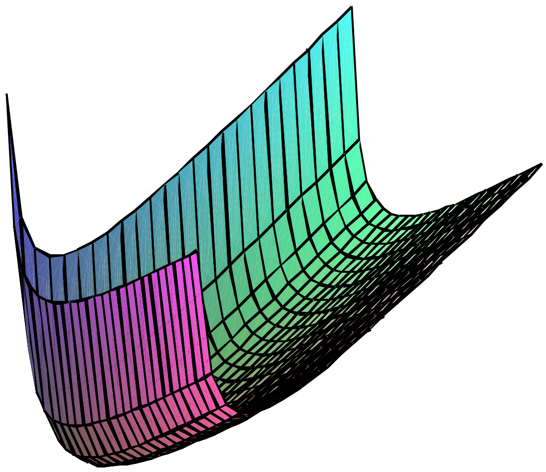
New Primal-dual Interior-point Methods Based on Kernel Functions



New Primal-dual Interior-point Methods Based on Kernel Functions

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. dr. ir. J. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
dinsdag 25 Oktober 2005 om 15.30 uur
door

Mohamed EL GHAMI

Certificat d'Etudes approfondies Mathematiques (C.E.A)
Universite Mohammed V, Rabat, Marokko
geboren te Tamsamane, Marokko.

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. ir. C. Roos

Toegevoegd promotor:

Dr. J.B.M. Melissen

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. C. Roos	Technische Universiteit Delft, promotor
Dr. J.B.M. Melissen	Technische Universiteit Delft, toegevoegd promotor
Prof. dr. G. J. Olsder	Technische Universiteit Delft
Prof. dr. ir. D. den Hertog	Tilburg University, Netherlands
Prof. dr. T. Terlaky	Mc-Master University, Hamilton, Canada
Prof. dr. Y. Nesterov	University Catholique de Louvain-la-Neuve, Belgium
Prof. dr. F. Glineur	University Catholique de Louvain-la-Neuve, Belgium
Prof. dr. ir. P. Wesseling	Technische Universiteit Delft, reservelid

Dit proefschrift kwam tot stand onder auspicien van:

THOMAS STIELTJES INSTITUTE
FOR MATHEMATICS

ISBN:90-9020025-8

Author email: `m.elghami@ewi.tudelft.nl`

To the memory of my grandmother and grandfathers

Acknowledgements

I am very grateful to my supervisors. Firstly, I would like to acknowledge Prof.dr.ir. C. Roos for his invaluable contributions and support during the period that this thesis was written, and Dr. J.B.M. Melissen for his help and discussion during the last two years of my Phd. I would also like to express my gratitude to Prof. Z.A. Guennoun, for his encouragement and authorization that has led to my research and stay in The Netherlands.

There are three former members of the group of Algorithms who I would specifically like to thank for the support they have given me over the years: Y. Bai, F.D. Barb and E. de Klerk.

Thanks are also due to all other members of our group. A special acknowledgement for: Ivan Ivanov is in place because of his help with the implementation of the algorithm presented in this thesis.

My research was financially supported by the Dutch Organization for Scientific Research (NWO grant 613.000.110), wich is kindly acknowledged.

Moving towards more personal acknowledgements, I would like to express aggregated thanks towards all my members of my family and friends[1] — I wish to thank in particular Ahmed Agharbi (for his friendship); Said Hamdioui (for changing my life from worse to bad); Hassan Maatok (he know way).

Finally, I wish to thank my parents A. El Ghami and F. Bouri and my sister H. El Ghami, my brothers Omar and Makki and my wife S. Belgami, who gave me enduring support during all those years, and I want also to honor my daughter Dounia in these acknowledgements.

[1] You know who you are.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

The study of Interior-Point Methods (IPMs) 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 IPM 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, IPMs for Linear Optimization (LO) have become quite mature in theory, and have been applied to practical LO problems with extraordinary success. In this chapter, a short survey of the fields of linear optimization and interior point methods is presented. Based on the simple model of standard linear optimization problems , some basic concepts of interior point methods and various strategies used in the algorithm are introduced. The scope of this thesis follows at the end of the chapter.

1.2 A short history of Linear Optimization

Linear optimization is one of the most widely applied mathematical techniques.[1] The last 15 years gave rise to revolutionary developments, both in computer technology and in algorithms for LO. As a consequence, LO-problems that 15 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 IPMs for LO .

[1] This section is based on a historical review in [Roo02].

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 LO-problems was the simplex method, proposed by Dantzig [Dan63], 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 LO-problem via the simplex method is polynomially bounded in terms of the size of a problem instance. Klee and Minty [Kle72], have shown in 1972 that in the process of solving the problem

$$\begin{aligned} \text{maximize} \sum_{j=1}^{n} 10^{n-j} x_j & \\ \text{s.t.} \left(2 \sum_{j=1}^{i-1} 10^{i-j} x_j \right) + x_i &\leq 100^{i-1}, \quad (i, j = 1, \ldots, n), \\ x_j &\geq 0. \end{aligned} \tag{1.2.1}$$

the simplex method goes through $2^n - 1$ vertices.[2] This shows that the worst-case behavior of the simplex method is exponential.

The first polynomial method for solving LO problems was proposed by Khachiyan, in 1979. It is the so-called ellipsoid method [Kha79]. It is based on the ellipsoid technique for nonlinear optimization developed by Shor [Sho87]. With this technique, Khachiyan proved that LO 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 [Kar84]. 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 LO 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

[2] Recently in [Dez04a; Dez04b] the authors prove that by adding an exponential number of redundant inequalities, the central path-following interior point methods visits small neighborhoods of all the vertices of the Klee-Minty cube.

this method was related to classical methods like the affine scaling method of Dikin [Dik67; Dik74; Dik88], the logarithmic barrier method of Frisch [Fri55; Fri56], and the center method of Huard [Hua67], and that the last two methods, when tuned properly, could also be proved to be polynomial. Moreover, it turned out that the IPM-approach to LO has a natural generalization to the related field of convex nonlinear optimization, which resulted in a new stream of research and an excellent monograph of Nesterov and Nemirovski [Nes93]. This monograph opened the way into other new subfields of optimization, like semidefinite optimization and second order cone optimization, with important applications in system theory, discrete optimization, and many other areas. For a survey of these developments the reader may consult Vandenberghe and Boyd [Boy96], and the book of Ben-Tal and Nemirovski [BT01].

1.3 Primal-dual interior point methods for LO

In this section we proceed by describing primal-dual interior point methods for LO and some recent results [Pen01; Pen02a; Pen02b]. There are many different ways to represent an problem. The two most popular and widely used representations are the standard and the canonical[3] forms. It is well known [Gol89] that any LO problem can be converted into standard or canonical form. In this thesis we consider the *standard linear* optimization problem

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

where $A \in \mathbf{R}^{m\times n}$ is a real $m \times n$ matrix of rank m, and $x, c \in \mathbf{R}^n$, $b \in \mathbf{R}^m$. The dual problem of (P) is given by

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

with $y \in \mathbf{R}^m$ and $s \in \mathbf{R}^n$.
The two problems (P) and (D) share the matrix A and the vectors b and c in their description. But the role of b and c has been interchanged: the objective vector c of (P) is the right-hand side vector of (D), and, similarly, the right-hand side vector b of (P) is the objective vector of (D). Moreover, the constraint matrix in (D) is the transposed matrix A^T, where A is the constraint matrix in (P). It is well known [Roo05], that finding an optimal solution of (P) and (D) is equivalent to solving the non-linear system of equations

$$\begin{aligned} Ax &= b, \quad x \geq 0, \\ A^T y + s &= c, \quad s \geq 0, \\ xs &= 0. \end{aligned} \tag{1.3.1}$$

[3] In the canonical form of the LO problem all constraints are inequality constraints, e.g., $\min\left\{c^T x \;:\; Ax \geq b,\, x \geq 0\right\}$.

The first equation requires that x is feasible for (P), and the second equation that the pair (y, s) is feasible for (D), whereas the third equation is the so-called *complementarity condition* for (P) and (D); here xs denotes the coordinatewise product of the vectors x and s, i.e.

$$xs = [x_1 s_1; x_2 s_2; \ldots ; x_n s_n].$$

We shall also use the notation

$$\frac{x}{s} = \left[\frac{x_1}{s_1}; \frac{x_2}{s_2}; \ldots ; \frac{x_n}{s_n}\right],$$

for each vector x and s such that $s_i \neq 0$, for all $1 \leq i \leq n$. For an arbitrary function $f : \mathbf{R} \to \mathbf{R}$, and an arbitrary vector x we will use the notation

$$f(x) = [f(x_1); f(x_2); \ldots ; f(x_n)].$$

The basic idea underlying primal-dual IPMs is to replace the third (non-linear) equation in (1.3.1) by the nonlinear equation $xs = \mu\mathbf{1}$, with parameter $\mu > 0$ and with $\mathbf{1}$ denoting the all-one vector $(1; 1; \ldots ; 1)$. The system (1.3.1) now becomes:

$$\begin{aligned} Ax &= b, \quad x \geq 0, \\ A^T y + s &= c, \quad s \geq 0, \\ xs &= \mu\mathbf{1}. \end{aligned} \tag{1.3.2}$$

Note that if x and s solve this system then these vectors are necessarily positive. Therefore, in order for (1.3.2) to be solvable there needs to exist a triple (x^0, y^0, s^0) such that

$$Ax^0 = b, \qquad x^0 > 0, \qquad A^T y^0 + s^0 = c, \quad s^0 > 0. \tag{1.3.3}$$

We assume throughout that both (P) and (D) satisfy this condition, which is known as the *interior-point condition* (IPC). For this and some of the properties mentioned below, see, e.g., [Roo05].

Satisfaction of the IPC can be assumed without loss of generality. In fact we may, and will, even assume that $x^0 = s^0 = \mathbf{1}$ [Roo05]. From (1.3.3) we observe that these x^0 and s^0, for some appropriate y^0, solve (1.3.2) when $\mu = 1$. If the IPC holds, the parameterized system (1.3.2) has a unique solution $(x(\mu), y(\mu), s(\mu))$ for each $\mu > 0$; $x(\mu)$ is called the μ-center of (P) and $(y(\mu), s(\mu))$ is the μ-center of (D). The set of μ-centers (with $\mu > 0$) defines a homotopy path, which is called *the central path* of (P) and (D) [Meg89; Son86]. If $\mu \to 0$ then the limit of the central path exists. This limit satisfies the complementarity condition , and hence yields optimal solutions for (P) and (D) [Roo05].

IPMs follow the central path approximately. Let us briefly indicate how this works. Without loss of generality we assume that $(x(\mu), y(\mu), s(\mu))$ is known for some positive μ. For example, due to the above choice, we may assume this to be the case for $\mu = 1$, with $x(1) = s(1) = \mathbf{1}$. We then decrease μ to $\mu_+ := (1-\theta)\mu$, for some $\theta \in (0,1)$ and apply Newton's method to iteratively solve the non-linear equations (1.3.2). So for each step we have to solve the following Newton system.

$$\begin{aligned} A\Delta x &= 0, \\ A^T\Delta y + \Delta s &= 0, \\ s\Delta x + x\Delta s &= \mu_+\mathbf{1} - xs. \end{aligned} \tag{1.3.4}$$

Because A has full row rank, the system (1.3.4) uniquely defines a search direction $(\Delta x, \Delta s, \Delta y)$ for any $x > 0$ and $s > 0$; this is the so-called Newton direction and this direction is used in all existing implementations of the primal-dual method. The first two equations take care of primal and dual feasibility after a (small enough) step along the Newton direction, whereas the third equation serves to drive the new iterates to the μ_+-centers. The third equation is called the *centering equation*.

By taking a step along the search direction, with the step size defined by a line search rule, one constructs a new triple (x, y, s), with $x > 0$ and $s > 0$. If necessary, we repeat the procedure until we find iterates that are close enough to $(x(\mu), y(\mu), s(\mu))$. Then μ is again reduced by the factor $1-\theta$ and we apply Newton's method targeting at the new $\mu-$centers, and so on. This process is repeated until μ is small enough, say until $n\mu \le \epsilon$; at this stage we have found $\epsilon-$solutions of the problems (P) and (D).

In this thesis we follow [Bai04a; Pen00a; Pen00b; Pen01; Roo05; Ye97] and reformulate this approach by defining the same search direction in a different way. To make this clear we associate to any triple (x, s, μ), with $x > 0$ and $s > 0$ and $\mu > 0$, the vector

$$v := \sqrt{\frac{xs}{\mu}}. \tag{1.3.5}$$

Note that if x is primal feasible and s is dual feasible then the pair (x, s) coincides with the μ-center $(x(\mu), s(\mu))$ if and only if $v = \mathbf{1}$.
Introducing the notations

$$\bar{A} := \frac{1}{\mu}AV^{-1}X = AS^{-1}V, \tag{1.3.6}$$

$$V := \operatorname{diag}(v),\ X := \operatorname{diag}(x),\ S := \operatorname{diag}(s), \tag{1.3.7}$$

and defining the *scaled search directions* d_x and d_s according to

$$d_x := \frac{v\Delta x}{x}, \qquad d_s := \frac{v\Delta s}{s}, \tag{1.3.8}$$

the system (1.3.4), can be rewritten as

$$\begin{aligned} \bar{A}d_x &= 0, \\ \bar{A}^T\Delta y + d_s &= 0, \\ d_x + d_s &= v^{-1} - v. \end{aligned} \tag{1.3.9}$$

Note that d_x and d_s are orthogonal vectors, since d_x belongs to the null space of the matrix $\bar{A}$ and d_s to its row space. Hence, we will have $d_x = d_s = 0$ if and only if $v^{-1} - v = 0$, which is equivalent to $v = \mathbf{1}$. We conclude that $d_x = d_s = 0$ holds if and only if the pair (x, s) coincides with the μ-center $(x(\mu), s(\mu))$.

We make another crucial observation. The third equation in (1.3.9) is called the *scaled centering equation*. The right-hand side $v^{-1} - v$ in the scaled centering equation equals minus the gradient of the function

$$\Psi_c(v) := \sum_{i=1}^{n} \left(\frac{v_i^2 - 1}{2} - \log v_i \right). \tag{1.3.10}$$

Not that $\nabla^2 \Psi_c(v) = \operatorname{diag}\left(\mathbf{1} + v^{-2}\right)$ and that this matrix is positive definite, so $\Psi_c(v)$ is strictly convex. Moreover, since $\nabla \Psi_c(\mathbf{1}) = 0$, it follows that $\Psi_c(v)$ attains its minimal value at $v = \mathbf{1}$, with $\Psi_c(\mathbf{1}) = 0$. Thus it follows that $\Psi_c(v)$ is nonnegative everywhere and vanishes if and only if $v = \mathbf{1}$, i.e., if and only if $x = x(\mu)$ and $s = s(\mu)$. The μ-centers $x(\mu)$ and $s(\mu)$ can therefore be characterized as the minimizers of $\Psi_c(v)$.

1.3.1 Primal-dual interior point methods based an kernel functions

Now we are ready to describe the idea underlying the approach in this thesis. In the scaled centering equation, the last equation of (1.3.9), we replace the scaled barrier function $\Psi_c(v)$ by an arbitrary strictly convex function $\Psi(v)$, $v \in \mathbf{R}^n_{++}$ such that $\Psi(v)$ is minimal at $v = \mathbf{1}$ and $\Psi(\mathbf{1}) = 0$, where $\mathbf{R}^n_{++}$ denote a positive orthant. Thus the new scaled centering equation becomes

$$d_x + d_s = -\nabla \Psi(v). \tag{1.3.11}$$

As before, we will have $d_x = 0$ and $d_s = 0$ if and only if $v = \mathbf{1}$, i.e., if and only if $x = x(\mu)$ and $s = s(\mu)$, as it should be.

To simplify matters we restrict ourselves to the case where $\Psi(v)$ is separable with identical coordinate functions. Thus, letting ψ denote the function on the coordinates, we write

$$\Psi(v) = \sum_{i=1}^{n} \psi(v_i), \tag{1.3.12}$$

where $\psi(t) : D \to \mathbf{R}_+$, with $\mathbf{R}_{++} \subseteq D$, is strictly convex and minimal at $t = 1$, with $\psi(1) = 0$. In the present context we call the univariate function $\psi(t)$ the *kernel function* of $\Psi(v)$. We will always assume that the kernel function is twice differentiable. Observe that $\psi_c(t)$, given by

$$\psi_c(t) := \frac{t^2 - 1}{2} - \log t, \quad t > 0, \tag{1.3.13}$$

is the kernel function yielding the Newton direction, as defined by (1.3.9). In this general framework we call $\Psi(v)$ a *scaled barrier function*. An unscaled barrier function, whose domain is the (x, s, μ)-space, can be obtained via the definition

$$\Phi(x, s, \mu) = \Psi(v) = \sum_{i=1}^{n} \psi\left(v_i\right) = \sum_{i=1}^{n} \psi\left(\sqrt{\frac{x_i s_i}{\mu}}\right). \tag{1.3.14}$$

One may easily verify that by application of this definition to the kernel function in (1.3.13) we obtain — up to a constant factor and a constant term — the classical logarithmic barrier function.

Any proximity function $\Psi(v)$ gives rise to a primal-dual IPM, as described below in Figure 1.1. With $\bar{A}$ as defined in (1.3.7), the search direction in the algorithm is obtained by solving the system

$$\begin{array}{rcl} \bar{A}d_x & = & 0, \\ \bar{A}^T\Delta y + d_s & = & 0, \\ d_x + d_s & = & -\nabla\Psi(v), \end{array} \tag{1.3.15}$$

for d_x, Δy and d_s, and then computing Δx and Δs from

$$\Delta x = \frac{x d_x}{v}, \qquad \Delta s = \frac{s d_s}{v}, \tag{1.3.16}$$

according to (1.3.8).

The inner while loop in the algorithm is called *inner iteration* and the outer while loop *outer iteration*. So each outer iteration consists of an update of the barrier parameter and a sequence of one or more inner iterations.

It is generally agreed that the total number of inner iterations required by the algorithm is an appropriate measure for the efficiency of the algorithm. This number will be referred to as the *iteration complexity* of the algorithm. Usually the iteration complexity is described as a function of the dimension n of the problem and the accuracy parameter ϵ.

A crucial question is, of course, how to choose the parameters that control the algorithm, i.e., the proximity function $\Psi(v)$, the threshold parameter τ, the barrier update parameter θ, and the step size α, so as to minimize the iteration complexity.

In practice one distinguishes between *large-update methods* [Ans92; Gon92; Gon91; Her92; Jan94a; Koj93a; Koj93b; Tod96; Roo89], with $\theta = \Theta(1)$, and *small-update methods*, with $\theta = \Theta(1/\sqrt{n})$ [And96; Her94; Tod89].

Figures 1.2 and 1.3 exhibit the behavior of IPMs with large-update and small-update for a specific two-dimensional LO problem. These figures are drawn in xs-space. Note that in the xs-space the central path is represented by the straight line consisting of all vectors μe, $\mu > 0$. In these figures we have drawn the iterates for a simple problem and also the level curves for $\psi(v) = 1$ around the target points on the central path that are used during the algorithm.

Generic Primal-Dual Algorithm for LO

Input:
A kernnel function $\psi(t)$;
a threshold parameter $\tau > 0$;
an accuracy parameter $\epsilon > 0$;
a fixed barrier update parameter θ, $0 < \theta < 1$;
begin
$x := \mathbf{1}$; $s := \mathbf{1}$; $\mu := 1$;
while $n\mu > \epsilon$ **do**
begin
$\mu := (1-\theta)\mu$;
$v := \sqrt{\frac{xs}{\mu}}$;
while $\Psi(v) > \tau$ **do**
begin
$x := x + \alpha\Delta x$;
$s := s + \alpha\Delta s$;
$y := y + \alpha\Delta y$;
$v := \sqrt{\frac{xs}{\mu}}$;
end
end
end

Figure 1.1: *The algorithm.*

Until recently, only algorithms based on the logarithmic barrier function were considered. In this case, where the proximity function is the scaled logarithmic barrier function, as given by (1.3.10), the algorithm has been well investigated (see, e.g., [Gon92; Her94; Jan94b; Koj89; Mon89; Tod89]). The corresponding complexity results can be summarized as follows.

Theorem 1.3.1 (cf. [Roo05]). *If the kernel function is given by (1.3.13) and* $\tau = O(1)$*, then the algorithm requires*

$$O\left(\sqrt{n}\log\frac{n}{\epsilon}\right) \tag{1.3.17}$$

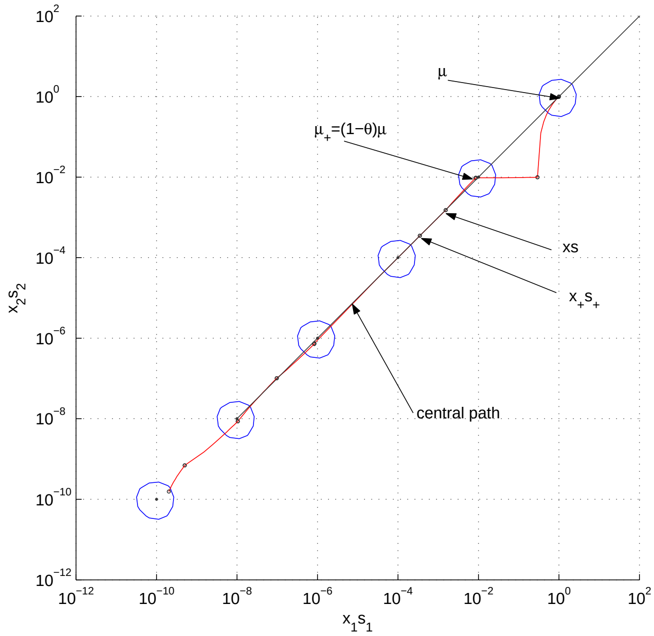


Figure 1.2: *Performance of a large-update IPM ($\theta = 0.99$).*

inner iterations if $\theta = \Theta\left(\frac{1}{\sqrt{n}}\right)$*, and*

$$O\left(n \log \frac{n}{\epsilon}\right)$$

inner iterations if $\theta = \Theta(1)$*. The output is a positive feasible pair* (x, s) *such that* $n\mu \leq \epsilon$ *and* $\Psi(v) = O(1)$*.*

As Theorem 1.3.1 makes clear, small-update methods theoretically have the best iteration complexity. Despite this, large-update methods are in practice much more efficient than small-update methods [And96]. This has been called the 'irony of IPMs [Ren01]. In fact, the observed iteration complexity of large-update methods is about $O\left(\log n \log \frac{n}{\epsilon}\right)$ in practice. This unpleasant gap between theory and practice has motivated many researchers to search for variants of large-update methods whose theoretical iteration complexity comes closer to what is observed in practice. As pointed out below, some progress has recently been made in this respect but, regrettably, it has to be admitted that we are still far from the desired result.

We proceed by describing some recent results. Note that if $\psi(t)$ is a kernel function then $\psi(1) = \psi'(1) = 0$, and hence $\psi(t)$ is completely determined by its

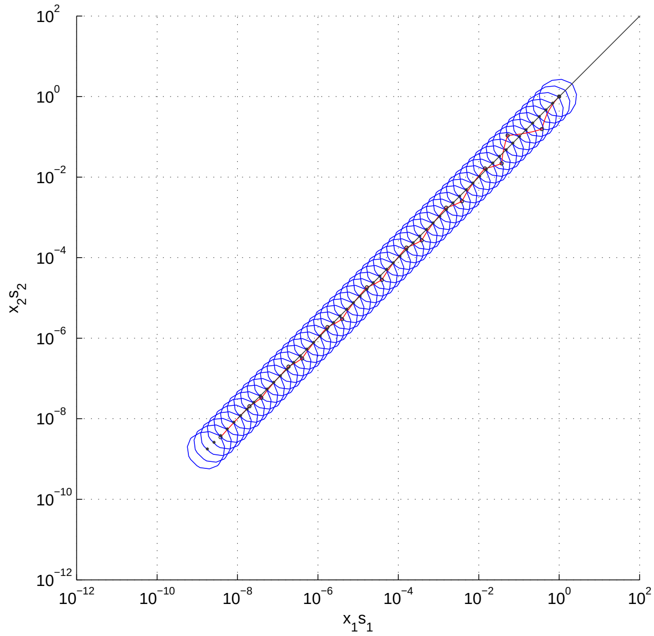


Figure 1.3: *Performance of a small-update IPM ($\theta = \frac{1}{\sqrt{2n}}$).*

second derivative:

$$\psi(t) = \int_1^t \int_1^\xi \psi''(\zeta)\, d\zeta d\xi. \tag{1.3.18}$$

In [Pen02a] the iteration complexity for large-update methods was improved to

$$O\left(\sqrt{n}\,(\log n)\log\frac{n}{\epsilon}\right), \tag{1.3.19}$$

which is currently the best result for such methods. This result was obtained by considering kernel functions that satisfy

$$\psi''(t) = \Theta(t^{p-1} + t^{-1-q}), \quad \forall t \in (0, \infty). \tag{1.3.20}$$

The analysis of an algorithm based on such a kernel function is greatly simplified if the kernel function also satisfies the following property:

$$\psi\left(\sqrt{t_1 t_2}\right) \le \tfrac{1}{2}\left[\psi\left(t_1\right) + \psi\left(t_2\right)\right], \quad \forall t_1, t_2 > 0. \tag{1.3.21}$$

The latter property has been given the name of *exponential convexity* (or shortly e-convexity) [Bai03a; Pen01]. In [Pen02a] kernel functions satisfying (1.3.20) and (1.3.21) were named *self-regular*. The best iteration complexity for large-update methods based on self-regular kernel functions is as given by (1.3.19) [Pen02a]. Subsequently, the same iteration complexity was obtained in [Pen01] in a more simple way for the specific self-regular function

$$\psi(t) = \frac{t^2-1}{2} + \frac{t^{1-q}-1}{q-1}, \quad q = \tfrac{1}{2}\log n.$$

1.4 The scope of this thesis

In this thesis we further explore the idea of IPMs based on kernel functions as described before.

In Chapter 2 we present a new class of barrier functions which are not necessary self-regular. This chapter is based on [Bai04a; Bai03a; Bai03b; Bai02b; Gha04b; Gha04a; Gha05a]. The proposed class is defined by some simple conditions on the kernel function and its first three derivatives. The best iteration bound for small- and large-update methods as given by (1.3.17) and (1.3.19) respectively are also achieved for kernel functions in this class.

In Chapter 3 we investigate the extension of primal-dual IPMs based on kernel functions studied in Chapter 2 to semidefinite optimization (SDO). The chapter is based on [Gha05b].

In Chapter 4 we report some numerical experiments. The aim of this section is to investigate the computational performance of IPMs based on various kernel functions. These tests indicate that the computational efficiency of an algorithm highly depends on the kernel function underlying the algorithm.

Finally, Chapter 5 contains some conclusions and recommendations for further research.

Chapter 2

Primal-Dual IPMs for LO Based on Kernel Functions

2.1 Introduction

As pointed out in Chapter 1, Peng, Roos, and Terlaky [Pen00a; Pen00b; Pen01; Pen02a; Pen02b] recently, introduced so-called self-regular barrier functions for primal-dual interior point methods (IPMs) for linear optimization. Each such barrier function is determined by its univariate self-regular kernel function. In this chapter we present a new class of barrier functions. The proposed class is defined by some simple conditions on the kernel function and its first three derivatives. As we will show, the currently best known bounds for both small- and large-update primal-dual IPMs are achieved by functions in the new class.

2.2 A new class of kernel functions

We call $\psi : (0, \infty) \to [0, \infty)$ a kernel function if ψ is twice differentiable and the following conditions are satisfied.

(i) $\psi'(1) = \psi(1) = 0$;

(ii) $\psi''(t) > 0, \quad \text{for all} \quad t > 0.$

In this chapter we restrict our selves to functions that are coercive, i.e.,

(iii) $\lim_{t \downarrow 0} \psi(t) = \lim_{t \to \infty} \psi(t) = \infty.$

Clearly, (i) and (ii) say that $\psi(t)$ is a nonnegative strictly convex function such that $\psi(1) = 0$. Recall from (1.3.18) that this implies that $\psi(t)$ is completely determined by its second derivative:

$$\psi(t) = \int_1^t \int_1^\xi \psi''(\zeta)\, d\zeta d\xi. \tag{2.2.1}$$

Moreover, by (iii), $\psi(t)$ has the so called barrier property. Having such a function $\psi(t)$, its definition is extended to positive n-dimensional vectors v by (1.3.12), thus yielding the induced (scaled) barrier function $\Psi(v)$. The barrier function induces primal-dual barrier search directions, by using (1.3.11) as the centering equation. In the sequel we also use the norm-based proximity measure $\delta(v)$ defined by

$$\delta(v) = \tfrac{1}{2} \left\| \nabla\Psi\left(v\right) \right\| = \tfrac{1}{2} \left\| d_x + d_s \right\|. \tag{2.2.2}$$

Note that

$$\Psi\left(v\right) = 0 \Leftrightarrow \delta\left(v\right) = 0 \Leftrightarrow v = e.$$

In this chapter we consider more conditions on the kernel function, namely $\psi \in C^3$ and

$$\begin{aligned}
t\psi''(t) + \psi'(t) &> 0, \quad t < 1, && \text{(2.2.3-a)}\\
t\psi''(t) - \psi'(t) &> 0, \quad t > 1, && \text{(2.2.3-b)}\\
\psi'''(t) &< 0, \quad t > 0, && \text{(2.2.3-c)}\\
2\psi''(t)^2 - \psi'(t)\psi'''(t) &> 0, \quad t < 1, && \text{(2.2.3-d)}\\
\psi''(t)\psi'(\beta t) - \beta\psi'(t)\psi''(\beta t) &> 0, \quad t > 1, \quad \beta > 1. && \text{(2.2.3-e)}
\end{aligned}$$

Condition (2.2.3-a) is obviously satisfied if $t \geq 1$, since then $\psi'(t) \geq 0$. Similarly, condition (2.2.3-b) is satisfied if $t \leq 1$, since then $\psi'(t) \leq 0$. Also (2.2.3-d) is satisfied if $t \geq 1$ since then $\psi'(t) \geq 0$, whereas $\psi'''(t) < 0$. We conclude that conditions (2.2.3-a) and (2.2.3-d) are conditions on the barrier behavior of $\psi(t)$. On the other hand, condition (2.2.3-b) deals only with $t \geq 1$ and hence concerns the growth behavior of $\psi(t)$. Condition (2.2.3-e) is technically more involved; we will discuss it later.

Remark 2.2.1. *It is worth pointing out that the conditions (2.2.3-a)-(2.2.3-d) are logically independent. Table 2.1 shows five kernel functions and the signs indicate whether a condition is satisfied* $(+)$ *or not* $(-)$. ■

The next two lemmas make clear that conditions (2.2.3-a) and (2.2.3-b) admit a nice interpretation.

Lemma 2.2.2 (Lemma 2.1.2 in [Pen02b]). *The following three properties are equivalent:*

$\psi(t)$	(2.2.3-a)	(2.2.3-b)	(2.2.3-c)	(2.2.3-d)	(2.2.3-e)
$\frac{t^2-1}{2}+\frac{e^{-\sigma(t-1)}-1}{\sigma},\ \ \sigma \geq 1.$	$-$	$+$	$+$	$+$	$+$
$t-1-\log t$	$+$	$-$	$+$	$+$	$+$
$t^3+t^{-3}-2$	$+$	$+$	$-$	$+$	$+$
$8t^2-11t+1+\frac{2}{\sqrt{t}}-4\log t$	$+$	$+$	$+$	$-$	$+$
$\frac{1}{2}(t+2)(t-1)-\log t$	$+$	$-$	$+$	$+$	$-$

Table 2.1: *The conditions (2.2.3-a)-(2.2.3-d) are logically independent.*

(i) $\psi\left(\sqrt{t_1 t_2}\right) \leq \frac{1}{2}\left(\psi\left(t_1\right)+\psi\left(t_2\right)\right)$, *for all* $t_1, t_2>0$*;*
(ii) $\psi'(t)+t\psi''(t) \geq 0, \quad t>0$*;*
(iii) $\psi\left(e^{\xi}\right)$ *is convex.*

Proof. *(iii)* $\Leftrightarrow$ *(i)*: From the definition of convexity, we know that $\psi(\exp(\zeta))$ is convex if and only if for any $\zeta_1, \zeta_2 \in \mathbf{R}$, the following inequality holds

$$\psi\left(\exp\left(\frac{1}{2}\left(\zeta_1+\zeta_2\right)\right)\right) \leq \frac{1}{2}\left(\psi(\exp\left(\zeta_1\right))+\psi(\exp\left(\zeta_2\right))\right).$$

Letting $t_1=\exp\left(\zeta_1\right), t_2=\exp\left(\zeta_2\right)$, obviously one has $t_1, t_2 \in(0,+\infty)$, and the above relation can be rewritten as

$$\psi(\sqrt{t_1 t_2}) \leq \frac{1}{2}\left(\psi(t_1)+\psi(t_2)\right).$$

(iii) $\Leftrightarrow$ *(ii)*: The function $\psi(\exp(\zeta))$ is convex if and only if the second derivative with respect to ζ is nonnegative. This gives $\exp(2\zeta)\,\psi''(\exp(\zeta)) + \exp(\zeta)\,\psi'(\exp(\zeta)) \geq 0$. Substituting $t=\exp(\zeta)$, one gets $t\psi'(t)+t^2\psi''(t) \geq 0$ which is equivalent to $\psi'(t)+t\psi''(t) \geq 0$ for $t>0$. This completes the proof of the lemma. □

Lemma 2.2.3. *Let* $\psi(t)$ *be a twice differentiable function for* $t>0$*. Then the following three properties are equivalent:*

(i) $\psi\left(\sqrt{\frac{t_1^2+t_2^2}{2}}\right) \leq \frac{1}{2}\left(\psi\left(t_1\right)+\psi\left(t_2\right)\right)$, *for* $t_1, t_2>0$*;*

(ii) $t\psi''(t) - \psi'(t) \geq 0, \quad t > 0$;

(iii) $\psi\left(\sqrt{\xi}\right)$ *is convex.*

Proof. $(iii) \Leftrightarrow (i)$: We know that $\psi\left(\sqrt{\xi}\right)$ is convex if and only if for any $\xi_1, \xi_2 \in \mathbf{R}_+$, the following inequality holds:

$$\psi\left(\sqrt{\frac{1}{2}(\xi_1+\xi_2)}\right) \leq \frac{1}{2}\left(\psi\left(\sqrt{\xi_1}\right) + \psi\left(\sqrt{\xi_2}\right)\right).$$

By letting $t_1 = \sqrt{\xi_1}$, $t_2 = \sqrt{\xi_2}$, the above relation can be equivalently rewritten as

$$\psi\left(\sqrt{\frac{t_1^2+t_2^2}{2}}\right) \leq \frac{1}{2}\left(\psi\left(t_1\right) + \psi\left(t_2\right)\right).$$

$(iii) \Leftrightarrow (ii)$: The second derivative of $\psi\left(\sqrt{\xi}\right)$ is nonnegative if and only if $\frac{1}{4\xi^{\frac{3}{2}}}\left(\sqrt{\xi}\psi''(\sqrt{\xi}) - \psi'\left(\sqrt{\xi}\right)\right) \geq 0$. Substituting $t = \sqrt{\xi}$ gives $\frac{1}{4t^3}\left(t\psi''(t) - \psi'(t)\right) \geq 0$, which is equivalent to $t\psi''(t) - \psi'(t) \geq 0$, for $t > 0$. □

Following [Bai03a], we call the property described in Lemma 2.2.2 *exponential convexity*, or shortly *e-convexity*. This property will turn out to be very useful in the analysis of primal-dual algorithms based on kernel functions.
In the next lemma we show that if $\psi(t)$ satisfies (2.2.3-b) and (2.2.3-c), then $\psi(t)$ also satisfies condition (2.2.3-e).

Lemma 2.2.4. *If $\psi(t)$ satisfies (2.2.3-b) and (2.2.3-c), then $\psi(t)$ satisfies (2.2.3-e).*

Proof. For $t > 1$ we consider

$$f(\beta) := \psi''(t)\psi'(\beta t) - \beta\psi'(t)\psi''(\beta t), \quad \beta \geq 1.$$

Note that $f(1) = 0$. Moreover,

$$\begin{aligned} f'(\beta) &= t\psi''(t)\psi''(\beta t) - \psi'(t)\psi''(\beta t) - \beta t\psi'(t)\psi'''(\beta t) \\ &= \psi''(\beta t)\left(t\psi''(t) - \psi'(t)\right) - \beta t\psi'(t)\psi'''(\beta t) > 0. \end{aligned}$$

The last inequality follows since $\psi''(\beta t) > 0$, $t\psi''(t) - \psi'(t) > 0$, by (2.2.3-b), and $-\beta t\psi'(t)\psi'''(\beta t) > 0$, since $t > 1$, which implies $\psi'(t) > 0$, and $\psi'''(\beta t) < 0$, by (2.2.3-c). Thus it follows that $f(\beta) > 0$ for $\beta > 1$, proving the lemma. □

As a preparation for later, we present in the next section some technical results for the new class of kernel functions.

2.2.1 Properties of kernel functions

Lemma 2.2.5. *One has*

$$t\psi'(t) \geq \psi(t), \quad \text{if } t \geq 1.$$

Proof. Defining $g(t) := t\psi'(t) - \psi(t)$ one has $g(1) = 0$ and $g'(t) = t\psi''(t) \geq 0$. Hence $g(t) \geq 0$ for $t \geq 1$ and the lemma follows. □

Lemma 2.2.6. *If ψ is a kernel function that satisfies (2.2.3-c), then*

$$\psi(t) > \frac{1}{2}\,(t-1)\,\psi'(t) \quad \text{and} \quad \psi'(t) > (t-1)\,\psi''(t), \quad \text{if} \quad t > 1,$$

$$\psi(t) < \frac{1}{2}\,(t-1)\,\psi'(t) \quad \text{and} \quad \psi'(t) > (t-1)\,\psi''(t), \quad \text{if} \quad t < 1.$$

Proof. Consider the function $f(t) = 2\psi(t) - (t-1)\,\psi'(t)$. Then $f(1) = 0$ and $f'(t) = \psi'(t) - (t-1)\,\psi''(t)$. Hence $f'(1) = 0$ and $f''(t) = -\,(t-1)\,\psi'''(t)$. Using that $\psi'''(t) < 0$ it follows that if $t > 1$ then $f''(t) > 0$, whence $f'(t) > 0$ and $f(t) > 0$, and if $t < 1$ then $f''(t) < 0$, so $f'(t) > 0$ and $f(t) < 0$. From this the lemma follows. □

Lemma 2.2.7. *If $\psi(t)$ satisfies (2.2.3-c), then*

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

$$\frac{1}{2}\psi''(1)\,(t-1)^2 < \psi(t) < \frac{1}{2}\psi''(t)\,(t-1)^2\,, \quad t < 1.$$

Proof. Using Taylor's theorem and $\psi(1) = \psi'(1) = 0$, we obtain

$$\psi(t) = \frac{1}{2}\psi''(1)(t-1)^2 + \frac{1}{3!}\psi'''(\xi)(t-1)^3, \quad \xi > 0.$$

Since $\psi'''(t) < 0$ the second inequality for $t > 1$ and the first inequality for $t < 1$ in the lemma follows. The remaining two inequalities are an immediate consequence of Lemma 2.2.6. □

Lemma 2.2.8. *Suppose that $\psi(t_1) = \psi(t_2)$, with $t_1 \leq 1 \leq t_2$ and $\beta \geq 1$. Then*

$$\psi(\beta t_1) \leq \psi(\beta t_2).$$

Equality holds if and only if $\beta = 1$ or $t_1 = t_2 = 1$.

Proof. Consider

$$f(\beta) := \psi(\beta t_2) - \psi(\beta t_1).$$

One has $f(1) = 0$ and

$$f'(\beta) = t_2 \psi'(\beta t_2) - t_1 \psi'(\beta t_1).$$

Since $\psi''(t) \geq 0$ for all $t > 0$, $\psi'(t)$ is monotonically non-decreasing. Hence $\psi'(\beta t_2) \geq \psi'(\beta t_1)$. Substitution gives

$$f'(\beta) = t_2 \psi'(\beta t_2) - t_1 \psi'(\beta t_1) \geq t_2 \psi'(\beta t_2) - t_1 \psi'(\beta t_2) = \psi'(\beta t_2)\,(t_2 - t_1) \geq 0.$$

The last inequality holds since $t_2 \geq t_1$, and $\psi'(t) \geq 0$ for $t \geq 1$. This proves that $f(\beta) \geq 0$ for $\beta \geq 1$, and hence the inequality in the lemma follows. If $\beta = 1$ then we obviously have equality. Otherwise, if $\beta > 1$, and $f(\beta) = 0$, then the mean value theorem implies $f'(\xi) = 0$ for some $\xi \in (1, \beta)$. But this implies $\psi'(\xi t_2) = \psi'(\xi t_1)$. Since $\psi'(t)$ is strictly monotonic, this implies $\xi t_2 = \xi t_1$, whence $t_2 = t_1$. Since also $t_1 \leq 1 \leq t_2$, we obtain $t_2 = t_1 = 1$. $\Box$

Lemma 2.2.9. *Suppose that $\psi(t_1) = \psi(t_2)$, with $t_1 \leq 1 \leq t_2$. Then $\psi'(t_1) \leq 0$ and $\psi'(t_2) \geq 0$, whereas*

$$-\psi'(t_1) \geq \psi'(t_2).$$

Proof. The lemma is obvious if $t_1 = 1$ or if $t_2 = 1$, because then $\psi(t_1) = \psi(t_2) = 0$ implies $t_1 = t_2 = 1$. We may therefore assume that $t_1 < 1 < t_2$. Since $\psi(t_1) = \psi(t_2)$, Lemma 2.2.7 implies:

$$\frac{1}{2}\,(t_1 - 1)^2\, \psi''(1) < \psi(t_1) = \psi(t_2) < \frac{1}{2}\,(t_2 - 1)^2\, \psi''(1).$$

Hence, since $\psi''(1) > 0$, it follows that $t_2 - 1 > 1 - t_1$. Using this and Lemma 2.2.6, while assuming $-\psi'(t_1) < \psi'(t_2)$, we may write

$$\begin{aligned} \psi(t_2) > \frac{1}{2}\,(t_2 - 1)\,\psi'(t_2) \quad &> \quad \frac{1}{2}\,(1 - t_1)\,\psi'(t_2) \\ &> \quad -\frac{1}{2}\,(1 - t_1)\,\psi'(t_1) = \frac{1}{2}\,(t_1 - 1)\,\psi'(t_1) > \psi(t_1). \end{aligned}$$

This contradiction proves the lemma. $\Box$

2.2.2 Ten kernel functions

By way of example we consider in this thesis ten kernel functions, as listed in Table 2.2. Note that some of these kernel functions depend on a parameter (e.g., $\psi_2(t)$ depends on the parameter $q > 1$), and hence when the parameter is not specified, it represents a whole class of kernel functions.

i	kernel functions ψ_i
1	$\frac{t^2-1}{2} - \log t$
2	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q(q-1)} - \frac{q-1}{q}(t-1), \quad q>1$
3	$\frac{t^2-1}{2} + \frac{(e-1)^2}{e}\frac{1}{e^t-1} - \frac{e-1}{e}$
4	$\frac{1}{2}\left(t-\frac{1}{t}\right)^2$
5	$\frac{t^2-1}{2} + e^{\frac{1}{t}-1} - 1$
6	$\frac{t^2-1}{2} - \int_1^t e^{\frac{1}{\xi}-1} d\xi$
7	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q-1}, \ q>1$
8	$t - 1 + \frac{t^{1-q}-1}{q-1}, \quad q>1$
9	$\frac{t^{1+p}-1}{1+p} - \log t, \ p \in [0,1]$
10	$\frac{t^{p+1}-1}{p+1} + \frac{t^{1-q}-1}{q-1}, \quad p \in [0,1], \quad q>1$

Table 2.2: *Ten kernel functions.*

The first proximity function, $\psi_1(t)$, gives rise to the classical primal-dual logarithmic barrier function and is a special case of $\psi_9(t)$, for $p = 1$. The second kernel function ψ_2 is the special case of the prototype self-regular kernel function [Pen02b],

$$\Upsilon_{p,q}(t) = \frac{t^{1+p}-1}{1+p} + \frac{t^{-q+1}-1}{q(q-1)} - \frac{q-1}{q}(t-1), \quad p,q \geq 1, \tag{2.2.4}$$

i	ψ_i'	ψ_i''
1	$t - \frac{1}{t}$	$1 + \frac{1}{t^2}$
2	$t - 1 - \frac{t^{-q}-1}{q}$	$1 + t^{-q-1}$
3	$t - \frac{e^t(e-1)^2}{e(e^t-1)^2}$	$1 + \frac{(e-1)^2 e^t \left(e^t+1\right)}{e(e^t-1)^3}$
4	$t - \frac{1}{t^3}$	$1 + \frac{3}{t^4}$
5	$t - \frac{e^{\frac{1}{t}-1}}{t^2}$	$1 + \frac{1+2t}{t^4}\, e^{\frac{1}{t}-1}$
6	$t - e^{\frac{1}{t}-1}$	$1 + \frac{e^{\frac{1}{t}-1}}{t^2}$
7	$t - t^{-q}$	$1 + qt^{-q-1}$
8	$1 - t^{-q}$	qt^{-q-1}
9	$t^p - \frac{1}{t}$	$pt^{p-1} + \frac{1}{t^2}$
10	$t^p - t^{-q}$	$pt^{p-1} + qt^{-q-1}$

Table 2.3: *First two derivatives of the ten kernel functions.*

for $p = 1$. The third kernel function has been studied in [Bai03b]. The fourth kernel function has been studied in [Pen00a]; one may easily verify that it is a special case of $\psi_7(t)$, when taking $q = 3$. The fifth and sixth kernel functions have been studied in [Bai04a]. The seventh kernel function has been studied in [Pen01; Pen02b]. Also note that $\psi_1(t)$ is the limiting value of $\psi_7(t)$ when q approaches 1.

In each of the first seven cases we can write $\psi(t)$ as

$$\psi(t) = \frac{t^2-1}{2} + \psi_b(t), \tag{2.2.5}$$

where $\frac{t^2-1}{2}$ is the so-called *growth term* and $\psi_b(t)$ the *barrier term* of the kernel function. The growth term dominates the behavior of $\psi(t)$ when t goes to infinity, whereas the barrier term dominates its behavior when t approaches zero. Note that in all cases the barrier term is monotonically decreasing in t.

The three last kernel functions in the table differ from the first seven others in that the *growth terms*, i.e., $t-1$, $\frac{t^{1+p}-1}{q+1}$ and $\frac{t^{1+p}-1}{q+1}$, respectively, are not quadratic in t. ψ_8 was first introduced and analyzed in [Bai04a], ψ_9 was analyzed in [Gha04a] and ψ_{10} has been studied for second order cone optimization in [Bai04b].

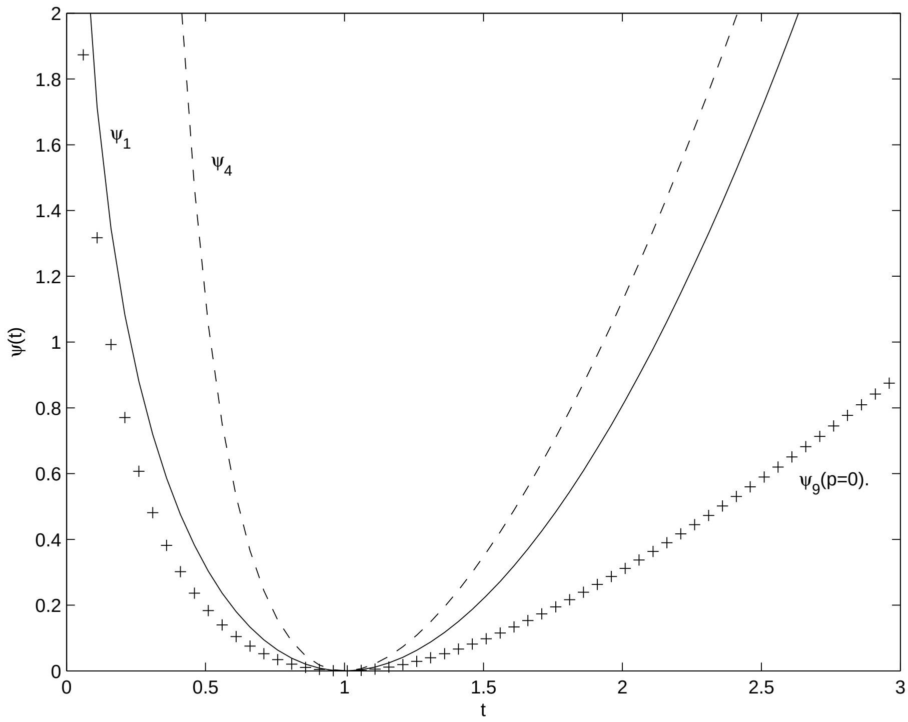


Figure 2.1: *Three different kernel functions.*

Figure 2.1 demonstrates the growth and barrier behavior of the three kernel functions ψ_1, ψ_4 and ψ_9 (with $p = 0$). From this figure we can see that the growth behaviors of ψ_1 and ψ_4 are quite similar as $t \longrightarrow \infty$, and that ψ_9 (with $p = 0$) grows much slower. However, when $t \longrightarrow 0$, ψ_1 and ψ_9 (with $p = 0$) are quite similar whereas ψ_4 grows much faster.

Now we proceed by showing that the ten kernel functions satisfy conditions (2.2.3-a), (2.2.3-c), (2.2.3-d), and (2.2.3-e). By using the information from Table 2.3 one may easily construct the entries in Table 2.4. It is almost obvious that all ten functions satisfy the condition (2.2.3-a) and from the second column in Table 2.5 we can see that the ten functions satisfy the condition (2.2.3-c). Also from the third column in Table 2.4 it is immediately seen that the first seven functions satisfy (2.2.3-b). Lemma 2.2.4 implies that the first seven functions satisfy also (2.2.3-e). The last column in Table 2.5 makes clear that ψ_8, ψ_9 and ψ_{10}, also satisfy (2.2.3-e). It remains to deal with (2.2.3-d). For this we use Table 2.6.

i	$t\psi_i''(t)+\psi_i'(t)$	$t\psi_i''(t)-\psi_i'(t)$
1	$2t$	$\frac{2}{t}$
2	$2t+\frac{q-1}{q}\left(t^{-q}-1\right)$	$\frac{(q+1)t^{-q}+q-1}{q}$
3	$2t+\frac{(e-1)^2}{e}\frac{(t+1)e^t+(t-1)e^{2t}}{(e^t-1)^3}$	$\frac{e^t(e-1)^2}{e(e^t-1)^2}\left(\frac{t\left(e^t+1\right)}{e^t-1}+1\right)$
4	$2t+\frac{2}{t^3}$	$\frac{4}{t^3}$
5	$2t+\frac{1+t}{t^3}e^{\frac{1}{t}-1}$	$\frac{1+3t}{t^3}e^{\frac{1}{t}-1}$
6	$2t+\frac{1-t}{t}e^{\frac{1}{t}-1}$	$\frac{1+t}{t}e^{\frac{1}{t}-1}$
7	$2t+(q-1)\,t^{-q}$	$(q+1)\,t^{-q}$
8	$1+(q-1)\,t^{-q}$	$-1+(q+1)\,t^{-q}$
9	$(1+p)\,t^p$	$(p-1)\,t^p+\frac{2}{t}$
10	$(p+1)\,t^p+(q-1)\,t^{-q}$	$(p-1)\,t^p+(q+1)\,t^{-q}$

Table 2.4: *The conditions (2.2.3-a) and (2.2.3-b).*

This table immediately shows that ψ_1, ψ_4, ψ_8, ψ_9, and ψ_{10} satisfy (2.2.3-d). The five remaining kernel functions also satisfy (2.2.3-d), as can be shown by simple, but rather technical arguments.

It may be noted that in [Pen02b] the kernel function $\psi(t)$ is defined to be self-regular if $\psi(t)$ is e-convex and, moreover,

$$\psi''(t)=\Theta\left(\Upsilon_{p,q}''(t)\right),$$

where $\Upsilon_{p,q}(t)$ was defined in (2.2.4). Since

$$\Upsilon_{p,q}''(t)=t^{p-1}+t^{-q-1},\quad \Upsilon_{p,q}'''(t)=(p-1)t^{p-2}-(q+1)t^{-q-2},$$

the prototype self-regular kernel function satisfies (2.2.3-c) only if $p\leq 1$. Note that the kernel functions ψ_2, ψ_4 and ψ_7 are self-regular.

It was observed in [Sal04a] that ψ_5 in Table 2.2 is the limit of the following sequence of functions

$$\psi_{(k)}(t)=\frac{t^2-1}{2}+\left(1+\frac{1}{k}\right)^{1-k}\left(\left(1+\frac{1}{kt}\right)^k-\left(1+\frac{1}{k}\right)^k\right),\qquad k=1,2,\ldots$$

i	$\psi_i'''(t)$	Condition (2.2.3-e)
1	$-\frac{2}{t^3}$	$\frac{2\left(\beta^2-1\right)}{\beta t}$
2	$-\left(q+1\right)t^{-q-2}$	$--$
3	$-\frac{e^t(e-1)^2}{e(e^t-1)^4}\left(e^{2t}+4e^t+1\right)$	$--$
4	$-\frac{12}{t^5}$	$\frac{4\left(\beta^4-1\right)}{\beta^3t^3}$
5	$-\frac{1+6t+6t^2}{t^6}e^{\frac{1}{t}-1}$	$--$
6	$-\frac{1+2t}{t^4}e^{\frac{1}{t}-1}$	$--$
7	$-q\left(q+1\right)t^{-q-2}$	$\frac{(q-1)\left(\beta^{q+1}-1\right)}{\beta^q t^q}$
8	$-q\left(q+1\right)t^{-q-2}$	$\frac{q\left(1-\beta^{-q}\right)}{t^{q+1}}$
9	$-p\left(1-p\right)t^{p-2}-\frac{2}{t^3}$	$\frac{t^p(1+p)\left(\beta^{p+1}-1\right)}{\beta t^2}$
10	$-p\left(1-p\right)t^{p-2}-q\left(q+1\right)t^{-q-2}$	$\frac{(p+q)(\beta^p-\beta^q)}{t^{q+1-p}}$

Table 2.5: *The conditions (2.2.3-c) and (2.2.3-e).*

By using Lemma 2.1.2 from [Pen02b], one can show that $\psi_{(k)}(t)$ is a S-R function for every $k \geq 1$. Furthermore, for any fixed $t > 0$, one has

$$\lim_{k\to\infty} \psi_{(k)}(t) = \frac{t^2-1}{2} + e^{\frac{1}{t}-1} - 1 = \psi_5(t).$$

This result implies that ψ_5 is the limit point of a sequence of S-R functions. Since ψ_5 itself is not S-R, it follows that the set of S-R functions is not closed. Note also that in our table only the first four kernel functions are S-R, and the two kernel functions ψ_8 and ψ_9 lie outside the closure of the set of S-R functions if $p < 1$.

2.3 Algorithm

In principle any kernel function gives rise to a primal-dual algorithm. The generic form of this algorithm is shown in Figure 1.1. The parameters τ, θ, and the step size α should be chosen in such a way that the algorithm is 'optimized' in the sense that the number of iterations required by the algorithm is as small as possible. Obviously, the resulting iteration bound will depend on the kernel function underlying the algorithm, and our main task becomes to find a kernel

i	$2\psi_i''(t)^2 - \psi_i'(t)\psi_i'''(t)$
1	$2 + \frac{6}{t^2}$
2	$2\left(1 + \frac{q}{t^{q+1}}\right)^2 + \frac{(q+1)}{t^{q+2}}\left(t - 1 + \frac{1-t^{-q}}{q}\right)$
3	$2\left(1 + \frac{(e-1)^2 e^t\left(e^t+1\right)}{e(e^t-1)^3}\right)^2 + \left(t - \frac{e^t(e-1)^2}{e(e^t-1)^2}\right)\left(\frac{e^t(e-1)^2}{e(e^t-1)^4}\left(e^{2t} + 4e^t + 1\right)\right)$
4	$2 + \frac{24}{t^4} + \frac{6}{t^8}$
5	$\frac{2}{t^8}\left(t^4 + (1+2t)\, e^{\frac{1}{t}-1}\right)^2 - \left(1 + 6t + 6t^2\right)\left(e^{\frac{1}{t}-1} - t^3\right) e^{\frac{1}{t}-1}$
6	$\frac{1}{t^4}\left(2\left(1 + e^{\frac{1}{t}-1}\right)^2 + (1+2t)\left(t - e^{\frac{1}{t}-1}\right) e^{\frac{1}{t}-1}\right)$
7	$2\left(1 + \frac{q}{t^{q+1}}\right)^2 + \frac{q(q+1)\left(t^{q+1}-1\right)}{t^{2(q+1)}}$
8	$q\left(q - 1 + (q+1)\, t^q\right) t^{-2(q+1)}$
9	$\frac{t^{1+p}\left(p^2+3p+2\right)+p(p+1)t^{2+2p}}{t^4}$
10	$\frac{p(p+1)t^{2p}+\left(q^2-p+4pq+p^2+q\right)t^{p-q}+q(q-1)t^{-2q}}{t^2}$

Table 2.6: *The condition (2.2.3-d).*

function that minimizes the iteration bound.

2.3.1 Upper bound for $\Psi(v)$ after each outer iteration

Note that at the start of each outer iteration of the algorithm, just before the update of μ, we have $\Psi(v) \le \tau$. By updating μ, the vector v is divided by $\sqrt{1-\theta}$, which generally leads to an increase in the value of $\Psi(v)$. Then, during the subsequent inner iterations, $\Psi(v)$ decreases until it passes the threshold τ again. Hence, during the course of the algorithm the largest values of $\Psi(v)$ occur just after the updates of μ. That is why in this section we derive an estimate for the effect of a μ-update on the value of $\Psi(v)$. In other words, with $\beta = \frac{1}{\sqrt{1-\theta}}$, we want to find an upper bound for $\Psi(\beta v)$ in terms of $\Psi(v)$.

It will become clear that in the analysis of the algorithm some inverse functions related to the underlying kernel functions and its first derivative play a crucial role. We introduce these inverse functions here.

We denote by $\varrho : [0,\infty) \to [1,\infty)$ and $\rho : [0,\infty) \to (0,1]$ the inverse functions of $\psi(t)$ for $t \geq 1$, and $-\frac{1}{2}\psi'(t)$ for $t \leq 1$, respectively. In other words

$$s = \psi(t) \quad \Leftrightarrow \quad t = \varrho(s), \quad t \geq 1, \tag{2.3.1}$$

$$s = -\tfrac{1}{2}\psi'(t) \quad \Leftrightarrow \quad t = \rho(s), \quad t \leq 1. \tag{2.3.2}$$

We have the following result.

Theorem 2.3.1. *For any positive vector v and any $\beta > 1$, we have*

$$\Psi(\beta v) \leq n\psi\left(\beta\varrho\left(\frac{\Psi(v)}{n}\right)\right).$$

Proof. We consider the following maximization problem:

$$\max_v \left\{\Psi(\beta v) \ : \ \Psi(v) = z\right\},$$

where z is any nonnegative number. The first order optimality conditions for this problem are

$$\beta\psi'(\beta v_i) = \lambda\psi'(v_i), \quad i = 1, \ldots, n, \tag{2.3.3}$$

where λ denotes the Lagrange multiplier. Since $\psi'(1) = 0$ and $\beta\psi'(\beta) > 0$, we must have $v_i \neq 1$ for all i. We even may assume that $v_i > 1$ for all i. To see this, let z_i be such that $\psi(v_i) = z_i$. Given z_i, this equation has two solutions: $v_i = v_i^{(1)} < 1$ and $v_i = v_i^{(2)} > 1$. As a consequence of Lemma 2.2.8 we have $\psi(\beta v_i^{(1)}) \leq \psi(\beta v_i^{(2)})$. Since we are maximizing $\Psi(\beta v)$, it follows that we may assume $v_i = v_i^{(2)} > 1$. This means that without loss of generality we may assume that $v_i > 1$ for all i. Note that then (2.3.3) implies $\beta\psi'(\beta v_i) > 0$ and $\psi'(v_i) > 0$, whence also $\lambda > 0$. Now defining

$$g(t) = \frac{\psi'(t)}{\psi'(\beta t)}, \quad t \geq 1,$$

we deduce from (2.3.3) that $g(v_i) = \frac{\beta}{\lambda}$ for all i. We proceed by showing that this implies that all v_i's are equal by proving that $g(t)$ is strictly monotonic. One has

$$g'(t) = \frac{\psi''(t)\psi'(\beta t) - \beta\psi'(t)\psi''(\beta t)}{\left(\psi'(\beta t)\right)^2}.$$

Using that $\psi(t)$ satisfies condition (2.2.3-e), we see that $g'(t) > 0$ for $t > 1$, since $\beta > 1$. Thus we have shown that $g(t)$ is strictly increasing. It thus follows that all v_i's are equal. Putting $v_i = t > 1$, for all i, we deduce from $\Psi(v) = z$ that

$n\psi(t) = z$. This implies that $t = \varrho(\frac{z}{n})$. Hence the maximal value that $\Psi(v)$ can attain is given by

$$\Psi(\beta t \mathbf{1}) = n\psi(\beta t) = n\psi\left(\beta\varrho\left(\frac{z}{n}\right)\right) = n\psi\left(\beta\varrho\left(\frac{\Psi(v)}{n}\right)\right).$$

This proves the theorem. □

Remark 2.3.2. *Note that the bound of Theorem 2.3.1 is sharp: one may easily verify that if $v = \beta\mathbf{1}$, with $\beta \geq 1$, then the bound holds with equality.* ■

As a result of Theorem 2.3.1 we have that if $\Psi(v) \leq \tau$ and $\beta = \frac{1}{\sqrt{1-\theta}}$ then

$$L_\psi(n, \theta, \tau) := n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right) \tag{2.3.4}$$

is an upper bound for $\Psi(\frac{v}{\sqrt{1-\theta}})$, the value of $\Psi(v)$ after the μ-update.

Corollary 2.3.3. *For any positive vector v and any $\beta > 1$, we have*

$$L_\psi(n, \theta, \tau) \leq \frac{n}{2}\,\psi''(1)\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}} - 1\right)^2.$$

Proof. Since $\frac{1}{\sqrt{1-\theta}} > 1$ and $\varrho(\frac{\tau}{n}) \geq 1$, the corollary follows from Theorem 2.3.1 by using Lemma 2.2.7. □

2.3.2 Decrease of the barrier function during an inner iteration

In this section, we compute a default step size α and the resulting decrease of the barrier function function. After a damped step we have

$$x_+ = x + \alpha\Delta x, \quad y_+ = y + \alpha\Delta y, \quad s_+ = s + \alpha\Delta s.$$

Hence, recalling from (1.3.5) and (1.3.16) that

$$v := \sqrt{\frac{xs}{\mu}}, \quad d_x := \frac{v\Delta x}{x}, \quad d_s := \frac{v\Delta s}{s},$$

we have

$$x_+ = x\left(e + \alpha\frac{\Delta x}{x}\right) = x\left(e + \alpha\frac{d_x}{v}\right) = \frac{x}{v}\left(v + \alpha d_x\right),$$

and

$$s_+ = s\left(e + \alpha\frac{\Delta s}{s}\right) = s\left(e + \alpha\frac{d_s}{v}\right) = \frac{s}{v}\left(v + \alpha d_s\right).$$

Thus we obtain, using $xs = \mu v^2$,

$$v_+^2 = \frac{x_+ s_+}{\mu} = (v + \alpha d_x)\,(v + \alpha d_s)\,. \tag{2.3.5}$$

Hence,

$$f(\alpha) := \Psi\,(v_+) - \Psi\,(v) = \psi\left(\sqrt{(v + \alpha d_x)\,(v + \alpha d_s)}\right) - \Psi\,(v)\,.$$

It is clear that $f(\alpha)$ is not necessarily convex[1] in α. To simplify the analysis we use a convex upper bound for $f(\alpha)$. Such a bound is obtained by using that $\psi(t)$ is e-convex. This gives

$$\begin{aligned}
\Psi\,(v_+) &= \Psi\left(\sqrt{(v + \alpha d_x)\,(v + \alpha d_s)}\right)\\
&= \sum_{i=1}^{n} \psi\left(\sqrt{(v_i + \alpha d_{xi})\,(v_i + \alpha d_{si})}\right)\\
&\leq \frac{1}{2}\left(\sum_{i=1}^{n} \psi\,(v_i + \alpha d_{xi}) + \sum_{i=1}^{n} \psi\,(v_i + \alpha d_{si})\right)\\
&= \tfrac{1}{2}\,(\Psi\,(v + \alpha d_x) + \Psi\,(v + \alpha d_s))\,.
\end{aligned}$$

Therefore $f(\alpha) \leq f_1(\alpha)$, where

$$f_1(\alpha) := \tfrac{1}{2}\,(\Psi\,(v + \alpha d_x) + \Psi\,(v + \alpha d_s)) - \Psi\,(v)\,,$$

which is convex in α, because $\Psi(v)$ is convex. Obviously,

$$f(0) = f_1(0) = 0.$$

Taking the derivative respect to α, we get

$$f_1'(\alpha) = \tfrac{1}{2}\sum_{i=1}^{n} \left(\psi'\,(v_i + \alpha d_{xi})\,d_{xi} + \psi'\,(v_i + \alpha d_{si})\,d_{si}\right).$$

This gives, using (1.3.11) and (2.2.2),

$$f_1'(0) = \tfrac{1}{2}\nabla\Psi(v)^T\,(d_x + d_s) = -\tfrac{1}{2}\nabla\Psi(v)^T\nabla\Psi(v) = -2\delta(v)^2. \tag{2.3.6}$$

Differentiating once more, we obtain

$$f_1''(\alpha) = \tfrac{1}{2}\sum_{i=1}^{n} \left(\psi''\,(v_i + \alpha d_{xi})\,{d_{xi}}^2 + \psi''\,(v_i + \alpha d_{si})\,{d_{si}}^2\right). \tag{2.3.7}$$

[1]Example: Let $\psi(t) = t + t^{-1} - 2 = \psi_8, (q = 2)$, and $n = 1$. For $v_1 = 1$, $d_{x1} = 1$, $d_{s1} = -\frac{1}{2}$, it is easy to verify that $\psi\left(\sqrt{(1+\alpha)\,(1 - \frac{1}{2}\alpha)}\right)$, is not convex.

Below we use the following notation:

$$v_1 := \min(v), \quad \delta := \delta(v).$$

Lemma 2.3.4. *One has* $f_1''(\alpha) \leq 2\delta^2 \, \psi''(v_1 - 2\alpha\delta)$.

Proof. Since d_x and d_s are orthogonal, (2.2.2) implies that $\|(d_x; d_s)\| = 2\delta$. Therefore, $\|d_x\| \leq 2\delta$ and hence $\|d_s\| \leq 2\delta$, and

$$v_i + \alpha d_{x\,i} \geq v_1 - 2\alpha\delta, \quad v_i + \alpha d_{s\,i} \geq v_1 - 2\alpha\delta, \quad 1 \leq i \leq n. \tag{2.3.8}$$

Due to (2.2.3-c), $\psi''(t)$ is monotonically decreasing, so from (2.3.7) we obtain

$$f_1''(\alpha) \leq \tfrac{1}{2}\, \psi''(v_1 - 2\alpha\delta) \sum_{i=1}^{n} \left({d_x}_i^2 + {d_s}_i^2 \right) = 2\delta^2 \, \psi''(v_1 - 2\alpha\delta).$$

This proves the lemma. □

Lemma 2.3.5. $f_1'(\alpha) \leq 0$ *holds if* α *satisfies the inequality*

$$-\psi'(v_1 - 2\alpha\delta) + \psi'(v_1) \leq 2\delta. \tag{2.3.9}$$

Proof. We may write, using Lemma 2.3.4, and also (2.3.6),

$$\begin{aligned} f_1'(\alpha) &= f_1'(0) + \int_0^\alpha f_1''(\xi)\, d\xi \\ &\leq -2\delta^2 + 2\delta^2 \int_0^\alpha \psi''(v_1 - 2\xi\delta)\, d\xi \\ &= -2\delta^2 - \delta \int_0^\alpha \psi''(v_1 - 2\xi\delta)\, d(v_1 - 2\xi\delta) \\ &= -2\delta^2 - \delta\left(\psi'(v_1 - 2\alpha\delta) - \psi'(v_1)\right). \end{aligned}$$

Hence, $f_1'(\alpha) \leq 0$ will certainly hold if α satisfies

$$-\psi'(v_1 - 2\alpha\delta) + \psi'(v_1) \leq 2\delta,$$

which proves the lemma. □

The next lemma uses the inverse function $\rho : [0, \infty) \to (0, 1]$ of $-\frac{1}{2}\psi'(t)$ for $t \in (0, 1]$, as introduced in (2.3.2).

Lemma 2.3.6. *The largest step size* α *that satisfies (2.3.9) is given by*

$$\bar{\alpha} := \frac{1}{2\delta}\left(\rho(\delta) - \rho(2\delta)\right). \tag{2.3.10}$$

Proof. We want α such that (2.3.9) holds, with α as large as possible. Since $\psi''(t)$ is decreasing, the derivative to v_1 of the expression at the left in (2.3.9) (i.e. $-\psi''(v_1 - 2\alpha\delta) + \psi''(v_1)$) is negative. Hence, fixing δ, the smaller v_1 is, the smaller α will be. One has

$$\delta = \tfrac{1}{2}\|\nabla\Psi(v)\| \geq \tfrac{1}{2}|\psi'(v_1)| \geq -\tfrac{1}{2}\psi'(v_1).$$

Equality holds if and only if v_1 is the only coordinate in v that differs from 1, and $v_1 \leq 1$ (in which case $\psi'(v_1) \leq 0$). Hence, the worst situation for the step size occurs when v_1 satisfies

$$-\tfrac{1}{2}\psi'(v_1) = \delta. \tag{2.3.11}$$

The derivative to α of the expression at the left in (2.3.9) equals

$$2\delta\psi''(v_1 - 2\alpha\delta) \geq 0,$$

and hence the left-hand side is increasing in α. So the largest possible value of α satisfying (2.3.9), satisfies

$$-\tfrac{1}{2}\psi'(v_1 - 2\alpha\delta) = 2\delta. \tag{2.3.12}$$

Due to the definition of ρ, (2.3.11) and (2.3.12) can be written as

$$v_1 = \rho(\delta), \quad v_1 - 2\alpha\delta = \rho(2\delta).$$

This implies,

$$\alpha = \frac{1}{2\delta}(v_1 - \rho(2\delta)) = \frac{1}{2\delta}(\rho(\delta) - \rho(2\delta)),$$

proving the lemma. □

Lemma 2.3.7. *Let $\bar{\alpha}$ be as defined in Lemma 2.3.6. Then*

$$\bar{\alpha} \geq \frac{1}{\psi''(\rho(2\delta))}. \tag{2.3.13}$$

Proof. By the definition of ρ,

$$-\psi'(\rho(\delta)) = 2\delta.$$

Taking the derivative to δ, we find

$$-\psi''(\rho(\delta))\,\rho'(\delta) = 2,$$

which implies that

$$\rho'(\delta) = -\frac{2}{\psi''(\rho(\delta))} < 0. \tag{2.3.14}$$

Hence ρ is monotonically decreasing in δ. An immediate consequence of (2.3.10) and (2.3.14) is

$$\bar{\alpha} = \frac{1}{2\delta}\int_{2\delta}^{\delta} \rho'(\sigma)\,d\sigma = \frac{1}{\delta}\int_{\delta}^{2\delta} \frac{d\sigma}{\psi''(\rho(\sigma))}. \tag{2.3.15}$$

To obtain a lower bound for $\bar{\alpha}$, we want to replace the argument of the last integral by its minimal value. So we want to know when $\psi''(\rho(\sigma))$ is maximal, for $\sigma \in [\delta, 2\delta]$. Due to (2.2.3-c), ψ'' is monotonically decreasing. So $\psi''(\rho(\sigma))$ is maximal when $\rho(\sigma)$ is minimal for $\sigma \in [\delta, 2\delta]$. Since ρ is monotonically decreasing this occurs when $\sigma = 2\delta$. Therefore

$$\bar{\alpha} = \frac{1}{\delta}\int_{\delta}^{2\delta} \frac{d\sigma}{\psi''(\rho(\sigma))} \geq \frac{1}{\delta}\frac{\delta}{\psi''(\rho(2\delta))} = \frac{1}{\psi''(\rho(2\delta))},$$

which proves the lemma. □

In the sequel we use the notation

$$\tilde{\alpha} = \frac{1}{\psi''(\rho(2\delta))}, \tag{2.3.16}$$

and we will use $\tilde{\alpha}$ as the default step size. By Lemma 2.3.7 we have $\bar{\alpha} \geq \tilde{\alpha}$.

Lemma 2.3.8. *If the step size α is such that $\alpha \leq \bar{\alpha}$ then*

$$f(\alpha) \leq -\alpha\,\delta^2. \tag{2.3.17}$$

Proof. Let $h(\alpha)$ be defined by

$$h(\alpha) := -2\alpha\delta^2 + \alpha\delta\psi'(v_1) - \frac{1}{2}\psi(v_1) + \frac{1}{2}\psi(v_1 - 2\alpha\delta).$$

Then

$$h(0) = f_1(0) = 0, \quad h'(0) = f_1'(0) = -2\delta^2, \quad h''(\alpha) = 2\delta^2\,\psi''(v_1 - 2\alpha\delta).$$

Due to Lemma 2.3.4, $f_1''(\alpha) \leq h''(\alpha)$. As a consequence, $f_1'(\alpha) \leq h'(\alpha)$ and $f_1(\alpha) \leq h(\alpha)$. Taking $\alpha \leq \bar{\alpha}$, with $\bar{\alpha}$ as defined in Lemma 2.3.6, we have

$$\begin{aligned} h'(\alpha) &= -2\delta^2 + 2\delta^2 \int_0^{\alpha} \psi''(v_1 - 2\xi\delta)\,d\xi \\ &= -2\delta^2 - \delta\left(\psi'(v_1 - 2\alpha\delta) - \psi'(v_1)\right) \leq 0. \end{aligned}$$

Since $h''(\alpha)$ is increasing in α, using Lemma A.1.3, we may write

$$f_1(\alpha) \leq h(\alpha) \leq \tfrac{1}{2}\alpha h'(0) = -\alpha\delta^2.$$

Since $f(\alpha) \leq f_1(\alpha)$, the proof is complete. □

By combining the results of Lemmas 2.3.7 and 2.3.8 we obtain

Theorem 2.3.9. *With $\tilde{\alpha}$ being the default step size, as given by (2.3.16), one has*

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''\left(\rho\left(2\delta\right)\right)}. \tag{2.3.18}$$

Lemma 2.3.10. *The right-hand side expression in (2.3.18) is monotonically decreasing in δ.*

Proof. Putting $t = \rho(2\delta)$, which implies $t \leq 1$, and which is equivalent to $4\delta = -\psi'(t)$, t is monotonically decreasing if δ increases. Hence, the right-hand expression in (2.3.18) is monotonically decreasing in δ if and only if the function

$$g(t) := \frac{\left(\psi'(t)\right)^2}{16\,\psi''(t)}$$

is monotonically decreasing for $t \leq 1$. Note that $g(1) = 0$ and

$$g'(t) = \frac{2\psi'(t)\psi''(t)^2 - \psi'(t)^2\psi'''(t)}{16\,\psi''(t)^2}.$$

Hence, since $\psi'(t) < 0$ for $t < 1$, $g(t)$ is monotonically decreasing for $t \leq 1$ if and only if

$$2\psi''(t)^2 - \psi'(t)\psi'''(t) \geq 0, \quad t \leq 1.$$

The last inequality is satisfied, due to condition (2.2.3-d). Hence the lemma is proved. □

Theorem 2.3.9 expresses the decrease of the barrier function value during a damped step, will step size $\tilde{\alpha}$, as a function of δ, and this function is monotonically decreasing in δ. In the sequel we need to express the decrease as a function of $\Psi(v)$. To this end we need a lower bound on $\delta(v)$ in terms of $\Psi(v)$. Such a bound is provided in the following section.

2.3.3 Bound on $\delta(v)$ in terms of $\Psi(v)$

The following theorem gives a lower bound of $\delta(v)$ in terms of $\Psi(v)$.

Theorem 2.3.11. *One has*

$$\delta(v) \geq \tfrac{1}{2}\psi'\left(\varrho\left(\Psi(v)\right)\right).$$

Proof. The statement in the lemma is obvious if $v = \mathbf{1}$ since then $\delta(v) = \Psi(v) = 0$. Otherwise we have $\delta(v) > 0$ and $\Psi(v) > 0$. To deal with the nontrivial case we consider, for $\omega > 0$, the problem

$$z_\omega = \min_v \left\{ \delta(v)^2 = \tfrac{1}{4}\sum_{i=1}^n \psi'(v_i)^2 \ : \ \Psi(v) = \omega \right\}.$$

The first order optimality condition are

$$\tfrac{1}{2}\psi'(v_i)\psi''(v_i) = \lambda\psi'(v_i), \quad i = 1, \ldots, n,$$

where $\lambda \in \mathbf{R}$ is the Lagrange multiplier. From this we conclude that we have either $\psi'(v_i) = 0$ or $\psi''(v_i) = 2\lambda$, for each i. Since $\psi''(t)$ is monotonically decreasing (2.2.3-c), this implies that all v_i's for which $\psi''(v_i) = 2\lambda$ have the same value. Denoting this value as t, and observing that all other coordinates have value 1 (since $\psi'(v_i) = 0$ for these coordinates), we conclude that, after reordering the coordinates, v has the form

$$v = (\underbrace{t; \ldots; t}_{k \text{ times}}; \underbrace{1; \ldots; 1}_{n-k \text{ times}}).$$

Now $\Psi(v) = \omega$ implies $k\psi(t) = \omega$. Given k, this uniquely determines $\psi(t)$, whence we have

$$4\delta(v)^2 = k\left(\psi'(t)\right)^2, \quad \psi(t) = \frac{\omega}{k}.$$

Note that the equation $\psi(t) = \frac{\omega}{k}$ has two solutions, one smaller than 1 and one larger than 1. By Lemma 2.2.9, the larger value gives the smallest value of $(\psi'(t))^2$. Since we are minimizing $\delta(v)^2$, we conclude that $t > 1$ (since $\omega > 0$). Hence we may write

$$t = \varrho\left(\frac{\omega}{k}\right),$$

where, as before, ϱ denotes the inverse function of $\psi(t)$ for $t \geq 1$. Thus we obtain that

$$4\delta(v)^2 = k\left(\psi'(t)\right)^2, \quad t = \varrho\left(\frac{\omega}{k}\right). \tag{2.3.19}$$

The question is now which value of k minimizes $\delta(v)^2$. To investigate this we take the derivative with respect to k of (2.3.19) extended to $k \in \mathbb{R}$. This gives

$$\frac{d\,4\delta(v)^2}{dk} = \left(\psi'(t)\right)^2 + 2k\psi'(t)\psi''(t)\frac{dt}{dk}. \tag{2.3.20}$$

From $\psi(t) = \frac{\omega}{k}$ we derive that

$$\psi'(t)\frac{dt}{dk} = -\frac{\omega}{k^2} = -\frac{\psi(t)}{k},$$

which gives

$$\frac{dt}{dk} = -\frac{\psi(t)}{k\psi'(t)}.$$

Substitution into (2.3.20) gives

$$\frac{d\,4\delta(v)^2}{dk} = \left(\psi'(t)\right)^2 - 2k\psi'(t)\psi''(t)\frac{\psi(t)}{k\psi'(t)} = \left(\psi'(t)\right)^2 - 2\psi(t)\psi''(t).$$

Defining $f(t) = \left(\psi'(t)\right)^2 - 2\psi(t)\psi''(t)$ we have $f(1) = 0$ and

$$f'(t) = 2\psi'(t)\psi''(t) - 2\psi'(t)\psi''(t) - 2\psi(t)\psi'''(t) = -2\psi(t)\psi'''(t) > 0.$$

We conclude that $f(t) > 0$ for $t > 1$. Hence $\frac{d\delta(v)^2}{dk} > 0$, so $\delta(v)^2$ increases when k increases. Since we are minimizing $\delta(v)^2$, at optimality we have $k = 1$. Also using that $\psi(t) \geq 0$, we obtain from (2.3.19) that

$$\min_v \left\{\delta(v) \;:\; \Psi(v) = \omega\right\} = \tfrac{1}{2}\psi'\left(t\right) = \tfrac{1}{2}\psi'\left(\varrho(\omega)\right) = \tfrac{1}{2}\psi'\left(\varrho\left(\Psi(v)\right)\right).$$

This completes the proof of the theorem. □

Remark 2.3.12. *The bound of Theorem 2.3.11 is sharp. One may easily verify that if v is such that all coordinates are equal to 1 except one coordinate which is greater than or equal to 1, then the bound holds with equality.* ■

Corollary 2.3.13. *One has*

$$\delta(v) \geq \frac{\Psi(v)}{2\,\varrho\left(\Psi(v)\right)}.$$

Proof. Using Theorem 2.3.11, i.e., $\delta(v) \geq \frac{1}{2}\psi'(\varrho(\Psi(v)))$, we obtain from Lemma 2.2.5 that

$$\delta(v) \geq \frac{1}{2}\psi'(\varrho(\Psi(v))) \geq \frac{\psi\left(\varrho(\Psi(v))\right)}{2\,\varrho\left(\Psi(v)\right)} = \frac{\Psi(v)}{2\,\varrho\left(\Psi(v)\right)}.$$

This proves the corollary. □

Remark 2.3.14. *It is worth noting that the proof of Theorem 2.3.11 implies that our kernel functions satisfy the inequality (cf. [[Pen02b, page 37])*

$$\psi'(t)^2 - 2\psi(t)\psi''(t) \geq 0, \quad t \geq 1.$$

■

Combining the results of Theorem 2.3.9 and Theorem 2.3.11 we obtain

$$f(\tilde{\alpha}) \leq -\frac{\left(\psi'\left(\varrho\left(\Psi(v)\right)\right)\right)^2}{4\psi''\left(\rho\left(\psi'\left(\varrho\left(\Psi(v)\right)\right)\right)\right)}. \tag{2.3.21}$$

This expresses the decrease in $\Psi(v)$ during an inner iteration completely in terms of the kernel function ψ, its first and second derivatives and the inverse functions ρ and ϱ.

To conclude this section we show in Table 2.7 the dependence of the results obtained so far on the conditions (2.2.3-a)–(2.2.3-e).

Conditions	Theorem (Th) or Lemma (L)
(2.2.3-a)	L 2.2.2
(2.2.3-b)	L 2.2.3, L 2.2.4
(2.2.3-c)	L 2.2.4, L 2.2.6, L 2.2.7, L 2.3.7, L 2.3.4, Th 2.3.11
(2.2.3-d)	L 2.3.10
(2.2.3-e)	Th 2.3.1

Table 2.7: *Use of conditions (2.2.3-a)-(2.2.3-e).*

2.4 Complexity

In this section we derive the complexity results for primal-dual interior point methods based on kernel functions satisfying the conditions (2.2.3-a), (2.2.3-c), (2.2.3-d) and (2.2.3-e).

After the update of μ to $(1-\theta)\mu$ we have, by Theorem 2.3.1 and (2.3.4),

$$\Psi(v_+) \leq L_\psi(n,\theta,\tau) = n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right). \tag{2.4.1}$$

We need to count how many inner iterations are required to return to the situation where $\Psi(v) \leq \tau$. We denote the value of $\Psi(v)$ after the μ-update as Ψ_0, and the subsequent values are denoted as Ψ_k, $k = 1, 2, \ldots$. A (negative) upper bound for the decrease during each inner iteration is provided by (2.3.21).

In the sequel we use that there exist a positive constants κ and γ, $\gamma \in [0,1]$ such that the right hand side expression in (2.3.21) satisfies

$$\frac{\left(\psi'\left(\varrho\left(\Psi(v)\right)\right)\right)^2}{4\psi''\left(\rho\left(\psi'\left(\varrho\left(\Psi(v)\right)\right)\right)\right)} \geq \kappa\Psi(v)^{1-\gamma}. \tag{2.4.2}$$

This holds because, since $\Psi(v) \geq \tau > 0$, $\gamma = 1$ and

$$\kappa = \frac{\left(\psi'\left(\varrho\left(\tau\right)\right)\right)^2}{4\psi''\left(\rho\left(\psi'\left(\varrho\left(\tau\right)\right)\right)\right)} \leq \frac{\left(\psi'\left(\varrho\left(\Psi(v)\right)\right)\right)^2}{4\psi''\left(\rho\left(\psi'\left(\varrho\left(\Psi(v)\right)\right)\right)\right)}.$$

satisfy (2.4.2). But our aim is to find smaller values of γ. The reason is this following lemma.

Lemma 2.4.1. *If K denotes the number of inner iterations, we have*

$$K \leq \frac{\Psi_0^\gamma}{\kappa\gamma}.$$

Proof. The definition of K implies $\Psi_{K-1} > \tau$ and $\Psi_K \leq \tau$ and

$$\Psi_{k+1} \leq \Psi_k - \kappa\Psi_k^{1-\gamma}, \qquad k = 0, 1, \cdots, K-1.$$

Yet we apply Lemma A.1.2, with $t_k = \Psi_k$. This yields the desired inequality. □

The last lemma provides an estimate for the number of inner-iterations in terms of Ψ_0 and the constants κ and γ. Recall that Ψ_0 is bounded above according to (2.4.1). An upper bound for the total number of iterations is obtained by multiplying (the upper bound for) the number K by the number of barrier parameter updates, which is bounded above by (cf. [Roo05, Lemma II.17, page 116])

$$\frac{1}{\theta}\log\frac{n}{\epsilon}.$$

Thus we obtain the following upper bound on the total number of iterations:

$$\frac{\Psi_0^\gamma}{\theta\kappa\gamma}\log\frac{n}{\epsilon} \leq \frac{1}{\theta\kappa\gamma}\left(n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right)\right)^\gamma \log\frac{n}{\epsilon}. \tag{2.4.3}$$

2.5 Application to the ten kernel functions

We apply the results of the previous sections, especially the upper bound (2.4.3) for the total number of iterations, to obtain iteration bounds for large-update and small-update methods based on the ten kernel functions ψ_i $i \in \{1, \ldots, 10\}$ introduced before. Since these kernel functions satisfy the conditions (2.2.3-a),(2.2.3-c), (2.2.3-d) and (2.2.3-e), we may simply apply the scheme in Figure 2.2 to each of the ten kernel functions. The subsequent steps in this scheme are justified by earlier results as indicated in Table 2.8.

Step	Based on
1	Equation (2.3.2)
2	Theorem 2.3.9
3	Equation (2.3.1)
4	Theorem 2.3.11
5	Step 3 and step 4
6	Theorem 2.3.1
7	Equation (2.4.3)
8	Small- and large-updates methods

Table 2.8: *Justification of the validity of the scheme in Figure 2.2.*

Using the scheme in Figure 2.2, our aim is to compute iteration bounds for large- and small-update methods based on the ten kernel functions. Large-update methods are characterized by $\tau = O(n)$ and $\theta = \Theta(1)$. It may be noted that we could also take smaller values of τ, e.g., $\tau = O(1)$, but one may easily check from the outcome of our analysis that this would not affect the order of magnitude of the bounds. Small-update methods are characterized by $\tau = O(1)$ and $\theta = \Theta\left(\frac{1}{\sqrt{n}}\right)$.

2.5.1 Some technical lemmas

Before dealing with each of the functions separately we derive some lemmas that will turn out to be useful. This is especially true if the inverse functions ϱ and ρ cannot be computed explicitly.

Step 0: Specify a kernel function $\psi(t)$; an update parameter θ, $0 < \theta < 1$; a threshold parameter τ; and an accuracy parameter ϵ.

Step 1: Solve the equation $-\frac{1}{2}\psi'(t) = s$ to get $\rho(s)$, the inverse function of $-\frac{1}{2}\psi'(t), t \in (0, 1]$. If the equation is hard to solve, derive a lower bound for $\rho(s)$.

Step 2: Calculate the decrease of $\Psi(v)$ during an inner iteration in terms of δ for the default step size $\tilde{\alpha}$ from

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''(\rho(2\delta))}.$$

Step 3: Solve the equation $\psi(t) = s$ to get $\varrho(s)$, the inverse function of $\psi(t), t \geq 1$. If the equation is hard to solve, derive lower and upper bounds for $\varrho(s)$.

Step 4: Derive a lower bound for δ in terms of $\Psi(v)$ by using

$$\delta(v) \geq \tfrac{1}{2}\psi'\left(\varrho\left(\Psi(v)\right)\right).$$

Step 5: Using the results of step 3 and step 4 find a valid inequality of the form

$$f(\tilde{\alpha}) \leq -\kappa \Psi(v)^{1-\gamma}$$

for some positive constants κ and γ, with $\gamma \in (0, 1]$ as small as possible.

Step 6: Calculate the upper bound of Ψ_0 from

$$\Psi_0 \leq L_\psi(n, \theta, \tau) = n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right) \leq \frac{n}{2}\psi''(1)\left(\frac{\varrho(\frac{\tau}{n})}{\sqrt{1-\theta}} - 1\right)^2.$$

Step 7: Derive an upper bound for the total number of iterations by using that

$$\frac{\Psi_0^\gamma}{\theta\kappa\gamma}\log\frac{n}{\epsilon}$$

is an upper bound for this number.

Step 8: Set $\tau = O(n)$ and $\theta = \Theta(1)$ to calculate a complexity bound for large-update methods, and set $\tau = O(1)$ and $\theta = \Theta(\frac{1}{\sqrt{n}})$ to calculate a complexity bound for small-update methods.

Figure 2.2: *Scheme for analyzing a kernel-function-based algorithm.*

Lemma 2.5.1. *When $\psi(t) = \psi_i(t)$ and $1 \le i \le 7$, then*

$$\sqrt{1+2s} \le \varrho(s) \le 1 + \sqrt{2s}. \tag{2.5.1}$$

Proof. The inverse function of $\psi(t)$ for $t \in [1, \infty)$ is obtained by solving t from the equation $\psi(t) = s$, for $t \ge 1$. In almost all cases it is hard to solve this equation explicitly. However, we can easily find a lower and an upper bound for t and this suffices for our goal. First one has

$$s = \psi(t) = \frac{t^2-1}{2} + \psi_b(t) \le \frac{t^2-1}{2},$$

where $\psi_b(t)$ denotes the barrier term. The inequality is due to the fact that $\psi_b(1) = 0$ and $\psi_b(t)$ is monotonically decreasing. It follows that

$$t = \varrho(s) \ge \sqrt{1+2s}.$$

For the second inequality we derive from (2.2.1) and $\psi''(t) \ge 1$ that

$$s = \psi(t) = \int_1^t \int_1^\xi \psi''(\zeta)\, d\zeta d\xi \ge \int_1^t \int_1^\xi d\zeta d\xi = \frac{1}{2}(t-1)^2,$$

which implies

$$t = \varrho(s) \le 1 + \sqrt{2s}.$$

This completes the proof. □

Lemma 2.5.2. *When $\psi(t) = \psi_i(t)$ with $i \in \{8, 10\}$, and $q \ge 2$, then*

$$t \le 1 + \sqrt{t\psi(t)}, \quad t \ge 1.$$

Proof. Defining $f(t) = t\psi(t) - (t-1)^2$ one has $f(1) = 0$ and $f'(t) = \psi(t) + t\psi'(t) - 2(t-1)$. Hence $f'(1) = 0$ and $f''(t) = 2\psi'(t) + t\psi''(t) - 2$. Since $f''(t) = (q-2)t^{-q} + pt^p + 2\,(t^p - 1) \ge 0$ for $\psi_8(t)$, and $f''(t) = (q-2)t^{-q} \ge 0$ for $\psi_{10}(t)$, the lemma follows. □

Lemma 2.5.3. *Let $1 \le i \le 7$. Then one has*

$$L_\psi(n, \theta, \tau) \le \frac{\psi''(1)}{2} \frac{\left(\sqrt{2\tau} + \theta\sqrt{n}\right)^2}{1-\theta}.$$

Hence, if $\tau = O(1)$ and $\theta = \Theta(\frac{1}{\sqrt{n}})$, then $\Psi_0 = O(\psi''(1))$.

Proof. By Lemma 2.5.1 we have $\varrho(s) \leq 1 + \sqrt{2s}$. Hence, also using (2.4.1) we have

$$L_\psi(n,\theta,\tau) = n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right) \leq n\psi\left(\frac{1+\sqrt{\frac{2\tau}{n}}}{\sqrt{1-\theta}}\right).$$

Applying Lemma 2.2.7 we obtain

$$\begin{aligned} L_\psi(n,\theta,\tau) \leq \frac{n\psi''(1)}{2}\left(\frac{1+\sqrt{\frac{2\tau}{n}}}{\sqrt{1-\theta}} - 1\right)^2 &\leq \frac{n\psi''(1)}{2}\left(\frac{\theta+\sqrt{\frac{2\tau}{n}}}{\sqrt{1-\theta}}\right)^2 \\ &= \frac{\psi''(1)}{2}\frac{\left(\sqrt{2\tau}+\theta\sqrt{n}\right)^2}{1-\theta}, \end{aligned}$$

where we also used

$$1 - \sqrt{1-\theta} = \frac{\theta}{1+\sqrt{1-\theta}} \leq \theta. \tag{2.5.2}$$

This proves the lemma. □

Lemma 2.5.4. *Let* $\underline{\rho} : [0,\infty) \to (0,1]$ *be the inverse function of the restriction of* $-\psi_b'(t)$ *to the interval* $(0,1]$. *When* $\psi(t) = \psi_i(t)$ *and* $1 \leq i \leq 7$, *then*

$$\rho(s) \geq \underline{\rho}(1+2s).$$

Proof. Let $t = \rho(s)$. Due to the definition of ρ as the inverse function of $-\frac{1}{2}\psi'(t)$ for $t \leq 1$ this means that

$$-2s = \psi'(t) = t + \psi_b'(t), \quad t \leq 1.$$

Since $t \leq 1$ this implies

$$-\psi_b'(t) = t + 2s \leq 1 + 2s.$$

Since $-\psi_b'(t)$ is monotonically decreasing in all seven cases, it follows from this that

$$t = \rho(s) \geq \underline{\rho}(1+2s),$$

proving the lemma. □

One more remark is in order. At the start of each inner iteration we have $\Psi(v) \geq \tau$. We will always assume that τ is large enough to guarantee that then $\delta(v) \geq 1$.

Lemma 2.5.5. *Let $\psi(t) = \psi_i(t)$ with $1 \leq i \leq 7$. If $\tau \geq 12$ then $\delta(v) \geq 1$ at the start of each inner iteration.*

Proof. By Corollary 2.3.13 we have $\delta(v) \geq \frac{\Psi(v)}{2\varrho(\Psi(v))}$. Using the right-hand side of (2.5.1), we get

$$\delta(v) \geq \frac{\Psi(v)}{2\varrho\left(\Psi(v)\right)} \geq \frac{\Psi(v)}{2\left(1+\sqrt{2\Psi(v)}\right)} \geq \frac{\tau}{2\left(1+\sqrt{2\tau}\right)} \geq \frac{12}{2\left(1+\sqrt{24}\right)} \approx 1.01... > 1.$$

Thus the lemma follows. □

We will use the same argument as in Lemma 2.5.5 $\psi(t) = \psi_i(t)$ and $8 \leq i \leq 10$, by changing the upper bound of ϱ. This upper bound will be derived in the next section.

2.5.2 Analysis of the ten kernel functions

Example 1. Consider the case where $\psi(t) = \psi_1(t)$:

$$\psi(t) = \frac{t^2-1}{2} - \log t.$$

Step 1: The inverse function $t = \rho(s)$ of $-\frac{1}{2}\psi'(t)$ for $t \in (0, 1]$ follows by solving t from the equation $-\frac{1}{2}\psi'(t) = s$:

$$-\left(t - \frac{1}{t}\right) = 2s.$$

This enables us to compute $\rho(s)$ exactly:

$$t = \rho(s) = \frac{1}{s + \sqrt{1+s^2}}.$$

Step 2: It follows that for the default step size $\tilde{\alpha}$ we have

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''\left(\rho(2\delta)\right)} = -\frac{\delta^2}{1 + \frac{1}{\rho(2\delta)^2}} = -\frac{\delta^2}{1 + \left(2\delta + \sqrt{1+4\delta^2}\right)^2} \leq -\frac{1}{19}.$$

For the last inequality we assumed that $\delta \geq 1$.
Step 3 and **Step 4**: are not needed.
Step 5: We have

$$f(\alpha) \leq -\kappa\psi(v)^{1-\gamma},$$

with $\kappa = \frac{1}{19}$ and $\gamma = 1$.

Step 6: We have to derive an upper bound for $\Psi(v)$ just after a μ-update. By using Lemma 2.5.3, with $\psi''(1) = 2$, we obtain

$$\Psi_0 \leq \frac{\left(\theta\sqrt{n} + \sqrt{2\tau}\right)^2}{1-\theta}.$$

Step 7: The total number of iterations is bounded above by

$$19\frac{\left(\theta\sqrt{n} + \sqrt{2\tau}\right)^2}{\theta(1-\theta)} \log\frac{n}{\epsilon}.$$

Step 8: For large-update methods (with $\tau = O(n)$ and $\theta = \Theta(1)$) the right hand side expression is $O\left(n \log\frac{n}{\epsilon}\right)$. For small-update methods (with $\tau = O(1)$ and $\theta = \Theta\left(\frac{1}{\sqrt{n}}\right)$) the right hand side expression is $O\left(\sqrt{n}\log\frac{n}{\epsilon}\right)$.

Note that in this case $\psi(t)$ is the kernel function of the logarithmic barrier function and that the iteration bounds agree with the bounds in Theorem 1.3.1.

Example 2. Consider the case where $\psi(t) = \psi_2(t)$

$$\psi(t) = \frac{t^2-1}{2} + \frac{t^{1-q}-1}{q(q-1)} - \frac{q-1}{q}(t-1), \qquad q > 1.$$

Step 1: The inverse function of $-\psi_b'(t) = 1 + \frac{t^{-q}-1}{q}$ is given by $\underline{\rho}(s) = \frac{1}{(1+q(s-1))^{\frac{1}{q}}}$. Hence, by Lemma 2.5.4,

$$\rho(s) \geq \frac{1}{(1+2qs)^{\frac{1}{q}}}.$$

Step 2: It follows that

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''\left(\rho(2\delta)\right)} = -\frac{\delta^2}{1+\frac{1}{\rho(2\delta)^{q+1}}} \leq -\frac{\delta^2}{1+(1+4q\delta)^{\frac{q+1}{q}}}.$$

Step 3: By Lemma 2.5.1 the inverse function of $\psi(t)$ for $t \in [1,\infty)$ satisfies

$$\sqrt{1+2s} \leq \varrho\left(s\right) \leq 1 + \sqrt{2s}.$$

Thus we have, omitting the argument v,

$$\varrho\left(\Psi(v)\right) \geq \sqrt{1+2\Psi}.$$

Step 4: Now using that $\delta(v) \geq \frac{1}{2}\psi'\left(\varrho\left(\Psi(v)\right)\right)$, and assuming $\Psi \geq \tau \geq 1$, we obtain

$$\delta \geq \frac{1}{2}\left(\sqrt{1+2\Psi} - 1 + \frac{1}{q}\left(1 - \frac{1}{(1+2\Psi)^q}\right)\right) \geq \frac{1}{2}\left(\sqrt{1+2\Psi}-1\right) = \frac{\Psi}{1+\sqrt{1+2\Psi}}.$$

Step 5: Substitution this, after some elementary reductions we arrive at

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{1+(1+4q\delta)^{\frac{q+1}{q}}} \leq -\frac{1}{53\,q}\Psi^{\frac{q-1}{2q}}.$$

Thus it follows that

$$\Psi_{k+1} \leq \Psi_k - \kappa {\Psi_k}^{1-\gamma}, \qquad k = 0, 1, \cdots, K-1,$$

with $\kappa = \frac{1}{53\,q}$ and $\gamma = \frac{q+1}{2q}$, and where K denotes the number of inner iterations. Hence the number K of inner iterations is bounded above by

$$K \leq \frac{\Psi_0^{\gamma}}{\kappa\gamma} = \frac{106\,q^2}{q+1}\,\Psi_0^{\frac{q+1}{2q}} \leq 106\,q\,\Psi_0^{\frac{q+1}{2q}}.$$

Step 6: To estimate Ψ_0 we use Lemma 2.5.3, with $\psi''(1) = 2$. Thus we obtain,

$$\Psi_0 \leq \frac{\left(\theta\sqrt{n}+\sqrt{2\tau}\right)^2}{1-\theta},$$

Step 7: The total number of iterations is bounded above by

$$\frac{106\,q}{\theta}\left(\frac{\left(\theta\sqrt{n}+\sqrt{2\tau}\right)^2}{1-\theta}\right)^{\frac{q+1}{2q}} \log\frac{n}{\epsilon}.$$

Step 8: For large-update methods (with $\tau = O(n)$ and $\theta = \Theta(1)$) the right hand side expression is

$$O\left(q\,n^{\frac{q+1}{2q}}\,\log\frac{n}{\epsilon}\right).$$

For small-update methods (with $\tau = O(1)$ and $\theta = \Theta\left(\frac{1}{\sqrt{n}}\right)$) the right hand side expression is $O\left(q\sqrt{n}\,\log\frac{n}{\epsilon}\right)$.

Example 3. Consider the case where $\psi(t) = \psi_3(t)$:

$$\psi(t) = \frac{t^2-1}{2} + \frac{(e-1)^2}{e}\frac{1}{e^t-1} - \frac{e-1}{e}$$

Step 1: In this example its hard to calculate explicitly the lower bound of the inverse function of $-\frac{1}{2}\psi'(t) = \frac{e^t(e-1)^2}{e(e^t-1)^2} - t$. Letting $t = \rho(2\delta)$, we have $-\psi'(t) = 4\delta$, since $\rho : [0,\infty) \to (0,1]$ denotes the inverse function of the restriction of $-\frac{1}{2}\psi'(t)$ to the interval $(0,1]$. Therefore, we may write $\frac{e^t(e-1)^2}{e(e^t-1)^2} - t = 4\delta$, then we have

$$\frac{e^t}{(e^t-1)^2} = \frac{e}{(e-1)^2}(4\delta+t) \leq \frac{e}{(e-1)^2}(4\delta+1).$$

Step 2: It follows that

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''(t)} = -\frac{\delta^2}{1+\frac{(e-1)^2 e^t (e^t+1)}{e(e^t-1)^3}} = -\frac{\delta^2}{1+\frac{(e-1)^2}{e}\left(e^{\frac{1}{2}t}+e^{-\frac{1}{2}t}\right)\left(\frac{e^t}{(e^t-1)^2}\right)^{\frac{3}{2}}}$$

By substitution we get

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{1+\frac{\sqrt{e}}{(e-1)}\left(e^{\frac{1}{2}t}+e^{-\frac{1}{2}t}\right)(4\delta+1)^{\frac{3}{2}}}.$$

Note that $0 < t \leq 1$. Hence $e^{\frac{1}{2}t}+e^{-\frac{1}{2}t} \leq e^{\frac{1}{2}}+e^{-\frac{1}{2}} = 2.2553 < \frac{5}{2}$, and since $\frac{\sqrt{e}}{(e-1)} = 0.959517 < 1$. Thus we conclude that

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{1+\frac{5}{2}(4\delta+1)^{\frac{3}{2}}} \leq -\frac{\delta^{\frac{1}{2}}}{37}.$$

For the last inequality we assumed that $\delta \geq 1$.
Step 3: Using the same argument as in Example 2, step 3, we have

$$\varrho\left(\Psi(v)\right) \geq \sqrt{1+2\Psi}.$$

Step 4: Now using that $\delta(v) \geq \frac{1}{2}\psi'\left(\varrho\left(\Psi(v)\right)\right)$, and assuming $\Psi \geq \tau \geq 1$, we obtain

$$\delta \geq \frac{1}{2}\left(\sqrt{1+2\Psi} - \frac{e^{\sqrt{1+2\Psi}}(e-1)^2}{e\left(e^{\sqrt{1+2\Psi}}-1\right)^2}\right) \geq \frac{1}{2}\left(\sqrt{1+2\Psi}-1\right) = \frac{\Psi}{1+\sqrt{1+2\Psi}}.$$

Step 5: Substitution this, after some elementary reductions we arrive at

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{1+\frac{5}{2}(4\delta+1)^{\frac{3}{2}}} \leq -\frac{\Psi^{\frac{1}{4}}}{74}.$$

Thus it follows that

$$\Psi_{k+1} \leq \Psi_k - \kappa {\Psi_k}^{1-\gamma}, \qquad k = 0, 1, \cdots, K-1,$$

with $\kappa = \frac{1}{74}$ and $\gamma = \frac{3}{4}$, and where K denotes the number of inner iterations. Hence the number K of inner iterations is bounded above by

$$K \leq \frac{\Psi_0^{\gamma}}{\kappa\gamma} = \frac{296}{3}\Psi_0^{\frac{3}{4}} \leq 99\Psi_0^{\frac{3}{4}}.$$

Step 6: To estimate Ψ_0 we use Lemma 2.5.3, with $\psi''(1) = 1 + \frac{e+1}{e-1}$. Thus we obtain,

$$\Psi_0 \leq \frac{\psi''(1)}{2} \frac{\left(\theta\sqrt{n} + \sqrt{2\tau}\right)^2}{1-\theta},$$

Step 7: The total number of iterations is bounded above by

$$\frac{99}{\theta} \left(\frac{\left(\theta\sqrt{n} + \sqrt{2\tau}\right)^2}{1-\theta} \right)^{\frac{3}{4}} \log \frac{n}{\epsilon}.$$

Step 8: For large-update methods (with $\tau = O(n)$ and $\theta = \Theta(1)$) the right hand side expression is $O\left(n^{\frac{3}{4}} \log \frac{n}{\epsilon}\right)$.

For small-update methods (with $\tau = O(1)$ and $\theta = \Theta\left(\frac{1}{\sqrt{n}}\right)$) the right hand side expression is $O\left(\sqrt{n} \log \frac{n}{\epsilon}\right)$.

Example 4. Consider the case where $\psi(t) = \psi_4(t)$:

$$\psi(t) = \frac{1}{2}\left(t - \frac{1}{t}\right)^2 = \frac{t^2-1}{2} + \frac{t^{-2}-1}{2}.$$

Step 1: The inverse function of $-\psi_b'(t) = \frac{1}{t^3}$ is given by $\underline{\rho}(s) = \frac{1}{s^{\frac{1}{3}}}$. Hence, by Lemma 2.5.4,

$$\rho(s) \geq \frac{1}{(1+2s)^{\frac{1}{3}}}.$$

Step 2: It follows that for the default step size $\tilde{\alpha}$ we have

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''(\rho(2\delta))} = -\frac{\delta^2}{1 + \frac{3}{\rho(2\delta)^4}} \leq -\frac{\delta^2}{1 + 3\,(1+4\delta)^{\frac{4}{3}}} \leq -\frac{\delta^{\frac{2}{3}}}{27}.$$

For the last inequality we assumed that $\delta \geq 1$.

Step 3: One may easily versify that the inverse function of $\psi(t)$ for $t \in [1, \infty)$ is given by

$$\varrho\,(s) = \sqrt{\frac{s}{2}} + \sqrt{1 + \frac{s}{2}}. \qquad (2.5.3)$$

Omitting the argument v, we have

$$\varrho\,(\Psi(v)) = \sqrt{\frac{\Psi}{2}} + \sqrt{1 + \frac{\Psi}{2}} \geq 2\sqrt{\frac{\Psi}{2}} = \sqrt{2\Psi}.$$

Step 4: Now using that $\delta(v) \geq \frac{1}{2}\psi'\left(\varrho\left(\Psi(v)\right)\right)$, and assuming $\Psi \geq \tau \geq 1$, we obtain

$$\delta \geq \frac{1}{2}\left(\sqrt{2\Psi} - \frac{1}{\left(\sqrt{2\Psi}\right)^3}\right) \geq 0.5303\Psi^{\frac{1}{2}} \geq \frac{1}{2}\Psi^{\frac{1}{2}}.$$

Step 5: Substitution gives

$$f(\tilde{\alpha}) \leq -\frac{\delta^{\frac{2}{3}}}{27} \leq -\frac{\left(0.5303\Psi^{\frac{1}{2}}\right)^{\frac{2}{3}}}{27} = -0.02333\Psi^{\frac{1}{3}} \leq -\frac{1}{42}\Psi^{\frac{1}{3}}.$$

Thus it follows that

$$\Psi_{k+1} \leq \Psi_k - \kappa{\Psi_k}^{1-\gamma}, \qquad k = 0, 1, \cdots, K-1,$$

with $\kappa = \frac{1}{42}$ and $\gamma = \frac{2}{3}$, and where K denotes the number of inner iterations. Hence the number K of inner iterations is bounded above by

$$K \leq \frac{\Psi_0^{\gamma}}{\kappa\gamma} = \frac{\Psi_0^{\frac{2}{3}}}{\frac{1}{42} \times \frac{2}{3}} = 63\,\Psi_0^{\frac{2}{3}}.$$

Step 6: To estimate Ψ_0, i.e., the value of $\Psi(v)$ just after a μ-update we could have used Lemma 2.5.3. A little sharper result is obtained as follows. By (2.4.1):

$$\Psi_0 \leq L_\psi(n, \theta, \tau) = n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right).$$

Using (2.5.3), substitution yields, after some elementary reductions,

$$\begin{aligned}\Psi_0 \leq \frac{\left(\tau + n\theta + \sqrt{\tau^2 + 2n\tau}\right)^2}{(1-\theta)\left(\sqrt{\tau} + \sqrt{2n+\tau}\right)^2} &= \frac{1}{1-\theta}\left(\sqrt{\tau} + \frac{n\theta}{\sqrt{\tau} + \sqrt{2n+\tau}}\right)^2 \\ &\leq \frac{1}{1-\theta}\left(\sqrt{\tau} + \frac{n\theta}{\sqrt{2n}}\right)^2.\end{aligned}$$

Step 7: Thus the total number of iterations is bounded above by

$$\frac{63}{\theta(1-\theta)^{\frac{2}{3}}}\left(\sqrt{\tau} + \frac{\theta\sqrt{n}}{\sqrt{2}}\right)^{\frac{4}{3}} \log\frac{n}{\epsilon}.$$

Step 8: For large-update methods the right-hand side expression is $O\left(n^{\frac{2}{3}} \log \frac{n}{\epsilon}\right)$.

For small-update methods the right-hand side expression is $O\left(\sqrt{n}\, \log \frac{n}{\epsilon}\right)$.

Example 5. Consider the case where $\psi(t) = \psi_5(t)$:

$$\psi(t) = \frac{t^2-1}{2} + e^{\frac{1}{t}-1} - 1.$$

Step 1: The inverse function of $-\psi_b'(t) = \frac{e^{\frac{1}{t}-1}}{t^2}$ is such that

$$\underline{\varrho}(s) = t \quad \Leftrightarrow \quad \frac{e^{\frac{1}{t}-1}}{t^2} = s, \quad t \leq 1.$$

It follows that $e^{\frac{1}{t}-1} = st^2 \leq s$ whence $\underline{\varrho}\,(s) = t \geq \frac{1}{1+\log s}$. Hence, by Lemma 2.5.4,

$$\rho(s) \geq \underline{\varrho}\,(1+2s).$$

Step 2: Since $\psi''(t)$ is monotonically decreasing we thus have

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''\left(\rho(2\delta)\right)} \leq -\frac{\delta^2}{\psi''\left(\underline{\varrho}\,(1+4\delta)\right)}.$$

Now putting $t = \underline{\varrho}\,(1+4\delta)$ we have $t \leq 1$ and may write

$$\begin{aligned} f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''\,(t)} &= -\frac{\delta^2}{1+\frac{1+2t}{t^4}\,e^{\frac{1}{t}-1}} \leq -\frac{\delta^2}{1+\frac{3}{t^4}\,e^{\frac{1}{t}-1}} \\ &= -\frac{\delta^2}{1+\frac{3(1+4\delta)}{t^2}} \leq -\frac{\delta^2}{1+\frac{15\delta}{t^2}}. \end{aligned}$$

Since

$$\frac{1}{t^2} = \frac{1}{\left(\underline{\varrho}\,(1+4\delta)\right)^2} \leq \left(1+\log\left(1+4\delta\right)\right)^2$$

we finally get

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{1+15\delta\left(1+\log\left(1+4\delta\right)\right)^2}.$$

Step 3: By Lemma 2.5.1 the inverse function of $\psi(t)$ for $t \in [1, \infty)$ satisfies

$$\sqrt{1+2s} \leq \varrho\,(s) \leq 1+\sqrt{2s}.$$

Omitting the argument v, we have

$$\varrho\left(\Psi(v)\right) \geq \sqrt{1+2\Psi}.$$

Step 4: Now using that $\delta(v) \geq \frac{1}{2}\psi'\left(\varrho\left(\Psi(v)\right)\right)$, we obtain

$$\delta \geq \frac{1}{2}\left(\sqrt{1+2\Psi} - \frac{e^{\frac{1}{\sqrt{1+2\Psi}}-1}}{1+2\Psi}\right) \geq \frac{1}{2}\left(\sqrt{1+2\Psi}-1\right) = \frac{\Psi}{1+\sqrt{1+2\Psi}}.$$

Step 5: Substitution gives, after some elementary reductions, while assuming $\Psi_0 \geq \Psi \geq \tau \geq 1$,

$$f(\tilde{\alpha}) \leq -\frac{\Psi^{\frac{1}{2}}}{44\left(1+\log(1+\sqrt{2\Psi})\right)^2} \leq -\frac{\Psi^{\frac{1}{2}}}{44\left(1+\log(1+\sqrt{2\Psi_0})\right)^2}.$$

Thus it follows that

$$\Psi_{k+1} \leq \Psi_k - \kappa {\Psi_k}^{1-\gamma}, \qquad k = 0, 1, \cdots, K-1,$$

with $\kappa = \frac{1}{44\left(1+\log(1+\sqrt{2\Psi_0})\right)^2}$ and $\gamma = \frac{1}{2}$, and where K denotes the number of inner iterations. Hence the number K of inner iterations is bounded above by

$$K \leq \frac{\Psi_0^\gamma}{\kappa\gamma} = 88 \left(1 + \log(1 + \sqrt{2\Psi_0})\right)^2 \Psi_0^{\frac{1}{2}}.$$

Step 6: We use Lemma 2.5.3, with $\psi''(1) = 4$, to estimate Ψ_0. This gives

$$\Psi_0 \leq 2 \frac{\left(\theta\sqrt{n} + \sqrt{2\tau}\right)^2}{1-\theta}.$$

Substitution in the expression for K gives

$$K \leq 88\sqrt{2} \left(1 + \log\left(1 + 2\,\frac{\theta\sqrt{n} + \sqrt{2\tau}}{\sqrt{1-\theta}}\right)\right)^2 \frac{\theta\sqrt{n} + \sqrt{2\tau}}{\sqrt{1-\theta}}.$$

Step 7: Thus the total number of iterations is bounded above by

$$88\sqrt{2} \left(1 + \log\left(1 + 2\,\frac{\theta\sqrt{n} + \sqrt{2\tau}}{\sqrt{1-\theta}}\right)\right)^2 \frac{\theta\sqrt{n} + \sqrt{2\tau}}{\theta\sqrt{1-\theta}} \log \frac{n}{\epsilon}$$

Step 8: For large-update methods the right hand side expression becomes $O\left(\sqrt{n}\,(\log n)^2 \log \frac{n}{\epsilon}\right)$.

For small-update methods the right hand side expression becomes $O\left(\sqrt{n} \log \frac{n}{\epsilon}\right)$.

Example 6. We proceed with $\psi(t) = \psi_6(t)$:

$$\psi(t) = \frac{t^2-1}{2} - \int_1^t e^{\frac{1}{\xi}-1} d\xi.$$

Step 1: The inverse function of $-\psi_b'(t) = e^{\frac{1}{t}-1}$ is given by $\underline{\rho}(s) = \frac{1}{1+\log s}$. Hence, by Lemma 2.5.4,

$$\rho(s) \geq \frac{1}{1 + \log(1+2s)}.$$

Step 2: It follows that

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''\left(\rho(2\delta)\right)} = -\frac{\delta^2}{1 + \frac{e^{\frac{1}{\rho(2\delta)}-1}}{\rho(2\delta)^2}} \leq -\frac{\delta^2}{1 + (1+4\delta)\left(1 + \log(1+4\delta)\right)^2}.$$

Step 3: By Lemma 2.5.1 the inverse function of $\psi(t)$ for $t \in [1, \infty)$ satisfies

$$\sqrt{1+2s} \leq \varrho\,(s) \leq 1 + \sqrt{2s}.$$

Omitting the argument v, we have

$$\varrho\left(\Psi(v)\right) \geq \sqrt{1+2\Psi}.$$

Step 4: Now using that $\delta(v) \geq \frac{1}{2}\psi'\left(\varrho\left(\Psi(v)\right)\right)$, we obtain

$$\delta \geq \frac{1}{2}\left(\sqrt{1+2\Psi} - e^{\frac{1}{\sqrt{1+2\Psi}}-1}\right) \geq \frac{1}{2}\left(\sqrt{1+2\Psi}-1\right) = \frac{\Psi}{1+\sqrt{1+2\Psi}}.$$

Step 5: Substitution gives, after some elementary reductions, while assuming $\Psi_0 \geq \Psi \geq \tau \geq 1$,

$$f(\tilde{\alpha}) \leq -\frac{\Psi^{\frac{1}{2}}}{21\left(1+\log(1+\sqrt{\Psi})\right)^2} \leq -\frac{\Psi^{\frac{1}{2}}}{21\left(1+\log(1+\sqrt{\Psi_0})\right)^2}.$$

Thus it follows that

$$\Psi_{k+1} \leq \Psi_k - \kappa {\Psi_k}^{1-\gamma}, \qquad k = 0, 1, \cdots, K-1,$$

with $\kappa = \frac{1}{21\left(1+\log(1+\sqrt{\Psi_0})\right)^2}$ and $\gamma = \frac{1}{2}$, and where K denotes the number of inner iterations. Hence the number K of inner iterations is bounded above by

$$K \leq \frac{\Psi_0^\gamma}{\kappa\gamma} = 42\left(1+\log(1+\sqrt{\Psi_0})\right)^2 \Psi_0^{\frac{1}{2}}.$$

Step 6: We use Lemma 2.5.3, with $\psi''(1) = 2$, to estimate Ψ_0. This gives

$$\Psi_0 \leq \frac{\left(\theta\sqrt{n}+\sqrt{2\tau}\right)^2}{1-\theta}.$$

Substitution in the expression for K gives

$$K \leq 42\left(1+\log\left(1+\frac{\theta\sqrt{n}+\sqrt{2\tau}}{\sqrt{1-\theta}}\right)\right)^2 \frac{\theta\sqrt{n}+\sqrt{2\tau}}{\sqrt{1-\theta}}.$$

Step 7: Thus the total number of iterations is bounded above by

$$42\left(1+\log\left(1+\frac{\theta\sqrt{n}+\sqrt{2\tau}}{\sqrt{1-\theta}}\right)\right)^2 \frac{\theta\sqrt{n}+\sqrt{2\tau}}{\theta\sqrt{1-\theta}} \log\frac{n}{\epsilon}.$$

Step 8: For large-update methods the bound becomes $O\left(\sqrt{n}\left(\log n\right)^2 \log\frac{n}{\epsilon}\right)$. and for small-update methods we get the iteration bound $O\left(\sqrt{n}\log\frac{n}{\epsilon}\right)$.

Example 7. Consider the case where $\psi(t) = \psi_7(t)$:

$$\psi(t) = \frac{t^2-1}{2} + \frac{t^{1-q}-1}{q-1}, \;\; q > 1.$$

Step 1: The inverse function of $-\psi_b'(t) = \frac{1}{t^q}$ is given by $\underline{\rho}(s) = \frac{1}{s^{\frac{1}{q}}}$. Hence, by Lemma 2.5.4,

$$\rho(s) \geq \frac{1}{(1+2s)^{\frac{1}{q}}}.$$

Step 2: For the default step size $\tilde{\alpha}$ we thus have

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''\left(\rho(2\delta)\right)} = -\frac{\delta^2}{1+\frac{q}{\rho(2\delta)^{q+1}}} \leq -\frac{\delta^2}{1+q\left(1+4\delta\right)^{\frac{q+1}{q}}}.$$

Step 3: By Lemma 2.5.1 the inverse function of $\psi(t)$ for $t \in [1, \infty)$ satisfies

$$\sqrt{1+2s} \leq \varrho\left(s\right) \leq 1 + \sqrt{2s}.$$

Omitting the argument v, we have

$$\varrho\left(\Psi(v)\right) \geq \sqrt{1+2\Psi} \geq \sqrt{2\Psi}.$$

Step 4: Now using that $\delta(v) \geq \frac{1}{2}\psi'\left(\varrho\left(\Psi(v)\right)\right)$, and assuming $\Psi \geq \tau \geq 1$, we obtain

$$\delta \geq \frac{1}{2}\left(\sqrt{2\Psi} - \frac{1}{\left(\sqrt{2\Psi}\right)^q}\right) \geq \frac{1}{2}\left(\sqrt{2\Psi} - \frac{1}{\sqrt{2\Psi}}\right) \geq \frac{1}{2}\left(\frac{\Psi}{2}\right)^{\frac{1}{2}}.$$

Step 5: Substitution gives

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{1+q\left(1+4\delta\right)^{\frac{q+1}{q}}} \leq -\frac{\Psi}{8\left(1+q\left(1+\sqrt{2\Psi}\right)^{\frac{q+1}{q}}\right)} \leq -\frac{1}{56\,q}\Psi^{\frac{q-1}{2q}}.$$

Thus it follows that

$$\Psi_{k+1} \leq \Psi_k - \kappa{\Psi_k}^{1-\gamma}, \qquad k = 0, 1, \cdots, K-1,$$

with $\kappa = \frac{1}{56\,q}$ and $\gamma = \frac{q+1}{2q}$, and where K denotes the number of inner iterations. Hence the number K of inner iterations is bounded above by

$$K \leq \frac{\Psi_0^\gamma}{\kappa\gamma} = \frac{112\,q^2}{q+1}\,\Psi_0^{\frac{q+1}{2q}} \leq 112\,q\,\Psi_0^{\frac{q+1}{2q}}.$$

Step 6: We now have to estimate Ψ_0, i.e., the value of $\Psi(v)$ just after a μ-update. Since $\psi''(1) = q + 1$, when using Lemma 2.5.3 we get a factor $q + 1$ that we can avoid as follows. By (2.4.1):

$$\Psi_0 \leq L_\psi(n, \theta, \tau) = n\psi\left(\frac{\rho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right) \leq n\psi\left(\frac{1+\sqrt{\frac{2\tau}{n}}}{\sqrt{1-\theta}}\right).$$

If $t \geq 1$ then $\psi(t) \leq \frac{1}{2}(t^2 - 1)$. Using this we obtain

$$\Psi_0 \leq \frac{n}{2}\left(\frac{\left(1+\sqrt{\frac{2\tau}{n}}\right)^2}{1-\theta} - 1\right) = \frac{n}{2}\,\frac{\theta + \frac{2\tau}{n} + 2\sqrt{\frac{2\tau}{n}}}{1-\theta} = \frac{\theta n + 2\tau + 2\sqrt{2\tau n}}{2\,(1-\theta)}.$$

Step 7: Thus the total number of iterations is bounded above by

$$\frac{112\,q}{\theta}\left(\frac{\theta n + 2\tau + 2\sqrt{2\tau n}}{2\,(1-\theta)}\right)^{\frac{q+1}{2q}} \log\frac{n}{\epsilon}.$$

Step 8: For large-update methods the right hand side expression is $O\left(q\, n^{\frac{q+1}{2q}}\, \log \frac{n}{\epsilon}\right)$ and for small-update methods $O\left(qn^{\frac{3q+1}{4q}}\, \log \frac{n}{\epsilon}\right)$.

The bound for small-update methods can be improved. For this we go back to Step 6.

Step 6: Using Lemma 2.5.3, with $\psi''(1) = q + 1$, to estimate Ψ_0, we obtain

$$\Psi_0 \leq \frac{q+1}{2}\,\frac{\left(\theta\sqrt{n} + \sqrt{2\tau}\right)^2}{1-\theta}.$$

Step 7: Thus, using $(q+1)^{\frac{q+1}{2q}} \leq q + 1$, the total number of iteration is bounded above by

$$\frac{K}{\theta}\log\frac{n}{\epsilon} \leq \frac{112q(q+1)}{\theta}\left(\frac{\left(\theta\sqrt{n} + \sqrt{2\tau}\right)^2}{2\,(1-\theta)}\right)^{\frac{q+1}{2q}} \log\frac{n}{\epsilon}.$$

Step 8: For small-update methods we thus get the bound $O\left(q^2\sqrt{n}\,\log\frac{n}{\epsilon}\right)$.

Example 8. We proceed with $\psi(t) = \psi_8(t)$, which is a special case of ψ_{10} $(p = 0)$:

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

Step 1: To obtain the inverse function $t = \rho(s)$ of $-\frac{1}{2}\psi'(t)$ for $t \in (0,1]$ we need to solve t from the equation $-\frac{1}{2}\psi'(t) = s$:

$$-1 + \frac{1}{t^q} = 2s, \quad t \in (0,1].$$

This implies

$$t = \rho(s) = \frac{1}{(2s+1)^{\frac{1}{q}}}.$$

Step 2: Since $\psi''(t) = qt^{-q-1}$, it follows that

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''(\rho(2\delta))} = -\frac{\delta^2 \rho(2\delta)^{q+1}}{q} = -\frac{\delta^2}{q\,(4\delta+1)^{\frac{q+1}{q}}} \leq -\frac{\delta^2}{q\,(4\delta+1)^{\frac{q+1}{q}}}.$$

Step 3: This is the only case where the growth term is linear. As a consequence we can not use Lemma 2.5.1 and in particular the treatment of ϱ, the inverse function of $\psi(t)$ for $t \in [1,\infty)$, needs special attention. One has

$$t = \varrho\,(s) \quad \Leftrightarrow \quad \psi(t) = t - 1 + \frac{t^{1-q}-1}{q-1} = s, \quad t \geq 1.$$

Using that $t \geq 1$ one easily sees that

$$1 + s \leq \varrho\,(s) \leq s + \frac{q}{q-1}.$$

We have

$$\varrho(\Psi(v)) \geq \Psi(v) + 1.$$

Step 4: Now using that $\delta(v) \geq \frac{1}{2}\psi'(\varrho(\Psi(v)))$, omitting the argument v, and assuming $\Psi \geq \tau \geq 1$, we obtain

$$\delta \geq \frac{1}{2}\left(1 - \frac{1}{(\Psi+1)^q}\right) \geq \frac{1}{4}.$$

Step 5: Substitution yields

$$f(\tilde{\alpha}) \leq -\frac{1}{16\,q\,2^{\frac{q+1}{q}}} \leq -\frac{1}{64\,q}.$$

Hence we have

$$\Psi_{k+1} \leq \Psi_k - \frac{1}{64\,q}, \quad k = 0, 1, \ldots, K-1,$$

where K denotes the number of the inner iterations. Thus the number K is bound above by

$$K \leq 64q\Psi_0.$$

Step 6: We finally have to estimate Ψ_0, i.e., to derive an upper bound for $\Psi(v)$ just after a μ-update. To this end we use (2.4.1):

$$\Psi_0 \leq n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right) \leq n\psi\left(\frac{\frac{\tau}{n}+\frac{q}{q-1}}{\sqrt{1-\theta}}\right) \leq n\left(\frac{\frac{\tau}{n}+\frac{q}{q-1}}{\sqrt{1-\theta}}-1\right).$$

By using (2.5.2), we have

$$\Psi_0 \leq n\left(\frac{\theta+\frac{\tau}{n}+\frac{1}{q-1}}{\sqrt{1-\theta}}\right).$$

Step 7: The total number of iterations is bound above by

$$64q\frac{\theta n+\tau+\frac{n}{q-1}}{\theta\sqrt{1-\theta}}\log\frac{n}{\epsilon}.$$

Step 8: If $\tau = O(n)$ and $\theta = \Theta(1)$, then assuming that q is bounded away from 1, the right-hand side expression is $O\left(qn\log\frac{n}{\epsilon}\right)$. Note that if $q = O(1)$, then this iteration bound is the same as the bound for the logarithmic barrier function.

For small-update methods the iteration bound obtained in this way is

$$O\left(qn\log\frac{n}{\epsilon}\right).$$

It may be noted that the bad iteration bound for small-update methods is due to the fact that the upper bound used for $\varrho(s)$ is not tight at $s=0$: it should be equal to $\varrho(0)=1$ when $s=0$. We will see below that an upper bound that is tight at $s=0$ will lead to a better iteration bound. By Lemma 2.5.2,

$$t \leq 1+\sqrt{t\psi(t)}.$$

Step 3: Substituting $t \leq \psi(t)+\frac{q}{q-1}$ we obtain

$$t=\varrho(s) \leq 1+\sqrt{s^2+\frac{q}{q-1}s}.$$

Note this upper bound is tight at $s=0$.

We recall that the number K of inner iterations is bound above by

$$K \leq 64q\Psi_0.$$

Step 6: To estimate Ψ_0, we use (2.3.4) and Lemma 2.2.7, with $\psi''(1)=q$. We then obtain

$$\begin{aligned}\Psi_0 &\leq n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right) \leq n\psi\left(\frac{1+\sqrt{\frac{\tau^2}{n^2}+\frac{q}{q-1}\frac{\tau}{n}}}{\sqrt{1-\theta}}\right) \\ &\leq \frac{qn}{2}\left(\frac{1+\sqrt{\frac{\tau^2}{n^2}+\frac{q}{q-1}\frac{\tau}{n}}}{\sqrt{1-\theta}}-1\right)^2.\end{aligned}$$

Using (2.5.2) and $\frac{q}{q-1} \leq 2$ this can be simplified to

$$\Psi_0 \leq \frac{qn}{2}\left(\frac{\theta+\sqrt{\frac{\tau^2}{n^2}+\frac{2\tau}{n}}}{\sqrt{1-\theta}}\right)^2 = \frac{q\left(\theta\sqrt{n}+\sqrt{\frac{\tau^2}{n}+2\tau}\right)^2}{2(1-\theta)}.$$

Step 7: We conclude that the total number of iterations is bounded above by

$$\frac{K}{\theta}\log\frac{n}{\epsilon} \leq 32\frac{q^2\left(\theta\sqrt{n}+\sqrt{\frac{\tau^2}{n}+2\tau}\right)^2}{\theta(1-\theta)}\log\frac{n}{\epsilon}.$$

Step 8: This makes clear that the iteration bound for small-update methods is $O\left(q^2\sqrt{n}\,\log\frac{n}{\epsilon}\right)$.

Example 9. We proceed with $\psi(t) = \psi_9(t)$:

$$\psi(t) = \frac{t^{p+1}-1}{p+1} - \log t, \quad p \in [0,1].$$

Step 1: To obtain the inverse function $t = \rho(s)$ of $-\frac{1}{2}\psi'(t)$ for $t \in (0,1]$ we need to solve t from the equation

$$-t^p + \frac{1}{t} = 2s, \quad t \in (0,1].$$

Using that $\frac{1}{t} = 2s + t^p \leq 2s + 1$, this implies

$$t = \rho(s) \geq \frac{1}{2s+1}.$$

Step 2: It follows that

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''(\rho(2\delta))} = -\frac{\delta^2}{p\left(\rho(2\delta)\right)^{p-1}+\frac{1}{(\rho(2\delta))^2}} \leq -\frac{1}{p\left(4\delta+1\right)^{1-p}+\left(4\delta+1\right)^2}.$$

Since $p\,(4\delta+1)^{1-p} \leq (4\delta+1)^2$, for $p \in [0,1]$, it follows that

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{2\,(4\delta+1)^2}.$$

Step 3: As in Example 8, we can not use Lemma 2.5.1 and the estimate of ϱ, the inverse function of $\psi(t)$ for $t \in [1,\infty)$, needs special treatment. One has

$$\frac{t^{1+p}-1}{1+p} - \log t = s, \qquad t \geq 1,$$

whence

$$\frac{t^{1+p}-1}{1+p} = \log t + s,$$

which implies the following lower bound for $\varrho(s) = t$:

$$\varrho(s) = t \geq (1+(1+p)\,s)^{\frac{1}{1+p}}, \qquad p \in [0,1]. \tag{2.5.4}$$

On the other hand, using $\frac{(t-1)^2}{2t} \leq \psi(t) = s$. Hence we have the following inequality.

$$t^2 - 2t\,(1+s) + 1 \leq 0.$$

This implies

$$1 + s - \sqrt{s^2+2s} \leq t \leq 1 + s + \sqrt{s^2+2s}.$$

Therefore, we certainly have the following upper and lower bounds for $t = \varrho(s)$:

$$(1+(1+p)\,s)^{\frac{1}{1+p}} \leq t = \varrho(s) \leq 1 + s + \sqrt{s^2+2s}.$$

Step 4: Now using that $\delta(v) \geq \frac{1}{2}\psi'(\varrho(\Psi(v)))$, omitting the argument v, and assuming $\Psi \geq \tau \geq 1$, we obtain

$$\delta(v) \geq \frac{1}{2}\left((1+(1+p)\,\Psi)^{\frac{p}{1+p}} - \frac{1}{(1+(1+p)\,\Psi)^{\frac{1}{1+p}}}\right),$$

then

$$\delta(v) \geq \frac{(1+p)\,\Psi(v)}{2\,(1+(1+p)\,\Psi)^{\frac{1}{1+p}}} \geq \frac{(1+p)\,\Psi(v)}{2\,(1+\Psi(v))}.$$

The last inequality follows from Bernoulli's inequality. Note that if $\Psi \geq \tau \geq 1$, substitution gives

$$\delta(v) \geq \frac{1}{4}. \tag{2.5.5}$$

Step 5: Using (9), we have

$$f(\tilde{\alpha}) \leq -\frac{1}{2\left(4\delta+1\right)^2} \leq -\frac{1}{128}.$$

Hence we have

$$\Psi_{k+1} \leq \Psi_k - \frac{1}{128}, \quad k = 0, 1, \ldots, K-1,$$

where K denotes the number of the inner iterations. Thus the number K is bound above by

$$K \leq 128\Psi_0.$$

Step 6: To estimate Ψ_0, the value of $\Psi(v)$ just after a μ-update, we use (2.4.1):

$$\Psi_0 \leq n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right) \leq \frac{n\left(1+p\right)}{2}\left(\frac{1+\frac{\tau}{n}+\sqrt{\left(\frac{\tau}{n}\right)^2+\frac{2\tau}{n}}}{\sqrt{1-\theta}}-1\right)^2.$$

By using

$$1-\sqrt{1-\theta} = \frac{\theta}{1+\sqrt{1-\theta}} \leq \theta,$$

we obtain

$$\begin{aligned} \Psi_0 &\leq \frac{n\left(1+p\right)}{2\left(1-\theta\right)}\left(\theta+\frac{\tau}{n}+\sqrt{\left(\frac{\tau}{n}\right)^2+\frac{2\tau}{n}}\right)^2 \\ &= \frac{\left(1+p\right)}{2\left(1-\theta\right)}\left(\theta\sqrt{n}+\frac{\tau}{\sqrt{n}}+\sqrt{\frac{\tau^2}{n}+2\tau}\right)^2. \end{aligned}$$

Step 7: Thus the total number of iterations is bound above by

$$\frac{K}{\theta}\log\frac{n}{\epsilon} \leq 64\left(1+p\right)\frac{\left(\theta\sqrt{n}+\frac{\tau}{\sqrt{n}}+\sqrt{\frac{\tau^2}{n}+2\tau}\right)^2}{\theta\left(1-\theta\right)}\log\frac{n}{\epsilon}.$$

Step 8: For large-update methods the iteration bound becomes $O\left(\left(1+p\right)n\log\frac{n}{\epsilon}\right)$ and for small-update methods $O\left(\left(1+p\right)\sqrt{n}\,\log\frac{n}{\epsilon}\right)$.

Example 10. We finally consider $\psi(t) = \psi_{10}(t)$:

$$\frac{t^{p+1}-1}{p+1}+\frac{t^{1-q}-1}{q-1}, \quad p \in [0,1], \quad q > 1.$$

Step 1: To obtain the inverse function $t = \rho(s)$ of $-\frac{1}{2}\psi'(t)$ for $t \in (0,1]$ we need to solve t from the equation

$$-t^p + t^{-q} = 2s, \quad t \in (0,1].$$

Using that $t^{-q} = 2s + t^p \leq 2s + 1$, this implies

$$t = \rho(s) \geq \frac{1}{(2s+1)^{\frac{1}{q}}}.$$

Step 2: It follows that

$$\begin{aligned} f(\tilde{\alpha}) &\leq -\frac{\delta^2}{\psi''(\rho(2\delta))} = -\frac{\delta^2}{p\,(\rho(2\delta))^{p-1} + \frac{1}{(\rho(2\delta))^{q+1}}} \\ &\leq -\frac{1}{p\,(4\delta+1)^{\frac{1-p}{q}} + q\,(4\delta+1)^{\frac{q+1}{q}}}. \end{aligned}$$

Since $(4\delta+1)^{1-p} \leq (4\delta+1)^{q+1}$, for $p \in [0,1]$, and $q \geq 1$, it follows that

$$f(\tilde{\alpha}) \leq -\frac{1}{(p+q)\,(4\delta+1)^{\frac{q+1}{q}}}. \tag{2.5.6}$$

Step 3: If $p < 1$ then the growth term of $\psi_8(t)$ is not quadratic, hence we can not use Lemma 2.5.1 to derive $\varrho(s)$, the inverse function of $\psi(t)$ for $t \in [1, \infty)$. Putting $t = \varrho(s)$ we have

$$\frac{t^{1+p}-1}{1+p} = \frac{1-t^{1-q}}{q-1} + s.$$

From this we easily get the following lower and upper bounds for $\varrho(s) = t$:

$$(1+(1+p)\,s)^{\frac{1}{1+p}} \leq \varrho(s) = t \leq \left((1+p)\,s + \frac{p+q}{q-1}\right)^{\frac{1}{1+p}}. \tag{2.5.7}$$

Step 4: Now using that $\delta(v) \geq \frac{1}{2}\psi'(\varrho(\Psi(v)))$, and ψ' is monotonically increasing for $t \geq 1$, we may replace $\varrho(\Psi(v))$ by a smaller value. Thus omitting the argument v, and assuming $\Psi \geq \tau \geq 1$, we obtain

$$\begin{aligned} \delta(v) &\geq \frac{1}{2}\left((1+(1+p)\,\Psi)^{\frac{p}{1+p}} - \frac{1}{(1+(1+p)\,\Psi)^{\frac{q}{1+p}}}\right) \\ &\geq \frac{1}{2}\left((1+(1+p)\,\Psi)^{\frac{p}{1+p}} - \frac{1}{(1+(1+p)\,\Psi)^{\frac{1}{1+p}}}\right), \end{aligned}$$

whence

$$\delta(v) \geq \frac{(1+p)\,\Psi(v)}{2\,(1+(1+p)\,\Psi)^{\frac{1}{1+p}}} \geq \frac{\Psi(v)}{2\,(1+2\Psi(v))^{\frac{1}{1+p}}} \geq \frac{1}{6}\Psi(v)^{\frac{p}{1+p}}.$$

Step 5: Since the right hand sid expression in (2.5.6) is monotonically decreasing in δ, also using (9), we obtain

$$f(\tilde{\alpha}) \leq -\frac{\Psi^{\frac{2p}{1+p}}}{36\,(p+q)\left(\frac{2}{3}\Psi^{\frac{p}{1+p}}+1\right)^{\frac{q+1}{q}}} \leq -\frac{1}{100\,(p+q)}\Psi^{\frac{p(q-1)}{q(1+p)}}.$$

Hence we have

$$\Psi_{k+1} \leq \Psi_k - \kappa{\Psi_k}^{1-\gamma}, \qquad k = 0, 1, \cdots, K-1,$$

with $\kappa = \frac{1}{100(p+q)}$ and $\gamma = \frac{q+p}{q(1+p)}$, and where K denotes the number of the inner iterations. Thus the number K is bounded above by

$$K \leq 100\,(1+p)\,q\Psi_0^{\frac{q+p}{q(1+p)}}.$$

Step 6: Now we need to find an upper bound of Ψ_0. To this end we use (2.4.1) and $\psi(t) \leq \frac{t^{1+p}}{1+p}$ for $t \geq 1$. This gives

$$\Psi_0 \leq n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right) \leq n\psi\left(\frac{\left(\frac{(1+p)\tau}{n}+\frac{q+p}{q-1}\right)^{\frac{1}{1+p}}}{\sqrt{1-\theta}}\right) \leq \frac{\left((1+p)\,\tau+\frac{q+p}{q-1}n\right)}{(p+1)\,(1-\theta)^{\frac{p+1}{2}}}.$$

Step 7: Thus we obtain an upper bound for the total number of iterations, namely,

$$\frac{100\,(1+p)\,q}{\theta\,(1-\theta)^{\frac{p+q}{2q}}}\left(\frac{(1+p)\,\tau+\frac{q+p}{q-1}n}{1+p}\right)^{\frac{p+q}{q(1+p)}}\log\frac{n}{\epsilon}.$$

Step 8: For large-update methods the right hand side expression becomes

$$O\left(qn^{\frac{p+q}{q(1+p)}}\log\frac{n}{\epsilon}\right),$$

and for small-update methods

$$O\left(q\sqrt{n}n^{\frac{p+q}{q(1+p)}}\log\frac{n}{\epsilon}\right).$$

The last bound can be sharpened, as we show below. Just as in Example 8 we go back to Step 3, and use Lemma 2.5.2 to derive bounds for the inverse function ϱ of $\psi(t)$. We have

$$t \leq 1+\sqrt{t\psi(t)}.$$

Step 3: Substituting $t \leq \left((1+p)\,\psi(t) + \frac{q+p}{q-1}\right)^{\frac{1}{1+p}}$ we obtain

$$t = \varrho(s) \leq 1 + \sqrt{s}\left((1+p)\,s + \frac{p+q}{q-1}\right)^{\frac{1}{2(1+p)}}.$$

Step 6: Thus we obtain the following upper bound of Ψ_0.

$$\Psi_0 \leq \frac{n\,(p+q)}{2}\left(\frac{1+\sqrt{\frac{\tau}{n}}\left((1+p)\,\frac{\tau}{n}+\frac{p+q}{q-1}\right)^{\frac{1}{2(1+p)}}}{\sqrt{1-\theta}} - 1\right)^2.$$

By using $1-\sqrt{1-\theta} \leq \theta$, by (2.5.2), we get

$$\begin{aligned}
\Psi_0 &\leq \frac{n\,(p+q)}{2\,(1-\theta)}\left(\theta + \sqrt{\frac{\tau}{n}}\left((1+p)\,\frac{\tau}{n}+\frac{p+q}{q-1}\right)^{\frac{1}{2(1+p)}}\right)^2 \\
&= \frac{(p+q)}{2\,(1-\theta)}\left(\theta\sqrt{n} + \tau\left((1+p)\,\frac{\tau}{n}+\frac{p+q}{q-1}\right)^{\frac{1}{2(1+p)}}\right)^2.
\end{aligned}$$

Step 7: Thus the total number of iterations is bound above by

$$\frac{100\,(1+p)\,q}{\theta}\left(\left(\frac{(p+q)}{2\,(1-\theta)}\right)\left(\theta\sqrt{n} + \tau\left((1+p)\,\frac{\tau}{n}+\frac{p+q}{q-1}\right)^{\frac{1}{2(1+p)}}\right)^2\right)^{\frac{(p+q)}{q(1+p)}} \log\frac{n}{\epsilon}.$$

Since $\frac{p+q}{q(1+p)} \leq 1$ for all $p \in [0,1]$ and $q \geq 2$, the bound can be simplified to

$$\frac{50q\,(1+p)\,(p+q)}{\theta\,(1-\theta)}\left(\theta\sqrt{n} + \tau\left((1+p)\,\frac{\tau}{n}+\frac{p+q}{q-1}\right)^{\frac{1}{2(1+p)}}\right)^2 \log\frac{n}{\epsilon}.$$

Step 8: For small-update methods and $p \in [0,1]$, the right hand side expression is $O\left(q^2\sqrt{n}\,\log\frac{n}{\epsilon}\right)$.

2.5.3 Summary of results

The various iteration bounds for small-update methods are listed in Table 2.9. Note that the small-update methods based on the kernel functions considered in this chapter all have the same complexity as the small-update method based on the logarithmic barrier function, namely $O\left(\sqrt{n}\log\frac{n}{\epsilon}\right)$. As is well known this is up till now the best iteration bound for methods solving LO problems.

i	kernel functions ψ_i	iteration complexity	references
1	$\frac{t^2-1}{2} - \log t$	$O\left(\sqrt{n}\log\frac{n}{\epsilon}\right)$	[And96; Her94; Tod89]
2	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q(q-1)} - \frac{q-1}{q}(t-1), \quad q>1$	$O\left(q\sqrt{n}\log\frac{n}{\epsilon}\right)$	[Pen02a; Pen02b]
3	$\frac{t^2-1}{2} + \frac{(e-1)^2}{e}\frac{1}{e^t-1} - \frac{e-1}{e}$	$O\left(\sqrt{n}\log\frac{n}{\epsilon}\right)$	[Bai02a]
4	$\frac{1}{2}\left(t-\frac{1}{t}\right)^2$	$O\left(\sqrt{n}\log\frac{n}{\epsilon}\right)$	[Pen00a]
5	$\frac{t^2-1}{2} + e^{\frac{1}{t}-1} - 1$	$O\left(\sqrt{n}\log\frac{n}{\epsilon}\right)$	[Bai04a]
6	$\frac{t^2-1}{2} - \int_1^t e^{\frac{1}{\xi}-1}d\xi$	$O\left(\sqrt{n}\log\frac{n}{\epsilon}\right)$	[Bai04a]
7	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q-1},\ q>1$	$O\left(q^2\sqrt{n}\log\frac{n}{\epsilon}\right)$	[Pen01; Pen02b]
8	$t-1+\frac{t^{1-q}-1}{q-1}, \quad q>1$	$O\left(q^2\sqrt{n}\log\frac{n}{\epsilon}\right)$	[Bai02b]
9	$\frac{t^{1+p}-1}{1+p} - \log t,\ p\in[0,1]$	$O\left(\sqrt{n}\log\frac{n}{\epsilon}\right)$	[Gha04a]
10	$\frac{t^{p+1}-1}{p+1} + \frac{t^{1-q}-1}{q-1}, \quad p\in[0,1], \quad q>1$	$O\left(\sqrt{n}\log\frac{n}{\epsilon}\right)$	New

Table 2.9: *Complexity results for small-update methods.*

For large-update methods, the resulting iteration bounds are summarized in the second columns of Table 2.10. For ψ_2 and ψ_7 the bound is minimal if we choose $q=\frac{1}{2}\log n$, and for ψ_8 the bound is minimal if we choose $p=1$ and $q=\frac{1}{2}\log n$. This gives the best bound known so far for large-update interior-point methods: $O\left(\sqrt{n}\,(\log n)\log\frac{n}{\epsilon}\right)$.

2.6 A kernel function with finite barrier term

We conclude this chapter by showing that by refining the analysis also other kernel functions than those satisfying our conditions can be used to define very efficient

i	kernel functions ψ_i	iteration complexity	references
1	$\frac{t^2-1}{2} - \log t$	$O\left(n \log \frac{n}{\epsilon}\right)$	[And96; Her94; Tod89]
2	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q(q-1)} - \frac{q-1}{q}(t-1), \quad q>1$	$O\left(qn^{\frac{q+1}{2q}} \log \frac{n}{\epsilon}\right)$	[Pen02a; Pen02b]
3	$\frac{t^2-1}{2} + \frac{(e-1)^2}{e}\frac{1}{e^t-1} - \frac{e-1}{e}$	$O\left(n^{\frac{3}{4}} \log \frac{n}{\epsilon}\right)$	[Bai02a]
4	$\frac{1}{2}\left(t-\frac{1}{t}\right)^2$	$O\left(n^{\frac{2}{3}} \log \frac{n}{\epsilon}\right)$	[Pen00a]
5	$\frac{t^2-1}{2} + e^{\frac{1}{t}-1} - 1$	$O\left(\sqrt{n}\, \log^2 n \log \frac{n}{\epsilon}\right)$	[Bai04a]
6	$\frac{t^2-1}{2} - \int_1^t e^{\frac{1}{\xi}-1} d\xi$	$O\left(\sqrt{n}\, \log^2 n \log \frac{n}{\epsilon}\right)$	[Bai04a]
7	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q-1}, \ q>1$	$O\left(qn^{\frac{q+1}{2q}} \log \frac{n}{\epsilon}\right)$	[Pen01; Pen02b]
8	$t-1+\frac{t^{1-q}-1}{q-1}, \quad q>1$	$O\left(qn \log \frac{n}{\epsilon}\right)$	[Bai02b]
9	$\frac{t^{1+p}-1}{1+p} - \log t, \ p \in [0,1]$	$O\left(n \log \frac{n}{\epsilon}\right)$	[Gha04a]
10	$\frac{t^{p+1}-1}{p+1} + \frac{t^{1-q}-1}{q-1}, \quad p\in[0,1], \quad q>1$	$O\left(qn^{\frac{p+q}{q(1+p)}} \log \frac{n}{\epsilon}\right)$	New

Table 2.10: *Complexity results for large-update methods.*

IPMs. By way of example we consider kernel functions of the following form.

$$\psi_{p,\sigma}(t) = \frac{t^{1+p}-1}{p+1} + \frac{e^{\sigma(1-t)}-1}{\sigma}, \quad p \in [0,1], \quad \sigma \geq 1. \tag{2.6.1}$$

Note that all kernel functions in Table 2.2 have the properties $\lim_{t\downarrow 0} \psi(t) = \infty$ and $\lim_{t\to\infty} \psi(t) = \infty$. The function $\psi_{p,\sigma}$ has the second property, but lacks the first property, because

$$\lim_{t\downarrow 0} \psi_{p,\sigma}(t) = \psi(0) = \frac{e^\sigma - 1}{\sigma} - \frac{1}{p+1} < \infty.$$

Let note that the case $p=1$ has been considered before in [Bai03a].

Figure (2.3) shows the graph of: $\psi_1(t) = \frac{t^2-1}{2} - \log(t)$, and $\psi_{1,2}(t) = \frac{t^2-1}{2} + \frac{e^{-2(t-1)}-1}{2}$.

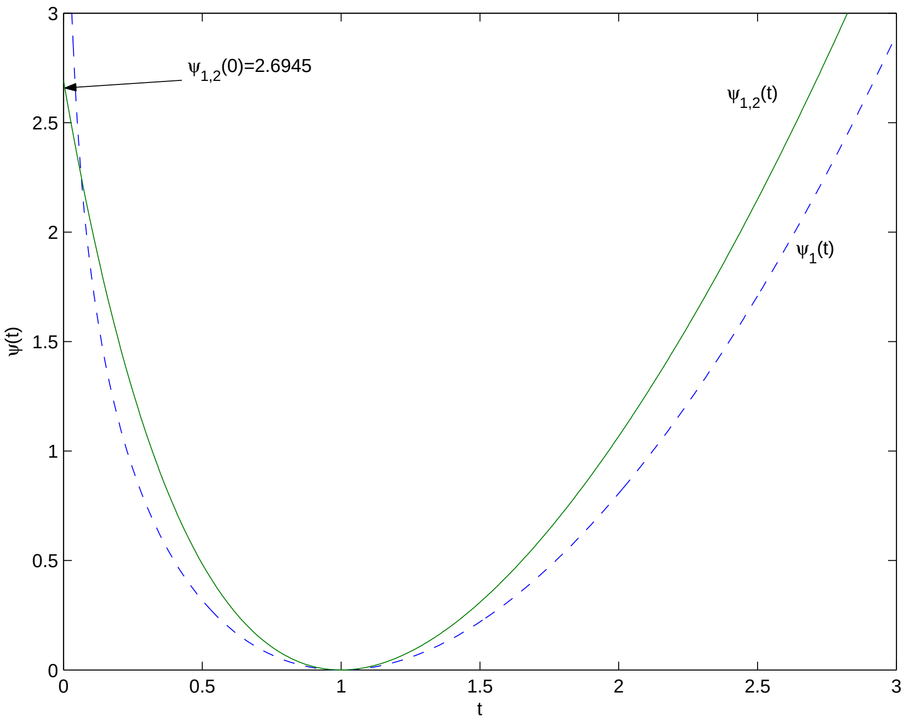


Figure 2.3: *Figure of $\psi_{1,2}$ and ψ_1.*

2.6.1 Properties

The first three derivatives of ψ are given by

$$\psi'(t) = t^p - e^{\sigma(1-t)}, \tag{2.6.2}$$
$$\psi''(t) = pt^{p-1} + \sigma e^{\sigma(1-t)}, \tag{2.6.3}$$
$$\psi'''(t) = -p\,(1-p)\,t^{p-2} - \sigma^2 e^{\sigma(1-t)}. \tag{2.6.4}$$

It follows that $\psi(1) = \psi'(1) = 0$ and $\psi''(t) \geq 0$.

Lemma 2.6.1. *Let ψ be as defined in (2.6.1). Then,*

$$t\psi''(t) + \psi'(t) \geq 0, \quad \text{if } t \geq \frac{1}{\sigma}, \tag{2.6.5-a}$$
$$\psi'''(t) < 0, \quad \text{if } t > 0, \tag{2.6.5-b}$$
$$\psi''(t)\psi'(\beta t) - \beta\psi'(t)\psi''(\beta t) \geq 0, \quad \text{if } t \geq 1, \quad \beta \geq 1. \tag{2.6.5-c}$$

Proof. Using (2.6.2) and (2.6.3) we write, also using $t \geq \frac{1}{\sigma}$,

$$\psi'(t) + t\psi^{''}(t) = (1+p)\,t^p + (t\sigma - 1)\,e^{\sigma(1-t)} \geq 0.$$

Thus (2.6.5-a) follows. Inequality (2.6.5-b) immediately follows from (2.6.4). By (2.6.2) and (2.6.3),

$$\psi''(t)\psi'(\beta t) - \beta\psi'(t)\psi''(\beta t) = \sigma\,(\beta - 1)\,e^{-\sigma(t-2+\beta t)} + t^{p-1}g(\beta) \geq t^{p-1}g(\beta),$$

where

$$g(\beta) = \beta^p\,(p + \sigma t)\,e^{-\sigma(t-1)} - (\beta\sigma t + p)\,e^{-\sigma(\beta t-1)}.$$

One has $g(1) = 0$ and

$$g'(\beta) = p\beta^{p-1}\,(p + \sigma t)\,e^{-\sigma(t-1)} + \sigma t\,(\sigma\beta t + p - 1)\,e^{\sigma(\beta t-1)} \geq 0,$$

because $\sigma\beta t \geq 1$. Hence $g(\beta) \geq 0$. From this (2.6.5-c) follows. □

We see that $\psi_{p,\sigma}(t)$ is not e-convex for all $t > 0$, but only if $t \geq \frac{1}{\sigma}$. This means that we must ensure that t is large enough, before using inequality (2.6.5-a).

Lemma 2.6.2. *If $\sigma \geq 2$, then one has*

$$t\psi(t) \geq (t-1)^2\,, \quad \textit{for} \;\; t \geq 1.$$

Proof. Defining $g(t) := t\psi(t) - (t-1)^2$ one has $g(1) = 0$ and $g'(t) = \psi(t) + t\psi'(t) - 2(t-1)$. Hence $g'(1) = 0$ and $g''(t) = 2\psi'(t) + t\psi''(t) - 2$. Since $g''(t) = 2\,(t^p - 1) + (t\sigma - 2)\,e^{-\sigma(t-1)} + pt^p \geq 0$, the lemma follows. □

2.6.2 Fixing the value of σ

After the update of μ to $(1-\theta)\mu$ we have $v_+ = \frac{v}{\sqrt{1-\theta}}$. Application of Theorem 2.3.1 yields that

$$\Psi(v_+) \leq L(n,\theta,\tau) := n\psi\left(\frac{\rho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right). \tag{2.6.6}$$

As before $L(n,\theta,\tau)$ is a uniform upper bound for $\Psi(v)$ during the course of the algorithm, since during subsequent inner iteration the value of $\Psi(v)$ decreases, as will follow below.

Lemma 2.6.3. *Suppose that* $L(n, \theta, \tau) \geq 9$ *and* $\Psi(v) \leq L(n, \theta, \tau)$. *If* $\sigma \geq 1 + 2\log(L(n, \theta, \tau) + 1)$, *then* $v_i > \frac{3}{2\sigma}$, *for all* $i = 1, \ldots, n$.

Proof. First note that $\Psi(v) \leq L(n, \theta, \tau)$ implies that $\psi(v_i) \leq L(n, \theta, \tau)$, for each $i = 1, \ldots, n$. Hence, putting $t = v_i$, we have

$$\frac{t^{1+p} - 1}{1 + p} + \frac{1}{\sigma}\left(e^{-\sigma(t-1)} - 1\right) \leq L(n, \theta, \tau).$$

It follows that

$$\frac{1}{\sigma}\left(e^{-\sigma(t-1)} - 1\right) \leq L(n, \theta, \tau) + \frac{1 - t^{1+p}}{1 + p} \leq L(n, \theta, \tau) + 1. \tag{2.6.7}$$

This implies

$$e^{1-\sigma t} \leq \frac{1 + \sigma\left(L(n, \theta, \tau) + 1\right)}{e^{\sigma - 1}}.$$

Since the right-hand side expression is monotonically decreasing in σ and $\sigma \geq 1 + 2\log(L(n, \theta, \tau) + 1)$, it follows that

$$e^{1-\sigma t} \leq \frac{1 + \left(1 + 2\log\left(L(n, \theta, \tau) + 1\right)\right)\left(L(n, \theta, \tau) + 1\right)}{\left(L(n, \theta, \tau) + 1\right)^2}.$$

The expression at the right-hand side is monotonically decreasing in $L(n, \theta, \tau)$. The value at $L(n, \theta, \tau) = 9$ is $0.5705... < e^{-\frac{1}{2}}$. Thus we obtain that $e^{1-\sigma t} < e^{-\frac{1}{2}}$, which implies $1 - \sigma t < -\frac{1}{2}$, or $t > \frac{3}{2\sigma}$, proving the lemma. □

Note that at the start of each inner iteration $\tau < \Psi(v) \leq L(n, \theta, \tau)$. To ensure that $L(n, \theta, \tau)$ satisfies the conditions of Lemma 2.6.3, we assume from now that $L(n, \theta, \tau) \geq 9$. and we choose

$$\sigma = 1 + 2\log\left(L(n, \theta, \tau) + 1\right) \geq 1 + 2\log 10 \approx 5.61. \tag{2.6.8}$$

2.6.3 Lower bound for $\delta(v)$ in terms of $\Psi(v)$

In this section we establish a lower bound of $\delta(v)$ in terms of $\Psi(v)$.

Lemma 2.6.4. *If* $\Psi(v) \geq 1$, *then*

$$\delta(v) \geq \frac{1}{6}\Psi(v)^{\frac{p}{1+p}}. \tag{2.6.9}$$

Proof. The proof of this lemma uses Corollary 2.3.13. So we have to estimate the inverse function ϱ of ψ for $t \in [1, \infty)$. This is obtained by solving t from the equation

$$\psi(t) = \frac{t^{1+p} - 1}{1+p} + \frac{e^{-\sigma(t-1)} - 1}{\sigma} = s, \qquad t \geq 1.$$

Assuming $s \geq 1$, and using $\sigma \geq 1$, we get

$$\frac{t^{1+p} - 1}{1+p} = s + \frac{1 - e^{-\sigma(t-1)}}{\sigma} \leq s + \frac{1}{\sigma} \leq s + 1 \leq 2s,$$

whence

$$t^{1+p} \leq 1 + 2\,(1+p)\,s \leq 3\,(1+p)\,s,$$

and therefore

$$\varrho(s) = t \leq (3\,(1+p)\,s)^{\frac{1}{1+p}} \leq 3s^{\frac{1}{1+p}}, \quad \text{for} \quad p \in [0,1], \quad s \geq 1.$$

Assuming $\Psi(v) \geq 1$, we thus have

$$\varrho(\Psi(v)) \leq 3\Psi(v)^{\frac{1}{1+p}}.$$

Now, using Corollary 2.3.13, we obtain

$$\delta(v) \geq \frac{\Psi(v)}{2\rho\,(\Psi(v))} \geq \frac{1}{6}\Psi(v)^{\frac{p}{1+p}}.$$

This proves the lemma. □

Note that if $\Psi(v) \geq 1$, substitution in (2.6.9) gives

$$\delta(v) \geq \frac{1}{6}. \tag{2.6.10}$$

2.6.4 Decrease of the proximity during a (damped) Newton step

After a damped step, with step size α, we have as before,

$$f(\alpha) := \Psi\,(v_+) - \Psi\,(v) := \Psi\left(\sqrt{(v + \alpha d_x)\,(v + \alpha d_s)}\right) - \Psi\,(v)\,.$$

For the moment we assume that the step size α is such that:

$$v_i + \alpha d_{xi} \geq \frac{1}{\sigma}, \quad v_i + \alpha d_{si} \geq \frac{1}{\sigma}, \quad 1 \leq i \leq n. \tag{2.6.11}$$

Later in the proof we show that this assumption is valid (cf. (2.6.16)). Now ψ is e-convex, so we can use the same argument as before to derive the following results

$$\widetilde{\alpha} = \frac{1}{\psi''\left(\rho(2\delta)\right)} \leq \bar{\alpha} \quad \text{and} \quad f(\widetilde{\alpha}) \leq -\frac{\delta^2}{\psi''\left(\rho(2\delta)\right)}, \tag{2.6.12}$$

where $\bar{\alpha}$ satisfies

$$\frac{1}{\psi''\left(\rho\left(2\delta\right)\right)} \leq \bar{\alpha} \leq \frac{1}{\psi''\left(\rho\left(\delta\right)\right)}.$$

Lemma 2.6.5. *Let $\rho : [0, -\frac{1}{2}\psi'(0)) \to (0, 1]$ denote the inverse function of $-\frac{1}{2}\psi'(t)$ restricted to the interval $(0, 1]$, and $\widetilde{\alpha}$ as in (2.6.12) and $\Psi(v) \geq 1$. Then*

$$f(\widetilde{\alpha}) \leq -\frac{\delta}{16\sigma}. \tag{2.6.13}$$

Proof. To obtain the inverse function $t = \rho(s)$ of $-\frac{1}{2}\psi'(t)$ for $t \in [\frac{1}{\sigma}, 1]$ we need to solve t from the equation

$$-\psi(t) = -\left(t^p - e^{\sigma(1-t)}\right) = 2s.$$

This implies, using $t \leq 1$,

$$e^{\sigma(1-t)} = 2s + t^p \leq 2s + 1.$$

Hence, putting $t = \rho(2\delta)$, which is equivalent to $4\delta = -\psi'(t)$, we get

$$e^{\sigma(1-t)} \leq 4\delta + 1. \tag{2.6.14}$$

Thus we have, using $t \geq \frac{1}{\sigma}$, and $p \in [0, 1]$,

$$\widetilde{\alpha} = \frac{1}{\psi''(t)} = \frac{1}{pt^{p-1} + \sigma e^{\sigma(1-t)}} \geq \frac{1}{p\sigma^{1-p} + \sigma e^{\sigma(1-t)}} \geq \frac{1}{\sigma\left(1 + e^{\sigma(1-t)}\right)}.$$

Also using (2.6.14) we get, using (2.6.10) (i.e., $6\delta \geq 1$),

$$\widetilde{\alpha} \geq \frac{1}{\sigma\left(2 + 4\delta\right)} = \frac{1}{2\sigma\left(1 + 2\delta\right)} \geq \frac{1}{2\sigma\left(6\delta + 2\delta\right)} = \frac{1}{16\sigma\delta}.$$

Hence

$$f(\widetilde{\alpha}) \leq -\frac{\delta^2}{16\sigma\delta} = -\frac{\delta}{16\sigma}.$$

Thus the theorem follows. □

In the sequel

$$\hat{\alpha} = \frac{1}{16\sigma\delta} \tag{2.6.15}$$

will be our default step size. Finally, to validate the above analysis we need to show that $\hat{\alpha}$ satisfies (2.6.11). This is now easy. Using (2.3.8) and Lemma 2.6.3, we may write

$$v_1 - 2\hat{\alpha}\delta \geq \frac{3}{2\sigma} - \frac{2\delta}{16\delta\sigma} \geq \frac{3}{2\sigma} - \frac{1}{8\sigma} = \frac{11}{8\sigma} > \frac{1}{\sigma}. \tag{2.6.16}$$

Using (2.6.9), by substitution in (2.6.13), gives

$$f(\tilde{\alpha}) \leq -\frac{\delta}{16\sigma} \leq -\frac{\Psi^{\frac{p}{1+p}}}{96\sigma}.$$

Lemma 2.6.6. *If K denotes the number of inner iterations between two subsequent updates of μ, we have*

$$K \leq 96\sigma\,(1+p)\,\Psi_0^{\frac{1}{1+p}}.$$

Proof. The definition of K implies $\Psi_{K-1} > \tau$ and $\Psi_K \leq \tau$ and

$$\Psi_{k+1} \leq \Psi_k - \kappa\,(\Psi_k)^{1-\beta}, \qquad k = 0, 1, \cdots, K-1,$$

with $\kappa = \frac{1}{96\sigma}$ and $\beta = \frac{1}{1+p}$. Application of Lemma A.1.2, with $t_k = \Psi_k$ yields the desired inequality. □

2.6.5 Complexity

In this section we will derive the complexity bounds for large-update methods and small-update methods. Using $\psi_0 \leq L(n,\theta,\tau)$, and Lemma 2.6.6 we obtain the following upper bound on the total number of iterations:

$$\frac{96\sigma\,(1+p)\,L(n,\theta,\tau)^{\frac{1}{1+p}}}{\theta}\log\frac{n}{\epsilon} \leq \frac{192\sigma L(n,\theta,\tau)^{\frac{1}{1+p}}}{\theta}\log\frac{n}{\epsilon}, \qquad p \in [0,1]. \tag{2.6.17}$$

Large-update methods

We just established that (2.6.17) is an upper bound for the total number of iterations, where the number $L(n,\theta,\tau)$ is as given in (2.6.6):

$$L(n,\theta,\tau) = n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right). \tag{2.6.18}$$

To estimate this number we need an upper bound for the inverse function ϱ of $\psi(t)$ for $t \in [1, \infty)$. So, if $\psi(t) = s$, $t \geq 1$, we need an upper bound for t. Using $\sigma \geq 1$, one has

$$\frac{t^{1+p} - 1}{1+p} = \psi(t) + \frac{1 - e^{-\sigma(t-1)}}{\sigma} \leq \psi(t) + \frac{1}{\sigma} \leq 1 + \psi(t),$$

which gives

$$t \leq \left((1+p)\left(\psi(t)+1\right)+1\right)^{\frac{1}{1+p}} \leq \left(2\psi(t)+3\right)^{\frac{1}{1+p}} = (2s+3)^{\frac{1}{1+p}}, \tag{2.6.19}$$

for all $p \in [0, 1]$. By Lemma 2.6.2 we have

$$t \leq 1 + \sqrt{t\psi(t)}.$$

Now substituting (2.6.19) we obtain

$$\varrho(s) = t \leq 1 + (2s+3)^{\frac{1}{2(1+p)}} \sqrt{s}. \tag{2.6.20}$$

Using

$$\psi(t) = \frac{t^{1+p} - 1}{p+1} + \frac{e^{\sigma(1-t)} - 1}{\sigma} \leq \frac{t^{1+p}}{1+p} \leq t^{1+p}, \quad \text{for} \quad t \geq 1,$$

and (2.6.20), by substitution in (2.6.18) we obtain

$$L(n, \theta, \tau) \leq \frac{n\varrho\left(\frac{\tau}{n}\right)^{1+p}}{(1-\theta)^{\frac{1+p}{2}}} \leq \frac{n}{(1-\theta)^{\frac{1+p}{2}}} \left(1 + \left(\frac{2\tau}{n} + 3\right)^{\frac{1}{2(1+p)}} \sqrt{\frac{\tau}{n}}\right)^{1+p}.$$

Using (2.6.17), thus the total number of iterations is bounded above by

$$\frac{K}{\theta} \log \frac{n}{\epsilon} \leq \frac{192\sigma}{\theta\sqrt{1-\theta}} \left(n\left(1 + \left(\frac{2\tau}{n} + 3\right)^{\frac{1}{2(1+p)}} \sqrt{\frac{\tau}{n}}\right)^{1+p}\right)^{\frac{1}{1+p}} \log \frac{n}{\epsilon}.$$

A large-update methods uses $\tau = O(n)$ and $\theta = \Theta(1)$ and $\sigma = O(\log n)$. The right-hand side expression is then $O\left(n^{\frac{1}{1+p}} (\log n) \log \frac{n}{\epsilon}\right)$, as easily may be verified.

Small-update methods

For small-update methods one has $\tau = O(1)$ and $\theta = \Theta\left(\frac{1}{\sqrt{n}}\right)$. Using Lemma 2.2.7, with $\psi''(1) = p + \sigma$, we then obtain

$$L(n, \theta, \tau) = n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right) \leq \frac{n\,(p+\sigma)}{2}\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}} - 1\right)^2 \leq n\sigma\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}} - 1\right)^2,$$

for $\sigma \geq 1$, and $p \in [0,1]$. Using (2.6.20), then

$$L(n,\theta,\tau) \leq n\sigma \left(\frac{1+\sqrt{2}\sqrt{\left(\frac{\tau}{n}\right)^2+\frac{\tau}{n}}}{\sqrt{1-\theta}} - 1 \right)^2 .$$

Using $1-\sqrt{1-\theta} = \frac{\theta}{1+\sqrt{1-\theta}} \leq \theta$, this leads to

$$L(n,\tau,\theta) \leq \sigma \frac{\left(\theta\sqrt{n}+\sqrt{2}\sqrt{\frac{\tau^2}{n}+\tau}\right)^2}{1-\theta} = \sigma O(1). \tag{2.6.21}$$

Using (2.6.8) (i.e., $\sigma = 1+2\log\left(1+L(n,\theta,\tau)\right)$), by (2.6.21) we have

$$\sigma \leq 1+2\log\left(1+\sigma O(1)\right). \tag{2.6.22}$$

This implies that $\sigma = O(1)$. Then $L(n,\theta,\tau) = O(1)$. Using (2.6.17), thus the total number of iterations is bounded above by

$$\frac{K}{\theta}\log\frac{n}{\epsilon} \leq \frac{192}{\theta}O(1)\log\frac{n}{\epsilon} = O\left(\sqrt{n}\log\frac{n}{\epsilon}\right).$$

Chapter 3

Primal-Dual IPMs for SDO based on kernel functions

3.1 Introduction

A semidefinite optimization problem (SDO) is a convex optimization problem in the space of symmetric matrices. We consider the standard semidefinite programming problem

$$(SDP) \quad p^* = \inf_X \{\mathrm{Tr}(CX) : \mathrm{Tr}(A_i X) = b_i (1 \leq i \leq m), X \succeq 0\},$$

and its dual problem (SDD)

$$(SDD) \quad d^* = \sup_{y,S} \left\{ b^T y : \sum_{i=1}^{m} y_i A_i + S = C, S \succeq 0 \right\},$$

where C and A_i are symmetric $n \times n$ matrices, $b, y \in \mathbf{R}^m$, and $X \succeq 0$ means that X is symmetric positive semidefinite and $\mathrm{Tr}(A)$ denotes the trace of A (i.e., the sum of its diagonal elements). The matrices A_i are further assumed to be linearly independent (without loss of generality). Recall that for any two $n \times n$ matrices, A and B

$$\mathrm{Tr}(A^T B) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ij}.$$

Interior point methods $(IPMs)$ provide a powerful approach for solving SDO problems. A comprehensive list of publications on this topic can be found in the SDO homepage maintained by Alizadeh [Ali91]. The pioneering works in this direction are due to Alizadeh [Ali91; Ali95] and Nesterov and Nemirovskii [Nes93]. Most IPMs for SDO can be viewed as natural extensions of $IPMs$ for LO, and have similar polynomial complexity results. However, to obtain valid search directions is much more difficult than in the LO case. Below we describe how the usual search directions are obtained for primal-dual methods for solving SDO problems. Our aim is to show in this section that the kernel-function-based approach that we presented for LO in Chapter 2, can be applied also to SDO problems, then yielding a wide class of new methods. For self-regular kernel functions this has been earlier in [Pen02a; Pen02b]. Just as in the LO case, the new methods have the same iteration complexity when small-updates are used, but the iteration complexity is better for large-updates methods.

3.1.1 Classical search direction

We assume that a strictly feasible pair $(X \succ 0, S \succ 0)$ exists, which is the interior-point condition (IPC) for SDO. This ensures the existence of an optimal primal-dual pair (X^*, S^*) with zero duality gap. Hence one can write the optimality conditions for the primal-dual pair of problems as follows.

$$\begin{aligned} \operatorname{Tr}(A_iX) &= b_i, \quad i = 1, \dots, m \\ \sum_{i=1}^{m} y_iA_i + S &= C \\ XS &= 0 \\ X, S &\succeq 0. \end{aligned} \tag{3.1.1}$$

The basic idea of primal-dual $IPMs$ is to replace the above complementarity condition $XS = 0$ by the parameterized equation

$$XS = \mu E; \quad X, S \succ 0,$$

where E denotes the $n \times n$ identity matrix and $\mu > 0$. The resulting system has a unique solution for each $\mu > 0$. This solution is denoted by $(X(\mu), y(\mu), S(\mu))$ for each $\mu > 0$; $X(\mu)$ is called the μ-center of (SDP) and $(y(\mu), S(\mu))$ is the μ-center of (SDD). The set of μ-centers (with $\mu > 0$) defines a homotopy path, which is called *the central path* of (SDP) and (SDD) [Kle97; Pen02a; Pen02b]. The principal idea of $IPMs$ is to follow this central path and approach the optimal set of SDP as μ goes to zero. Newton's method amounts to linearizing the system

(3.1.1), thus yielding the following system of equations.

$$
\begin{aligned}
\operatorname{Tr}(A_i \Delta X) &= b_i, \quad i = 1, \ldots, m. \\
\sum_{i=1}^{m} \Delta y_i A_i + \Delta S &= 0 \\
X \Delta S + \Delta X S &= \mu E - XS.
\end{aligned}
\tag{3.1.2}
$$

This so-called Newton system has a unique solution $(\Delta X, \Delta y, \Delta S)$. Note that ΔS is symmetric, due to the second equation in (3.1.2). However, a crucial point is that ΔX may be not symmetric. Many researchers have proposed various ways of 'symmetrizing' the third equation in the Newton system so that the new system has a unique symmetric solution. All these proposals can be described by using a symmetric nonsingular scaling matrix P and by replacing (3.1.2) by the system

$$
\begin{aligned}
\operatorname{Tr}(A_i \Delta X) &= b_i, \quad i = 1, \ldots, m \\
\sum_{i=1}^{m} \Delta y_i A_i + \Delta S &= 0 \\
\Delta X + P \Delta S P^T &= \mu S^{-1} - X
\end{aligned}
\tag{3.1.3}
$$

Now ΔX is automatically a symmetric matrix. Some popular choices for the matrix P are listed in Table 3.1.

P	References
$X^{\frac{1}{2}} \left(X^{\frac{1}{2}} S X^{\frac{1}{2}} \right)^{-\frac{1}{2}} X^{\frac{1}{2}}$	[Nes97]
X^{-1}	[Koj94; Mon97]
S	[Koj94; Mon97]
I	[Ali96]

Table 3.1: *Choices for the scaling matrix P.*

3.1.2 Nesterov-Todd direction

In this thesis we consider the symmetrization schema of Nesterov-Todd. So we use

$$P = X^{\frac{1}{2}} \left(X^{\frac{1}{2}} S X^{\frac{1}{2}} \right)^{-\frac{1}{2}} X^{\frac{1}{2}} = S^{-\frac{1}{2}} \left(S^{\frac{1}{2}} X S^{\frac{1}{2}} \right)^{\frac{1}{2}} S^{-\frac{1}{2}}, \tag{3.1.4}$$

where the last equality can be easily verified. Let $D = P^{\frac{1}{2}}$, where $P^{\frac{1}{2}}$ denotes the symmetric square root of P. Now, the matrix D can be used to scale X and S to the same matrix V, defined by [Stu99; Kle02]:

$$V := \frac{1}{\sqrt{\mu}} D^{-1} X D^{-1} = \frac{1}{\sqrt{\mu}} DSD. \tag{3.1.5}$$

Obviously the matrices D and V are symmetric, and positive definite. Let us further define

$$\bar{A}_i := DA_iD, \quad i = 1, 2, \ldots, m;$$

and

$$D_X := \frac{1}{\mu} D^{-1} \Delta X D^{-1}; \quad D_S := \frac{1}{\mu} D \Delta S D \tag{3.1.6}$$

Then it follows from (3.1.3)

$$\begin{aligned} \operatorname{Tr}(\bar{A}_i D_X) &= 0, i = 1, \ldots, m. \\ \sum_{i=1}^{m} \Delta y_i \bar{A}_i + D_S &= 0 \\ D_X + D_S &= V^{-1} - V. \end{aligned} \tag{3.1.7}$$

In the sequel, we use the following notational conventions. Throughout this chapter, $\|\cdot\|$ denotes the 2-norm of a vector. The nonnegative and the positive orthants are denoted as $\mathbf{R}_+^n$ and $\mathbf{R}_{++}^n$, respectively, and $\mathbf{S^n}$, $\mathbf{S^n_+}$, and $\mathbf{S^n_{++}}$ denote the cone of symmetric, symmetric positive semidefinite and symmetric positive definite $n \times n$ matrices, respectively. For any $V \in \mathbf{S^n_{++}}$, we denote by $\lambda(V)$ the vector of eigenvalues of V arranged in increasing order, that is, $\lambda_1(V) \leq \lambda_2(V) \leq, \ldots, \lambda_n(V)$. For any matrix A, we denote by $\eta_1(A) \leq \eta_2(A) \leq, \ldots, \leq \eta_n(A)$ the singular values of A; if A is symmetric, then one has $\eta_i(A) = |\lambda_i(A)|$, $i = 1, 2, \ldots, n$. Finally, if $z \in \mathbf{R}_+^n$ and $f : \mathbf{R}_+ \rightarrow \mathbf{R}_+$, then $f(z)$ denotes the vector in $\mathbf{R}_+^n$ whose i-th component is $f(z_i)$, with $1 \leq i \leq n$.

3.2 New search direction

In this section we introduce the definition of a matrix function [Hor85; Rud78].

Definition 3.2.1. Let X be a symmetric matrix, and let

$$X = Q_X^{-1}\mathrm{diag}(\lambda_1(X), \lambda_2(X), \ldots, \lambda_n(X))Q_X, \tag{3.2.1}$$

be an eigenvalue decomposition of X, where $\lambda_i(X)$, $1 \leq i \leq n$ denote the eigenvalues of X, and Q_X is orthogonal. If $\psi(t)$ is any univariante function whose domain contains $\{\lambda_i(X); 1 \leq i \leq n\}$ then the matrix function $\psi(X)$ is defined by

$$\psi(X) = Q_X^{-1}\mathrm{diag}(\psi(\lambda_1(X)), \psi(\lambda_2(X)), \ldots, \psi(\lambda_n(X)))Q_X. \tag{3.2.2}$$

Define the barrier function $\Psi(X)$ as follows [Pen02b].

$$\Psi(X) := \sum_{i=1}^{n} \psi(\lambda_i(X)) = \mathrm{Tr}(\psi(X)). \tag{3.2.3}$$

In this chapter, when we use the function $\psi(.)$ and its first three derivatives $\psi'(.)$, $\psi''(.)$, and $\psi'''(.)$ without any specification, it denotes a matrix function if the argument is a matrix and a univariate function (from $\mathbf{R}$ to $\mathbf{R}$) if the argument is in $\mathbf{R}$.

Following [Pen02a; Pen02b] we describe the kernel-function-based approach to SDO. Given the kernel function $\psi(t)$ and the associated $\psi(V)$ and $\psi'(V)$ as defined in Definition (3.2.1), we replace the right-hand side $V - V^{-1}$ in the third equation in (3.1.7) by $-\psi'(V)$. Thus we consider the following system.

$$\begin{aligned} \mathrm{Tr}(\bar{A}_i D_X) &= 0, i = 1, \quad \ldots, m. \\ \sum_{i=1}^{m} \Delta y_i \bar{A}_i + D_S &= 0 \\ D_X + D_S &= -\psi'(V). \end{aligned} \tag{3.2.4}$$

Having D_X and D_S, $\triangle X$ and $\triangle S$ can be calculated from (3.1.6). Due to the orthogonality of $\triangle X$ and $\triangle S$, it is trivial to see that $D_X \perp D_S$, and so

$$\mathrm{Tr}(D_X D_S) = \mathrm{Tr}(D_S D_X) = 0. \tag{3.2.5}$$

The algorithm considered in this chapter is described in Figure 3.1.

Just as in the LO case, the parameters τ, θ, and the step size α should be chosen in such a way that the algorithm is 'optimized' in the sense that the

Generic Primal-Dual Algorithm for SDO

Input:
a kernel function $\psi(t)$;
a threshold parameter $\tau > 0$;
an accuracy parameter $\epsilon > 0$;
a barrier update parameter θ, $0 < \theta < 1$;
begin
$X := X_0$; $S := S_0$; $\mu := \mu_0$;
while $n\mu \geq \epsilon$ **do**
begin
$\mu := (1-\theta)\mu$;
while $\Phi(X, S, \mu) \geq \tau$ **do**
begin
Solve system (3.1.3) for $\triangle X$, $\triangle y$, $\triangle S$;
Determine a step size α;
$X := X + \alpha \triangle X$;
$S := S + \alpha \triangle S$;
$y := y + \alpha \Delta y$;
end
end
end

Figure 3.1: *Generic primal-dual interior-point algorithm for SDO.*

number of iterations required by the algorithm is as small as possible. Obviously, the resulting iteration bound will depend on the kernel function underlying the algorithm, and our main task becomes to find a kernel function that minimizes the iteration bound.

The chapter is organized as follows. In Section 3.3 we start by deriving some properties of the barrier function $\Psi(V)$. The estimate of the step size and the decrease behavior of the barrier function are discussed in Section 3.4. Finally, the total iteration bound of the algorithm and the complexity results are derived in Section 3.5.

3.3 Properties of $\Psi(V)$ and $\delta(V)$

In this section we extend two theorems from Chapter 2. The first theorem follows from (2.2.3-e) and the second theorem is a consequence of (2.2.3-c). The proofs of the following theorems are based on Theorems 2.3.1 and 2.3.11 in Chapter 2.

Lemma 3.3.1. *Let $\psi(t)$ be a kernel function. Let the matrix functions $\psi(X)$, $\Psi(X)$ be defined by (3.2.2) and (3.2.3) respectively. Then $\Psi(X)$ is strictly convex with respect to $X \succ 0$, and $\psi(E) = \psi'(E) = 0_{n\times n}$.*

Proof. See [Pen02a; Pen02b].

Theorem 3.3.2. *With ϱ as defined in (2.3.1), as the inverse function of ψ on $[1, \infty)$, we have for any positive definite matrix V, and any $\beta > 1$,*

$$\Psi(\beta V) \le n\psi\left(\beta\varrho\left(\frac{\Psi(V)}{n}\right)\right).$$

Proof. We consider the following maximization problem:

$$\max_V \left\{ \Psi(\beta V) = \sum_{i=1}^n \psi(\beta\lambda_i(V)) \ : \ \Psi(V) = \sum_{i=1}^n \psi(\lambda_i(V)) = r \right\},$$

where r is any nonnegative number. Let $v_i := \lambda_i(V)$, $1 \le i \le n$. Then $v > 0$ and

$$\Psi(\beta V) = \sum_{i=1}^n \psi(\beta\lambda_i(V)) = \sum_{i=1}^n \psi(\beta v_i) = \Psi(\beta v).$$

Using Theorem 2.3.1 we get

$$\begin{aligned}
\Psi(\beta v) \le n\psi\left(\beta\varrho\left(\frac{\Psi(v)}{n}\right)\right) &= n\psi\left(\beta\varrho\left(\frac{\sum_{i=1}^n \psi(v_i)}{n}\right)\right) \\
&= n\psi\left(\beta\varrho\left(\frac{\sum_{i=1}^n \psi(\lambda_i(V))}{n}\right)\right) \\
&= n\psi\left(\beta\varrho\left(\frac{\Psi(V)}{n}\right)\right).
\end{aligned}$$

This proves the theorem. □

The next theorem gives a lower bound on the norm-based proximity measure $\delta(V)$, defined by

$$\delta(V) = \tfrac{1}{2}\|\psi'(V)\| = \frac{1}{2}\sqrt{\sum_{i=1}^n \psi'(\lambda_i(V))^2} = \frac{1}{2}\left\|D_X + D_S\right\|, \tag{3.3.1}$$

in terms of $\Psi(V)$. Since $\Psi(V)$ is strictly convex and attains its minimal value zero at $V = E$, we have

$$\Psi\left(V\right) = 0 \quad \Leftrightarrow \quad \delta\left(V\right) = 0 \quad \Leftrightarrow \quad V = E.$$

Theorem 3.3.3. *Let ϱ be as defined in (2.3.1). Then*

$$\delta(V) \geq \tfrac{1}{2}\psi'\left(\varrho\left(\Psi(V)\right)\right).$$

Proof. The statement in the lemma is obvious if $V = E$ since then $\delta(V) = \Psi(V) = 0$. Otherwise we have $\delta(V) > 0$ and $\Psi(V) > 0$. To deal with the nontrivial case we consider, for $\gamma > 0$, the problem

$$z_\gamma = \min_V \left\{ \delta(V)^2 = \tfrac{1}{4}\sum_{i=1}^{n} \psi'(\lambda_i(V))^2 \;:\; \Psi(V) = \gamma \right\}.$$

Again, let $v_i := \lambda_i(V)$, $1 \leq i \leq n$. Then $v > 0$ and

$$\tfrac{1}{4}\sum_{i=1}^{n} \psi'(\lambda_i(V))^2 = \tfrac{1}{4}\sum_{i=1}^{n} \psi'(v_i)^2 = \delta(v)^2.$$

Using Theorem 2.3.11 we get

$$\begin{aligned}
\delta(v) \geq \tfrac{1}{2}\psi'\left(\varrho\left(\Psi(v)\right)\right) &= \tfrac{1}{2}\psi'\left(\varrho\left(\sum_{i=1}^{n}\psi(v_i)\right)\right) \\
&= \tfrac{1}{2}\psi'\left(\varrho\left(\sum_{i=1}^{n}\psi(\lambda_i(V))\right)\right) \\
&= \tfrac{1}{2}\psi'\left(\varrho\left(\Psi(V)\right)\right).
\end{aligned}$$

This completes the proof of the theorem. □

3.4 Analysis of the algorithm

In the analysis of the algorithm the concept of exponential convexity is again a crucial ingredient. We start with two technical lemmas.

Lemma 3.4.1 (Lemma 3.3.14 (c) in [Hor85]). *Let $A, B \in \mathbf{S}^n$ be two nonsingular matrices and $f(t)$ a real-valued function such that $f(e^t)$ is a convex function. One has*

$$\sum_{i=1}^{n} f(\eta_i(AB)) \leq \sum_{i=1}^{n} f(\eta_i(A)\eta_i(B)), \tag{3.4.1}$$

where $\eta_i(A)$ *and* $\eta_i(B)$*,* $i = 1, 2, \dots, n$ *denote the singular values of* A *and* B*, respectively.*

Lemma 3.4.2 (Lemma 5.1 in [Wan04]). *Let* A*,* $A + B \in \mathbf{S}^{\mathbf{n}}_{+}$*. Then one has*

$$\lambda_i(A+B) \geq \lambda_1(A) - |\lambda_n(B)|\,, \quad i = 1, 2, \dots, n. \tag{3.4.2}$$

Lemma 3.4.3. *Let* V_1 *and* V_2 *are two symmetric positive definite. Then*

$$\Psi\left((V_1^{\frac{1}{2}} V_2 V_1^{\frac{1}{2}})^{\frac{1}{2}}\right) \leq \frac{1}{2}\left(\Psi(V_1) + \Psi(V_2)\right). \tag{3.4.3}$$

Proof. For any nonsingular matrix $U \in \mathbf{S}^n$, we have

$$\eta_i(U) = \left(\lambda_i(U^T U)\right)^{\frac{1}{2}} = \left(\lambda_i(UU^T)\right)^{\frac{1}{2}}, \quad i = 1, 2, \dots, n.$$

From this, we can write

$$\eta_i(V_1^{\frac{1}{2}} V_2^{\frac{1}{2}}) = \left(\lambda_i(V_1^{\frac{1}{2}} V_2 V_1^{\frac{1}{2}})\right)^{\frac{1}{2}} = \lambda_i\left((V_1^{\frac{1}{2}} V_2 V_1^{\frac{1}{2}})^{\frac{1}{2}}\right), \quad i = 1, 2, \dots, n.$$

Since V_1 and V_2 are symmetric positive definite, using Lemma 3.4.1 one has

$$\begin{aligned}
\Psi\left((V_1^{\frac{1}{2}} V_2 V_1^{\frac{1}{2}})^{\frac{1}{2}}\right) &= \sum_{i=1}^{n} \psi\left(\eta_i(V_1^{\frac{1}{2}} V_2^{\frac{1}{2}})\right) \leq \sum_{i=1}^{n} \psi\left(\eta_i(V_1^{\frac{1}{2}})\eta_i(V_2^{\frac{1}{2}})\right) \\
&\leq \frac{1}{2}\sum_{i=1}^{n}\left(\psi\left(\eta_i^2(V_1^{\frac{1}{2}})\right) + \psi\left(\eta_i^2(V_2^{\frac{1}{2}})\right)\right) \\
&= \frac{1}{2}\sum_{i=1}^{n}\left(\psi\left(\eta_i(V_1)\right) + \psi\left(\eta_i(V_2)\right)\right) = \frac{1}{2}\left(\Psi(V_1) + \Psi(V_2)\right).
\end{aligned}$$

The second inequality follows from the exponential convexity of $\psi(t)$. This completes the proof of lemma. □

3.4.1 Decrease of the barrier function during a (damped) Newton step

In this section we start to compute the step size. After a damped step, with step size α, using (3.1.6) we have

$$\begin{aligned}
&X_+ = X + \alpha \triangle X = X + \alpha\sqrt{\mu} D D_X D = \sqrt{\mu} D\left(V + \alpha D_X\right) D, \\
&y_+ = y + \alpha \Delta y, \\
&S_+ = S + \alpha \triangle S = X + \alpha\sqrt{\mu} D^{-1} D_S D^{1} = \sqrt{\mu} D^{-1}\left(V + \alpha D_S\right) D^{-1}.
\end{aligned}$$

One has [Pen02b]

$$V_+ = \frac{1}{\sqrt{\mu}} \left(D^{-1} X_+ S_+ D\right)^{\frac{1}{2}}. \tag{3.4.4}$$

Note that V_+^2 is unitarily similar to the matrix $X_+^{\frac{1}{2}} S_+ X_+^{\frac{1}{2}}$ and thus to

$$(V + \alpha D_X)^{\frac{1}{2}} (V + \alpha D_S) (V + \alpha D_X)^{\frac{1}{2}}.$$

This implies that the eigenvalues of V_+ are the same as those of the matrix

$$\tilde{V}_+ := \left((V + \alpha D_X)^{\frac{1}{2}} (V + \alpha D_S) (V + \alpha D_X)^{\frac{1}{2}}\right)^{\frac{1}{2}}. \tag{3.4.5}$$

By the definition of $\Psi(V)$, we have $\Psi(V_+) = \Psi\left(\tilde{V}_+\right)$.
Our aim is to find an upper bound for

$$f(\alpha) := \Psi(V_+) - \Psi(V) = \Psi\left(\tilde{V}_+\right) - \Psi(V).$$

To do this we will use Lemma 3.4.3, so

$$\begin{aligned} \Psi\left(\tilde{V}_+\right) &= \Psi\left(\left((V + \alpha D_X)^{\frac{1}{2}} (V + \alpha D_S) (V + \alpha D_X)^{\frac{1}{2}}\right)^{\frac{1}{2}}\right) \\ &\leq \tfrac{1}{2} \left[\Psi(V + \alpha D_X) + \Psi(V + \alpha D_S)\right]. \end{aligned}$$

Thus we have $f(\alpha) \leq f_1(\alpha)$, where

$$f_1(\alpha) := \tfrac{1}{2} \left[\Psi(V + \alpha D_X) + \Psi(V + \alpha D_S)\right] - \Psi(V)$$

is convex in α, since Ψ is convex. Obviously, $f(0) = f_1(0) = 0$. Taking the derivative to α, we get

$$f_1'(\alpha) = \tfrac{1}{2} \mathrm{Tr}\left(\psi'(V + \alpha D_X) D_X + \psi'(V + \alpha D_S) D_S\right).$$

This gives, using the last equalities in (3.2.4) and (3.3.1),

$$f_1'(0) = \tfrac{1}{2} \mathrm{Tr} \psi'(V) (D_X + D_S) = -\tfrac{1}{2} \mathrm{Tr}\left(\psi'(V)^2\right) = -2\delta(V)^2. \tag{3.4.6}$$

Differentiating once more, we obtain

$$f_1''(\alpha) = \tfrac{1}{2} \mathrm{Tr}\left(\psi''(V + \alpha D_X) D_X^2 + \psi''(V + \alpha D_S) D_S^2\right). \tag{3.4.7}$$

Below we use the following notation:

$$\delta := \delta(V).$$

Lemma 3.4.4. *One has*

$$f_1''(\alpha) \leq 2\delta^2 \, \psi'' \left(\lambda_1(V) - 2\alpha\delta\right).$$

Proof. The last equality in (3.2.4) and (3.3.1) imply that $\|D_X + D_S\|^2 = \|D_X\|^2 + \|D_S\|^2 = 4\delta^2$. Thus we have $|\lambda_n(D_X)| \leq 2\delta$ and $|\lambda_n(D_S)| \leq 2\delta$. Using Lemma 3.4.2 and $V + \alpha D_X \succeq 0$, therefore,

$$\lambda_i(V + \alpha D_X) \geq \lambda_1(V) - \alpha \, |\lambda_n(D_X)| \geq \lambda_1(V) - 2\alpha\delta, \quad 1 \leq i \leq n, \tag{3.4.8}$$

$$\lambda_i(V + \alpha D_S) \geq \lambda_1(V) - \alpha \, |\lambda_n(D_S)| \geq \lambda_1(V) - 2\alpha\delta, \quad 1 \leq i \leq n. \tag{3.4.9}$$

Since ψ'' is monotonically decreasing, using the above inequalities, we get

$$\psi''(\lambda_i(V + \alpha D_X)) \leq \psi''(\lambda_1(V) - 2\alpha\delta), \tag{3.4.10}$$

and

$$\psi''(\lambda_i(V + \alpha D_S)) \leq \psi''(\lambda_1(V) - 2\alpha\delta). \tag{3.4.11}$$

This implies that

$$\psi''(V + \alpha D_X) \preceq \psi''(\lambda_1(V) - 2\alpha\delta)E, \tag{3.4.12}$$

$$\psi''(V + \alpha D_S) \preceq \psi''(\lambda_1(V) - 2\alpha\delta)E. \tag{3.4.13}$$

Now, using (3.2.5), and $\|D_X\|^2 + \|D_S\|^2 = 4\delta^2$, by (3.4.7) we obtain

$$\begin{aligned} f_1''(\alpha) &\leq \tfrac{1}{2} \, \psi'' \left(\lambda_1(V) - 2\alpha\delta\right) \sum_{i=1}^{n} \left(\lambda_i(D_X^2) + \lambda_i(D_X^2)\right) \\ &= 2\delta^2 \, \psi'' \left(\lambda_1(V) - 2\alpha\delta\right). \end{aligned}$$

This proves the lemma. □

Putting $v_i = \lambda_i(X)$, $1 \leq i \leq n$, we have

$$f_1''(\alpha) \leq 2\delta^2 \, \psi'' \left(v_1 - 2\alpha\delta\right),$$

which is the same inequality as in Lemma 2.3.4. From this stage on we can apply exactly the same argument as in the LO case to obtain the following results which require no further proof.

Lemma 3.4.5. *One has $f_1'(\alpha) \leq 0$ if α satisfies the inequality*

$$-\psi' \left(\lambda_1(V) - 2\alpha\delta\right) + \psi' \left(\lambda_1(V)\right) \leq 2\delta. \tag{3.4.14}$$

Lemma 3.4.6. *With ρ as defined in (2.3.2), as the inverse function of $-\frac{1}{2}\psi'(t)$ for $t \in (0,1]$, the step size*

$$\bar{\alpha} := \frac{1}{2\delta}\left[\rho(\delta) - \rho(2\delta)\right] \tag{3.4.15}$$

is the largest possible solution of inequality (3.4.14).

Lemma 3.4.7. *Let ρ and $\bar{\alpha}$ be as defined in Lemma 3.4.6. Then*

$$\frac{1}{\psi''(\rho(2\delta))} \le \bar{\alpha} \le \frac{1}{\psi''(\rho(\delta))}.$$

As in the LO case, we use

$$\widetilde{\alpha} = \frac{1}{\psi''(\varrho(2\delta))}, \tag{3.4.16}$$

as the default step size. By Lemma 3.4.7 we have $\widetilde{\alpha} \le \bar{\alpha}$.

Lemma 3.4.8. *If the step size α is such that $\alpha \le \bar{\alpha}$ then*

$$f(\alpha) \le -\alpha\,\delta^2. \tag{3.4.17}$$

Theorem 3.4.9. *Let ρ be as defined in (2.3.2) and $\widetilde{\alpha}$ as in (3.4.16). Then*

$$f(\widetilde{\alpha}) \le -\frac{\delta^2}{\psi''(\rho(2\delta))}. \tag{3.4.18}$$

The right-hand side expression in (3.4.18) is monotonically decreasing in δ, due to (2.2.3-d).

Using the results of Theorems 3.4.9, 3.3.3 we obtain

$$f(\tilde{\alpha}) \le -\frac{\left(\psi'\left(\varrho\left(\Psi(V)\right)\right)\right)^2}{4\psi''\left(\rho\left(\psi'\left(\varrho\left(\Psi(V)\right)\right)\right)\right)}. \tag{3.4.19}$$

This expresses the decrease in $\Psi(V)$ during an inner iteration completely in terms of $\psi(t)$, its first and second derivatives and the inverse functions ρ and ϱ.

3.5 Iteration bounds

In this section we derive the complexity bounds for large-update methods and small-update methods. Similarly to linear case in Chapter 2, we obtain the following upper bound on the total number of iterations.

$$\frac{\Psi_0^\gamma}{\theta\kappa\gamma}\log\frac{n}{\epsilon} \le \frac{1}{\theta\kappa\gamma}\left(n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right)\right)^\gamma \log\frac{n}{\epsilon}. \tag{3.5.1}$$

3.5.1 Application to the ten kernel functions

It may be clear that just as in the LO case we can use the scheme of Figure 2.2 to analyze the behavior of our algorithm for SDO, as given in Figure 3.1. For any given kernel function $\psi(t)$, this will yield exactly the same complexity results as in the LO case. For the sake of completeness we summarize these results in Table 3.2, both for small-update and for large-update methods.

i	kernel functions ψ_i	Large-update	Small-update
1	$\frac{t^2-1}{2} - \log t$	$O\left(n \log \frac{n}{\epsilon}\right)$	$O\left(\sqrt{n} \log \frac{n}{\epsilon}\right)$
2	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q(q-1)} - \frac{q-1}{q}(t-1), \quad q>1$	$O\left(qn^{\frac{q+1}{2q}} \log \frac{n}{\epsilon}\right)$	$O\left(q\sqrt{n} \log \frac{n}{\epsilon}\right)$
3	$\frac{t^2-1}{2} + \frac{(e-1)^2}{e}\frac{1}{e^t-1} - \frac{e-1}{e}$	$O\left(n^{\frac{3}{4}} \log \frac{n}{\epsilon}\right)$	$O\left(\sqrt{n} \log \frac{n}{\epsilon}\right)$
4	$\frac{1}{2}\left(t - \frac{1}{t}\right)^2$	$O\left(n^{\frac{2}{3}} \log \frac{n}{\epsilon}\right)$	$O\left(\sqrt{n} \log \frac{n}{\epsilon}\right)$
5	$\frac{t^2-1}{2} + e^{\frac{1}{t}-1} - 1$	$O\left(\sqrt{n}\ \log^2 n \log \frac{n}{\epsilon}\right)$	$O\left(\sqrt{n} \log \frac{n}{\epsilon}\right)$
6	$\frac{t^2-1}{2} - \int_1^t e^{\frac{1}{\xi}-1} d\xi$	$O\left(\sqrt{n}\ \log^2 n \log \frac{n}{\epsilon}\right)$	$O\left(\sqrt{n} \log \frac{n}{\epsilon}\right)$
7	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q-1},\ q>1$	$O\left(qn^{\frac{q+1}{2q}} \log \frac{n}{\epsilon}\right)$	$O\left(q^2\sqrt{n} \log \frac{n}{\epsilon}\right)$
8	$t - 1 + \frac{t^{1-q}-1}{q-1}, \quad q>1$	$O\left(qn \log \frac{n}{\epsilon}\right)$	$O\left(q^2\sqrt{n} \log \frac{n}{\epsilon}\right)$
9	$\frac{t^{1+p}-1}{1+p} - \log t,\ p \in [0,1]$	$O\left(n \log \frac{n}{\epsilon}\right)$	$O\left(\sqrt{n} \log \frac{n}{\epsilon}\right)$
10	$\frac{t^{p+1}-1}{p+1} + \frac{t^{1-q}-1}{q-1}, \quad p \in [0,1], \quad q>1$	$O\left(qn^{\frac{p+q}{q(1+p)}} \log \frac{n}{\epsilon}\right)$	$O\left(\sqrt{n} \log \frac{n}{\epsilon}\right)$

Table 3.2: *Complexity results for large- and small-update methods for SDO.*

Chapter 4

Numerical results

The aim of this chapter is to investigate the influence of the choice of the kernel function $\psi(t)$ on the computational behavior of the generic primal-dual algorithm for LO, as given in Figure 1.1.

The kernel functions that we used in our experiments are the ten kernel functions ψ_i, $i \in \{1, \ldots, 10\}$, as presented in Table 2.2, and the kernel function $\psi_{p,\sigma}$ with finite barrier term that was considered in Section 2.6. When there are parameters involved in the definition of a kernel function we used several values of these parameters as indicated in Table 4.1 below. These values were chosen after some preliminary experiments that showed that these values gave the most promising iteration counts for the respective kernel functions. This left us with 26 different kernel functions. For the test problems we used problems from the well-known library Netlib.[1] To limit the number of test problems we used only the problems in this library that are known to have optimal solutions. This left us with 95 test problems. More information on these problems can be found in Appendix B.

To get a first impression of the iteration bounds for the several kernel functions we applied the algorithm to a selection of ten of these problems. After this first round of experiments we run all the mentioned 95 problems from the Netlib library with the kernel functions that gave the best performance on the aforementioned set of ten problems.

We used a straightforward implementation of our algorithm in MATLAB.[2]

[1] http://www-fp.mcs.anl.gov/otc/Guide/TestProblems/index.html
[2] http://www.matlab.com

Kernel function	Parameter values
ψ_2	$q \in \{1.5, 2, 2.5\}$
ψ_7	$q \in \{1.5, 2, 2.5\}$
ψ_8	$q \in \{1.5, 2\}$
ψ_9	$p \in \{0.5, 0.8\}$
ψ_{10}	$p \in \{0.5, 0.8\}, \ q \in \{1.5, 2\}$
$\psi_{p,\sigma}$	$(p, \sigma) \in \{(0, 1), (0.5, 1), (0.8, 1), (1, 1), (1, 1.5), (1, 2), (1, 2.5)\}$

Table 4.1: *Choice of parameters.*

We employed the self-dual embedding model [Roo05] to enable us to start the algorithm as indicated in Figure 1.1, namely with $x = s = \mathbf{1}$ and $\mu = 1$. Our experiments were performed on a standard PC with a Pentium 4 processor and with 1 GB internal memory. Since we wanted to compare iteration numbers for several kernel functions, and since these numbers depend on the parameters τ, θ and the accuracy parameter ϵ, we fixed these parameters in our experiments to $\tau = 1$, $\theta = 0.99$ and $\epsilon = 10^{-8}$. In this way the iteration numbers depend only on the kernel function and the problem instance.

The results of the first round of experiments are given in five tables (Table 4.2 to Table 4.6). For each of the ten problems we used bold font to highlight the best, i.e., the smallest, iteration number. This information is summarized in Table 4.7, which gives for each of the ten problems the smallest iteration number, and for which kernel function(s) this was achieved. From Table 4.7 we conclude that the smallest iteration numbers were realized by five kernel functions. For these five kernel functions we solved in the second round the mentioned 95 problems in the Netlib library. The results of the second round are listed in two tables (Table 4.8 to Table 4.10).

In the first round we encountered a problem with kernel function ψ_5, indicated by question marks in the corresponding column of Table 4.3. The reason is the occurrence of the expression $e^{\frac{1}{t}}$ in the definition of this kernel function. For values of t smaller than ≈ 0.0014 the value of this expression goes beyond the size of numbers that can be handled by MATLAB. Since such small values occur in the vector v during the execution of the algorithm for some of the test problems, the programme failed to run for nine of the ten problems in the first round. Only for

LP Problem	Number of iterations for ψ_1, ψ_2, and ψ_3				
	ψ_1	ψ_2 $(q = 1.5)$	ψ_2 $(q = 2)$	ψ_2 $(q = 2.5)$	ψ_3
ADLITTLE	**23**	**23**	24	24	39
AFIRO	**16**	**16**	17	17	28
DEGEN2	**24**	26	26	26	46
DEGEN3	**30**	31	33	37	71
Grow15	**37**	39	38	40	77
MAROS	74	76	82	85	90
SC105	**18**	19	19	19	36
SC205	**22**	**22**	23	25	36
SCTAP2	**26**	28	30	31	41
SHELL	**46**	49	50	54	81
Total of N.It	316	329	342	358	545

Table 4.2: *Iteration numbers for ψ_1, ψ_2, and ψ_3.*

SC205 it found the solution, but when comparing the number of iterations with the iteration count for the other kernel functions, the result is not very promising.

LP Problem	Number of iterations for ψ_4, ψ_5, ψ_6 and ψ_7					
	ψ_4	ψ_5	ψ_6	ψ_7 $(q = 1.5)$	ψ_7 $(q = 2)$	ψ_7 $(q = 2.5)$
ADLITTLE	25	?	24	**23**	25	26
AFIRO	17	?	18	**16**	17	18
DEGEN2	25	?	27	25	25	25
DEGEN3	34	?	37	**30**	34	34
Grow15	41	?	43	39	41	46
MAROS	87	?	92	69	87	87
SC105	21	?	19	19	21	24
SC205	24	45	23	23	24	25
SCTAP2	31	?	32	27	31	32
SHELL	52	?	55	49	52	55
Total of N.It	357	?	370	320	357	372

Table 4.3: *Iteration numbers for ψ_4, ψ_5, ψ_6 and ψ_7.*

From the first four tables we may draw a few conclusions. First, the numbers of iterations obtained by using ψ_8, which has a linear growth term, are almost among the worst. For ψ_9, it becomes clear that smaller values of the parameter p influence the iteration count negatively. Hence, $p = 1$ seems to be the best possible choice, which gives ψ_1, the kernel function of the logarithmic barrier function. Furthermore, the kernel functions ψ_3, ψ_4, ψ_5, ψ_6, ψ_8 ψ_9 and ψ_{10} never

LP Problem	Number of iterations for ψ_8 and ψ_9			
	ψ_8 $(q = 1.5)$	ψ_8 $(q = 2)$	ψ_9 $(p = 0.5)$	ψ_9 $(p = 0.8)$
ADLITTLE	67	82	30	25
AFIRO	63	71	25	19
DEGEN2	69	85	34	27
DEGEN3	101	193	37	32
Grow15	97	116	45	39
MAROS	107	126	74	71
SC105	70	81	28	22
SC205	72	84	30	23
SCTAP2	97	97	34	30
SHELL	101	96	55	51
Total of N.It	844	905	392	339

Table 4.4: *Iteration numbers for ψ_8 and ψ_9.*

LP Problem	Number of iterations for ψ_{10}			
	$(p = 0.5, q = 1.5)$	$(p = 0.5, q = 2)$	$(p = 0.8, q = 1.5)$	$(p = 0.8, q = 2)$
ADLITTLE	34	35	27	29
AFIRO	29	29	20	21
DEGEN2	29	42	29	31
DEGEN3	42	47	33	37
GROW15	49	55	43	46
MAROS	77	95	80	80
SC105	30	33	22	25
SC205	34	38	26	28
SCTAP2	38	46	31	34
SHELL	61	67	52	56
Total of N.It	423	487	363	387

Table 4.5: *Iteration numbers for ψ_{10}.*

give the smallest iteration number.

Special attention deserves the finite barrier kernel function $\psi_{p,\sigma}$. In Chapter 2 we mentioned that it differs from the other kernel functions in the sense that it has a finite value at the boundary of the feasible region. The results in Table 4.6 show that for $\psi_{1,1}$ the iteration numbers in all ten cases are the same (or almost the same) as for the kernel function ψ_1 of the classical logarithmic barrier function.

In the second round we used only the kernel functions that appear in the second column of Table 4.7, namely ψ_1, ψ_2 and ψ_7, both with $q = 1.5$, and $\psi_{1,1}$ and $\psi_{1,1.5}$. The iteration counts for the 95 problems are listed in Table 4.8, Table

LP Problem	Number of iterations for $\psi_{p.\sigma}$						
	$\psi_{0,1}$	$\psi_{0.5,1}$	$\psi_{0.8,1}$	$\psi_{1,1}$	$\psi_{1,1.5}$	$\psi_{1,2}$	$\psi_{1,2.5}$
ADLITTLE	60	29	26	24	24	25	25
AFIRO	43	24	20	**16**	17	17	18
DEGEN2	56	34	28	**24**	26	26	26
DEGEN3	66	39	36	32	31	31	31
GROW15	70	47	40	**37**	38	38	38
MAROS	94	77	77	**62**	64	68	65
SC105	51	28	22	**18**	19	20	21
SC205	54	31	24	**22**	**22**	23	24
SCTAP2	61	35	31	**26**	27	27	27
SHELL	78	57	53	50	52	51	49
Total of N.It	633	401	355	311	320	326	324

Table 4.6: *Iteration numbers for some finite barrier functions.*

Problem	Best result	Kernel functions
ADLITTLE	23	ψ_1, $\psi_2(q = 1.5)$, $\psi_7(q = 1.5)$
AFIRO	16	ψ_1, $\psi_2(q = 1.5)$, $\psi_7(q = 1.5)$, $\psi_{1,1}$
DEGEN2	24	ψ_1, $\psi_{1,1}$
DEGEN3	30	ψ_1, $\psi_7(q = 1.5)$
GROW15	37	ψ_1, $\psi_{1,1}$
MAROS	62	$\psi_{1,1}$
SC105	18	ψ_1, $\psi_{1,1}$
SC205	22	ψ_1, $\psi_2(q = 1.5)$, $\psi_{1,1}$, $\psi_{1,1.5}$
SCTAP2	26	ψ_1, $\psi_{1,1}$
SHELL	46	ψ_1

Table 4.7: *Smallest iteration numbers and corresponding kernel function(s)*

4.9 and Table 4.10. The results in these three tables justify the conclusion that the kernel functions $\psi_{p,\sigma}$ deserve further investigation. Their performance seems quite promising. These kernel functions are new and have not yet been optimized for practical use.[3]

[3] To get a method that is useful from a practical point of view, one has to embed the generic algorithm of Figure 1.1 in a predictor-corrector scheme as proposed by Mehrotra. A nice example of this is a recent paper of Zhu et al. [Zhu03]. They propose a Mehrotra-type predictor-corrector

LP Problem	Number of iterations				
	ψ_1	ψ_2 $(q = 1.5)$	ψ_7 $(q = 1.5)$	$\psi_{1,1}$	$\psi_{1,1.5}$
25FV47	**71**	75	75	**71**	72
80BAU3B	**100**	104	104	103	104
ADLITTLE	**23**	**23**	**23**	24	24
AFIRO	**16**	**16**	**16**	**16**	17
AGG	43	44	44	**42**	43
AGG2	**36**	37	37	39	38
AGG3	**39**	45	41	43	41
BANDM	39	39	40	**38**	40
BEACONFD	**23**	**23**	24	25	25
BLEND	**19**	20	**19**	**19**	20
BNL1	69	70	69	**68**	**68**
BNL2	76	76	79	76	**75**
BOEING1	**42**	44	45	44	44
BOEING2	**35**	**35**	36	36	36
BORE3D	39	39	39	**36**	38
BRANDY	40	41	42	39	**38**
CAPRI	**42**	44	44	**42**	**42**
CYCLE	86	93	90	**77**	86
CZPROB	78	84	86	77	**76**
D2Q06C	**107**	112	112	110	109
D6CUBE	**38**	39	40	39	40
DEGEN2	**24**	26	25	**24**	26
DEGEN3	**30**	31	**30**	32	31
DFL001	**81**	85	85	83	83
E226	**41**	43	44	42	43
ETAMACRO	66	66	67	**64**	**64**
FFFFF800	**64**	66	65	**64**	**64**
FINNIS	60	60	61	**56**	**56**
FIT1D	**32**	33	**32**	33	33
FIT1P	**33**	**33**	35	34	34
FIT2D	42	43	42	**41**	42
FIT2P	**40**	42	43	**40**	**40**
FORPLAN	**40**	43	43	48	46

Table 4.8: *Iteration numbers for the five best kernel functions* (I).

algorithm, in which the search direction determined by the kernel function ψ_7 plays a role. The algorithm starts with a predictor step based on the so-called primal-dual affine scaling direction. If the maximum step size in this direction is sufficiently large, then the algorithm performs after the predictor step a Mehrotra corrector step, based on the classical primal-dual Newton direction, followed by a backtracking line search technique to keep the iterates in a certain neighborhood of the central path. If the maximum feasible step size in the predictor step is not large enough,

LP Problem	Number of iterations				
	ψ_1	ψ_2 ($q = 1.5$)	ψ_7 ($q = 1.5$)	$\psi_{1,1}$	$\psi_{1,1.5}$
GANGES	43	43	**42**	44	44
GFRD-PNC	**32**	35	38	34	36
GREENBEA	**119**	125	130	123	125
GREENBEB	**121**	124	126	123	**121**
GROW15	**37**	39	39	**37**	38
GROW22	41	39	41	40	**38**
GROW7	**35**	36	37	**35**	**35**
ISRAEL	**36**	37	39	37	37
KB2	**30**	**30**	**30**	**30**	**30**
LOTFI	**29**	30	32	31	32
MAROS	74	76	69	**62**	64
MAROS-R7	**37**	**37**	38	**37**	39
MODSZK1	**49**	50	51	51	51
NESM	75	**74**	75	75	75
PEROLD	73	72	75	73	**71**
PILOT	102	102	104	100	**99**
PILOT.JA	**74**	76	76	76	76
PILOT.WE	137	133	132	**125**	131
PILOT4	71	76	76	**70**	**70**
PILOT87	145	149	150	147	**144**
PILOTNOV	56	58	59	**55**	56
RECIPE	**19**	21	21	21	21
SC105	**18**	19	19	**18**	19
SC205	**22**	**22**	23	**22**	**22**
SC50A	18	**17**	18	**17**	18
SC50B	17	17	17	**16**	17
SCAGR25	**32**	33	36	33	33
SCAGR7	**25**	26	26	26	26
SCFXM1	**42**	44	44	43	43
SCFXM2	**52**	**52**	54	**52**	53
SCFXM3	57	**51**	54	57	56

Table 4.9: *Iteration numbers for the five best kernel functions* (II).

then the algorithm uses the search direction determined by the kernel function ψ_7 to bring the iterate closer to the central path. From their paper it is clear that the iteration numbers presented in the last three tables for the Netlib set of problems can be reduced by about 50%.

LP Problem	Number of iterations				
	ψ_1	ψ_2 $(q = 1.5)$	ψ_7 $(q = 1.5)$	$\psi_{1,1}$	$\psi_{1,1.5}$
SCORPION	**33**	34	36	35	35
SCRS8	51	51	53	**50**	52
SCSD1	**32**	41	46	39	33
SCSD6	50	**45**	54	67	61
SCSD8	45	**36**	50	39	41
SCTAP1	**36**	42	42	**36**	**36**
SCTAP2	**26**	28	27	**26**	27
SCTAP3	**28**	31	29	**28**	29
SEBA	54	56	57	**53**	54
SHARE1B	48	50	50	**47**	48
SHARE2B	**22**	**22**	23	24	23
SHELL	**46**	49	49	50	52
SHIP04L	**30**	**30**	**30**	**30**	31
SHIP04S	**28**	29	31	30	30
SHIP08L	36	34	34	**32**	33
SHIP08S	**27**	30	30	28	28
SHIP12L	**48**	53	55	50	**48**
SHIP12S	**41**	44	44	42	43
SIERRA	**40**	44	45	42	44
STAIR	**33**	**33**	35	34	34
STANDATA	**29**	31	30	30	30
STANDGUB	**29**	31	30	30	30
STANDMPS	**35**	39	38	38	38
STOCFOR1	27	27	27	**25**	**25**
STOCFOR2	65	**64**	66	68	68
STOCFOR3	**120**	121	121	122	123
TRUSS	**61**	62	65	63	64
TUFF	**40**	41	42	41	41
VTP-BASE	**28**	29	**28**	29	**28**
WOOD1P	37	35	40	**33**	34
WOODW	**60**	74	78	62	63

Table 4.10: *Iteration numbers for the five best kernel functions (III).*

Chapter 5

Conclusions

5.1 Conclusions and Remarks

This thesis was inspired by recent work on so-called self-regular barrier functions for primal-dual interior-point methods for linear optimization, second order cone and semidefinite optimization [Pen02a; Pen02b]. Each such barrier function is determined by its (univariate) self-regular kernel function. We introduce a new class of kernel functions which differs from the class of self-regular kernel functions. The class is defined by some simple conditions on the kernel function which concern the growth and the barrier behavior of the kernel function. These properties enable us to derive many new and tight estimates that greatly simplify the analysis of IPMs based on these kernel functions. An important conclusion from the analysis is that inverse functions of suitable restrictions of the kernel function and its first derivative more or less determine the behavior of the corresponding IPMs .

In Chapter 2 we consider ten specific (classes of) kernel functions belonging to the new class, and using the new estimates present a complete complexity analysis for each of these functions. Some of these functions are self-regular and others are not. Three of the functions are special in the sense that there growth term is not quadratic. We also present the analysis of a kernel function with finite barrier term. Iterations bounds both for large- and small-update methods are derived. It is shown that small-update methods based on the new kernel functions all have the same complexity as the classical primal-dual IPM, namely $O(\sqrt{n}\,\log\frac{n}{\epsilon})$. For large-update methods the best obtained bound is $O(\sqrt{n}\ (\log n)\ \log \frac{n}{\epsilon})$, which is up till now the best known bound for such methods.

The results of Chapter 2 for LO can be easily extended to other conic optimization problem classes like second order cone and semidefinite optimization. We show this in Chapter 3, where we deal with semidefinite optimization and show that at some point the analysis boils down to exactly the same analysis as for the LO case.

Similar results hold for the case of SOCO, but this is not considered in this thesis.

In Chapter 4 some numerical results are presented. These results show that by using a kernel functions $\psi_{p,\sigma}$, with finite barrier term, the best iteration complexity was achieved in most of the test problems, especially for the kernel functions $\psi_{1,1}$ and $\psi_{1,1.5}$. Their practical performance seems quite promising for LO.

5.2 Directions for further research

Future research might focuss on one of the following questions.

- Does there exist a kernel function for which the complexity of large-update methods is the same as for small-update methods?[1]
- How do the methods for SDO perform in practice?
- Is it possible to design dual (or primal) IPMs for LO based on the new class of kernel functions? If the result is positive, how to extend these results to SOCO and SDO problems?
- Can we design primal-dual IPMs for SDO based on scaling techniques other than the Nesterov-Todd scaling?
- It is possible to extend the present work to more general nonlinear optimization problems?
- In [Sal03; Sal05b] the authors presented a new complexity analysis of Infeasible IPMs for linear optimization based on a specific self-regular kernel

[1] It might be mentioned that a large-update method with such complexity was obtained recently in [Ai04] for monotone linear complementarity problems by using search directions that are different from the usual Newton direction.

function (namely, ψ_7 with $q = \log(n)$). Can we design Infeasible IPMs based on different kernel functions, as presented in this thesis, for linear and nonlinear optimization?

- Recently, Peng et al. [Pen05] and Salahi et al. [Sal04b; Sal05a] studied intensively the so-called self-regular predictor-corrector IPMs for LO. Can we design predictor-corrector IPMs for LO and nonlinear optimization based on non-self-regular kernel functions, as presented in this thesis?

Appendix A

Technical Lemmas

A.1 Three technical lemmas

We need three simple technical results. For completeness' sake we include their (short) proofs. The first lemma is needed only in the proof of the second lemma, which is interesting in itself.

Lemma A.1.1 (Lemma 2.1 in [Pen02a]). *If* $\alpha \in [0, 1]$, *then*

$$(1+t)^{\alpha} \leq 1 + \alpha t, \quad \forall\, t \geq -1. \tag{A.1.1}$$

Proof. Consider the function $f(t) = (1+t)^{\alpha} - 1 - \alpha t$ for $t \geq -1$. One has $f'(t) = \alpha(1+t)^{\alpha-1} - \alpha$ and $f''(t) = \alpha(\alpha-1)(1+t)^{\alpha-2}$. Since $f''(t) \leq 0$, $f(t)$ is concave. Since $f'(0) = 0$, the function f is maximal at $t = 0$. Finally, since $f(0) = 0$, the lemma follows. □

Lemma A.1.2 (Proposition 2.2 in [Pen02a]). *Let* $t_0, t_1, \cdots, t_K$ *be a sequence of positive numbers such that*

$$t_{k+1} \leq t_k - \kappa t_k^{1-\gamma}, \qquad k = 0, 1, \cdots, K-1, \tag{A.1.2}$$

where $\kappa > 0$ *and* $0 < \gamma \leq 1$*. Then* $K \leq \left\lfloor \frac{t_0^\gamma}{\kappa\gamma} \right\rfloor$.

Proof. Using (A.1.2), we may write

$$0 < t_{k+1}^\gamma \leq \left(t_k - \kappa t_k^{1-\gamma}\right)^\gamma = t_k^\gamma \left(1 - \kappa t_k^{-\gamma}\right)^\gamma \leq t_k^\gamma \left(1 - \kappa\gamma t_k^{-\gamma}\right) = t_k^\gamma - \kappa\gamma,$$

where the second inequality follows from (A.1.1). Hence, for each k, $t_k^\gamma \leq t_0^\gamma - k\gamma\kappa$. Taking $k = K$ we obtain $0 < t_0^\gamma - K\gamma\kappa$, which implies the lemma. □

Lemma A.1.3 (Lemma 3.12 in [Pen02a]). *Let* $h(t)$ *be a twice differentiable convex function with* $h(0) = 0$, $h'(0) < 0$ *and let* $h(t)$ *attain its (global) minimum at* $t^* > 0$*. If* $h''(t)$ *is increasing for* $t \in [0, t^*]$ *then*

$$h(t) \leq \frac{th'(0)}{2}, \quad 0 \leq t \leq t^*.$$

Proof. Using the hypothesis of the lemma we may write

$$\begin{aligned} h(t) &= \int_0^t h'(\xi)d\xi = h'(0)t + \int_0^t \int_0^\xi h''(\zeta)d\zeta d\xi \leq h'(0)t + \int_0^t \xi h''(\xi)d\xi \\ &= h'(0)t + \int_0^t \xi dh'(\xi) = h'(0)t + \left(\xi h'(\xi)\right)\big|_0^t - \int_0^t h'(\xi)d\xi \\ &\leq h'(0)t - \int_0^t dh'(\xi) = h'(0)t - h(t). \end{aligned}$$

This implies the lemma. □

Appendix B

The Netlib-Standard Problems

Name	Rows	Columns	Nonzeros	Optimal value
25FV47	822	1571	11127	$4.5018458883E+03$
80BAU3B	2263	9799	29063	$9.8723216072E+05$
ADLITTLE	57	97	465	$2.2549496316E+05$
AFIRO	28	32	88	$-4.6475314286E+02$
AGG	489	163	2541	$-3.5991767287E+07$
AGG2	517	302	4515	$-2.0239252356E+07$
AGG3	517	302	4531	$1.0312115935E+07$
BANDM	306	472	2659	$-1.5862801845E+02$
BEACONFD	174	262	3476	$3.3592485807E+04$
BLEND	75	83	521	$-3.0812149846E+01$
BNL1	644	1175	6129	$1.9776292856E+03$
BNL2	2325	3489	16124	$1.8112365404E+03$
BORE3D	234	315	1525	$1.3730803942E+03$
BRANDY	221	249	2150	$1.5185098965E+03$
CAPRI	272	353	1786	$2.6900129138E+03$
CYCLE	1904	2857	21322	$-5.2263930249E+00$
CZPROB	930	3523	14173	$2.1851966989E+06$
D2Q06C	2172	5167	35674	$1.2278423615E+05$
D6CUBE	416	6184	43888	$3.1549166667E+02$
DEGEN2	445	534	4449	$-1.4351780000E+03$
DEGEN3	1504	1818	26230	$-9.8729400000E+02$
DFL001	6072	12230	41873	$1.12664E+07?$
E226	224	282	2767	$-1.8751929066E+01$
ETAMACRO	401	688	2489	$-7.5571521774E+02$
FFFFF800	525	854	6235	$5.5567961165E+05$
FINNIS	498	614	2714	$1.7279096547E+05$
FIT1D	25	1026	14430	$-9.1463780924E+03$
FIT1P	628	1677	10894	$9.1463780924E+03$
FIT2D	26	10500	138018	$-6.8464293294E+04$
FIT2P	3001	13525	60784	$6.8464293232E+04$
FORPLAN	162	421	4916	$-6.6421873953E+02$

Table B.1: *The Netlib-Standard Problems (I).*

Name	Rows	Columns	Nonzeros	Optimal value
GANGES	1310	1681	7021	$-1.0958636356E+05$
GFRD-PNC	617	1092	3467	$6.9022359995E+06$
GREENBEA	2393	5405	31499	$-7.2462405908E+07$
GREENBEB	2393	5405	31499	$-4.3021476065E+06$
GROW15	301	645	5665	$-1.0687094129E+08$
GROW22	441	946	8318	$-1.6083433648E+08$
GROW7	141	301	2633	$-4.7787811815E+07$
ISRAEL	175	142	2358	$-8.9664482186E+05$
KB2	44	41	291	$-1.7499001299E+03$
LOTFI	154	308	1086	$-2.5264706062E+01$
MAROS	847	1443	10006	$-5.8063743701E+04$
MAROS-R7	3137	9408	151120	$1.4971851665E+06$
MODSZK1	688	1620	4158	$3.2061972906E+02$
NESM	663	2923	13988	$1.4076073035E+07$
PEROLD	626	1376	6026	$-9.3807580773E+03$
PILOT	1442	3652	43220	$-5.5740430007E+02$
PILOT.JA	941	1988	14706	$-6.1131344111E+03$
PILOT.WE	723	2789	9218	$-2.7201027439E+06$
PILOT4	411	1000	5145	$-2.5811392641E+03$
PILOT87	2031	4883	7304	$3.0171072827E+02$
PILOTNOV	976	2172	13129	$-4.4972761882E+03$
RECIPE	92	180	758	$-2.6661600000E+02$
SC105	106	103	281	$-5.2202061212E+01$
SC205	206	203	552	$-5.2202061212E+01$
SC50A	51	48	131	$-6.4575077059E+01$
SC50B	51	48	119	$-7.0000000000E+01$
SCAGR25	452	500	2029	$-1.4753433061E+07$
SCAGR7	130	140	553	$-2.3313892548E+08$
SCFXM1	331	457	2612	$1.8416759028E+04$
SCFXM2	661	914	5229	$3.6660261565E+04$
SCFXM3	991	1371	7846	$5.4901254550E+04$

Table B.2: *The Netlib-Standard Problems* (II).

Name	Rows	Columns	Nonzeros	Optimal value
SCORPION	389	358	1708	$1.871248227E+03$
SCRS8	491	1169	4029	$9.0429998619E+02$
SCSD1	78	760	3148	$8.6666666743E+00$
SCSD6	148	1350	5666	$5.0500000078E+01$
SCSD8	398	2750	11334	$9.0499999993E+02$
SCTAP1	301	480	2052	$1.4122500000E+03$
SCTAP2	1091	1880	8124	$1.7248071429E+03$
SCTAP3	118	225	1182	$1.4240000000E+03$
SEBA	516	1028	4874	$1.5711600000E+04$
SHARE1B	118	225	1182	$-7.6589318579E+04$
SHARE2B	97	79	730	$-4.1573224074E+02$
SHELL	537	1775	4900	$1.2088253460E+09$
SHIP04L	403	2118	8450	$1.7933245380E+06$
SHIP04S	403	1458	5810	$1.7987147004E+06$
SHIP08L	779	4283	17085	$1.9090552114E+06$
SHIP08S	779	2387	9501	$1.9200982105E+06$
SHIP12L	1152	5427	21597	$1.4701879193E+06$
SHIP12S	1152	1763	10941	$1.4892361344E+04$
SIERRA	1228	2036	9252	$1.5394362184E+07$
STAIR	357	467	3857	$-2.5126695119E+02$
STANDATA	360	1075	3038	$1.2576995000E+03$
STANDGUB	362	1184	3147	$1.2576995000E+03$
STANDMPS	468	1075	3686	$1.4060175000E+03$
STOCFOR1	118	111	474	$-4.1131976219E+04$
STOCFOR2	2158	2031	9492	$-3.9024408538E+04$
STOCFOR3	16676	15695	74004	$-3.9976661576E+04$
TRUSS	1001	8806	36642	$4.5881584719E+05$
TUFF	334	587	4523	$2.9214776509E+01$
VTP-BASE	199	203	914	$1.2983146246E+05$
WOOD1P	245	2594	70216	$1.4429024116E+00$
WOODW	1099	8405	37478	$1.3044763331E+00$

Table B.3: *The Netlib-Standard Problems* (III).

Bibliography

[Ai04] W. Ai and S. Zhang. An $O(\sqrt{n}\log\frac{(x^0)^T s^0}{\epsilon}$ iteration primal-dual path-following method, based on wide neighborhoods and large updates, for monotone linear complementarity problems. Technical Report 2004/01, Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, 2004.

[Ali91] F. Alizadeh. *Combinatorial Optimization with Interior Point Methods and Semi-Definite Matrices*. Ph.D. thesis, University of Minnesota, Minneapolis, Minnesota, USA, 1991.

[Ali95] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM J Optim*, vol. 5(1):pp. 13–51, 1995.

[Ali96] F. Alizadeh, J.-P. Haeberley, and M. Overton. Primal-dual interior point methods for semidefinite programming: convergence rates, stability and numirical results. Technical report 271, NYU compter Science Dept, New York University, 1996.

[And96] E. Andersen, J. Gondzio, C. Mészáros, and X. Xu. Implementation of interior point methods for large scale linear programming. In T. Terlaky, editor, *Interior Point Methods of Mathematical Programming*, pp. 189–252. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[Ans92] K. Anstreicher and R. Bosch. Long steps in a $O(n^3L)$ algorithm for linear programming. *Mathematical Programming*, vol. 54:pp. 251–265, 1992.

[Bai02a] Y. Bai, M. E. Ghami, and C. Roos. A primal-dual interior-point algorithm for linesr optimization based on a new proximity function. *Optimization Methods & Software*, vol. 17(6):pp. 985–1008, 2002.

[Bai02b] Y. Bai and C. Roos. A primal-dual interior-point method based on a new kernel function with linear growth rate, 2002. To appear in *Proceedings Perth Industrial Optimization meeting*.

[Bai03a] Y. Bai, M. E. Ghami, and C. Roos. A new efficient large-update primal-dual interior-point method based on a finite barrier. *SIAM Journal on Optimization*, vol. 13(3):pp. 766–782, 2003.

[Bai03b] Y. Bai and C. Roos. A polynomial-time algorithm for linear optimization based on a new simple kernel function, 2003. Appeared in Proceedings of the Hangzhou meeting.

[Bai04a] Y. Bai, M. E. Ghami, and C. Roos. A comparative study of kernel functions for primal-dual interior-point algorithms in linear optimization. *SIAM Journal on Optimization*, vol. 15(1):pp. 101–128, 2004.

[Bai04b] Y. Bai, G. Wang, and C. Roos. A new primal-dual interior-point method algorithm for second-order cone optimization, 2004. Manuscript.

[Boy96] S. Boyd and L. Vandenberghe. Semidefinite programming. *SIAM Review*, vol. 38(1):pp. 49–96, 1996.

[BT01] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization. Analysis, Algorithms, Engineering Applications*. MPS-SIAM Series on Optimization. SIAM, Philadelphia, USA, 2001.

[Dan63] G. Dantzig. *Linear Programming and Extensions*. Princeton University Press, Princeton, New Jersey, 1963.

[Dez04a] A. Deza, E. Nematollahi, R. Peyghami, and T. Terlaky. The central path visits all the vertices of the Klee-Minty cube. Technical Report 2004/11, Advanced Optimization Laboratory, Mc Master University, Hamilton, Ontario, Canada, 2004.

[Dez04b] A. Deza, E. Nematollahi, and T. Terlaky. How good are interior point methods? Klee-Minty cubes tighten-complexity bounds. Technical Report 2004/20, Advanced Optimization Laboratory, Mc Master University, Hamilton, Ontario, Canada, 2004.

[Dik67] I. Dikin. Iterative solution of problems of linear and quadratic programming. *Doklady Akademii Nauk SSSR*, vol. 174:pp. 747–748, 1967. Translated in : *Soviet Mathematics Doklady*, 8:674–675, 1967.

[Dik74] I. Dikin. On the convergence of an iterative process. *Upravlyaemye Sistemi*, vol. 12:pp. 54–60, 1974. (In Russian).

[Dik88] I. Dikin. Letter to the editor. *Mathematical Programming*, vol. 41:pp. 393–394, 1988.

[Fri55] R. Frisch. The logarithmic potential method for solving linear programming problems. Memorandum, University Institute of Economics, Oslo, 1955.

[Fri56] K. Frisch. La resolution des problemes de programme lineaire par la methode du potential logarithmique. *Cahiers du Seminaire D'Econometrie*, vol. 4:pp. 7–20, 1956.

[Gha04a] M. E. Ghami, Y. Bai, and C. Roos. A polynomial-time interior-point algorithm based on generalized logarithmic barrier function. *Manuscript*, 2004.

[Gha04b] M. E. Ghami, J. Melissen, and C. Roos. Primal-dual interior-point methods for linear optimization based on a new class of kernel functions, 2004. Submitted To *Optimization Methods & Software*.

[Gha05a] M. E. Ghami and C. Roos. Primal-dual interior point methods based on a new kernel function, 2005. Accepted in special issue of the International Journal *RAIRO.*

[Gha05b] M. E. Ghami and C. Roos. Primal-dual IPMs for semidefinite optimization based on kernel functions, 2005. Manuscript.

[Gol89] D. Goldfarb and M. Todd. Linear Programming. In G. Nemhauser, A. Rinnooy Kan, and M. Todd, editors, *Optimization*, vol. 1 of *Handbooks in Operations Research and Management Science*, pp. 141–170. North Holland, Amsterdam, The Netherlands, 1989.

[Gon91] C. Gonzaga. Large step path-following methods for linear programming, Part I : Barrier function method. *SIAM Journal on Optimization*, vol. 1:pp. 268–279, 1991.

[Gon92] C. Gonzaga. Path following methods for linear programming. *SIAM Review*, vol. 34(2):pp. 167–227, 1992.

[Her92] D. den Hertog, C. Roos, and J.-P. Vial. A complexity reduction for the long–step path–following algorithm for linear programming. *SIAM Journal on Optimization*, vol. 2:pp. 71–87, 1992.

[Her94] D. Hertog. *Interior Point Approach to Linear, Quadratic and Convex Programming*, vol. 277 of *Mathematics and its Applications.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[Hor85] R. Horn and C. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, UK, 1985.

[Hua67] P. Huard. Resolution of mathematical programming with nonlinear constraints by the method of centers. In J. Abadie, editor, *Nonlinear Programming*, pp. 207–219. North Holland, Amsterdam, The Netherlands, 1967.

[Jan94a] B. Jansen, C. Roos, and T. Terlaky. The theory of linear programming : Skew symmetric self-dual problems and the central path. *Optimization*, vol. 29:pp. 225–233, 1994.

[Jan94b] B. Jansen, C. Roos, T. Terlaky, and J.-P. Vial. Primal-dual algorithms for linear programming based on the logarithmic barrier method. *Journal of Optimization Theory and Applications*, vol. 83:pp. 1–26, 1994.

[Kar84] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, vol. 4:pp. 373–395, 1984.

[Kha79] L. G. Khachiyan. A polynomial algorithm in linear programming. *Doklady Akademiia Nauk SSSR*, vol. 244:pp. 1093–1096, 1979. Translated into English in *Soviet Mathematics Doklady* 20, 191–194.

[Kle72] V. Klee and G. Minty. How good is the simplex algorithm? *Acadimic Press in Inequalities-III, O Shisha, ed New York*, pp. 159–175, 1972.

[Kle97] E. de Klerk. *Interior Point Methods for Semidefinite Programming.* Ph.D. thesis, TU Delft, The Netherlands, 1997.

[Kle02] E. de Klerk. *Aspects of Semidefintie Programming*, vol. 65 of *Applied Optimization.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.

[Koj89] M. Kojima, S. Mizuno, and A. Yoshise. A primal-dual interior point algorithm for linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming: Interior Point and Related Methods*, pp. 29–47. Springer Verlag, New York, 1989.

[Koj93a] M. Kojima, Y. Kurita, and S. Mizuno. Large-step interior point algorithms for linear complementarity problems. *SIAM Journal on Optimization*, vol. 3(2):pp. 398–412, 1993.

[Koj93b] M. Kojima, N. Megiddo, and S. Mizuno. Theoretical convergence of large-step-primal-dual interior point algorithms for linear programming. *Mathematical Programming*, vol. 59:pp. 1–22, 1993.

[Koj94] M. Kojima, S. Shindoh, and S. Hara. Interior point methods for the monotone linear complementarity problem in symmetric matrices. Research reports on information sciences, series b: Operations research, Dept. of Information Sciences, Tokyo Institute of Technology, Tokyo, Japan, 1994.

[Meg89] N. Megiddo. Pathways to the optimal set in linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming: Interior Point and Related Methods*, pp. 131–158. Springer Verlag, New York, 1989. Identical version in : *Proceedings of the 6th Mathematical Programming Symposium of Japan, Nagoya, Japan*, pages 1–35, 1986.

[Mon89] R. Monteiro and I. Adler. Interior-path following primal-dual algorithms : Part I : Linear programming. *Mathematical Programming*, vol. 44:pp. 27–41, 1989.

[Mon97] R. Monteiro. Primal-dual path-following algorithms for semidefinite programming. *SIAM Journal on Optimization*, vol. 7:pp. 663–678, 1997.

[Nes93] Y. Nesterov and A. Nemirovskii. *Interior Point Polynomial Methods in Convex Programming : Theory and Algorithms.* SIAM Publications. SIAM, Philadelphia, USA, 1993.

[Nes97] Y. Nesterov and M. Todd. Self-scaled barriers and interior-point methods for convex programming. *Mathematics of Operations Research*, vol. 22(1):pp. 1–42, 1997.

[Pen00a] J. Peng, C. Roos, and T. Terlaky. A new class of polynomial primal-dual methods for linear and semidefinite optimization. *European Journal of Operations Research*, 2000. Appeared.

[Pen00b] J. Peng, C. Roos, and T. Terlaky. Primal-dual interior-point methods for second-order cone optimization based on self-regular proximities. *SIAM Journal on Optimization*, 2000. See ORpaper69.

[Pen01] J. Peng, C. Roos, and T. Terlaky. A new and efficient large-update interior-point method for linear optimization. *Journal of Computational Technologies*, vol. 6(4):pp. 61–80, 2001.

[Pen02a] J. Peng, C. Roos, and T. Terlaky. Self-regular functions and new search directions for linear and semidefinite optimization. *Mathematical Programming*, vol. 93:pp. 129–171, 2002.

[Pen02b] J. Peng, C. Roos, and T. Terlaky. *Self-Regularity: A New Paradigm for Primal-Dual Interior-Point Algorithms.* Princeton University Press, 2002.

[Pen05] J. Peng, T. Terlaky, and Y. B. Zhao. A self-regularity based predictor-corrector algorithm for linear optimization. *SIAM Journal on Optimization*, vol. 15(4):pp. 1105–1127, 2005.

[Ren01] J. Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization*, vol. 1 of *MPS/SIAM Series on Optimization.* SIAM, Philadelphia, USA, 2001. ISBN 0-89871-502-4.

[Roo89] C. Roos and J.-P. Vial. Long steps with the logarithmic penalty barrier function in linear programming. In J. Gabszevwicz, J. Richard, and L. Wolsey, editors, *Economic Decision–Making : Games, Economics and Optimization, dedicated to J.H. Dreze*, pp. 433–441. Elsevier Science Publisher B.V., Amsterdam, The Netherlands, 1989.

[Roo02] C. Roos. Interior-point methods for linear optimization. In P. Pardalos and M. Resende, editors, *Handbook of Applied Optimization*, pp. 21–39. Oxford University Press, New York, USA, 2002.

[Roo05] C. Roos, T. Terlaky, and J.-P. Vial. *Theory and Algorithms for Linear Optimization. An Interior-Point Approach.* Springer Science, 2005.

[Rud78] W. Rudin. *Principles of Mathematical Analysis.* Mac-Graw Hill Book Company, New York, 1978.

[Sal03] M. Salahi, T. Terlaky, and G. Zhang. copmlexity of self-regular proximity based infeasible IPMs. Technical Report 2003/3, Advanced Optimization Laboratory, Mc Master University, Hamilton, Ontario, Canada, 2003.

[Sal04a] M. Salahi, R. Sotirov, and T. Terlaky. On self-regular IPMs. *TOP: Sociedat de Estadística e Investigación Operativa*, vol. 12(2):pp. 209 – 300, 2004.

[Sal04b] M. Salahi and T. Terlaky. Adaptive large neighborhood self-regular predictor-corrector IPMs for LO. Technical Report 2004/7, Advanced Optimization Laboratory, Mc Master University, Hamilton, Ontario, Canada, 2004.

[Sal05a] M. Salahi, J. Peng, and T. Terlaky. On Mehrotra-type predictor-corrector algorithms. Technical Report 2005/4, Advanced Optimization Laboratory, Mc Master University, Hamilton, Ontario, Canada, 2005.

[Sal05b] M. Salahi, M. R. Peyghami, and T. Terlaky. New copmlexity analysis of IIPMs for linear optimization based on specific self regular function. Technical Report 2005/10, Advanced Optimization Laboratory, Mc Master University, Hamilton, Ontario, Canada, 2005.

[Sho87] N. Shor. Quadratic optimization problems. *Soviet Journal of Computer and System Sciences*, vol. 25:pp. 1–11, 1987.

[Son86] G. Sonnevend. An "analytic center" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. In A. Prékopa, J. Szelezsán, and B. Strazicky, editors, *System Modelling and Optimization : Proceedings of the 12th IFIP-Conference held in Budapest, Hungary, September 1985*, vol. 84 of *Lecture Notes in Control and Information Sciences*, pp. 866–876. Springer Verlag, Berlin, West–Germany, 1986.

[Stu99] J. Sturm and S. Zhang. Symmetric primal-dual path following algorithms for semidefinite programming. *Applied Numirical Mathematics*, vol. 29:pp. 301–315, 1999.

[Tod89] M. Todd. Recent developments and new directions in linear programming. In M. Iri and K. Tanabe, editors, *Mathematical Programming : Recent Developments and Applications*, pp. 109–157. Kluwer Academic Press, Dordrecht, The Netherlands, 1989.

[Tod96] M. Todd and Y. Ye. A lower bound on the number of iterations of long-step and polynomial interior-point linear programming algorithms. *Annals of Operations Research*, vol. 62:pp. 233–252, 1996.

[Wan04] G. Wang, Y. Bai, and C. Roos. Primal-dual interior-point algorithms for semidefinite optimization based on a simple kernel function, 2004. To appear in *International Journal of Mathematical Algorithms*.

[Ye97] Y. Ye. *Interior Point Algorithms, Theory and Analysis.* John Wiley & Sons, Chichester, UK, 1997.

[Zhu03] X. Zhu, J. Peng, T. Terlaky, and G. Zhang. On implementing self-regular proximity based feasible IPMs. Technical Report 2, Advanced Optimization Lab. Department of Computing and Software, Mc Master University, 2003. Http://www.cas.mcmaster.ca/ oplab/publication.

Summary

New Primal-dual Interior-point Methods Based on Kernel Functions
by Mohamed El Ghami

Two important classes of polynomial-time interior-point method (IPMs) are *small-* and *large-update* methods, respectively. The theoretical complexity bound for large-update methods is a factor $\sqrt{n}$ worse than the bound for small-update methods, where n denotes the number of (linear) inequalities in the problem. In practice the situation is opposite: implementations of large-update methods are much more efficient than those of small-update methods. This so-called *irony of IPMs* motivated the present work.

Recently J. Peng C. Roos and T. Terlaky were able to design new IPMs with large-updates whose complexity is only a factor $\log n$ worse than for small-update methods. This means that the factor $\sqrt{n}$ was reduced to $\log n$, thus significantly reducing the gap between the theoretical behavior of large- and small-update methods. They made use of so-called self-regular barrier (or proximity) functions. Each such barrier function is determined by its (univariate) self-regular kernel function.

In these thesis we introduce a new class of kernel functions which differs from the class of self-regular kernel functions. The class is defined by some simple conditions on the kernel function which concern the growth and the barrier behavior of the kernel function. These properties enable us to derive many new and tight estimates that greatly simplify the analysis of IPMs based on these kernel functions.

In Chapter 2 we consider ten specific (classes of) kernel functions belonging to the new class, and using the new estimates present a complete complexity analysis for each of these functions. Some of these functions are self-regular and others are not. Iterations bounds both for large- and small-update methods are derived. It is shown that small-update methods based on the new kernel functions all have the same complexity as the classical primal-dual IPM, namely $O(\sqrt{n}\,\log\frac{n}{\epsilon})$. For large-update methods the best obtained bound is $O(\sqrt{n}\,(\log n)\,\log\frac{n}{\epsilon})$, which is up till now the best known bound for such methods.

The results of Chapter 2 for LO are extended to semidefinite optimization in Chapter 3, where we it is shown that at some point the analysis boils down to exactly the same analysis as for the LO case.

In Chapter 4 some numerical results are presented. These results show that one of the new kernel functions, with finite barrier term and with the best possible theoretical complexity, performs surprisingly well in our experiments.

Mohamed El Ghami

Samenvatting

Nieuwe Primaal-duale Methoden gebaseerd op Kernfuncties
van Mohamed El Ghami

Twee belangrijke klassen van polynomiale inwendige punt methoden (IPMn) zijn *small-* en *large-update* methoden, respectievelijk. De theoretische complexiteitsgrens voor large-update methoden is een factor $\sqrt{n}$ slechter dan de grens voor small-update methoden, waar n staat voor het aantal (lineaire) ongelijkheden in het probleem. In de praktijk is de situatie juist omgekeerd: implementaties van large-update methoden zijn veel efficiënter dan die van small-update methoden. Deze zogenaamde *irony of IPMs* motiveerde het voorliggende onderzoek.

Recentelijk waren J. Peng C. Roos en T. Terlaky in staat om nieuwe IPMn met large-updates te ontwerpen waarvan de theoretische complexiteit slechts een factor $\log n$ slechter is dan die voor small-update methoden. Dit betekent dat de factor $\sqrt{n}$ werd teruggebracht tot $\log n$; aldus werd het verschil in theoretisch gedrag van large- en small-update methoden aanzienlijk verkleind. Zij maakten gebruik van zogenaamde zelf-reguliere barrière (of afstands-)functies. Iedere zodanige barrière functie wordt bepaald door zijn zelf-reguliere kernfunctie.

In dit proefschrift introduceren wij een nieuwe klasse van kernfuncties die verschilt van de klasse van zelf-reguliere kernfuncties. Deze klasse wordt gedefinieerd door middel van enkele eenvoudige voorwaarden ten aanzien van het groei-gedrag en en het barrière-gedrag van de kernfunctie. Deze voorwaarden stellen ons in staat om een groot aantal nieuwe en scherpe schatttingen af te leiden die de analyse van IPMn gebaseerd op kernfuncties aanzienlijk vereenvoudigen.

In Hoofdstuk 2 beschouwen wij tien specifieke (klassen van) kernfuncties die behoren tot de nieuwe klasse. Met behulp van de nieuwe afschattingen maken wij een volledige complexiteitsanalyse voor ieder van deze tien kernfuncties. Sommige van de tien kernfuncties zijn zelf-regulier, en andere niet. Wij leiden iteratiegrenzen af voor zowel large-als small-update methoden. Het blijkt dat small-update methoden gebaseerd op de tien kernfuncties allemaal dezelfde complexiteit hebben als de klassieke primaal-duale IPM, namelijk $O(\sqrt{n}\log\frac{n}{\epsilon})$. Voor large-update methoden is de best verkregen iteratiegrens $O(\sqrt{n}\,(\log n)\log\frac{n}{\epsilon})$, wat tot op heden de best bekende grens is voor zulke methoden.

De resultaten van Hoofdstuk 2 voor LO worden gegeneraliseerd tot semidefiniete optimalisering in Hoofdstuk 3, alwaar blijkt dat vanaf zeker moment in de analyse deze samenvalt met die voor LO.

In Hoofdstuk 4 presenteren wij de resultaten van numerieke experimenten. Deze resultaten maken duidelijk dat één van de nieuwe kernfuncties, met een begrensde barrière term en met de best mogelijke theoretische complexiteit, verrassend goed presteert in deze experimenten.

Mohamed El Ghami

Curriculum Vitae

Mohamed EL Ghami was born in Tamsamane, Nador, Morocco on May 1, 1969.

He finished secondary school in 1989 at the technical school Maghreb Arab Oujda, Morocco. He studied two years Mechanical engineering at INSET Rabat, Morocco. After that he started his studies at University Mohammed I Oujda, Morocco and received the Bachelor degree “Licence” in Mathematics, with concentration in Operation Research and Statistics in 1995. He finished his Master degree “CEA” in Applied Mathematics, with specialization in Nonlinear Analysis and Optimization in December 1996, at Mohammed V University, Rabat, Morocco. From January 1997 until December 1999 he was doing research at this university and also teaching Mathematics at private secondary schools.

During the academic year 2000-2001, he visited the Department of Operation Research, Delft University of Technology, to do research in the group of Algorithms, chaired by Prof. dr. ir. C. Roos. From May 2001, he started his Ph.D. study at the Faculty of Electrical Engineering, Mathematics and Computer Science. His research was part of the project *New Barrier Functions for Cone-Optimization*, financed by a Dutch NWO-grant (No. 613.000.110). This research has led to the present Ph.D thesis. He has published several papers in international journals.

Index