Interior-point methods based on kernel functions

Kees Roos
e-mail: C.Roos@ewi.tudelft.nl
URL: http://www.isa.ewi.tudelft.nl/~roos

University of Isfahan
Isfahan, Iran
January 31, A.D. 2005
Outline

★ Some historical remarks
★ The self-dual linear optimization (LO) problem $(SP)$
★ Central path and optimal partition
★ Algorithm with full Newton steps
★ The scaled Newton direction
★ New search directions from kernel functions
★ Generic Large-update Primal-Dual Path-following Method
★ Complexity results
★ Conditions on the kernel function
★ Some fundamental lemmas
★ Concluding remarks
★ References
Milestones in the history of linear optimization

Standard linear optimization problem

\[(P) \quad \min \left\{ c^T x : Ax = b, \ x \geq 0 \right\} \ , \ A \text{ is } m \times n \]

1947: Simplex Method (Dantzig)
1955: Logarithmic Barrier Method (Frisch)
1967: Affine Scaling Method (Dikin)
1967: Center Method (Huard)
1968: Barrier Methods (Fiacco, McCormick)
1972: Exponential example (Klee and Minty)
1979: Ellipsoid Method (Khachiyan)
1984: Projective Method (Karmarkar)  \Rightarrow \text{Interior Point Methods for}
                                Convex Optimization (1989)
                                Semidefinite Optimization (1994)
                                Second Order Cone Optimization (1994)
                                Discrete Optimization (1994)
Some keyplayers

Farkas Gyula (1847–1930)
The self-dual linear optimization (LO) problem \((SP)\)

\[
(SP) \quad \min \left\{ q^T x : s(x) = Mx + q \geq 0, \quad x \geq 0 \right\}
\]

with \(M^T = -M\) and \(q \geq 0\).

- Every LO problem \((P)\) has a dual problem \((D)\) associated to it.
- The dual problem of \((SP)\) is identical to \((SP)\), i.e., \((SP)\) is self-dual.
- Every problem \((P)\) and its dual \((D)\) can be embedded in a problem of the form \((SP)\).
- We may assume w.l.o.g. that \(e = Me + q\), i.e., \(e\) is feasible and \(s(e) = e\).
- A solution \(x\) of \((SP)\) is strictly complementary if \(x + s(x) > 0\).
- Solutions of \((P)\) and \((D)\) follow from a strictly complementary solution of \((SP)\).
- Well known solvers like SeDuMi and MOSEK are based on this approach.
Some simple properties of $(SP)$

$$(SP) \quad \min \{ q^T x : s(x) = Mx + q \geq 0, \ x \geq 0 \}$$

$M$ is $n \times n$ and $M^T = -M$; $q \geq 0$.

**Feasible set:** $\mathcal{P} := \{ x : x \geq 0, s(x) \geq 0 \}$

**Lemma 1** $x = 0$ is feasible and optimal.

**Lemma 2** (Orthogonality property) For every $x \in \mathbb{R}^n$ one has $x^T M x = 0$.

Proof: $x^T M x = (x^T M x)^T = x^T M^T x = -x^T M x \Rightarrow 2x^T M x = 0 \Rightarrow x^T M x = 0$.

**Objective value:** $q^T x = x^T (s(x) - Mx) = x^T s(x)$.

**Optimal set:** $\mathcal{P}^* := \{ x : x \geq 0, s(x) \geq 0, xs(x) = 0 \}$.

**Lemma 3** Let $x, y \in \mathcal{P}$. Then $x$ and $y$ are optimal if and only if $xs(y) = ys(x) = 0$.

Proof: $(x - y)^T M (x - y) = 0 \Rightarrow (x - y)^T (s(x) - s(y)) = 0$. Hence $x^T s(y) + y^T s(x) = x^T s(x) + y^T s(y)$ and this vanishes if and only if $x$ and $y$ are optimal.
Central path and optimal partition

**Theorem 1** For each $\mu > 0$ there exist a unique $x \in \mathcal{P}$ such that $xs(x) = \mu e$.

The solution is denoted as $x(\mu)$ and $s(\mu) := s(x(\mu))$.

**Central path**: $\{x(\mu) : \mu > 0\}$.

Note that

$$q^T x(\mu) = x(\mu)^T s(\mu) = n\mu.$$

We introduce index sets $B$ and $N$ as follows:

$$B := \{i : x_i > 0 \text{ for some } x \in \mathcal{P}^*\}$$

$$N := \{i : s_i(x) > 0 \text{ for some } x \in \mathcal{P}^*\}$$

$$T := \{i : x_i = s_i(x) = 0 \text{ for all } x \in \mathcal{P}^*\}.$$

$B$ and $N$ are disjoint by Lemma 2.

**Theorem 2** The set $T$ is empty and $B \cap N = \emptyset$. Moreover, if $\mu \downarrow 0$, then the limit $x^*$ of $x(\mu)$ exists and is strictly complementary: $x^*_B > 0$ and $s_N(x^*) > 0$. 
Definition of the Newton step

Suppose we are given a positive feasible pair \((x, s)\), with \(s = s(x)\) and \(\mu > 0\), and that we want to find the \(\mu\)-center \(x(\mu)\), i.e., we want to solve

\[
Mx + q = s > 0, \quad x > 0
\]

\[
x s = \mu e.
\]

Thus we need displacements \(\Delta x, \Delta s\) such that \(x + \Delta x, s + \Delta s\) satisfy:

\[
M(x + \Delta x) + q = s + \Delta s > 0, \quad x + \Delta x > 0
\]

\[
(x + \Delta x)(s + \Delta s) = \mu e.
\]

Neglecting for the moment the inequality constraints, this system can be rewritten as:

\[
M \Delta x = \Delta s,
\]

\[
s \Delta x + x \Delta s + \Delta x \Delta s = \mu e - xs.
\]

Neglecting the quadratic term \(\Delta x \Delta s\) we obtain the linear system

\[
M \Delta x = \Delta s,
\]

\[
s \Delta x + x \Delta s = \mu e - xs.
\]

This system uniquely defines the Newton direction \(\Delta x\) at \(x\) and the corresponding \(\Delta s\).
Algorithm with full Newton steps

Input:
An accuracy parameter $\epsilon > 0$;  
a barrier update parameter $\theta$, $0 < \theta < 1$.

\begin{align*}
\begin{array}{l}
x := e; \ s := e; \ \mu := 1; \\
\text{while } n\mu \geq \epsilon \text{ do} \\
\quad \begin{array}{l}
\mu := (1 - \theta)\mu; \\
x := x + \Delta x;
\end{array}
\end{array}
\end{align*}

Theorem 3 If $\theta = 1/(2\sqrt{n})$, then the algorithm with full Newton steps requires at most 
\[
\left\lceil 2\sqrt{n} \log \frac{n}{\epsilon} \right\rceil
\] 
iterations. The output is a feasible $x > 0$ such that $q^T x \leq \epsilon$.

Theorem 4 Taking $\epsilon$ small enough, the solution provided by the algorithm can be rounded to a strictly complementary solution of $(SP)$. The required number of iterations is not more than $5\sqrt{n} L$, where $L$ is the length of a binary encoding of $(SP)$. 
The scaled Newton direction

With \( s = s(x) \), we define for each \( x \in \text{int } \mathcal{P} \),

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

The classical (Newton) search direction \( \Delta x \) for primal-dual interior-point methods is obtained by solving

\[
  M \Delta x - \Delta s = 0, \quad M \Delta x + \Delta s = \mu e - xs. \quad \iff \quad \bar{M} d_x - d_s = 0, \quad \bar{M} d_x + d_s = v^{-1} - v,
\]

where

\[
  \bar{M} = V S^{-1} MV^{-1} X, \quad V = \text{diag } (v), \quad X = \text{diag } (x), \quad S = \text{diag } (s).
\]

The second equation in these systems is called the centering equation. The right hand sides vanish if and only if

\[
  v = e \quad \iff \quad xs = \mu e \quad \iff \quad x = x(\mu),
\]

and only then we have

\[
  \Delta x = \Delta s = 0 \quad \text{and} \quad d_x = d_s = 0.
\]
New search directions from kernel functions

A twice differentiable function $\psi : (0, \infty) \rightarrow [0, \infty)$ is called a kernel function if
\[
\psi(1) = 0, \quad \psi'(1) = 0, \quad \psi''(t) > 0, \quad \forall t > 0.
\]

Due to $\psi(1) = \psi'(1) = 0$ any kernel function is determined by its second derivative:
\[
\psi(t) = \int_1^t \int_1^{\xi} \psi''(\zeta) \, d\zeta \, d\xi.
\]

We extend the definition of $\psi$ to positive vectors $v \in \mathbb{R}_{++}^n$ by defining
\[
\psi(v) := (\psi(v_1), \ldots, \psi(v_n)).
\]

For any kernel function $\psi(t)$, the function
\[
\Psi(v) := e^T \psi(v) = \sum_{i=1}^{n} \psi(v_i)
\]
is a strictly convex function of $v > 0$ which is minimal at $v = e$ and with $\Psi(e) = 0$.

We replace the centering equation in the scaled system by
\[
d_x + d_s = -\nabla \Psi(v) = -\psi'(v).
\]

We have
\[
v = e \iff d_x = d_s = 0.
\]

If $\psi(t) = (t^2 - 1)/2 - \log t$, this yields the Newton direction, since $-\psi'(t) = t^{-1} - t$. 
Further motivation

The Hessian of $\Psi(v)$ at $e$ is a positive multiple of the identity matrix: $\nabla^2 \Psi(e) = \rho I$, for some positive $\rho$. This implies that

$$-\frac{1}{\rho} \nabla \Psi(v)$$

is a good approximation, at least in the neighborhood of $e$, of the Newton direction and hence the steepest descent direction is an excellent search direction for minimizing $\Psi(v)$. For (small) displacements $\Delta x$ and $\Delta s$, in the $x$- and $s$-space, respectively, let $\Delta v$ be the resulting change in $v$. Then we have

$$\Delta v = \sqrt{\frac{(x + \Delta x)(s + \Delta s)}{\mu}} - \sqrt{\frac{xs}{\mu}} \approx \frac{x\Delta s + s\Delta x}{2\sqrt{\mu xs}} = \frac{x\Delta s + s\Delta x}{2\mu v}.$$

Hence, since we want this to be about the steepest descent direction, we arrive at the equation

$$\frac{x\Delta s + s\Delta x}{2\mu v} = -\frac{1}{\rho} \nabla \Psi(v).$$

Due to the definition of $v$, and up to the positive factor $2/\rho$ this can be rewritten as

$$s\Delta x + x\Delta s = -\mu v \nabla \Psi(v).$$

Thus, adding the conditions for maintaining primal and dual feasibility after a step we find the following linear system for the search directions $\Delta x$, $\Delta s$ and $\Delta y$:

$$A \Delta x = 0,$$
$$A^T \Delta y + \Delta s = 0,$$
$$s\Delta x + x\Delta s = -\mu v \nabla \Psi(v) = -\mu v \psi'(v).$$

The last equation is equivalent to $d_x + d_s = -\psi'(v)$. 

Optimization Group
New promising direction?

The equation

\[ d_x + d_s = -\nabla \Psi(v) = -\psi'(v) \]

expresses that the search direction is the steepest descent direction for the barrier function \( \Psi(v) \).

One might wonder if we should not replace the steepest descent direction by the Newton direction and use instead

\[ d_x + d_s = -\left( \nabla^2 \Psi(v) \right)^{-1} \nabla \Psi(v) = -\frac{\psi'(v)}{\psi''(v)}. \]

For the logarithmic barrier function this would give

\[ d_x + d_s = -\frac{v - v^{-1}}{e + v^{-2}} = \frac{e - v^2}{e + v^2} v. \]

This direction may deserve further investigation.
Graphical illustration of some kernel functions

\[
\frac{t^2-1}{2} - \log t
\]

\[
\frac{t^2-1}{2} + \frac{t^{1-q-1}}{q-1}, \quad q = 2
\]

\[
\frac{t^2-1}{2} + \frac{t^{1-q-1}}{q-1}, \quad q = 4
\]

\[
\frac{t^2-1}{2} + \frac{1}{e^{\frac{t}{\tau}}-e}
\]
Generic Large-update Primal-Dual Path-following Method

**Input:**
- accuracy parameter $\epsilon > 0$;
- threshold parameter $\tau > 0$;
- update parameter $\theta$, $0 < \theta < 1$.

begin

$x := e; s := e; \mu := 1$

while $q^T x \geq \epsilon$ do (outer iteration)

$\mu := (1 - \theta) \mu$

while $\Psi(v) \geq \tau$ do (inner iteration)

$x := x + \alpha \Delta x$

$v = \sqrt{\frac{xs(x)}{\mu}}$

endwhile (inner iteration)

endwhile (outer iteration)

end

N.B. The step size $\alpha$ has to be chosen such that $e^T \psi(v)$ decreases a sufficient amount.
Primal-Dual Path-Following Method

One outer iteration.

Optimization Group
Aim of the research

The aim of our research is to find a kernel function that minimizes the iteration complexity of the algorithm. We distinguish between small-update and large-update methods, which have $\theta = O\left(\frac{1}{\sqrt{n}}\right)$ and $\theta = O(1)$ respectively.

Given a kernel function $\psi(t)$, in the analysis of the induced algorithm we have to face the following items:

- growth behavior (increase of $\Psi(v)$ due to a barrier parameter update);
- step size $\alpha$ and decrease of $\Psi(v)$ during an inner iteration;
- estimate of the number $K$ of inner iterations between two successive barrier parameter updates;
- (total) iteration complexity: $\frac{K}{\theta} \log \frac{n}{\epsilon}$.

Under four additional assumptions on the kernel function and its first three derivatives one can show that the iteration complexity can be expressed in the kernel function and its first and second derivatives.
Scheme for analyzing a kernel function

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

**Step 2:** Calculate the decrease of $\Psi(v)$ in terms of $\delta$: $f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''(\rho(2\delta))}.$

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

**Step 4:** Derive a lower bound for $\delta$ in terms of $\Psi(v)$ from $\delta(v) \geq \frac{1}{2}\psi' (\varrho (\Psi(v)))$.

**Step 5:** Using the results of step 2 and step 4 find positive constants $\kappa$ and $\gamma$, with $\gamma \in (0, 1]$ such that $f(\tilde{\alpha}) \leq -\kappa \Psi(v)^{1-\gamma}$.

**Step 6:** Calculate a global upper bound $\Psi_0$ for $\Psi(v)$ from $\Psi_0 \leq n\Psi \left( \varrho (\frac{\tau}{n}) \right)$.

**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 $\theta = \Theta(1)$ to obtain 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.
## Complexity results

<table>
<thead>
<tr>
<th>$i$</th>
<th>kernel function $\psi_i(t)$</th>
<th>small-update</th>
<th>large-update</th>
<th>ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\frac{t^2-1}{2} - \log t$</td>
<td>$O(\sqrt{n}) \log \frac{n}{\epsilon}$</td>
<td>$O(n) \log \frac{n}{\epsilon}$</td>
<td>RTV</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{2} (t - \frac{1}{t})^2$</td>
<td>$O(\sqrt{n}) \log \frac{n}{\epsilon}$</td>
<td>$O\left(n^{\frac{3}{2}}\right) \log \frac{n}{\epsilon}$</td>
<td>PRT</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q-1}$, $q &gt; 1$</td>
<td>$O(q^2 \sqrt{n}) \log \frac{n}{\epsilon}$</td>
<td>$O\left(qn^{\frac{q+1}{2q}}\right) \log \frac{n}{\epsilon}$</td>
<td>PRT</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q-1} - \frac{q-1}{q} (t - 1)$, $q &gt; 1$</td>
<td>$O(q \sqrt{n}) \log \frac{n}{\epsilon}$</td>
<td>$O\left(qn^{\frac{q+1}{2q}}\right) \log \frac{n}{\epsilon}$</td>
<td>PRT</td>
</tr>
<tr>
<td>5</td>
<td>$\frac{t^2-1}{2} + \frac{e^t - e}{e}$</td>
<td>$O(\sqrt{n}) \log \frac{n}{\epsilon}$</td>
<td>$O(\sqrt{n} \log^2 n) \log \frac{n}{\epsilon}$</td>
<td>BER</td>
</tr>
<tr>
<td>6</td>
<td>$\frac{t^2-1}{2} - \int_1^t e^{\frac{1}{\xi} - 1} d\xi$</td>
<td>$O(\sqrt{n}) \log \frac{n}{\epsilon}$</td>
<td>$O(\sqrt{n} \log^2 n) \log \frac{n}{\epsilon}$</td>
<td>BER</td>
</tr>
<tr>
<td>7</td>
<td>$t - 1 + \frac{t^{1-q}-1}{q-1}$, $q &gt; 1$</td>
<td>$O\left(q^2 \sqrt{n}\right) \log \frac{n}{\epsilon}$</td>
<td>$O\left(qn^{\frac{q+1}{2q}}\right) \log \frac{n}{\epsilon}$</td>
<td>BR</td>
</tr>
<tr>
<td>8</td>
<td>$\frac{t^2-1}{2} + q \left(e^{\frac{1}{\sqrt{n}} - 1} - 1\right)$, $q \geq 1$</td>
<td>$O\left(qn\right) \log \frac{n}{\epsilon}$</td>
<td>$O\left(\sqrt{n} \left(\log n\right)^{q+1}\right) \log \frac{n}{\epsilon}$</td>
<td>AMR</td>
</tr>
<tr>
<td>9</td>
<td>$\frac{t^2-1}{2} - \int_1^t e^{\frac{1}{\sqrt{\xi}} - 1} d\xi$, $q \geq 1$</td>
<td>$O\left(\sqrt{n} \left(\log n\right)^{q+1}\right) \log \frac{n}{\epsilon}$</td>
<td>$O\left(\sqrt{n} \left(\log n\right)^{q+1}\right) \log \frac{n}{\epsilon}$</td>
<td>AMR</td>
</tr>
<tr>
<td>10</td>
<td>$t - 1 - \log t$</td>
<td>$O(n) \log \frac{n}{\epsilon}$</td>
<td>$O(n) \log \frac{n}{\epsilon}$</td>
<td>AMR</td>
</tr>
</tbody>
</table>

In all cases the iteration bound for small-update methods is $O(\sqrt{n} \log \frac{n}{\epsilon})$.

The best 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 $O\left(\sqrt{n} \left(\log n\right)^{q+1}\right) \log \frac{n}{\epsilon}$, which is currently the best known bound for large-update methods.
Alternative directions (Zs. Darway, 2002)

Recall that in each inner iteration we want to find the $\mu$-center $x(\mu)$, i.e., we want to solve

$$Mx + q = s > 0, \ x > 0$$

$$xs = \mu e.$$

Let $\phi : (0, \infty) \to (0, \infty)$ be a function that has an inverse function $\phi^{-1}$. Then we may replace the second equation by

$$\phi \left( \frac{x s}{\mu} \right) = \phi(e).$$

Linearizing this equation we obtain

$$\phi \left( \frac{x s}{\mu} \right) + \frac{s}{\mu} \phi' \left( \frac{x s}{\mu} \right) \Delta x + \frac{x}{\mu} \phi' \left( \frac{x s}{\mu} \right) \Delta s = \phi(e),$$

which is equivalent to

$$d_x + d_s = -\frac{\phi(v^2) - \phi(e)}{v \phi'(v^2)}.$$  

**Theorem 5** (Zs. Darway, 2002) *If $\phi(t) = \sqrt{t}$, the algorithm with full steps requires

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

iterations. The output is a feasible $x > 0$ such that $q^T x \leq \epsilon$ ($\theta = 1/(2\sqrt{n})$).*
Darway’s directions are induced by kernel functions

Let \( \psi(t) \) be a kernel function such that

\[
\psi'(t) = \frac{\phi(t^2) - \phi(1)}{t \phi'(t^2)}.
\]

Then it is obvious that the two search directions coincide. Using \( \psi(1) = 0 \) it follows that

\[
\psi(t) = \int_{1}^{t} \frac{\phi(\xi^2) - \phi(1)}{\xi \phi'(\xi^2)} d\xi.
\]

Two cases are considered by Darway:

**Case 1**: \( \phi(t) = t \). Then \( \phi'(t) = 1 \), and hence

\[
\psi(t) = \int_{1}^{t} \frac{\xi^2 - 1}{\xi} d\xi = \frac{t^2 - 1}{2} - \log t,
\]

which is the kernel function for the classical Newton direction.

**Case 2**: \( \phi(t) = \sqrt{t} \). Then \( \phi'(t) = 1/(2\sqrt{t}) \), and hence

\[
\psi(t) = \int_{1}^{t} \frac{\xi - 1}{\xi^{1/2}} d\xi = 2 \int_{1}^{t} (\xi - 1) d\xi = (t - 1)^2.
\]

This is also a kernel function, and it is easy to derive the same complexity bound as Darway.
Our conditions on the kernels function

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

\[ t\psi''(t) + \psi'(t) > 0, \quad t < 1, \]  
\[ t\psi''(t) - \psi'(t) > 0, \quad t > 1, \]  
\[ \psi'''(t) < 0, \quad t > 0, \]  
\[ 2\psi'(t)^2 - \psi'(t)\psi'''(t) > 0, \quad t < 1, \]  
\[ \psi''(t)\psi'(\beta t) - \beta \psi'(t)\psi''(\beta t) > 0, \quad t > 1, \beta > 1. \]

Note that conditions (1.a) and (1.b) require that \( \psi(t) \) is three times differentiable.

Furthermore, condition (1.a) is obviously satisfied if \( t \geq 1 \), since then \( \psi'(t) \geq 0 \) and, similarly, condition (1.b) is satisfied if \( t \leq 1 \), since then \( \psi'(t) \leq 0 \). Also (1.c) is obviously satisfied if \( t \geq 1 \) since then \( \psi'(t) \geq 0 \), whereas \( \psi'''(t) < 0 \).

We conclude that conditions (1.a) and (1.b) are conditions on the **barrier behavior** of \( \psi(t) \).

On the other hand, condition (1.d) only deals with \( t \geq 1 \) and hence concerns the **growth behavior** of \( \psi(t) \). Condition (1.e) is technically more involved; we will discuss it later.
Some simple lemmas

Lemma 4 If the kernel function $\psi(t)$ satisfies (??), then

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

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

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 5 If the kernel function $\psi(t)$ satisfies (??), then

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

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

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 ??.
Lemma 6 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 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.
Theorem 6 Let $\varrho : [0, \infty) \rightarrow [1, \infty)$ be the inverse function of $\psi(t)$ for $t \geq 1$. Then we have for any positive vector $v$ and any $\beta \geq 1$:

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

Proof: First we consider the case where $\beta > 1$. We consider the following maximization problem:

$$\max_{v} \{ \Psi(\beta v) : \Psi(v) = z \},$$

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

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

where $\lambda$ 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 ?? 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 $i$. 

Optimization Group
Note that then (??) 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 (??) that $g(v_i) = \frac{\beta}{\lambda}$ for all $i$. One has

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

At this stage we use that $\psi(t)$ satisfies condition (??). 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

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

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

•
Lemma 7 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 ?? 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 ??, while assuming $-\psi'(t_1) < \psi'(t_2)$, we may write

$$\psi(t_2) > \frac{1}{2} (t_2 - 1) \psi'(t_2) > \frac{1}{2} (1 - t_1) \psi'(t_2) > -\frac{1}{2} (1 - t_1) \psi'(t_1)$$

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

This contradiction proves the lemma. •
Theorem 7  One has

$$\delta(v) := \frac{1}{2} \| \psi'(v) \| \geq \frac{1}{2} \psi' (\varrho (\Psi(v))).$$

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 $\omega > 0$, the problem

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

The first order optimality condition is

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

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

$$v = (t, \ldots, t, 1, \ldots, 1).$$
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 ??, 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, $\varrho$ denotes the inverse function of $\psi(t)$ for $t \geq 1$. Thus we obtain that

$$4\delta(v)^2 = k \left( \psi'(t) \right)^2, \quad t = \varrho \left( \frac{\omega}{k} \right). \quad (3)$$

The question is now which value of $k$ minimizes $\delta(v)^2$. To investigate this, we take the derivative with respect to $k$ of (3) extended to $k \in \mathbb{R}$. This gives

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

(4)
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 (??) gives

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

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 (??) that

$$\min_v \{\delta(v) : \Psi(v) = \omega\} = \frac{1}{2} \psi'(t) = \frac{1}{2} \psi' (\varphi(\omega)) = \frac{1}{2} \psi' (\varphi (\Psi(v))).$$

This completes the proof of the theorem.
Concluding remarks

- Up till now all polynomial-time interior-point methods for LO are based on the use of the logarithmic barrier function. Many authors (e.g., Hamala, Nazareth) tried to explain the ‘good’ behavior logarithmic barrier function in comparison with other barrier function (inverse barrier, entropy function, etc.). Most barriers in this talk are not logarithmic; they are also not necessarily self-regular.

- For small-update methods all kernel function considered so far yield the best possible iteration bound known today: $O \left( \sqrt{n} \log \frac{n}{\epsilon} \right)$. This is exactly the same bound as for the classical methods, based on the logarithmic barrier function.

- For large-update methods the bounds are usually better than for the logarithmic barrier function.

- The extension to SDO and SOCO deserves further investigation. Probably this is a purely technical, although not obvious matter.

- Computational experiments have been done by Peng and Terlaky, so far only for so-called self-regular functions.
Some references


