Lecture Notes in Mathematics

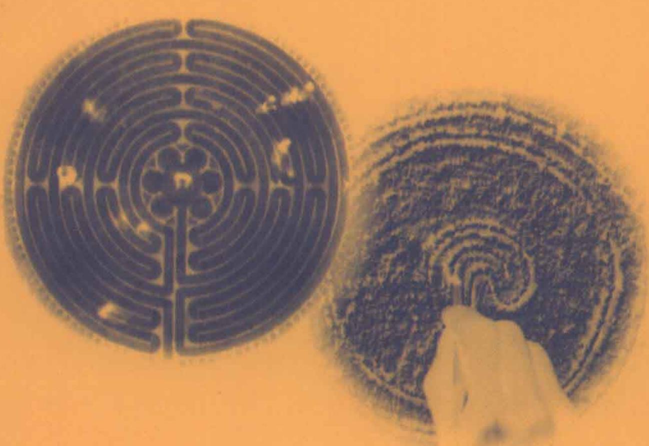
89

Immanuel M. Bomze   Roger Fletcher
Vladimir Demyanov   Tamás Terlaky

Nonlinear Optimization

Cetraro, Italy 2007

Editors:
Gianni Di Pillo, Fabio Schoen



Springer

FONDAZIONE CIME ROBERTO CONTI

Immanuel M. Bomze · Vladimir Demyanov
Roger Fletcher · Tamás Terlaky

Nonlinear Optimization

Lectures given at the
C.I.M.E. Summer School
held in Cetraro, Italy
July 1–7, 2007

Editors:
Gianni Di Pillo
Fabio Schoen

Springer

Fondazione
CIME
Roberto Conti

Editors
Gianni Di Pillo
Sapienza, Università di Roma
Dipto. Informatica e Sistemistica
Via Ariosto, 25
00185 Rome
Italy
dipillo@dis.uniroma1.it

Fabio Schoen
Università degli Studi di Firenze
Dipto. Sistemi e Informatica
Via di Santa Marta, 3
50139 Firenze
Italy
fabio.schoen@unifi.it

***Authors:** see List of Contributors*

ISSN: 0075-8434 e-ISSN: 1617-9692
ISBN: 978-3-642-11338-3 e-ISBN: 978-3-642-11339-0
DOI: 10.1007/978-3-642-11339-0
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010923242

Mathematics Subject Classification (2000): 90C30; 90C26; 90C25; 90C55

Cover images: stills images from video "labirinto" - antonellabussanich@yahoo.fr, © VG Bild-Kunst, Bonn 2010

Cover design: WMXDesign GmbH

Printed on acid-free paper

springer.com

Preface

This volume collects the expanded notes of four series of lectures given on the occasion of the CIME course on *Nonlinear Optimization* held in Cetraro, Italy, from July 1 to 7, 2007.

The Nonlinear Optimization problem of main concern here is the problem of determining a vector of *decision variables* $x \in \mathbb{R}^n$ that minimizes (maximizes) an *objective function* $f(\cdot) : \mathbb{R}^n \to \mathbb{R}$, when x is restricted to belong to some *feasible set* $\mathcal{F} \subseteq \mathbb{R}^n$, usually described by a set of *equality and inequality constraints*: $\mathcal{F} = \{x \in \mathbb{R}^n : h(x) = 0, h(\cdot) : \mathbb{R}^n \to \mathbb{R}^m;\ g(x) \leq 0, g(\cdot) : \mathbb{R}^n \to \mathbb{R}^p\}$; of course it is intended that at least one of the functions f, h, g is nonlinear. Although the problem can be stated in very simple terms, its solution may result very difficult due to the analytical properties of the functions involved and/or to the number n, m, p of variables and constraints. On the other hand, the problem has been recognized to be of main relevance in engineering, economics, and other applied sciences, so that a great lot of effort has been devoted to develop methods and algorithms able to solve the problem even in its more difficult and large instances.

The lectures have been given by eminent scholars, who contributed to a great extent to the development of Nonlinear Optimization theory, methods and algorithms. Namely, they are:

- Professor Immanuel M. BOMZE, University of Vienna, Austria
- Professor Vladimir F. DEMYANOV, St. Petersburg State University, Russia
- Professor Roger FLETCHER, University of Dundee, UK
- Professor Tamás TERLAKY, McMaster University, Hamilton, Ontario, Canada (now at Lehigh University, Bethlehem, PA - USA).

The lectures given by Roger Fletcher deal with a basic framework for treating the Nonlinear Optimization problem in the smooth case, that is the Sequential Quadratic Programming (SQP) approach. The SQP approach can be considered as an extension to constrained problems of Newton's method for unconstrained minimization. Indeed, the underlying idea of the SQP approach is that of applying Newton's method to solve the nonlinear equations given by the first order necessary conditions for optimality. In order to fully develop the idea, the required optimality conditions for the constrained problem are recalled. Then the basic SQP method is introduced and some issues of the method are discussed: in particular the requirement of avoiding the evaluation of second order derivatives, and the occurrence of infeasibility in solving the QP subproblems. However the basic SQP method turns out to be only locally convergent, even if with a superlinear convergence rate. Therefore the need arises of some globalization strategy, that retains the good convergence rate. In particular, two classes of globalization strategies are considered, the first one using some merit function, the second one resorting to some filter method. Filter methods are of main concern in the context of the course, since they have been introduced and developed by Roger Fletcher himself. A last section of the lecture notes deals with the practical problem of interfacing a model of the nonlinear programming problem with a code for its solution. Different modelling languages are mentioned, and a short introduction to AMPL is provided.

The lectures given by Tamás Terlaky, whose chapter is co-authored by Imre Pólik, focus on the Interior Point Methods (IPM), that arose as the main novelty in linear optimization in the eighties of the last century. The interesting point is that the IPM, originally developed for linear optimization, is deeply rooted in nonlinear optimization, unlike the simplex method used until before. It was just the broadening of the horizon from linear to nonlinear, that allowed to describe for the first time an algorithm for linear optimization not only with polynomial complexity but also with competitive performances. The lecture notes first review the IPM for linear optimization, by introducing the self dual-model into which every linear optimization problem can be embedded; the basic notion of central path is defined, and its existence and convergence are analyzed; it is shown that, by a rounding procedure on the central path, a solution of the problem can be found in a polynomial number of arithmetic operations. On these bases, a general scheme of IP algorithms for linear optimization is provided, and several implementation issues are considered. Then, the more general problem of conic optimization is addressed, relying on the fact that most of theoretical results and algorithmic considerations valid for the linear case carry over to the conic case with only minor modifications. Moreover conic optimization represents a step in the pathway from linear to nonlinear optimization. The interest in conic optimization is motivated by important applications, like robust linear optimization, eigenvalue optimization, relaxing of binary variables. In particular, two special classes of conic optimization problems are considered, namely second order conic optimization and semidefinite optimization, and for each class a well suited IPM is described. Finally, the interior point approach is extended to nonlinear optimization, by employing the key of a reformulation of the nonlinear optimization problem as a nonlinear complementarity problem. In this way a central path can be defined also for the nonlinear case, even if its existence and convergence require stronger assumptions than in the linear or conic cases, and complexity results hold only in the convex case. The analytical and algorithmic analysis of IPM is complemented by an overview of existing software implementations, pointing out that some of them are available in leading commercial packages. A challenging list of open questions, concerning mainly algorithmic issues, concludes these lecture notes.

The methods mentioned before are able to find only local solutions of the Nonlinear Optimization problem. In his lectures Immanuel Bomze considers the much more difficult problem of Global Optimization, that is the problem of finding global, rather than local, solutions. In order to fully explain how a gap in difficulty of the problem arises, he makes reference to the simplest nonlinear optimization problem, that is quadratic programming, minimizing a quadratic objective function under linear constraints. If the quadratic objective function is nonconvex, this problem may have so many local non global solutions that any enumerative strategy is not viable. A particular feature of the nonconvex quadratic programming is that necessary and sufficient global optimality conditions can be stated, which not only provide a certificate of optimality for a current tentative solution, but also an improving feasible point if the conditions are not satisfied. These conditions rely on the notion of copositivity, which is central in algorithmic developments. Moreover, additional optimality conditions can be stated in terms of nonsmooth analysis, thus establishing a link with the contents of the lectures by Vladimir Demyanov introduced below. A particular instance of a quadratic programming problem is the so-called Standard Quadratic Programming Problem (StQP), where the feasible set is the unitary simplex. StQP is used to illustrate the basic techniques available for searching global solutions; among these, the well known branch-and-bound approach borrowed from combinatorial optimization. Again, StQP is used to illustrate approaches by which the problem may be in some way reformulated, relaxed or approximated in order to obtain a good proxy of its exact global solution. Finally, a section deals with detecting copositivity, a problem known to be in general NP-hard.

In the Nonlinear Optimization problems considered up to now, the functions f, h, g, are assumed to be smooth, that is at least continuously differentiable. In his lectures, Vladimir Demyanov faces the much more difficult case of nonsmooth optimization. The smooth case can be characterized as the "kingdom of gradient", due to the main role played by the notion of gradient in establishing optimality conditions and in detecting improving feasible solutions. Therefore, a first challenge, when moving outside of that kingdom, is to provide analytical notions able to perform, in some way, the same role. To this aim, different definitions of differentiability and of set-valued subdifferential are introduced, where each element of the subdifferential is, in some sense, a generalized gradient. On these bases, it is possible to establish optimality conditions for nonsmooth optimization problems, not only when the decision variable belongs to the usual $\mathbb{R}^n$ finite dimensional space, but also when it belongs to a more general metric or normed space. More in particular, first the case of unconstrained optimization problems, and then the case of constrained optimization problems are considered. It is remarkable the fact that a nonsmooth constrained optimization problem can always be transformed into a nonsmooth unconstrained optimization problem by resorting to an exact nondifferentiable penalty functions that accounts for the constraints. Therefore, an amazing feature of nonsmooth optimization is that, in principle, the presence of constraints does not add analytical difficulties with respect to the unconstrained case, as it happens if the same exact penalty approach is adopted in smooth optimization.

The course took place in the wonderful location of San Michele Hotel in Cetraro and was attended by 34 researchers from 9 different countries. The course was organized in 6 days of lectures, with each lecturer presenting his course material in 5 parts. The course was indeed successful for its scientific interest and for the friendly environment - this was greatly facilitated by the beauty of the course location and by the professional and warm atmosphere created by the organizers and by all of the staff of Hotel San Michele.

We are very grateful with CIME for the opportunity given of organizing this event and for the financial as well as logistic support; we would like to thank in particular CIME Director, prof. Pietro Zecca, for his continuous encouragement and friendly support before, during and after the School; we also would like to thank Irene Benedetti for her help and participation during the School, and all of the staff of CIME, who made a great effort for the success of this course. In particular we would like to thank Elvira Mascolo, CIME Scientific Secretary, for her precious work in all parts of the organization of the School, and Francesco Mugelli who maintained the web site.

Gianni Di Pillo and Fabio Schoen

Contributors

Immanuel M. Bomze Department of Statistics and Operations Research, University of Vienna, 1010 Wien, Austria, immanuel.bomze@univie.ac.at

Vladimir F. Demyanov Applied Mathematics Dept., St.-Petersburg State University, Staryi Peterhof, St.-Petersburg, 198504 Russia, vfd@ad9503.spb.edu

Roger Fletcher Department of Mathematics, University of Dundee, Dundee DD1 4HN, fletcher@maths.dundee.ac.uk

Imre Pólik Department of Industrial and Systems Engineering, Lehigh University, 200 West Packer Avenue, 18015-1582, Bethlehem, PA, USA, imre@polik.net

Tamás Terlaky Department of Industrial and Systems Engineering, Lehigh University, 200 West Packer Avenue, 18015-1582, Bethlehem, PA, USA, terlaky@lehigh.edu

Contents

Global Optimization: A Quadratic Programming Perspective

Immanuel M. Bomze

Special thanks to Gianni Di Pillo and Fabio Schoen, for their continuous support and valuable remarks

Introduction

Global optimization is a highly active research field in the intersection of continuous and combinatorial optimization (a basic web search delivers over a million hits for this phrase and for its British cousin, *Global Optimisation*). A variety of methods have been devised to deal with this problem class, which – borrowing biological taxonomy terminology in a very superficial way – may be divided roughly into the two domains of *exact/rigorous methods* and *heuristics*, the difference between them probably being that you can prove less theorems in the latter domain. Breaking the domain of exact methods into two phyla of *deterministic methods* and *stochastic methods*, we may have the following further taxonomy of the deterministic phylum:

$$\text{exhaustive methods} \begin{cases} \text{passive/direct, streamlined enumeration} \\ \text{homotopy, trajectory methods} \end{cases}$$

$$\text{methods using global structure} \begin{cases} \text{smoothing, filling, parameter continuation} \\ \text{hierarchical funnel, difference-of-convex} \end{cases}$$

I.M. Bomze (✉)
Department of Statistics and Operations Research, University of Vienna,
1010 Wien, Austria
e-mail: immanuel.bomze@univie.ac.at

G. Di Pillo et al. (eds.), *Nonlinear Optimization*, Lecture Notes in Mathematics 1989, DOI 10.1007/978-3-642-11339-0_1,

$$\text{iterative improvement methods} \begin{cases} \text{escape, tunneling, deflation, aux.functions} \\ \text{successive approximation, minorants} \end{cases}$$

implicit enumeration methods: branch & bound

In these notes, we will focus on a problem class which serves as an application model for some of the above techniques, but which mathematically nevertheless is of surprisingly simple structure – basically the next step after Linear Optimization, namely Quadratic Optimization. Despite the fact that curvature of the objective function is constant and that constraints are linear, quadratic problems exhibit all basic difficulties you may encounter in global optimization: a multitude of inefficient local solutions; global solutions with a very narrow domain of attraction for local solvers; and instances where you encounter very early the optimal solution, but where you find a certificate for global optimality of this solutions, or even only a satisfying rigorous bound very late – a case of particular nuisance in applications.

The contents of these notes are organized as follows: Section 1 deals with local and global optimality conditions in the quadratic world. Due to the constant curvature of the objective function, conditions for both local and global optimality can be formulated in a compact way using the notion of copositivity. It turns out that this class also allows for closing the gap between necessary and sufficient conditions in most cases. ε-subdifferential calculus is used to analyse these conditions in a more general framework, going beyond Quadratic Optimization to the quite general theory of difference-of-convex (d.c.) optimization. To emphasize how close continuous global optimization is tied to discrete problems, we investigate a particular class of quadratic problems, the so-called Standard Quadratic Problems (StQPs) which simply consist of extremizing a quadratic form of the standard simplex – and yet form an NP hard problem class with immediate applications in combinatorial optimization. We continue the study of StPQs in Section 2, to exemplify some basic global optimization techniques like determining escape directions and rigorous bounds, as well as the basic steps in branch-and-bound. Section 3 is devoted to different approaches to global quadratic optimization, namely relaxation and approximation, but also exact reformulation. As an example for the latter, we discuss an emerging branch of optimization which receives rapidly increasing interest in contemporary scientific community: copositive optimization. Again applied to StQPs, the copositive reformulation means that a global quadratic optimization problem is rewritten as a linear programming problem over a convex cone of matrices, thereby completely avoiding the problem of inefficient local solutions. The hardness of the problem is completely shifted to sheer feasibility, and this new aspect opens up a variety of different methods to approach the global solution of the original problem (the StQP in our case). The cone of copositive matrices is known since the mid-fifties of the last century, however, algorithmic approaches to detect whether or not a given matrix satisfies this condition, are much more recent. The last Section 4 is devoted to these aspects, and also discusses some complexity issues. A by now well-established technology for conic optimization which gained momentum since practical implementations of interior-point methods were available, is Semidefinite Programming (SDP), where the matrix cone is that of positive-semidefinite matrices. Since checking positive-semidefiniteness of a given matrix is relatively easy, SDPs can be solved to any prescribed accuracy in polynomial time. Section 4 also describes how SDP-based bounds arising from approximating copositivity via SDP technology can be reinterpreted in the decomposition context of the earlier Section 2.

In the sequel, we will employ the following notation: $^\top$ stands for transposition of a (column) vector in n-dimensional Euclidean space $\mathbb{R}^n$; for two such vectors $\{x, y\} \subset \mathbb{R}^n$, we denote by $x \leq y$ the fact that $x_i \leq y_i$ for all i. The letters o, O, 0 stand for the zero vector, matrix, or number, respectively, all of appropriate dimension. The positive orthant is denoted by $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : o \leq x\}$, the $n \times n$ identity matrix by I_n, with i-th column e_i (the ith standard basis vector). $e = \sum_i e_i \in \mathbb{R}^n$ is the all-ones vector and $E_n = ee^\top$ the all-ones $n \times n$ matrix. For a finite set A, we denote its cardinality by $|A|$. If v is a vector in $\mathbb{R}^n$, we denote by $\operatorname{Diag} v$ the diagonal $n \times n$ matrix D with $d_{ii} = v_i$, for $i = 1, \ldots, n$. Conversely, for an $n \times n$ matrix B, $\operatorname{diag} B = [b_{ii}]_i \in \mathbb{R}^n$ denotes the n-dimensional vector formed by the diagonal elements of B. Finally, we abbreviate by $S \succeq O$ the fact that a symmetric $n \times n$ matrix S is positive-semidefinite (psd), and by $N \geq O$ the fact that N has no negative entries.

1 Global Optimization of Simplest Structure: Quadratic Optimization

Let $Q = Q^\top$ be a symmetric $n \times n$ matrix, A an $m \times n$ matrix and $b \in \mathbb{R}^m$. The feasible set of a quadratic optimization problem (QP) is a polyhedron which can be described as the intersection of finitely many half-spaces:

$$M = \{x \in \mathbb{R}^n : Ax \leq b\}.$$

Hence let us consider here

$$\min \left\{ f(x) = \tfrac{1}{2} x^\top Q x + c^\top x : x \in M \right\}, \tag{1.1}$$

as the basic model of a QP. To conveniently formulate optimality conditions, we frequently employ the *Lagrangian function*

$$L(x; u) = f(x) + u^\top (Ax - b), \quad u \in \mathbb{R}^m_+$$

being the vector of (non-negative) Lagrange multipliers for the (inequality) constraints. In our case, the gradient of L with respect to x at a point $(\overline{x}; u)$ is affine (i.e., a linear map plus a constant vector) and reads $\nabla_x L(\overline{x}; u) = Q\overline{x} + c + A^\top u$.

If Q is positive-semidefinite ($x^\top Qx \geq 0$ for all $x \in \mathbb{R}^n$), then (1.1) is (relatively) easy: the basin(s) of attraction of the global solution(s) are universal, and virtually all local solvers will deliver it. Note that the first-order optimality conditions

$$\nabla_x L(\overline{x}; u) = o \quad \text{and} \quad u^\top (A\overline{x} - b) = 0, \quad (\overline{x}, u) \in M \times \mathbb{R}^m_+ \tag{1.2}$$

are in this case necessary and sufficient for global optimality, and can be recast into a *Linear Complementarity Problem* which may be solved by complementary pivoting methods like Lemke's algorithm [24].

However, if Q is indefinite, then even with a few simple constraints, problem (1.1) is really difficult. As a running example, we will consider the problem to determine the farthest point from a point c, in the hypercube $M = [-1, 1]^n$:

Example 1. Let $Q = -I_n$, and $A = [I_n | - I_n]^\top$ with $b = [e^\top | - e^\top]^\top$. If $c_i \in [-1, 1]$, then it is easily seen that all $y \in \mathbb{R}^n$ with $y_i \in \{-1, c_i, 1\}$, all n, are KKT points, and that all $x \in \{-1, 1\}^n$ (the vertices of the hypercube) are (local) solutions. If $c = o$, all vertices are evidently global solutions. However, if we consider $c \neq o$, this renders a unique global solution, while now all other $(2^n - 1)$ local solutions are inefficient; even more drastically, $(3^n - 1)$ KKT points, i.e., solutions of (1.2) are inefficient. We slightly simplify calculations by restricting ourselves to $c = -\mu e$ where $0 < \mu < 1$. The unique global solution then is the positive vertex $x^* = e$ of M.

1.1 Local Optimality Conditions in QPs

The first-order KKT conditions (1.2) help us to single out finitely many candidates (3^n in Example 1) for being optimal solutions. Note that the complementary slackness conditions – in our case $u^\top (A\overline{x} - b) = 0$ – at $(\overline{x}, u) \in M \times \mathbb{R}^m_+$ always mean $L(\overline{x}; u) = f(\overline{x})$ in terms of the Lagrangian, while primal-dual feasibility always implies $L(x; u) \leq f(x)$, by construction of L, for all $(x, u) \in M \times \mathbb{R}^m_+$.

Now, to remove $3^n - 2^n$ candidates in Example 1 above, we have to employ second-order optimality conditions, using constant curvature of the objective and/or the Lagrangian: both functions have the same Hessian matrix $D^2_x L(x; u) = D^2 f(x) = Q$ for all $(x, u) \in M \times \mathbb{R}^m_+$.

The local view of M from $\overline{x}$ is captured by the tangent cone, which due to linearity of constraints coincides with the cone of feasible directions at $\overline{x}$,

$$\begin{aligned}\Gamma(\overline{x}) &= \{v \in \mathbb{R}^n : \overline{x} + tv \in M \text{ for all small enough } t > 0\,\} \\ &= \mathbb{R}_+(M - \overline{x}) \\ &= \{v \in \mathbb{R}^n : (Av)_i \leq 0 \text{ if } (A\overline{x})_i = b_i, \quad \text{all } i \in \{1,\dots,m\}\}\,. \end{aligned} \tag{1.3}$$

If $\overline{x}$ is a local solution, then a decrease along the feasible direction v is impossible, and this follows if $v^\top Qv \geq 0$ for all feasible directions $v \in \Gamma(\overline{x})$:

$$\begin{aligned} f(x) - f(\overline{x}) &\geq L(x;u) - L(\overline{x};u) \\ &= v^\top \nabla_x L(\overline{x};u) + \tfrac{1}{2}\,v^\top Qv \\ &= v^\top o + \tfrac{1}{2} v^\top Qv \geq 0. \end{aligned}$$

However, this condition is too strong: no locality is involved at all ! Hence we have to repeat the argument directly, with f replacing L. To account for locality, we also put $x = \overline{x} + tv$ with $t > 0$ small. Note that the KKT condition (1.2) implies the weak first-order ascent condition $v^\top \nabla f(\overline{x}) \geq 0$ for the increment function $\theta_v(t) = f(\overline{x} + tv) - f(\overline{x})$ and

$$f(x) - f(\overline{x}) = \theta_v(t) = tv^\top \nabla f(\overline{x}) + \frac{t^2}{2}\, v^\top Qv > 0, \tag{1.4}$$

if $t > 0$ small and $v^\top \nabla f(\overline{x}) > 0$, even if $v^\top Qv < 0$: *strict first-order ascent directions may be negative curvature directions.* Clearly, the sign of $v^\top Qv$ determines curvature of the univariate function θ_v which is convex if and only if $v^\top Qv \geq 0$, and strictly concave otherwise. In the latter case, the condition $\theta_v(t) \geq 0$ for all $t \in [0, \bar{t}\,]$ is equivalent to $\theta_v(\bar{t}\,) \geq 0$, as always $\theta_v(0) = 0$ holds.

Thus we concentrate on the *reduced tangent cone*

$$\Gamma_{\text{red}}(\overline{x}) = \left\{v \in \Gamma(\overline{x}) : v^\top \nabla f(\overline{x}) = 0\right\} \tag{1.5}$$

and stipulate only $v^\top Qv \geq 0$ for all $v \in \Gamma_{\text{red}}(\overline{x})$.

Theorem 1 (2nd order local optimality condition) [20, 23, 48]:
A KKT point $\overline{x}$ (i.e., satisfying (1.2) for some $u \in \mathbb{R}^m_+$) is a local solution to (1.1) if and only if

$$v^\top Qv \geq 0 \qquad \textit{for all } v \in \Gamma_{\text{red}}(\overline{x}), \tag{1.6}$$

*i.e., if Q is $\Gamma_{\text{red}}(\overline{x})$-*copositive. *If (1.6) is violated, then $v \in \Gamma_{\text{red}}(\overline{x})$ with $v^\top Qv < 0$ is a strictly improving feasible direction.*

Copositivity conditions of the form (1.6) will be central also later on in these notes. Here it may suffice to notice that this condition is clearly satisfied