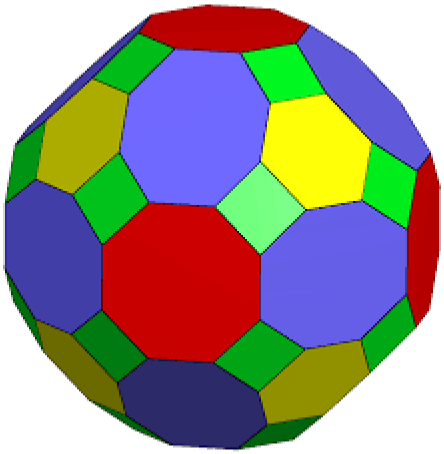
Workshop on Advances in Optimization



August 12-13, 2016

TKP Shinagawa Conference Center
Tokyo, JAPAN

The workshop will bring together a diverse group of researchers from both continuous and combinatorial optimization, theoretical and applied. We will celebrate the 60th birthday of Shinji Mizuno at the conference banquet.

Organizers: Antoine Deza, Tomonari Kitahara, and Noriyoshi Sukegawa

This workshop is partially supported by a Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research (A) 26242027

Workshop on Advances in Optimization

Program

The workshop will be held in TKP Shinagawa Conference Center, room 4E (4th floor), Tokyo, Japan.

Friday, August 12

09h30 – 10h00 Reception/Coffee/Opening Remark

10h00 – 12h00 : Session A1

On the worst-case complexity of the gradient method with exact line search for strongly convex functions with Lipschitz gradients

Etienne de Klerk (Tilburg University)

Sharpe ratio optimization

Michael Metel (Université de Paris Sud)

A polynomial contraction algorithm for a class of convex problems

Sergei Chubanov (University of Siegen)

Polynomial Time Iterative Methods for Integer Programming

Shmuel Onn (Technion)

12h00 – 13h30 Lunch

13h30 – 15h30 : Session A2

Ellipsoid method vs accelerated gradient descent

Yin Tat Lee (Microsoft)

On the exact convergence rates of gradient methods for smooth (strongly) convex optimization

François Glineur (Université catholique de Louvain)

Mixed-integer SOCP in optimal contribution selection of tree breeding

Makoto Yamashita (Tokyo Institute of Technology)

Squared slack variables in nonlinear symmetric cone programming: Optimality conditions and augmented Lagrangians
Bruno Figueira Lourenço (Seikei University)

15h30 – 16h00 : Coffee

16h00 – 18h00 : Session A3
An improved two-stage optimization-based framework for unequal-areas facility layout
Miguel Anjos (Polytechnique Montréal)

Full-, twin- and Lagrangian-DNN relaxations for binary quadratic optimization Problems
Masakazu Kojima (Chuo University)

On semismooth Newton based augmented Lagrangian method for lasso-type problems
Kim Chuan Toh (National University of Singapore)

The simplex method for degenerate and nondegenerate linear programming problems
Shinji Mizuno (Tokyo Institute of Technology)

18h30 : Banquet celebrating Shinji Mizuno 60th birthday

Saturday, August 13

09h30 – 10h00 Reception/Coffee

10h00 – 12h00 : Session B1
Exact duals and short certificates of infeasibility and weak infeasibility in conic linear programming
Gabor Pataki (University of North Carolina at Chapel Hill)

Disjunctive conic cuts for hyperboloids and non-convex quadratic cones
Julio Góez (Norwegian School of Economics)

Analysis on conic programing via facial reduction
Masakazu Muramatsu (University of Electro-communications)

A second-order cone based approach for solving the trust region subproblem and its variants
Fatma Kilinc Karzan (Carnegie Mellon University)

12h00 – 13h30 Lunch

13h30 – 15h30 : Session B2

Centerpoints: A link between optimization and convex geometry
Amitabh Basu (Johns Hopkins University)

Feature subset selection for logistic regression via mixed integer optimization
Yuichi Takano (Senshu University)

Geometric Median in Nearly Linear Time
Aaron Sidford (Stanford University)

Random-edge is slower than random-facet on abstract cubes
Thomas Dueholm Hansen (Aarhus University)

15h30 – 16h00 : Coffee

16h00 – 18h00 : Session B3
Margin and overlap for convex bodies
David Bremner (University of New Brunswick)

Implementation of interior-point methods for LP using Krylov subspace methods preconditioned by inner iterations
Keiichi Morikuni (University of Tsukuba)

A comparative study of steepest descent methods for strongly convex quadratic functions
Mituhiro Fukuda (Tokyo Institute of Technology)

Linear optimization: Algorithms and conjectures
Tamás Terlaky (Lehigh University)

18h00 : Closing remarks

On the worst-case complexity of the gradient method with exact line search for strongly convex functions with Lipschitz gradients

Etienne de Klerk

Tilburg University

abstract

We consider the gradient (or steepest) descent method with exact line search when applied to a strongly convex function with Lipschitz continuous gradient. We establish the exact worst-case rate of convergence, and show that this worst-case behavior is exhibited by a certain convex quadratic function. The proof is computer-assisted, and uses SDP performance evaluation problems as introduced in the paper [Y. Drori and M. Teboulle]. Performance of first-order methods for smooth convex minimization: a novel approach. Mathematical Programming, 145(1-2):451-482, 2014]. Joint work with Francois Glineur and Adrien Benoit Taylor (UCL, Louvain-la-Neuve).

Sharpe ratio optimization

Michael Metel

McMaster University

abstract

The Sharpe ratio is a popular means of evaluating stock portfolio management performance, considering the risk adjusted return over a given time period. We review its formulation as a convex problem, as well as present an active-set algorithm with a numerical study.

A polynomial contraction algorithm for a class of convex problems

Sergei Chuvanov

University of Siegen

abstract

We consider a class of convex optimization problems which are reducible in polynomial time to the problem of finding a point in the intersection of two closed convex sets such that their intersection is a subset of a unit cube. We suppose that one of the convex sets is a polyhedron having some special form and that for each of the two sets there is a polynomial algorithm for computing a mapping which is non-expansive with respect to this set, i.e., which does not increase the distance from a given point to the respective convex set. We call such mappings conditionally non-expansive. This class of mappings contains e.g. orthogonal projections.

The method of alternating projections or, for short, the alternating projections is a simple and popular classical approach for this type of problems in the case when the respective conditionally non-expansive mappings are orthogonal projections. Each iteration of the alternating projections consists in projecting the current point onto one of the sets and then onto the other set. We will consider a modification of the alternating projections where the projections are replaced by conditionally non-expansive mappings. We call this method the alternating contractions.

Neither the alternating projections nor the alternating contractions are polynomial algorithms even if both sets in the definition of the convex feasibility problem are straight lines. This follows from well-known results on the alternating projections and related methods such as the relaxation method for linear inequalities.

Nevertheless, it turns out that we can develop a polynomial algorithm on the basis of the alternating contractions for a fairly large class of convex problems and everything we need is to run a certain number of instances of the alternating contractions simultaneously, each instance being applied to an appropriately defined subproblem of the original convex feasibility problem. At each iteration, we keep track of the progress of each instance of the alternating contractions and check whether the geometric center (the centroid) of the set of the points which are currently considered by the instances of the alternating contractions is a feasible solution to the convex feasibility problem. If the centroid is not feasible, then, at the next iteration, at least one of

the instances of the alternating contractions will guarantee a progress toward the set of feasible solutions of the respective subproblem. This means that if the centroid is still not feasible after a polynomial number of iterations of the described procedure, we can conclude that some of the subproblems are infeasible. This allows us to simplify or rescale the problem. In a polynomial number of simplifications and rescalings, we either find a feasible solution or prove that the convex feasibility problem is infeasible.

This method leads to new polynomial algorithms for linear programming, convex quadratic programming, and for a class of combinatorial problems which can be formulated as linear problems where the constraints are given by a separation oracle, including the maximum matching problem in general graphs.

Polynomial Time Iterative Methods for Integer Programming

Shmuel Onn

Technion

abstract

I will overview our theory of Graver-based iterative methods that enable to solve in polynomial time broad important classes of integer programs. I will then describe a recent drastic improvement that shows that these problems can be solved in cubic time and are fixed-parameter tractable. I will conclude by describing the state of the art on multicommodity flows.

Ellipsoid method vs accelerated gradient descent

Yin Tat Lee

Microsoft

abstract

We propose a new method for unconstrained optimization of a smooth and strongly convex function, which attains the optimal rate of convergence of Nesterov's accelerated gradient descent. The new algorithm has a simple geometric interpretation, loosely inspired by the ellipsoid method. We provide some numerical evidence that the new method can be superior to Nesterov's accelerated gradient descent.

On the exact convergence rates of gradient methods for smooth (strongly) convex optimization

François Glineur

Universite catholique de Louvain

abstract

This talk focuses on the exact worst-case performance of several black-box first-order methods for smooth (strongly) convex optimization. We consider (a) the classical gradient method for unconstrained optimization of a smooth convex function, (b) its variant for the constrained case (which projects on the convex feasible domain at each iteration), and (c) its proximal variant, which deals with a composite objective function where one term is smooth and the other admits an easy-to-compute proximal operator.

For each method, we provide exact worst-case rates when applied to strongly convex function functions, characterizing successively the distance to the solution, the norm of the gradient (suitably modified in the constrained and proximal cases) and the objective function accuracy after a certain number of iterations. This type of global (non-asymptotic) complexity result requires an assumption on the initial iterate, and here also we consider three variants (initial distance, initial gradient norm and initial objective accuracy), providing thus a complete picture.

We also discuss the use of a backtracking strategy to avoid knowledge of the smoothness and strong convexity constants, and the possibly relaxation of the strong convexity assumption.

Mixed-integer SOCP in optimal contribution selection of tree breeding

Makoto Yamashita

Tokyo Institute of Technology

abstract

Mathematical optimization are becoming an indispensable tool in tree breeding researches. Among many optimization problems in tree breeding, the purpose of optimal contribution problems is to determine the contributions of genotypes. We should obtain the contributions that attain the highest gain, but at the same time, we need a careful consideration on the genetic diversity to keep the long-term performance. It is recognized recently that the constraint for the genetic diversity can be expressed by a positive semidefinite matrix or a second-order cone.

The optimal contribution problems can be classified into two groups, unequally contribution problems and equally contribution problems. In equal contribution problems, the contribution of each selected genotype must be same. More precisely, if we choose N genotypes from the candidates, the contribution of each selected genotypes must be exactly 1/N. The equally contribution problems can be mathematically modeled as mixed- integer SOCP problems. However, the equally contribution problems are very hard mainly because the constraint that requires the sum of the contributions be unity. In this talk, we apply three conic relaxations (LP, SOCP and SDP relaxations) to the equally contribution problems and discuss their relations. We also combine a steep-ascent method with the conic relaxations.

From numerical experiments, we verify that our method can output suitable solutions in a practical time. Its computation time is much shorter than an existing method widely used in tree breeding and a branch-and- bound method for mixed-integer SOCP problems.

Squared slack variables in nonlinear symmetric cone programming: Optimality conditions and augmented Lagrangians

Bruno Figueira Lourenço

Seikei University

abstract

In this work we are interested in optimality conditions for *nonlinear symmetric cone programs* (NSCPs), which contains as special cases optimization problems with positive semidefinite constraints and second-order cone constraints. We continue the study initiated by Fukuda and Fukushima [1]and we reformulate an NSCP as a classical nonlinear problem (NLP) using *squared slack variables*. First, we study the correspondence between Karush-Kuhn-Tucker points and constraint qualifications between the original NSCP and its reformulation. Exploring this connection, we show that we can obtain second order optimality conditions (SOC) for NSCP in a fairly elementary way. As a by-product, we also obtain a new "sharp" criteria for testing membership in a symmetric cone in a way that encodes information about the rank of the element.

For the case where the underlying cone is the positive semidefinite matrices, we will discuss recent results [2] showing that the resulting SOCs are essentially equivalent to the ones derived by Shapiro using techniques from variational analysis [4], which suggests that the squared slack variables can encode a lot of the structure of the cone.

Finally, we will discuss how to explore the NLP reformulating in order to design and analyze augmented Lagrangian methods for NSCPs, while avoiding certain technical pitfalls described by Noll [3]. Moreover, through careful considerations, we can eliminate the slack variables from the final algorithm thus obtaining a method that acts directly on the original formulation.

References

[1] Ellen H. Fukuda and Masao Fukushima. The use of squared slack variables in nonlinear second-order cone programming. *To appear in the Journal of Optimization Theory and Applications*, 2016. URL: http://www.optimization-online.org/DB_HTML/2015/12/5249.html.

[2] Bruno F. Lourenço, Ellen H. Fukuda, and Masao Fukushima. Optimality

conditions for nonlinear semidefinite programming via squared slack variables. *To appear in Mathematical Programming*, 2016.
URL: http://arxiv.org/abs/1512.05507.

[3] Dominikus Noll. Local convergence of an augmented lagrangian method for matrix inequality constrained programming. *Optimization Methods and Software*, 22(5):777–802, 2007.

[4] Alexander Shapiro. First and second order analysis of nonlinear semidefinite programs. *Math. Program.*, 77(1):301–320, 1997.

An improved two-stage optimization-based framework for unequal-areas facility layout

Miguel Anjos

Polytechnique Montréal

abstract

The facility layout problem seeks the optimal arrangement of non-overlapping departments with unequal areas within a facility. We present an improved framework combining two mathematical optimization models. The first model is a nonlinear approximation that establishes the relative position of the departments, and the second model is an exact convex optimization formulation of the problem that determines the final layout. Aspect ratio constraints are taken into account by both models. Preliminary results show that the proposed framework is computationally efficient and consistently produces competitive, and often improved, layouts for instances with up to 100 departments.

Full-, twin- and Lagrangian-DNN relaxations for binary quadratic optimization Problems

Masakazu Kojima

Chuo University

abstract

We consider a binary quadratic optimization problem (BQOP): minimize a quadratic function $\boldsymbol{x}\boldsymbol{Q}\boldsymbol{x}^T + 2\boldsymbol{c}\boldsymbol{x}^T$ in the n-dimensional variable row vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ subject to the binary constraint $\boldsymbol{x} \in \{0,1\}^n$. This problem is known as one of the most basic combinatorial optimization problems. Starting from the standard doubly nonnegative (DNN) relaxation, we derive some DNN relaxations for lower bounds of the optimal value of BQOP. Among the standard-, the full-, the twin- and the composite-DNN relaxations of BQOP which we derive, the composite DNN-relaxation is the most effective one that generates the tightest lower bound for the optimal value of BQOP. To compute its lower bound, we can apply the accelerated bisection and projection algorithm [1] to the Lagrangian relaxation of the composite-DNN relaxation, which we call the Lagrangian-DNN relaxation of BQOP [3] . We also show some numerical results which compare the standard SDP relaxation, the standard-DNN relaxation and the composite-DNN relaxation applied to randomly generated fully dense BQOPs. This talk is based on the recent papers [2] .

References

[1] N. Arima, S. Kim, M. Kojima, and K.C. Toh. A robust Lagrangian-DNN method for a class of quadratic optimization problems. Research Report B-482, Tokyo Institute of Technology, Department of Mathematical and Computing Sciences, Oh-Okayama, Meguro-ku, Tokyo 152-8552, February 2016.

[2] S. Kim, M. Kojima, and K. C. Toh. Doubly nonnegative relaxations for quadratic and polynomial optimization problems with binary and box constraints. Research Rport B-483, Tokyo Institute of Technology, Department of Mathematical and Computing Sciences, Oh-Okayama, Meguro-ku, Tokyo 152-8552, July 2016.

[3] S. Kim, M. Kojima, and K. C. Toh. A Lagrangian-DNN relaxation: a fast method for computing tight lower bounds for a class of quadratic optimization problems. *Math. Program.*, 156:161–187, 2016.

On semismooth Newton based augmented Lagrangian method for lasso-type problems

Kim Chuan Toh

National University of Singapore

abstract

We develop fast and robust algorithms for solving large-scale convex composite optimization models with an emphasis on least square problems having lasso-type regularizers. Although there exist many solvers in the literature for lasso-type problems, in particular the standard lasso problems, so far no solver can efficiently handle difficult real large scale lasso-regularized regression problems. By relying on the piecewise linear-quadratic structure of the problems to realize the remarkable fast linear convergence property of the augmented Lagrangian algorithm, and by exploiting the superlinear convergence of the semismooth Newton-CG method, we are able to design a new algorithm, called SSNAL, to efficiently solve the aforementioned difficult problems. Global convergence and local linear convergence results for SSNAL are established. Numerical results, including the comparison of our approach and several state-of-the-art solvers, on real data sets are presented to demonstrate the high efficiency and robustness of our proposed algorithm in solving large-scale difficult problems. (This is based on joint work with Xudong Li and Defeng Sun.)

The simplex method for degenerate and nondegenerate linear programming problems

Shinji Mizuno

Tokyo Institute of Technology

abstract

It is well-known that the simplex method with Dantzig's rule terminates in a finite number of iterations if a linear programming problem is non-degenerate, and it may require infinite number of iterations otherwise. Whether the problem is degenerate of not, the number of distinct solutions generated by the simplex method is always finite.

In this talk, we discuss the number of iterations and the number of distinct solutions of the simplex method for both degenerate and non-degenerate cases.

Exact duals and short certificates of infeasibility and weak infeasibility in conic linear programming

Gabor Pataki

University of North Carolina at Chapel Hill

abstract

In conic linear programming – in contrast to linear programming – the Lagrange dual may not be a strong dual, and the corresponding Farkas? lemma may fail to prove infeasibility. Here we describe exact duals, and certificates of infeasibility and weak infeasibility for conic LPs which retain most of the simplicity of the Lagrange dual, but do not rely on any constraint qualification. Some of our exact duals generalize the SDP duals of Ramana, Klep and Schweighofer to the context of general conic LPs. Some of our infeasibility certificates generalize the row echelon form of a linear system of equations: they consist of a small, trivially infeasible subsystem obtained by elementary row operations. We prove analogous results for weakly infeasible systems. We obtain some fundamental geometric corollaries: an exact characterization of when the linear image of a closed convex cone is closed, and an exact characterization of nice cones. Our infeasibility certificates provide algorithms to generate all infeasible conic LPs over several important classes of cones; and all weakly infeasible SDPs in a natural class. Using these algorithms we generate a public domain library of infeasible and weakly infeasible SDPs. The status of our instances is easy to verify by inspection in exact arithmetic, but they turn out to be challenging for commercial and research solvers.

Disjunctive conic cuts for hyperboloids and non-convex quadratic cones

Julio Góez

Norwegian School of Economics

abstract

In recent years, the generalization of Balas disjunctive cuts for mixed integer linear optimization problems to mixed integer non-linear optimization problems has received significant attention. Among these studies, mixed integer second order cone optimization (MISOCO) is a special case. For MISOCO we have the disjunctive conic cuts approach. That generalization introduced the concept of disjunctive conic cuts (DCCs) and disjunctive cylindrical cuts (DCyCs). Specifically, It showed that under some mild assumptions the intersection of those DCCs and DCyCs with a closed convex set, given as the intersection of a second order cone and an affine set, is the convex hull of the intersection of the same set with a parallel linear disjunction. The key element in that analysis was the use of pencils of quadrics to find close forms for deriving the DCCs and DCyCs. In this work we use the same approach to show that the DCCs and DCyCs are also valid disjunctive conic inequalities for hyperboloids and non-convex quadratic cones when the disjunction is defined by parallel hyperplanes. Also, we show that for each of the branches of those sets, which are convex, the intersections with the DCCs or DCyCs still provides the convex hull of the intersection of the branches with a parallel linear disjunction.

Analysis on conic programing via facial reduction

Masakazu Muramatsu

University of Electro-communications

abstract

Recently we develop some new techniques using facial reduction algorithms in the analysis of conic programming problems. Through the techniques, we give fresh views on extended duals, weak infesibility, distance to polihedrality, and other properties inherent in conic programming. In this talk, we survey these results.

A second-order cone based approach for solving the trust region subproblem and its variants

Fatma Kilinc Karzan

Carnegie Mellon University

abstract

We study the trust region subproblem (TRS) of minimizing a nonconvex quadratic function over the unit ball with additional conic constraints. Despite having a nonconvex objective, it is known that the TRS and a number of its variants are polynomial-time solvable. In this talk, we follow a second-order cone based approach to derive an exact convex formulation of the TRS. As a result, our study highlights an explicit connection between the nonconvex TRS and smooth convex quadratic minimization, which allows for the application of cheap iterative methods such as Nesterov's accelerated gradient descent to the TRS. Under slightly stronger conditions, we give a low-complexity characterization of the convex hull of the epigraph of the nonconvex quadratic function intersected with the constraints defining the domain without any additional variables. We also explore the inclusion of additional hollow constraints to the domain of the TRS, and convexification of the associated epigraph.

Centerpoints: A link between optimization and convex geometry

Amitabh Basu

Johns Hopkins University

abstract

We introduce a concept that generalizes several different notions of a "centerpoint" in the literature. We develop an oracle-based algorithm for convex mixed-integer optimization based on centerpoints. Further, we show that algorithms based on centerpoints are "best possible" in a certain sense. Motivated by this, we establish several structural results about this concept and provide efficient algorithms for computing these points.

Feature subset selection for logistic regression via mixed integer optimization

Yuichi Takano

Senshu University

abstract

This talk concerns a method of selecting a subset of features for a logistic regression model. Information criteria, such as the Akaike information criterion and Bayesian information criterion, are employed as a goodness-of-fit measure. The purpose of our work is to establish a computational framework for selecting a subset of features with an optimality guarantee. For this purpose, we devise mixed integer optimization formulations for feature subset selection in logistic regression. Specifically, we pose the problem as a mixed integer linear optimization problem, which can be solved with standard mixed integer optimization software, by making a piecewise linear approximation of the logistic loss function. The computational results demonstrate that when the number of candidate features was less than 40, our method successfully provided a feature subset that was sufficiently close to an optimal one in a reasonable amount of time. Furthermore, even if there were more candidate features, our method often found a better subset of features than the stepwise methods did in terms of information criteria. This is a joint work with Toshiki Sato from University of Tsukuba, Ryuhei Miyashiro from Tokyo University of Agriculture and Technology, and Akiko Yoshise from University of Tsukuba.

Geometric Median in Nearly Linear Time

Aaron Sidford

Stanford University

abstract

The geometric median problem is one of the oldest non-trivial problems in computational geometry: given n points in $\mathbb{R}^d$ compute a point that minimizes the sum of Euclidean distances to the points. However, despite an abundance of research the previous fastest algorithms for computing a $(1+\epsilon)$-approximate geometric median with only a polylogarithmic dependence on epsilon have super-linear dependences on n and d in the running time.

In this talk I will show how to compute a $(1+\epsilon)$-approximate geometric median in time $O(nd\log^3(1/esp))$. To achieve this running time I will start with a a simple $O((nd)^{O(1)}\log(1/\epsilon))$ time interior point method and show how to improve it, ultimately building a long-step algorithm that is non-standard from the perspective of the interior point literature. This result is one of very few cases I am aware provably outperforming traditional interior point theory and the only I am aware of using interior point methods to obtain a nearly linear time algorithm for a canonical optimization problem that traditionally requires superlinear time.

This talk reflects joint work with Michael B. Cohen, Yin Tat Lee, Gary Miller, and Jakub Pachocki.

Random-edge is slower than random-facet on abstract cubes

Thomas Dueholm Hansen

Aarhus University

abstract

Random-Edge and Random-Facet are two very natural randomized pivoting rules for the simplex algorithm. The behavior of Random-Facet is fairly well understood. It performs an expected sub-exponential number of pivoting steps on any linear program, or more generally, on any Acyclic Unique Sink Orientation (AUSO) of an arbitrary polytope, making it the fastest known pivoting rule for the simplex algorithm. The behavior of Random-Edge is much less understood. We show that in the AUSO setting, Random-Edge is slower than Random-Facet. To do that, we construct AUSOs of the n-dimensional hypercube on which Random-Edge performs an expected number of $2^{\Omega(\sqrt{n \log n})}$ steps. This improves on a $2^{\Omega(\sqrt[3]{n})}$ lower bound of Matou***v***sek and Szabó. As Random-Facet performs an expected number of $2^{O(\sqrt{n})}$ steps on any n-dimensional AUSO, this establishes our result. Improving our $2^{\Omega(\sqrt{n \log n})}$ lower bound seems to require radically new techniques.

Margin and overlap for convex bodies

David Bremner

University of New Brunswick

abstract

A fundamental question about two convex bodies is whether they can be separated by a hyperplane. In this talk I will discuss various ways in which degree of seperability or non-seperability can be measured, and the computational complexity of the resulting problems in algorithmic geometry.

For the "classical" notion of separability in terms of a widest slab (margin) between the two bodies, there is known dual problem of finding the shortest translation so that the two bodies touch. It turns out there there is a very similar dual for the non-seperable case, namely translating until the two bodies are weakly seperable.

We can also measure the degree of (non-)seperability (overlap) by considering the degree to which the two bodies need to be grown (shrunk) before transitioning between separable and non-seperable. I'll discuss the usual scaling operation, the "reduced convex hull" of Bern and Eppstein, and the inner parallel body as candidate methods of shrinking. In addition to various computational complexity issues, I'll try to connect the translation and shrinking points of view.

Implementation of interior-point methods for LP using Krylov subspace methods preconditioned by inner iterations

Keiichi Morikuni

University of Tsukuba

abstract

We apply inner-iteration preconditioned Krylov subspace methods [3] to underdetermined systems of linear equations arising in interior-point methods for solving linear programming (LP) problems in the primal-dual formulation

$$\min_{\boldsymbol{x}} \boldsymbol{c}^{\mathsf{T}}\boldsymbol{x}, \quad \text{subject to} \quad A\boldsymbol{x} = \boldsymbol{b}, \quad \boldsymbol{x} \geq \boldsymbol{0}, \tag{1a}$$

$$\max_{\boldsymbol{y}} \boldsymbol{b}^{\mathsf{T}}\boldsymbol{y}, \quad \text{subject to} \quad A^{\mathsf{T}}\boldsymbol{y} + \boldsymbol{s} = \boldsymbol{c}, \quad \boldsymbol{s} \geq \boldsymbol{0}, \tag{1b}$$

where $A \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, $\boldsymbol{c} \in \mathbb{R}^n$, and $m \leq n$ [1]. Our implementation of infeasible primal-dual predictor-corrector interior-point methods [2] for (1) employs iterative solvers to determine a search direction. To determine the affine direction $(\Delta\boldsymbol{x}_{\mathrm{af}}, \Delta\boldsymbol{y}_{\mathrm{af}}, \Delta\boldsymbol{s}_{\mathrm{af}})$ in the predictor stage, and the corrector direction $(\Delta\boldsymbol{x}_{\mathrm{cc}}, \Delta\boldsymbol{y}_{\mathrm{cc}}, \Delta\boldsymbol{s}_{\mathrm{cc}})$ in the corrector stage, we have to solve Newton's equations. Let $\mathcal{A} = AS^{-1/2}X^{1/2}$, where $X = \mathrm{diag}(\boldsymbol{x})$, $S = \mathrm{diag}(\boldsymbol{s})$. Denote the residual for the primal problem (1a) by $\boldsymbol{r}_{\mathrm{p}} = \boldsymbol{b} - A\boldsymbol{x}$, and the residual for the dual problem (1b) $\boldsymbol{r}_{\mathrm{d}} = \boldsymbol{c} - A^{\mathsf{T}}\boldsymbol{y} - \boldsymbol{s}$. Then, the Newton's equations reduce to the minimum-norm solution problem

$$\min \|\Delta\boldsymbol{w}\|_2, \quad \text{subject to} \quad \mathcal{A}\Delta\boldsymbol{w} = \boldsymbol{f}, \tag{2}$$

or equivalently the normal equations of the second kind $\mathcal{A}\mathcal{A}^{\mathsf{T}}\Delta\boldsymbol{y} = \boldsymbol{f}$, $\Delta\boldsymbol{w} = \mathcal{A}^{\mathsf{T}}\Delta\boldsymbol{y}$, where $\boldsymbol{f} = \boldsymbol{b} + AS^{-1}X\boldsymbol{r}_{\mathrm{d}}$ and $\Delta\boldsymbol{w}_{=}\mathcal{A}^{\mathsf{T}}\Delta\boldsymbol{y}_{\mathrm{af}}$ for the predictor stage, and $\Delta\boldsymbol{w} = \mathcal{A}^{\mathsf{T}}\Delta\boldsymbol{y}_{\mathrm{cc}}$ and $\boldsymbol{f} = AS^{-1}\Delta X_{\mathrm{af}}\Delta S_{\mathrm{af}}\boldsymbol{e} - \sigma\mu AS^{-1}\boldsymbol{e}$ for the corrector stage. Here, $\mu, \sigma > 0$, and $\boldsymbol{e} = [1, 1, \ldots, 1]^{\mathsf{T}}$.

The solution of Newton's equations is the main computational task in each interior-point iteration. In the late phase of interior point iterations, the diagonal matrix $S^{-1}X$ has tiny and large values as a result of convergence. Thus, $\mathcal{A}$ becomes ill-conditioned and Krylov subspace methods may be slow to converge. To overcome this difficulty, we apply novel inner-iteration preconditioned Krylov subspace methods based on the conjugate gradient (CG)

method, the MINRES method, and the generalized minimal residual (GMRES) method.

Firstly, CG is an iterative method for solving linear systems of equations $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{A} \in \mathbf{R}^{N\times N}$ is a symmetric and positive (semi)definite matrix and $\mathbf{b} \in \mathcal{R}(\mathbf{A})$, where $\mathcal{R}(\mathbf{A})$ is the range of $\mathbf{A}$. CG starts with an initial guess $\mathbf{x}_0 \in \mathbb{R}^N$ and determines the kth iterate $\mathbf{x}_k \in \mathbb{R}^N$ by minimizing $\|\mathbf{x}_k - \mathbf{x}_*\|_{\mathbf{A}}^2 = (\mathbf{x}_k - \mathbf{x}_*)^\mathsf{T}\mathbf{A}(\mathbf{x}_k - \mathbf{x}_*)$ over the space $\mathbf{x}_0 + \mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$, where $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$, and $\mathbf{x}_*$ is a solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$. Secondly, MINRES is another iterative method for solving linear systems of equations $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{A} \in \mathbf{R}^{N\times N}$ is a symmetric matrix. MINRES with $\mathbf{x}_0$ determines the kth iterate $\mathbf{x}_k$ by minimizing $\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2$ over $\mathbf{x}_0 + \mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$. Thirdly, GMRES is an iterative method for solving $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{A} \in \mathbf{R}^{N\times N}$ is a square matrix. GMRES with $\mathbf{x}_0$ determines the kth iterate $\mathbf{x}_k$ by minimizing $\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2$ over $\boldsymbol{x}_0 + \mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$.

CG and MINRES applied to $\mathbf{A}\mathbf{x} = \mathbf{b}$, with $\mathbf{A} = \mathcal{A}\mathcal{A}^\mathsf{T}$, $\mathbf{x} = \Delta\boldsymbol{y}$, $\mathbf{b} = \boldsymbol{f}$, and $\Delta\boldsymbol{w} = \mathcal{A}^\mathsf{T}\Delta\boldsymbol{y}$ are regarded as the CGNE (conjugate gradient method for the normal error) method and the MRNE method, respectively. On the other hand, GMRES can deal with a square system transformed from the rectangular system $\mathcal{A}\Delta\boldsymbol{w} = \boldsymbol{f}$. Let $\mathcal{B} \in \mathbf{R}^{n\times m}$ be a preconditioning matrix for $\mathcal{A}$. Then, GMRES applied to $\mathcal{A}\mathcal{B}\boldsymbol{z} = \boldsymbol{f}$, $\Delta\boldsymbol{w} = \mathcal{B}\boldsymbol{z}$ is regarded as the right-preconditioned GMRES method (AB-GMRES) [3] .

Now we introduce inner-iteration preconditioning for CGNE, MRNE, and AB-GMRES. Let M be a nonsingular matrix such that $\mathcal{A}\mathcal{A}^\mathsf{T} = M - N$. Denote the inner-iteration matrix by $H = M^{-1}N$. Then, $C^{(\ell)} = \sum_{i=0}^{\ell-1} H^i M^{-1}$ is the inner-iteration preconditioning matrix for CGNE and MRNE, whereas $\mathcal{B}^{(\ell)} = \mathcal{A}^\mathsf{T}C^{(\ell)}$ is the preconditioning matrix for AB-GMRES. If $\lim_{i\to\infty} H^i$ exists, then AB-GMRES determines the solution to (2) for all $\boldsymbol{f} \in \mathcal{R}(\mathcal{A})$. The condition on H is satisfied for the successive overrelaxation (SOR) splitting $M = \omega^{-1}(D + \omega L)$, where $\mathcal{A}\mathcal{A}^\mathsf{T} = L + D + L^\mathsf{T}$ with D diagonal and L strictly lower triangular, i.e., if $\omega \in (0, 2)$, $\lim_{i\to\infty} H^i$ exists. The inner-iteration preconditioners can avoid explicitly computing and storing the preconditioning matrices, and enable us to overcome the severe ill-conditioning of the linear systems in the final phase of interior-point iterations. These Krylov subspace methods do not break down for LP problems with rank-deficient constraint matrices even when previous direct methods fail. Therefore, no preprocessing is necessary even if rows of the constraint matrix are not linearly independent. Fortran and C codes for Matlab-MEX of AB-GMRES preconditioned by NE-SOR inner iterations are given in `http://researchmap.jp/KeiichiMorikuni/Implementations/`.

The resulting interior-point solver is composed of three nested iterations. The outer-most layer is the predictor-corrector interior-point method; the middle layer is the Krylov subspace method; the inner-most layer is the stationary inner iterations. Among the three layers, only the outer-most one runs towards the required accuracy and the other two are terminated prematurely.

Numerical experiments on 125 instances from Netlib, QAPLIB and Mittelmann's LP benchmarks show that our implementation is more stable and robust than the standard direct methods SeDuMi and SDPT3. To the best of our knowledge, this is the first result that an interior-point method entirely based on iterative methods succeeds in solving a fairly large number of standard LP instances under the standard stopping criteria. Moreover, the proposed method outperforms the interior-point solver of the state-of-the-art commercial code MOSEK for LP problems with random dense rank-deficient ill-conditioned constraint matrices. The coefficient matrix can be dense such as in quadratic programs arising in training support vector machines, or linear programs arising in basis pursuit, and even when A is given as a sparse matrix, $AXS^{-1}A^{\intercal}$ can be dense or have a pattern of nonzero elements that renders the system difficult for direct methods.

References

[1] Y. Cui, K. Morikuni, T. Tsuchiya, and K. Hayami, *Implementation of Interior-Point Methods for LP Based on Krylov Subspace Iterative Solvers with Inner-Iteration Preconditioning*, NII Technical Report, National Institute of Informatics, NII-2016-003E, arXiv:1604.07491 [math.OC] (2016), pp. 1–27. (http://www.optimization-online.org/DB_HTML/2016/04/5421.html).

[2] S. Mehrotra, *On the implementation of a primal-dual interior point method*, SIAM J. Optim., 2 (1992), pp. 575–601.

[3] K. Morikuni and K. Hayami, *Convergence of inner-iteration GMRES methods for rank-deficient least squares problems*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 225–250.

A comparative study of steepest descent methods for strongly convex quadratic functions

Mituhiro Fukuda

Tokyo Institute of Technology

abstract

Several methods based only on the gradients of the objective function have been recently proposed for minimizing (strongly) convex (differentiable) functions. Among them, we focus on several variations of the steepest descent method which depend only on different step-sizes. A simple comparative numerical study of these variants for some special classes of strongly convex quadratic functions indicates a preference of some variants over the others. In particular, we considered the theoretical and original algorithm proposed by Gonzaga [G06+], Gonzaga and Schneider [CS16], and de Asmundis *et al.* [deAdiSHTZ14] based on the Yuan step-size [Y06]. A simple modification on the Gonzaga's algorithm give a promissing result for some examples.

Linear optimization: Algorithms and conjectures

Tamás Terlaky

Lehigh University

abstract

In this talks we briefly review some fundamental algorithmic concepts for Linear optimization. The algorithms include large families of algorithms: elimination and pivot algorithms, ellipsoid and interior point methods, variants of perceptron, the von Neumann and Chubanov's algorithms. Complexity and convergence of the algorithms will be discussed. Open problems and conjectures related to these algorithms, as well as related to generalizations of LO will be discussed. Finally, we consider how the various algorithms could utilize the readily available multi-core architectures