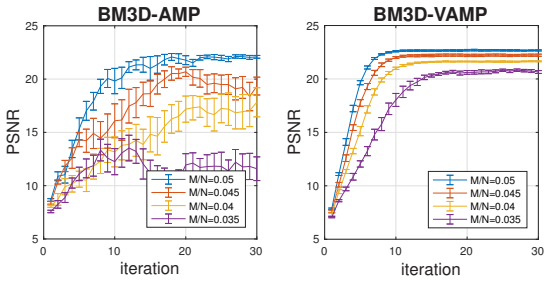


Fig. 1. PSNR versus iteration at several sampling ratios M/N for i.i.d. Gaussian $\boldsymbol{A}$.

TABLE I
AVERAGE PSNR AND RUNTIME FROM MEASUREMENTS WITH I.I.D. GAUSSIAN MATRICES AND ZERO NOISE AFTER 30 ITERATIONS

sampling ratio	10%		20%		30%		40%		50%	
	PSNR	time	PSNR	time	PSNR	time	PSNR	time	PSNR	time
ℓ_1-AMP	**17.7**	**0.5s**	**20.2**	1.0s	**22.4**	1.6s	24.6	2.3s	27.0	3.1s
ℓ_1-VAMP	17.6	**0.5s**	**20.2**	**0.9s**	**22.4**	**1.4s**	**24.8**	**1.8s**	**27.2**	**2.3s**
BM3D-AMP	**25.2**	**10.1s**	**30.0**	8.8s	**32.5**	8.6s	35.1	9.1s	37.4	9.8s
BM3D-VAMP	**25.2**	10.4s	**30.0**	**8.5s**	**32.5**	**8.2s**	**35.2**	**8.5s**	**37.7**	**8.8s**

TABLE II
AVERAGE PSNR AND RUNTIME FROM MEASUREMENTS WITH DHT-BASED MATRICES AND SNR=32 DB AFTER 10 ITERATIONS

condition no.	1		10		10^2		10^3		10^4	
	PSNR	time	PSNR	time	PSNR	time	PSNR	time	PSNR	time
ℓ_1-AMP	17.3	**0.02**	<0	—	<0	—	<0	—	<0	—
ℓ_1-VAMP	**17.4**	0.04	**17.4**	**0.04**	**15.6**	**0.03**	**14.7**	**0.03**	**14.4**	**0.03**
BM3D-AMP	**24.8**	**5.2s**	8.0	—	7.2	—	7.1	—	7.2	—
BM3D-VAMP	**24.8**	5.4s	**24.3**	**5.5s**	**22.6**	**5.3s**	**21.4**	**4.9s**	**20**	**4.5s**

REFERENCES

[1] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.
[2] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "From denoising to compressed sensing," *IEEE Trans. Inform. Theory*, vol. 62, no. 9, pp. 5117–5144, 2016.
[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.
[4] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *arXiv:1610.03082*, 2016.
[5] A. K. Fletcher and P. Schniter, "Learning and free energies for vector approximate message passing," *arXiv:1602.08207*, 2016.

Multilinear Low-Rank Tensors on Graphs & Applications

Nauman Shahid*, Francesco Grassi†, and Pierre Vandergheynst*
* LTS2, Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland, † Politecnico di Torino, Italy

***Abstract*—Current low-rank tensor literature lacks development in large scale processing and generalization to graphs [1]. Motivated by the fact that the first few eigenvectors of the k_{nn}-nearest neighbors graph provide a smooth basis for the data, we propose a novel framework "Multilinear Low-Rank Tensors on Graphs (MLRTG)". The applications of our scalable and approximate method include approximate and fast methods for tensor compression, Robust PCA, completion and clustering. We specifically focus on Graph Multilinear SVD (GMLSVD) in this work.**

Multilinear Low-Rank Tensors on Graphs (MLRTG): A tensor $\boldsymbol{\mathcal{Y}}^* \in \mathbb{R}^{n\times n\times n}$ is said to be Multilinear Low-Rank on Graphs (MLRTG) if it can be encoded in terms of the lowest k Laplacian eigenvectors as:

$$\text{vec}(\boldsymbol{\mathcal{Y}}^*) = (P_{1k} \otimes P_{2k} \otimes P_{3k})\,\text{vec}(\boldsymbol{\mathcal{X}}^*), \tag{1}$$

where $\text{vec}(\cdot)$ denotes the vectorization, $\otimes$ denotes the kronecker product, $P_{\mu k} \in \mathbb{R}^{n\times k}, \forall \mu$ are the lowest k eigenvectors of k_{nn}-graphs constructed between the rows of the matricized versions Y_μ of $\boldsymbol{\mathcal{Y}}^*$ and $\boldsymbol{\mathcal{X}}^* \in \mathbb{R}^{k\times k\times k}$ is the *Graph Core Tensor (GCT)*. We call the tuple (k,k,k), where $k \ll n$, as the *Graph Multilinear Rank* of $\boldsymbol{\mathcal{Y}}^*$ and refer to a tensor from the set of all possible MLRTG as $\boldsymbol{\mathcal{Y}} \in \mathbb{MLT}$.

For any $\boldsymbol{\mathcal{Y}} \in \mathbb{MLT}$, the GCT $\boldsymbol{\mathcal{X}}$ is the most useful entity. For a clean matricized tensor Y_1 it is straight-forward to determine the matricized $\boldsymbol{\mathcal{X}}$ as $X_1 = P_{1k}^\top Y_1 P_{2,3k}$, where $P_{2,3k} = P_{1k} \otimes P_{2k} \in \mathbb{R}^{n^2\times k^2}$. For the case of noisy $\boldsymbol{\mathcal{Y}}$, one seeks a robust $\boldsymbol{\mathcal{X}}$ which is not possible without an appropriate regularization on $\boldsymbol{\mathcal{X}}$. Hence, we propose to solve the following convex minimization problem:

$$\min_{\boldsymbol{\mathcal{X}}} \|Y_1 - P_{1k}X_1P_{2,3k}^\top\|_F^2 + \gamma \sum_\mu \|X_\mu\|_{*g(\Lambda_{\mu k})}, \tag{2}$$

where $\|\cdot\|_* g(\cdot)$ denotes the weighted nuclear norm and $g(\Lambda_{\mu k}) = \Lambda_{\mu k}^a, a \geq 1$, denotes the kernelized Laplacian eigenvalues as the weights for the nuclear norm minimization. Assuming the eigenvalues are sorted in ascending order, this corresponds to a higher penalization of higher singular values of X_μ which correspond to noise. Such a nuclear norm minimization on the full tensor (without weights) has appeared in earlier works [2]. However, note that in our case we lift the computational burden by minimizing only the core tensor $\boldsymbol{\mathcal{X}}$. Using $Y_1 = P_{1k}\hat{X}_1P_{2,3k}^\top$ in eq. (2), we get:

$$\min_{\boldsymbol{\mathcal{X}}} \|\hat{X}_1 - X_1\|_F^2 + \gamma \sum_\mu \|X_\mu\|_{*g(\Lambda_{\mu k})}. \tag{3}$$

Notice that the decomposition (1) is quite similar to the standard Mulitlinear SVD (MLSVD) [3]. In standard MLSVD, one aims to decompose a tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n\times n\times n}$ into factors $U_\mu \in \mathbb{R}^{n\times r}$ which are linked by a core $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{r\times r\times r}$. In our case the decomposition is given in terms of the pre-computed Laplacian eigenvectors $P_{\mu k}$ and $\boldsymbol{\mathcal{X}}$ is determined by GCTP eq. (3).

Algorithm for GMLSVD: For a tensor $\boldsymbol{\mathcal{Y}}$, one can compute GMLSVD in the following steps: 1) Compute the graph core tensor $\boldsymbol{\mathcal{X}}$ via eq. (3), 2) Perform the MLSVD of $\text{vec}(\boldsymbol{\mathcal{X}}) = (A_{1k} \otimes A_{2k} \otimes A_{3k})\,\text{vec}(\boldsymbol{\mathcal{R}})$, 3) Let the factors $V_\mu = P_{\mu k}A_{\mu k}$ and the core tensor is $\boldsymbol{\mathcal{R}}$. Thus, the MLSVD of $\boldsymbol{\mathcal{Y}}$ is given as $\text{vec}(\boldsymbol{\mathcal{Y}}) = (P_{1k}A_{1k} \otimes P_{2k}A_{2k} \otimes P_{3k}A_{3k})\,\text{vec}(\boldsymbol{\mathcal{X}})$. The Laplacian eigenvectors $P_{\mu k}$ can be computed with a cost $\mathcal{O}(nk^2)$ via the Power Method. Hence, GMLSVD scales with $\mathcal{O}(nk^2 + k^4)$ per iteration. This is advantageous for applications such as tensor Robust PCA ($\mathcal{O}(n^4)$ per iteration) which constitute the work in progress. Theoretical results for this method are in progress.

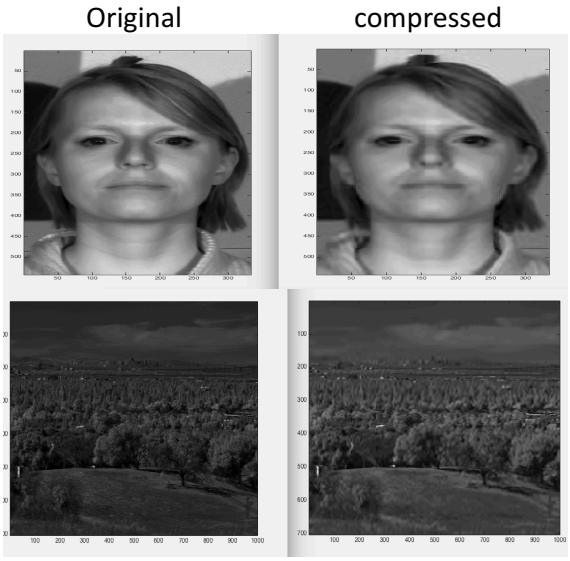


Figure 1. Compression of hyperspectral images via GMLSVD. Left plots show the grayscale image of the 1st spectral band of each dataset and the right show the same image after compression. Compression rates of 125 and 110 times are attained with an SNR of 25dB and 15dB for the two datasets.

Hyperspectral Image Compression : We report results for the compression of two hyperspectral image datasets collected from Stanford database: 1) face dataset $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{542\times 333\times 148}$, 250MB in size and 2) the landscape $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{702\times 1000\times 148}$, 650MB in size as shown in the left plots of Fig. 1. The right plots in each row show the 1st spectral band in grayscale for the compressed datasets. For the face dataset, we used a core tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{70\times 70\times 30}$ and achieved a compression of 125 times, while maintaining an SNR of 25dB. For the landscape dataset, we used a core tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{150\times 150\times 30}$ to achieve a compression of 110 times with an SNR of 15dB. Both datastes required less than 1 minute for compression.

References

[1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," Signal Processing Magazine, IEEE, vol. 30, no. 3, pp. 83–98, 2013. 1

[2] B. Huang, C. Mu, D. Goldfarb, and J. Wright, "Provable low-rank tensor recovery," Optimization-Online, vol. 4252, p. 2, 2014. 1

[3] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," SIAM review, vol. 51, no. 3, pp. 455–500, 2009. 1

Restoration of manifold-valued images by variational models

Miroslav Bačák*, Ronny Bergmann†, Ralf Hielscher‡ Friederike Laus†, Johannes Persch†, Gabriele Steidl† and Andreas Weinmann§
* Max Planck Institute for Mathematics in the Sciences, 04103 Leipzig, Germany.
† Department of Mathematics, University of Kaiserslautern, 67663 Kaiserslautern, Germany.
‡ Faculty of Mathematics, University of Chemnitz, 09107 Chemnitz, Germany.
§ University of Applied Sciences, 64295 Darmstadt, Germany.

***Abstract*—Recently non-smooth variational models were applied very successfully for the restoration and segmentation of images. One of the most popular models was proposed by Rudin, Osher and Fatemi in 1992 which was meanwhile improved and generalized by many authors. Nowadays we are confronted with „images" taking values in a Riemannian manifold. We show how various variational models, in particular those including second order differences can be generalized to the manifold-valued setting. We suggest certain algorithms to find the minimizer of the corresponding functions among them the cyclic proximal point algorithm, the Douglas-Rachford algorithm and half-quadratic minimization.**

I. Introduction

In various applications in image processing and computer vision the functions of interest take values in a Riemannian manifold. Examples are diffusion tensor imaging (DT-MRI) or texture processing with covariance matrices of Gaussian distributions where the data lives in the Riemannian manifold of positive definite matrices. Circular and spherical data appear in synthetic aperture radar (SAR), in color image processing based on non-flat color models or whenever directional information is handled. The motion group $\mathrm{SE}(3)$ and the rotation group $\mathrm{SO}(3)$ play a role in tracking, robotics, (scene) motion analysis and electron backscattered diffraction (EBSD). Due to the natural appearance of such nonlinear data spaces, processing manifold-valued data has gained a lot of interest in recent years.

We introduce a non-smooth variational model for the restoration of manifold-valued images using first and second order differences. The model can be seen as a second order generalization of the Rudin-Osher-Fatemi (ROF) functional [1] for images taking their values in a Riemannian manifold. For real-valued images, its discrete, anisotropic penalized form is given by

$$\mathcal{D}(u;f) + \alpha \mathrm{TV}(u) \tag{1}$$

where the *data fidelity term*

$$\mathcal{D}(u;f) := \frac{1}{2}\sum_{i,j} |f_{i,j} - u_{i,j}|^2,$$

measures the similarity between the wanted image u and the given data $f := (f_{i,j}) \in R^{N,M}$, and the total variation type *regularizing term*

$$\mathrm{TV}(u) := \alpha \sum_{i,j} \left(|u_{i,j} - u_{i+1,j}| + |u_{i,j} - u_{i,j+1}| \right),$$

takes care that important features such as edges are preserved. The regularization parameter $\alpha > 0$ steers the relation between both terms. Unfortunately, the model tends to produce staircasing: instead of reconstructing smooth areas as such, the reconstruction consists of constant plateaus with small jumps. An approach for avoiding this effect incorporates second order differences. For an overview we refer to [2].

Recently, primal-dual splitting algorithms were successfully applied in image processing mainly for two reasons: the functionals to minimize allow for simple proximal mappings within the method, and it turned out that the algorithms are highly parallelizable, see, e.g., [3]. Therefore these algorithms are among the most popular ones in variational image processing. However, it appears to be a hard task to transfer them to the manifold-valued setting. In this paper, we show how some approaches as the cyclic proximal point algorithm, the Douglas-Rachford splitting algorithm and the half quadratic minimization algorithm can be generalized to manifolds. Various numerical examples give a powerful proof of the concept.

II. Variational model

Let $\mathcal{M}$ be a complete, connected n-dimensional Riemannian manifold with geodesic distance $d_{\mathcal{M}}\colon \mathcal{M} \times \mathcal{M} \to R_{\geq 0}$. Given a noisy image $f \in \mathcal{M}^{N,M}$ we are looking for a denoised image

$$\mathcal{D}(u,f) := \frac{1}{2}\sum_{i,j} d_{\mathcal{M}}(f_{i,j}, u_{i,j})^2 + \alpha \mathrm{TV}(u)$$

and

$$\mathrm{TV}(u) := \sum_{i,j} \left(d_{\mathcal{M}}(u_{i,j}, u_{i+1,j}) + d_{\mathcal{M}}(u_{i,j}, u_{i,j+1}) \right).$$

More precisely, we update the above model by introducing another regularizing term which contains second order differences of manifold-valued samples and provide numerical algorithms for finding minimizers of these functions. The approach can be generalized to images with missing pixels. Details can be found in our papers [4], [5], [6], [7].

References

[1] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1, pp. 259–268, 1992.
[2] G. Steidl, "Combined first and second order variational approaches for image processing," *Jahresbericht der Deutschen Mathematiker-Vereinigung 2015*, vol. 117, no. 2, pp. 133–160, 2015.
[3] P. L. Combettes and J.-C. Pesquet, "A proximal decomposition method for solving convex variational inverse problems," *Inverse Problems*, vol. 24, no. 6, 2008.
[4] M. Bačák, R. Bergmann, G. Steidl, and A. Weinmann, "A second order non-smooth variational model for restoring manifold-valued images," *SIAM Journal on Scientific Computing*, vol. 38, no. 1, pp. 567 – 597, 2016.
[5] R. Bergmann, R. H. Chan, R. Hielscher, J. Persch, and G. Steidl, "Restoration of manifold-valued images by half-quadratic minimization," *Inverse Problems and Imaging*, vol. 10, no. 2, pp. 281–304, 2016.
[6] R. Bergmann, F. Laus, G. Steidl, and A. Weinmann, "Second order differences of cyclic data and applications in variational denoising," *SIAM Journal on Imaging Sciences*, vol. 7, no. 4, pp. 2916–2953, 2014.
[7] R. Bergmann, J. Persch, and G. Steidl, "A parallel DouglasRachford algorithm for restoring images with values in symmetric hadamard manifolds," *SIAM Journal on Imaging Sciences*, vol. 9, no. 3, pp. 901–937, 2016.

Blind Demixing and Deconvolution: Near-optimal Rate

Dominik Stoeger*, Peter Jung†, and Felix Krahmer*
* Zentrum Mathematik, Technische Universität München, 85748 Garching (Munich) , Germany.
† Communications and Information Theory Group, Technische Universität Berlin, 10587 Berlin, Germany

***Abstract*—We consider the problem of blind deconvolution of multiple signals from its superposition, also called *blind demixing and deconvolution*. One is given a signal $y = \sum_{i=1}^{r} w_i * x_i \in \mathbb{R}^L$ which is the superposition of r unknown source signals $\{x_i\}_{i=1}^r$ and convolution kernels $\{w_i\}_{i=1}^r$. The goal is to reconstruct the vectors w_i and x_i, which are elements of known, but random subspaces. The problem can be lifted into a low rank matrix recovery problem and then solved by a semi-definite program. We will present our theorem, which states that up to $\log$-factors, the number of required measurements scales with the number of the degree of freedoms of our system. This significantly improves results from [1].**

I. Introduction

Suppose that we are given a signal

$$y = \sum_{i=1}^{r} w_i * x_i \in \mathbb{R}^L.$$

Our goal is to reconstruct w_i and x_i. Without any further assumptions, this problem is highly underdetermined. We will assume that both w_i and x_i are elements of known subspaces. Thus, we may write $w_i = B_i h_i$ and $x_i = C_i m_i$, where $B_i \in \mathbb{R}^{L\times K_i}$ and $C_i \in \mathbb{R}^{L\times N_i}$. The matrices $(C_i)_{i=1}^r$ are chosen independently at random. Each matrix has independent Gaussian entries, i.e. $(C_i)_{j,k} \in \mathcal{N}(0,1)$. B_i will be the matrix, which extends h_i by zeros. This scenario is of importance in wireless communication, see e.g. [2], [3] .

By $F \in \mathbb{R}^{L\times L}$ we will denote the non-normalized Fourier transform, i.e. $(F)_{kl} = \exp\left(\frac{2\pi i k l}{L}\right)$. We set $\hat{y} = Fy$. Furthermore, if $b_{i,l}$ denotes the l-th row of FB_i and $c_{i,l}$ denotes the l-th column of $(FC_i)^*$ one can compute that for the l-th entry of $\hat{y}$

$$\hat{y}(l) = \frac{1}{\sqrt{L}} \sum_{i=1}^{r} b_{i,l}^* (h_i m_i^*) c_{i,l}.$$

This motivates us to introduce linear maps $\mathcal{A}_i : \mathbb{R}^{K\times N} \to \mathbb{C}^L$ given by

$$\mathcal{A}_i(X)(l) = \frac{1}{\sqrt{L}} b_{i,l}^* X c_{i,l}.$$

Note that this implies $\hat{y} = \sum_{i=1}^{r} \mathcal{A}_i(h_i m_i^*)$.

A convex recovery approach was pioneered in [4] for $r = 1$. In [1] the following convex program was proposed in order to recover the vectors h_i and m_i for $1 \le i \le r$:

$$\text{minimize} \sum_{i=1}^{r} \|Y_i\|_* \quad \text{subject to } \hat{y} = \sum_{i=1}^{r} \mathcal{A}_i(Y_i). \tag{1}$$

Setting $K = \max\limits_{1\le i\le r}$ and $N = \max\limits_{1\le i\le r} N_i$ they could show the following result: If the number of measurements scales essentially in the order of $r^2\left(K + \mu_h^2 N\right)$, then with high probability the convex program is successful. This means that X_0 is the unique minimizer, where

$$X_0 = (h_1 m_1^*, \cdots, h_i m_i^*, \cdots, h_r m_r^*).$$

The quantity $\mu_h \in [1, K]$ is a coherence parameter defined by

$$\mu_h^2 = \max_{1\le i\le r,\ 1\le l\le L} \frac{|h_i^* b_l|^2}{\|h_i\|_{\ell_2}}.$$

II. Our result

In [1], the authors discuss that their recovery guarantee might be too pessimistic. Indeed, their numerical experiments suggest that it seems to be enough if the number of measurements scales linearly in r. The following theorem puts this observation on solid theoretical ground.

Theorem 1 (Jung, Krahmer, Stoeger, 2016)**.** *Let $\alpha \ge 1$. Assume that*

$$L \ge C_\alpha r \left(K \log^2 K + N\mu_h^2\right) \log^2 L \log(\gamma_0 r), \tag{2}$$

where $\gamma_0 = \sqrt{N\left(\log\left(\frac{NL}{2}\right)\right) + \alpha \log L}$ and C_α is a constant only depending on α. Then with probability $1 - \mathcal{O}\left(L^{-\alpha}\right)$ the recovery program (1) is successful, i.e. X_0 is the unique minimizer of (1).

The main novelty in the proof is that the authors established that with high probability

$$(1-\delta)\sum_{i=1}^{r} \|X_i\|_F^2 \le \|\sum_{i=1}^{r} \mathcal{A}_i(X_i)\|_{\ell_2}^2 \le (1+\delta)\sum_{i=1}^{r} \|X_i\|_F^2$$

for all $X = (X_1, \cdots, X_r) \in T$, where

$$T = \Big\{ (u_1 m_1^* + h_1 v_1^*, \cdots, u_r m_r^* + h_r v_r^*,) : \\ u_1 \in \mathbb{R}^{K_1}, \cdots, u_r \in \mathbb{R}^{K_r}, v_1 \in \mathbb{R}^{N_1}, \cdots, v_r \in \mathbb{R}^{N_r} \Big\}$$

This is achieved by using methods developed in [5]. The complete proof has already been announced in [6] and will appear in [7].

References

[1] S. Ling and T. Strohmer, "Blind deconvolution meets blind demixing: Algorithms and performance bounds," 2015.
[2] G. Wunder, H. Boche, T. Strohmer, and P. Jung, "Sparse Signal Processing Concepts for Efficient 5G System Design," *IEEE Access*, vol. 3, pp. 195–208, 2015.
[3] P. Jung and P. Walk, "Sparse Model Uncertainties in Compressed Sensing with Application to Convolutions and Sporadic Communication," in *Compressed Sensing and its Applications*, H. Boche, R. Calderbank, G. Kutyniok, and J. Vybiral, Eds. Springer, 2015, pp. 1–29.
[4] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *Information Theory, IEEE Transactions on*, vol. 60, no. 3, pp. 1711–1732, 2014.
[5] F. Krahmer, S. Mendelson, and H. Rauhut, "Suprema of chaos processes and the restricted isometry property." *Commun. Pure Appl. Math.*, vol. 67, no. 11, pp. 1877–1904, 2014.
[6] D. Stoeger, P. Jung, and F. Krahmer, "Blind deconvolution and compressed sensing," in *Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa 2016)*, Aachen, Germany, September 2016.
[7] P. Jung, F. Krahmer, and D. Stoeger, "Blind Demixing and Deconvolution at Near-optimal Rate," in preparation.

Nonconvex Recovery of Low-Complexity Models

Ju Sun*, Qing Qu†, John Wright†
* Department of Mathematics, Stanford University
† Department of Electrical Engineering, Columbia University

***Abstract*—We study a family of nonconvex optimization problems which can be solved globally, from arbitrary initializations, using efficient algorithms. The key property that these functions have is that (i) every local minimum is global, and (ii) every saddle point has a direction of strict negative curvature. We describe results showing that this property obtains in problems of interest for signal and image processing, including the complete dictionary recovery problem, and the generalized phase retrieval problem.**

Many problems in signal and image processing are most naturally cast as *nonconvex* optimization problems. General nonconvex optimization is NP-hard: in fact, even finding a *local* minimum is hard – nevermind the global optimum. Nevertheless, in many practical situations, simple heuristic algorithms find high quality solutions. The ability of simple, local algorithms to find high-quality solutions for practical problems remains largely mysterious. This talk will recount recent efforts to close this gap, including results on *complete dictionary learning* (DL) and *generalized phase retrieval* (GPR). These results are described in more detail, e.g., in [1], [2], [3].

Perhaps the simplest example of a nonconvex optimization problem which can be solved globally using efficient methods is the problem of optimizing a quadratic form over the sphere

$$\min_{\boldsymbol{x}\in\mathbb{S}^{n-1}} \boldsymbol{x}^*\boldsymbol{M}\boldsymbol{x}, \tag{1}$$

where $\boldsymbol{M}$ is a symmetric matrix. Stationary points of this problem correspond to eigenvectors of $\boldsymbol{M}$. Moreover, the problem exhibits two properties which are critical to showing that it can be solved globally: (i) every local optimizer is global – a vector $\boldsymbol{x}$ is a local maximizer if and only if it is an eigenvector of $\boldsymbol{M}$ corresponding to the smallest eigenvalue. Moreover, (ii) every other critical point has a direction of strict negative curvature – i.e., an appropriate notion of the Hessian over $\mathbb{S}^{n-1}$ (the Riemannian Hessian) has a negative eigenvalue. The first property implies that it is sufficient to find a local optimum; the second implies that a wide range of simple algorithms efficiently find local minimizers – including the (Riemannian) trust region method [2] and noisy gradient descent [4]. Property (ii) has been referred to in the literature as a *strict saddle* [4] or *ridable saddle* property [6].

Perhaps surprisingly, properties (i) and (ii) obtain in a number of practical problems of interest for signal processing. We describe two examples. The first is the sparse dictionary learning problem, in which one seeks a concise representation for a data matrix $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p] \in \mathbb{R}^{n\times p}$. That is to say, we seek a factorization $\boldsymbol{Y} \approx \boldsymbol{Q}\boldsymbol{X}$, in which $\boldsymbol{Q}$ is the dictionary, and the coefficients $\boldsymbol{X}$ are as sparse as possible. The goal is to recover $(\boldsymbol{Q}, \boldsymbol{X})$ given $\boldsymbol{Y}$. Natural approaches to this problem are nonconvex; moreover, the problem admits large symmetry group – if $(\boldsymbol{Q}, \boldsymbol{X})$ is a solution, so is $(\boldsymbol{Q}\boldsymbol{\Gamma}, \boldsymbol{\Gamma}^*\boldsymbol{X})$, for any signed permutation $\boldsymbol{\Gamma}$. This symmetry makes the problem not amenable, in any natural way, to convex relaxation. To study the problem theoretically, a typical approach is to posit a generative model: $\boldsymbol{Y} = \boldsymbol{Q}_0\boldsymbol{X}_0$, and as whether algorithms efficiently recover $(\boldsymbol{Q}_0, \boldsymbol{X}_0)$ up to signed permutation. In particular, we study the problem under the assumption that $\boldsymbol{X}_0$ is a sparse random matrix: each entry is nonzero with some probability θ, and the nonzero entries are independent $\mathcal{N}(0,1)$ random variables. We also restrict $\boldsymbol{Q}_0$ to be a square and invertible matrix. Under these assumptions, the problem of learning $\boldsymbol{Q}_0$ and $\boldsymbol{X}_0$ can be reduced to a sequence of n-dimensional optimization problems, in which the goal is to recover a single column of $\boldsymbol{Q}_0$:

$$\min_{\boldsymbol{q}\in\mathbb{S}^{n-1}} \sum_{i=1}^{p} h_\mu\left(\boldsymbol{q}^*\boldsymbol{y}_i\right). \tag{2}$$

Here, h_μ is a smooth, sparsity encouraging function. It turns out that when $p \geq \mathrm{poly}(n)$, and $\theta < 1/3$, with high probability this problem has no spurious local minimizers. This implies that we can efficiently recover $\boldsymbol{Q}_0$ and $\boldsymbol{X}_0$ in this situation. A noteworthy aspect of this result is that the probability θ of an entry of $\boldsymbol{X}_0$ being nonzero can be a constant; most previous results on simple, practical methods required $\theta = O(1/\sqrt(n))$.

A second example in which properties (i) and (ii) obtains is the *generalize phase retrieval* problem, in which we attempt to recover a complex vector $\boldsymbol{x} \in \mathbb{C}^n$ from the moduli $y_i = |\boldsymbol{a}_i^*\boldsymbol{x}|$ of its projections onto a collection of vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{C}^n$. This property also exhibits a symmetry, which arises because the measurements are invariant to a global phase shift $\boldsymbol{x} \mapsto e^{i\phi}\boldsymbol{x}$. We study a natural optimization formulation,

$$\min_{\boldsymbol{z}} \tfrac{1}{2m}\sum_{i=1}^{m}\left(|\boldsymbol{a}_i^*\boldsymbol{z}|^2 - y_i^2\right)^2. \tag{3}$$

For this problem, properties (i) and (ii) obtain when the $\boldsymbol{a}_i$ are iid complex normal random vectors, and $m \geq Cn\log^3 n$. A contrast between this result and previous efforts on this problem is that it implies that simple iterative methods obtain the correct solution independent of initialization; previous results on nonconvex methods for this problem required careful initialization.

For both problems, we describe an approach to analysis which first examines the population (large sample) version of the objective function, demonstrates that it is well-structured for optimization, and then shows that with high probability the finite sample version is similarly well-structured. For both problems, these results imply that a variety of simple, efficient algorithms produce correct solutions.

References

[1] J. Sun, Q. Qu, and J. Wright, "Complete dictionary recovery over the sphere i: Overview and geometric picture," preprint, *http://arxiv.org/abs/1504.06785*, 2015.

[2] ——, "Complete dictionary recovery over the sphere ii: Recovery by riemannian trust-region method," preprint, *http://arxiv.org/abs/1504.06785*, 2015.

[3] ——, "A geometric analysis of phase retrieval," preprint, *https://arxiv.org/abs/1602.06664*, 2016.

[4] R. Ge, F. Huang, C. Jin, and Y. Yuan, "Escaping from saddle points - online stochastic gradient for tensor decomposition," *CoRR*, vol. abs/1503.02101, 2015. [Online]. Available: http://arxiv.org/abs/1503.02101

[5] J. D. Lee, M. Simchowitz, M. I. Jordan, and B. Recht, "Gradient descent only converges to minimizers," in *COLT 2016*.

[6] J. Sun, Q. Qu, and J. Wright, "When are nonconvex problems not scary?"