运筹与管理科学丛书 9

Kernel Function-based Interior-point Algorithms for Conic Optimization

(锥优化的基于核函数的内点算法)

Yanqin Bai

SCIENCE PRESS
Beijing

运筹与管理科学丛书 9

Yanqin Bai

Kernel Function-based Interior-point Algorithms for Conic Optimization

(锥优化的基于核函数的内点算法)

SCIENCE PRESS
Beijing

《运筹与管理科学丛书》已出版书目

1. 非线性优化计算方法　袁亚湘　著　2008年2月
2. 博弈论与非线性分析　俞　建　著　2008年2月
3. 蚁群优化算法　马良等　著　2008年2月
4. 组合预测方法有效性理论及其应用　陈华友　著　2008年2月
5. 非光滑优化　高　岩　著　2008年4月
6. 离散时间排队论　田乃硕　徐秀丽　马占友　著　2008年6月
7. 动态合作博弈　高红伟　〔俄〕彼得罗相　著　2009年3月
8. 锥约束优化——最优性理论与增广 Lagrange 方法　张立卫　著　2010年1月
9. Kernel Function-based Interior-point Algorithms for Conic Optimization
   Yanqin Bai　著　2010年7月

Contents

Chapter 1
Introduction

During the last two decades, major development in convex optimization were focusing on conic optimization, primarily, on linear, conic quadratic and semidefinite optimization. Conic programming allows to reveal rich structure which usually is possessed by a convex programm and to exploit this structure in order to process the programm efficiently.

Arkadi Nemirovski
Advance in Convex Optimization: Conic Programming,
Plenary talk in IMU 2009, Spain

After the path-breaking paper of Karmarkar [64], linear optimization (LO) became an active area of research. The resulting interior-point methods (IPMs) are now among the most effective methods for solving LO problems. For a survey we refer to recent books on the subject [123, 165, 170]. In this book we focus on the conic optimization problems which are extended from LO in a cone structure way and solve them by primal-dual IPMs based on kernel functions. It is generally agreed that these IPMs are most efficient from a computational point of view (see, e.g. Andersen et al. [10]). Below we first recall both the developments of conic optimization and IPMs.

1.1 Conic optimization problems

Conic optimization problems are a class of convex nonlinear optimization problems, lying between linear optimization problems and general convex nonlinear optimization problems. Among others, convex quadratic programming (QP) and quadratically constrained programming (QCP) problems, including most portfolio construction and risk budgeting problems, can be formulated as conic optimization problems.

Conic optimization addresses the problem of minimizing a linear objective function over the intersection of an affine set and a convex cone. The general form is as follows

$$(\text{CO}) \qquad \min_{x \in \mathbf{R}^n} \left\{ c^{\mathrm{T}} x \ : \ Ax - b \in \mathcal{K} \right\}.$$

The objective function is $c^{\mathrm{T}}x$, with objective vector $c \in \mathbf{R}^n$. Furthermore, $Ax - b$ represents an affine function from $\mathbf{R}^n$ to $\mathbf{R}^m$ and $\mathcal{K}$ denotes a convex cone in $\mathbf{R}^m$. Usually A is given as an $m \times n$ (constraint) matrix, and $b \in \mathbf{R}^m$. The importance of this class of problems is due to two facts: many nonlinear problems can be modeled as a conic optimization problem, and, secondly, under some weak conditions on the underlying cone $\mathcal{K}$, conic optimization problems can be solved efficiently.

The most easy and most well known case occurs when the cone $\mathcal{K}$ is the nonnegative orthant of $\mathbf{R}^m$, i.e. when $\mathcal{K} = \mathbf{R}^m_+$:

$$(\text{LO}) \qquad \min_{x \in \mathbf{R}^n} \left\{ c^{\mathrm{T}} x \ : \ Ax - b \in \mathbf{R}^m_+ \right\}.$$

This is nothing else as one of the standard forms of LO problem. Thus it becomes clear that LO is a special case of CO. It is well known that LO models cover numerous applications. Whenever applicable, LO allows to obtain useful quantitative and qualitative information on the problem at hand. The specific analytic structure of an LO problem gives rise to a number of general results which provide in many cases valuable insight and understanding. At the same time, this analytic structure underlies some specific computational techniques for LO; these techniques, which by now are perfectly well developed, allow to solve routinely quite large (with tens/hundreds of thousands of variables and constraints) LO problems. Nevertheless, there are many situations in reality which cannot be covered by LO models. To handle these "essentially nonlinear" cases, there is a strong need to extend the basic theoretical results and computational techniques known for LO beyond the bounds of LO.

When passing from a generic LO problem to its nonlinear extensions, we should expect to encounter some nonlinear components in the problem. Historically, this was done by putting the nonlinearity in the functions defining the problem, as done above in problem (P). In conic optimization, however, we replace the cone $\mathbf{R}^m_+$ in LO by a nonlinear convex cone $\mathcal{K}$, and hence the nonlinearity is now captured in the cone. In the next section we discuss some basic properties of relevant convex cones and we introduce two special cones that play prominent role in the context of conic optimization.

In the recent years, a lot of attention has been devoted to conic optimization. The reason is that the interior-point methods that were developed in the last two decades for LO (see, e.g., [123, 143, 165, 170]), and which revolutionized the field of LO, could be naturally extended to obtain polynomial-time methods for CO (see, e.g. [100]). This opened the way to a wide spectrum of new applications which cannot be captured by LO, e.g. in control theory, combinatorial optimization, etc. For a complete survey both of the theory of CO and its applications, we refer to the recent book [23].

The general form of a conic optimization problem is as given by (CO). In this section we start with a discussion of the conditions on the cone $\mathcal{K}$, and we review the three most important cones. Then we deal with the main duality results for CO. It will become clear that under some mild conditions the duality theory for CO closely resembles the well known duality theory for LO.

Recall that a subset $\mathcal{K}$ of $\mathbf{R}^m$ is a cone if

$$a \in \mathcal{K},\ \lambda \geqslant 0 \ \Rightarrow\ \lambda a \in \mathcal{K}, \tag{1}$$

and the cone $\mathcal{K}$ is a convex cone if moreover

$$a, a' \in \mathcal{K} \Rightarrow a + a' \in \mathcal{K}. \tag{2}$$

We will impose three more conditions on $\mathcal{K}$. Recall that CO is a generalization of LO. To obtain duality results for CO similar to those for LO, the cone $\mathcal{K}$ should inherit three more properties from the cone underlying LO, namely the nonnegative orthant:

$$\mathbf{R}^m_+ = \left\{ x = (x_1, \ldots, x_m)^{\mathrm{T}} \ :\ x_i \geqslant 0,\, i = 1, \ldots, m \right\}.$$

This cone is called the *linear cone.* The linear cone is not only a convex cone; it is also pointed, it is closed and it has a nonempty interior. These are exactly the three properties we need. We describe these properties now. A convex cone $\mathcal{K}$ is called pointed if it does not contain a line. This property can be stated equivalently as

$$a \in \mathcal{K},\ -a \in \mathcal{K} \Rightarrow a = 0. \tag{3}$$

A convex cone $\mathcal{K}$ is called closed if it is closed under taking limits:

$$a_i \in \mathcal{K}\ (i = 1, 2, \ldots),\ a = \lim_{i \to \infty} a_i \Rightarrow a \in \mathcal{K}. \tag{4}$$

Finally, denoting the interior of a cone $\mathcal{K}$ as $\mathrm{int}\mathcal{K}$, we will require that

$$\mathrm{int}\mathcal{K} \neq \varnothing. \tag{5}$$

This means that there exists a vector (in $\mathcal{K}$) such that a ball of positive radius centered at the vector is contained in $\mathcal{K}$. In conic optimization we only deal with cones $\mathcal{K}$ that enjoy all of the above properties. So we always assume that $\mathcal{K}$ is *a pointed and closed convex cone with a nonempty interior.* Apart from the linear cone, two other relevant examples of such cones are

(1) The *Lorentz cone*

$$L^m = \left\{ x \in \mathbf{R}^m \ : \ x_m \geqslant \sqrt{x_1^2 + \cdots + x_{m-1}^2} \right\}.$$

This cone is also called the *second-order cone*, or the *ice-cream cone.*

(2) The *positive semidefinite cone* $\mathbf{S}_+^m$. This cone "lives" in the space $\mathbf{S}^m$ of $m \times m$ symmetric matrices (equipped with the Frobenius inner product $\langle A, B\rangle = \mathrm{Tr}(AB) = \sum_{i,j} A_{ij}B_{ij}$) and consist of all $m \times m$ matrices A which are positive semidefinite, i.e.,

$$S_+^m = \left\{ A \in S^m \ : \ x^{\mathrm{T}}Ax \geqslant 0, \quad \forall x \in \mathbf{R}^m \right\}.$$

We assume that the cone $\mathcal{K}$ in (CO) is a direct product of the form

$$\mathcal{K} = \mathcal{K}^1 \times \cdots \times \mathcal{K}^m,$$

where each component $\mathcal{K}^i$ is either a linear, a Lorentz or a semidefinite cone.

1.2 Conic duality

Before we derive the duality theory for conic optimization, we need to define the *dual cone* of a convex cone $\mathcal{K}$:

$$\mathcal{K}_* = \left\{ \lambda \in \mathbf{R}^m \ : \ \lambda^{\mathrm{T}}a \geqslant 0, \forall a \in \mathcal{K} \right\}. \tag{6}$$

We recall the following result from [23].

Theorem 1.1 *Let $\mathcal{K} \subset \mathbf{R}^m$ is a nonempty cone. Then*

(i) *The set $\mathcal{K}_*$ is a closed convex cone.*
(ii) *If $\mathcal{K}$ has a nonempty interior (i.e., $\mathrm{int}\mathcal{K} \neq \varnothing$), then $\mathcal{K}_*$ is pointed.*
(iii) *If $\mathcal{K}$ is a closed convex pointed cone, then $\mathrm{int}\mathcal{K}_* \neq \varnothing$.*
(iv) *If $\mathcal{K}$ is a closed convex cone, then so is $\mathcal{K}_*$, and the cone dual to $\mathcal{K}_*$ is $\mathcal{K}$ itself.*

Corollary 1.2 *If $\mathcal{K} \subset \mathbf{R}^m$ is a closed pointed convex cone with nonempty interior then so is $\mathcal{K}_*$, and vice versa.*

One may easily verify that the three cones introduced in Section 1.1 are self-dual. The dual of a direct product of convex cones is the direct product of their duals, i.e.,

$$\mathcal{K} = \mathcal{K}^1 \times \cdots \times \mathcal{K}^m \Rightarrow \mathcal{K}_* = \mathcal{K}_*^1 \times \cdots \times \mathcal{K}_*^m.$$

As a consequence, any direct product of linear, Lorentz and semidefinite cones is self-dual.

1.3 From the dual cone to the dual problem

Now we are ready to deal with the dual problem for a conic problem (CO). We start with observing that whenever x is a feasible solution for (CO) then the definition of $\mathcal{K}_*$ implies $\lambda^{\mathrm{T}}(Ax - b) \geqslant 0$, for all $\lambda \in \mathcal{K}_*$, and hence x satisfies the scalar inequality

$$\lambda^{\mathrm{T}} Ax \geqslant \lambda^{\mathrm{T}} b, \quad \forall \lambda \in \mathcal{K}_*.$$

It follows that whenever $\lambda \in \mathcal{K}_*$ satisfies the relation

$$A^{\mathrm{T}}\lambda = c, \tag{7}$$

then one has

$$c^{\mathrm{T}}x = (A^{\mathrm{T}}\lambda)^{\mathrm{T}}x = \lambda^{\mathrm{T}}Ax \geqslant \lambda^{\mathrm{T}}b = b^{\mathrm{T}}\lambda,$$

for all x feasible for (CP). So, if $\lambda \in \mathcal{K}_*$ satisfies (7), then the quantity $b^{\mathrm{T}}\lambda$ is a lower bound for the optimal value of (CP). The best lower bound obtainable in this way is the optimal value of the problem

$$\text{(CD)} \qquad \max_{\lambda \in \mathbf{R}^m} \left\{ b^{\mathrm{T}}\lambda \ : \ A^{\mathrm{T}}\lambda = c, \, \lambda \in \mathcal{K}_* \right\}.$$

By definition, (CD) is the *dual problem* of (CO). Using Theorem 1.1 (iv), one easily verifies that the duality is symmetric: the dual problem is conic and the problem dual to the dual problem is the primal problem.

Indeed, from the construction of the dual problem it immediately follows that we have the weak duality property: if x is feasible for (CP) and λ is feasible for (CD), then

$$c^{\mathrm{T}}x - b^{\mathrm{T}}\lambda \geqslant 0.$$

The crucial question is, of course, if we have equality of the optimal values whenever (CO) and (CD) have optimal values. Different from the LO case, however, this is in general not the case, unless some additional conditions are satisfied. The following theorem clarifies the situation. For its proof we refer again to [23]. We call the problem (CO) *solvable* if it has a (finite) optimal value, and this value is attained.

Before stating the theorem it may be worth pointing out that a finite optimal value is not necessarily attained. For example, the problem

$$\min_{x,y\in \mathbf{R}} \left\{ x : \begin{pmatrix} x & 1 \\ 1 & y \end{pmatrix} \succeq 0 \right\}$$

has optimal value 0, but one may easily verify that this value is not attained. We need one more definition: if there exists an x such that $Ax - b \in \text{int}\mathcal{K}$ then we say that (CO) is *strictly feasible.* We have similar, and obvious, definitions for (CD) being solvable and strictly feasible, respectively.

Theorem 1.3 *Let the primal problem* (CO) *and its dual problem* (CD) *be as given above. The strictly feasibility and below bounded of one of the problems imply solvability of both of them and characterization of optimality. It can be stated as follows.*

- (i) (a) *If* (CO) *is below bounded and strictly feasible, then* (CD) *is solvable and the respective optimal values are equal.*
  - (b) *If* (CD) *is above bounded and strictly feasible, then* (CP) *is solvable, and the respective optimal values are equal.*
- (ii) *Suppose that at least one of the two problems* (CO) *and* (CD) *is bounded and strictly feasible. Then a primal-dual feasible pair* (x, λ) *is comprised of optimal solutions to the respective problems*
  - (a) *if and only if* $b^{\mathrm{T}}\lambda = c^{\mathrm{T}}x$ *(zero duality gap).*
  - (b) *if and only if* $\lambda^{\mathrm{T}}[Ax - b] = 0$ *(complementary slackness).*

Note that this result is slightly weaker than the corresponding result for the LO case. In the LO case the same theorem holds by putting everywhere "feasible" instead of "strictly feasible". The adjective "strictly" cannot be omitted here, however. For a more extensive discussion and some appropriate counterexamples we refer to [23].

1.4 Development of the interior-point methods

The study of interior-point methods is currently one of the most active research areas in optimization. The name interior-point methods originates from the fact that the points generated by an interior-point methods lie in the interior of the feasible region. This is in contrast with the famous and well-established simplex method where the iterates move along the boundary of the feasible region from one extreme point to another. Nowadays, interior-point methods for linear optimization have become quite mature in theory, and have been applied to practical linear optimization with extraordinary success. Linear optimization is one of the most widely applied mathematical techniques. The last fifteen years gave rise to revolutionary developments, both in computer technology and in algorithms for linear optimization. As a consequence, linear optimization that fifteen years ago required a computational time of one year, can now be solved within a couple of minutes. The achieved acceleration is due partly to advances in computer technology but significant part also to the new interior-point methods for Linear optimization.

During the 1940's, it became clear that an effective computational method was required to solve the many linear optimization problems that originated from logistical questions that had to be solved during World War II. The first practical method for solving Linear optimization was the simplex method, proposed by Dantzig [36], in 1947. This algorithm explicitly explores the combinatorial structure of the feasible region to locate a solution by moving from a vertex of the feasible set to an adjacent vertex while improving the value of the objective function. Since then, the method has been routinely used to solve problems in business, logistics, economics, and engineering. In an effort to explain the remarkable efficiency of the simplex method, using the theory of complexity, one has tried very hard to prove that the computational effort to solve an linear optimization problem via the simplex method is polynomially bounded in terms of the size of a problem instance. Klee and Minty, have shown that the worst-case behavior of the simplex method is exponential in [66].

The first polynomial method for solving linear optimization was proposed by Khachiyan [65], in 1979. It is the so-called ellipsoid method. It is based on the ellipsoid technique for nonlinear optimization developed by Shor [132]. With this technique, Khachiyan proved that linear optimization belongs to the class of polynomially solvable problems. Although this result had a great theoretical impact, it failed to keep up its promises in actual computational efficiency. A second proposal was made in 1984 by Karmarkar in [64]. Karmarkar's algorithm is also polynomial, with a better complexity bound than Khachiyan's, but it has the further advantage of being highly efficient in practice. After an initial controversy it has been established that for very large, sparse problems, subsequent variants of Karmarkar's method often outperform the simplex method. Though the field of linear optimization was then considered more or less mature, after Karmarkar's paper it suddenly surfaced as one of the most active areas of research in optimization. In the period 1984—1989 more than 1300 papers were published on the subject. Originally, the aim of the research was to get a better understanding of the so-called projective method of Karmarkar. Soon it became apparent that this method was related to classical methods like the affine scaling method of Dikin [38], the logarithmic barrier method of Frisch [41], and the center method of Huard, and that the last two methods, when tuned properly, could also be proved to be polynomial.

The interior-point methods developed for linear optimization could be naturally extended to obtain polynomial-time methods for conic optimization. In conic optimization, a linear function is minimized over the intersection of an affine space and a closed convex cone. The foundation for solving these problems by interior-point methods was laid by Nesterov and Nemirovskii [100]. These authors considered primal (and dual) interior point methods based on so-called self-concordant barrier functions. Later, Nesterov and Todd [101, 102] introduced symmetric primal-dual interior-point methods on a special class of cones called self-scaled cones, which allowed a symmetric treatment of the primal and the dual problem. Conic optimization includes solving problems such as linear optimization, semidefinite optimization and second order cone optimization problems. During the last two decades interior-point methods have proved to be a powerful tool to solve convex optimization problems, provided that we have a self-concordant computationally tractable barrier function for the underlying cone. Until recently all the barrier functions considered were so-called logarithmic barrier functions. However, there is a gap between the practical behavior of the algorithms and the theoretical performance results, where the practical behavior is better than the worst-case complexity analysis. This is especially true for the so-called large- update methods. If n denotes the number of variables in the problem, then the theoretical complexity analysis of large-update methods yielded an $O\left(n \log\left(\frac{n}{\varepsilon}\right)\right)$ iteration bound, where ε represents the desired accuracy of the solution. In practice, however, large-update methods are much more efficient than the so-called small-update methods for which the theoretical iteration bound is only $O\left(\sqrt{n} \log\left(\frac{n}{\varepsilon}\right)\right)$. So the current theoretical bounds differ by a factor $\sqrt{n}$, in favor of the small-update methods. This gap is significant.

As we mentioned before, several interior-point methods for linear optimization were extended to semidefinite optimization and second-order cone optimization. In fact, these optimization problems can be defined as minimizing a linear function over the intersection of an affine space and a closed convex cone. If the cone is the linear cone, the second order cone or the cone of real semidefinite positive symmetric matrices, then we have respectively a linear optimization problem, a second-order cone optimization problem or a semidefinite optimization problem. These three cones are the most relevant for the optimization field, and they were classified as belonging to the set of self-dual and homogenous cones, also called symmetric cones. Thus, many authors developed interior-point methods for conic optimization by generalizing existing interior-point methods for linear optimization.

1.5 Scope of the book

In preparing this book, a special efforts has been made to explore the idea of interior-point methods for conic optimization based on kernel functions.

In Chapter 2, the kernel functions are discussed and illustrated by eleven examples. Properties of kernel functions, including barrier properties, minima over its domain as well as their differentiability, are discussed in the chapter. Furthermore, the barrier functions based on kernel function are presented. These barrier function will be used as the search direction for designing and analyzing the primal-dual interior-point algorithms.

Chapter 3 presents a primal-dual interior-point algorithms for solving linear optimization problem based on the kernel functions, in which the KKT conditions of problem is perturbed and the barrier function determined by kernel function is essentially solved by Newton method. The iteration bounds for large- and small-update methods are discussed. Chapter 4 the algorithm is extended to solve $P^*(\kappa)$ linear complementarity problem.

The remaining chapters are Chapter 5 and 6 that present the extension of algorithm based on kernel functions to semidefinite optimization and second-order cone optimization, respectively. The complexity bound for large- and small-update methods are obtained, respectively. The numerical examples are shown that the algorithms are efficient.

Chapter 2

Kernel Functions

In the past two decades the barrier functions play an important rule in the development of interior-point methods (IPMs). The selection of a suitable barrier may help the construction of the scheme of efficient interior point algorithm. The key ingredient in kernel function-based interior-point algorithm is the construction of an univariate function for the central path of the optimization problem considered. Then n-dimensional function determined by this univariate function and called barrier function provides provably good proximation between the iterates and the central path. The excellent properties of kernel function and the barrier are easy to computing and derivative.

Kernel function is one of major theme of this book, and this chapter provides an elementary introduction. We introduce the donation of kernel functions and explore their properties which will be used for analysis of the interior-point algorithms. We will move on the barrier function determined by kernel function. We then demonstrate the properties of barrier functions which can be applied directly to compute the complexity bound of algorithms. In final section of this chapter we generalize the kennel function to finite kernel function and parametric kernel functions, respectively.

2.1 Definition of kernel functions and basic properties

Beginning with introduction of the notation of kernel functions in this section, we present their further conditions and develop some of their basic properties both of which will lead to the analysis for IPMs in the following sections.

Definition 2.1 *Let $\psi : \mathbf{R}_{++} \to \mathbf{R}_{+}$ be twice differentiable. A univariate function ψ is called a kernel function if it is satisfied with the following conditions:*

$$\psi'(1) = \psi(1) = 0; \tag{8}$$

$$\psi''(t) > 0; \tag{9}$$

$$\lim_{t\to 0} \psi(t) = \lim_{t\to\infty} \psi(t) = \infty. \tag{10}$$

Obviously, (8) and (9) say that $\psi(t)$ is a nonnegative strictly convex function and attains its minimizer at $t = 1$ with the optimal value that $\psi(1) = 0$. Furthermore, (8) and (9) implies that $\psi(t)$ is completely determined by its second derivative:

$$\psi(t) = \int_1^t \int_1^\xi \psi''(\zeta)\, \mathrm{d}\zeta \mathrm{d}\xi. \tag{11}$$

Moreover, (10) expresses that $\psi(t)$ is coercive and has the barrier property.

The prototype self-regular kernel function in [110] is given by

$$\Upsilon_{p,q}(t) = \frac{t^{p+1} - 1}{p(p+1)} + \frac{t^{1-q} - 1}{q(q-1)} + \frac{p-q}{pq}(t-1),$$

where $p \geqslant 1$ and $q > 1$. The parameter p is called the growth degree, and q the barrier degree of the kernel function.

The following are some examples of kernel functions:

(1) $\psi(t) = \dfrac{t^2 - 1}{2} - \log t, \quad t > 0.$

(2) $\psi(t) = \dfrac{t^2 - 1}{2} + \dfrac{t^{1-q} - 1}{q - 1}, \quad t > 0, \quad q > 1.$

(3) $\psi(t) = \dfrac{t^2 - 1}{2} - \displaystyle\int_1^t \mathrm{e}^{q(\frac{1}{\xi} - 1)} \mathrm{d}\xi, \quad t > 0, \quad q \geqslant 1.$

(4) $\psi(t) = t - 1 + \dfrac{t^{1-q} - 1}{q - 1}, \quad t > 0, \quad q > 1.$

2.2 The further conditions of kernel functions

In this section, we work with five more conditions on the kernel function, namely,

$$t\psi''(t) + \psi'(t) > 0, \quad t < 1, \tag{12}$$

$$t\psi''(t) - \psi'(t) > 0, \quad t > 1, \tag{13}$$

$$\psi'''(t) < 0, \quad t > 0, \tag{14}$$

$$2\psi''(t)^2 - \psi'(t)\psi'''(t) > 0, \quad t < 1, \tag{15}$$

$$\psi''(t)\psi'(\beta t) - \beta\psi'(t)\psi''(\beta t) > 0, \quad t > 1, \beta > 1. \tag{16}$$

Note that conditions (14) and (16) require that $\psi(t)$ is three times differentiable. Furthermore, condition (12) is obviously satisfied if $t \geqslant 1$, since then $\psi'(t) \geqslant 0$ and, similarly, condition (13) is satisfied if $t \leqslant 1$, since then $\psi'(t) \leqslant 0$. Also (15) is obviously satisfied if $t \geqslant 1$ since then $\psi'(t) \geqslant 0$, whereas $\psi'''(t) < 0$. We conclude