3rd IMA Conference on Numerical Linear Algebra and Optimisation
September 10-12, 2012, Birmingham, United Kingdom

Mini-Symposium: Gradient Methods for Large-Scale Optimisation

Organizer: Peter Richtárik

Gradient methods have in the past six years dominated research activity on large scale optimisation in application domains where it is modest dependence on problem dimension (scalability) and not accuracy of solution that is of primary concern. Old algorithms are being refurbished and redesigned and new ones developed and analyzed for novel applications in data intensive domains such as machine learning, compressed sensing, truss topology design, eigenvalue optimization, image reconstruction and stochastic programming.

This mini-symposium will bring together several leading and early-career researchers working on various aspects of gradient methods, including algorithm design and analysis, modelling and applications.

Session 1:

Michel Baes (ETH Zurich)
First-order methods for eigenvalue minimization

Guillaume Obozinski (INRIA)
Algorithms for structured sparsity

Peter Richtárik (Edinburgh)
Parallel block coordinate descent methods for huge-scale partially separable problems

Raphael Hauser (Oxford)
Sequential compressed sensing

Session 2:

Martin Takáč (Edinburgh)
Distributed block coordinate descent method: iteration complexity and efficient hybrid implementation

Vinh Xuan Doan (Warwick)
A proximal point algorithm for sequential feature extraction applications

Carola-Bibiane Schönlieb (Cambridge)
Domain decomposition for total variation regularisation and applications

Rachael Tappenden (Edinburgh)
Block coordinate descent applied to multistage stochastic optimization

Abstracts

Michel Baes (ETH Zurich)

First-order methods for eigenvalue minimization

Many semidefinite programming problems encountered in practice can be recast as minimizing the maximal eigenvalue of a convex combination of symmetric matrices. In this talk, we describe and analyze a series of first-order methods for solving this problem when the input matrices are large (of dimension 1000 to 10000 and more) and mildly sparse. We propose in this talk several accelerating strategies of optimal first-order methods. A first procedure consists in adopting an aggressive step-size selection in an appropriate smoothing technique, a strategy which proves to be particularly efficient for large-size problems. The second procedure is motivated by the fact that the fastest first-order methods for eigenvalue minimization require the computation of a matrix exponential at every iteration. We develop a randomized strategy to accelerate the computation of these exponentials, while maintaining strong probabilistic guarantees on the outcome of our algorithm. We show theoretically and experimentally that this method is faster than all the standard existing methods on a certain set of instances.

Guillaume Obozinski (INRIA)

Structured sparsity: successes and open challenges

The broadly used L1 regularization is initially motivated as providing a convex surrogate to the penalization of the number of non-zero coefficients in a model. It leads to nonsmooth optimization problems, which can however be solved quite efficiently by first-order methods. The last years have seen the emergence of the field of structured sparsity, which aims at identifying a model of small complexity, when the latter is not measured merely by the number of its nonzero parameters but by a more general complexity measure of the patterns formed by these ``selected" parameters. Various convex regularizations have been proposed that formalized the notion that prior information can be expressed through functions encoding the set of possible or encouraged patterns. The convex optimization problems arising in this setting have interesting characteristics from the point of view of algorithmic design: they are non-differentiable optimization problems, but they present structures, symmetries and partial separability properties that in many cases allow for efficient algorithms based on first order methods: in particular, proximal methods have been applied successfully to a certain number of these problems including groups sparsity, hierarchical sparsity and sparsity with overlap.

In this talk, I will present a new unifying point of view to structured sparsity, which, given any combinatorial function encoding the structured a priori, constructs a convex regularizer with the fundamental property that it provides the tightest convex relaxation to that penalty. Such tight regularizers are in many cases different from the ones that have been considered in the literature and come with new computational challenges: proximal methods and other state-of-the-art first order methods are not so easily applicable to them. In the particular case, where the combinatorial function is submodular however, I will show that the computation of the proximal operator can be done efficiently, hence leading to efficient algorithms.

Peter Richtárik (Edinburgh)

Parallel block coordinate descent methods for huge-scale partially separable problems

In this work we show that randomized block coordinate descent methods can be accelerated by parallelization when applied to the problem of minimizing the sum of a partially block separable smooth convex function and a simple block separable convex function. We give a generic algorithm and several variants thereof based on the way parallel computation is performed. In all cases we prove iteration complexity results, i.e., we give bounds on the number of iterations sufficient to approximately solve the problem with high probability. Our results generalize the intuitive observation that in the separable case the theoretical speedup caused by parallelization must be equal to the number of processors. We show that the speedup increases with the number of processors and with the degree of partial separability of the smooth component of the objective function. Our analysis also works in the mode when the number of blocks being updated at each iteration is random, which allows for modelling situations with variable (busy or unreliable) number of processors. We conclude with some encouraging computational results applied to huge-scale LASSO and sparse SVM instances.

Raphael Hauser (Oxford)

Sequential compressed sensing

We introduce a nonlinear sparse optimisation problem that arises in medical imaging and discuss an algorithm for its solution based on an iterative sequence of compressed sensing problems.

Martin Takáč (Edinburgh)

Distributed block coordinate descent method: iteration complexity and efficient hybrid implementation

In this work we propose solving huge-scale instances of regularized convex minimization problems using a distributed block coordinate descent method. We analyze the iteration complexity of the (synchronous) algorithm and show how it depends on the way the problem data is partitioned to the nodes. Several variations of the basic method are obtained based on the way updates are handled (P2P, broadcasting, asynchronicity). Finally, we report encouraging numerical results for an efficient hybrid MPI + Open MP implementation applied to LASSO and sparse support vector machine instances.

Xuan Vinh Doan (Warwick)

A proximal point algorithm for sequential feature extraction applications

We propose a proximal point algorithm to solve a problem that can be used to extract features in data. We use a modified alternating direction method to generate good starting points and then use the accelerated proximal gradient method in each iteration of the main algorithm. We also develop new stopping criteria for the proximal point algorithm, which is based on optimality conditions of epsilon-optimal solutions, with a theoretical guarantee. We test our algorithm with two image databases and show that we can use the proposed algorithm to extract appropriate common features from these images. This is joint work with Kim-Chuan Toh and Stephen Vavasis.

Carola-Bibiane Schönlieb (Cambridge)

Domain decomposition for total variation regularisation and applications

Domain decomposition methods were introduced as techniques for solving partial differential equations based on a decomposition of the spatial domain of the problem into several subdomains. The initial equation restricted to the subdomains defines a sequence of new local problems. The main goal is to solve the initial equation via the solution of the local problems. This procedure induces a dimension reduction which is the major responsible of the success of such a method. Indeed, one of the principal motivations is the formulation of solvers which can be easily parallelized.

In this presentation we shall develop a domain decomposition algorithm to the minimization of functionals with total variation constraints. In this case the interesting solutions may be discontinuous, e.g., along curves in 2D. These discontinuities may cross the interfaces of the domain decomposition patches. Hence, the crucial difficulty is the correct treatment of interfaces, with the preservation of crossing discontinuities and the correct matching where the solution is continuous instead. I will present our domain decomposition strategy, including convergence results for the algorithm and numerical examples for its application in image inpainting and magnetic resonance imaging.

Rachael Tappenden (Edinburgh)

Block coordinate descent applied to multistage stochastic optimization

Recently there has been much interest in block coordinate descent methods because of their ability to tackle large-scale optimization problems. Multistage stochastic programming problems are an example of very large optimization problems (where the size of the problem grows rapidly with the number of scenarios and time horizon) that display particular structure and sparsity patterns, namely, dual block angular structure. This work considers a scenario decomposition approach to multistage stochastic optimization and the application of the block coordinate descent method to determine the solution of the subproblems. Preliminary numerical results will also be presented.