A Comparative Study of New Barrier Functions for Primal-Dual Interior-Point Algorithms in Linear Optimization*

Y.Q. Bai M. El ghami C. Roos

February 14, 2003
September 1, 2003 (1st revision)

Faculty of Information Technology and Systems

Delft University of Technology

P.O. Box 5031, 2600 GA Delft, The Netherlands

e-mail: [Y.Bai,M.Elghami,C.Roos]@ewi.tudelft.nl

Abstract

Recently, so-called self-regular barrier functions for primal-dual interior-point methods (IPMs) for linear optimization were introduced. Each such barrier function is determined by its (univariate) self-regular kernel function. We introduce a new class of kernel functions. The class is defined by some simple conditions on the kernel function and its derivatives. These properties enable us to derive many new and tight estimates that greatly simplify the analysis of IPMs based on these kernel functions. Both in the algorithm and in its analysis we use a single neighborhood of the central path; the neighborhood naturally depends on the kernel function. An important conclusion 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. Based on the new estimates we present a simple and unified computational scheme for the complexity analysis of kernel function in the new class. We apply this scheme to seven specific kernel functions. Some of these functions are self-regular and others are not. One of the functions differs from the others, and from all self-regular functions, in the sense that its growth term is linear. Iteration 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 $O(\sqrt{n}\,\log\frac{n}{\varepsilon})$. For large-update methods the best obtained bound is $O(\sqrt{n}\,(\log n)\,\log\frac{n}{\varepsilon})$, which is up till now the best known bound for such methods.

Keywords: Linear optimization, interior-point method, primal-dual method, large-update method, polynomial complexity.

AMS Subject Classification: **90C05, 90C51**

*The first author is on leave from the Department of Mathematics, Shanghai University, Shanghai 200436, China. She and the second author kindly acknowledge the support of the Dutch Organization for Scientific Research (NWO grant 613.000.110).

1 Introduction

After the path-breaking paper of Karmarkar [5], Linear Optimization (LO) became an active area of research. The resulting Interior-point Methods (IPMs) are now among the most effective methods for solving LO problems. For a survey we refer to recent books on the subject [14, 16, 17]. In this paper we deal with so-called primal-dual IPMs. It is generally agreed that these IPMs are most efficient from a computational point of view (see, e.g. Andersen et al. [1]).

Up till now primal-dual IPMs all use the Newton direction as the search direction; this direction is closely related to the well-known primal-dual logarithmic barrier function. For a discussion of this relation we refer to, e.g., [2, 3, 11]. There is a gap between the practical behavior of these algorithms and the theoretical performance results, especially for so-called large-update methods. If n denotes the number of inequalities in the problem, then the theoretical iteration bound is $O(n \log(n/\varepsilon))$, where ε represents the desired accuracy of the solution. In practice, large-update methods are much more efficient than the so-called small-update methods for which the theoretical iteration bound is only $O(\sqrt{n}\, \log(n/\varepsilon))$. So the current theoretical bounds differ by a factor $\sqrt{n}$, in favor of the small-update methods. This significant gap between theory and practice has been referred to as the irony of IPMs [13, page 51].

Recently a new class of IPMs was introduced and the aforementioned gap could be narrowed up to a factor $\log n$. These methods do not use the classical Newton direction. Instead they use a direction that can be characterized as a steepest descent direction (in a scaled space) for a so-called *self-regular* barrier function [11, 12]. Any such barrier function is determined by a simple univariate self-regular function, called its *kernel function*. The prototype self-regular kernel function in [11, 12] is given by

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

where $p \geq 1$ and $q > 1$. The parameter p is called the *growth degree*, and q the *barrier degree* of the kernel function. It may be noted that the iteration bounds obtained in [11, 12] depend monotonically on p: the smaller p the better is the iteration bound. Hence it is stated in [12, page 63] that in practical implementations small values of p should be used.

In this paper we introduce a new class of kernel functions which are not necessarily self-regular; on the other hand all self-regular kernel functions $\Upsilon_{p,q}(t)$ with growth degree $p \leq 1$ belong to the new class. We develop some new analytic tools for the analysis of IPMs based on kernel functions from the new class. As a result the analysis in this paper is much simpler than in [11, 12], whereas the iteration bounds are at least as good. In addition, our analysis also applies to (self-regular) functions with growth degree $p \leq 1$. Both in the algorithm and in its analysis we use a single neighborhood of the central path; the neighborhood naturally depends on the kernel function.

The paper is organized as follows. In Section 2.1 we first briefly recall the notion of central path, which is a basic concept underlying all primal-dual IPMs. In Section 2.2 we describe the idea underlying the approach of the paper. A crucial observation is that any function that is strictly convex on the positive orthant and that attains its minimal value in the all-one vector e, determines in a natural way a primal-dual interior-point method, which is formally described in Section 2.3.

In Section 2.4 we define the notion of a (univariate) kernel function, and in Section 2.5 we introduce the class of kernel functions considered in this paper. The first property shared by kernel functions in this class is *exponential convexity*. The other properties control the growth and barrier behavior of the kernel function. Section 2.6 introduces seven specific kernel functions belonging to the new class.

In Section 3 we study the growth behavior of the new barrier functions. Section 4 is devoted to the analysis of the amount of decrease of the barrier function during an inner iteration. This analysis yields a default step size, and, in Section 5, the iteration complexity of the algorithm in terms of the kernel function and its first and second derivative. It also turns out that this complexity depends on the inverse function of (a restriction of) the kernel function itself and on the inverse function of (a restriction of) its first derivative. In Section 6 we apply our results to seven kernel functions in order to obtain iteration bounds for large-update methods; to make clear that the small-update methods based on the new kernel functions are as good as the classical logarithmic barrier method we also derive iteration bounds for small-update methods. Finally, Section 7 contains some concluding remarks, and directions for future research.

We use the following notational conventions. If $x, s \in \mathbb{R}^n$, then xs denotes the coordinatewise (or Hadamard) product of the vectors x and s. Furthermore, e denotes the all-one vector of length n. The nonnegative orthant and positive orthant are denoted as $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$, respectively. Finally, if $z \in \mathbb{R}^n_+$ and $f : \mathbb{R}_+ \to \mathbb{R}_+$, then $f(z)$ denotes the vector in $\mathbb{R}^n_+$ whose i-th component is $f(z_i)$, with $1 \leq i \leq n$. We write $f(x) = O(g(x))$ if $f(x) \leq cg(x)$ for some positive constant c and $f(x) = \Theta(g(x))$ if $c_1 g(x) \leq f(x) \leq c_2 g(x)$ for positive constants c_1 and c_2. Like in this paragraph, functions like f and g always have a local meaning in this paper.

2 Preliminaries

2.1 The central path

We deal with the LO-problem in standard format:

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

where $A \in \mathbb{R}^{m\times n}, \operatorname{rank}(A) = m,\ b \in \mathbb{R}^m,\ c \in \mathbb{R}^n$, and its dual problem

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

It is well know that finding an optimal solution of (P) and (D) is equivalent to solving the following system.

$$\begin{aligned} Ax &= b, \quad x \geq 0, \\ A^T y + s &= c, \quad s \geq 0, \\ xs &= 0. \end{aligned} \tag{2}$$

The basic idea of primal-dual IPMs is to replace the third equation in (2), the so-called *complementarity condition* for (P) and (D), by the parameterized equation $xs = \mu$e, with $\mu > 0$. Thus we consider the system

$$\begin{aligned} Ax &= b, \quad x \geq 0, \\ A^T y + s &= c, \quad s \geq 0, \\ xs &= \mu \mathrm{e}. \end{aligned} \tag{3}$$

Due to the last equation, any solution (x, y, s) of (3) will satisfy $x > 0$ and $s > 0$. So a solution exists only if (P) and (D) satisfy the *interior-point condition* (IPC), i.e., there exists (x^0, s^0, y^0) such that

$$Ax^0 = b,\ x^0 > 0, \quad A^T y^0 + s^0 = c, s^0 > 0. \tag{4}$$

Surprising enough, if the IPC is satisfied then a solution exists, for each $\mu > 0$, and this solution is unique. It is denoted as $(x(\mu), y(\mu), s(\mu))$ and we call $x(\mu)$ the μ-center of (P) and $(y(\mu), s(\mu))$ the μ-center of (D). The set of μ-centers (with μ running through all positive real numbers) gives a homotopy path, which is called *the central path* of (P) and (D). The relevance of the central path for LO was recognized first by Sonnevend [15] and Megiddo [6]. If $\mu \to 0$ then the limit of the central path exists and since the limit points satisfy the complementarity condition, the limit yields optimal solutions for (P) and (D).

From a theoretical point of view the IPC can be assumed without loss of generality. In fact we may, and will assume that $x^0 = s^0 = \mathrm{e}$. In practice, this can be realized by embedding the given problems (P) and (D) in a homogeneous self-dual problem which has two additional variables and two additional constraints. For this and the other properties mentioned above, see, e.g., [14].

2.2 A wide class of primal-dual path-following methods

IPMs follow the central path approximately. Following the approach in [2] we now describe how this goes. We assume that we are given a strictly convex function $\Psi(v)$, $v \in \mathbb{R}^n_{++}$, such that $\Psi(v)$ is minimal at $v = \mathrm{e}$ and $\Psi(\mathrm{e}) = 0$. We will argue below that any such function determines in a natural way an IPM.

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 assumption we may assume this for $\mu = 1$, with $x(1) = s(1) = \mathrm{e}$. We then decrease μ to $\mu := (1 - \theta)\mu$, for some fixed $\theta \in (0, 1)$ and we solve the following linear system in the unknown vectors Δx, Δs and Δy:

$$\begin{aligned} A\Delta x &= 0, \\ A^T\Delta y + \Delta s &= 0, \\ s\Delta x + x\Delta s &= -\mu\, v\nabla\Psi(v), \end{aligned} \tag{5}$$

where

$$v := \sqrt{\frac{xs}{\mu}}. \tag{6}$$

We proceed by showing that system (5) uniquely defines a search direction $(\Delta x, \Delta s, \Delta y)$ and that the search direction vanishes if and only if $\Psi(v) = 0$, and this holds if and only if $v = \mathrm{e}$, i.e., if and only if $(x, y, s) = (x(\mu), y(\mu), s(\mu))$. To this end, using the vector v defined in (6), we introduce scaled versions of the displacements Δx and Δs as follows:

$$d_x := \frac{v\Delta x}{x}, \qquad d_s := \frac{v\Delta s}{s}. \tag{7}$$

Now one easily checks that the system (5), which defines the search directions, can be rewritten as

$$\begin{aligned} \bar{A}d_x &= 0, \\ \tfrac{1}{\mu}\bar{A}^T\Delta y + d_s &= 0, \\ d_x + d_s &= -\nabla\Psi(v), \end{aligned} \tag{8}$$

where $\bar{A} = AV^{-1}X$, with $V = \operatorname{diag}(v), X = \operatorname{diag}(x)$. The last equation in the above system is called the *scaled centering equation.* It states that the sum of the scaled search directions d_x and d_s is equal to $-\nabla\Psi(v)$, the steepest descent direction of $\Psi(v)$. Note that d_x and d_s are orthogonal vectors, since the vector d_x belongs to the null space and d_s to the row space of the matrix $\bar{A}$. Thus it follows that the scaled search directions d_x and d_s form an orthogonal decomposition of the steepest descent direction of the function $\Psi(v)$. Note that since d_x and d_s are orthogonal, we will have

$$d_x = d_s = 0 \quad \Leftrightarrow \quad \nabla\Psi(v) = 0 \quad \Leftrightarrow \quad v = \mathrm{e} \quad \Leftrightarrow \quad \Psi(v) = 0,$$

i.e., if and only if $x = x(\mu)$ and $s = s(\mu)$, as it should. Hence, if $(x, y, s) \neq (x(\mu), y(\mu), s(\mu))$ then $(\Delta x, \Delta s, \Delta y)$ is nonzero. By taking a step along the search direction, with the step size α defined by some line search rules, one constructs a new triple (x, y, s) according to

$$x_+ = x + \alpha\Delta x, \quad y_+ = y + \alpha\Delta y, \quad s_+ = s + \alpha\Delta s. \tag{9}$$

If necessary, this procedure is repeated until we find iterates (x, y, s) that are 'close' enough to $(x(\mu), y(\mu), s(\mu))$. Then μ is again reduced by the factor $1 - \theta$ and we apply the above method targeting at the new μ-centers, and so on. This process is repeated until μ is small enough, say until $n\mu \leq \varepsilon$; at this stage we have found an ε-solution of the problems (P) and (D).

2.3 A generic primal-dual algorithm

We can now describe the algorithm in a more formal way. The generic form of the algorithm is shown in Figure 1. It is clear from this description that closeness of (x, y, s) to $(x(\mu), y(\mu), s(\mu))$ is measured by the value of $\Psi(v)$, with τ as a threshold value: if $\Psi(v) \leq \tau$ then we start a new *outer iteration* by performing a μ-update, otherwise we enter an *inner iteration* by computing the search directions at the current iterates with respect to the current value of μ and apply (9) to get new iterates.

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. The choice of the *barrier update parameter* θ plays an important role both in theory and practice of IPMs. Usually, if θ is a constant independent of the dimension n of the problem, for instance $\theta = \frac{1}{2}$, then we call the algorithm a *large-update* (or *long-step*) method. If θ depends on the dimension of the problem, such as $\theta = 1/\sqrt{n}$, then the algorithm is named a *small-update* (or *short-step*) method.

The choice of the step size α $(0 < \alpha \leq 1)$ is another crucial issue in the analysis of the algorithm. It has to be taken such that the closeness of the iterates to the current μ-center improves by a sufficient amount. In the theoretical analysis the step size α is usually given a value that depends on the closeness of the current iterates to the μ-center.

2.4 Kernel functions

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

$$\Psi(v) = \sum_{i=1}^{n} \psi(v_i), \tag{10}$$

Generic Primal-Dual Algorithm for LO

Input:
A threshold parameter $\tau > 0$;
an accuracy parameter $\varepsilon > 0$;
a fixed barrier update parameter θ, $0 < \theta < 1$;
begin
$x := \mathbf{e}$; $s := \mathbf{e}$; $\mu := 1$;
while $n\mu \geq \varepsilon$ **do**
begin
$\mu := (1-\theta)\mu$;
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: Generic algorithm

where $\psi(t) \; : \; D \to \mathbb{R}_+$, with $\mathbb{R}_{++} \subseteq D$, is strictly convex and minimal at $t = 1$, with $\psi(1) = 0$. We call the univariate function $\psi(t)$ the *kernel function* of the barrier function $\Psi(v)$. 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.

Except for the kernel function considered in [3], all kernel functions considered so far are twice differentiable and go to infinity if either $t \downarrow 0$ or $t \to \infty$. Thus they satisfy

$$\psi'(1) = \psi(1) = 0; \tag{11-a}$$

$$\psi''(t) > 0; \tag{11-b}$$

$$\lim_{t \downarrow 0} \psi(t) = \lim_{t \to \infty} \psi(t) = \infty. \tag{11-c}$$

Clearly, (11-a) and (11-b) say that $\psi(t)$ is a nonnegative strictly convex function such that $\psi(1) = 0$. Note that this implies that $\psi(t)$ is completely determined by its second derivative:

$$\psi(t) = \int_1^t \int_1^\xi \psi''(\zeta)\, d\zeta d\xi. \tag{12}$$

Moreover, (11-c) expresses that $\psi(t)$ is coercive and has the barrier property. Having such a kernel function $\psi(t)$, its definition is extended to positive n-dimensional vectors v by (10), thus yielding the induced (scaled) barrier function $\Psi(v)$. In the sequel we assume in this paper that a kernel function satisfies (11-a)–(11-c).

As we indicated in the previous section $\Psi(v)$ not only serves to define a search direction, but also as a measure of closeness of the current iterates to the μ-center. In the analysis of the algorithm we also use the *norm-based proximity measure* $\delta(v)$ defined by

$$\delta(v) := \frac{1}{2} \|\nabla\Psi(v)\| = \frac{1}{2} \|d_x + d_s\| . \tag{13}$$

Note that since $\Psi(v)$ is strictly convex and minimal at $v = \mathrm{e}$ we have

$$\Psi(v) = 0 \Leftrightarrow \delta(v) = 0 \Leftrightarrow v = \mathrm{e}.$$

Thus the algorithm considered in this paper uses the barrier function $\Psi(v)$ to measure closeness of the iterates to the μ-center; as will become clear below, in the analysis of the algorithm $\delta(v)$ serves as a second proximity measure. Both measures are naturally determined by the kernel function.

2.5 Further conditions on the kernel function

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

$$t\psi''(t) + \psi'(t) > 0, \quad t < 1, \tag{14-a}$$
$$t\psi''(t) - \psi'(t) > 0, \quad t > 1, \tag{14-b}$$
$$\psi'''(t) < 0, \quad t > 0, \tag{14-c}$$
$$2\psi''(t)^2 - \psi'(t)\psi'''(t) > 0, \quad t < 1, \tag{14-d}$$
$$\psi''(t)\psi'(\beta t) - \beta\psi'(t)\psi''(\beta t) > 0, \quad t > 1,\ \beta > 1. \tag{14-e}$$

Note that conditions (14-c) and (14-d) require that $\psi(t)$ is three times differentiable. Furthermore, condition (14-a) is obviously satisfied if $t \geq 1$, since then $\psi'(t) \geq 0$ and, similarly, condition (14-b) is satisfied if $t \leq 1$, since then $\psi'(t) \leq 0$. Also (14-d) is obviously satisfied if $t \geq 1$ since then $\psi'(t) \geq 0$, whereas $\psi'''(t) < 0$. We conclude that conditions (14-a) and (14-d) are conditions on the barrier behavior of $\psi(t)$. On the other hand, condition (14-b) only deals with $t \geq 1$ and hence concerns the growth behavior of $\psi(t)$. Condition (14-e) is technically more involved; we will discuss it later.

The next two lemmas make clear that conditions (14-a) and (14-b) admit a nice interpretation. The proof of the first lemma can be found in [12]; the proof of the second lemma is of the same spirit and is left to the reader.

Lemma 2.1 (Lemma 2.1.2 in [12]) *Let $\psi(t)$ be a twice differentiable function for $t > 0$. Then the following three properties are equivalent:*

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

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

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

Lemma 2.2 *Let $\psi(t)$ be a twice differentiable function for $t > 0$. Then the following three properties are equivalent:*

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

(ii) $t\psi''(t) - \psi'(t) \geq 0, \quad t > 0$;
(iii) $\psi\left(\sqrt{\xi}\right)$ *is convex.*

Following [8], we call the property described in Lemma 2.1 *exponential convexity*, or shortly *e-convexity*. This property has been proven to be very useful in the analysis of primal-dual algorithms based on kernel functions (cf. [10, 11, 12]). We recall from [12, Lemma 2.1.3] that if $\psi(t)$ is e-convex then $\psi(t^a)$ is e-convex as well, for every $\alpha \in \mathbb{R}$. Also, for any $\beta_0 \in \mathbb{R}$ the function

$$\beta_0 \log t + \sum_{i=1}^{m} \beta_i \left(t^{\alpha_i} - 1\right)$$

is e-convex whenever $\beta_i \geq 0$ and $\alpha_i \in \mathbb{R}$ for $i = 1, \ldots, m$ [12, Propostion 2.1.4].

It may be noted that the kernel function $\psi(t)$ is defined to be self-regular in [11, 12] if $\psi(t)$ is e-convex and moreover

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

where $\Upsilon_{p,q}(t)$ denotes the prototype self-regular kernel function (1). 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 kernel function $\Upsilon_{p,q}(t)$ (and hence each self-regular kernel function) satisfies (14-c) only if $p \leq 1$.

Remark 2.3 It is worth pointing out that the conditions (14-a)-(14-d) are logically independent. The following table shows four kernel functions and the signs indicate whether a condition is satisfied (+) or not (−).

$\psi(t)$	(14-a)	(14-b)	(14-c)	(14-d)
$3t^2 - 2t - 2 + \frac{1}{t^2} - 2\log t$	−	+	+	+
$(t+2)(t-1) - 3\log t$	+	−	+	+
$t^3 + t^{-3} - 2$	+	+	−	+
$8t^2 - 11t + 1 + \frac{2}{\sqrt{t}} - 4\log t$	+	+	+	−

The last condition, however, is not independent from the other conditions. One has

Lemma 2.4 *If $\psi(t)$ satisfies (14-b) and (14-c) then $\psi(t)$ satisfies (14-e).*

Proof: For $t > 1$ we consider the function

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

Since $f(1) = 0$, the lemma will follow if $f'(\beta) > 0$ for $\beta > 1$. This can be shown as follows:

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

The last inequality follows since $\psi''(\beta t) \geq 0$, $t\psi''(t) - \psi'(t) \geq 0$, by (14-b), and $-\beta t\psi'(t)\psi'''(\beta t) > 0$, since $t > 1$, which implies $\psi'(t) > 0$, and $\psi'''(\beta t) < 0$, by (14-c). This proves the lemma. □

The following two lemmas deal with consequence of condition (14-c).

Lemma 2.5 *If the kernel function $\psi(t)$ satisfies (14-c), then*

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

Proof: Consider the function $f(t) = 2\psi(t) - (t-1)\psi'(t)$. One has $f(1) = 0$ 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 $f''(t) < 0$, whence $f'(t) > 0$ and $f(t) < 0$. From this the lemma follows. □

Lemma 2.6 *If the kernel function $\psi(t)$ satisfies (14-c), then*

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

Proof: By using Taylor's theorem and $\psi(1) = \psi'(1) = 0$, we obtain

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

where $1 < \xi < t$ if $t > 1$ and $t < \xi < 1$ if $t < 1$. Since $\psi'''(\xi) < 0$ the second inequality for $t > 1$ and the first inequality for $t < 1$ in the lemma follow. The remaining two inequalities are an immediate consequence of Lemma 2.5. □

The theory developed in this paper applies to kernel functions that satisfy the conditions (14-a), (14-c), (14-d) and (14-e). In this paper we call any such function an *eligible kernel function*. Due to Lemma 2.4, any kernel function satisfying the first four conditions, (14-a) to (14-d) , is eligible.

2.6 Seven kernel functions

By way of example we consider in this paper the seven kernel functions in the second column of Table 1. In this section we show that these kernel functions are eligible. For the first 6 functions this is almost immediate: they satisfy the conditions (14-a) to (14-d), as we now show. The 5th and 6th columns in Table 1 make clear that each of these kernel functions satisfy the conditions (14-a) and (14-b). From Table 2 it follows that they also satisfy conditions (14-c) and (14-d). It is obvious from the 2nd column that (14-c) is satisfied. That also (14-d) is satisfied is not always immediately clear from the given expressions in the 4th column in Table 2; we leave it as an exercise to the reader to verify that these expressions indeed are nonnegative for $t \leq 1$.

The seventh kernel function requires special attention. It is clear from the tables that is satisfies (14-a), (14-c) and (14-d), but it fails to satisfy (14-b). To prove that this function is eligible it suffices to show that it satisfies (14-e). This, however, is easy. Some straightforward computations yield that for $\psi = \psi_7$ one has

$$\psi''(t)\psi'(\beta t) - \beta\psi'(t)\psi''(\beta t) = \frac{q\,(1-\beta^{-q})}{t^{q+1}},$$

i	$\psi_i(t)$	$\psi_i'(t)$	$\psi_i''(t)$	$t\psi_i''(t) + \psi_i'(t)$	$t\psi_i''(t) - \psi_i'(t)$
1	$\frac{t^2-1}{2} - \log t$	$t - \frac{1}{t}$	$1 + \frac{1}{t^2}$	$2t$	$\frac{2}{t}$
2	$\frac{1}{2}\left(t - \frac{1}{t}\right)^2$	$t - \frac{1}{t^3}$	$1 + \frac{3}{t^4}$	$2t + \frac{2}{t^3}$	$\frac{4}{t^3}$
3	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q-1}$	$t - t^{-q}$	$1 + qt^{-q-1}$	$2t + (q-1)t^{-q}$	$(q+1)t^{-q}$
4	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q(q-1)} - \frac{q-1}{q}(t-1)$	$t - 1 - \frac{t^{-q}-1}{q}$	$1 + t^{-q-1}$	$2t + \frac{q-1}{q}\left(t^{-q} - 1\right)$	$\frac{q-1}{q} + \frac{q+1}{q}\, t^{-q}$
5	$\frac{t^2-1}{2} + \frac{e^{\frac{1}{t}}-e}{e}$	$t - \frac{e^{\frac{1}{t}-1}}{t^2}$	$1 + \frac{1+2t}{t^4}\, e^{\frac{1}{t}-1}$	$2t + \frac{1+t}{t^3}\, e^{\frac{1}{t}-1}$	$\frac{1+3t}{t^3}\, e^{\frac{1}{t}-1}$
6	$\frac{t^2-1}{2} - \int_1^t e^{\frac{1}{\xi}-1} d\xi$	$t - e^{\frac{1}{t}-1}$	$1 + \frac{e^{\frac{1}{t}-1}}{t^2}$	$2t + \frac{1-t}{t}\, e^{\frac{1}{t}-1}$	$\frac{1+t}{t}\, e^{\frac{1}{t}-1}$
7	$t - 1 + \frac{t^{1-q}-1}{q-1}$	$1 - t^{-q}$	qt^{-q-1}	$1 + (q-1)t^{-q}$	$-1 + (q+1)t^{-q}$

Table 1: The seven kernel functions and the conditions (14-a) and (14-b) ($q > 1$).

i	$\psi_i(t)$	$\psi_i'''(t)$	$2\psi_i''(t)^2 - \psi_i'(t)\psi_i'''(t)$
1	$\frac{t^2-1}{2} - \log t$	$-\frac{2}{t^3}$	$2 + \frac{6}{t^2}$
2	$\frac{1}{2}\left(t - \frac{1}{t}\right)^2$	$-\frac{12}{t^5}$	$2 + \frac{24}{t^4} + \frac{6}{t^8}$
3	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q-1}$	$-q(q+1)t^{-q-2}$	$2\left(1 + \frac{q}{t^{q+1}}\right)^2 + \frac{q(q+1)\left(t^{q+1}-1\right)}{t^{2(q+1)}}$
4	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q(q-1)} - \frac{q-1}{q}(t-1)$	$-(q+1)t^{-q-2}$	$2\left(1 + \frac{1}{t^{q+1}}\right)^2 + \frac{(q+1)\left(t-1+\frac{1-t^{-q}}{q}\right)}{t^{q+2)}}$
5	$\frac{t^2-1}{2} + \frac{e^{\frac{1}{t}}-e}{e}$	$-\frac{1+6t+6t^2}{t^6}\, e^{\frac{1}{t}-1}$	$\frac{2\left(t^4+(1+2t)e^{\frac{1}{t}-1}\right)^2-(1+6t+6t^2)\left(e^{\frac{1}{t}-1}-t^3\right)e^{\frac{1}{t}-1}}{t^8}$
6	$\frac{t^2-1}{2} - \int_1^t e^{\frac{1}{\xi}-1} d\xi$	$-\frac{1+2t}{t^4}\, e^{\frac{1}{t}-1}$	$2\left(1 + \frac{e^{\frac{1}{t}-1}}{t^2}\right)^2 + (1+2t)\frac{e^{\frac{1}{t}-1}\left(t-e^{\frac{1}{t}-1}\right)}{t^4}$
7	$t - 1 + \frac{t^{1-q}-1}{q-1}$	$-q(q+1)t^{-q-2}$	$q\left(q - 1 + (q+1)t^q\right)t^{-2(q+1)}$

Table 2: The seven kernel functions and the conditions (14-c) and (14-e).

which is clearly positive if $t > 1$ and $\beta > 1$.

Thus we have shown that each of the seven kernel functions is eligible. We now briefly discuss these functions. The first function, $\psi_1(t)$, is the kernel function of the classical primal-dual logarithmic barrier function. This can be seen as follows. The barrier function induced by $\psi(t) = \psi_1(t)$ is

$$\Psi(v) = \sum_{i=1}^{n} \psi(v_i) = \sum_{i=1}^{n} \left(\frac{v_i^2 - 1}{2} - \log v_i\right) = \frac{1}{2}\sum_{i=1}^{n}\left(v_i^2 - 1 - \log v_i^2\right).$$

Because of (6), by defining $\Phi(x,s,\mu) := \Psi(v)$ the barrier function can be expressed in x and s:

$$\Phi(x,s,\mu) = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i s_i}{\mu} - 1 - \log\frac{x_i s_i}{\mu}\right) = \frac{1}{2\mu}\left(x^T s - \mu\sum_{i=1}^{n}\log x_i s_i\right) + \frac{n\log\mu - n}{2}.$$

The expression between the brackets is the well known primal-dual logarithmic barrier function for (P) and (D). It differs from $\Psi(v)$ only by some additive and multiplicative 'constants', depending on n and μ, but not on x and s. Note that for any kernel function, the above definition of $\Phi(x,s,\mu)$ gives rise to a function that goes to infinity if, for some i, $x_i s_i$ goes either to infinity or to 0, and hence $\Phi(x,s,\mu)$ is a barrier function for the feasible domain. It may be worth emphasizing that the search direction in this paper, as defined by (5), is obtained neither by solving a barrier problem, nor as a primal-dual Newton direction induced by $\Phi(x,s,\mu)$.

The second kernel function has been studied in [9]; one may easily verify that it is a special case of $\psi_3(t)$, by taking $q = 3$. The third one has been studied for general $q > 1$ in [8, 12]. It is in turn a special case of a so-called self-regular kernel function, as studied in [11, 12]. Also note that $\psi_1(t)$ is the limiting value of $\psi_3(t)$ when q approaches 1. The fourth kernel function is the special case of $\Upsilon_{p,q}(t)$, for $p = 1$. The fifth and sixth kernel functions are new, at least to the knowledge of the authors. In each of the first six cases we can write $\psi(t)$ as

$$\psi(t) = \frac{t^2 - 1}{2} + \psi_b(t), \tag{15}$$

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 last kernel function ($\psi_7(t)$, with $q > 1$) differs from all the other kernel function in that its growth term, i.e., $t - 1$, is linear in t. It was first introduced and analyzed in [4].

3 Growth behavior

Note that at the start of each outer iteration of the algorithm, just before the update of μ with the factor $1-\theta$, we have $\Psi(v) \le \tau$. Due to the update of μ the vector v is divided by the factor $\sqrt{1-\theta}$, with $0 < \theta < 1$, which in general 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 $\Psi(v)$. We start with the following lemma.

Lemma 3.1 *Suppose that $\psi(t_1) = \psi(t_2)$, with $t_1 \le 1 \le t_2$ and $\beta \ge 1$. Then*

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

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 increasing. 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 $t_2 = t_1 = 1$. This completes the proof. □

We have the following result.

Theorem 3.2 *Let $\varrho : [0, \infty) \to [1, \infty)$ be the inverse function of $\psi(t)$ for $t \geq 1$. Then we have for any positive vector v and any $\beta \geq 1$:*

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

Proof: First we consider the case where $\beta > 1$. 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

$$\beta\psi'(\beta v_i) = \lambda\psi'(v_i), \quad i = 1, \ldots, n, \tag{16}$$

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 3.1 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 $v_i = v_i^{(2)} > 1$. Thus we have shown that without loss of generality we may assume that $v_i > 1$ for all i. Note that then (16) implies $\beta\psi'(\beta v_i) > 0$ and $\psi'(v_i) > 0$, whence also $\lambda > 0$. Now defining $g(t)$ according to

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

we deduce from (16) that $g(v_i) = \frac{\beta}{\lambda}$ for all i. One has

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

At this stage we use that $\psi(t)$ satisfies condition (14-e). Due to this we have $g'(t) > 0$, for $t > 1$ and $\beta > 1$. So $g(t)$ is strict monotonically increasing. Hence it follows that all v_i's are mutually equal. Putting $v_i = t > 1$, for all i, we deduce from $\Psi(v) = z$ that $n\psi(t) = z$. This implies $t = \varrho(\frac{z}{n})$. Hence the maximal value that $\Psi(v)$ can attain is given by

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

This proves the theorem if $\beta > 1$. For the case $\beta = 1$ it suffices to observe that both sides of the inequality in the theorem are continuous in β. □

Corollary 3.3 *Using the notation of Theorem 3.2, we have*

$$\Psi(\beta v) \leq \frac{n}{2}\psi''(1)\left(\beta\varrho\left(\frac{\Psi(v)}{n}\right) - 1\right)^2.$$

Proof: Since $\beta \geq 1$ and $\varrho(\frac{\Psi(v)}{n}) \geq 1$, the corollary follows from Theorem 3.2 by using also Lemma 2.6. □

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

$$L_\psi(n, \theta, \tau) := n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right) \tag{17}$$

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

Remark 3.4 Note that the bound of Theorem 3.2 is sharp: one may easily verify that if $v = \beta e$, with $\beta \geq 1$, then the bound holds with equality.

4 Decrease of the barrier function during an inner iteration

4.1 Computation of the step size and the decrease

After a damped step we have

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

Hence, with

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

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

$$v_+^2 = \frac{x_+ s_+}{\mu} = \left(v + \alpha d_x\right)\left(v + \alpha d_s\right). \tag{18}$$

Hence, since ψ is e-convex,

$$\Psi\left(v_{+}\right)=\Psi\left(\sqrt{\left(v+\alpha d_{x}\right)\left(v+\alpha d_{s}\right)}\right) \leq \tfrac{1}{2}\left(\Psi\left(v+\alpha d_{x}\right)+\Psi\left(v+\alpha d_{s}\right)\right).$$

Defining

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

we thus have $f(\alpha) \leq f_1(\alpha)$, where

$$f_1(\alpha):=\tfrac{1}{2}\left(\Psi\left(v+\alpha d_{x}\right)+\Psi\left(v+\alpha d_{s}\right)\right)-\Psi\left(v\right).$$

Obviously,

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

Taking the derivative with respect to α, we get

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

This gives, using the third equation in (8) and (13),

$$f_1'(0)=\tfrac{1}{2}\nabla\Psi(v)^T\left(d_x+d_s\right)=-\tfrac{1}{2}\nabla\Psi(v)^T\nabla\Psi(v)=-2\delta(v)^2. \tag{19}$$

Differentiating once more, we obtain

$$f_1''(\alpha)=\tfrac{1}{2}\sum_{i=1}^{n}\left(\psi''\left(v_i+\alpha d_{xi}\right)d_{xi}^{\,2}+\psi''\left(v_i+\alpha d_{si}\right)d_{si}^{\,2}\right). \tag{20}$$

Below we use the following notation:

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

Lemma 4.1 *One has* $f_1''(\alpha)\leq 2\delta^2\,\psi''\left(v_1-2\alpha\delta\right)$*.*

Proof: Since d_x and d_s are orthogonal, (13) implies that $\|(d_x,d_s)\|=2\delta$. Hence we have $\|d_x\|\leq 2\delta$ and $\|d_s\|\leq 2\delta$. Therefore,

$$v_i+\alpha d_{xi}\geq v_1-2\alpha\delta, \quad v_i+\alpha d_{si}\geq v_1-2\alpha\delta, \quad 1\leq i\leq n.$$

Due to (14-c), $\psi''(t)$ is monotonically decreasing. Therefore, from (20) we obtain

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

This proves the lemma. □

Lemma 4.2 $f_1'(\alpha)\leq 0$ *holds certainly if* α *satisfies the inequality*

$$-\psi'\left(v_1-2\alpha\delta\right)+\psi'\left(v_1\right)\leq 2\delta. \tag{21}$$

Proof: We may write, using Lemma 4.1, and also (19),

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

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

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

which proves the lemma. □

Lemma 4.3 *Let $\rho : [0, \infty) \to (0, 1]$ denote the inverse function of the restriction of $-\frac{1}{2}\psi'(t)$ to the interval $(0, 1]$. Then the largest step size α that satisfies (21) is given by*

$$\bar{\alpha} := \frac{1}{2\delta}\left(\rho\left(\delta\right) - \rho\left(2\delta\right)\right) \tag{22}$$

Proof: We want α such that (21) holds, with α as large as possible. Since $\psi''(t)$ is decreasing, the derivative with respect to v_1 of the expression at the left in (21) (i.e. $-\psi''\left(v_1 - 2\alpha\delta\right) + \psi''\left(v_1\right)$) is negative. Hence, fixing δ, the smaller v_1 is, the smaller α will be. One has

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

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 $\psi'\left(v_1\right) \leq 0$). Hence, the worst situation for the step size occurs when v_1 satisfies

$$-\tfrac{1}{2}\psi'\left(v_1\right) = \delta. \tag{23}$$

The derivative with respect to α of the expression at the left in (21) equals $2\delta\psi''\left(v_1 - 2\alpha\delta\right) \geq 0$, and hence the left hand side is increasing in α. So the largest possible value of α satisfying (21), satisfies

$$-\tfrac{1}{2}\psi'\left(v_1 - 2\alpha\delta\right) = 2\delta. \tag{24}$$

Due to the definition of ρ, (23) and (24) can be written as

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

respectively. This implies,

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

proving the lemma. □

Lemma 4.4 *Let ρ and $\bar{\alpha}$ be as defined in Lemma 4.3. Then*

$$\bar{\alpha} \geq \frac{1}{\psi''\left(\rho\left(2\delta\right)\right)}. \tag{25}$$

Proof: By the definition of ρ,

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

Taking the derivative with respect to δ, we find

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

which gives

$$\rho'(\delta) = -\frac{2}{\psi''(\rho(\delta))} < 0. \tag{26}$$

Hence ρ is monotonically decreasing. An immediate consequence of (22) is

$$\bar{\alpha} = \frac{1}{2\delta}\int_{2\delta}^{\delta} \rho'(\sigma)\,d\sigma = \frac{1}{\delta}\int_{\delta}^{2\delta} \frac{d\sigma}{\psi''(\rho(\sigma))}, \tag{27}$$

where we also used (26). 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 (14-c), ψ'' is monotonically decreasing. So $\psi''(\rho(\sigma))$ is maximal for $\sigma \in [\delta, 2\delta]$ when $\rho(\sigma)$ is minimal. 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

$$\tilde{\alpha} = \frac{1}{\psi''(\rho(2\delta))} \tag{28}$$

and we will use $\tilde{\alpha}$ as the default step size. By Lemma 4.4 we have $\bar{\alpha} \geq \tilde{\alpha}$.

Lemma 4.5 *If the step size α is such that $\alpha \leq \bar{\alpha}$ then*

$$f(\alpha) \leq -\alpha\,\delta^2. \tag{29}$$

Proof: Let the univariate function h be such that

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

Due to Lemma 4.1, $f_1''(\alpha) \leq h''(\alpha)$. 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 4.3, we have

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

Since $h''(\alpha)$ is increasing in α, using Lemma A.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. □

Combining the results of Lemma 4.4 and Lemma 4.5 we obtain

Theorem 4.6 *With $\tilde{\alpha}$ being the default step size, as given by (28), one has*

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''\left(\rho\left(2\delta\right)\right)}. \tag{30}$$

Lemma 4.7 *The right hand side expression in (30) 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 (30) is monotonically decreasing in δ if and only if the function

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

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 (14-d). Hence the lemma is proved. □

We want 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.

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

We need the following lemma.

Lemma 4.8 *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 $t_2 = 1$, because then $\psi(t_1) = \psi(t_2) = 0$ implies $t_1 = t_2 = 1$. So we may assume that $t_1 < 1 < t_2$. Since $\psi(t_1) = \psi(t_2)$, Lemma 2.6 implies:

$$\tfrac{1}{2}\left(t_1 - 1\right)^2 \psi''(1) < \psi(t_1) = \psi(t_2) < \tfrac{1}{2}\left(t_2 - 1\right)^2 \psi''(1).$$

Hence, since $\psi''(1) > 0$, it follows that $t_2 - 1 > 1 - t_1$. Using this and Lemma 2.5, while assuming $-\psi'(t_1) < \psi'(t_2)$, we may write

$$\psi(t_2) > \tfrac{1}{2}(t_2 - 1)\psi'(t_2) > \tfrac{1}{2}(1 - t_1)\psi'(t_2) > -\tfrac{1}{2}(1 - t_1)\psi'(t_1) = \tfrac{1}{2}(t_1 - 1)\psi'(t_1) > \psi(t_1).$$

This contradiction proves the lemma. □

Theorem 4.9 *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 $v = \mathrm{e}$ 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 $\omega > 0$, 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 is

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

where $\lambda \in \mathbb{R}$. 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, 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 4.8, 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

$$4\delta(v)^2 = k\left(\psi'(t)\right)^2, \quad t = \varrho\left(\frac{\omega}{k}\right). \tag{31}$$

The question is now which value of k minimizes $\delta(v)^2$. To investigate this, we take the derivative with respect to k of (31) extended to $k \in \mathbb{R}$. This gives

$$\frac{d\,4\delta(v)^2}{dk} = \left(\psi'(t)\right)^2 + 2k\psi'(t)\psi''(t)\frac{dt}{dk}. \tag{32}$$

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 (32) 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 $f(t) := (\psi'(t))^2 - 2\psi(t)\psi''(t)$ 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 (31) that

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

This completes the proof of the theorem. □

Remark 4.10 The bound of Theorem 4.9 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.

Remark 4.11 It is worth noting that the proof of Theorem 4.9 implies that our kernel functions satisfy the inequality (cf. [12, page 37])

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

Combining the results of Theorem 4.6 and Theorem 4.9 we obtain

$$f(\tilde{\alpha}) \leq -\frac{(\psi'(\varrho(\Psi(v))))^2}{4\psi''(\rho(\psi'(\varrho(\Psi(v)))))}. \tag{33}$$

This expresses the decrease in $\Psi(v)$ during an inner iteration completely in ψ, its first and second derivatives and the inverse functions ρ and ϱ. The entities arising in the right hand side of (33) are depicted in Figure 2.

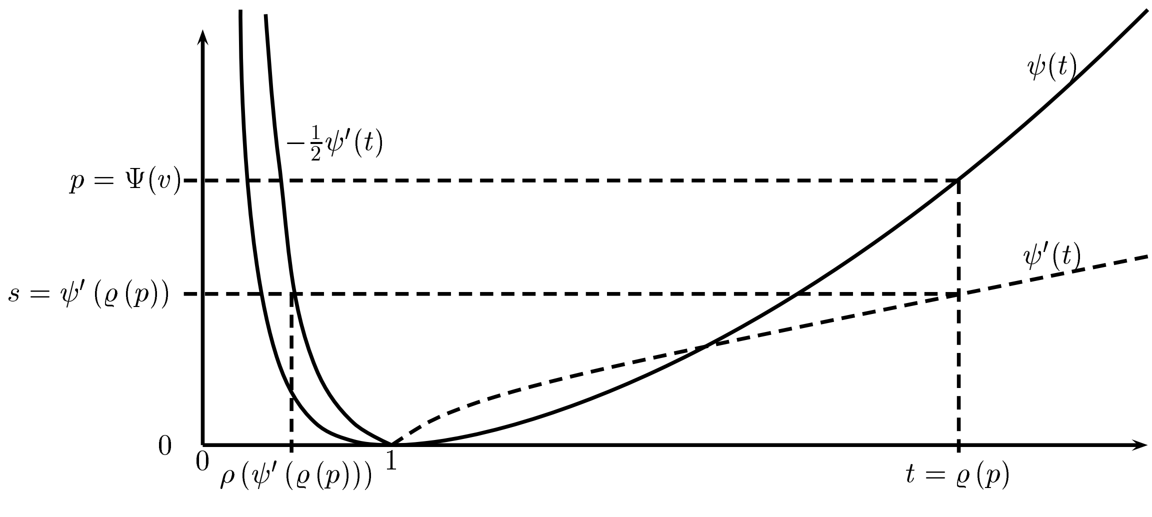

Figure 2: Graphical illustration of the entities arising in (33).

5 Iteration bounds

After the update of μ to $(1-\theta)\mu$ we have, by Theorem 3.2 and (17),

$$\Psi(v_+) \leq L_\psi(n,\theta,\tau) = n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right). \tag{34}$$

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, $k = 1, 2, \ldots$. The decrease on each inner iteration is given by (33). In the sequel we assume that the expression in the right hand side expression of (33) satisfies

$$\frac{\left(\psi'\left(\varrho\left(\Psi(v)\right)\right)\right)^2}{4\psi''\left(\rho\left(\psi'\left(\varrho\left(\Psi(v)\right)\right)\right)\right)} \geq \kappa\Psi(v)^{1-\gamma} \tag{35}$$

for some positive constants κ and γ, with $\gamma \in (0,1]$. It may be worth noting at this point that property (35) has to be checked for each case individually. When dealing with the seven specific examples later on, we will show how appropriate values of κ and γ are obtained in each case.

Lemma 5.1 *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.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 (34). 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. [14] Lemma II.17, page 116)

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

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

$$\frac{\Psi_0^\gamma}{\theta\kappa\gamma}\log\frac{n}{\varepsilon} \leq \frac{1}{\theta\kappa\gamma}\left(n\psi\left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}}\right)\right)^\gamma \log\frac{n}{\varepsilon}. \tag{36}$$

6 Application to the seven kernel functions

6.1 Introduction

We apply the results of the previous sections to obtain iteration bounds for large- and small-update methods based on the seven kernel functions introduced before. Thus, for each of the kernel functions basically we will do the following.

Step 0: Input a kernel function ψ; 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 $\Psi(v)$ 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 positive constants κ and γ, with $\gamma \in (0,1]$, such that

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

Step 6: Calculate the uniform upper bound Ψ_0 for $\Psi(v)$ from

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

Step 7: Derive an upper bound for the total number of iterations from

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

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

Using the above scheme, our aim is to compute iteration bounds for large- and small-update methods based on the seven 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 effect the order of magnitude of the bounds. Small-update methods are characterized by $\tau = O(1)$ and $\theta = \Theta(\frac{1}{\sqrt{n}})$.

One more remark is in order. At the start of each inner iteration we have $\Psi(v) \geq \tau$. By Theorem 4.9 this implies that $\delta(v) \geq \frac{1}{2}\psi'\left(\varrho\left(\tau\right)\right)$. We always assume that $\tau \geq 1$, and that τ is large enough to ensure that $\delta(v) \geq 1$ at the start of each inner iteration.

By way of example, we perform in the next section the complete analysis for $\psi(t) = \psi_6(t)$. Later on, we discuss more briefly the analysis for the remaining kernel functions. In the analysis we make use of three more lemmas. The first two lemmas are useful in deriving bounds for the inverse functions ϱ and ρ, in case these functions cannot be computed explicitly. The third lemma sometimes gives a better estimate for Ψ_0 in Step 6. The lemmas apply only to the first six kernel functions.

Lemma 6.1 *Let $\psi(t) = \psi_i(t)$, with $1 \leq i \leq 6$, and let $\underline{\rho} : [0,\infty) \to (0,1]$ be the inverse function of the restriction of $-\psi_b'(t)$ to the interval $(0,1]$, with $\psi_b(t)$ as defined in (15). Then one has*

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

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 six cases, it follows from this that

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

proving the lemma. □

To illustrate the use of this lemma, consider for example, the case $i = 2$. Then

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

whence $\psi_b(t) = \frac{t^{-2}-1}{2}$. 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 6.1,

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

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 used our assumption that $\delta \geq 1$ at the start of each inner iteration.

Lemma 6.2 *When $\psi(t) = \psi_i(t)$ and $1 \leq i \leq 6$, then*

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

Proof: The inverse function of $\psi(t)$ for $t \in [1, \infty)$ is obtained by solving t from the equation $\psi(t) = s$, for $t \geq 1$. In almost all ceases 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

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

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

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

For the second inequality we use that $\psi_i''(t) \geq 1$ for $1 \leq i \leq 6$, as is clear from Table 1. Also using (12) we may write

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

which implies

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

This completes the proof. □

Lemma 6.3 *Let $1 \leq i \leq 6$. Then one has*

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

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

Proof: By Lemma 6.2 we have $\varrho(s) \leq 1 + \sqrt{2s}$. Hence, also using (34) we have

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

Applying Lemma 2.6 we obtain

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

where we also used

$$1-\sqrt{1-\theta} = \frac{\theta}{1+\sqrt{1-\theta}} \leq \theta. \tag{37}$$

This proves the lemma. □

6.2 Example: analysis of methods based on $\psi_6(t)$

Consider $\psi(t) = \psi_6(t)$:

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

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 6.1,

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

It follows that

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

By Lemma 6.2 the inverse function of $\psi(t)$ for $t \in [1,\infty)$ satsifies

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

Thus we have, omitting the argument v,

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

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

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

Substitution gives, after some elementary reductions, while using $\Psi_0 \geq \Psi \geq \tau \geq 1$,

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

Thus it follows that

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

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

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

We use Lemma 6.3, 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

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

Thus the total number of iterations is bounded above by

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

For large-update methods (when $\tau = O(n)$ and $\theta = \Theta(1)$) the right hand side expression becomes $O\left(\sqrt{n}\,(\log n)^2 \log\frac{n}{\varepsilon}\right)$, and for small-update methods (when $\tau = O(1)$ and $\theta = \Theta(\frac{1}{\sqrt{n}})$) the right hand side expression becomes $O\left(\sqrt{n}\log\frac{n}{\varepsilon}\right)$.

6.3 Analysis of all the examples

In this section we survey the analysis results for all the kernel functions $\psi_i(t)$, $1 \leq i \leq 7$. The analysis for each of the kernel functions goes in the same way as in the previous section for $\psi_6(t)$. To save space, we do not present details of the computations for the other kernel functions. We only give the outcome of each step of our computational scheme for each of the kernel functions. For readers who want to see the details of the computations, we refer to an earlier draft of this paper that can be downloaded from `http://ssor.twi.tudelft.nl/~roos/wpapers.html`. The outcome of the steps 1 to 7 is given in the Tables 3, 4 and 5.

6.4 Complexity results

To get complexity results we finally have to perform Step 8 in our scheme. Setting $\tau = O(n)$ and $\theta = \Theta(1)$ we obtain the iteration bound for large-update methods, and setting $\tau = O(1)$

i	kernel function $\psi_i(t)$	$\rho(s)$	$f(\tilde{\alpha}) \leq$
1	$\frac{t^2-1}{2} - \log t$	$= \frac{1}{s+\sqrt{1+s^2}}$	$-\frac{1}{19}$
2	$\frac{1}{2}\left(t - \frac{1}{t}\right)^2$	$\geq \frac{1}{(1+2s)^{\frac{1}{3}}}$	$-\frac{\delta^{\frac{2}{3}}}{27}$
3	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q-1}$	$\geq \frac{1}{(1+2s)^{\frac{1}{q}}}$	$-\frac{\delta^2}{1+q(1+4\delta)^{\frac{q+1}{q}}}$
4	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q(q-1)} - \frac{q-1}{q}(t-1)$	$\geq \frac{1}{(1+2qs)^{\frac{1}{q}}}$	$-\frac{\delta^2}{1+(1+4q\delta)^{\frac{q+1}{q}}}$
5	$\frac{t^2-1}{2} + \frac{e^{\frac{1}{t}}-e}{e}$	$\geq \frac{1}{1+\log(1+2s)}$	$-\frac{\delta^2}{1+6\delta(1+\log(1+4\delta))^2}$
6	$\frac{t^2-1}{2} - \int_1^t e^{\frac{1}{\xi}-1} d\xi$	$\geq \frac{1}{1+\log(1+2s)}$	$-\frac{\delta^2}{1+(1+4\delta)(1+\log(1+4\delta))^2}$
7	$t - 1 + \frac{t^{1-q}-1}{q-1}$	$= \frac{1}{(1+2s)^{\frac{1}{q}}}$	$-\frac{\delta^2}{q(4\delta+1)^{\frac{q+1}{q}}}$

Table 3: Results of Step 1 and 2.

i	$\varrho(s)$	$\delta(v) \geq$	γ	κ
1	not needed	not needed	1	$\frac{1}{19}$
2	$= \sqrt{\frac{s}{2}} + \sqrt{1+\frac{s}{2}}$	$\frac{1}{2}\Psi^{\frac{1}{2}}$	$\frac{2}{3}$	0.02333
3	$\sqrt{1+2s} \leq \varrho(s) \leq 1 + \sqrt{2s}$	$\frac{1}{2}\left(\frac{\Psi}{2}\right)^{\frac{1}{2}}$	$\frac{q+1}{2q}$	$\frac{1}{56\,q}$
4	$\sqrt{1+2s} \leq \varrho(s) \leq 1 + \sqrt{2s}$	$\frac{\Psi}{1+\sqrt{1+2\Psi}}$	$\frac{q+1}{2q}$	$\frac{1}{53\,q}$
5	$\sqrt{1+2s} \leq \varrho(s) \leq 1 + \sqrt{2s}$	$\frac{\Psi}{1+\sqrt{1+2\Psi}}$	$\frac{1}{2}$	$\frac{1}{19\left(1+\log(1+\sqrt{\Psi_0})\right)^2}$
6	$\sqrt{1+2s} \leq \varrho(s) \leq 1 + \sqrt{2s}$	$\frac{\Psi}{1+\sqrt{1+2\Psi}}$	$\frac{1}{2}$	$\frac{1}{21\left(1+\log(1+\sqrt{\Psi_0})\right)^2}$
7	$1 + s \leq \varrho(s) \leq 1 + \sqrt{s^2 + \frac{q}{q-1}s}$	$\frac{1}{2}\left(1 - \frac{1}{(\Psi+1)^q}\right)$	1	$\frac{1}{64\,q}$

Table 4: Results of Step 3, 4 and 5.

and $\theta = \Theta(\frac{1}{\sqrt{n}})$ we obtain the iteration bound for small-update methods. The resulting iteration bounds are summarized in the 3rd and 5th column of Table 6.

Small-update methods based on the seven kernel functions all have the same complexity as the small-update method based on the logarithmic barrier function, namely $O\left(\sqrt{n}\log\frac{n}{\varepsilon}\right)$; as is well known this is up till now the best iteration bound for methods solving LO problems. It must

i	$\Psi_0 \leq$	$\frac{K}{\theta} \leq$
1	$\frac{\left(\theta\sqrt{n}+\sqrt{2\tau}\right)^2}{1-\theta}$	$19\frac{\left(\theta\sqrt{n}+\sqrt{2\tau}\right)^2}{\theta(1-\theta)}$
2	$\frac{1}{1-\theta}\left(\sqrt{\tau}+\frac{n\theta}{\sqrt{2n}}\right)^2$	$\frac{65}{\theta(1-\theta)^{\frac{2}{3}}}\left(\sqrt{\tau}+\frac{\theta\sqrt{n}}{\sqrt{2}}\right)^{\frac{4}{3}}$
3	$\frac{q+1}{2}\frac{\left(\theta\sqrt{n}+\sqrt{2\tau}\right)^2}{1-\theta}$	$\frac{112\,q(q+1)}{\theta}\left(\frac{\left(\theta\sqrt{n}+\sqrt{2\tau}\right)^2}{2(1-\theta)}\right)^{\frac{q+1}{2q}}$
4	$\frac{\left(\theta\sqrt{n}+\sqrt{2\tau}\right)^2}{1-\theta}$	$\frac{106\,q}{\theta}\left(\frac{\left(\theta\sqrt{n}+\sqrt{2\tau}\right)^2}{1-\theta}\right)^{\frac{q+1}{2q}}$
5	$2\frac{\left(\theta\sqrt{n}+\sqrt{2\tau}\right)^2}{1-\theta}$	$38\sqrt{2}\left(1+\log\left(1+\sqrt{2}\,\frac{\theta\sqrt{n}+\sqrt{2\tau}}{\sqrt{1-\theta}}\right)\right)^2\frac{\theta\sqrt{n}+\sqrt{2\tau}}{\theta\sqrt{1-\theta}}$
6	$\frac{\left(\theta\sqrt{n}+\sqrt{2\tau}\right)^2}{1-\theta}$	$42\left(1+\log\left(1+\frac{\theta\sqrt{n}+\sqrt{2\tau}}{\sqrt{1-\theta}}\right)\right)^2\frac{\theta\sqrt{n}+\sqrt{2\tau}}{\theta\sqrt{1-\theta}}$
7	$\frac{q\left(\theta\sqrt{n}+\sqrt{\frac{\tau^2}{n}+2\tau}\right)^2}{2(1-\theta)}$	$32\frac{q^2\left(\theta\sqrt{n}+\sqrt{\frac{\tau^2}{n}+2\tau}\right)^2}{\theta(1-\theta)}$

Table 5: Results of Step 6 and 7.

i	kernel function $\psi_i(t)$	small-update	ref.	large-update	ref.
1	$\frac{t^2-1}{2}-\log t$	$O\left(\sqrt{n}\right)\,\log\frac{n}{\varepsilon}$	e.g., [14]	$O\left(n\right)\,\log\frac{n}{\varepsilon}$	e.g., [14]
2	$\frac{1}{2}\left(t-\frac{1}{t}\right)^2$	$O\left(\sqrt{n}\right)\,\log\frac{n}{\varepsilon}$	[7]	$O\left(n^{\frac{2}{3}}\right)\,\log\frac{n}{\varepsilon}$	[7]
3	$\frac{t^2-1}{2}+\frac{t^{1-q}-1}{q-1}$	$O\left(q^2\sqrt{n}\right)\,\log\frac{n}{\varepsilon}$	[8]	$O\left(qn^{\frac{q+1}{2q}}\right)\,\log\frac{n}{\varepsilon}$	[8, 12]
4	$\frac{t^2-1}{2}+\frac{t^{1-q}-1}{q(q-1)}-\frac{q-1}{q}(t-1)$	$O\left(q\sqrt{n}\right)\,\log\frac{n}{\varepsilon}$	[11]	$O\left(qn^{\frac{q+1}{2q}}\right)\,\log\frac{n}{\varepsilon}$	[11, 12]
5	$\frac{t^2-1}{2}+\frac{e^{\frac{1}{t}}-e}{e}$	$O\left(\sqrt{n}\right)\,\log\frac{n}{\varepsilon}$	new	$O\left(\sqrt{n}\,\log^2 n\right)\,\log\frac{n}{\varepsilon}$	new
6	$\frac{t^2-1}{2}-\int_1^t e^{\frac{1}{\xi}-1}d\xi$	$O\left(\sqrt{n}\right)\,\log\frac{n}{\varepsilon}$	new	$O\left(\sqrt{n}\,\log^2 n\right)\,\log\frac{n}{\varepsilon}$	new
7	$t-1+\frac{t^{1-q}-1}{q-1}$	$O\left(q^2\sqrt{n}\right)\,\log\frac{n}{\varepsilon}$	new	$O\left(qn\right)\,\log\frac{n}{\varepsilon}$	[4]

Table 6: Complexity results for large- and small-update methods.

be noted that where appropriate (i.e. for $i \in \{3,4,7\}$) one has to take $q = O(1)$ to achieve this bound; moreover the derivation of the small-update bound for $i = 7$ is valid only for $q \geq 2$.

Now let us consider the bounds for large-update methods in Table 6. It should be mentioned that the large-update bound in this table for $i = 3$ is based on estimates for Ψ_0 and $\frac{K}{\theta}$ that differ from the estimates in Table 5; for this case we used that

$$\Psi_0 \leq \frac{\theta n + 2\tau + 2\sqrt{2\tau n}}{2\,(1-\theta)}, \qquad \frac{K}{\theta} \leq \frac{112\,q}{\theta} \left(\frac{\theta n + 2\tau + 2\sqrt{2\tau n}}{2\,(1-\theta)} \right)^{\frac{q+1}{2q}}.$$

One may easily verify that the best iteration bound for large-update methods is obtained for $i \in \{3,4\}$ by taking $q = \frac{1}{2}\log n$. This gives the iteration bound $O\left(\sqrt{n}\,(\log n)\log\frac{n}{\varepsilon}\right)$, which is currently the best known bound for large-update methods.

7 Concluding Remarks

This paper was inspired by recent work on so-called self-regular barrier functions for primal-dual interior-point methods (IPMs) for linear optimization [11, 12]. Each such barrier function is determined by its (univariate) self-regular kernel function. We introduced 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. It has become clear that inverse functions of suitable restrictions of the kernel function and its first derivative play a key role in the behavior of the corresponding IPMs. Moreover, due to the new estimates the analysis of an IPM based on a given kernel function is greatly simplified and uses tools only from a first year calculus course.

Future research might focus on finding a kernel function for which the complexity of large-update methods is equal to (or even better than) $O(\sqrt{n}\log\frac{n}{\varepsilon})$, or show that such a kernel function does not exist. In this respect it might be of interest to show that the new class of kernel functions contains many functions not considered so far. For example,

$$\psi(t) = \frac{t^2-1}{2} + \frac{6}{\pi}\tan\frac{\pi(1-t)}{4t+2}$$

also is a member of the new class.

Also the extensions to Semidefinite Optimization and Second Order Cone Optimization deserve to be investigated. At present no computational results exist for the methods presented in this paper. This will be another issue for future research.

Acknowledgement

The authors kindly acknowledge the help of the associate editor and two anonymous referees in improving the readability of the paper.

References

[1] E.D. Andersen, J. Gondzio, Cs. Mészáros, and X. Xu. Implementation of interior point methods for large scale linear programming. In T. Terlaky, editor, *Interior Point Methods of Mathematical Programming*, pages 189–252. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[2] Y.Q. Bai, M. El ghami, and C. Roos. A primal-dual interior-point algorithm for linear optimization based on a new proximity function. *Optimization Methods & Software*, 17(6):985–1008 (electronic), 2002.

[3] Y.Q. Bai, M. El ghami, and C. Roos. A new efficient large-update primal-dual interior-point method based on a finite barrier. *SIAM J. Optim.*, 13(3):766–782 (electronic), 2003.

[4] Y.Q. 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.*

[5] N.K. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[6] N. Megiddo. Pathways to the optimal set in linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming: Interior Point and Related Methods*, pages 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.

[7] J. Peng, C. Roos, and T. Terlaky. New complexity analysis of the primal-dual Newton method for linear optimization. *Ann. Oper Res.*, 99:23–39 (2001), 2000. Applied mathematical programming and modeling, IV (Limassol, 1998).

[8] J. Peng, C. Roos, and T. Terlaky. A new and efficient large-update interior-point method for linear optimization. *Journal of Computational Technologies*, 6(4):61–80, 2001. ISSN 1560-7534.

[9] 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*, 143(2):234–256, 2002.

[10] J. Peng, C. Roos, and T. Terlaky. Primal-dual interior-point methods for second-order conic optimization based on self-regular proximities. *SIAM J. Optim.*, 13(1):179–203 (electronic), 2002.

[11] J. Peng, C. Roos, and T. Terlaky. Self-regular functions and new search directions for linear and semidefinite optimization. *Mathematical Programming*, 93:129–171, 2002.

[12] J. Peng, C. Roos, and T. Terlaky. *Self-Regularity. A New Paradigm for Primal-Dual Interior-Point Algorithms.* Princeton University Press, 2002.

[13] J. Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization*, volume 1 of *MPS/SIAM Series on Optimization.* SIAM, Philadelphia, USA, 2001. ISBN 0-89871-502-4.

[14] C. Roos, T. Terlaky, and J.-Ph. Vial. *Theory and Algorithms for Linear Optimization. An Interior-Point Approach.* John Wiley & Sons, Chichester, UK, 1997.

[15] 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*, volume 84 of *Lecture Notes in Control and Information Sciences*, pages 866–876. Springer Verlag, Berlin, West–Germany, 1986.

[16] S. J. Wright. *Primal-Dual Interior-Point Methods.* SIAM, Philadelphia, USA, 1997.

[17] Y. Ye. *Interior Point Algorithms, Theory and Analysis.* John Wiley & Sons, Chichester, UK, 1997.

Appendix

A 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 (Lemma 2.1 in [9]) *If $\alpha \in [0,1]$, then*

$$(1+t)^\alpha \leq 1 + \alpha t, \quad \forall\; t \geq -1. \tag{38}$$

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.2 (Proposition 2.2 in [9]) *Let $t_0, t_1, \cdots, t_K$ be a sequence of positive numbers such that*

$$t_{k+1} \leq t_k - \kappa t_k^{1-\gamma}, \qquad k = 0, 1, \cdots, K-1, \tag{39}$$

where $\kappa > 0$ and $0 < \gamma \leq 1$. Then $K \leq \left\lfloor \frac{t_0^\gamma}{\kappa\gamma} \right\rfloor$.

Proof: Using (39), 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 (38). 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.3 (Lemma 3.12 in [11]) *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. □