7th IMA Conference on Numerical Linear Algebra and Optimization

University of Birmingham

Birmingham, B15 2TT, UK

Wednesday 29 June - Friday 1 July 2022

CONFERENCE ABSTRACT BOOK

Table of Contents

***Fast and accurate optimization on the orthogonal
manifold without retraction, Pierre Ablin***

We consider the problem of minimizing a function over the manifold of orthogonal matrices. The majority of algorithms for this problem compute a direction in the tangent space, and then use a retraction to move in that direction while staying on the manifold. Unfortunately, the numerical computation of retractions on the orthogonal manifold always involves some expensive linear algebra operation, such as matrix inversion, exponential or square-root. These operations quickly become expensive as the dimension of the matrices grows. To bypass this limitation, we propose the landing algorithm which does not use retractions. The algorithm is not constrained to stay on the manifold but its evolution is driven by a potential energy which progressively attracts it towards the manifold. One iteration of the landing algorithm only involves matrix multiplications, which makes it cheap compared to its retraction counterparts. We provide an analysis of the convergence of the algorithm, and demonstrate its promises on large-scale and deep learning problems, where it is faster and less prone to numerical errors than retraction-based methods.

Circulant Preconditioning of the Fast Gradient Method for Predictive Control

Ian McInerney *(The University of Manchester & Imperial College London)* & Eric C. Kerrigan & George A. Constantinides *(Imperial College London)*

Recently, Model Predictive Control (MPC) has grown in popularity due to its ability to incorporate operating constraints in the computation of an optimal control action. At its core, MPC solves an optimization problem to determine the next control action, which for the Constrained Linear Quadratic Regulator (CLQR) is a Quadratic Program (QP). This popularity has led to MPC being implemented on smaller processors and FPGAs that do not contain hardware to accelerate floating-point computations and that can take advantage of many parallelization opportunities in the problem structure and algorithms.

We present a preconditioner for the Fast Gradient Method (FGM) when the FGM is used to solve the condensed linear MPC problem. We exploit the block Toeplitz structure in the Hessian of the linear MPC problem's QP to create a block circulant preconditioning matrix that is easy to compute and that can accelerate the convergence of the FGM by up to 9x. To preserve the parallelization opportunities inside the FGM's projection step, we symmetrically precondition the Hessian and restrict the circulant preconditioner to be block diagonal, meaning the overall preconditioner can be defined by a single matrix $L \in \mathbb{R}^{m \times m}$, (where m is the number of inputs to the system being controlled).

We define A and B to be the system's state transition and input matrices, respectively, K to be a prestabilizing controller for the system, and P to be the solution to the discrete algebraic Ricatti equation or the discrete Lyapunov equation for the system. The blocks of the preconditioner can be computed in closed form, with L given by the lower-triangular Cholesky decomposition of M with

$$M := B'PB - B'K'R - RKB + R.$$

We show in numerical examples that this block circulant preconditioner has equivalent performance to the start-of-the-art preconditioner computed using a Semidefinite Program (SDP), giving between a 2x and 9x speedup for the FGM. The computation of the proposed preconditioner is orders of magnitude faster than the SDP-based preconditioner, with our preconditioner computed in milliseconds for large MPC problems instead of the minutes required for the SDP-based preconditioner.

Low-rank matrix recovery with Ky Fan 2-k-norm[*]

Xuan Vinh Doan[†‡]   Stephen Vavasis[§]

February 28, 2020

Matrix recovery problem concerns the construction of a matrix from incomplete information of its entries. This problem has a wide range of applications such as recommendation systems with incomplete information of users' ratings. Finding these low-rank matrices are theoretically difficult due to their non-convex properties. Computationally, it is important to study the tractability of these problems given the large scale of datasets considered in practical applications. In this paper, we propose Ky Fan 2-k-norm-based models for the non-convex low-rank matrix recovery problem which aims to recover matrices which are not recoverable by the convex relaxation using the nuclear norm [4]. Similar to the nuclear norm, the dual Ky Fan 2-k-norm with $k > 1$ can be used to compute the k-approximation of a matrix $\boldsymbol{A}$ (Proposition 2.9, [1]), which demonstrates its low-rank property. A general difference of convex algorithm (DCA) is developed to solve these models. The sub-problem of the DCA algorithm is a convex optimization problem which can be reformulated as a semidefinite optimization problem (SDP) using the formulation of the dual Ky Fan 2-k-norm derived in [1]. Furthermore, to handle large instances, we develop a proximal point algorithm (PPA) framework to solve the Ky Fan 2-k-norm minimization sub-problem, which is similar to that of nuclear norm minimization problem proposed in [3]. We derive a new analytical formulation of the dual Ky Fan 2-k-norm and use it to compute the proximal operator in the proposed PPA framework. Numerical results show that the proposed models achieve much higher recoverability rates as compared to the nuclear norm relaxation and the alternating bilinear optimization approach discussed in [2]. The results also demonstrate that the proposed DCA algorithm together with the PPA framework is efficient in handling larger instances.

References

[1] X. V. Doan and S. Vavasis. Finding the largest low-rank clusters with Ky Fan 2-k-norm and ℓ_1-norm. *SIAM Journal on Optimization*, 26(1):274–312, 2016.

[2] P. Jain, P. Netrapalli, and S. Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, pages 665–674. ACM, 2013.

[3] Y.-J. Liu, D. Sun, and K.-C. Toh. An implementable proximal point algorithmic framework for nuclear norm minimization. *Mathematical Programming*, 133(1-2):399–436, 2012.

[4] B. Recht, M. Fazel, and P. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.

[†]Warwick Business School, University of Warwick, Coventry, CV4 7AL, United Kingdom, xuan.doan@wbs.ac.uk,

[‡]Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB, United Kingdom.

[§]Department of Combinatorics and Optimization, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada, vavasis@uwaterloo.ca.

An apocalypse-free first-order low-rank optimization algorithm, Guillaume Olikier

We consider the problem of minimizing a differentiable function with locally Lipschitz continuous gradient on the real determinantal variety, and present a first order algorithm designed to find stationary points of that problem. This algorithm applies steps of steepest descent with back tracking line search on the variety, as proposed by Schneider and Uschmajew (2015), but by taking the numerical rank into account to perform suitable rank reductions. We prove that this algorithm produces sequences of iterates the accumulation points of which are stationary, and therefore does not follow the so-called apocalypses described by Levin Kileel, and Boumal

(2021).

Multivariate Least–Squares Approximations in Irregular Domains via Vandermonde with Arnoldi

[*]Kate Wenqi Zhu, [†]Yuji Nakatsukasa

Many problems in computational science call for the approximation of smooth, multivariate functions. Vandermonde matrices arise frequently in polynomial approximations. However, Vandermonde matrices are exponentially ill–conditioned and often results in an unstable approximation.

We introduce and explore the *multivariate Vandermonde with Arnoldi (V+A) method*, which is based on least–squares approximation, for approximating continuous, multivariate functions on d–dimensional irregular domains. The V+A method addresses the ill–conditioning of the Vandermonde approximation by creating a set of discrete orthogonal basis with respect to a discrete measure. The V+A method is simple and general. It relies only on the sample points from the domain and requires no prior knowledge of the domain.

A crucial insight of the V+A technique is that the algorithm generates a discrete orthogonal basis with respect to a discrete measure. This paper establishes the link between the discrete orthogonal polynomials and the continuous orthogonal polynomials. Under the natural assumptions of the Riemann sum integral, we give an original proof that the discrete orthogonal polynomials generated by equispaced sample points in any real interval are well approximated by the scaled Legendre polynomials with $M = \mathcal{O}(N^2)$ sample points.

It is well known that least–squares approximations become inaccurate when the number of sample points M (from a sub-optimal distribution, e.g. equispaced points) is close to the dimension of the approximation space N. We show that, for a large class of domains, the V+A method gives a well–conditioned and accurate N–dimensional least–squares approximation using $M = \mathcal{O}(N^2)$ equispaced sample points or $M = \mathcal{O}(N^2 \log N)$ random sample points. We analyze the accuracy of the V+A approximation using two different approaches (i.e. the Lebesgue constant approach and the admissible mesh approach) and provide an error bound which is tighter than the bound discussed by Calvia and Levenberg [2].

Using a suitable weighting measure for the least–squares system reduces the number of samples needed to achieve an accurate approximation. Leveraging on the weighting measure in [1, 3], we propose a new variant of the weighted least–squares algorithm that uses the multivariate V+A. This algorithm is stable with high probability and only takes $M = \mathcal{O}(N \log N)$ sample points to give a near–optimal approximation. Our numerical results confirm that this method gives a more accurate approximation than the original method in [1].

This work is joint with Yuji Nakatsukasa (Oxford University).

---
[*]wenqi.zhu@maths.ox.ac.uk, University of Oxford, UK
[†]nakatsukasa@maths.ox.ac.uk, University of Oxford, UK

A Fully Discrete, Riemanninan Approach to PDE-Constrained Shape Optimization

Estefanía Loayza-Romero
joint work with: Roland Herzog

March 25, 2022

It is well known among practitioners that the numerical solution of shape optimization problems constrained by partial differential equations (PDEs) often exhibits several difficulties. In particular, when the PDE is discretized by a finite element method, and the underlying mesh is used to represent the shape of the domain to be optimized directly, one often experiences a degeneracy of the mesh quality as the optimization progresses. The degeneracy manifests itself in some of the mesh cells thinning in the sense that at least one of its heights approaches zero.

In this talk, we focus on two-dimensional problems, where the PDE is discretized using a finite element method, and the underlying mesh represents the discrete shape. We study the set of all possible configurations of node positions a mesh of a given connectivity can attain. Then, using the language of simplicial complexes, we provide theoretical evidence that this set is a smooth manifold, and we term it the *manifold of planar triangular meshes* [3]. Moreover, we construct a complete Riemannian metric for the manifold of planar triangular meshes. This construction is based in the result presented in [1]. In practical terms using this metric allows us to update the meshes following geodesics in any direction and as long as we want without jeopardizing their quality.

Finally, alongside the newly proposed notion of the complete manifold of planar triangular meshes, we focus on the theoretical and computational aspects of discretized PDE-constrained shape optimization problems. We provide numerical evidence that such problems generally possess no solutions within the manifold of planar triangular meshes, even when the shape functional is bounded below. To overcome this drawback, we introduce a penalty functional which, briefly speaking, controls the mesh quality. When added to the shape functional, it renders well-posed discrete shape optimization problems, i.e., they possess at least one globally optimal solution. Subsequently, we solve the penalized problem using four different variants of the Riemannian steepest descent method, [2].

References

[1] William B. Gordon. "An analytical criterion for the completeness of Riemannian manifolds". In: *Proceedings of the American Mathematical Society* 37 (1973), pp. 221–225. DOI: 10.2307/2038738.

[2] Roland Herzog and Estefania Loayza-Romero. *A Discretize-Then-Optimize Approach to PDE-Constrained Shape Optimization*. 2021. arXiv: 2109.00076.

[3] Roland Herzog and Estefania Loayza-Romero. *A manifold of planar triangular meshes with complete Riemannian metric*. 2020. arXiv: 2012.05624.

Computing second-order points under equality constraints using Fletcher's augmented Lagrangian

Florentin Goyens

March 18, 2022

As part of the mini-symposium: Algorithms on manifolds.
Joint work with: Armin Eftekhari, Nicolas Boumal.

Given a Euclidean space $\mathcal{E}$ with inner product $\langle \cdot , \cdot \rangle$ and associated norm $\|\cdot\|$, we consider a generic smooth nonconvex problem with equality constraints

$$\min_x f(x) \text{ subject to } h(x) = 0, \tag{P}$$

where $f\colon \mathcal{E} \to \mathbb{R}$ and $h\colon \mathcal{E} \to \mathbb{R}^m$ are smooth. The feasible set is denoted by $\mathcal{M} = \{x \in \mathcal{E} : h(x) = 0\}$ and is assumed to be a smooth manifold. One emerging motivation for Problem (P) is the minimization over Riemannian manifolds using infeasible methods, which is valuable when computing retractions is prohibitively expensive, as is the case with orthogonality constraints when the number of columns is large. Fletcher (1970) introduced the penalty function

$$g(x) := \mathcal{L}_\beta(x, \lambda(x)) = f(x) - \langle h(x), \lambda(x) \rangle + \frac{\beta}{2} \|h(x)\|^2$$

with the least-squares multipliers $\lambda(x) = (\mathrm{D}h(x)^*)^\dagger [\nabla f(x)]$, where the dagger denotes a pseudo-inverse. The function g, now known as Fletcher's augmented Lagrangian, is smooth on the set $\mathcal{D} = \{x \in \mathcal{E} : \mathrm{rank}(\mathrm{D}h(x)) = m\}$. For any $x \in \mathcal{D}$, we show that the set $\mathcal{M}_x = \{y \in \mathcal{E} : h(y) = h(x)\}$ is a submanifold of $\mathcal{E}$ contained in $\mathcal{D}$. This allows to define the Riemannian gradient and Hessian of f with respect to the manifold $\mathcal{M}_x$ for $x \in \mathcal{D}$, which we write $\mathrm{grad}_{\mathcal{M}_x} f(x)$ and $\mathrm{Hess}_{\mathcal{M}_x} f(x)$. We introduce a natural definition of approximate second-order criticality for (P). A point $x \in \mathcal{D}$ is defined as an $(\varepsilon_0, \varepsilon_1, \varepsilon_2)$-approximate second-order critical point of (P) if

$$\|h(x)\| \leq \varepsilon_0, \qquad \left\|\mathrm{grad}_{\mathcal{M}_x} f(x)\right\| \leq \varepsilon_1 \qquad \text{and} \qquad \mathrm{Hess}_{\mathcal{M}_x} f(x) \succeq -\varepsilon_2 \mathrm{Id}. \tag{ε-SOCP}$$

Likewise, an $(\varepsilon_0, \varepsilon_1)$-approximate first-order critical point (FOCP) satisfies the first two conditions above. We establish conditions under which approximate first- and second-order critical points of Fletcher's augmented Lagrangian are also approximate minimizers of (P) in the sense of (ε-SOCP).

We propose an algorithm based on the minimization of Fletcher's augmented Lagrangian which produces an $(\varepsilon_1, 2\varepsilon_1)$-FOCP of (P) in at most $\mathcal{O}\left(\varepsilon_1^{-2}\right)$ iterations and an $(\varepsilon_1, 2\varepsilon_1, \varepsilon_2 + C\varepsilon_1)$-SOCP of (P) in at most $\mathcal{O}\left(\max\{\varepsilon_1^{-2}, \varepsilon_2^{-3}\}\right)$ iterations, where $C \geq 0$ is a constant depending on the function h. This matches the optimal worst-case complexity with respect to ε, and is the first infeasible algorithm to achieve such rate in this setting for a similar notion of approximate optimality.

References

Fletcher, R. (1970). A class of methods for nonlinear programming with termination and convergence properties. Integer and nonlinear programming, pages 157–173.

1

Algorithms for Deterministically Constrained Stochastic Optimization, Frank Curtis

Abstract: I will present the recent work by my research group on the design, analysis, and implementation of algorithms for solving continuous nonlinear optimization problems that involve a stochastic objective function and deterministic constraints. The talk will focus on our sequential quadratic optimization (commonly known as SQP) methods for cases when the constraints are defined by nonlinear systems of equations, which arise in various applications including optimal control, PDE-constrained optimization, and network optimization problems. One might also consider our techniques for training machine learning (e.g., deep learning) models with constraints. Much of our recent work focuses on the "fully stochastic" regime in which only stochastic gradient estimates are employed, for which we have derived convergence in expectation results and worst-case iteration complexity bounds that are on par with stochastic gradient methods for the unconstrained setting. I will also discuss the various extensions that my group is exploring, including an approach that allows inexact subproblem solutions obtained via iterative linear algebra techniques, along with other related open questions.

New indicators for the early termination of the linear solver in Interior Point Methods

Filippo Zanetti   Jacek Gondzio

University of Edinburgh

Abstract

When an iterative method is applied to solve the linear equation system in interior point methods (IPMs), the attention is usually placed on accelerating their convergence by designing appropriate preconditioners, but the linear solver is applied as a black box solver with a standard termination criterion which asks for a sufficient reduction of the residual in the linear system. Such an approach often leads to an unnecessary "oversolving" of linear equations. In this talk, it is shown how an IPM can preserve the polynomial worst-case complexity when relying on an inner termination criterion that is not based on the residual of the linear system. Moreover, a practical criterion is derived from a deep understanding of IPM needs. The new technique has been adapted to the Conjugate Gradient (CG) and to the Minimum Residual method (MINRES) applied in the IPM context. The new criterion has been tested on a set of quadratic optimization problems including compressed sensing, image processing and instances with partial differential equation constraints, and it has been compared to standard residual tests with variable tolerance. Evidence gathered from these computational experiments shows that the new technique delivers significant improvements in terms of inner (linear) iterations and those translate into significant savings of the IPM solution time.

Title: Label-free Domain Adaptation on Riemannian Manifolds
Presenter: Ronen Talmon, Technion – Israel Institute of Technology

Abstract:
Domain adaptation is a key scientific task in modern data analysis. The need for domain adaptation arises since data are often acquired in multiple domains, under different environmental conditions, using various acquisition equipment, and at different sites. Many contexts of domain adaptation are considered in the literature. In our work, we consider a label-free domain adaptation and the following goal. We attempt to mitigate the difference between related data sets by providing a new representation of their union, such that any subsequent downstream task applied to the union could be unaware of their original partition and could treat them as one homogeneous set.

We focus on the problem of label-free domain adaptation of data living on Riemannian manifolds. We take a purely geometric approach, which is often referred to as Procrustes Analysis, and introduce a label-free domain adaptation method that is based on three geometric operations: translation, scaling, and rotation. We consider three Riemannian manifolds: (i) the manifold of symmetric positive definite (SPD) matrices with the so-called affine-invariant metric, (ii) the manifold of symmetric positive semi-definite (SPSD) matrices, which is viewed as a product space of the SPD manifold and the Grassmann manifold, and (iii) the Lorentz model of hyperbolic space. In each manifold, we implement our method using the respective Riemannian geometry and theoretically analyze its properties. We showcase the efficacy of the method in several applications involving batch effect removal from gene expression datasets, fusion of hyperspectral images, and multi-subject adaptation of EEG recordings.

Randomized GMRES with Singular Vectors Based Deflated Restarting

Yongseok Jang[1], Cédric Content[1], Emeric Martin[1], and Laura Grigori[2]

[1]DAAA, ONERA, The French Aerospace Lab, Châtillon, France
`{yongseok.jang,cedric.content,emeric.martin}@onera.fr`
[2]Alpines, Inria, Sorbonne Université, Université de Paris, CNRS, Laboratoire Jacques-Louis Lions, Paris, France
`laura.grigori@inria.fr`

Abstract

For high dimensional spaces, a randomized Gram-Schmidt algorithm is beneficial in computational costs as well as numerical stability. We apply this dimension reduction technique by random sketching to Krylov subspace methods, e.g. to GMRES. We propose a flexible variant of GMRES with the randomized Gram-Schmidt based Arnoldi iteration to produce a set of basis vectors of the Krylov subspace. Even though the Krylov basis is no longer l_2 orthonormal, its random projection onto the low dimensional space shows l_2 orthogonality. As a result, it is observed the numerical stability which turns out to be independent of the dimension of the problem even in extreme scale problems. On the other hand, as the Harmonic Ritz values are commonly used in GMRES with deflated restarting to improve convergence, we consider another deflation strategy, for instance disregarding the singular vectors associated with the smallest singular values. We thus introduce a new algorithm of the randomized flexible GMRES with SVD based deflated restarting. At the end, we carry out some numerical experiments in the context of compressible turbulent flow simulations. Our proposed approach exhibits a quite competitive numerical behavior to existing methods while reducing computational costs.

Keywords— Gram-Schmidt orthogonalization · Randomization · Random sketching · Krylov subspace methods · Deflation · implicit restarting · SVD

1

Interpolating on the Stiefel manifold under a one-parameter family of metrics.

Ralf Zimmermann, IMADA, University of Southern Denmark

zimmermann@imada.sdu.dk

Abstract:

In this talk, we present efficient, customized matrix algorithms for solving the geodesic endpoint problem on the Stiefel manifold for a one-parameter family of Riemannian metrics that was first studied in [HMSL21]. This family includes the canonical and the Euclidean metric as the most prominent members. To solve the (local) geodesic endpoint problem is to compute the Riemannian logarithm. We provide a unified, structured, reduced formula for the Stiefel geodesics that simultaneously works for the full family of metrics under consideration. Moreover, we develop a unified method to tackle the geodesic endpoint problem numerically.

The findings are illustrated by numerical experiments, where we consider interpolation problems on the Stiefel manifold. We investigate, i.a., the numerical performance of the proposed methods and the dependency on the metric parameter. The talk is based on the joint work [ZH21] with Knut Hüper, Julius-Maximilians-Universität Würzburg.

References

[HMSL21] K. Hüper, I. Markina, and F. Silva Leite. A Lagrangian approach to extremal curves on Stiefel manifolds. *Journal of Geometrical Mechanics*, 13(1):55–72, 2021.

[ZH21] R. Zimmermann and K. Hüper. Computing the Riemannian logarithm on the Stiefel manifold: metrics, methods and performance, 2021. to appear in SIMAX.

Iterative methods for interior point algorithms appliedto the L2 Optimal Transport Problem, Enrica Facca

The Optimal Transport problem studies to optimal strategy for moving one non-negative density into another, minimizing the total cost of displacement. When the transport cost per unit mass is the squared distance, the problem can be rewritten in the so-called Benamou-Brenier formulation, a PDE-constrained optimization problem. This formulation provides a natural way to interpolate between probability densities, with applications in image processing, inverse problems, and generation of adaptive grids.

The Benamou-Brenier problem has been solved for a long time using first-order methods, that typically require a large number of iterations to obtain accurate solutions. In "Computation of Optimal Transport with Finite Volumes" (2021) Natale and Todeschi proposed a second-order approach based on interior point methods to solve a finite-dimensional approximation of the Benamou-Brenier problem obtained using Finite Volumes. This approach has been shown to converge to the optimal solution with a fixed number of interior point steps, independently on the size of the discretized problem.

However, these results were obtained using direct linear solver for solving the saddle point systems involved in Newton iteration used by the interior point method. This poses serious limitations to the application of the proposed method to large problems, thus iterative methods must be adopted. We present the difficulties arising in the design of efficient preconditioners for these saddle point linear systems, and some solutions we devised to cope with them.

Optimal control of bifurcation structures

Nicolas Boullé

Abstract

Many problems in engineering can be understood as controlling the bifurcation structure of a given device. For example, one may wish to delay the onset of instability, or bring forward a bifurcation to enable rapid switching between states. In this talk, we will describe a numerical technique for controlling the bifurcation diagram of a nonlinear partial differential equation by varying the shape of the domain or a parameter in the equation. Our aim is to delay or advance a given branch point to a target parameter value. The algorithm consists of solving an optimization problem constrained by an augmented system of equations that characterize the location of the branch points. The flexibility and robustness of the method also allows us to advance or delay a Hopf bifurcation to a target value of the bifurcation parameter, as well as controlling the oscillation frequency. We will apply this technique on systems arising from biology, fluid dynamics, and engineering, such as the FitzHugh–Nagumo model, Navier–Stokes, and hyperelasticity equations.

Factorized structure of the long-range two-electron integrals tensor and its
application in quantum chemistry

[1]Igor Chollet, [1]*Siwar Badreddine, [2]Laura Grigori.

*INRIA Paris, Alpines group, joint with Laboratoire J.L. Lions, Sorbonne University, France.

[1] Laboratoire de Chimie Théorique, 4 place Jussieu, France.

[2]INRIA Paris, Alpines group, joint with Laboratoire J.L. Lions, Sorbonne University, France.

One of the most challenging problems in computational chemistry is the accurate simulation of the electronic and molecular systems which implies solving the Schrodinger equation for a general N-body system. However, finding an exact, analytic solution of the Schrodinger equation becomes intractable for systems with more than one electron. Therefore, to further simplify the resolution, several additional approximations are considered. These approximations yield to involving the evaluation of the six-dimensional Two-Electron Integrals TEI. Such integrals have a singular function (Coulomb potential) and they are the entries of a fourth-order tensor that we need to compress. Therefore, we consider in our work an approach that relies on the range-separation of the Coulomb potential that splits it into a smooth range-part that we call the long-range potential and a complementary diverging part.

In this work, we propose two new compression methods for the efficient treatment of the long-range part as well as evaluation approaches for the resulting compressed TEI tensor. The first procedure exploits the tensorized structure of the compressed TEI obtained through two-dimensional Chebyshev interpolation combined with Gaussian quadrature, we call this first approach LTA for Long-range Tensorized Approximation. We derive error bounds and theoretical complexities for the approximation process we use. For the second method, we demonstrate that our kernel which is the long-range potential is asymptotically smooth so that we can benefit from fast hierarchical methods (especially Fast Multipole Methods FMM). The resulting factorized representations of the TEI tensor from both approaches are beneficial to speed up the tensor contractions involved in the electronic structure calculations. In fact, the storage and time complexity of the presented methods are analyzed and numerical comparisons are provided, exhibiting both the high efficiency of LTA and the linear complexity of the FMM-based approach.

Preconditioning Optimal Control of Incompressible Viscous Fluid Flow

Santolo Leveque & John Pearson *(University of Edinburgh)*

Optimal control problems with PDEs as constraints arise very often in scientific and industrial problems. Due to the difficulties arising in their numerical solution, researchers have put a great effort into devising robust solvers for this class of problems. An example of a highly challenging problem attracting significant attention is the (distributed) control of incompressible viscous fluid flow problems. In this case, the physics may be described, for very viscous flow, by the (linear) incompressible Stokes equations, or, in case the convection of the fluid plays a non-negligible role in the physics, by the (non-linear) incompressible Navier–Stokes equations. In particular, as the PDEs given in the constraints are non-linear, in order to obtain a solution of Navier–Stokes control problems one has to iteratively solve linearizations of the problems until a prescribed tolerance on the non-linear residual is achieved.

In this talk, we present novel, efficient, and robust preconditioned iterative methods for the solution of the time-dependent incompressible Stokes and Navier–Stokes control problems, with backward Euler discretization in time. The proposed preconditioner is based on a saddle-point type of approximation. We employ an inner iteration for the $(1,1)$-block accelerated by preconditioned iterative methods for heat and convection–diffusion control problems when employing backward Euler in time. In addition, in order to derive an approximation for the Schur complement we employ a potent and flexible commutator argument applied to an appropriate block matrix. The flexibility of the commutator argument, which is a generalization of the technique derived in [1], allows one to alternatively apply a Crank–Nicolson discretization in time, as well as to solve the stationary regimes for the problems considered. We show the effectiveness and robustness of our approach through a range of numerical experiments.

This talk is based on the work in [2] and [3].

References

[1] D. Kay, D. Loghin and A. J. Wathen, *A Preconditioner for the Steady-State Navier–Stokes Equations*, SIAM Journal on Scientific Computing **24**, 237–256, 2002.

[2] S. Leveque and J. W. Pearson, *Parameter-Robust Preconditioning for Unsteady Stokes Control Problems*, Proceedings in Applied Mathematics and Mechanics **21**, e202100131, 2021.

[3] S. Leveque and J. W. Pearson, *Parameter-Robust Preconditioning for Oseen Iteration Applied to Stationary and Instationary Navier–Stokes Control*, to appear in SIAM Journal on Scientific Computing, arXiv preprint arXiv:2108.00282.

Mode-matching analysis of eigenvalue problems arising in the study of two-dimensional acoustic waveguide structures

Rab Nawaz

Department of Mathematics, COMSATS University Islamabad, Pakistan.

Abstract:

Structural acoustics has gained significant intention towards the development of many engineering and applied mathematics problems alike. The major part of these contributions includes designing structures with the aim to control noise emanated from multiple sources. These naturally involve the wave scattering features at a discontinuity both in material properties and structures as investigated in the literature while studying the dissipative devices that are modelled for computational purpose. We study a mode-matching analysis of a two-dimensional waveguide problem subject to wave bearing boundaries. The underlying mathematical model characterizes the system to be non Sturm-Liouville whose solution suggests that how a choice of structure involving wave bearing boundaries is handled computationally. We aim to discuss the performance of reflected and transmitted regions of trifurcated lined waveguide backed by a line walled cavity involving multiple step discontinuities. Precisely we analyze the viability of the mode-matching technique and support our results with apposite numerical experiments. While performing the power balance the expressions for energy fluxes in different regimes are also the key finding for this study.

7th IMA Conference on Numerical Linear Algebra and Optimization, June 24-26, 2020
University of Birmingham, Birmingham, B15 2TT, UK

17

Title
Deflation in saddle-point systems following a Golub-Kahan bidiagonalization
Andrei Dumitrasc

Abstract
Deflation techniques are typically used to shift isolated clusters of small eigenvalues in order to obtain a tighter distribution and a smaller condition number. Such changes induce a positive effect in the convergence behavior of Krylov subspace methods, which are among the most popular iterative solvers for large sparse linear systems. We develop a deflation strategy for saddle-point matrices by taking advantage of their underlying block structure. The vectors used for deflation come from an elliptic singular value decomposition relying on the generalized Golub-Kahan bidiagonalization process. The block targeted by deflation is the off-diagonal one since it features a problematic singular value distribution for certain applications. One example is the Stokes flow in elongated channels, where the off-diagonal block has several small, isolated singular values, depending on the length of the channel. Applying deflation to specific parts of the saddle-point system is important when using solvers such as CRAIG, which operates on individual blocks rather than the whole system. The theory is developed by extending the existing framework for deflating square matrices before applying a Krylov subspace method like MINRES. Numerical experiments confirm the merits of our strategy and lead to interesting questions about using approximate vectors for deflation.

Primal-Dual Newton Proximal Method
for Convex Quadratic Programs, and Beyond

Alberto De Marchi[*]

Convex quadratic programs (QPs) have been studied since the 1950s; they arise in a huge variety of applications and play a key role in nonlinear optimization. This talk discusses a primal-dual numerical method for convex quadratic programming, which builds upon and weaves together the proximal point algorithm and a damped semismooth Newton's method. The approach aims at a robust and fast method, able to cope with degeneracy and large-scale instances, and with warm-starting (and possibly hot-starting) capabilities. A peculiar design choice is that of avoiding the introduction of nonlinearities to handle the inequality constraints, as it usually happens in interior point or Newton-type methods based on nonlinear complementarity functions. Instead, the method exploits projections onto the constraint set, that is just a hyperbox.

The outer primal-dual proximal regularization generates a sequence of feasible, strictly convex QPs. Each and every subproblem can be efficiently tackled by solving the associated KKT conditions via a semismooth Newton procedure. After characterizing the relationships between the subproblems and the primal-dual augmented Lagrangian function, it is demonstrated how the latter can be adopted as a merit function for the inner procedure, thus guaranteeing global and finite convergence. In particular, this observation allows to globalize Newton's method despite the nonsmoothness introduced by projections, while avoiding additional nonlinearities due to, e.g., the Fisher-Burmeister function. Then, the inner Newton's scheme can exploit sparse symmetric linear solvers and multi-rank factorization updates. Moreover, the linear systems are always solvable independently from the problem data and exact linesearch can be performed. While requiring only convexity, the proposed method can handle degenerate problems, provides a mechanism for infeasibility detection, and can exploit warm starting.

We discuss implementation details of our solver `QPDO`, an open-source software package written in C. Numerical tests and benchmarks against state-of-the-art QP solvers showed the robustness of the approach, yet with competitive performance in terms of iterations and computation time.

The results obtained for convex QPs can be readily extended to general convex programs. Furthermore, the combination of proximal regularization and semismooth Newton can be adopted to directly tackle nonlinear nonconvex programs. In this context, tracing back to the tight connection with proximal methods, the augmented Lagrangian framework offers insights and useful perspectives. In particular, the primal and dual regularization terms are associated with a discretized gradient/antigradient Lagrangian flow with separate time scales. Moreover, it is possible to interpret the inertia correction of the Newton matrix as the result of a primal-dual proximal regularization, effectively bringing the proposed approach into the nonconvex setting.

Reference

A. De Marchi, On a primal-dual Newton proximal method for convex quadratic programs. *Comput Optim Appl* 81, 369–395 (2022). DOI: 10.1007/s10589-021-00342-y

[*]Institute of Applied Mathematics and Scientific Computing, Department of Aerospace Engineering, Universität der Bundeswehr München, D-85577 Neubiberg/Munich, Germany.
EMAIL: alberto.demarchi@unibw.de, ORCID: 0000-0002-3545-6898.

Computing the square root of a low-rank perturbation of the scaled identity matrix, Xiaobo Liu

We consider the problem of computing the square root of a perturbation of the scaled identity matrix, $A = \alpha I_n + UV^*$, where U and V are $n\times k$ matrices with $k \le n$. This problem arises in various applications, including computer vision and optimization methods for machine learning. We derive a new formula for the pth root of A that involves a weighted sum of powers of the pth root of the $k\times k$ matrix $\alpha I_k + V^*U$. This formula is particularly attractive for the square root, since the sum has just one term when $p = 2$. We also derive a new class of Newton iterations for computing the square root that exploit the low-rank structure. We test these new methods on random matrices and on positive definite matrices arising in applications. Numerical experiments show that the new approaches can yield much smaller residual than existing alternatives and can be significantly faster when the perturbation UV^* has low rank.

Primal Dual Regularized IPM: a Proximal Point perspective

S. Cipolla[1], J.Gondzio[1]

[1]University of Edinburgh, School of Mathematics
`scipolla@ed.ac.uk, j.gondzio@ed.ac.uk`

Abstract

Computational evidence suggests that the Primal-Dual Regularization for Interior Point Methods (IPMs) is a successful technique able to stabilize and to speed-up the linear algebra used in IPM implementations [1]. On the other hand, many issues remain open when IPMs are used in their primal-dual regularized form and, in particular, to the best of our knowledge, the known convergence theory requires strong assumptions on the uniform boundedness of the Newton directions [2]. Recently, the study of the interaction of primal-dual regularized IPMs with the Augmented Lagrangian Method and the Proximal Point Algorithm has permitted to prove the convergence when the regularization parameter is driven to zero at a suitable speed [3].

In this talk, we will show that it is possible to naturally frame the primal-dual regularized IPMs in the context of the Proximal Point Algorithm [4]. Among the benefits of the proposed approach, we will show how convergence can be guaranteed without any supplementary assumptions and how the rate of convergence can be explicitly estimated in relation to (fixed) regularization parameter. Additionally, numerical results proving the efficiency and reliability of the proposed approach will be presented.

References

[1] Altman, A., Gondzio, J. (1999) Regularized symmetric indefinite systems in interior point methods for linear and quadratic optimization, Optimization Methods and Software, 11:1-4, 275-302.

[2] Friedlander, M. P., Orban D. (2012) A primal-dual regularized interior-point method for convex quadratic programs, Mathematical Programming Computation 4.1 : 71-107.

[3] Pougkakiotis, S., Gondzio J. (2021) An interior point-proximal method of multipliers for convex quadratic programming, Computational Optimization and Applications 78.2 : 307-351.

[4] Cipolla, S., Gondzio J. (In preparation) Proximal stabilized Interior Point Methods for quadratic programming and *low-frequency-updates* preconditioning techniques.

1

A semismooth Newton-proximal method of multipliers for ℓ_1-regularized convex quadratic programming

Spyridon Pougkakiotis*, `s.pougkakiotis@ed.ac.uk`,
*The University of Edinburgh.

Joint work with:
Prof. Jacek Gondzio*.

Abstract:

In this talk we present a method for the solution of ℓ_1-regularized convex quadratic optimization problems. It is derived by suitably combining a proximal method of multipliers strategy with a semi-smooth Newton method. The resulting linear systems are solved using a Krylov-subspace method, accelerated by appropriate general-purpose preconditioners, which are shown to be optimal with respect to the proximal parameters. We show that the method achieves global convergence under feasibility assumptions. Furthermore, under additional standard assumptions, the method achieves global linear and local superlinear convergence.

Several practical considerations are discussed. In particular, we introduce proximal terms for each variable of the problem, which yields sub-problems that can be efficiently tackled by standard semi-smooth Newton schemes available in the literature. Furthermore, we argue that practical efficiency is further improved by appropriately warm-starting the algorithm using a proximal alternating direction method of multipliers. We provide computational evidence which demonstrates the efficiency and robustness of the approach, as well as its ability to solve large-scale instances, by testing it over standard L^1-regularized PDE-constrained optimization problems.

PDE-constrained optimization for multiscale particle dynamics, Jonna Roden

There are many industrial and biological processes, such as beer brewing, nano-separation of colloids and bird flocking, which can be described by integro-PDEs. These PDEs describe the dynamics of a 'particle' density within a fluid bath, under the influence of diffusion, external forces, and particle interactions. They often include nonlinear, nonlocal boundary conditions.

A key challenge is to optimize these types of processes, which requires tools from PDE-constrained optimization. In this talk I will introduce a numerical method to solve this class of optimal control problems, which combines pseudospectral methods and spectral elements with a Newton-Krylov algorithm. This provides a tool for the fast and accurate solution of the resulting optimality systems.
In particular, this framework allows for the solution of (integro-)PDE models and optimal control problems on complex domains, which is a crucial feature in accurately describing various (industry) applications. Finally, some examples of current work and future industrial applications will be given. This is joint work with Ben Goddard and John Pearson.

A Sequential Homotopy Method for Mathematical Programming Problems, Andreas Potschka

We propose a sequential homotopy method for the solution of mathematical programming problems in Hilbert spaces under the Guignard constraint qualification. The method is equivalent to projected backward Euler timestepping on a projected gradient/antigradient flow of the augmented Lagrangian. The projected backward Euler equations can be interpreted as the necessaryoptimality conditions of a primal-dual proximal regularization of the originalproblem. The regularized problems are always feasible, satisfy a strongconstraint qualification guaranteeing uniqueness of Lagrange multipliers, yieldunique primal solutions provided that the stepsize is sufficiently small, andcan be solved by a continuation in the stepsize. We demonstrate its efficiencyfor challenging PDE optimization problems using a semismooth Newton method.

An Ordering Metaheuristic for Symmetric Sparse Direct Solvers

Marcel Jacobse, Christof Büskens
Center for Industrial Mathematics, University of Bremen, Germany

Abstract

Matrix ordering is well-known to be an important issue for the performance of direct methods for the solution of sparse symmetric linear equation systems [1]. Most prominent are minimum degree and nested dissection orderings that aim to reduce the amount of fill-in that occurs during Cholesky and LDL^T factorizations. These heuristics are viewed to provide good orderings in comparably short runtime.

In contrast, we examine a method that is intended to spend much more time to try to find better orderings. We consider the permutation optimization problem of minimum fill-in and tackle it with a genetic algorithm. New candidates are generated using mutation and recombination operators for the permutation representation [3]. The fitness of these candidate permutations is evaluated by computing the number of nonzeros in the Cholesky factor by performing the necessary parts of the symbolic factorization, for example using HSL_MC78 [4]. The population of permutations is then updated, preferring candidates with a low number of nonzeros and discarding others. For the initial population, the aforementioned minimum degree and nested dissection methods are obvious candidates.

Unsurprisingly, many iterations of this randomized approach are required to achieve significant progress. Thus, due to the prohibitive runtime, it would certainly not be useful in practice if the task is to only factorize a single matrix. Instead, the idea is to use it in applications that require the repeated solution of linear systems with a sparsity pattern that does not change. This setting often arises in numerical optimization, for example in interior point methods [6]. Especially in online optimal control applications, extra effort could be spent offline in advance to find a better ordering, that could then be used online to allow for faster computation of controls.

An advantage of the metaheuristic shows when considering that factorization performance is influenced by more than just the amount of fill-in. For instance, many sparse solvers employ supernodes to better use the memory hierarchy of modern computers by working on dense submatrices [1]. Because supernodes are formed from neighboring columns with similar sparsity patterns in the factor, the possibility to choose good supernodes also depends on the chosen ordering. In addition to minimizing the amount of fill-in, a second goal may therefore be to find orderings that yield efficient supernode structures. In the general framework of the outlined genetic algorithm, such a second objective is quite simple to add. The problem then becomes multiobjective and our genetic algorithm produces sets of nondominated permutations [3].

For indefinite matrices, another aspect to be considered is the numerical stability of the factorization. With different approaches for pivoting and with dependency on the actual matrix values rather than just the sparsity pattern, this proves to be more difficult to incorporate. Nevertheless we suggest different approaches, for example inspired by symmetric weighted matching [2] or related to saddle-point [7] and quasidefinite matrices [8].

We present results of the genetic algorithm for small to medium sized test matrices. For evaluation, we benchmark the runtime of different direct solvers such as HSL_MA97 [5] using the found orderings in comparison to the orderings given by state-of-the-art minimum degree and nested dissection implementations. We observe runtime reductions of up to about 30% with the improved orderings.

"A Primal-Dual Augmented Lagrangian Method for the Solution of Semidefinite Programs

Arefeh Kavand, Michal Kocvara, Michel Stingl

In this talk, a Penalty/Barrier-Multiplier Augmented Lagrangian (PBM-AL) method for the solution of Semidefinite Programs (SDP) is introduced. We present a dual convergence analysis of the method. For this, the relation to a proximal point algorithm with a specific distance function computed from the Fenchel conjugate of the Penalty-Barrier function is investigated. Based on the dual convergence results also convergence for the primal (PBM-AL) can be established.

We further suggest to augment the (PBM-AL) framework by a primal-dual approach to solve sub-problems. An early stopping criterion can be applied for the inner iteration which renders the overall method more stable. We show that the primal-dual Newton systems can be solved in two steps involving the solution of a linear system with the primal variable and a modified multiplier update. We further demonstrate that the linear system can be approached using either direct solvers or the preconditioned conjugate gradient method.

We also remark on how these changes affect convergence theory.

For the CG method, a preconditioner is suggested, which assumes that the matrix multiplier is low-rank. The efficiency of the overall algorithm is demonstrated by numerical experiments for different classes of SDP problems."

Multiobjective Optimization without Scalarization: Heterogeneous Problems with Expensive Functions
Gabriele Eichfelder

In multiobjective optimization, one considers optimization problems with several competing objective functions. For instance, in engineering, a design often has to be stable and light at the same time. A classical approach to such optimization problems is to formulate suitable parameter-dependent single-objective replacement problems, called scalarization, such as considering a weighted sum of the objective functions. Then, the parameters are varied and the scalarized problems are solved iteratively.

However, many multiobjective optimization problems have a structure where a scalarization is not a suitable approach for an efficient procedure. In this talk, we give an introduction to the basic concepts and classical approaches in multiobjective optimization. Then, we present such classes of multiobjective optimization problems where it is better not to scalarize. For specific heterogeneous problems, where one of the objective functions is assumed to be an expensive black-box function while the other objectives are analytically given, we give more details on a numerical approach. That method uses the basic trust region concept by restricting the computations in every iteration to a local area. The objective functions are replaced by suitable models which reflect the heterogeneity of the objective functions.

Anymatrix: an extensible MATLAB matrix collection

by Mantas Mikaitis and Nicholas J. Higham

Department of Mathematics, The University of Manchester, Manchester, M13 9PL, UK

Anymatrix is a MATLAB toolbox that provides an extensible collection of matrices with the ability to search the collection by matrix properties. Each matrix is implemented as a MATLAB function and the matrices are arranged in groups. Compared with previous collections, Anymatrix offers three novel features. First, it allows a user to share a collection of matrices by putting them in a group, annotating them with properties, and placing the group on a public repository, for example on GitHub; the group can then be incorporated into another user's local Anymatrix installation. Second, it provides a tool to search for matrices by their properties, with Boolean expressions supported. Third, it provides organization into sets, which are subsets of matrices from the whole collection appended with notes, which facilitate reproducible experiments. Anymatrix `v1.1` comes with 150 built-in matrices organized into 7 groups with 49 recognized properties. The authors continue to extend the collection and welcome contributions from the community. It is available at `https://github.com/mmikaitis/anymatrix`. Below are some examples of usage in MATLAB.

```
% List all matrix IDs in the collection.
>> m = anymatrix('all')
m =
  150×1 cell array
    {'contest/baitsample'    }
    {'contest/curvature'     }
    {'contest/erdrey'        }
    {'contest/geo'           }
    {'contest/gilbert'       }
[...]

% List available groups.
>> g = anymatrix('groups')
g =
  7×1 cell array
    {'contest' }
    {'core'    }
    {'gallery' }
    {'hadamard'}
    {'matlab'  }
    {'nessie'  }
    {'regtools'}

% List the matrices in a particular group.
>> g = anymatrix('groups','core')
g =
  26×1 cell array
    {'core/augment'       }
    {'core/beta'          }
    {'core/blockhouse'    }
    {'core/collatz'       }
    {'core/dembo9'        }
    {'core/edelman27'     }
    {'core/fourier'       }
    {'core/gfpp'          }
[...]
```

```
% Show properties of the beta matrix.
>> p = anymatrix('properties','core/beta')
p =
  12×1 cell array
    {'built-in'            }
    {'infinitely divisible'}
    {'integer'             }
    {'nonnegative'         }
    {'positive'            }
    {'positive definite'   }
    {'real'                }
[...]

% Generate the 5-by-5 beta matrix.
>> A = anymatrix('core/beta',5)
A =
     1     2     3     4     5
     2     6    12    20    30
     3    12    30    60   105
     4    20    60   140   280
     5    30   105   280   630

% List all matrices that have integer
% entries and are positive definite.
>> anymatrix('properties', ...
     'integer and positive definite')
ans =
  7×1 cell array
    {'core/beta'      }
    {'core/wilson'    }
    {'gallery/gcdmat' }
    {'gallery/minij'  }
    {'gallery/moler'  }
    {'gallery/pei'    }
    {'matlab/pascal'  }
```

Speaker: Adilet Otemissov

Joint with: Coralia Cartis and Estelle Massart

Title: A dimensionality reduction technique for global optimization via random subspace embeddings

Abstract:

In an attempt to improve the scalability of global optimization problems, we propose a general random subspace algorithmic framework, which tackles the (high-dimensional) global optimization problem by repeatedly and possibly adaptively solving reduced subproblems with the domains restricted to be along randomly embedded low-dimensional subspaces. We analyse the convergence of the proposed framework for Lipschitz continuous functions using tools from conic integral geometry and random matrix theory. We then particularise the framework and analysis for the class of functions with low effective dimensionality, which are constant along an (unknown) linear subspace and only vary over the effective (complement) subspace. We show that for these functions the convergence could potentially be exponentially better and that, under certain assumptions, it does not depend on the ambient dimension. Our numerical experiments on functions with low effective dimensionality illustrate two things: 1) the improved scalability of the framework when coupled with state-of-the-art global — and even local — optimization solvers for the subproblems 2) the ability of the framework to effectively estimate the (initially unknown) effective dimensionality of the objective function.

Further references:

https://arxiv.org/abs/2003.09673

https://arxiv.org/abs/2009.10446

https://arxiv.org/abs/2107.12102

Learning in High-Dimensional Feature Spaces Using ANOVA-Based Fast Matrix-Vector Multiplication

Theresa Wagner, TU Chemnitz

(joint work with Martin Stoll and Franziska Nestler)

March 24, 2022

Abstract

Kernel matrices are crucial in many learning tasks and typically dense and large-scale. Depending on the dimension of the feature space even the computation of all its entries in reasonable time becomes a challenging task. For such dense matrices the cost of a matrix-vector product scales quadratically with the dimensionality, if no customized methods are applied. In basically all kernel methods, a linear system must be solved. Our approach exploits the computational power of the non-equispaced fast Fourier transform (NFFT), which is of linear complexity for fixed accuracy. The ANOVA kernel has proved to be a viable tool to group the features into smaller pieces that are then amenable to the NFFT-based summation technique. Multiple kernels based on lower-dimensional feature spaces are combined, such that kernel-vector products can be realized by this fast approximation algorithm. Based on a feature grouping approach this can be embedded into a CG solver within a learning method and we nearly reach a linear scaling. This approach enables to run learning tasks using kernel methods for large-scale data on a standard laptop computer in reasonable time without or very benign loss of accuracy. It can be embedded into methods that rely on kernel matrices or even graph Laplacians. Examples are support vector machines or graph neural networks that can then benefit from having the fast matrix-vector products available.

1

The Continuous Stochastic Gradient Method: A New Approach for Stochastic Optimization

Lukas Pflug, Michael Stingl,

Applied Mathematics (Continuous Optimization), Department of Mathematics,
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Cauerstr. 11, 91058 Erlangen, Germany
lukas.pflug@fau.de, michael.stingl@fau.de

To solve problems in machine learning as well as in PDE constrained optimization with a high (or infinite) number of scenarios, stochastic optimization methods are commonly used. Existing schemes are based on the concept of the *stochastic gradient method* (SG) [2]. In this presentation, a new scheme, the so called *continuous stochastic gradient method* (CSG) [1], which is a generalization of SG and its relative, the *stochastic average gradient method* (SAG) [3] is discussed.

In generalization of [1], here we aim to solve optimization problems of the generic form:

$$\min_{u \in \mathcal{U}_{ad}} J(u) \qquad \text{with} \qquad J(u) := \int_{\mathcal{X}} j(u, x) \, \mathrm{d}\mu(x).$$

In the machine learning context, j denotes the loss functional, u is the parameter vector for a neural network and x is a random variable, who's realizations correspond to inputs in the set of input-output pairs. On the other hand, in the PDE context, u plays the role of a design or control, j denotes a state dependent cost and x a random variable which models the parameter dependence of the PDE. Moreover $\mathcal{U}_{ad}$ is the set of admissible network parameters or controls and $\mathcal{X}$ is the space in which the random variable is defined with its probability measure μ.

In contrast to SG and SAG the new CSG method relies on the Lipschitz continuity of $\nabla_1 j$ with respect to both variables, x and u. With this assumption we define an algorithm, which is able to reuse already computed gradient information and, with the help of this, approximate the full gradient better and better in the course of the iterations. This is exploited to prove that both, the objective functional J as well as its gradient ∇J convergence almost surely.

A crucial ingredient of the algorithm is the availability of weights $(\alpha_k^n)_{k=1,\dots,n}$, with the help of which in iteration n the full gradient evaluated at the iterate u_n is approximated in the following sense:

$$\nabla J(u_n) = \int_{\mathcal{X}} \nabla_1 j(u_n, x) \, \mathrm{d}\mu(x) \approx \sum_{k=1}^{n} \alpha_k^n \nabla_1 j(u_k, x_k).$$

In the presentation it will be shown how these weights can be computed in the case where the probability measure μ is known as well as in the case where only μ-distributed data are accessible.

Finally the effectiveness and efficiency of the CSG approach will be demonstrated in comparison to SG and SAG by means of a few examples covering both, machine learning and PDE constrained optimization.

References

[1] Lukas Pflug, Niklas Bernhardt, Max Grieshammer, and Michael Sting. Special issue: A new stochastic gradient method for the efficient solution of structural optimization problems with infinitely many state problems. *Structural and Multidisciplinary Optimization*, 2020 (in press).

[2] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

[3] Mark Schmidt, Nicolas Le Roux, and Francis Bach. Minimizing finite sums with the stochastic average gradient. *Mathematical Programming*, 162(1):83–112, Mar 2017.

Randomized, Adaptive Linear System Solvers

Vivak Patel

Solving consistent linear systems remains a critical operation in scientific and engineering applications. As the size of systems or the sheer number of systems that need to be solved grow, faster and approximate linear system solvers have become essential to scalability. Recently, randomized linear system solvers have become of interest as they can compress the information in the original linear system in a problem-blind fashion, which can then be used to inexpensively and approximately solve the original linear system. Moreover, by iterating on this procedure, randomized linear system solvers will converge exponentially fast to the solution of the original system.

Despite the desirable nature of a problem-blind compression of linear systems, randomized linear system solvers can suffer from poor performance if they fail to either account for problem structure or hardware considerations. For instance, a randomized linear system solver that samples equations of a linear system that are not contiguously held in memory will incur expensive Input and Output (I/O) costs. Owing to this issue, randomized adaptive linear system solvers have been explored, especially those that take advantage of efficient block operations. However, such randomized adaptive linear system solvers are likely to be without theoretical foundations with only a handful of exceptions, which certainly do not cover the entire spectrum of possible reasonable methods. As a result, the vast majority of randomized adaptive linear system solvers cannot be used reliably, which limits the proliferation and deployment of such solvers.

To address this limitation, we need a general, flexible theory for such randomized, adaptive solvers. In this talk, we supply such a theory. We begin by distilling three key properties satisfied by the majority of this class of linear system solvers, and showing that solvers that satisfy these properties will converge exponentially to a solution. Owing to our theory, not only do we provide a unified treatment of over 80 years of methods, we also enable the reliable development and deployment of randomized, adaptive linear system solvers.

Robust and reliable numerical linear algebra – applications and implementations

Zlatko Drmač

Department of Mathematics, Faculty of Science, University of Zagreb
`https://web.math.pmf.unizg.hr/~drmac/`

Abstract.

The goal of this talk is twofold.

First, we show how to deploy the state of the art numerical linear algebra perturbation theory and algorithms to curb ill-conditioning that precludes successful numerical implementation of sophisticated methods in applications. Some instances of ill-conditioning (loss of accuracy due to large matrix condition numbers) do not reflect the intrinsic sensitivity of the problem, but are due to a particular algorithm executed in finite precision arithmetic. Suitably constructed algorithms can compute with particularly structured backward errors that render such ill-conditioning harmless and deliver more accurate results with a more favorable measure of sensitivity. For instance, we can compute the (generalized) eigenvalues of positive definite matrices (pencils) to the best possible accuracy - as accurately as the data deserves - independent of the traditional condition number.
As case studies we use selected problems from the computational analysis of nonlinear systems (Koopman operator based spectral analysis and data driven identification/learning of nonlinear dynamical systems). In particular, we show how to compute with notoriously ill-conditioned Vandermonde and Cauchy matrices, and how to solve linear least squares problems with the matrix that involves the Khatri-Rao product of a triangular and a Vandermonde matrix. Further, we discuss computing matrix approximation of the infinitezimal generator of the Koopman operator semigroup. This involves matrix logarithm of potentially severely ill-conditioned matrices. We show how implicit preconditioning helps alleviate the problem.

Secondly, we discuss the issue of the reliability of the numerical software. We use a case study example (rank revealing QR factorization) to show how a numerical instability and a pure software bug can remain undetected for decades in the state of the art libraries such as LAPACK, ScaLAPACK, Matlab, despite many tests (all passed) and extensive usage.

We conclude with a general discussion on the importance of numerical analysis of computational algorithms, that must include the concrete software implementation.

HiGHS: Beyond the Idiot crash
Julian Hall

HiGHS has emerged in the past couple of years as the world's best open-source linear optimization software. The story began in 2016 with funding to study the Idiot crash in the Clp open-source simplex solver for linear programming (LP) problems. As well as introducing HiGHS, this talk will discuss how the point obtained by the Idiot crash can best be used to accelerate the solution of LP problems via the HiGHS simplex solver.

Russell Luke

title: Convergence of Proximal Splitting Algorithms in Positively Curved Spaces

abstract: In linear spaces, proximal splitting methods are widely employed to solve large-scale problems efficiently.
One of the main theoretical properties guaranteeing convergence of these methods is nonexpansivity - i.e.
Lipschitz continuity with constant 1; this property comes ``for free'' in the
presence of convexity. In positively curves spaces, like the surface of the sphere,
proximal mappings generated from convex functions are no longer nonexpansive. Nevertheless, we can show
that common fixed point iterations built from prox mappings (e.g. prox-prox, Krasnoselsky–Mann relaxations, nonlinear projected-gradients) converge locally linearly under related, but weaker assumptions.
To show this, we develop a theory of fixed point mappings that violate the usual assumptions of nonexpansiveness in p-uniformly convex spaces.

Randomized Algorithms for Tikhonov Regularization in Linear Least Squares

Maike Meier and Yuji Nakatsukasa

Tikhonov regularization (or ridge regression) is a regularization technique for linear least squares (LLS) problems. Regularization may be necessary if the design matrix $A \in \mathbb{R}^{m \times n}$ is ill-conditioned or if the problem is underdetermined. The regularized problem takes the form

$$\min_{x \in \mathbb{R}^n} \left\{ \|Ax - b\|_2^2 + \lambda \|x\|_2^2 \right\}, \tag{1}$$

for a regularization parameter $\lambda > 0$. The optimal regularization parameter is usually unknown a priori and is to be determined in an ad hoc manner, which may involve solving (1) for a number of regularization parameters $\lambda_1, \ldots, \lambda_N$. We describe two algorithms based on sketching to efficiently solve this set of problems. The first is applicable in a general setting, the second is capable of exploiting (approximate) low-rank structure in A. The algorithms compute preconditioners for (1) such that LSQR converges in $\mathcal{O}(\log(1/\varepsilon))$ iterations for ε accuracy. Our algorithms are applicable in both the underdetermined ($m \ll n$) and the overdetermined ($m \gg n$) setting. We present convergence results for both algorithms based on the structural conditions presented in [3].

Firstly, we propose a Cholesky-based sketch-to-precondition algorithm that only requires one sketch for a set of N regularization parameters λ. The two main contributions are 1) the use of a 'partly exact' [2] sketch specific to Tikhonov regularization with multiple λ-parameters and 2) the use of the more efficient Cholesky decomposition instead of the QR decomposition or the SVD. The complexity of solving for N parameters to ε relative accuracy with preconditioned LSQR is $\mathcal{O}(mn \log(\max(m, n)) + N(\min(m, n)^3 + mn \log(1/\varepsilon)))$.

Secondly, we introduce an algorithm that exploits (approximate) low-rank structure in A. The quantity of interest here is the statistical dimension, $\mathrm{sd}_\lambda(A) = \sum 1/(1 + \lambda/\sigma_i(A)^2)$. This has received significant attention in the randomized NLA literature as recent work [1, 2] shows the sketch dimension can be of the same order as the statistical dimension (as opposed to proportionality to $\min(m, n)$). Assuming a small sketch, we present a scheme that can solve for N values of λ_i in $\mathcal{O}(mn \log(\max(m, n)) + \min(m, n) \, \mathrm{sd}_{\min \lambda_i}(A)^2 + Nmn \log(1/\epsilon))$ operations, avoiding $\mathcal{O}(N \min(m, n)^3)$ work as only one decomposition in total is required. Our preconditioners have a low-rank structure, with rank proportional to $\mathrm{sd}_{\lambda_i}(A)$, that allow them to be applied quickly. We are able to truncate the preconditioners to the appropriate rank for different λ_i without additional cost using singular value estimation based on [4].

Although our work is close in spirit to [1] (translated to an LLS context) and [3], these works are concerned with solving the normal equations. The scheme we propose does not require the computation of the Gram matrix, resulting in a more stable scheme. As far as the authors are aware, this is the only sketch-to-precondition algorithm that efficiently makes use of a sketch dimension smaller than $\min(m, n)$ while avoiding the normal equations. The details of both algorithms, and the corresponding convergence analyses, can be found in [5].

[1] H. Avron, K. Clarkson, and D. Woodruff. Faster kernel ridge regression using sketching and preconditioning. *SIAM J. Matrix Anal. Appl*, 38 (4): 1116-1138, 2017

[2] H. Avron, K. Clarkson, and D. Woodruff. Sharper bounds for regularized data fitting. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 81: 27:1-27:22, 2017

[3] A. Chowdhury, J. Yang, and P. Drineas. An iterative, sketching-based framework for ridge regression. *International Conference on Machine Learning (ICML)*, 3: 1595-1626, 2018

[4] M. Meier and Y. Nakatsukasa. Fast randomized numerical rank estimation. *arXiv:2105.07388*, 2021

[5] M. Meier and Y. Nakatsukasa. Randomized algorithms for Tikhonov regularization in linear least squares. *arXiv:2203.07329*, 2022

IFISS: A computational laboratory for investigating incompressible flow problems
David Silvester

Abstract

The Incompressible Flow & Iterative Solver Software (IFISS) package contains software that can be run with MATLAB or Octave to create a computational laboratory for the interactive numerical study of incompressible flow problems. It includes algorithms for discretisation by mixed finite element methods and a posteriori error estimation of the computed solutions, together with state-of-the-art preconditioned iterative solvers for the resulting discrete linear equation systems. In this talk we give a flavour of the main features and illustrate its applicability using several case studies. We will demonstrate that the software is a valuable tool in the present era of open science and reproducible research.