University of BRISTOL

SuSTaIn EdgeCutter One Day Workshop on High-dimensional Statistics, Inverse Problems and Convex Analysis

Workshop Programme
22nd March 2016

Organisers:

Marcelo Pereyra, Bristol
Carola-Bibiane Schönlieb, Cambridge

Supported by SuSTaIn

SuSTaIn

Workshop Programme

8:45	Registration & Coffee
9:15	Welcome
9:30	**Martin Wainwright** (Berkeley) *Statistics meets Optimization: Fast randomized algorithms for large data sets*
10:20	**Volkan Cevher** (EPFL) *A universal primal-dual convex optimization framework*
11:15	Water break
11:20	**Gabriel Peyré** (Paris-Dauphine) *Exact Support Recovery for Sparse Spikes Deconvolution*
12:10	Lunch and Poster Session
13:30	**Mario Figuiredo** (Lisbon) *Learning with Strongly Correlated Variables: Ordered Weighted l1 Regularization*
14:20	**Simon Arridge** (UCL) *Regularisation Methods for Joint Image Reconstruction*
15:10	Coffee Break
15:40	**Yi Yu** (Bristol/ Cambridge) *Capture the neglected -- recovery of the dependence structure in directed and dynamic networks*
16:30	**Carola-Bibiane SchĘnlieb** (Cambridge) *Bilevel optimisation for variational regularisation models*
17:20	Closing remarks
17:30	Scheduled end

Abstracts (in order of appearance)

Martin Wainwright, Berkeley
Statistics meets Optimization: Fast randomized algorithms for large data sets

Large-scale data sets are now ubiquitous throughout engineering and science, and present a number of interesting challenges at the interface between statistics and optimization. In this talk, we discuss the use of randomized dimensionality reduction techniques, also known as sketching, for obtaining fast but approximate solutions to large-scale convex programs. Using information-theoretic techniques, we first reveal a surprising deficiency of the most widely used sketching technique. We then show how a simple iterative variant leads to a much faster algorithm, and one which adapts to the intrinsic dimension of the solution space. Moreover, we show how it generalizes naturally to a randomized version of the Newton algorithm with provable guarantees. Based on joint work with Mert Pilanci, UC Berkeley.

Volkan Cevher, EPFL
A universal primal-dual convex optimization framework

This talk proposes a new primal-dual algorithmic framework for a prototypical constrained convex optimization template. The algorithmic instances of ourframework are universal since they can automatically adapt to the unknown Holder continuity properties within the template. They are also guaranteed to have optimal convergence rates in the objective residual and the feasibility gap for each smoothness level. In contrast to existing primal-dual algorithms, ourframework avoids the proximity operator of the objective function altogether. We instead leverage computationally cheaper, Fenchel-type operators, which are the main workhorses of the generalized conditional gradient (GCG)-type methods.

In contrast to the GCG-type methods, our framework does not require the objective function to be differentiable, and can also process additional general linear inclusion constraints. Our analysis technique unifies Nesterov's universal gradient methods and GCG-type methods to address the more broadly applicable primal-dual setting. Along the way, we also show that GCG-type methods can also work for non Lipschitz objectives. We provide numerical evidenceto demonstrate the scalability of our framework beyond trillion dimensions in diverse applications, from Quantum Tomography to Phase Retrieval and from clustering to matrix completion.

Gabriel Peyré, Paris-Dauphine

Exact Support Recovery for Sparse Spikes Deconvolution

In this talk, I study sparse spikes deconvolution over the space of measures, following several recent works (see for instance [2,3]). For non-degenerate sums of Diracs, we show that, when the signal-to-noise ratio is large enough, total variation regularization of measures (which the natural extension of the l1 norm of vectors to the setting of measures) recovers the exact same number of Diracs. We also show that both the locations and the heights of these Diracs converge toward those of the input measure when the noise drops to zero. The exact speed of convergence is governed by a specific dual certificate, which can be computed by solving a linear system. These results extend those obtained by [2]. We also draw connections between the performances of sparse recovery on acontinuous domain and on a discretized grid. When the measure is positive, it is known that l1-type methods always succeed when there is no noise. We show that exact support recovery is still possible when there is noise. The signal-to- noise ratio should then scale like $1/t^{2N-1}$ where there are N spikes separated by a distance t. This reflects the intrinsic explosion of the ill-posedness of the problem [4].

This is joint work with Vincent Duval and Quentin Denoyelle, see [1,4] for more details. Bibliography:[1] V. Duval, G. Peyré, Exact Support Recovery for Sparse Spikes Deconvolution, Foundation of Computational Mathematics, 15(5), pp. 1315–1355, 2015.[2] E. J. Candès and C.Fernandez-Granda. Towards a mathematical theory of super-resolution. Communications on Pure and Applied Mathematics, 67(6), 906-956, 2013.[3] K. Bredies and H.K. Pikkarainen. Inverse problems in spaces of measures. ESAIM: Control, Optimisation and Calculus of Variations, 19:190-218, 2013.[4] Q. Denoyelle, V. Duval, G. Peyré. Support Recovery for Sparse Deconvolution of Positive Measures. Preprint Arxiv:1506.08264, 2015

Mario Figueiredo, Lisbon

Learning with Strongly Correlated Variables: Ordered Weighted l1 Regularization

In high-dimensional linear regression (and other supervised learning problems), it is common to encounter several highly correlated variables (a.k.a. covariates, predictors, features). Adopting standard sparsity-inducing regularization (namely l1, or LASSO) in such situations may be unsatisfactory, as it results in the selection of arbitrary convex combinations of those features, maybe even of an arbitrary subset thereof. However, especially in scientific applications, it is desirable to explicitly identify all the covariates that are relevant for a given task, as well as explicitly identify groups/clusters of such highly correlated covariates. This talk addresses the recently introduced sorted weighted L1 regularizer, which has been proposed for this purpose. We review several convex optimization aspects concerning this regularizer, namely efficient methods to compute its proximity operator. In the analysis front, we give sufficient conditions for exact feature clustering (under squared error, absolute error, and logistic losses) and characterize its statistical performance.

Simon Arridge, UCL
Regularisation Methods for Joint Image Reconstruction. S.R.Arridge Joint Work with M.Ehrhardt

Several imaging problems consider multiple images simultaneously. Examples include colour and multispectral imaging, hybrid imaging in medical imaging (such as PET-MRI, and SPECT-CT), as well as geophysical imaging (electrical and acoustic properties reconstruction). The use of variational regularisation techniques for inverse problems in these applications can treat each image channel separately or jointly. In this talk we consider methods based on the joint information of multiple images in terms of both their geometry, and their statistics. For the former we propose methods based on parallel level sets, and for the latter methods based on both joint entropy, and on multispectral probabilistic diffusion. Examples are shown on model problems and for medical imaging applications.

Yi Yu, Bristol/ Cambridge
Capture the neglected -- recovery of the dependence structure in directed and dynamic networks.

Directed and/or dynamic networks are of increasing interest but also under-studied. In this paper, we aim to recover the dependence structures across both time course and network structure. Multivariate counting processes with recurrent survival analysis techniques are used to model the individual behaviour with time dependent properties; whilst the network structure is modelled as the graphical model. The contribution of this paper is two-fold: a) a sandwich covariance estimator is available for more robust statistical inference and b) a conditional independent technique in undirected network is integrated into the directed network making the dependence structure characterisable and computationally feasible.

Carola-Bibiane Schönlieb, Cambridge

Bilevel optimisation for variational regularisation models

When assigned with the task of reconstructing an image from imperfect data the first challenge one faces is the derivation of a truthful image and data model. In the context of regularised reconstructions, some of this task amounts to selecting an appropriate regularisation term for the image, as well as an appropriate distance function for the data fit. This can be determined by the a-priori knowledge about the image, the data and their relation to each other. The source of this knowledge is either our understanding of the type of images we want to reconstruct and of the physics behind the acquisition of the data or we can thrive to learn parametric models from the data itself. The common question arises: how can we optimise our model choice? In this talk we discuss a bilevel optimization method for computing optimal parameters in variational regularisation models. In particular, we will consider optimal parameter derivation for total variation denoising with multiple noise distributions, optimising total generalised variation regularisation for its application in photography, and learning good spatial-temporal regularisation for dynamic image reconstruction.
This is joint work with M. Benning, L. Calatroni, C. Chung, J. C. De Los Reyes, T. Valkonen, and V.Vladic

University of BRISTOL

School of Mathematics, University Walk, Bristol BS8 1TW, UK
Astrostatistics-workshop@bristol.ac.uk
www.sustain.bris.ac.uk/ws-astrostatistics/