Mathematisches Forschungsinstitut Oberwolfach

Report No. 2/2005

Optimization and Applications

Organised by
Florian Jarre (Düsseldorf)
Claude Lemaréchal (Grenoble)
Jochem Zowe (Erlangen)

January 9th – January 15th, 2005

Abstract. Proceedings of a workshop devoted to optimization problems, their theory and resolution, and above all applications of them. The topics covered existence and stability of solutions; design, analysis, development and implementation of algorithms; applications in mechanics, telecommunications, medicine, operations research.

Mathematics Subject Classification (2000): 49-06, 65K, 90C.

Introduction by the Organisers

This workshop was attended by 47 participants with broad geographic representation from all continents. It was enlightened by this cheerful atmosphere special to Oberwolfach, and which the participants contributed to by their mutual friendship and esteem.

The theme was optimization problems, their theory and resolution, and above all applications of them. Some emphasis was put on the still recent subject of optimization over the cone of positive semidefinite matrices (SDP). In the following 20 extended abstracts, one finds

– several papers devoted to theory *per se*: exitence and stability of solutions, analysis of algorithms;
– a good number on methodology of resolution: design, development, implementation of algorithms;
– a majority dealing with applications; some such applications come from other branches of mathematics (combinatorics, dynamic systems), but most from the

real world: operations research, telecommunications, production and transportation, medicine, model management.

We thank Gail Pieper (Argonne National Laboratory) for her careful proofreading of every abstract; her editorial work substantially improved the presentation of this material.

Workshop: Optimization and Applications

Table of Contents

Abstracts

An Overview of Nonsmooth Dynamical Systems

Vincent Acary

1. A Very Short Introduction to Moreau's Sweeping Process

Nonsmooth dynamical systems (NSDS) are a very special kind of dynamical systems, characterized by a nonsmooth evolution with respect to time and by a set of nonsmooth generalized equations. The so-called Moreau sweeping process [10, 11] is a special kind of differential inclusion with a maximal monotone operator [5], which appears to be a nice formulation for the unilateral dynamics:

$$\begin{cases} \dot{x}(t) + f(x,t) = \lambda(t), \quad x(t) \in \mathbb{R}^n, \lambda(t) \in \mathbb{R}^m \\ -\lambda(t) \in N_{\Phi(t)}(x(t)), \end{cases} \tag{1}$$

where f is a smooth vector field and $N_{\Phi(t)}$ is the normal cone to an admissible set $\Phi(t)$ for the state x. We refer to [2] for the equivalence with other types of order-one NSDS and to [7] for a review of various extensions of Moreau's sweeping process and associated mathematical results with weaker assumptions on the regularity of the solution.

Many examples of NSDS come from the engineering sciences. In electrical engineering, networks with idealized components (diodes, saturation, relays, etc.) are easily formulated as in (1). If the dynamics is linear, *linear complementarity systems* are often considered:

$$\begin{cases} \dot{x} = Ax + B\lambda, \quad x \in \mathbb{R}^n, \lambda \in \mathbb{R}^m \\ w = Cx + D\lambda \\ 0 \leqslant w \perp \lambda \geqslant 0. \end{cases} \tag{2}$$

For passive systems, this formulation is equivalent to (1) when $D = 0$ (see [6]); but we will see that it is not well defined in all cases. In mechanical engineering, M. Schatzman [16] has given a correct meaning to motion with measure acceleration for Lagrangian systems with unilateral contact, and J.-J. Moreau has extended the sweeping process to the second-order system [12]

$$-du + f(t, q(t)) \in N_{T(q)}(u(t)). \tag{3}$$

Here $u = \dot{q}$ is the velocity assumed to be a right-continuous function of bounded variations, and $T(q)$ is the tangent cone to the admissible set Φ at the position $q(t)$. The acceleration is replaced by a differential measure du, which may be viewed as the derivative in the sense of distributions of the velocity u (for more details see [13]). This compact formulation is powerful, not only from the computational point of view, but also from the pure mathematical point of view [9, 17, 4]. In control engineering, the standard problem of controlling a dynamical system with state constraints yields a dual problem that is also an NSDS. Numerous applications also

exist in biology, in economics, and in any problems where a constraint is imposed on the state variable.

Numerical methods inherit from the approach chosen to investigate NSDS. Two major approaches are widespread: the hybrid approach and the nonsmooth approach.

The *hybrid approach* considers NSDS as a hybrid multimodal dynamical system. In each mode, separated by two events, sufficient regularity is assumed on the system to accommodate standard analysis and classical numerical methods. This approach results in a family of computational schemes called *event driven*. In this framework, one cannot establish a general convergence proof, and the accumulation of events in finite time cannot be circumvented.

The *nonsmooth approach* is based on Moreau's sweeping process and its variants. The key idea is to write a suitable approximation of measures on a finite interval, which yields efficient and robust numerical schemes, called *time stepping*. The first such algorithm was the "catching-up algorithm" [11]. In the framework of multibody dynamics, the derived algorithm is a "nonsmooth contact dynamics" method [14, 15, 8], which can treat several thousands of 3D frictional contact conditions. The time step is no longer driven by events but is simply fixed by an a priori error criterion. Accumulations of events, or large numbers thereof in finite time, are handled without difficulty. Furthermore, convergence analysis of this family of schemes gives a constructive existence proof for rather complicate systems [9, 17].

2. Higher Relative-Degree Moreau's Sweeping Process

In joint work with B. Brogliato and D. Goeleven [1, 3], higher relative-degree systems are studied[1]. If the relative degree is at least 3, a generalized solution is a distribution of order larger than 2. Therefore, a positivity constraint on λ is meaningless. To circumvent this problem, we propose a new formulation of such systems as a measure-differential inclusion of higher order. This derivation is rather technical; we refer the reader to [3] for more details.

With this formulation, we give a precise meaning to solutions as distributions generated by a finite set of differential measures. Global existence and uniqueness results are also proved for a certain class of regular functions (analytical in every right neighborhood) and under a monotonicity assumption. Moreover, an efficient time-stepping scheme is designed; its convergence proof is currently under study. Applications for such types of systems include electrical and mechanical systems with feedback control, as well as the design of an indirect framework for solving optimal control problems with state constraints, based on necessary conditions.

[1]In systems of the form (2), the relative degree r between the output w and the multiplier λ may be defined as the rank of the first nonzero element in the sequence of Markov parameters $(D, CB, CAB, CA^2B, \ldots)$. A clear analogy exists between this relative degree and the differential index in differential algebraic equations.

References

[1] V. Acary and B. Brogliato. Higher order Moreau's sweeping process. In P. Alart, O. Maisonneuve, and R. T. Rockafellar, eds., *Nonsmooth Mechanics and Analysis: Theoretical and Numerical Advances, Colloquium in Honor of the 80th Birthday of J. J. Moreau (2003)*. Kluwer, 2005.

[2] V. Acary, B. Brogliato, A. Daniilidis, and C. Lemaréchal. On the equivalence between complementarity systems, projected systems and unilateral differential inclusions. RR 5107, Inria, 2004. `www.inria.fr/rrrt/rr-5107.html`. In revision for *Systems and Control Letters*.

[3] V. Acary, B. Brogliato, and D. Goeleven. Higher order Moreau's sweeping process: Mathematical formulation and numerical simulation. In revision for *Mathematical Programming A*, 2005.

[4] P. Ballard. The dynamics of discrete mechanical systems with perfect unilateral constraints. *Archives of Rational Mechanics and Analysis*, 154:199–274, 2000.

[5] H. Brezis. *Opérateurs maximaux monotones et semi-groupe de contraction dans les espaces de Hilbert*. North Holland, 1973.

[6] B. Brogliato. Absolute stability and the Lagrange-Dirichlet theorem with monotone multivalued mappings. *Systems and Control Letters*, 51:343–353, 2004.

[7] C. Castaing, T. X. Duc Ha, and M. Valadier. Evolution equations governed by the sweeping process. *Set-Valued Analysis*, 1:109–139, 1993.

[8] M. Jean. The nonsmooth contact dynamics method. *Computer Methods in Applied Mechanics and Engineering*, 177:235–257, 1999. Special issue on Computational Modeling of Contact and Friction, edited by J. A. C. Martins and A. Klarbring.

[9] M. D. P. Monteiro Marques. *Differential inclusions in nonsmooth mechanical problems: Shocks and dry friction*, volume 9 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser, 1993.

[10] J. J. Moreau. Rafle par un convexe variable, première partie, exposé 15 (deuxième partie, exposé 3, 1972). *Séminaire d'analyse convexe*, 1971.

[11] J. J. Moreau. Evolution problem associated with a moving convex set in a Hilbert space. *Journal of Differential Equations*, 26:347–374, 1977.

[12] J. J. Moreau. Standard inelastic shocks and the dynamics of unilateral constraints. In G. Del Piero and F. Maceri, eds., *Unilateral Problems in Structural Analysis*, number 288 in CISM Courses and Lectures, pages 173–221. Springer Verlag, 1985.

[13] J. J. Moreau. Bounded variation in time. In J. J. Moreau, P. D. Panagiotopoulos, and G. Strang, eds., *Topics in Nonsmooth Mechanics*, pages 1–74, Bikhäuser, 1988.

[14] J. J. Moreau. Unilateral contact and dry friction in finite freedom dynamics. In J. J. Moreau and P. D. Panagiotopoulos, eds., *Nonsmooth Mechanics and Applications*, number 302 in CISM Courses and Lectures, pages 1–82. Springer Verlag, 1988.

[15] J. J. Moreau. Numerical aspects of the sweeping process. *Computer Methods in Applied Mechanics and Engineering*, 177:329–349, 1999. Special issue on Computational Modeling of Contact and Friction, edited by J. A. C. Martins and A. Klarbring.

[16] M. Schatzman. A class of nonlinear differential equations of second order in time. *Nonlinear Analysis*, 2(3):355–373, 1978.

[17] D. Stewart. Convergence of a time-stepping scheme for rigid-body dynamics and resolution of Painlevé's problem. *Archives of Rational Mechanics and Analysis*, 145:215–260, 1998.

Managing Models in Simulation-Based Design Optimization

Natalia Alexandrov
(joint work with Robert Michael Lewis)

Advances in numerical modeling and computational power enable increasingly accurate simulation of physical and engineering phenomena. However, the enormous cost of repeated high-fidelity simulations, such as the Navier-Stokes equations or those based on fine computational meshes, makes the use of high-fidelity models impractical in the context of single-discipline or multidisciplinary design optimization.

Engineering designers have traditionally alleviated this difficulty by the combined use of high-fidelity and low-fidelity models in heuristic procedures, with low-fidelity models ranging from data-fitting models to simplified physics models. See, for instance, [8] for a review of approximations in structural optimization and [15] for a review of models in aerodynamic optimization. Some approaches to optimization with variable-fidelity models attempt to create the "best" low-fidelity model and optimize it, while others update the model by using heuristic procedures. See, for example, [10] and the references therein. Convergence to high-fidelity solutions is not guaranteed, in general.

Approximation and model management optimization (AMMO) [3, 1, 6, 7, 2, 4, 5] combine the use of general variable-fidelity models with analytically substantiated algorithms to improve tractability of design with high-fidelity models while preserving provable convergence properties. We consider the following general design problem. Given a set of design variables x, the analysis computes a set of quantities $u(x)$ of engineering interest. The computational model used in computing $u(x)$ often involves the solution of a set of (coupled) differential equations, as in the case of computing the flow around an airplane. The design problem is

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x, u(x)) \\ \text{subject to} & h(x, u(x)) = 0 \\ & g(x, u(x)) \geq 0 \\ & x_l \leq x \leq x_u. \end{array}$$

In conventional optimization, the analysis supplies the optimizer with objective and constraint function and derivative information. The optimizer builds local approximations of the objective and constraints—usually first- or second-order Taylor series—to compute new designs x. The basic idea of AMMO is to replace the local Taylor series model in the optimization subproblems with general models that satisfy consistency conditions with respect to the high-fidelity model. This is motivated as follows.

In general, the favorable situation in optimization occurs when the trends in the low-fidelity model responses coincide with those of the high-fidelity model. Thus, while the absolute difference between the low- and high-fidelity responses may be large, it is the *trend* in the low-fidelity model that is of significance to optimization. In practice, the trends in the low-fidelity problem may behave differently from

those in the high-fidelity problem. To address this, AMMO borrows from existing engineering practice to transform the low-fidelity responses to produce a better approximation of the trends of the high-fidelity problem.

The first-order AMMO idea can be used in conjunction with any gradient-based optimization algorithm. Here is an example of an AMMO algorithm for bound constrained minimization.

Initialize x_c, Δ_c
Do until convergence:
Select model a_c with $a_c(x_c) = f_{hi}(x_c)$; $\nabla a_c(x_c) = \nabla f_{hi}(x_c)$
Solve approximately for $s_c = x - x_c$:

$$\begin{aligned} \underset{s}{\text{minimize}} \quad & a_c(x_c + s) \\ \text{subject to} \quad & x_l \leq x_c + s \leq x_u \\ & \| s \|_\infty \leq \Delta_c \end{aligned}$$

Compute $\rho_c \equiv \dfrac{f_{hi}(x_c) - f_{hi}(x_c + s_c)}{f_{hi}(x_c) - a_c(x_c + s_c)}$
Accept s_c if $f_{hi}(x_c) > f_{hi}(x_c + s_c)$; otherwise reject
Update Δ_c and x_c based on the value of ρ_c
End do

The ratio ρ_c of the actual reduction in the high-fidelity objective to the predicted reduction obtained by computing with the corrected low-fidelity model measures the performance of the corrected low-fidelity model. The trust radius Δ_c is updated, based on the value of ρ_c, according to the standard trust-region practice. The step of choosing the model a_c and the nature of the trust-region subproblem distinguish this algorithm from conventional optimization.

The first-order consistency conditions in this case require the model a_c used in the optimization subproblem to satisfy $a_c(x_c) = f_{hi}(x_c)$ and $\nabla a_c(x_c) = \nabla f_{hi}(x_c)$. These ensure that a_c mimics the local behavior of a Taylor series model around the current best design x_c. This, in turn, can be used to prove that the overall optimization process will converge to a constrained stationary point of the high-fidelity objective f_{hi}.

Although exact consistency is unnecessary, a number of easily computed corrections ensure consistency, for instance, a technique we call the β-correction [13, 11]. Given the high-fidelity objective $f_{hi} = f$ and any low-fidelity approximation f_{lo} of the objective f_{hi}, we correct f_{lo} as follows. Define

$$\beta(x) = \frac{f_{hi}(x)}{f_{lo}(x)}\,, \quad \beta_c(x) = \beta(x_c) + \nabla\beta(x_c)^T(x - x_c)\,.$$

Then $a_c(x) = \beta_c(x) f_{lo}(x)$ satisfies the first-order consistency conditions. Convergence analysis of the resulting AMMO schemes relies on the consistency conditions and standard assumptions for the convergence analysis of the underlying optimization algorithm.

Practical efficiency of any particular AMMO scheme depends on the ability to transfer the computational load to the lower-fidelity computations and the relative

cost of low-fidelity model with respect to high-fidelity model. AMMO has been implemented by several research groups, mostly in the context of optimization of systems governed by computational fluid dynamics. Threefold to sevenfold savings in terms of high-fidelity evaluations have been observed in accumulated numerical experience.

Model management has also been explored in the context of derivative-free optimization (e.g., [9]), as well as in the special algorithmic context of multigrid-based approaches (e.g., [14] and [12]).

References

[1] N. M. Alexandrov. On managing the use of surrogates in general nonlinear optimization, AIAA Paper 98-4798, September 1998.

[2] N. M. Alexandrov, J. E. Dennis, Jr., R. M. Lewis, and V. Torczon. A trust region framework for managing the use of approximation models in optimization. *Structural Optimization,* 15:16–23, 1998.

[3] N. M. Alexandrov and R. M. Lewis. A trust region framework for managing approximation models in engineering optimization. AIAA Papers 96-4101 and 96-4102, September 1996.

[4] N. M. Alexandrov and R. M. Lewis. An overview of first-order model management for engineering optimization. *Optimization and Engineering,* 2:413–430, 2001.

[5] N. M. Alexandrov and R. M. Lewis. First-order approximation and model management in optimization. In *Large-Scale PDE-Constrained Optimization,* edited by L. T. Biegler, O. Ghattas, M. Heinkenschloss, and B. van Bloemen Waanders, Springer, 2003, pp. 63–79.

[6] N. M. Alexandrov, R. M. Lewis, C. R. Gumbert, L. L. Green, and P. A. Newman. Optimization with variable-fidelity models applied to wing design. AIAA Paper 2000-0841, January 2000.

[7] N. M. Alexandrov, E. J. Nielsen, R. M. Lewis, and W. K. Anderson. First-order model management with variable-fidelity physics applied to multi-element airfoil optimization. AIAA Paper 2000-4886, September 2000.

[8] J.-F. M. Barthelemy and R. T. Haftka. Approximation concepts for optimum structural design—a review. *Structural Optimization,* 5:129–144, 1993.

[9] A. J. Booker, J. E. Dennis, Jr., P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset. A rigorous framework for optimization of expensive functions by surrogates. *Structural Optimization,* 17:1–13, 1999.

[10] S. Burgee, A. Giunta, V. Balabanov, B. Grossman, W. Mason, R. Narducci, R. Haftka, and L. Watson. A coarse-grained parallel variable-complexity multidisciplinary optimization problem. *The International Journal of Supercomputing Applications and High Performance Computing,* 10:269–299, 1996.

[11] K. J. Chang, R. T. Haftka, G. L. Giles, and P.-J. Kao. Sensitivity-based scaling for approximating structural response. *Journal of Aircraft,* 30:283–288, 1993.

[12] S. Gratton, A. Sartenaer, and P. Toint. Recursive trust-region methods for multiscale nonlinear optimization (Part I): Global convergence and complexity. Report 04/06, CERFACS and Dept. of Mathematics, University of Namur, 2005.

[13] R. T. Haftka. Combining global and local approximations. *AIAA Journal,* 29:1523–1525, 1991.

[14] R. M. Lewis and S. G. Nash. A multigrid approach to the optimization of systems governed by differential equations. AIAA Paper 2000-4890, 2000.

[15] J. C. Newman, III, A. C. Taylor, III, R. W. Barnwell, P. A. Newman, and G. J.-W. Hou. Overview of sensitivity analysis and shape optimization for complex aerodynamic configurations. *Journal of Aircraft,* 36:87–96, 1999.

Musin's Proof of the Kissing Number in Dimension Four

Kurt M. Anstreicher

For $n \geq 3$ let $S^{n-1} = \{x \in \mathbb{R}^n : x^T x = 1\}$, and $-1 < z < 1$. A finite set $C = \{x_i\}_{i=1}^M \subset S^{n-1}$ is called a *spherical z-code* if $x_i^T x_j \leq z$ for all $i \neq j$. For $z = \frac{1}{2}$, $\{x_i\}$ correspond to contact points between S^{n-1} and M non-overlapping spheres of radius one that are all incident to S^{n-1}. Maximizing the number M of such spheres is called the *kissing problem* in dimension n, and the maximal M is called the *kissing number*.

The kissing problem in dimension 3 has a long history, going back to a discussion between Isaac Newton and David Gregory in 1694. Newton apparently believed that for $n = 3$ the kissing number was 12, while Gregory thought that 13 might be possible. In fact 13 is *not* possible but this is surprisingly difficult to prove. The first complete proof is credited to Schütte and van der Waerden [8] in 1953, and a subsequent proof by Leech [5] (see also [9]) in 1956 is now standard. The idea of Leech's proof is relatively simple but there are many details, some non-trivial, that require verification. The proof also appears to be impossible to extend to higher dimensions, for example to $n = 4$.

For general n and z there are several approaches that provide upper bounds on the size M of a z-code $C \subset S^{n-1}$; see for example [2, 9]. For small n the best results are typically obtained using the *Delsarte bounds*, which are based on a combination of harmonic analysis and linear programming [2, 3, 4, 6, 9]. For $z = \frac{1}{2}$ this approach leads to a complete characterization of maximal codes in dimensions $n = 8$ and 24 [1, 2], but for $n = 3$ the result is a bound of 13. For $n = 4$ a $\frac{1}{2}$-code with $M = 24$ is known, but the Delsarte bound is 25. The determination of the kissing number in dimension 4 has been an outstanding open problem for many years.

A 2003 paper of Musin [7] appears to have settled the kissing problem in dimension 4 via an interesting extension of the Delsarte bounds. To describe Musin's result we need to introduce the original linear programming bounds. The *distance distribution* of a code $C = \{x_i\}_{i=1}^M$ is the function $\lambda(\cdot) : [-1, 1] \to \mathbb{R}_+$ defined as

$$\lambda(s) = \frac{|\{(i,j) : x_i^T x_j = s\}|}{M}. \tag{1}$$

It follows that if C is a spherical z-code the distance distribution satisfies $\lambda(s) \geq 0$, $-1 \leq s \leq z$, $\lambda(1) = 1$, and $\sum_{-1 \leq s \leq z} \lambda(s) = M - 1$. Let $\Phi_k(\cdot)$, $k = 0, 1, \ldots$ denote the Gegenbauer, or ultraspherical, polynomials $\Phi_k(t) = P_k^{(\beta,\beta)}(t)/\binom{k+\beta}{k}$, where $P_k^{(\beta,\beta)}(\cdot)$ is the Jacobi polynomial with $\beta = (n-3)/2$. The normalization of $\Phi_k(\cdot)$ is chosen so that $\Phi_k(1) = 1$ for all k. Using techniques from harmonic analysis it can be shown ([4], [2, Chapters 9, 13], [9, Chapter 8]) that

$$1 + \sum_{-1 \leq s \leq z} \lambda(s)\Phi_k(s) \geq 0, \quad k = 1, 2, \ldots. \tag{2}$$

From (1) and (2), using $k = 1, \ldots, K$, the Delsarte bound on M is obtained via the semi-infinite linear programming problem

$$\begin{aligned} \text{LP}: \qquad \max \quad & \sum_{-1\le s\le z} \lambda(s) \\ \text{s.t.} \quad & \sum_{-1\le s\le z} \lambda(s)\Phi_k(s) \ge -1, \quad k = 1, \ldots, K, \\ & \lambda(s) \ge 0, \quad -1 \le s \le z. \end{aligned}$$

The dual of LP is the problem

$$\begin{aligned} \text{LD}: \qquad \min \quad & \sum_{k=1}^{K} f_k \\ \text{s.t.} \quad & \sum_{k=1}^{K} f_k\Phi_k(s) \le -1, \qquad -1 \le s \le z, \\ & f_k \ge 0, \quad k = 1, \ldots, K. \end{aligned}$$

Let $f(s) = 1 + \sum_{k=1}^{K} f_k\Phi_k(s)$, where $f_k \ge 0$, $k = 1, \ldots, K$. For the Delsarte bound, obtained from LD rather than LP, f_k are chosen to minimize $f(1)$ subject to $f(s) \le 0$, $s \in [-1, z]$. Musin's idea is to generalize $f(\cdot)$ in such a way that more detail regarding the possible structure of $\lambda(\cdot)$ can be exploited. Note that if $\lambda(\cdot)$ is the distance distribution of a spherical z-code, then

$$\begin{aligned} \sum_{-1\le s\le z} \lambda(s) \;&=\; \sum_{k=1}^{K} f_k + \sum_{-1\le s\le z} \lambda(s)\left[1 + \sum_{k=1}^{K} f_k\Phi_k(s)\right] \\ &\quad - \sum_{k=1}^{K} f_k\left[1 + \sum_{-1\le s\le z} \lambda(s)\Phi_k(s)\right] \;\le\; \sum_{k=1}^{K} f_k + \sum_{-1\le s\le z} \lambda(s)f(s), \end{aligned}$$

and therefore

$$M \;=\; 1 + \sum_{-1\le s\le z} \lambda(s) \;\le\; f(1) + \sum_{-1\le s\le z} \lambda(s)f(s).$$

Musin allows $f(\cdot)$ to be *nonnegative and decreasing* on $[-1, t_0)$ and nonpositive on $[t_0, z]$, for some $t_0 < -z$. We refer to such an $f(\cdot)$ as a *Musin polynomial*. Let $\lambda_i(s) = |\{j : x_i^T x_j = s\}|$, so that $\lambda(s) = [\sum_{i=1}^{M} \lambda_i(s)]/M$. Then if $f(\cdot)$ is a Musin polynomial,

$$\begin{aligned} M \;&\le\; f(1) + \sum_{-1\le s\le t_0} \lambda(s)f(s) \;=\; f(1) + \sum_{i=1}^{M} \sum_{-1\le s\le t_0} [\lambda_i(s)/M]f(s) \\ &\le\; f(1) + \max_{i=1,\ldots,M} \sum_{-1\le s\le t_0} \lambda_i(s)f(s). \end{aligned}$$

For given (n, z, t_0), define Y_m to be the set of spherical z-codes $\{y_i\}_{i=0}^{m}$ in S^{n-1} that satisfy the additional condition $y_0^T y_i \le t_0$, $i = 1, \ldots m$. Let μ be the maximum

m so that Y_m is nonempty, and define

$$h_m = \max_{y \in Y_m} \sum_{i=1}^{m} f(y_0^T y_i), \quad m = 1, \ldots, \mu. \tag{3}$$

Musin's final bound is then given by

$$M \leq f(1) + \max_{m=1,\ldots,\mu} h_m. \tag{4}$$

The computation of Musin's bound (4) has two main steps. The first requires *construction* of an appropriate Musin polynomial $f(\cdot)$. For fixed (n, z, t_0) this involves the solution of a problem similar to LD, with added terms corresponding to certain candidate configurations in Y_m, $m = 1, \ldots, \mu$. The second step is *verification* of the bound (4) via computation of h_m, $m = 1, \ldots, \mu$. While the construction of $f(\cdot)$ is relatively straightforward, verification of the bound is more complex since the optimization problem (3) defining each h_m is highly nonconvex. There are, however, several factors that facilitate the verification phase. First, if n is small and $|t_0|$ is not too small then μ cannot be too large. For example for $z = \frac{1}{2}$ and a value $t_0 \approx -0.6$, the value of μ in dimensions 3 and 4 is 4 and 6, respectively [7]. Second, Musin is able to partially characterize the structure of a maximizing set Y_m^*, simplifying the computation of h_m. For $m \leq n$ Musin shows that Y_m^* must be a regular spherical simplex containing $-y_0$. For $m > n$ the structure of Y_m^* is more complex, but for dimensions 3 and 4 the form is sufficiently constrained to permit the computation of h_m. For $z = \frac{1}{2}$ the result is the first proof that $M < 25$ in dimension 4, and a new proof that $M < 13$ in dimension 3. In some higher dimensions Musin constructs a polynomial $f(\cdot)$ that appears to give an improvement over the Delsarte bound, but verification of the bound in these dimensions is currently too difficult to complete. Simplification of the verification phase is an interesting topic of ongoing research.

References

[1] E. Bannai and N.J.A. Sloane, *Uniqueness of certain spherical codes*, Canadian J. Math. **33** (1981), 437-449.

[2] J.H. Conway and N.J.A. Sloane, *Sphere Packings, Lattices and Groups*, Third edition, Springer, New York, 1999.

[3] P. Delsarte, J.-M. Goethals and J.J. Seidel, *Spherical codes and designs*, Geom. Dedicat. **6** (1977), 363-388.

[4] G.A. Kabatiansky and V.I. Levenshtein, *Bounds for packings on a sphere and in space*, Problems of Information Transmission **14** (1978), 1-17.

[5] J. Leech, *The problem of the thirteen spheres*, Math. Gazette **40** (1956), 22-23.

[6] V.I. Levenshtein, *Universal bounds for codes and designs*, In *Handbook of Coding Theory*, Volume 1, V.S. Pless and W.C. Huffman, eds., North-Holland, Amsterdam, 1998, 499-648.

[7] O.R. Musin, *The kissing number in four dimensions*, preprint, Institute for Mathematical Study of Complex Systems, Moscow, 2003.

[8] K. Schütte and B.L. van der Waerdan, *Das Problem der dreizehn Kugeln*, Math. Annalen **53** (1953), 325-334.

[9] C. Zong, *Sphere Packings*, Springer, Berlin, 1999.

Some New Results on the Total Least Squares Problem

Amir Beck
(joint work with Aharon Ben-Tal and Marc Teboulle)

Many problems in data fitting and estimation give rise to an overdetermined system of linear equations $Ax \approx b$, where both the matrix $A \in \mathbb{R}^{m\times n}$ and the vector $b \in \mathbb{R}^m$ are contaminated by noise. The total least squares (TLS) approach to this problem [8, 10] is to seek a perturbation matrix E and a perturbation vector w that minimize $\|E\|^2 + \|w\|^2$ subject to the consistency equation; that is, E and w are the solutions of the optimization problem

$$\text{(TLS)} \quad \min_{w,E,x} \{\|E\|^2 + \|w\|^2 : (A+E)x = b + w\}.$$

One of the main reasons for the wide use of TLS is the fact that the problem has essentially an explicit solution, expressed by the singular value decomposition (SVD) of the augmented matrix (A, b) [8, 10].

We consider here two variants of the TLS problem:

- **The regularized TLS (RTLS) problem**, in which a quadratic constraint on the vector x is introduced:

  $$\text{(RTLS)} \quad \min_{w,E,x} \{\|E\|^2 + \|w\|^2 : (A+E)x = b + w, L^2 \le x^T Q x \le U^2\},$$

  where Q is a positive definite matrix.

- **The structured TLS (STLS) problem**, in which an additional linear constraint on the perturbation matrix is enforced:

  $$\text{(STLS)} \quad \min_{w,E,x} \{\|E\|^2 + \|w\|^2 : (A+E)x = b + w, \mathcal{L}(E) = 0\}.$$

  Here $\mathcal{L}$ is a linear operator. The linear structure discussed here is the *block circulant structure*.

A key difficulty with both problems is their nonconvexity. All current known methods to solve them (see, e.g., [13, 9, 12, 1]) converge to a point satisfying first-order necessary optimality conditions. In [4, 2] we show that a global optimal solution to both problems can be found efficiently.

The analysis of (RTLS) relies on a different formulation of the problem. Specifically, if we fix x and minimize with respect to E and w, (RTLS) becomes:

$$\text{(RTLS')} \quad \min_x \left\{ \frac{\|Ax-b\|^2}{\|x\|^2+1} : L^2 \le x^T Q x \le U^2 \right\}.$$

A simple observation, which goes back to Dinkelbach [7] and which will enable us to solve (RTLS'), is the following.

Observation: The following two statements are equivalent for a given α:

(1) $\min_{L^2 \le x^T Q x \le U^2} \frac{\|Ax-b\|^2}{\|x\|^2+1} \le \alpha$.

(2) $\min_{L^2 \le x^T Q x \le U^2} \{\|Ax-b\|^2 - \alpha(\|x\|^2+1)\} \le 0$.

The second problem, of minimizing an indefinite quadratic function subject to a double-sided quadratic constraint, can be reduced to the following convex problem by using an argument of Ben-Tal and Teboulle [6]:

$$\text{(CP)} \quad \min_{v_j \ge 0} \left\{ \sum_{j=1}^{n} \lambda_j v_j - |f_j| \sqrt{v_j} + c : L^2 \le \sum_{j=1}^{n} v_j \le U^2 \right\}.$$

By using convex duality, the solution of (CP) can be obtained just by solving at most two *single-variable* convex optimization problems.

The above discussion gives rise to an efficient algorithm for solving (RTLS'). An ϵ-global optimal solution to (RTLS') is calculated by solving a sequence of very simple convex minimization problems parametrized by a single parameter. The overall computational effort of the algorithm is $O(n^3 \log \epsilon^{-1})$.

In [2] we study (STLS), where the matrix A has either a *block circulant* (BC) structure or a (more special) *elementary block circulant* structure (EBC):

$$A = \underbrace{\begin{pmatrix} A_0 & A_1 & \cdots & A_{N-1} \\ A_{N-1} & A_0 & \cdots & A_{N-2} \\ \vdots & \vdots & & \vdots \\ A_1 & A_2 & \cdots & A_0 \end{pmatrix}}_{BC}, \quad A = \underbrace{\begin{pmatrix} A_0 & A_1 & \cdots & A_1 \\ A_1 & A_0 & \cdots & A_1 \\ \vdots & \vdots & & \vdots \\ A_1 & A_1 & \cdots & A_0 \end{pmatrix}}_{EBC},$$

where $A_0, A_1, \ldots, A_{N-1}$ are $m \times n$ matrices. The BC and EBC structures appear in the context of multichannel signal estimation [5, 3], image restoration [11] and more. We show that by applying the discrete Fourier transform, (STLS) decomposes into N *unstructured* (TLS) problems. The N solutions of these problems are then assembled to generate the optimal global solution of (STLS). Similar results are obtained for elementary block circulant matrices. Here the optimal solution is obtained by assembling two solutions: one of an *unstructured* TLS problem and the second of a multidimensional TLS problem.

References

[1] T. J. Abatzoglou, J. M. Mendel, and G. A. Harada. The constrained total least squares technique and its applications to harmonic superresolution. *IEEE Trans. Signal Processing*, 39(5):1070–1087, May 1991.

[2] A. Beck and A. Ben-Tal. A global solution for the structured total least squares problem with block circulant matrices. *SIAM J. Matrix Anal. Appl.* to appear.

[3] A. Beck, A. Ben-Tal, and Y. C. Eldar. Robust mean-squared error estimation of multiple signals in linear systems affected by model and noise uncertainties. Preprint.

[4] A. Beck, A. Ben-Tal, and M. Teboulle. Finding a global optimal solution for a quadratically constrained fractional quadratic problem with applications to the regularized total least squares. Preprint.

[5] A. Beck, Y. C. Eldar, and A. Ben-Tal. Minimax mean-squared error estimation of multichannel signals. Preprint.

[6] A. Ben-Tal and M. Teboulle. Hidden convexity in some nonconvex quadratically constrained quadratic programming. *Mathematical Programming*, 72(1):51–63, 1996.
[7] W. Dinkelbach. On nonlinear fractional programming. *Management Sciences*, 13:492–498, 1967.
[8] G. H. Golub and C. F. Van Loan. An analysis of the total least-squares problem. *SIAM J. Numer. Anal.*, 17(6):883–893, Dec. 1980.
[9] H. Guo and R. Renaut. A regularized total least squares algorithm. In *Total Least Squares and Errors-in-Variables Modeling*, pages 57–66. Kluwer, 2002.
[10] S. Van Huffel and J. Vandewalle. *The Total Least-Squares Problem: Computational Aspects and Analysis*, volume 9 of *Frontier in Applied Mathematics*. Philadelphia, PA: SIAM, 1991.
[11] A. K. Katsaggelos, K. T. Lay, and N. P. Galatsanos. A general framework for frequency domain multi-channel signal processing. *IEEE Trans. Image Processing*, 2(3):417–420, July 1993.
[12] J. B. Rosen, H. Park, and J. Glick. Total least norm formulation and solution for structured problems. *Siam J. Matrix Anal. Appl.*, 17(1):110–126, 1996.
[13] D. Sima, S. Van Huffel, and G. H. Golub. Regularized total least squares based on quadratic eigenvalue prblem solvers. Technical Report SCCM-04-10, Stanford Univ., 2003.

UMTS Radio Network Evaluation and Design

Andreas Eisenblätter
(joint work with Hans-Florian Geerdes, Thorsten Koch)

The Universal Mobile Telecommunications System (UMTS) is a third-generation cellular system for mobile telecommunications. UMTS supports all services of the worldwide-predominant GSM and GPRS networks and is more powerful, more flexible, and more radio-spectrum efficient than its predecessors. The present short exposition reports on activities within the EU-funded project Momentum [7] and within the DFG Research Center Matheon: Mathematics for Key Technologies [1].

UMTS is a *Wideband Code Division Multiple Access* system [5]. Radio transmissions are generally not separated in the time or the frequency domain. Complex coding schemes are used to distinguish different radio transmissions at a receiver. The capability to properly detect the desired carrier signal, however, requires that the ratio between the carrier signal and interfering signals (CIR) does not drop below some threshold value. Interference thus needs to be carefully controlled during network planning and operation because it is a limiting factor for network capacity [6].

The Problem. A central part in the initial deployment and the subsequent expansions of a UMTS radio network is to decide about the location and configuration of the base stations, including their antennas. Among others, the type of an antenna (and thus its radiation pattern), the mounting height, and its main radiation direction– *azimuth* in the horizontal plane and the *tilt* in the vertical plane –have to be decided. Basically, the problem is to design a network of sufficient coverage and capacity.

Each antenna provides coverage and capacity to the network, but the two are coupled through intercell interference and depend on the current occupation of

the network. In consequence, even assessing the merits of a given network design with respect to coverage and capacity is not easy. Moreover, there is no canonical optimization goal. Irrespective of such subtleties, mathematical UMTS radio network optimization models typically have aspects of facility location/set covering (trying to establish coverage) and (multiple) knapsack (trying to reflect coupled cell capacity).

A System Model. Concerning network evaluation, closed linear characterizations of the receive and transmit powers for UMTS cells have recently been developed. Here, they are presented in a simple version to illustrate the strong interrelation among cells.

The transmit powers for the up- and downlink of all mobiles can be derived from these characterizations. The powers are obtained under the assumptions that *perfect power control* applies to each dedicated radio link (signals are not stronger than necessary) [5], that no cell is in overload, and that no restrictions are imposed on the transmit powers of mobiles. Only the downlink case (base station to mobile communication) is addressed here. Similar expressions exist for the uplink case.

For cell i, let $\bar{p}_i^{\downarrow}$ denote the total transmit power in the cell, $\check{p}_i^{\downarrow}$ the cumulative power of all common channels (not subject to power control), and $p_i^{(\eta)}$ the power the cell would need to emit for its users if all receptions were interference free, that is, the power needed for overcoming receiver noise only. Moreover, let $C^{\downarrow}$ denote a square matrix containing the interference coupling within cells (on the diagonal) and between cells. A linear equation system allows the derivation of the total cell transmit powers given the other terms:

$$\bar{p}^{\downarrow} = \left(I - C^{\downarrow}\right)^{-1} \cdot \left(p^{(\eta)} + \check{p}^{\downarrow}\right). \tag{1}$$

The matrix $C^{\downarrow}$ and the vector $p_i^{(\eta)}$ are defined below, based on a measure for the impact of serving a user. For mobile m, let $\alpha_m^{\downarrow} \in (0,1]$ denote the service-specific transmit activity, $\mu_m^{\downarrow}$ the service-specific CIR requirement, and $\bar{\omega}_m \in [0,1]$ the fraction of own-cell signals received as interference due to loss of code orthogonality in the radio propagation environment. Moreover, let $\gamma_{im}^{\downarrow}$ denote the end-to-end attenuation between cell i and mobile m, and let M_i be the set of users served by cell i. Mobile m can be associated with a *downlink user load* $l_m^{\downarrow}$ defined as

$$l_m^{\downarrow} := (\alpha_m^{\downarrow}\, \mu_m^{\downarrow})/(1 + \bar{\omega}_m\, \alpha_m^{\downarrow}\, \mu_m^{\downarrow}). \tag{2}$$

This quantity is the key to defining the *traffic noise power* $p^{(\eta)}$ and the *downlink coupling matrix* $C^{\downarrow}$:

$$p_j^{(\eta)} := \sum_{m \in M_j} \frac{\eta_m}{\gamma_{jm}^{\downarrow}}\, l_m^{\downarrow}, \quad C_{ii}^{\downarrow} := \sum_{m \in M_i} \bar{\omega}_m\, l_m^{\downarrow}, \quad C_{ij}^{\downarrow} := \sum_{m \in M_i} \frac{\gamma_{jm}^{\downarrow}}{\gamma_{im}^{\downarrow}}\, l_m^{\downarrow}\ (i \neq j). \tag{3}$$

Definitions (2) and (3) can be extended to spatial (average) traffic intensity.

Optimization Models. Various mixed integer programming models have been proposed in recent years. Two distinct approaches are sketched. On the one hand are models that try to closely reflect the expected network performance. They

consider traffic snapshots and try to find a network optimized for these snapshots; see, for example, [4]. On the other hand, a recent development is based on (1) with average traffic intensities [2]. Here, the goal is basically to obtain a network with good "average" performance.

Challenges. The challenge of developing effective optimization techniques is preceded by two other challenges. One involves devising a reasonable system model ([7, 4, 2], still ongoing), and the other involves collecting sufficient real-world data [7, 3] to guide the modeling and to perform computational studies on other optimization models and methods.

The robustness of solutions, as well as the fact that input is statistical data, becomes increasingly important. What if the expectations to the traffic distribution are off by 10%? Advanced optimization and evaluation techniques should certify their superiority to simple planning heuristics in these respects.

In practice, however, the simple heuristic methods often outperform (or at least compete with) heavy-duty optimization techniques, simply because the latter fail on large instances. An amazing gap still exists between our state-of-the-art optimization techniques and what is needed to prevail for regular planning and optimization tasks in radio network planning.

References

[1] DFG Research Center Matheon: Mathematics for Key Technologies: Modelling, simulation, and optimization of read-world processes. Project B4. `http://www.matheon.de` 2002–2005.

[2] A. Eisenblätter and H.-F. Geerdes. A novel view on cell coverage and coupling for UMTS radio network evaluation and design. Technical report, Zuse Institute Berlin (ZIB), 2004. Submitted to INOC'05 Conference.

[3] A. Eisenblätter, H.-F. Geerdes, U. Türke, and T. Koch. Momentum data scenarios for radio network planning and simulation (extended abstract). In *Proc. of WiOpt'04*, Cambridge, UK, 2004.

[4] A. Eisenblätter, T. Koch, A. Martin, T. Achterberg, A. Fügenschuh, A. Koster, O. Wegel, and R. Wessäly. Modelling feasible network configurations for UMTS. In G. Anandalingam and S. Raghavan, eds, *Telecommunications Network Design and Management*. Kluwer, 2002.

[5] H. Holma and A. Toskala. *WCDMA for UMTS*. Wiley, 2001.

[6] J. Laiho, A. Wacker, and T. Novosad, eds. *Radio Network Planning and Optimization for UMTS*. Wiley, 2001.

[7] MOMENTUM project. Models and simulations for network planning and control of UMTS. `http://momentum.zib.de`, 2001–2003. IST-2000-28088.

Optimizing Call Center Staffing by Using Simulation and Analytic-Center Cutting-Plane Methods

Marina A. Epelman

(joint work with Júlíus Atlason, Shane Henderson)

We describe a method for selecting staff levels at an inbound call center, or a similar service facility, that minimizes cost while simultaneously ensuring satisfactory customer service. Simulation is used to report service-level performance for

a given set of staffing levels, and an analytic-center cutting-plane method guides the selection of staffing levels.

Define the vector of staffing levels as $y \in \mathbb{Z}_+^p$, where p is the number of time periods in the planning horizon. The call center staffing problem can then be formulated as a MINLP as follows:

$$\min\{f(y) : g_i(y) \geq 0,\ i = 1, \ldots, p,\ y \in \mathbb{Z}_+^p\}. \tag{1}$$

Here the constraints $g_i(y) \geq 0,\ i = 1, \ldots, p$, express the quality-of-service requirements in each period, which typically represent "π percent of calls are answered within τ seconds." The objective function $f(y)$ represents the minimal cost of covering the desired staffing level y by predetermined feasible staff schedules, or *shifts,* and can be evaluated by solving an integer linear covering problem.

The main challenge in solving (1) lies in the difficulty of identifying feasible staffing levels, or even verifying feasibility of a particular y, especially when there is significant linkage in performance between different periods. To this end, we use simulation with n independent replications to approximately evaluate performance of the call center at proposed staffing level y. Thus, after selecting n, we replace problem (1) with its *sample average approximation:*

$$\min\{f(y) : \bar{g}_i(y, n) \geq 0,\ i = 1, \ldots, p,\ y \in \mathbb{Z}_+^p\}. \tag{2}$$

Use of common random numbers ensures consistency of function evaluations between simulations. See [1] for convergence analysis of solutions of (2) as $n \to \infty$.

It has been empirically confirmed that service-level functions in each period exhibit behavior similar to that of pseudoconcave functions of continuous variables. We say that a function $h(\cdot)$ is *discrete pseudoconcave* if for any $\hat{y} \in \mathbb{Z}_p^+$, there exists a vector $q(\hat{y}) \in \mathbb{R}^p$ such that for any $y \in \mathbb{Z}_p^+$,

$$q(\hat{y})^T(y - \hat{y}) \leq 0 \Rightarrow h(y) \leq h(\hat{y}). \tag{3}$$

We refer to vectors $q(\cdot)$ as *pseudogradients.* Since the constraint functions of (1) and (2) appear to be discrete pseudoconcave, it is appealing to apply a cutting plane-type algorithm to solve the MINLP (2).

Although several such methods for solving MINLPs have been proposed, they cannot be directly applied to (2). Indeed, in these algorithms it is assumed that the constraint functions are, in fact, differentiable functions of continuous variables; the integrality restrictions on the variables are, in a sense, exogenous. In such a setting the concept of a convex (continuous) nonlinear relaxation of the integer program is straightforward, and feasibility cuts are generated simply by using the gradients of these continuous functions. In our setting, however, the service-level functions and their sample average approximations are not defined at noninteger values of y, and devising their continuous extension is nontrivial at best.

To address this difficulty, we introduce the first application of an Analytic Center Cutting Plane Method (ACCPM) to MINLPs with discrete pseudoconcave constraints. Our method differs from algorithms described above in two main ways (see [6] for review of ACCPMs):

– As in ACCPM, the analytic center of the polyhedral localization set is computed at every iteration. Next, however, the integer point in the localization set nearest to the analytic center is found by solving an IP; feasibility of this integer point $\hat{y}$ is then checked by simulation.
– To add a feasibility cut if $\hat{y}$ found above violates one or more of the constraints, we estimate a pseudogradient of the violated constraint at $\hat{y}$ and generate a cut using inequality (3). Among many approaches tested, forward finite differencing was shown to perform the best in approximating the pseudogradients (see [3]).

Under reasonable assumptions, this algorithm will converge finitely to the optimal solution of (2).

We have successfully implemented the algorithm above and tested its performance on the call center staffing problem. Our benchmarks were the heuristics based on the analytical queueing methods developed by Green et al. ([4, 5]), widely considered to be the state of the art in multiperiod staffing. Experiments on a call center modeled as an $M(t)/M/s(t)$ queue show that our algorithm outperforms, or at least equals, these heuristics in every case in which shift structure is explicitly considered, which is the setting we are primarily interested in (see [2]). Of course, these extremely appealing properties have to be traded off against the computational cost of the procedure, which is not inconsiderable. Nonetheless, it is a robust procedure that can be applied in a near-black-box fashion. Moreover, the complex structure of realistic call centers imposes a limit on the applicability of queueing-based heuristics. In such cases, simulation and our simulation-based algorithm are a viable alternative.

References

[1] J. Atlason, M. A. Epelman, and S. G. Henderson. Call center staffing with simulation and cutting plane methods. *Ann. Oper. Res.,* 127:333–358, 2004.

[2] J. Atlason, M. A. Epelman, and S. G. Henderson. Optimizing call center staffing using simulation and ACCPMs. IOE Tech. Report 04-09, University of Michigan, 2004.

[3] J. Atlason, M. A. Epelman, and S. G. Henderson. Using simulation to approximate subgradients of convex performance measures in service systems. In *Proceedings of the 2003 Winter Simulation Conference,* edited by S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, 2003.

[4] L. V. Green, P. J. Kolesar, and J. Soares. Improving the SIPP approach for staffing service systems that have cyclic demands. *Oper. Res.,* 49:549–564, 2001.

[5] L. V. Green, P. J. Kolesar, and J. Soares. An improved heuristic for staffing telephone call centers with limited operating hours. *Prod. and Oper. Management,* 12:46–61, 2003.

[6] J. E. Mitchell. Polynomial interior point cutting plane methods. *Optim. Methods Softw.,* 18:507–534, 2003.

Breast Cancer Epidemiology: Calibrating Simulations via Optimization

Michael C. Ferris
(joint work with Geng Deng, Dennis G. Fryback, Vipat Kuruchittham)

We investigate the use of optimization and data mining techniques for calibrating the input parameters to a discrete event simulation code. In the context of a breast-cancer epidemiology model we show how a hierarchical classifier can accurately predict those parameters that ensure the simulation replicates benchmark data within 95% confidence intervals. We formulate an optimization model that evaluates solutions based on an integer valued score function. The scores are determined from a simulation run (and are therefore subject to stochastic variations), and are expensive to calculate.

The Wisconsin Breast Cancer Epidemiology Simulation uses detailed individual-woman-level discrete event simulation of four processes (breast cancer natural history, detection, treatment and nonbreast cancer mortality among U.S. women) to replicate breast cancer incidence rates according to the Surveillance, Epidemiology, and End Results (SEER) Program data from 1975 to 2000. Incidence rates are calculated for four different stages of tumor growth, namely, in situ, localized, regional, and distant; these correspond to increasing size and/or progression of the disease. Each run involves the simulation of 3 million women and takes approximately 8 minutes to execute on a 1 GHz Pentium machine with 1 GB of RAM.

The four simulated processes overlap in complex ways, and thus it is very difficult to formulate analytical models of their interactions. However, each can be modeled by simulation; these models need to take into account the increase in efficiency of screening processes that has occurred since 1975, the changes in nonscreen detection as a result of increased awareness of the disease, and a variety of other changes during that time. The simulations are grounded in mathematical and statistical models that are formulated by using a parameterization. For example, the natural history process in the simulation can be modeled by using a Gompertzian growth model that is parameterized by a mean and variance typically unknown exactly but for which a range of reasonable values can be estimated. The overall simulation facilitates interaction between the various components, but it is extremely difficult to determine values for the parameters that ensure the simulation replicates known data patterns across the time period studied. In all, there are 37 parameters, most of which interact with each other and are constrained by linear relationships. Further details can be found in [1, 3].

A score is calculated that measures how well the simulation output replicates an estimate of the incidence curves in each of the four growth stages. Using SEER and Wisconsin Cancer Reporting System (WCRS) data, we generate an envelope that captures the variation in the data that might naturally be expected in a population of the size we simulated. For the 26 years considered, the four growth stages give a total of 104 points, each of which is tested to see whether it lies in the envelope.

The number of points outside the envelope is summed to give the score (0 is ideal). While one could argue that distance to the envelope might be a better measure, such calculations are scale dependent and were not investigated. Unfortunately, the score function also depends on the "history" of breast cancer incidence and mortality that is generated in the simulation based on a random seed value ω. We will adopt the notation $f_\omega(v)$, where v represents the vector of parameters and ω indexes the replication. While we are interested in the distribution (over ω) of $f_\omega(v)$, we will focus here on the problem

$$\min_v \max_\omega f_\omega(v).$$

The purpose of this study is to determine parameter values v that generate small values for the scoring function. Prior to the work described here, acceptance sampling had been used to fit the parameters. Essentially, the simulation was run tens of thousands of times with randomly chosen inputs to determine a set of good values. With over 450,000 simulations, only 363 were found that had a score no more than 10. That is, for a single replication ω, 363 vectors v had $f_\omega(v) \leq 10$.

Our first goal was to generate many more vectors v with scores no more than 10. To do this, we attempted to use the given scoring function data to generate a classifier that quickly predicts whether a given vector v is in

$$L(\lambda) = \{v | f_\omega(v) \leq \lambda\}, \text{for a fixed replication } \omega.$$

We typically use $\lambda = 5$ to indicate good fit and $\lambda = 10$ for acceptable parameter choices. Our approach is as follows:

– Split the data into a training (90%) and testing (10%) set.
– Given the training set, generate a (hierarchical) classifier that predicts membership of $L(\lambda)$. Validate this classifier on the testing set.
– Generate 100,000 potential values for v, uniformly at random.
– For those vectors v that the classifier predicts are in $L(\lambda)$, evaluate $f_\omega(v)$ via simulation.

Since the classifier is cheap to evaluate, this process facilitates a more efficient exploration of the parameter space. Clearly, instead of using a single replication ω, we could replace $f_\omega(v)$ by $\max_{\omega\in\Omega} f_\omega(v)$ where $\Omega = \{\omega_1, \ldots, \omega_m\}$ for some $m > 1$. In fact this approach was carried out. The difficulty is that we require replication data (for our experiments we choose $m = 10$) and we update the definition of $L(\lambda)$ appropriately. However, the process we follow is identical to that outlined here.

In our setting, v has dimension 37. Using expert advice, we allowed only 9 dimensions to change; the other 28 values were fixed to the feasible values that have highest frequency of occurrence over the "positive" samples. For example, if v_{37} can take possible values from $[\phi_1, \phi_2, \ldots, \phi_n]$, then we set the value of v_{37} to be $\arg\min_{i=1}^n \frac{P_i}{W_i}$, where P_i and W_i are the number of appearances of ϕ_i in the positive and whole sample set. This is similar to using a naive Bayesian classifier to determine which value has the highest likelihood to be "positive". Our experiments showed this choice of values outperformed even the values that experts deemed appropriate for these 28 values; a posteriori analysis confirmed their superiority.

We generated a hierarchical classifier. The key difficulty in generating a classifier is the fact that we have a vast majority of "negative" data points (i.e., vectors $v \notin L(\lambda)$). By successively projecting our training data into two-dimensional slices, we identified two pairs of planes (meanGamma/varGamma and onsetProp/lag) in which only "negative" data points in our training set were present outside a small band of values. The top level of the classifier labels points outside these bands as "negative". The remaining points (within the bands, the positive and negative points are intermingled) are classified by using the following procedure.

Given a particular training set A, a variety of support vector machine classifiers can be generated by solving an optimization problem for values u and γ, and using the kernel classifier [7]

$$K(v', A')u - \gamma \leq 0$$

to imply that a new point v is "negative", where K is a given kernel function. We used the following kernels: linear, polynomial degree 2, polynomial degree 3, and Gaussian. We also used the C45 decision tree classifier and k-nearest neighbor classifier with $k = 5$. All of these classifiers are publicly available [6, 9].

Furthermore, the one-sided sampling approach [4] was used to generate a number of different training sets; the sampling approach iteratively removes "negative" points in a rigorously defined manner, and we stop this process when there are approximately 500 "negative" points remaining (there are around 300 "positive" points in each training set). The resulting classifier is evaluated on the testing set by using the measures

$$TP = \frac{\#\text{ correctly classified positives}}{\text{total }\#\text{ of positives}} \text{ and } TN = \frac{\#\text{ correctly classified negatives}}{\text{total }\#\text{ of negatives}}.$$

(Note that cross-validation accuracy is inappropriate to use in this setting because it can be made large by classifying all points as "negative" on account of the imbalanced nature of the data.) Classifiers are discarded if the value of TP is less than 0.9 (typically TN is around 0.4). This value was chosen to guarantee the probability of removing positive points in error is small. We also generate training sets by resampling with replacement.

For a uniform sample of 100,000 potential values of v, the naive banding classifier removes all but 8 640. Each of the above classifiers was used successively to determine whether the point v was "negative" (and hence removed from consideration); if not, v was passed onto the next classifier. This process was repeated until the number of points being removed decreased to zero. At that stage there were 788 points that were hypothesized to be "positive". These 788 points were tested using simulation, and 65% were found to be "positive". This is a significant improvement over the random sampling scheme.

A further sequence of classifiers was determined from a training set generated by using the above sampling schemes, but where we adjusted the number of "negative" points in the training set so that the resulting values for TP and TN were approximately 0.6 and 0.7. These additional classifiers have a larger chance of removing "positive" samples in error, but they reduce the number of remaining points in our sample much more quickly. For our example set the remaining 788

points was reduced to 220 points. Evaluating these remaining points by simulation, 195 were found to be in $L(10)$. Thus, with very high success rate (89%), our classifier is able to predict values of v that have a low score $f_\omega(v)$.

We employed the classifier technique above to generate a large number of samples in $L(30)$. Given these samples, we used the DACE toolbox [5] to fit a kriging model to the data, which we consider a surrogate function [2] for our objective. We used the Nelder-Mead simplex method [8] to optimize this surrogate and generated several local minimizers for this function based on different trial starting points. These local minimizers were evaluated by simulation. To improve our results further, we updated the surrogate function with the simulation results of the local minimizers and repeated the optimization. The parameter values found by using this process outperform all previous values found. Furthermore, expert analysis of various output curves generated from the simulation results with the best set of parameter values confirms the quality of this solution.

While our procedure is somewhat ad hoc, the following conclusions are evident:

– The classifier technique is cheap to use and predicts good parameter values very accurately without performing additional simulations.
– A hierarchical classifier significantly improves classification accuracy.
– Imbalanced training data has a detrimental effect on classifier behavior. Ensuring the data is balanced in size is crucial before generating classifiers.

The classifier facilitates easy generation of parameter settings within a given level set of score values and potentially allows investigation of such level sets and good parameter settings from a biological perspective. Future work will investigate characterizing the level set more precisely with the aim of enhancing biological understanding of the model parameters.

References

[1] CISNET, `http://cisnet.cancer.gov/profiles/`.
[2] J. E. Dennis, A. Booker, P. Frank, D. Serafini, V. Torczon, and M. Trosset, A rigorous framework for optimization of expensive functions by surrogates. *Structural Optimization,* 17(1):1–13, 1999.
[3] D. G. Fryback, N. K. Stout, M. A. Rosenberg, A. Trentham-Dietz, V. Kuruchittham, and P. L. Remington. The Wisconsin breast cancer epidemiology simulation model. Preprint, January 2005.
[4] M. Kubat and S. Matwin. Addressing the curse of imbalanced data sets: One sided sampling. In *Proceedings of the Fourteenth International Conference on Machine Learning*, 1997, pp. 179–186.
[5] S. N. Lophaven, H. B. Nielsen, and J. Søndergaard. DACE – A Matlab kriging toolbox, Informatics and Mathematical Modelling. Technical University of Denmark, DTU, 2002.
[6] J. Ma, Y. Zhao, and A. Stanley. OSU SVM classifier Matlab Toolbox. `http://www.ece.osu.edu/~maj/osu_svm/`.
[7] O. L. Mangasarian, Generalized support vector machines. In *Advances in Large Margin Classifiers,* edited by A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, MIT Press, Cambridge, MA, 2000, pp. 135–146.
[8] J. A. Nelder and R. Mead, A simplex method for function minimization. *Computer Journal,* 7:308–313, 1965.

[9] J. Weston, A. Elisseeff, G. Bakir, and F. Sinz. The Spider, Matlab Machine Learning Toolbox. `http://www.kyb.tuebingen.mpg.de/bs/people/spider/`.

L^1-Optimal Boundary Control of a String to Rest in Finite Time

Martin Gugat

Hyperbolic partial differential equations often appear as models in engineering, for example as systems of conservation laws that model fluid flow.

The control of such systems is usually possible only with boundary controls, which in the mathematical model corresponds to control via the boundary conditions. To get some insight into the nature of optimal controls for such systems, we consider the following problem of optimal Dirichlet boundary control for the wave equation:

$$\text{(P)} \qquad \begin{cases} \min\limits_{u_1,u_2\in L^1(0,T)} \displaystyle\int_0^T |u_1(t)| + |u_2(t)|\, dt \quad \text{suject to} \\ y_{tt}(x,t) = c^2 y_{xx}(x,t), & (x,t)\in(0,L)\times(0,T) \\ y(0,t) = u_1(t), \quad y(L,t) = u_2(t), & t\in(0,T) \\ y(x,0) = y_0(x), \quad y_t(x,0) = y_1(x), & x\in(0,L) \\ y(x,T) = 0, \quad y_t(x,T) = 0, & x\in(0,L). \end{cases}$$

The functions y_0 and y_1 are given, as well as the real numbers $T > 0$, $L > 0$, $c > 0$.

In general, this problem does not have a unique solution. An explicit representation of all solutions is given in the following theorem.

Theorem 1. *Assume that $T \geq t_0 = L/c$, that $y_0 \in L^1(0,L)$, and that $Y_1(x) = \int_0^x y_1(s)\, ds \in L^1(0,L)$. For $t \in (0,t_0)$, let*

$$\begin{aligned} \alpha_0(t) &= y_0(ct) + (1/c)\int_0^{ct} y_1(s)\, ds, \\ \beta_0(t) &= y_0(L-ct) - (1/c)\int_0^{L-ct} y_1(s)\, ds. \end{aligned}$$

Choose a real number r that minimizes

$$I(r) = \frac{1}{2}\int_0^{t_0} |\alpha_0(t) - r| + |\beta_0(t) + r|\, dt. \tag{1}$$

Let $k = \max\{j\in\mathbb{N} : jt_0 \leq T\}$ and $\Delta = T - kt_0$. For $j\in\{0,\dots,k\}$ and $t\in(0,\Delta)$, let $\lambda_j(t) \geq 0$, $\nu_j(t) \geq 0$ almost everywhere be such that $\lambda_j(\alpha_0 - r) \in L^1(0,\Delta)$, $\nu_j(\beta_0 + r) \in L^1(0,\Delta)$, and

$$\sum_{j=0}^{k} \lambda_j(t) = 1 = \sum_{j=0}^{k} \nu_j(t) \quad \textit{almost everywhere on } (0,\Delta).$$

For $j \in \{0, \dots, k-1\}$ and $t \in (\Delta, t_0)$, let $\mu_j(t) \geq 0$, $\omega_j(t) \geq 0$ almost everywhere be such that $\mu_j(\alpha_0 - r) \in L^1(\Delta, t_0)$, $\omega_j(\beta_0 + r) \in L^1(\Delta, t_0)$ and

$$\sum_{j=0}^{k-1} \mu_j(t) = 1 = \sum_{j=0}^{k-1} \omega_j(t) \quad \textit{almost everywhere on } (\Delta, t_0).$$

Then the optimal solutions of (P) are the controls u_1, u_2 described by (2) below:

$$(2) \quad \begin{cases} u_1(t + jt_0) = \lambda_j(t)[\alpha_0(t) - r]/2 & \text{if } j \text{ is even and } t \in (0, \Delta), \\ u_1(t + jt_0) = \mu_j(t)[\alpha_0(t) - r]/2 & \text{if } j \text{ is even and } t \in (\Delta, t_0), \\ u_1(t + jt_0) = -\nu_j(t)[\beta_0(t) + r]/2 & \text{if } j \text{ is odd and } t \in (0, \Delta), \\ u_1(t + jt_0) = -\omega_j(t)[\beta_0(t) + r]/2 & \text{if } j \text{ is odd and } t \in (\Delta, t_0), \\ u_2(t + jt_0) = \nu_j(t)[\beta_0(t) + r]/2 & \text{if } j \text{ is even and } t \in (0, \Delta), \\ u_2(t + jt_0) = \omega_j(t)[\beta_0(t) + r]/2 & \text{if } j \text{ is even and } t \in (\Delta, t_0), \\ u_2(t + jt_0) = \lambda_j(t)[-\alpha_0(t) + r]/2 & \text{if } j \text{ is odd and } t \in (0, \Delta), \\ u_2(t + jt_0) = \mu_j(t)[-\alpha_0(t) + r]/2 & \text{if } j \text{ is odd and } t \in (\Delta, t_0). \end{cases}$$

The minimal value of (P) is given by the integral $I(r)$ of (1) with an optimal choice of r. The solution of (P) is unique if and only if its minimal value is zero.

A proof of this result is given in [2]. This proof is based on the traveling waves solution of the wave equation. It shows how the structure of the optimal controls is related to the characteristic curves.

The solutions of the corresponding problems for L^p-norms with $p \in [2, \infty)$ are given in [3], where a proof based on Fourier series and moment problems is presented. For these problems, where

$$\int_0^T |u_1(t)|^p + |u_2(t)|^p \, dt$$

is minimized, the optimal controls are uniquely determined and have the same structure as in (2) but only with

$$\lambda_j(t) = \nu_j(t) = 1/(k+1), \quad \mu_j(t) = \omega_j(t) = 1/k.$$

This is true also for $p \in (1, 2)$, but then the proof uses the technique of [2].

For the corresponding L^∞ problem where the objective function is an essential supremum, the controls u_1, u_2 described as in the case $p \in (1, \infty)$ give the element of minimal L^2 norm in the solution set (see [3]).

Theorem 1 shows that, if the problem data T, L, or c is changed, the structure of the solution set can change if $T = kt_0$. Note that, even for C^∞ initial data, the optimal state may have jumps generated by the discontinuities of the optimal controls.

Problem (P) is related to control problems for nonlinear hyperbolic systems; see [1]. We hope that Theorem 1 helps to exploit the structure of optimal controls for quasilinear hyperbolic systems.

References

[1] M. Gugat. Boundary controllability between sub- and supercritical flow. *SIAM Journal on Control and Optimization,* 42:1056–1070, 2003.

[2] M. Gugat. L^1-Optimal boundary control of a string to rest in finite time. In *Recent Advances in Optimization (Proc. of the 12th French-German-Spanish Conference on Optimization).* Lecture Notes in Economics and Mathematical Systems, Springer (2005), to appear.

[3] M. Gugat, G. Leugering, and G. Sklyar. L^p optimal boundary control for the wave equation. *SIAM Journal on Control and Optimization,* 2005, to appear.

Joint Online Truck Scheduling and Inventory Management for Multiple Warehouses[1]

Christoph Helmberg

(joint work with Stefan Röhl)

Consider the following real-world problem. Given several warehouses connected by a shuttle service of several trucks for shipping pallets of stored articles between them, and given an online stream of orders that are stochastic and that have to be handled within short time at specific warehouses, find, online, a schedule of truck routes and truck loads so that (one hopes) all products are available at the right place ahead of processing time.

For this problem we suggest an approach based on convex relaxation of an integer programming formulation and demonstrate its practical suitability on real world data of our industrial partner eCom Logistik GmbH & Co. KG. For up to three warehouses and roughly 40 000 articles, the method computes a schedule within five to twelve minutes. In long-term simulations it reduces the average number of pallets that have to be transported on short notice because of demand to less than half the number of the semi-automatic approach currently in use.

Several issues are of relevance in this problem. An appropriate stochastic optimization model is required that links the success probability of the inventory of the warehouses to the truck rides. The model must be solvable within short time in order to be suitable for online computations, and the approach must be sufficiently robust to compensate frequent external changes in orders and uncertainties in the logistic transportation process.

In our method we follow the classical approach to model large-scale transportation or network design problems as multicommodity flow problems (see, e.g., [19, 16, 14]). These can be decomposed and solved efficiently with Lagrangian relaxation by combining min-cost flow algorithms (see, e.g., [1]) and bundle methods (see, e.g., [11, 4]). In particular, we model the rides of the trucks as well as the flow of pallets between warehouses by time-discretized networks coupled via linear capacity constraints. Our main contribution is the development of a convex piecewise-linear cost function that models the stochastic quality of the warehouse configurations as follows. We assume that for each article p and each warehouse

[1]This work was supported by research grant 03HEM2B4 of the German Federal Ministry of Education and Research (responsibility for the content rests with the authors).

w, a probability distribution $F_p^w : \mathbb{Z} \to [0,1]$ is given that assigns to a number α of pallets of article p the probability that demand at w for p over a specified period of time will not exceed α pallets. Our cost function penalizes the reduction of α available pallets by one by $g(F_p^w(\alpha - 1))$, where $g : [0,1] \to \mathbb{R}_+$ is required to be nonnegative, nonincreasing, and convex. Choosing g identical for all products p and warehouses w yields a reasonable priority on the sequence in which pallets should be transported. If we choose $g(x) = 1 - x$ and evaluate the cost function at the end of the planning period, we minimize the expected number of pallets that still have to be transported after the last truck ride. In practice, we compute estimates for F_p^w using the empirical distribution based on past demand (see, e.g., [8, 18]), choose $g(x) \approx -\log x$ to enforce transportation of pallets that are needed with high probability, and evaluate at several time steps. Even moderately accurate solutions for this cost structure give rise to reasonable schedules. Indeed, within five to ten minutes a schedule based on rounding, an approximate solution to linear programs with up to 1.5 million variables and 500,000 constraints is obtained on a Linux PC, using MCF [15] for solving the min-cost-flow problems and ConicBundle, which is an outgrowth of [9], for finding appropriate multipliers. Solving the linear program by a state-of-the-art simplex solver would need between half an hour and four hours on the same machine without leading to significantly better rounded solutions.

Within online environments, robustness hinges on reacting flexibly to new situations rather than sticking to past decisions. Consequently, our method does not keep any information on previous solutions but operates solely on status messages of the logistic operating system (the message system has been developed jointly with our industrial partner for this purpose). Therefore, the method is capable of continuing independent on what part of its proposed solution has been accepted by the human planner. The approach has been thoroughly tested in a detailed simulation study over a period of 100 days on the real-world online data stream of our industrial partner.

There is a vast literature on inventory management and logistics (see, e.g., [7]), yet we found few references that deal with both problems at the same time; and none of them treat both problems in sufficient detail for our purposes. In some works the transportation process is assumed to be instantaneous (see e.g. [12, 13, 5, 3, 20]), in others the stochastic part is fixed (see e.g. [2, 6]) or considerations are reduced to only one product [17]. To the best of our knowledge the approach proposed is the first that deals jointly with inventory management of multiple products and interwarehouse logistics involving vehicle routing with transportation times. The full paper [10], code, and data are available at `www.tu-chemnitz.de/mathematik/discrete/projects/warehouse_trucks/`

References

[1] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1993.

[2] S. Anily and A. Federgruen. Two-echelon distribution systems with vehicle routing costs and central inventories. *Oper. Res.*, 41(1):37–47, 1993.

[3] T. W. Archibald, S. A. E. Sassen, and L. C. Thomas. An optimal policy for a two depot inventory problem with stock transfer. *Manage. Sci.*, 43(2):173–183, 1997.

[4] S. Feltenmark and K. C. Kiwiel. Dual applications of proximal bundle methods, including Lagrangian relaxation of nonconvex problems. *SIAM J. Optim.*, 10(3):697–721, 2000.

[5] E. H. Frazelle, S. T. Hackman, U. Passy, and L. K. Platzman. The forward-reserve problem. In *Proceedings of the 1992 IBM Europe Institute on Optimization Solutions*, pages 43–61, Oberlech, Austria, 1994.

[6] F. Fumero and C. Vercellis. Synchronized development of production, inventory, and distribution schedules. *Transp. Sci.*, 33(3):330–340, 1999.

[7] S. Graves, A. Rinnooy Kan, and P. Zipkin, editors. *Logistics of production and inventory*, volume 4 of *Handbooks in Operations Research and Management Science*. North-Holland, Amsterdam, 1993.

[8] P. Hall. *The Bootstrap and Edgeworth Expansions*. Springer, New York, 1992.

[9] C. Helmberg and K. C. Kiwiel. A spectral bundle method with bounds. *Math. Programming*, 93(2):173–194, 2002.

[10] C. Helmberg and S. Röhl. A Case Study of Joint Online Truck Scheduling and Inventory Management for Multiple Warehouses Preprint 2005-3, Technische Universität Chemnitz, Fakultät für Mathematik, January 2005.

[11] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms II*, volume 306 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 1993.

[12] P. Köchel. Ein dynamisches Mehr-Lager-Modell mit Transportbeziehungen zwischen den Lagern. *Math. Operationsforsch. Statist., Ser. Optimization*, 13:267–286, 1982.

[13] P. Köchel. Optimal adaptive inventory control for a multi-location model with redistribution. *Optimization*, 19:525–537, 1988.

[14] B. Korte and J. Vygen. *Combinatorial optimization. Theory and algorithms*, volume 21 of *Algorithms and Combinatorics*. Springer, Berlin, 2 edition, 2002.

[15] A. Löbel. *MCF Version 1.2 – A network simplex Implementation*. Konrad-Zuse-Zentrum für Informationstechnik Berlin, Jan. 2000. Available at `http://www.zib.de/Optimization/Software/Mcf` (free of charge for academic use).

[16] M. Padberg. *Linear optimization and extensions*, volume 12 of *Algorithms and Combinatorics*. Springer, 2 edition, 1999.

[17] M. I. Reiman, R. Rubio, and L. M. Wein. Heavy traffic analysis of the dynamic stochastic inventory-routing problem. *Transp. Sci.*, 33(4):361–380, 1999.

[18] J. Shao. *Mathematical Statistics*, Springer Texts in Statistics. Springer-Verlag, New York, Berlin, Heidelberg, 2 edition, 2003.

[19] A. Schrijver. *Combinatorial Optimization*, volume 24 of *Algorithms and Combinatorics*. Springer, 2003.

[20] J. P. van den Berg, G. P. Sharp, A. J. R. M. (Noud) Gademann, and Y. Pochet. Forward-reserve allocation in a warehouse with unit-load replenishments. *Europ. J. Oper. Res.*, 111:98–113, 1998.

Stability in Nonlinear Optimization under Constraint Degeneracy

Diethard Klatte
(joint work with Bernd Kummer)

The stationary solution map X of a perturbed nonlinear program is studied. We characterize stability properties of X under constraint degeneracy (which here means that the linear independence constraint qualification (LICQ) is not satisfied at the solutions under consideration), and we compare this with known stability results in the nondegenerate case. We restrict our study here to a basic model: a nonlinear program under *canonical perturbations*, namely,

$$\mathrm{P}(p) \qquad \begin{array}{l} f(x) - a^{\mathsf{T}}x \to \min_x \\ \text{s.t.} \;\; g_i(x) \le b_i \;\; (i = 1, \dots, m) \end{array} , \quad p = (a, b) \text{ perturbation vector},$$

where $f, g_i \in C^2(\mathbb{R}^n, \mathbb{R})$ $\forall i$. In a similar way, certain perturbed variational inequalities could be handled, and additional equality constraints (or constraints with $C^{1,1}$ functions) or nonlinear parameterizations could be included.

Notation. Following Kojima [6], we write the Karush-Kuhn-Tucker (KKT) conditions for $\mathrm{P}(p)$, $p = (a, b)$ as

$$(1) \qquad \begin{array}{lcccll} \Phi_1(x,y) := Df(x) & + & \sum_{i=1}^m y_i^+ Dg_i(x) & = & a\,, & y_i^+ = \max\{0, y_i\}, \\ \Phi_{2i}(x,y) := \; g_i(x) & - & y_i^- & = & b_i\,, & y_i^- = \min\{0, y_i\}. \end{array}$$

We will study the *stationary solution set mapping*

$$X(p) := \{x \mid \exists y : \; (x, y) \text{ satisfies } (1)\}$$

for $(p, x) \in \operatorname{gph} X$ near $(0, x^0)$, where $x^0 \in X(0)$ is some given point. The *multiplier set* associated with $(0, x^0)$ is

$$Y^0 := \{y \,|\, (x^0, y) \text{ satisfies } (1) \text{ at } p = 0\}.$$

Without loss of generality suppose throughout $g(x^0) = 0$. Further, let

$$A_i = Dg_i(x^0)^{\mathsf{T}} \text{ (rows)}, \;\; L(x,y) = f(x) + \sum\nolimits_{i=1}^m y_i^+ g_i(x), \;\; Q(y) = D^2_{xx} L(x^0, y).$$

We are interested in the following three stability notions: X is said to be *locally upper Lipschitz (l.u.L.) at* $(0, x^0) \in \operatorname{gph} X$ if there exist $\mu, \varepsilon > 0$ such that

$$\|x - x^0\| \le \mu \|p\| \quad \forall x \in X(p) \cap B(x^0, \varepsilon), \; p \text{ near } 0,$$

in particular, $X(0) \cap B(x^0, \varepsilon) = \{x^0\}$ holds, and $X(p) \cap B(x^0, \varepsilon) = \emptyset$ for $p \neq 0$ is possible; X is called *strongly Lipschitz-stable (s.L.s.)* at $(0, x^0) \in \operatorname{gph} X$ if, for some $\varepsilon > 0$, $X(\cdot) \cap B(x^0, \varepsilon)$ is *single-valued and Lipschitz* near $p = 0$; X is called *strongly stable in Kojima's sense* at $(0, x^0)$ if, w.r.t. small quadratic perturbations of the objective and small righthand side perturbations of the constraints, $X(\cdot) \cap B(x^0, \varepsilon)$ is (for some $\varepsilon > 0$) single valued and continuous near $p = 0$.

Obviously, if X is s.L.s. at $(0, x^0)$, then X is l.u.L. at $(0, x^0)$ and fulfills $X(p) \cap B(x^0, \varepsilon) \neq \emptyset$ for some $\varepsilon > 0$ and all p near 0 (and hence the Mangasarian-Fromovitz constraint qualification (MFCQ) holds at x^0; see [5]), while the opposite direction already fails for linear programs under MFCQ.

Stability under LICQ. Suppose that LICQ holds at x^0; hence Y^0 is a singleton, say $Y^0 = \{y^0\}$. We recall some well-known results for this case.

Supposing in addition strict complementarity, the KKT system for (x, y, p) near $(x^0, y^0, 0)$ reduces to a C^1 equality system. Thus Φ^{-1} is locally single-valued and lies in C^1 (with the well-known formula for $D\Phi^{-1}$) if and only if $D\Phi(x^0, y^0)$ is nonsingular. This classical result (see [3]) uses the standard inverse function theorem.

Avoiding strict complementarity, Jongen et al. [2] have shown that X is s.L.s. at $(0, x^0)$ if and only if X is strongly stable in Kojima's sense at $(0, x^0)$.

At this point, let us answer a question of the discussion at the Oberwolfach meeting: it is not enough to require that X is locally single valued and continuous *under canonical perturbations only* to imply strong stability in Kojima's sense. The following one-dimensional parametric unconstrained program is a counterexample:

$$x^4 - ax - cx^2 \;\to\; \min_{x\in\mathbb{R}}, \quad |a|,\ |c| \text{ small.}$$

Next we characterize l.u.L. and s.L.s. of X under LICQ by linearization of the KKT system: *X is s.L.s at $(0, x^0)$* if and only if the system

$$\begin{array}{lllll} (i) & Q(y^0)u + \sum_{i=1}^m \alpha_i A_i{}^{\mathsf{T}} & = & 0, & \\ (ii) & y_i^0\ A_i\, u & = & 0 & (\forall i) \\ (iii) & \alpha_i\ A_i\, u & \geq & 0 & (\forall i) \end{array}$$

has the unique trivial solution $(u, \alpha) = 0$; see [4] for the above form. Several equivalent conditions are well known; see, for example, [7, 2, 1, 4]. Further, *X is l.u.L. at* $(0, x^0)$ if and only if the system

$$(i), (ii), (iii) \quad \text{and } (iv)\ A_i\, u \leq 0 \leq \alpha_i \text{ (for all } i : y_i^0 = 0)$$

has the unique trivial solution $(u, \alpha) = 0$; see, for example, [4].

Both characterizations may be rewritten in terms of quadratic auxiliary programs. Further, they show that under LICQ *and strict complementarity* at x^0, the three stability concepts coincide. However, under LICQ *without strict complementarity*, s.L.s. and l.u.L. already differ for (solvable) quadratic convex programs.

Stability under possible constraint degeneracy. The formal negation of the properties l.u.L. and s.L.s., respectively, leads to the violation of injectivity of appropriate generalized derivatives of X; for details of the corresponding concepts see [8, 4]. Thus, the negation of the local upper Lipschitz property gives

$$\begin{array}{ll} \textit{singularity 1:} & \exists \xi^k \in X(\pi^k) \text{ with } \xi^k \to x^0 \text{ and } \pi^k \to 0 \\ & \text{such that } \xi^k \neq x^0 \text{ and } \|\pi^k\|/\|\xi^k - x^0\| \;\to\; 0. \end{array}$$

This means that, by taking a cluster point u of $u^k = \frac{\xi^k - x^0}{\|\xi^k - x^0\|}$, some $u \neq 0$ belongs to $CX(0, x^0)(0)$, the *contingent derivative* of X at $(0, x^0)$ in direction 0.

Strong Lipschitz stability of X at $(0, x^0)$ is violated iff one has *local unsolvability* (i.e., there are $\varepsilon > 0$ and $p^k \to 0$ with $X(p^k) \cap B(x^0, \varepsilon) = \emptyset$), or

$$\begin{array}{ll} \textit{singularity 2:} & \exists x^k \in X(p^k),\ \xi^k \in X(\pi^k) \text{ with } x^k,\ \xi^k \to x^0 \text{ and } p^k,\ \pi^k \to 0 \\ & \text{such that } x^k \neq \xi^k \text{ and } \|\pi^k - p^k\|/\|\xi^k - x^k\| \;\to\; 0. \end{array}$$

The latter means, in terms of the *strict graphical (Thibault) derivative* TX, that some $u \neq 0$ belongs to $TX(0, x^0)(0)$.

A subtle analysis of CX and TX in terms of the original problem gives the following results (note that the above conditions under LICQ were obtained similarly). In the following, suppose MFCQ at x^0; for partial use of a weaker CQ see [5].

Singularity 1 holds if and only if there are $y^0 \in Y^0$ and (u, α) with $u \neq 0$ such that (y^0, u, α) satisfies the above system $(i), (ii), (iii)$ and (iv); see [4, Chapt. 8] for this result and further refinements. In [5], we prove that *singularity 2 holds* if and only if there are $y^0 \in Y^0$ and $u \neq 0$ such that

$$(ii) \quad y_i^0 \, A_i \, u \;=\; 0 \qquad (\forall i)$$

and for certain sequences $x^k \to x^0$ and α^k, one has

$$\begin{array}{lll} (i)' & Q(y^0)u + \sum \alpha_i^k Dg_i(x^k) & \to \quad 0, \\ (iii)' & \alpha_i^k \; A_i \, u & \geq \quad 0 \quad (\forall i). \end{array}$$

While the above characterizations under LICQ and that of singularity 1 use only information for P(0) at $\{x^0\} \times Y^0$, our condition for singularity 2, and hence for s.L.s. of X at $(0, x^0)$ under constraint degeneracy, depends on limits of data at a sequence $x^k \to x^0$. In [5], we discuss examples with convex polynomial problem functions which show that s.L.s. is not invariant when the problem functions are replaced by their quadratic approximations at x^0.

It is standard that *local solvability* may be guaranteed by appropriate second-order optimality conditions; see, for example, [1, 4]. In [5] we prove in the case of linear constraints under MFCQ that nonsingularity 2 automatically implies $X(p) \cap B(x^0, \varepsilon) \neq \emptyset$ for some $\varepsilon > 0$. For further stability results under constraint degeneracy (relations to Kojima's strong stability and the Aubin property, simplifications for linearly constrained programs, and so on), we again refer to [5].

References

[1] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems.* Springer, New York, 2000.

[2] H. Th. Jongen, T. Möbert, J. Rückmann, and K. Tammer. On inertia and Schur complement in optimization. *Lin. Algebra Appl.,* 95"97–109, 1987.

[3] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* Wiley, New York, 1968.

[4] D. Klatte and B. Kummer. *Nonsmooth Equations in Optimization.* Kluwer, Dordrecht-Boston-London, 2002.

[5] D. Klatte and B. Kummer. Strong Lipschitz stability of stationary solutions for nonlinear programs and variational inequalities. *SIAM J. Optim.*, to appear, 2005.

[6] M. Kojima. Strongly stable stationary solutions in nonlinear programs. In *Analysis and Computation of Fixed Points*, edited by S. M. Robinson, Academic Press, New York, 1980, pp. 93–138.

[7] S. M. Robinson. Strongly regular generalized equations. *Math. Oper. Res.,* 5:43–62, 1980.

[8] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.

Recent Progress in the NLP-SDP Code PENNON

Michal Kočvara
(joint work with Michael Stingl)

We report on recent advances in the computer program PENNON, aimed at solving optimization problems with nonlinear and semidefinite constraints. The problems are of the type

$$\begin{aligned} \text{(NLP-SDP)} \qquad & \min_{x\in\mathbb{R}^n} f(x) \\ & \text{s.t.} \qquad h_i(x) \le 0, \qquad i = 1,\dots,m_h \\ & \qquad\quad A(x) \preccurlyeq 0\,, \end{aligned}$$

where $f,\ h_i : \mathbb{R}^n \to \mathbb{R}$ are twice continuously differentiable and $A : \mathbb{R}^n \to \mathbb{S}^{m_A}$ is generally nonconvex. The principal idea of the algorithm is to replace the inequality constraints by penalized ones. With $p_i > 0$ for $i \in \{1,\dots,m\}$, we have

$$h_i(x) \le 0 \iff p_i\varphi(h_i(x)/p_i) \le 0, \quad i = 1,\dots,m_h$$

and

$$A(x) \preccurlyeq 0 \Longleftrightarrow \Phi_P(A(x)) \preccurlyeq 0\,,$$

where φ and Φ_P are smooth penalty functions satisfying a number of properties (see [4]). For the penalized problem we write the Lagrangian

$$F(x,u,U,p,P) = f(x) + \sum_{i=1}^{m_h} u_i p_i \varphi(h_i(x)/p_i) + \langle U, \Phi_P\left(A(x)\right)\rangle_{\mathbb{S}_{m_A}} \tag{1}$$

and define the algorithm, a variant of the method of multipliers:

PENNON algorithm:

$$\begin{aligned} &(i) && \text{Find } x^{k+1} \text{satisfying } \|\nabla_x F(x,u^k,U^k,p^k,P^k)\| \le \varepsilon^k \\ &(ii) && u_i^{k+1} = u_i^k\varphi'(h_i(x^{k+1})/p_i^k), \quad i = 1,\dots,m_h \\ & && U^{k+1} = D_A\Phi_p(A(x);U^k) \\ &(iii) && p_i^{k+1} < p_i^k,\ i = 1,\dots,m_h \\ & && P^{k+1} < P^k\,. \end{aligned}$$

The algorithm was originally proposed and analyzed for convex nonlinear programs (NLPs) by Polyak [7], later refined by Ben-Tal and Zibulevsky [1], generalized for nonconvex NLP by Breitfeld and Shanno [2], and recently revised by Polyak and Griva [8]. For details on the current SDP-NLP algorithm as used in PENNON, see [4, 5].

The approximate unconstrained minimization in Step (i) is performed by the modified Newton or trust-region method. In both cases, we find a search direction by solving the equation $Hd = -g$, where g and H is the gradient and Hessian of the Lagrangian (1) at a given point. In the standard version of the code we use (sparse or dense) Cholesky factorization to solve this system. Recently, we have implemented preconditioned conjugate gradient (PCG) method. The advantage

is twofold, depending on the structure of the matrix H and on the type of the constraints.

PCG and Large-Scale Sparse NLP. Large-scale NLP problems often have sparse structure, leading to sparse H and allowing us to use efficient variants of sparse Cholesky factorization. If, however, the gradient of just one constraint is dense, the matrix H becomes dense, too. In such a case, we replace the Cholesky factorization by PCG. In PCG, we need to compute only the Hessian-vector product Hz; we do not need to compute H explicitly. In the above situation, we evaluate and store only the "sparse part" of H (call it H_{sp}), and the dense gradients ∇h_i. The product is then performed by the formula of the type

$$Hz = H_{\mathrm{sp}} z + \sum_{\mathrm{dense}} \gamma_i \gamma_i^T z \,.$$

This strategy allows us to solve large "sparse-dense" problems. For instance, problem `lane-emden40` from the COPS3 collection [3] ($n = 19241, m_h = 81$) can be solved in 1 min 40 sec using 300 MB of memory, as compared to 26 min and 1600 MB needed by the Cholesky-based version.

PCG and Medium-Scale Dense SDP. Most linear SDP problems arising from various applications lead to a dense matrix H, even if the problem data matrices are sparse. For linear problems with $A(x) = \sum_{i=1}^{n} x_i A_i$ the complexity of Hessian computation is $O(m_A^3 n + m_A^2 n^2)$ for dense A_i and $O(m_A^2 n + K^2 n^2)$ for sparse A_i, where K is the maximum number of nonzeros in A_i, $i = 1, \ldots, n$. The complexity of the Cholesky factorization for a dense matrix is $O(n^3)$. For problems with $n \gg m$, the Cholesky method is expected to become the bottleneck of the code. In this case, the use of approximate PCG (with expected complexity of $O(n^2)$) can improve the efficiency significantly. Further, we need only the product with $\nabla^2 F(x_k)$, which can be replaced by a finite-difference formula

$$\nabla^2 F(x_k) z \approx \frac{\nabla F(x_k + \theta z) - \nabla F(x_k)}{\theta}$$

with $\theta = (1 + \|x_k\|_2 \sqrt{\varepsilon})$. Thus, in each PCG iteration, we need to compute just one gradient; this may significantly decrease the demands on the memory and CPU time (see [6] for details).

Table 1 gives comparison of the standard code with the PCG-based code with and without exact Hessian computation. Tested are problems with $n \gg m$ from [9], arising from maximum clique problems on randomly generated graphs (`theta*`) and maximum clique problems from the Second DIMACS Implementation Challenge.

References

[1] A. Ben-Tal and M. Zibulevsky. Penalty/barrier multiplier methods for convex programming problems. *SIAM J. Optim.*, 7:347–366, 1997.

[2] M. Breitfeld and D. Shanno. A globally convergent penalty-barrier algorithm for nonlinear programming and its computational performance. Technical report, Rutcor Research Report, Rutgers University, New Jersey, 1994.

TABLE 1. Results for selected TOH problems. PENSDP: standard code with Cholesky algorithm; P-PCG(diag): code with CG algorithm and diagonal preconditioner; P-A-PCG(BFGS): code with CG algorithm, approximate Hessian computations and BFGS preconditioner. CPU/it: time per a Newton iteration. Times in seconds; Sun UltraSparc III 1200 MHz with 4 GB RAM.

			PENSDP		P-PCG(diag)		P-A-PCG(BFGS)	
Problem	n	m	CPU	CPU/it	CPU	CPU/it	CPU	CPU/it
ham_7_5_6	1 793	128	104	3.2	19	0.7	4	0.1
ham_9_8	2305	512	266	9.8	138	5.3	210	4.7
ham_8_3_4	16129	256	71264	2036.1	2983	80.1	104	2.7
ham_9_5_6	53761	512	memory		memory		1984	37.4
theta42	5986	200	3978	104.6	391	9.3	51	1.2
theta6	4375	300	1719	42.9	197	5.3	108	2.0
theta62	13390	300	51359	1222.8	3779	77.1	196	4.3
theta8	7905	400	8994	243.0	783	19.1	263	5.
theta82	23872	400	memory		memory		650	14.4
theta10	12470	500	30610	956.5	6571	126.4	492	10.
theta102	37467	500	memory		memory		1948	47.5
theta103	62516	500	memory		memory		6149	149.9
theta104	87845	500	memory		memory		8400	215.3
theta12	17979	600	timed out		14098	223.7	843	16.2
theta123	90020	600	memory		memory		11733	266.66
theta162	127600	800	memory		memory		50098	927.74
keller4	5101	171	3724	66.5	297	6.5	52	1.1
sanr200-0.7	6 033	200	4210	107.9	393	9.1	52	1.2

[3] E. D. Dolan, J. J. Moré, and T. S. Munson. Benchmarking Optimization Software with COPS 3.0. Technical report, ANL/MCS-TM-273, Argonne National Laboratory, Mathematics and Computer Science Division, February 2004.

[4] M. Kočvara and M. Stingl. PENNON—a code for convex nonlinear and semidefinite programming. *Optim. Methods and Software*, 18:317–333, 2003.

[5] M. Kočvara and M. Stingl. Solving nonconvex SDP problems of structural optimization with stability control. *Optim. Methods and Software*, 19:595–609, 2004.

[6] M. Kočvara and M. Stingl. On the solution of large-scale SDP problems by the modified barrier method using iterative solvers. Research report 304, Institute of Applied Mathematics, University of Erlangen, 2005.

[7] R. A. Polyak. Modified barrier functions: Theory and methods. *Math. Prog.*, 54:177–222, 1992.

[8] R. A. Polyak and I. Griva. Nonlinear rescaling in constrained optimization: Primal-dual approach. Technical report, George Mason University, 2004.

[9] K. C. Toh. Solving large scale semidefinite programs via an iterative solver on the augmented systems. *SIAM J. Optim.*, 14:670–698, 1999.

Local Structure and Algorithms in Nonsmooth Optimization

Adrian S. Lewis
(joint work with James V. Burke and Michael L. Overton)

The Belgian Chocolate Problem. Illustrating the difficulty of control design problems, Blondel [1] proposed the following problem in 1994:

> Given a real δ, find stable real polynomials p and q such that the polynomial $r(s) = (s^2 - 2\delta s + 1)p(s) + (s^2 - 1)q(s)$ is also stable.

(We call a polynomial p *stable* if its *abscissa* $\alpha(p) = \max\{\operatorname{Re} s : p(s) = 0\}$ is nonpositive.) Clearly the problem is unsolvable if $\delta = 1$, since then $r(1) = 0$; more delicate results (summarized in [7]) show it remains unsolvable for $\delta < 1$ close to 1. Blondel offered a prize of 1 kg of Belgian chocolate for the case $\delta = 0.9$, a problem solved by using randomized search in [7].

To illustrate the theme of this talk, we first outline (based on joint work with D. Henrion) a more systematic optimization approach to the chocolate problem. We fix the degrees of the polynomials p and q (say 3). Without loss of generality, suppose p is monic, and consider the resulting problem

$$(CP) \qquad \min\{\alpha(pqr) : p, q \text{ cubic, } p \text{ monic.}\}$$

A feasible solution with negative objective value would solve Blondel's problem.

A Simple Nonsmooth Algorithm. For nonsmooth optimization problems like (CP), it is convenient to have on hand a simply-implementable, intuitive, robust algorithm for minimizing a nonsmooth function f. We present such a method in [3]. To motivate it, suppose for simplicity f (unlike the abscissa α) is Lipschitz.

Fundamental for good behavior in nonsmooth optimization is the *regularity* of the function f at points x, which means we can write the directional derivative as

$$f'(x;d) = \limsup_{y \to x} \nabla f(x)^T d, \text{ for all } d$$

(noting the almost everywhere differentiability of f on its domain in $\mathbb{R}^n$). Both convex and smooth functions are regular. Assuming regularity, we can check that the steepest descent direction at x is

$$-\lim_{\epsilon \downarrow 0} \operatorname{argmin}\big\{\|d\| : d \in \operatorname{co}\{\nabla f(y) : y \in x + \epsilon B\}\big\},$$

where B denotes the unit ball. The *gradient sampling* algorithm of [3] approximates this direction by a random vector

$$G_\epsilon^m(x) = -\operatorname{argmin}\big\{\|d\| : d \in \operatorname{co}\{\nabla f(Y_i) : i = 1, 2, \ldots, m\}\big\},$$

for some fixed radius ϵ, fixed $m > n$, and independent, uniformly distributed, random points $Y_i \in x + \epsilon B$. (In practice, we add the point x.) The algorithm then performs a simple line search along this direction, and repeats.

The Performance of Gradient Sampling. The gradient sampling algorithm is intuitive and straightforward to implement when function and gradient evaluations are cheap. Experiments on a wide variety of examples are promising [3]. Rigorous justifications include

– the almost sure convergence of the search direction $G_\epsilon^m(x)$ to a "robust" steepest descent direction as the sample size m grows [2], and
– convergence results for the algorithm under a variety of underlying assumptions and implementation regimes (for reducing the radius ϵ, for example) [3].

Among these results, however, the following fact is particularly suggestive of the "smoothing" effect of the algorithm.

Theorem 1. *The expectation of the search direction $G_\epsilon^m(x)$ depends continuously on the point x.*

We sketch a proof suggested by S. Henderson. First, we sample the points Y_i corresponding to the current point x, as above. Next, we construct random points Y_i' corresponding to a perturbed point x', but "coupled" with the points Y_i as follows. If $Y_i \in x' + \epsilon B$, then we set $Y_i' = Y_i$; otherwise we choose Y_i' uniformly distributed on the set $(x' + \epsilon B) \setminus (x + \epsilon B)$. The resulting random points Y_i' are mutually independent, and uniformly distributed on the ball $x' + \epsilon B$, as required. Since the set $(x + \epsilon B) \setminus (x' + \epsilon B)$ has measure $O(\|x - x'\|)$, the sets $\{Y_i\}$ and $\{Y_i'\}$ (and hence the vectors $G_\epsilon^m(x)$ and $G_\epsilon^m(x')$) are identical with probability $1 - O(\|x - x'\|)$. On the other hand, even if this latter event does not occur, since f is Lipschitz, the vector $G_\epsilon^m(x) - G_\epsilon^m(x')$ is uniformly bounded. In summary, the expectation of this latter vector must be $O(\|x - x'\|)$.

Solving the Chocolate Problem. The gradient sampling algorithm suggests numerically that the solution of the problem (CP) for any value of δ near 0.9 has a distinctive structure: the polynomial q is a constant, and the polynomial r has a negative real zero of order five. Armed with this observation, a simple hand calculation reveals a unique feasible solution of this form under the assumption $\delta < \frac{1}{2}\sqrt{2 + \sqrt{2}} \approx 0.924$, in particular solving Blondel's problem.

A nice exercise in nonsmooth calculus verifies our numerical observation that the above solution is indeed a local minimizer for the problem (CP), at least when we further restrict the polynomial q to be constant. The requisite nonsmooth chain rule we need relies heavily on the following striking result [4].

Theorem 2. *The abscissa α is regular throughout the set of degree-k polynomials.*

Structural Persistence in Nonsmooth Optimization. The persistent solution structure for the chocolate problem (CP) as the parameter δ varies illustrates another important feature of concrete nonsmooth optimization problems, akin to active set phenomena in nonlinear programming. For classical nonlinear programs, the second-order sufficient conditions have several important consequences:

(i) The current point is a strict local minimizer.
(ii) As we perturb the problem's parameters, this minimizer varies smoothly on an "active" manifold.

(iii) We can calculate perturbed minimizers via smooth systems of equations.

Properties (ii) and (iii) do not rely fundamentally on second-order theory; indeed, they also hold for a broad class of nonsmooth functions introduced in [6].

For simplicity once again, we restrict attention to Lipschitz functions f. We call f *partly smooth* relative to the *active manifold* $\mathcal{M}$ if f is regular throughout $\mathcal{M}$ and the directional derivative $f'(x;d)$ is continuous as x varies on $\mathcal{M}$, with

$$f'(x;-d) > -f'(x;d) \text{ whenever } 0 \neq d \perp \mathcal{M} \text{ at } x.$$

This last condition enforces a "vee-shape" on the graph of f around a "ridge" corresponding to $\mathcal{M}$. Partial smoothness holds, for example, for the function $x \mapsto \max\{x_i\}$, the Euclidean norm, and the maximum eigenvalue of a symmetric matrix, and the property is typically preserved under smooth composition, generating a wealth of applications. Furthermore, critical points of partly smooth functions typically satisfy the sensitivity properties (ii) and (iii) above.

The structural persistence we first observed numerically in the chocolate problem (CP) is explained by the following refinement of Theorem 2. We associate with any polynomial p a *list of multiplicities* of those zeroes of p with real part equal to the abscissa, listed in order of decreasing imaginary part.

Theorem 3. *The abscissa α is partly smooth relative to any manifold of polynomials having a fixed list of multiplicities.*

By contrast with the sensitivity properties (ii) and (iii) above, convenient checks for property (i) (strict local minimality) do typically involve second-order analysis. For partly smooth functions f, the extra assumption we need is *prox-regularity* [5]. This property requires, locally, that the nearest-point projection onto the epigraph $\{(x,r) : r \geq f(x)\}$ be unique (as typically holds if f is the pointwise maximum of some smooth functions, for example). The question of the prox-regularity of the abscissa α remains open. The essential ingredient is the following question, with which we end.

Question 1. *Does every degree-k polynomial $p(s)$ near the polynomial s^k have a unique nearest stable polynomial?*

References

[1] V. D. Blondel. *Simultaneous Stabilization of Linear Systems.* Springer, Berlin, 1994.

[2] J. V. Burke, A. S. Lewis, and M. L. Overton. Approximating subdifferentials by random sampling of gradients. *Math. of O.R.,* 27:567–584, 2002.

[3] J. V. Burke, A. S. Lewis, and M. L. Overton. A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM J. Opt.*, to appear.

[4] J..V. Burke and M. L. Overton. Variational analysis of the abscissa mapping for polynomials. *SIAM J. Control Opt.,* 39:1651–1676, 2001.

[5] W. L. Hare and A. S. Lewis. Identifying active sets via partial smoothness and prox-regularity. *J. Convex Analysis,* to appear.

[6] A. S. Lewis. Active sets, nonsmoothness and sensitivity. *SIAM J. Opt.,* 13:702–725, 2003.

[7] V. V. Patel, G. Deodhare, and T. Viswanath. Some applications of randomized algorithms for control system design. *Automotica,* 28:2085–2092, 2002.

The Return of the Active Set Method

Sven Leyffer (Argonne National Laboratory)

For solving nonlinear optimization problems, two competing iterative approaches are available: active set methods and interior-point methods. Current implementations of interior methods often outperform active set methods in terms of speed. On the other hand, active set methods are more robust and better suited for warm starts, which are important for solving integer optimization problems [8, 9]. Consequently, we have recently become interested in new active set approaches, which are reviewed in this note.

1. Active Set Methods for Quadratic Programs

Consider the quadratic programming (QP) problem

$$\underset{x}{\text{minimize}}\ \frac{1}{2}x^T H x + g^T x \quad \text{subject to}\ \ A^T x = b\ \ \text{and}\ \ l \le x \le u,$$

where $A \in \mathbb{R}^{n\times m}$ has full rank and H is symmetric but not necessarily positive definite. Our new active set approach has two main components.

First, we identify an estimate of the optimal active set by approximately minimizing the augmented Lagrangian

$$L(x, y, \rho) := \frac{1}{2}x^T H x + g^T x - y^T(A^T x - b) + \frac{1}{2}\rho\|A^T x - b\|^2,$$

in the box $l \le x \le u$, where y are the multipliers of $A^T x = b$ and ρ_k is the penalty parameter. This step provides a Cauchy-point x_c^k, a first-order multiplier estimate $y_c^k = y^k - \rho_k(A^T x_c^k - b)$, and an active set estimate $\mathcal{A}_k := \{i : [x_c^k]_i = l_i \ \text{ or } [x_c^k]_i = u_i\}$. This step is similar to the iterates generated by LANCELOT [3].

Next, we solve an equality constraint QP in the remaining inactive variables indexed by $\mathcal{I} := \{1, \ldots, n\}\ \mathcal{A}$ by computing an approximate solution to the first-order conditions

$$\begin{bmatrix} H_{\mathcal{I},\mathcal{I}} & -A_{:,\mathcal{I}} \\ A_{:,\mathcal{I}}^T & \end{bmatrix} \begin{pmatrix} \Delta x_{\mathcal{I}} \\ \Delta y \end{pmatrix} = -\begin{pmatrix} [\nabla x L(x_k^c, y_k^c, 0)]_{\mathcal{I}} \\ A^T x_k^c - b \end{pmatrix},$$

where $H_{\mathcal{I},\mathcal{I}}$ is the submatrix of H corresponding to rows and columns of $\mathcal{I}$. We then perform a backtracking line-search along $(x_c^k + \alpha\Delta x_{\mathcal{I}}, y_c^k + \alpha\Delta y)$ to ensure global convergence. We show that if $\alpha = 1$, then the two steps are equivalent to a Newton step on the first-order conditions.

Global convergence is enforced through the use of a filter [6, 7]. A filter $\mathcal{F}$ is a list of pairs of constraint violation $h^l := \|A^T x^l - b\|$ and projected gradient error $\theta^l := \|\nabla_x \mathcal{L}^l - z^l\|$ (z^l are the multipliers of the box constraints). A new point x^k is acceptable to the filter if $h^k \le \beta h^l$, or $\theta^k \le \beta\theta^l$ for all $l \in \mathcal{F}$. The backtracking line-search reduces α until an acceptable iterate is found. This acceptable iterate may be added to the filter.

Traditional proofs for augmented Lagrangian methods use two forcing sequences $\eta_k \searrow 0$ and $\omega_k \searrow 0$ to control progress in h^k and θ^k, respectively, and guide the

penalty parameter. Recently, it has been shown [4] that $\eta_k \searrow 0$ is sufficient to ensure convergence for certain QPs. The filter approach removes the need for any forcing sequence whose choice may be problematic in practice. Preliminary numerical experience is encouraging, and we are able to detect the optimal active set in a modest number of iterations.

2. Active Set Methods for Nonlinear Programs

Recently, researchers have expressed renewed interest in sequential linear programming (SLP) methods for nonlinear optimization problems such as

$$\underset{x}{\text{minimize}} \ f(x) \quad \text{subject to} \ \ c(x) \geq 0;$$

see [5, 2, 1]. These SLP methods solve a trust-region LP around the current iterate x^k, given by

$$\underset{d}{\text{minimize}} \ g_k^T d \quad \text{subject to} \ \ c_k + A_k^T d \geq 0 \ \text{ and } \ \|d\|_\infty \leq \Delta_k,$$

where $g_k = \nabla f(x^k)$, $c_k = c(x^k)$, and $A_k = \nabla c(x^k)^T$. The solution of this LP provides an estimate of the active inequality constraints, which is used to define an equality constrained QP to compute a second-order step.

One problematic aspect of this approach is the use of the ℓ_∞ trust-region. It has been observed that while the active constraints corresponding to $c(x) \geq 0$ settle down, the active trust-region bounds do not, and this feature may cause the LP solver to perform many wasteful pivots even close to the solution.

We propose an alternative trust-region subproblem based on penalizing an elliptic or ℓ_2 trust-region. This gives rise to the following active set identification problem

$$\underset{d}{\text{minimize}} \ \mu g_k^T d + \frac{1}{2} d^T d \quad \text{subject to} \ \ c_k + A_k^T d \geq 0.$$

It can be shown that the dual of this problem is a bound-constrained QP in the multipliers y,

$$\underset{y}{\text{minimize}} \ \frac{1}{2} y^T A^T A y - (c - \mu A^T g)^T y + \frac{\mu^2}{2} g^T g \quad \text{subject to} \ \ y \geq 0.$$

Convergence of a filter algorithm along the lines of [2] can be shown. The proof exploits a piecewise quadratic relationship between the penalty parameter μ and the ℓ_2 trust-region radius Δ.

3. Conclusions

We have introduced two active set identification strategies for optimization. Both schemes can be implemented in a matrix-free format, requiring only matrix-vector operations and iterative linear system solves. We believe that this is an important ingredient for a successful large-scale active set strategy.

References

[1] R. H. Byrd, N. I. M. Gould, J. Nocedal, and R. A. Waltz. An algorithm for nonlinear optimization using linear programming and equality constrained subproblems. *Mathematical Programming, Series B*, 100(1):27–48, 2004.

[2] C.M. Chin and R. Fletcher. On the global convergence of an SLP-filter algorithm that takes EQP steps. *Mathematical Programming*, 96(1):161–177, 2003.

[3] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods*. MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2000.

[4] Z. Dostal, A. Friedlander, and S. A. Santos. Augmented Lagrangian with adaptive precision control for quadratic programming with equality constraints. *Computational Optimization and Appllications*, 14(1):37–53, 1999.

[5] R. Fletcher and E. Sainz de la Maza. Nonlinear programming and nonsmooth optimization by successive linear programming. *Math. Programming*, 43:235–256, 1989.

[6] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Math. Programming*, 91:239–270, 2002.

[7] R. Fletcher, S. Leyffer, and Ph. L. Toint. On the global convergence of a filter-SQP algorithm. *SIAM J. Opt.*, 13(1):44–59, 2002.

[8] C. A. Floudas. *Nonlinear and Mixed–Integer Optimization*. Topics in Chemical Engineering. Oxford University Press, New York, 1995.

[9] I. E. Grossmann. Review of nonlinear mixed-integer and disjunctive programming techniques. *Optimization and Engineering*, 3:227–252, 2002.

Stability Optimization of Periodic Walking and Running Motions

Katja D. Mombaur

(joint work with H. G. Bock, J. P. Schlöder, R. W. Longman)

The model-based mathematical investigation of walking and running motions plays an important role in many areas of application, such as the design of walking robots, computer graphics, sports, and medicine. Mathematical models of walking involve distinct model phases with possibly different degrees of freedom, each described by a different set of nonlinear differential and algebraic equations with invariants

$$
\begin{array}{rcl}
(\dot{q}^T(t), \dot{v}^T(t)) & = & (v(t)^T, a^T(t)) \\
\begin{pmatrix} M(q(t),v(t),p) & G^T(q(t),p) \\ G(q(t),p) & 0 \end{pmatrix} \cdot \begin{pmatrix} a \\ \lambda \end{pmatrix} & = & \begin{pmatrix} f(q(t),v(t),u(t),p) \\ \gamma(q(t),v(t),p) \end{pmatrix} \\
g_{pos} = g(q(t),p) = 0 & & g_{vel} = G(q(t),p) \cdot \dot{q}(t) = 0.
\end{array}
$$

Phase boundaries are implicitly defined by the roots of some switching functions $s_i(t, q(t), v(t), p) = 0$. At these points, there may be discontinuities in the righthand side of the linear system, that is, $\Delta f(q, v, u, p), \Delta\gamma(q, v, p)$ (which translates into discontinuities in the accelerations Δa), or even in the velocities, $\Delta v(t, q, v, u, p)$, which are part of the state variables $x^T = (q^T, v^T)$). Walking problems also involve a number of complex linear and nonlinear, coupled and decoupled equality and inequality constraints; one example is the periodicity constraint on the state variables (or a subset thereof) $\tilde{x}(T) = \tilde{x}(0)$.

A crucial aspect of walking and running is stability. Stabilization by means of feedback control is a complex issue, and the large amount of online computations performed by the feedback system restricts the speed of operation. As an alternative approach we therefore investigate the existence of open-loop stable (i.e., self-stabilizing) systems that – under small perturbations – asymptotically converge back to the precomputed periodic orbit without any feedback corrections at all, relying just on the inherent dynamic and kinematic properties.

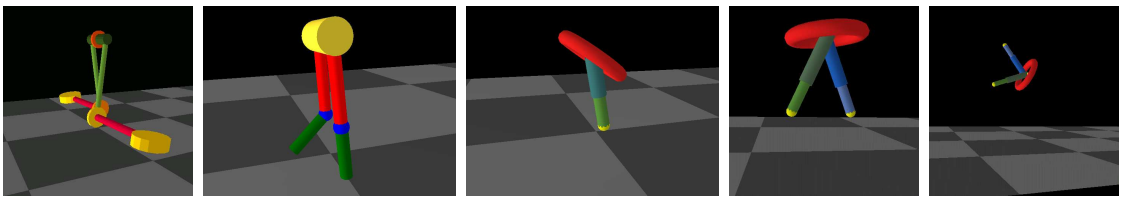
The determination of open-loop stable solutions for such complex systems as walking robots is possible only by means of optimization. We have developed a numerical method for the optimization of open-loop stability of periodic systems based on a two-level approach [1]. In the outer loop, a stability optimization is performed with the model parameters left free for variation. Stability is defined by using a generalization of Lyapunov's first method to multiphase systems with discontinuities: the spectral radius $|\lambda|_{max}$ of the nonsymmetric monodromy matrix C_x must be smaller than 1 for the system to be stable. This requirement leads to the optimization criterion

$$\min_p |\lambda(C_x(p))|_{max},$$

which is nondifferentiable, may be non-Lipschitz at points of multiple maximum eigenvalue, and involves the derivatives of the Poincaré mapping; it thus represents a difficult nonstandard optimization criterion. The task of the inner loop is a generation of an energy optimal periodic gait solution for the model parameter values given by the outer loop, that is, a solution of the optimal control problem

$$\begin{aligned}
\text{multiphase DAE-model} \qquad & \min_{x,u,T} \int_0^T \|u\|_2^2 dt \\
& \text{subject to} \\
x(\tau_j^+) &= h(x(\tau_j^-)) \quad \text{for } j = 1, \ldots, n_{ph} \\
g_j(t, x(t), u(t), p) &\geq 0 \qquad \text{for } t \in [\tau_{j-1}, \tau_j] \\
r_{eq/ineq}(x(0), \ldots, x(T), p) &= / \geq 0.
\end{aligned}$$

For the solution of this problem we built on the optimal control methods based on the direct boundary value problem approach developed by Bock and Plitt [2] and Leineweber [3] and adapted these methods to handle 3-index DAEs. The outer-loop nondifferentiable problem has been solved by using a direct search method.



This method has been used to design a number of fundamentally new open-loop stable robot configurations with one and two legs performing walking, running, and gymnastics motions; some are shown in the figure (for details, see, e.g., [4, 5, 6]). The range of application of this method is not restricted to robots and

biomechanical systems however; it can be used for any periodic dynamic process with discontinuities, for instance, from chemistry or economics.

References

[1] K. D. Mombaur, H. G. Bock, J. P. Schlöder, and R. W. Longman. Open-loop stable solution of periodic optimal control problems in robotics. *Zeitschrift für Angewandte Mathematik und Mechanik (ZAMM)*, to appear, 2005.

[2] H. G. Bock and K.-J. Plitt. A multiple shooting algorithm for direct solution of optimal control problems. In *Proceedings of the 9th IFAC World Congress*, Budapest, 1984, pp. 242–247.

[3] D. B. Leineweber. Efficient reduced SQP methods for the optimization of chemical processes described by large sparse DAE models, VDI-Fortschrittbericht, Reihe 3, no. 613, 1999.

[4] K. D. Mombaur, R. W. Longman, H. G. Bock, and J. P. Schlöder. Open-loop stable running. *Robotica,* 23:21–33, 2005.

[5] M. J. Coleman, M. Garcia, K. D. Mombaur, and A. Ruina. Prediction of stable walking for a toy that cannot stand. *Physical Review E,* 2:2001.

[6] K. D. Mombaur, H. G. Bock, J. P. Schlöder, and R. W. Longman. Human-like actuated walking that is asymptotically stable without feedback. In *Proc. IEEE International Conference on Robotics and Automation.* Seoul, 2001, pp. 4128–4133.

Global Performance of the Newton Method

Yurii Nesterov

(joint work with Boris Polyak)

We consider a special strategy for regularizing the standard Newton method as applied to the unconstrained minimization problem

$$f(x) \quad \rightarrow \quad \min \ : \ x \in E,$$

where E is a finite-dimensional real vector space and f is two times continuously differentiable function whose Hessian is Lipschitz continuous on some convex open set $\mathcal{F} \subseteq E$:

$$\|f''(x) - f''(y)\| \leq L\|x - y\|, \quad x, y \in \mathcal{F}.$$

The classical Newton iterate $y_f(x)$ is defined as a minimizer of the quadratic model of f around x:

$$\begin{aligned} y_f(x) &= \operatorname*{argmin}_y \ V_{f,x}(y) \\ V_{f,x}(y) &= f(x) + \langle f'(x), y - x\rangle + \tfrac{1}{2}\langle f''(x)(y - x), y - x\rangle. \end{aligned}$$

Sometimes this iterate is not well defined. Therefore we suggest using the following regularization [1]:

$$c_{f,M}(x) \quad = \quad \arg\min_y \left[V_{f,x}(y) + \tfrac{M}{6}\|y - x\|^3\right],$$

where M is a positive parameter. The advantages of this construction follow from two observations:

– If $M \geq L$, then for any $y \in \mathcal{F}$ we have

$$f(y) \leq V_{f,x}(y) + \tfrac{M}{6}\|y - x\|^3.$$

– The point $c_{f,M}(x)$ can be computed by standard techniques from linear algebra from a convex univariate minimization problem (as could be inferred from [2]).

Thus, iterating

$$x_{k+1} = c_{f,M}(x_k), \quad k = 0, 1, \ldots, \tag{1}$$

we obtain a sequence that satisfies all naturally expected properties of a second-order process:

– any limit point satisfies second-order necessary optimality conditions;

– the local rate of convergence is quadratic;

– the global rate of convergence in terms of the norm of the gradient is better than that of the gradient method.

Moreover, for some nontrivial classes of (nonconvex) problems, we managed to justify for (1) a *global* rate of convergence. For example, if f is star-convex, then (1) converges globally as $O(\frac{1}{k^2})$, where k is the iteration counter.

A similar idea can be used for modifying the Gauss-Newton method [3] as applied to the system of nonlinear equations

$$F(x) = 0,$$

with $F : E_1 \to E_2$. Introducing for E_i the norms $\|\cdot\|_i$, $i = 1, 2$, we can define the corresponding *operator norm* $\|\cdot\|_{12}$:

$$\|A\|_{12} = \max_{x \in E_1} \{\|Ax\|_2 : \ \|x\|_1 \leq 1\}.$$

Let us assume that F has a Lipschitz-continuous Jacobian:

$$\|F'(x) - F'(y)\|_{12} \leq L\|x - y\|_1, \quad x, y \in \mathcal{F}.$$

Then, we can define

$$g_{F,M}(x) \quad = \quad \arg\min_y \left[\ \|F(x) + F'(x)(y - x)\|_2 + \tfrac{M}{2}\|y - x\|_1^2 \ \right],$$

where M is a positive parameter. As before, this object is computable in polynomial time. Moreover, if $M \geq L$, then for any $y \in \mathcal{F}$ we have

$$\|F(y)\|_2 \quad \leq \quad \|F(x) + F'(x)(y - x)\|_2 + \tfrac{M}{2}\|y - x\|_1^2.$$

This inequality is essential for global and local convergence analysis of the process

$$x_{k+1} = g_{F,M}(x_k), \quad k = 0, 1, \ldots.$$

References

[1] Yu. Nesterov and B. T. Polyak. Cubic regularization of a Newton scheme and its global performance. CORE Discussion paper #2003/41, 2003. Submitted to *Math. Prog.*

[2] B. T. Polyak. Convexity of quadratic transformations and its use in control and optimization. JOTA, 99(3), p.553-583, 1998.

[3] Yu. Nesterov. Modified Gauss-Newton scheme with worst-case guarantees for its global performance. CORE Discussion Paper #2003/86, 2003. Submitted to *SIOPT*.

A Boundary-Point Method for Semidefinite Programs

Franz Rendl
(joint work with I. Dukanovic, J. Povh, A. Wiegele)

1. Semidefinite Programs and Projection onto the Positive Semidefinite Cone

Recently, Burer and Vandenbussche [2] proposed to solve lift-and-project relaxations of binary integer problems by exploiting the augmented Lagrangian technique applied to the primal positive semidefinite program (SDP). We will apply the idea to the dual SDP; see also [4, 5]. This approach seems to be more natural, and also gives an interesting interpretation of the augmented Lagrangian approach as a boundary-point method.

We consider the following primal-dual pair of SDP problems, given by symmetric $n \times n$ matrices C and A_i, $i = 1, \ldots, m$ and a vector $b \in \mathbb{R}^m$:

$$\text{(P)} \qquad \max\langle C, X\rangle\,, \quad \text{such that } A(X) = b,\ X \succeq 0\,,$$

whose dual is

$$\text{(D)} \qquad \min b^T y \quad \text{such that } A^T(y) - C = Z \succeq 0\,.$$

We make the usual assumption that both problems have strictly feasible points, so that strong duality holds and (X, y, Z) is optimal if and only if

$$X \succeq 0\,, \quad A(X) = b\,, \quad Z \succeq 0\,, \quad A^T(y) - Z = C\,, \quad ZX = 0\,.$$

We have problems in mind where the size n of the primal matrix is not too large, say $n \leq 1000$, but the number m of constraints could be arbitrary.

If W is a symmetric matrix with eigenvalue decomposition $W = P\Lambda P^T$, we partition P and the diagonal matrix Λ according to positive and negative eigenvalues: $\Lambda = (\Lambda_1, \Lambda_2)$, $P = (P_1, P_2)$ with $\mathrm{diag}\,(\Lambda_1) \geq 0$, $\mathrm{diag}\,(\Lambda_2) < 0$. Thus $W = P_1\Lambda_1 P_1^T + P_2\Lambda_2 P_2^T = W_+ + W_-$. It is well known that

$$\min_{Z \succeq 0} \|Z - W\|^2$$

is attained at $Z = W_+ = P_1\Lambda_1 P_1^T$. In other words, the projection of W onto the cone of semidefinite matrices is given by setting the negative eigenvalues of W to zero.

2. An Augmented Lagrangian Approach to Solve (D)

We apply the augmented Lagrangian method to solve (D). Thus we introduce a Lagrange multiplier X for the dual equations $Z + C - A^T(y) = 0$ and consider for fixed $\sigma > 0$ the augmented Lagrangian L_σ:

$$L_\sigma(y, Z; X) := b^T y + \langle X, Z + C - A^T(y)\rangle + \frac{\sigma}{2}\|Z + C - A^T(y)\|^2.$$

Define

$$W(y) := A^T(y) - C - \frac{1}{\sigma}X, \tag{1}$$

so that $L_\sigma(y, Z; X) = b^T y + \frac{\sigma}{2}\|Z - W(y)\|^2 - \frac{1}{2\sigma}\|X\|^2$, and let

$$f(y, Z) := b^T y + \frac{\sigma}{2}\|Z - W(y)\|^2 . \tag{2}$$

The augmented Lagrangian method to solve (D) consists in minimizing $f(y, Z)$ (approximately), to get y and $Z \succeq 0$. Then X is updated to $X + \sigma(Z + C - A^T(y))$ and the whole process is iterated until convergence; see [1].

Clearly, the crucial step here is minimizing $f(y, Z)$, so we take a closer look at the optimality conditions of this problem.

3. Optimality Conditions for the Inner Minimization

To minimize $f(y, Z)$ of (2) subject to $y \in \mathbb{R}^m$, $Z \succeq 0$ is a convex quadratic SDP. After introducing a Lagrange multiplier $V \succeq 0$ for the constraint $Z \succeq 0$, we get the Lagrangian

$$L(y, Z, V) := f(y, Z) - \langle V, Z\rangle$$

and the following KKT necessary conditions for optimality:

$$\nabla_y L = 0\,, \quad \nabla_Z L = 0\,, \quad V \succeq 0\,, \quad Z \succeq 0\,, \quad VZ = 0\,.$$

Since the problem is convex and the Slater condition holds, these conditions are also sufficient for optimality. Expanding the gradient conditions, we note that y, Z is optimal if and only if there exists V such that

$$A(A^T(y)) = A(Z + C) + \tfrac{1}{\sigma}(A(X) - b)\,, \tag{3}$$

$$V = X - \sigma W(y)\,,\ Z \succeq 0\,,\ V \succeq 0\,,\ VZ = 0\,. \tag{4}$$

For y fixed, the problem $\min_{Z\succeq 0} f(y, Z)$ is a projection onto the cone of semidefinite matrices. Therefore, Z must also satisfy the projection condition

$$Z = W(y)_+. \tag{5}$$

Thus we can reformulate the necessary and sufficient conditions for optimality as follows: (y, Z, V) is optimal if and only if

$$A(A^T(y)) = A(Z + C) + \frac{1}{\sigma}(A(X) - b)\,, \quad Z = W(y)_+\,, \quad V = -\sigma W(y)_-\,.$$

Keeping Z constant, we get y from the linear system (3), while keeping y constant, we get Z from (5). Note that here we alternate solving a system of order m, and a projection onto the cone S_+ of SDP matrices. Using (1) and (5), the update on X is given by $X + \sigma(Z + C - A^T(y)) = -\sigma W(y)_- = V \succeq 0$.

These updates for X and Z motivate us to call this a boundary-point method, as both Z and X are on the boundary of the cone of semidefinite matrices. Moreover $ZX = 0$ holds throughout. Hence, once feasibility with respect to the primal and dual linear equations is reached, we have an optimal solution.

Finally, the matrix variables X and Z can be eliminated, since they can be defined through y. The problem is really to minimize over $y \in \mathbb{R}^m$ the function

$h(y) := b^T y + \frac{\sigma}{2}\|W(y)_-\|^2$; it is convex differentiable with Lipschitz continuous gradient given by $\nabla h(y) = b + \sigma A(W(y)_-)$. We could also minimize $h(y)$ directly, using for instance the method of Nesterov [6].

4. Application to Compute $\vartheta(G)$

The theta number $\vartheta(G)$, associated to a graph $G = (V, E)$, is the optimal value of the following SDP; see for instance [3]:

$$\vartheta(G) = \max\langle J, X\rangle\,, \quad \text{such that} \quad x_{ij} = 0\,\forall [ij] \notin E\,,\ \mathrm{tr}(X) = 1\,,\ X \succeq 0\,.$$

Let $n = |V|$ and $m = \binom{n}{2} - |E|$, that is, m denotes the number of equations $x_{ij} = 0$. We consider graphs with edge density 0.5, which are the hardest for standard methods because $m \approx |E(G)|$. We solve the resulting SDP to about six digits of relative accuracy and get the following computational results on a laptop (1.7 Mhz, 1 GB RAM, Matlab under Linux). The computation times are essentially determined by the number of eigenvalue decompositions. There are typically less than 1 000 of these. Current research focuses on further applications of this approach and will be reported elsewhere.

n	200	300	400	500	600
m	10000	22500	40000	62500	90100
time (sec.)	120	350	880	1700	3060

References

[1] D. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic Press, New York, 1982.

[2] S. Burer and D. Vandenbussche. Solving lift-and-project relaxations of binary integer programs. Technical report, 2004.

[3] L. Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25:1–7, 1979.

[4] J. Malick. A dual approach to semidefinite least-squares problems. *SIAM J. Matrix Analysis and Applications*, 26:272.284, 2004.

[5] J. Malick. Proximal algorithms for semidefinite programming. Technical report, Inria Rhône-Alpes, in preparation.

[6] Y. Nwesterov. *Introductory Lectures on Convex Optimization: a Basic Course*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

A Full-Newton Step $O(n)$ Infeasible Interior-Point Algorithm for Linear Optimization

Kees Roos

Interior-point methods (IPMs) for solving linear optimization (LO) problems were initiated by Karmarkar [6]. They not only have polynomial complexity but are also highly efficient in practice. One may distinguish between *feasible* IPMs and *infeasible* IPMs (IIPMs). Feasible IPMs start with a strictly feasible interior point and maintain feasibility during the solution process. An elegant and theoretically sound method to find a strictly feasible starting point is to use a homogeneous embedding model, by introducing artificial variables. This technique was presented

first by Ye et al. [29] and studied by many others [1, 4, 5, 8, 11, 12, 14, 15, 16, 18, 21, 24, 25, 27, 29, 30]. Only a few commercial software packages are based on this approach: to our knowledge, only MOSEK [2], SeDuMi [19], and Xpress$^{\mathrm{MP}}$.

Most of the existing software packages use an IIPM. IIPMs start with an arbitrary positive point, and feasibility is reached as optimality is approached. The first IIPMs were proposed by Lustig [9] and Tanabe [20]. Global convergence was shown by Kojima et al. [7], whereas Zhang [30] proved an $O(n^2L)$ iteration bound for IIPMs under certain conditions. Mizuno [10] introduced a primal-dual IIPM and proved global convergence of the algorithm. A discussion and analysis of IIPMs can be found in the book by Wright [26] and, with less detail, in the books by Ye [28] and Vanderbei [22]. The performance of existing IIPMs highly depends on the choice of the starting point, which makes these methods less robust than the methods using the homogeneous embedding technique.

As usual, we consider the linear optimization (LO) problem in the standard form

$$\text{(P)} \qquad \min\left\{c^T x : Ax = b, \quad x \geq 0\right\},$$

with its dual problem

$$\text{(D)} \qquad \max\left\{b^T y : A^T y + s = c, \quad s \geq 0\right\}.$$

Here $A \in \mathbb{R}^{m\times n}$; $b, y \in \mathbb{R}^m$ and $c, x, s \in \mathbb{R}^n$. Without loss of generality we assume that $\operatorname{rank}(A) = m$. The vectors x, y and s are the vectors of variables. The best-known iteration bound for IIPMs is

$$\text{(1)} \qquad O\left(n \log \frac{\max\left\{\left(x^0\right)^T s^0,\ \left\|b - Ax^0\right\|,\ \left\|c - A^T y^0 - s^0\right\|\right\}}{\varepsilon}\right).$$

Here $(x^0, y^0, s^0) > 0$ denotes the starting point; $b - Ax^0$ and $c - A^T y^0 - s^0$ are the initial primal and dual residue vectors, respectively, whereas ε is an upper bound for the duality gap and the norms of residual vectors upon termination of the algorithm. It is assumeed in this result that an optimal solution (x^*, y^*, s^*) exists with $\|(x^*, s^*)\| \leq \zeta$ and that the initial iterate is $(x^0, y^0, s^0) = \zeta(e, 0, e)$.

Until 2003, the search directions used in all primal-dual IIPMs were computed from the linear system

$$\begin{aligned} A\Delta x &= b - Ax \\ A^T\Delta y + \Delta s &= c - A^T y - s \\ s\Delta x + x\Delta s &= \mu e - xs, \end{aligned}$$

which yields tho so-called primal-dual Newton search directions Δx, Δy, and Δs.

To describe the idea underlying our algorithm, we make some remarks with a historical flavor. In feasible IPMs, feasibility of the iterates is given, and the ultimate goal is to get iterates that are optimal. There is a well-known IPM that aims to reach optimality in one step, namely, the affine-scaling method. But everybody familiar with IPMs knows that this does not yield a polynomial-time

method. The past two decades have made it very clear that, to get a polynomial-time method, one should be less greedy and work with a search direction that moves the iterates only slowly in the direction of optimality. Only then can one take full profit of the efficiency of Newton's method, which is the working horse in all IPMs.

In IIPMs, the iterates are not feasible; and, apart from reaching optimality, one needs to strive for feasibility. This is reflected by the choice of the search direction, as defined above: when moving from x to $x^+ := x + \Delta x$, the new iterate x^+ satisfies the primal feasibility constraints, except possibly the nonnegativity constraint. In fact, in general x^+ will have negative components; and, to keep the iterates positive, one is forced to take a damped step of the form $x^+ := x + \alpha\Delta x$, where $\alpha < 1$ denotes the stepsize. But this same phenomenon occurred with the affine-scaling method in feasible IPMs. There the best complexity results clearly hold for methods that are much less greedy and that use full Newton steps (with $\alpha = 1$). Striving to reach feasibility in one step might be too optimistic and may deteriorate the overall behavior of a method. One may better exercise a little patience and move more slowly in the direction of feasibility. Therefore, in our approach, the search directions are designed in such a way that a full Newton step reduces the sizes of the residual vectors with the same speed as the duality gap. The outcome of our analysis confirms that this is a good strategy. It yields a full-Newton step method with the same complexity as given by (1). We conjecture, however, that a more careful analysis will reduce this bound by a factor of $\sqrt{n}$.

References

[1] E. D. Andersen and Y. Ye. A computational study of the homogeneous algorithm for large-scale convex optimization. *Computational Applications and Optimization*, 10:243–269, 1998.

[2] E. D. Andersen and K. D. Andersen. The MOSEK interior-point optimizer for linear programming: An implementation of the homogeneous algorithm. In S. Zhang, H. Frenk, C. Roos, and T. Terlaky, eds., *High Performence Optimization Techniques*, pages 197–232. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.

[3] Y.Q. Bai, M. El Ghami, and C. Roos. A comparative study of kernel functions for primal-dual interior-point algorithms in linear optimization. *SIAM J. Opt.*, 15(1):101–128 (electronic), 2004.

[4] J. F. Bonnans and F. A. Potra. On the convergence of the iteration sequence of infeasible path following algorithms for linear complementarity problems. *Math. O.R.*, 22(2):378–407, 1997.

[5] R. M. Freund. A potential-function reduction algorithm for solving a linear program directly from an infeasible 'warm start'. *Math. Programming*, 52:441–466, 1991.

[6] N. Karmarkar. New polynomial-time algorithm for linear programming. *Combinattorica 4*, pages 373–395, 1984.

[7] M. Kojima, N. Megiddo, and S. Mizuno. A primal-dual infeasible-interior-point algorithm for linear programming. *Math. Programming 61*, pages 263–280, 1993.

[8] M. Kojima, T. Noma, and A. Yoshise. Global convergence in infeasible-interior-point algorithms. *Math. Programming, Series A*, 65:43–72, 1994.

[9] I. J. Lustig. Feasible issues in a primal-dual interior-point method. *Math. Programming 67*, pages 145–162, 1990/91.

[10] S. Mizuno. Polynomiality of infeasible-interior-point algorithms for linear programming. *Math. Programming 67*, pages 109–119, 1994.

[11] S. Mizuno, M. J. Todd, and Y. Ye. A surface of analytic centers and infeasible- interior-point algorithms for linear programming. *Math. O.R.*, 20:135–162, 1995.
[12] Yu. Nesterov, M. J. Todd, and Y. Ye. Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems. *Math. Programming*, 84(2, Ser. A):227–267, 1999.
[13] J. Peng, C. Roos, and T. Terlaky. *Self-Regularity. A New Paradigm for Primal-Dual Interior-Point Algorithms*. Princeton University Press, Princeton, New Jersey, 2002.
[14] F. A. Potra. A quadratically convergent predictor-corrector method for solving linear programs from infeasible starting points. *Math. Programming*, 67(3, Ser. A):383–406, 1994.
[15] F. A. Potra. An infeasible-interior-point predictor-corrector algorithm for linear programming. *SIAM J. Opt.*, 6(1):19–32, 1996.
[16] C. Roos, T. Terlaky, and J.-Ph. Vial. *Theory and Algorithms for Linear Optimization. An Interior-Point Approach*. John Wiley & Sons, Chichester, UK, 1997.
[17] M. Salahi, T. Terlaky, and G. Zhang. The complexity of self-regular proximity based infeasible IPMs. Technical Report 2003/3, Advanced Optimization Laboratory, McMaster University, Hamilton, Ontario, Canada, 2003.
[18] R. Sheng and F. A. Potra. A quadratically convergent infeasible-interior-point algorithm for LCP with polynomial complexity. *SIAM J. Opt.*, 7(2):304–317, 1997.
[19] J. F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods & Software*, 11:625–653, 1999.
[20] K. Tanabe. Centered Newton method for linear programming: Interior and "exterior" point method (in Janpanese). In K. Tone, ed., *New Methods for Linear Programming,* volume 3, pages 98–100, The Institute of Statistical Mathematics, Tokyo, 1990.
[21] M. J. Todd and Y. Ye. Approximate Farkas lemmas and stopping rules for iterative infeasible-point algorithms for linear programming. *Math. Programming*, 81(1, Ser. A):1–21, 1998.
[22] R. J. Vanderbei. *Linear Programming: Foundations and Extensions*. Kluwer Academic Publishers, Boston, Massachusetts, 1996.
[23] R. S. Westfall. *Never at Rest: A Biography of Isaac Newton*. Cambridge University Press, Cambridge, UK, 1964.
[24] S. J. Wright. An infeasible-interior-point algorithm for linear complementarity problems. *Math. Programming*, 67(1):29–52, 1994.
[25] S. J. Wright. A path-following infeasible-interior-point algorithm for linear complementarity problems. *Optimization Methods & Software*, 2:79–106, 1993.
[26] S.J. Wright. *Primal-Dual Interior-Point Methods*. SIAM, Philadelphia, Pennsylvania, 1996.
[27] F. Wu, S. Wu, and Y. Ye. On quadratic convergence of the $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. *Annals of Operations Research*, 87:393–406, 1999.
[28] Y. Ye. *Interior Point Algorithms, Theory and Analysis*. John Wiley & Sons, Chichester, UK, 1997.
[29] Y. Ye, M.J. Todd, and S. Mizuno. An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. *Math. O.R.*, 19:53–67, 1994.
[30] Y. Zhang. On the convergence of a class of infeasible-interior-point methods for the horizontal linear complementarity problem. *SIAM J. Opt.*, 4:208–227, 1994.

On Recursive Multiscale Trust-Region Algorithms for Unconstrained Minimization

Annick Sartenaer and Philippe L. Toint
(joint work with Serge Gratton)

Many large-scale finite-dimensional minimization programs arise from the discretization of infinite-dimensional problems, such as optimal-control problems defined in terms of either ordinary or partial differential equations. We report here on a potentially efficient new class of algorithms using this structure and briefly discuss a first set of numerical experiments.

A simple first approach is to use coarser grids in order to compute approximate solutions, which can then be used as starting points for the optimization problem on a finer grid (see [10, 3, 5, 4], for instance). However, potentially more efficient techniques are inspired from the multigrid paradigm in the solution of partial differential equations and associated systems of linear algebraic equations (see, for example, [6, 7]). The work presented here was in particular motivated by the "generalized truncated Newton algorithm" presented in Fisher [9], a talk by Moré [13], the contributions by Lewis and Nash [11, 12] and the computational success of the low/high-fidelity model management techniques of Alexandrov, Lewis, and coauthors [1, 2].

We consider the solution of the unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1}$$

where f is a twice-continuously differentiable objective function which maps $\mathbb{R}^n$ into $\mathbb{R}$ and is bounded below. The trust-region methods which we study produce, given an initial point x_0, a sequence $\{x_k\}$ of iterates (one hopes) converging to a local first-order critical point for the problem. At each iterate x_k, these methods build a (typically quadratic) model $m_k(x_k + s)$ of $f(x_k + s)$. This model is then assumed to be adequate in a "trust region", defined as a sphere of radius $\Delta_k > 0$ centered at x_k, and a step s_k is then computed that sufficiently reduces this model in the region. The objective function is computed at the trial point $x_k + s_k$ and this trial point is accepted as the next iterate if and only if the achieved reduction in f is sufficiently large compared to the predicted reduction in m_k. The value of the radius is then updated to ensure that it is decreased when the trial point cannot be accepted as the next iterate, and is increased otherwise. Obtaining sufficient decrease on this model then amounts to (approximately) solving the problem $\min_{\|s\|\leq\Delta_k} m_k(x_k + s)$.

We investigate what can be done to reduce the cost of solving (1) if one attempts to exploit the knowledge of simplified expressions of the objective function, when available. More specifically, we assume that we know a collection of functions $\{f_i\}_{i=0}^r$ such that each f_i is a twice-continuously differentiable function from $\mathbb{R}^{n_i}$ to $\mathbb{R}$ (with $n_i \geq n_{i-1}$), the connection with our original problem being that $n_r = n$ and $f_r(x) = f(x)$ for all $x \in \mathbb{R}^n$. We will also assume that, for each $i = 1, \ldots, r$, f_i is "more costly" to minimize than f_{i-1}. This may be because f_i has more variables

than f_{i-1} (as would typically be the case if the f_i represent increasingly finer discretizations of the same infinite-dimensional objective) or because the structure (in terms of partial separability, sparsity, or eigenstructure) of f_i is more complex than that of f_{i-1}, or for any other reason. To fix terminology, we will refer to a particular i as a *level*. Of course, for f_{i-1} to be useful at all in minimizing f_i, there should be some relation between the variables of these two functions. We henceforth assume that, for each $i = 1, \ldots, r$, there exist a full-rank linear operator R_i from $\mathbb{R}^{n_i}$ into $\mathbb{R}^{n_{i-1}}$ and another full-rank operator P_i from $\mathbb{R}^{n_{i-1}}$ into $\mathbb{R}^{n_i}$ such that $P_i = R_i^T$, where P_i and R_i are interpreted as restriction and prolongation between a fine and a coarse grid. The idea is then to use f_{r-1} to construct an alternative model h_{r-1} for $f_r = f$ in the neighborhood of the current iterate that is cheaper than m_k and to use this alternative model to define the step in the trust-region algorithm whenever possible. If more than two levels are available ($r > 1$), this procedure can be done recursively, the approximation process stopping at level 0, where the usual quadratic model is always used. The resulting algorithm can then be specified as a variant of the basic trust-region algorithm of [8].

We briefly describe the global convergence theory associated with this algorithm and show convergence from arbitrary starting points to first-order critical points under classical assumptions. We also discuss an associated dimension-independent worst-case complexity result. We next present a first numerical application for one of the possible implementations. This implementation specifies the nature of the nonrecursive iterations, which fall into two classes: smoothing iterations, aimed at decreasing high-frequency components of the gradient, and damping iterations, which decrease their low-frequency components (an important issue is to modify these iterations so as to ensure "sufficient decrease" in the sense of the Cauchy condition). Other implementation issues concern the form of the recursive iterations, ranging from free form (where the optimization at lower levels is governed purely by accuracy requirements) to fixed cycles (such as the V and W cycles inspired by multigrid techniques). The efficiency of the method is demonstrated on a minimum-surface problem with highly oscillatory boundary conditions. Problems of this type involving up to 1.1 million variables were solved by the new algorithm in Matlab on an oldish laptop PC (Pentium 4 Mobile, 1.6 GHz). Perspectives are described, which are both numerous and interesting.

References

[1] N. M. Alexandrov and R. L. Lewis. An overview of first-order model management for engineering optimization. *Optimization and Engineering*, 2:413–430, 2001.

[2] N. M. Alexandrov, R. L. Lewis, C. R. Gumbert, L. L. Green, and P. A. Newman. Approximation and model management in aerodynamic optimization with variable fidelity models. *J. Aircraft*, 38(6):1093–1101, 2001.

[3] R. E. Bank, P. E. Gill, and R. F. Marcia. Interior point methods for a class of elliptic variational inequalities. In T. Biegler, O. Ghattas, M. Heinkenschloss, and B. Van Bloemen Waanders, eds., *High Performance Algorithms and Software for Nonlinear Optimization*, pages 218–235, Springer Verlag, Heidelberg, 2003.

[4] S. J. Benson, L. C. McInnes, J. Moré, and J. Sarich. Scalable algorithms in optimization: Computational experiments. Preprint ANL/MCS-P1175-0604, Mathematics and Computer Science, Argonne National Laboratory, Argonne, Illinois, 2004. To appear in the *Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization (MA&O) Conference,* August 30-September 1, 2004.
[5] J. T. Betts and S. O. Erb. Optimal low thrust trajectory to the moon. *SIAM J. Applied Dynamical Systems*, 2(2):144–170, 2003.
[6] A. Brandt. Multi-level adaptative solutions to boundary value problems. *Math. Computation*, 31(138):333–390, 1977.
[7] W. L. Briggs, V. E. Henson, and S. F. McCormick. *A Multigrid Tutorial*, 2nd ed. SIAM, Philadelphia, Pennsylvania, 2000.
[8] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods*. Number 01 in MPS-SIAM Series on Optimization. SIAM, Philadelphia, Pennsylvania, 2000.
[9] M. Fisher. Minimization algorithms for variational data assimilation. In *Recent Developments in Numerical Methods for Atmospheric Modelling*, pages 364–385. ECMWF, 1998.
[10] A. Griewank and Ph. L. Toint. Local convergence analysis for partitioned quasi-Newton updates. *Num. Mathematik*, 39:429–448, 1982.
[11] M. Lewis and S. G. Nash. Practical aspects of multiscale optimization methods for VLSI-CAD. In Jason Cong and Joseph R. Shinnerl, eds., *Multiscale Optimization and VLSI/CAD*, pages 265–291. Kluwer Academic Publ., Dordrecht, The Netherlands, 2002.
[12] M. Lewis and S. G. Nash. Model problems for the multigrid optimization of systems governed by differential equations. *SIAM J. Sci. Comp.*, to appear, 2005.
[13] J. J. Moré. Terascale optimal PDE solvers. Talk at the ICIAM 2003 Conference, Sydney, Australia, 2003.

Complexity Analysis of Interior-Point Algorithms and Geometric Properties of the Central Trajectory for Linear Programming

Takashi Tsuchiya
(joint work with Renato D. C. Monteiro)

We study the geometrical structure of the central trajectory associated with a linear programming (LP) problem using the same techniques developed in [2] to establish a new complexity bound for the Mizuno-Todd-Ye predictor-corrector (MTY-PC) primal-dual interior-point algorithm.

We consider the LP problem

$$\text{minimize}_x \; c^T x \;\; \text{subject to } Ax = b,\; x \geq 0, \tag{1}$$

and its associated dual problem

$$\text{maximize}_{(y,s)} \; b^T y \;\; \text{subject to } A^T y + s = c,\; s \geq 0, \tag{2}$$

where $A \in \mathbb{R}^{m\times n}$, $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$ are given, and the vectors $x, s \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ are the unknown variables. We assume that both (1) and (2) have interior feasible solutions.

The MTY-PC algorithm solves (1) and (2) by closely following the central trajectory, which is defined as the set of solutions of the following family of systems of bilinear equations in (x, s, y), parametrized by $\nu > 0$:

$$x \circ s = \nu \mathbf{1},\; Ax = b,\; A^T y + s = c,\; x > 0,\; s > 0. \tag{3}$$

Here $\circ$ denotes the Hadamard product of vectors, that is, componentwise product of two vectors. The vector of all ones is denoted by $\mathbf{1}$. Given $\beta \in (0,1)$, we define the 2-norm neighborhood of the central trajectory with opening β as

$$N(\beta) \equiv \left\{ (x,y,s) \text{ feasible solution} : \left\| \frac{x \circ s}{\mu(x,s)} - e \right\| \leq \beta \right\},$$

where $\mu(x,s) = x^T s/n$.

We now describe the MTY-PC algorithm with opening $\beta \in [0, 1/2]$. One iteration of the algorithm consists of a predictor step followed by a corrector step. At the beginning of an iteration, an iterate lying in the smaller neighborhood $N(\beta^2)$ is assumed given. A predictor step is then performed, which consists of moving along the Newton direction for system (3) with $\nu = 0$ until a point lying in the boundary of the larger neighborhood $N(\beta)$ is located. Next, a corrector step from the latter point is performed to obtain the next iterate of the MTY-PC algorithm lying in the smaller neighborhood $N(\beta^2)$. This corrector step consists of taking a full Newton step toward the central trajectory point with the same duality gap as the base point. Classical results about the MTY-PC algorithm establish that it has the iteration-complexity bound $O(\sqrt{n}\log(1/\varepsilon))$ for reducing the duality gap by a factor of ε, as well as asymptotic quadratic convergence.

Set $\bar{\chi}_A = \max\{\|G^{-1}A\| : G \in \widetilde{G}\}$, where $\widetilde{G}$ denotes the set of all $m \times m$ nonsingular submatrices of A. This quantity is known as *the condition number* of A and its properties have been studied by several authors starting from Dikin [1] (see references in [4, 6] for more details). It is known that $\bar{\chi}_A$ is bounded by 2^{L_A} if A is integral, where L_A is the input size of A. Furthermore, let

$$\bar{\chi}_A^* \equiv \inf\{\bar{\chi}_{AD} : D \text{ is a positive definite diagonal matrix}\}.$$

We established in [4] the following complexity result for the MTY-PC algorithm.

Theorem 1 *Given $0 < \varepsilon \leq 1$ and an initial point $(x^0, s^0, y^0) \in N(\beta^2)$ with $\beta \in (0, 1/4]$, the MTY-PC algorithm generates an iterate $(x^k, s^k, y^k) \in N(\beta^2)$ satisfying $\mu(x^k, s^k) \leq \varepsilon\mu(x^0, s^0)$ in $O\left(\min[n^2 \log\log(1/\varepsilon), \log(1/\varepsilon)] + n^{3.5}\log(\bar{\chi}_A^* + n)\right)$ iterations.*

This result was strongly motivated by Vavasis and Ye's seminal work [6], where a polynomial-time interior-point algorithm with an iteration-complexity bound of $O(n^{3.5}\log(\bar{\chi}_A + n))$ is developed (see also [3]). Their analysis is based on the clever notion of a crossover event, and their algorithm occasionally uses a special step, namely, the layered least squares (LLS) step, in order to make substantial progress along the straight parts of the central trajectory. Crossover events and LLS steps are also used in a crucial way in the proof of the above theorem and the subsequent results about the geometrical structure of the central trajectory.

Vavasis and Ye [6] made an interesting observation that *the central trajectory consists of $O(n^2)$ "long and straight" parts and that its remaining curved part is relatively short*, in that it can be traversed by a standard short-step path-following algorithm in $O(n^{3.5}\log(\bar{\chi}_A + n))$ iterations. This observation was essentially motivated by the technique used to establish the iteration-complexity of their method and was not formally justified, in that the notion of straightness of the central

trajectory was *not* defined rigorously. The purpose of this work is to relate the new complexity analysis of the interior-point algorithms developed in [3, 4, 6] with the geometrical structure of the central trajectory in a more concrete way.

To this end, the following integral over the central trajectory plays a fundamental role:

$$F(\nu_0, \nu_1) = \int_{\nu_1}^{\nu_0} \frac{\kappa(\nu)}{\nu}\, d\nu,$$

where $\kappa(\nu) = \|\nu\dot{x}(\nu)\dot{s}(\nu)\|^{1/2}$. This integral was first introduced by Sonnevend, Stoer and Zhao [5] to analyze the iteration-complexity of a predictor-corrector type path-following algorithm (see also [7]). We refer to the function $\kappa(\cdot)$ as the "curvature" of the central trajectory. The following result gives an estimation of the number of iterations of the MTY-PC algorithm in terms of the above integral.

Proposition 2 *For given $0 < \mu_1 < \mu_0$, denote by $\#(\mu_0, \mu_1, \beta)$ the number of iterations of the MTY-PC algorithm with opening β needed to reduce the duality gap from μ_0 to μ_1. Then,*

$$\lim_{\gamma \to 0} \frac{F(\mu_0, \mu_1)/\sqrt{\beta}}{\#(\mu_0, \mu_1, \beta)} = 1.$$

One of the interpretations of the above formula is that, as the opening β becomes sufficiently small, the curvature $\kappa(\nu)$ divided by $\beta^{1/2}$ provides an estimate on the number of iterations needed to reduce the current duality gap ν by a factor of e^{-1}. Thus, κ is a reasonable measure of straightness of the central trajectory. The classical complexity analysis yields $\kappa \leq \sqrt{n}$, and therefore we have $F(\mu, \varepsilon\mu) = O(\sqrt{n}\log(1/\varepsilon))$ as a trivial bound, where $0 < \varepsilon \leq 1$.

The main results we have obtained about the behavior of the curvature and the integral are summarized in the following two theorems.

Theorem 3 *For any constant $\bar{\kappa} \in (0, \sqrt{n}]$, there exists $l \leq n(n-1)/2$ closed intervals $I_k = [d_k, e_k], k = 1, ..., l$, such that:*

i) $d_k \geq e_{k+1}$ for all $k = 1, ..., l-1$;

ii) $\{\nu > 0 : \kappa(\nu) \geq \bar{\kappa}\} \subseteq \cup_{k=1}^{l} I_k$;

iii) $\log(e_k/d_k) = O(n\log(\bar{\chi}_A^* + n) + n\log\bar{\kappa}^{-1})$ for all $k = 1, ..., l$.

Theorem 4 *Let $\nu > 0$ and $0 < \varepsilon \leq 1$. Then, we have*

$$F(\nu, \varepsilon\nu) = O(n^{3.5}\log(\bar{\chi}_A^* + n) + \min[n^{3.5}\log\log(1/\varepsilon), \log(1/\varepsilon)]).$$

A few remarks are now in order. Theorem 3 bounds the length (in the logarithmic scale) of the curved part of the central trajectory. The intervals I_k $(i = 1, ..., l)$ contain the parameters ν with large curvature (i.e., such that $\kappa(\nu) \geq \bar{\kappa}$) and the theorem claims that the total length of all the I_k's is reasonably bounded. Theorem 4 gives an upper bound on the integral F. The bound consists of two terms: the first one is independent of ε, and the second one grows very slowly as a function of ε or independent of n.

In view of existing results about the limiting behavior of the central trajectory, it is not difficult to see that $F(\infty, 0)$ is well defined as an improper integral and is bounded by a constant depending on (A, b, c). Therefore, an interesting question is

whether there is an upper bound on $F(\infty, 0)$ depending only on A. We conjecture that $F(\infty, 0) = O(n^{3.5} \log(\bar{\chi}_A^* + n))$.

References

[1] I. I. Dikin. On the speed of an iterative process. *Upravlyaemye Sistemi*, 12:54–60, 1974.

[2] S. Mizuno, M. J. Todd, and Y. Ye. On adaptive-step primal-dual interior-point algorithms for linear programming. *Math. O.R.*, 18:964–981, 1993.

[3] R. D. C. Monteiro and T. Tsuchiya. A variant of the Vavasis-Ye layered-step interior-point algorithm for linear programming. *SIAM J. Opt.*, 13:1054–1079, 2003.

[4] R. D. C. Monteiro and T. Tsuchiya. New iteration-complexity bound for the MTY predictor-corrector algorithm. *SIAM J. Opt.*, 15:319–347, 2005.

[5] G. Sonnevend, J. Stoer, and G. Zhao. On the complexity of following the central path of linear programs by linear extrapolation II, *Math. Programming*, 52:527–553, 1951.

[6] S. Vavasis and Y. Ye. A primal-dual accelerated interior point method whose running time depends only on A. *Math. Programming*, 74:79–120, 1996.

[7] G. Zhao and J. Stoer. Estimating the complexity of path-following methods for solving linear programs by curvature integrals. *Applied Mathematics and Optimization*, 27:86–103, 1993.

Newton-Type Preconditioned Multilevel Methods for Infinite-Dimensional Complementarity Problems with Applications

Michael Ulbrich

In this study, we give an overview of recent results and work in progress on Newton-type methods for the following infinite-dimensional *Mixed Complementarity Problem* (MCP):

$$\begin{aligned} &\text{(a)} && G(y,u) = 0 \\ &\text{(b)} && u \in \mathcal{K},\ F(y,u) \in \mathcal{K}^+,\ \langle F(y,u), u\rangle_{U^*,U} = 0. \end{aligned} \tag{1}$$

Here $G : Y \times U \to Z$ and $F : Y \times U \to V$ are continuously Fréchet differentiable operators; Y, Z are Banach spaces; and $U \subset \mathcal{M}(\Omega)$ is a function space, $\mathcal{M}(\Omega)$ denoting the space of real-valued measurable functions on the bounded measurable set $\Omega \subset \mathbb{R}^n$. Furthermore, $V \subset U^*$ is a (generalized) function space, where U^* is the dual space of U, $\mathcal{K} \subset U$ is the cone of a.e. nonnegative functions in U, and $\mathcal{K}^+ \subset U^*$ denotes the dual cone of $\mathcal{K}$.

When $V \subset \mathcal{M}(\Omega)$, the complementarity condition (1) (b) can be written pointwise as

$$u \geq 0, \quad F(u) \geq 0, \quad uF(u) = 0 \quad \text{a.e. on } \Omega.$$

We can further transform this problem into the operator equation

$$H(y,u) \stackrel{\text{def}}{=} \begin{pmatrix} G(y,u) \\ \Phi(y,u) \end{pmatrix} = 0, \tag{2}$$

where $\Phi(y,u)(\cdot) \stackrel{\text{def}}{=} \phi(u(\cdot), F(y,u)(\cdot))$. Here, $\phi : \mathbb{R}^2 \to \mathbb{R}$ is an NCP-function, that is, satisfies $\phi(a,b) = 0$ iff $a, b \geq 0$ and $ab = 0$. Commonly used NCP-functions, for example, $\phi_{\min}(a,b) = \min(a,b)$, are usually nonsmooth, but Lipschitz continuous and semismooth. As a consequence, the operator Φ is in general nonsmooth and this is inherited by H.

For simplicity, we now focus on an important special case of (1), the *nonlinear complementarity problem* (NCP). Assuming again $V \subset \mathcal{M}(\Omega)$, one can state the NCP as follows:

$$u \geq 0, \quad F(u) \geq 0, \quad uF(u) = 0 \quad \text{a.e. on } \Omega. \tag{3}$$

The reformulated problem (2) then reduces to the nonsmooth equation

$$\Phi(u) = 0, \quad \text{with} \quad \Phi(u)(\cdot) := \phi(u(\cdot), F(u)(\cdot)). \tag{4}$$

We present a rigorous analysis for the case where $U = L^p(\Omega)$, $V = L^q(\Omega)$, $p \in [2, \infty]$, $q \in [1, p]$, $1/p + 1/q \leq 1$. Under appropriate assumptions (which are mild if $p > q$ and require a particular structure of ϕ and F if $p = q$), we then can show that $\Phi : U \to V$ is $\partial\Phi$-semismooth [2] (this concept is introduced below), that is,

$$\sup_{M \in \partial\Phi(u+s)} \|\Phi(u+s) - \Phi(u) - Ms\|_V = o(\|s\|_U) \quad (\|s\|_U \to 0),$$

where $\partial\Phi(u) \subset \mathcal{L}(U, V)$ is a suitably chosen generalized differential. Under slightly stronger assumptions, we can prove semismoothness of order $\kappa > 0$. Here, we propose to call a continuous operator $H : W_1 \to W_2$ between Banach spaces *∂H-semismooth* at w if the following holds:

$$\sup_{M \in \partial H(w+s)} \|H(w+s) - H(w) - Ms\|_{W_2} = o(\|s\|_{W_1}) \quad (\|s\|_{W_1} \to 0).$$

The operator H is called ∂H-semismooth of order $\kappa > 0$ if $o(\|s\|_{W_1})$ can be strengthened to $O(\|s\|_{W_1}^{1+\kappa})$. We then can show that the Newton-type iteration

$$\textit{Given } w^k\textit{, choose } M_k \in \partial H(w^k)\textit{, solve } M_k s^k = -H(w^k)\textit{, set } w^{k+1} = w^k + s^k$$

converges locally q-superlinearly (with order $1 + \kappa$ if H is semismooth of order $\kappa > 0$) to a solution $w^* \in W_1$ if all the operators M_k are uniformly bounded invertible (we call *regularity condition* this property and related ones).

With regard to the specific operator Φ, the structure of its associated generalized differential implies that the regularity condition usually can be verified only in the case $U = V$. Fortunately, this choice of spaces is possible if F is continuously Fréchet differentiable and has the structure $F(u) = \gamma u + A(u)$, where $A : U \to L^r(\Omega)$ is locally Lipschitz continuous for some $r > p$ and if we make the particular choice $\phi(a, b) = \min(\gamma a, b)$ for the NCP-function. Alternatively, we can use a more general class of NCP-functions, obtain semismoothness of Φ with a norm gap (i.e., $p > q$), and then have to augment the Newton iteration by a smoothing step [2]. Here, however, we follow the first approach.

The following theoretical questions are addressed:

– Sufficient conditions for regularity [3, 4], that is, conditions implying

$$\left\|M^{-1}\right\|_{V,U} \leq C \quad \forall\, M \in \partial\Phi(u),\ \|u - u^*\|_U < \varepsilon.$$

Two variants are given for the case $U = V = L^2(\Omega)$, where the main assumption is a coercivity condition for $F'(u^*)$ either on L^2 or on an appropriate tangent space.

– A mesh independence results [1]. Here the NCP (3) and piecewise constant discretizations $(3)_h$, $h > 0$ denoting the mesh size, are considered. The corresponding reformulations (4) and $(4)_h$, respectively, induce semismooth Newton iterations that generate sequences u^k and u_h^k, respectively. Under very natural assumptions, we show that, for any given rate θ, there exists a universal radius $\delta_\theta > 0$ and a mesh size $h_\theta > 0$ such that the Newton iterates (u^k) and (u_h^k), $h \le h_\theta$, converge with at least q-linear rate θ to a solution u^* and corresponding discrete solutions u_h^*, respectively, whenever the initial points lie within δ_θ-balls centered at u^* and u_h^*, respectively.
– Verification of the assumptions for concrete problems.

In our study, a strong focus is placed on applications in the field of optimization with PDEs and on efficient implementations of the proposed method for large-scale discretizations. In particular, we discuss control-constrained semilinear elliptic control problems and 3D elastic two-body contact problems. The first problem class directly fits in our theoretical framework, whereas for the contact problem a regularization procedure is required to make the theory rigorously applicable. Error estimates for the regularized solutions are also presented [5].

For both, the control problem and the contact problem we show that multigrid solvers for elliptic PDEs can be used to solve the linear operator equations arising in each iteration of the semismooth Newton method very efficiently. To this end, we observe that appropriate block eliminations in the Newton systems result in an elliptic system of PDEs to which multigrid methods can be applied. For instance, the semismooth Newton system for the (regularized) elastic contact problem can be reduced to a system involving the linear operator

$$\begin{pmatrix} E + \gamma^{-1} B^*(d \cdot B) & 0 \\ -d \cdot B & \gamma I \end{pmatrix},$$

where E is the elliptic differential operator of the elasticity equations, B is the normal trace operator on the potential contact boundary Γ_C, γ is the regularization parameter, and $d(x) \in \{0, 1\}$, $x \in \Gamma_C$. Since the upper left block is a system of elliptic differential operators, we can apply a multigrid method to solve the Newton system with optimal complexity. A similar approach can be taken for the elliptic optimal control problem. The multigrid method can either be used as a direct solver or, which is more efficient for harder problems (e.g., the elasticity problem), one or more multigrid cycles can serve as a preconditioner for Krylov subspace iterative solvers. Also, a nested iteration can be used to take additional advantage of the multilevel grid hierarchy. Starting on the coarsest grid, approximate solutions of the discretized optimization problem are transported to the next finer grid to obtain good initial points for the Newton process on this grid and so forth.

Numerical results for both problems are presented. They support the superlinear convergence results for the Newton iteration (less than 10 iterations are needed), its mesh-independent behavior, and they prove the efficiency of the multigrid Newton approach.

References

[1] M. Hintermüller and M. Ulbrich, *A mesh-independence result for semismooth Newton methods*, Math. Programming **101** (2004), 151–184.

[2] M. Ulbrich, *Semismooth Newton methods for operator equations in function spaces*, SIAM J. Optim. **13** (2003), 805–842.

[3] M. Ulbrich, *Nonsmooth Newton-like methods for variational inequalities and constrained optimization problems in function spaces*, Habilitation thesis, Fakultät für Mathematik, TU München, 2001.

[4] M. Ulbrich, *On a nonsmooth Newton method for nonlinear complementarity problems in function space with applications to optimal control*, in M. C. Ferris, O. L. Mangasarian, and J.-S. Pang (eds.), Complementarity: Applications, Algorithms and Extensions, Kluwer Academic Publishers, 2001, 341–360.

[5] M. Ulbrich and S. Ulbrich, *A multigrid semismooth Newton method for contact problems in linear elasticity*, Technical report in preparation, Fachbereich Mathematik, Universität Hamburg, 2005.

Primal-Dual Interior-Point Multigrid Methods for PDE-Constrained Optimization

Stefan Ulbrich

We present recent results from [1] on interior-point methods for PDE-constrained optimization problems of the form

$$\min_{(y,u)\in Y\times U} f(y,u) \quad \text{subject to } c(y,u)=0,\ u\in B:=\{v\in U:\ l\le u\le r\}$$

where $U = L^p(\Omega)$, $2 \le p < \infty$ is the control space, $\Omega \subset \mathbb{R}^d$ is bounded and measurable, Y is the state space and the bounds l, r lie in $L^\infty(\Omega)$, $r - l \ge \nu > 0$. The constraint $c(y, u) = 0$ is the appropriate formulation of a (system of) PDE(s) with boundary and/or initial conditions. We assume that $f : Y \times U \to \mathbb{R}$ and $c : Y \times U \to Z$ are twice locally Lipschitz-continuously differentiable, that there exists a unique solution $y = y(u) \in Y$ of $c(y, u) = 0$ for all $u \in B$ and that c_y has a bounded inverse.

As in the finite-dimensional case the basic concept of the proposed interior point method consists in following the central path by using damped Newton steps inside a neighborhood of the central path. However, the algorithmic details and the convergence analysis are strongly influenced by the infinite-dimensional nature of the problem and differ significantly from the finite-dimensional case.

Starting from the first order optimality conditions we introduce the central path as the solution path of the relaxed optimality conditions with perturbed complementarity condition. Although the associated logarithmic barrier function is not a barrier function in the classical sense, since C has no interior point in $L^s(\Omega)$ for $s < \infty$, the perturbed optimality conditions are nevertheless the optimality system for a corresponding barrier problem.

The proposed interior point algorithm follows the central path approximately by applying damped primal-dual Newton steps, which are projected on a wide neighborhood of the central path. The convergence of the algorithm is analyzed

for convex linear-quadratic problems that have the following smoothing property: the reduced gradient $\hat{g}(y,u)$ corresponding to the control component has the form $\hat{g}(y,u) = \alpha u + \hat{g}_s(y,u)$, where $\hat{g}_s$ is locally bounded in $L^q(\Omega)$ for some $q > p$ and the reduced Hessian $\hat{H}$ has the structure $\hat{H}(y,u) = \alpha I + \hat{H}_s(y,u)$, where $\hat{H}_s(y,u)$ is locally bounded in $\mathcal{L}(L^p(\Omega), L^q(\Omega))$. These structural assumptions are satisfied for several important applications, e.g., distributed and Neumann boundary control of elliptic and parabolic PDEs, distributed control of hyperbolic PDEs and the incompressible Navier-Stokes equations, and regularized contact problems.

In the case $q = \infty$ we prove global linear convergence. The case $q < \infty$ leads to a norm gap that is closed by using a smoothing step. We show that the resulting method converges globally linear und locally superlinear under a moderate strict complementarity assumption. An extension to nonlinear problems can be obtained in a vicinity of a solution satisfying appropriate regularity assumptions.

Moreover, we show how multigrid methods can be applied to solve the Newton system efficiently. We demonstrate this for control constrained elliptic problems and 3D elastic two body contact problems. Numerical results are shown for these problems which demonstrate the efficiency of the approach.

References

[1] S. Ulbrich, *Primal-dual interior point methods for PDE-constrained optimization*, Technical report in preparation, Fachbereich Mathematik, Technical University of Darmstadt, Germany, 2005.

Reporter: Claude Lemaréchal

Participants

Dr. Vincent Acary
vincent.acary@inrialpes.fr
INRIA Rhone-Alpes
655 avenue de l'Europe
F-38334 Montbonnot, St. Ismier, Cedex

Prof. Dr. Wolfgang Achtziger
Wolfgang.Achtziger@uni-dortmund.de
Fachbereich Mathematik
Universität Dortmund
44221 Dortmund

Prof. Dr. Natalia Alexandrov
n.alexandrov@nasa.gov
NASA Langley Research Center
Mail Stop 159
Hampton VA 23681-2199
USA

Prof. Dr. Kurt M. Anstreicher
kurt-anstreicher@uiowa.edu
Department of Management Sciences
The University of Iowa
Iowa City IA 52242-1000
USA

Dr. Amir Beck
becka@tx.technion.ac.il
Faculty of Industrial Engineering & Management
Technion
Israel Institute of Technology
Haifa 32000
ISRAEL

Prof. Dr. Aharon Ben-Tal
abental@ie.technion.ac.il
Faculty of Industrial Engineering & Management
Technion
Israel Institute of Technology
Haifa 32000
ISRAEL

Dr. Andreas Eisenblätter
eisenblaetter@zib.de
Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)
Takustr. 7
14195 Berlin

Prof. Dr. Marina Epelman
mepelman@umich.edu
Department of Industrial and Operations Engineering
University of Michigan, Room 2845
1205 Beal Avenue
Ann Arbor MI 48109-2117
USA

Prof. Dr. Michael Ferris
ferris@cs.wisc.edu
Computer Sciences Department
University of Wisconsin-Madison
1210 West Dayton St.
Madison, WI 53706-1685
USA

Prof. Dr. Roger Fletcher
fletcher@maths.dundee.ac.uk
Dept. of Mathematics
University of Dundee
23 Perth Road
GB-Dundee, DD1 4HN

Prof. Dr. Robert M. Freund
rfreund@mit.edu
MIT
Sloan School of Management
50 Memorial Dr.
Cambridge, MA 02142-1347
USA

Prof. Dr. Roland W. Freund
freund@math.ucdavis.edu
Department of Mathematics
University of California
1, Shields Avenue
Davis, CA 95616-8633
USA

Prof. Dr. Jean-Charles Gilbert
jean-charles.gilbert@inria.fr
INRIA
BP 105
F-78153 Rocquencourt

Dr. Martin Gugat
gugat@am.uni-erlangen.de
Institut für Angewandte Mathematik
Universität Erlangen
Martensstr. 3
91058 Erlangen

Prof. Dr. Margareta Halicka
halicka@fmph.uniba.sk
UAM FMFI
UK
Mlynska dolina
842 48 Bratislava
Slovakia

Prof. Dr. Matthias Heinkenschloss
heinken@caam.rice.edu
heinken@rice.edu
CAAM Dept. MS 134
Rice University
6100 S. Main Street
Houston, TX 77005-1892
USA

Prof. Dr. Christoph Helmberg
helmberg@mathematik.tu-chemnitz.de
Fakultät für Mathematik
Technische Universität Chemnitz
09107 Chemnitz

Prof. Dr. Florian Jarre
jarre@opt.uni-duesseldorf.de
Mathematisches Institut
Heinrich-Heine-Universität
Gebäude 25.22
Universitätsstraße 1
40225 Düsseldorf

Prof. Dr. Christian Kanzow
kanzow@mathematik.uni-wuerzburg.de
Institut für Angewandte Mathematik
und Statistik
Universität Würzburg
Am Hubland
97074 Würzburg

Prof. Dr. Diethard Klatte
klatte@ior.unizh.ch
Institut für Operations Research
Universität Zürich
Moussonstr. 15
CH-8044 Zürich

Dr. Michal Kocvara
kocvara@utia.cas.cz
Inst. of Info. Theory & Automation
Academy of Sciences of the Czech R.
Pod Vodarenskou vezi 4
P.O.Box 18
182 08 Prague 8
CZECH REPUBLIC

Dr. Friedemann Leibfritz
leibfr@uni-trier.de
Abteilung Mathematik
Fachbereich IV
Universität Trier
54286 Trier

Prof. Dr. Claude Lemarechal
claude.lemarechal@inria.fr
INRIA Rhone-Alpes
655 avenue de l'Europe
F-38334 Montbonnot, St. Ismier, Cedex

Prof. Dr. Adrian Lewis
aslewis@orie.cornell.edu
ORIE
Cornell University
Rhodes Hall 214
Ithaca, NY 14853
USA

Dr. Sven Leyffer
leyffer@mcs.anl.gov
Mathematics and Computer Science Division
Argonne National Laboratory
9700 South Cass Avenue
Argonne, IL 60439-4844
USA

Prof. Dr. Shinji Mizuno
mizuno@me.titech.ac.jp
Dep.of Industrial Engineering and Mangement
Tokyo Institute of Technology
2-12-1 Oh-okayama Meguro-ku
Tokyo 152-8552
JAPAN

Dr. Katja Mombaur
Katja.Mombaur@IWR.Uni-Heidelberg.de
Interdisziplinäres Zentrum für Wissenschaftliches Rechnen
Universität Heidelberg
Im Neuenheimer Feld 368
69120 Heidelberg

Prof. Dr. Renato D.C. Monteiro
renato.monteiro@isye.gatech.edu
School of Industrial and Systems Engineering
Georgia Institute of Technology
765 Ferst Drive
Atlanta GA 30332-0205
USA

Prof. Dr. Yurii Nesterov
nesterov@core.ucl.ac.be
Center for Operations Research and Econometrics
Universite de Louvain
34 Voie du Roman Pays
B-1348 Louvain-la-Neuve

Prof. Dr. Florian-A. Potra
potra@math.umbc.edu
Dept. of Mathematics and Statistics
University of Maryland
Baltimore County
1000 Hilltop Circle
Baltimore, MD 21250
USA

Prof. Dr. Franz Rendl
franz.rendl@uni-klu.ac.at
Institut für Mathematik
Universität Klagenfurt
Universitätsstr. 65-67
A-9020 Klagenfurt

Prof. Dr. James Renegar
renegar@mac.com
renegar@orie.cornell.edu
School of Operations Research and Industrial Engineering
Cornell University
Upson Hall
Ithaca, NY 14853-7901
USA

Prof. Dr. Kees Roos
c.roos@ewi.tudelft.nl
EWI/ISA/ALG
Delft University of Technology
Mekelweg 4
NL-2628 CD Delft

Prof. Annick Sartenaer
annick.sartenaer@fundp.ac.be
as@math.fundp.ac.be
Departement de Mathematique
Universite de Namur
(FUNDP)
Rempart de la Vierge 8
B-5000 Namur

Prof. Dr. Mike Saunders
saunders@stanford.edu
Management Science and Engineering
Terman Engineering Center
Room 330
Stanford University
Stanford CA 94305-4026
USA

Prof. Dr. Josef Stoer
jstoer@mathematik.uni-wuerzburg.de
Institut für Angewandte Mathematik
und Statistik
Universität Würzburg
Am Hubland
97074 Würzburg

Prof. Dr. Jean-Jacques Strodiot
jean-jacques.strodiot@fundp.ac.be
Departement de Mathematique
Universite de Namur
(FUNDP)
Rempart de la Vierge 8
B-5000 Namur

Prof. Dr. Michael J. Todd
miketodd@cs.cornell.edu
miketodd@orie.cornell.edu
School of Operations Research and
Industrial Engineering
Cornell University
Upson Hall
Ithaca, NY 14853-7901
USA

Prof. Dr. Philippe L. Toint
philippe.toint@fundp.ac.be
Department of Mathematics
The University of Namur
FUNDP
61, rue de Bruxelles
B-5000 Namur

Prof. Dr. Takashi Tsuchiya
tsuchiya@sun312.ism.ac.jp
The Institute of Stat. Mathematics
4-6-7 Minami Azabu, Minato-ku
Tokyo 106-8569
JAPAN

Prof. Dr. Reha Tütüncü
reha@cmu.edu
Department of Mathematical Sciences
Carnegie Mellon University
6113 Wean Hall
Pittsburgh PA 15213
USA

Prof. Dr. Michael Ulbrich
ulbrich@math.uni-hamburg.de
Fachbereich Mathematik
Universität Hamburg
20141 Hamburg

Prof. Dr. Stefan Ulbrich
ulbrich@mathematik.tu-darmstadt.de
Fachbereich Mathematik
TU Darmstadt
Schloßgartenstr. 7
64289 Darmstadt

Prof. Dr. Jean Philippe Vial
jean-philippe.vial@hec.unige.ch
HEC-Geneve
Section of Management Studies
University of Geneva
Bd du Pont d'Arve 40
CH-1211 Geneve 4

Prof. Dr. Stephen J. Wright
swright@cs.wisc.edu
Computer Sciences Department
University of Wisconsin-Madison
1210 West Dayton St.
Madison, WI 53706-1685
USA

Prof. Dr. Jochem Zowe
imi_imi@t-online.de
Laimbacher Str. 73
95447 Bayreuth