

Advances in Interior Point Methods

Tamás Terlaky

McMaster University

Department Computing and Software

Based on joint work with:

J. Peng, McMaster, K. Roos, T.U.D.

Optimization–2001 Aveiro, July 22-26, 2001

AdvOB



- **History of Linear Optimization**
- **LO: Fundamentals of Algorithms**
 - The primal and dual LO problem
 - Duality Theorems
 - Optimality and algorithms
- **Interior Point Algorithms**
 - Central path
 - Primal-dual path following IPMs
 - New Self-Regularity based IPMs
- **IPMs for SOCO**
 - Algebra of Second Order Cones
 - SOCO and the Central path
 - New Self-Regularity based IPMs
- **Summary**

Notation: Let $x, s \in \mathbb{R}^n$ then
 $x^T s$ denotes the scalar product;
 xs denotes the componentwise product.



Standard form LO model

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

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

A is an $m \times n$ matrix with $\text{rank}(A) = m$.

18xx: Gauss elimination

1827: Fourier-Motzkin(1936) elimination

1894: Farkas' Lemma

1947: **Simplex Method (Dantzig)**

1955: Logarithmic Barrier Method (Frisch)

1956: Strictly complementary solution (GT)

1968: Barrier Methods (Fiacco, McCormick)

1972: **Exponential example (Klee and Minty)**

1979: **Ellipsoid Method (Kachijan)**

1984: Criss-cross methods

1984: **Projective Method (Karmarkar)**

1990: Self-concordancy (Nesterov-Nemrovski)

2000: Self-regularity (Peng, Roos, Terlaky)



Standard form LO model, Fundamentals

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

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

A is an $m \times n$ matrix with $\text{rank}(A) = m$.

Proposition 1 (Weak duality) *If $x \in \mathbb{R}^n$ is primal, $y \in \mathbb{R}^m$ is dual feasible then*

$$c^T x \geq b^T y,$$

where the equality is satisfied iff $x^T s = 0$.

Corollary 1 (Complementarity) *If $x \in \mathbb{R}^n$ is primal, $y \in \mathbb{R}^m$ is dual feasible and $x^T s = 0$ then x and y are primal and dual optimal, respectively.*

Optimality: Optimal solutions can be given as the set of solutions of the system:

$$Ax = b$$

$$x \geq 0$$

$$A^T y + s = c$$

$$s \geq 0.$$

$$x_i s_i = 0 \quad \forall i$$



Standard form LO model: Optimality Conditions

$$Ax = b$$

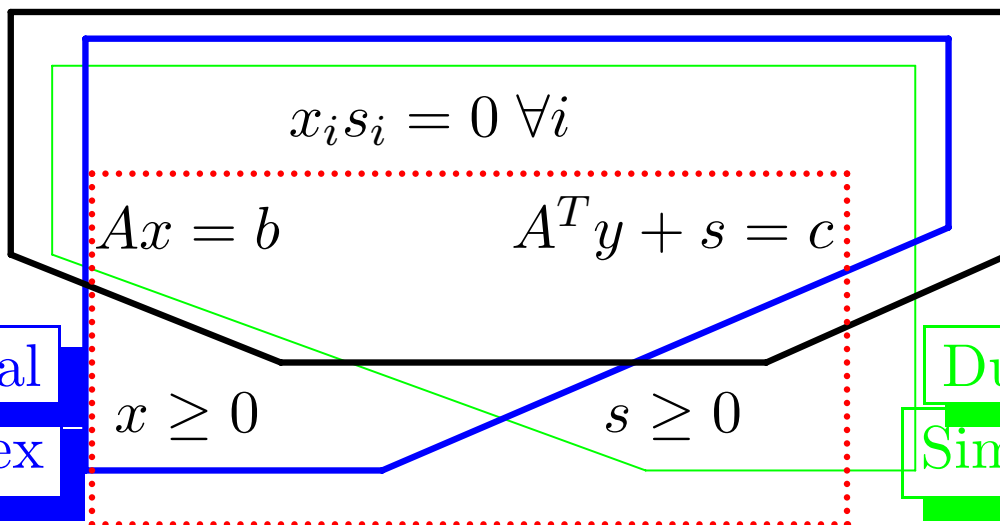
$$A^T y + s = c$$

$$xs = 0$$

$$x \geq 0$$

$$s \geq 0.$$

Criss-Cross Method



Primal
Simplex

Dual
Simplex

Interior Point Methods

Algorithmic concepts



We want to solve:

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

Here xs denotes the coordinatewise product.

IPMs use interior solutions,

i.e. (x, s, y) such that

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

We need the interior point condition (IPC):

there exists (x^0, s^0, y^0) such that

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

Known that w.l.o.g. the IPC can be assumed.



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

$x(\mu)$ the μ -center of (P)

$(y(\mu), s(\mu))$ is the μ -center of (D) .

The path $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$ is analytic.

The Newton step

First update $\mu := (1 - \theta)\mu$, for some $\theta \in (0, 1)$.

Then solve the Newton system

$$\begin{aligned} A\Delta x &= 0, \\ A^T \Delta y + \Delta s &= 0, \\ s\Delta x + x\Delta s &= \mu e - xs, \end{aligned}$$

and get a unique search direction $(\Delta x, \Delta s, \Delta y)$

Then take a step along the search direction,

such that the new triple (x, y, s)

is ‘close’ to the target $(x(\mu), y(\mu), s(\mu))$



Let $\delta(xs, \mu)$ be a proximity measure, e.g. let

$$\delta(xs, \mu) = \left\| \frac{xs}{\mu} - e \right\| \text{ or } \delta(xs, \mu) = \left\| \sqrt{\frac{xs}{\mu}} - \sqrt{\frac{\mu}{xs}} \right\|.$$

Input:

A proximity parameter $\tau > 0$;

an accuracy parameter $\epsilon > 0$;

a variable damping factor α ;

a barrier update parameter $0 < \theta < 1$;

(x^0, s^0) and $\mu^0 = 1$ such that $\delta(x^0 s^0, \mu^0) \leq \tau$.

begin

$x := x^0$; $s := s^0$; $\mu := \mu^0$;

while $n\mu \geq \epsilon$ **do**

begin

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

while $\delta(xs, \mu) \geq \tau$ **do**

Solve the Newton system to get $\Delta x, \Delta y, \Delta s$,

begin

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

$s := s + \alpha\Delta s$;

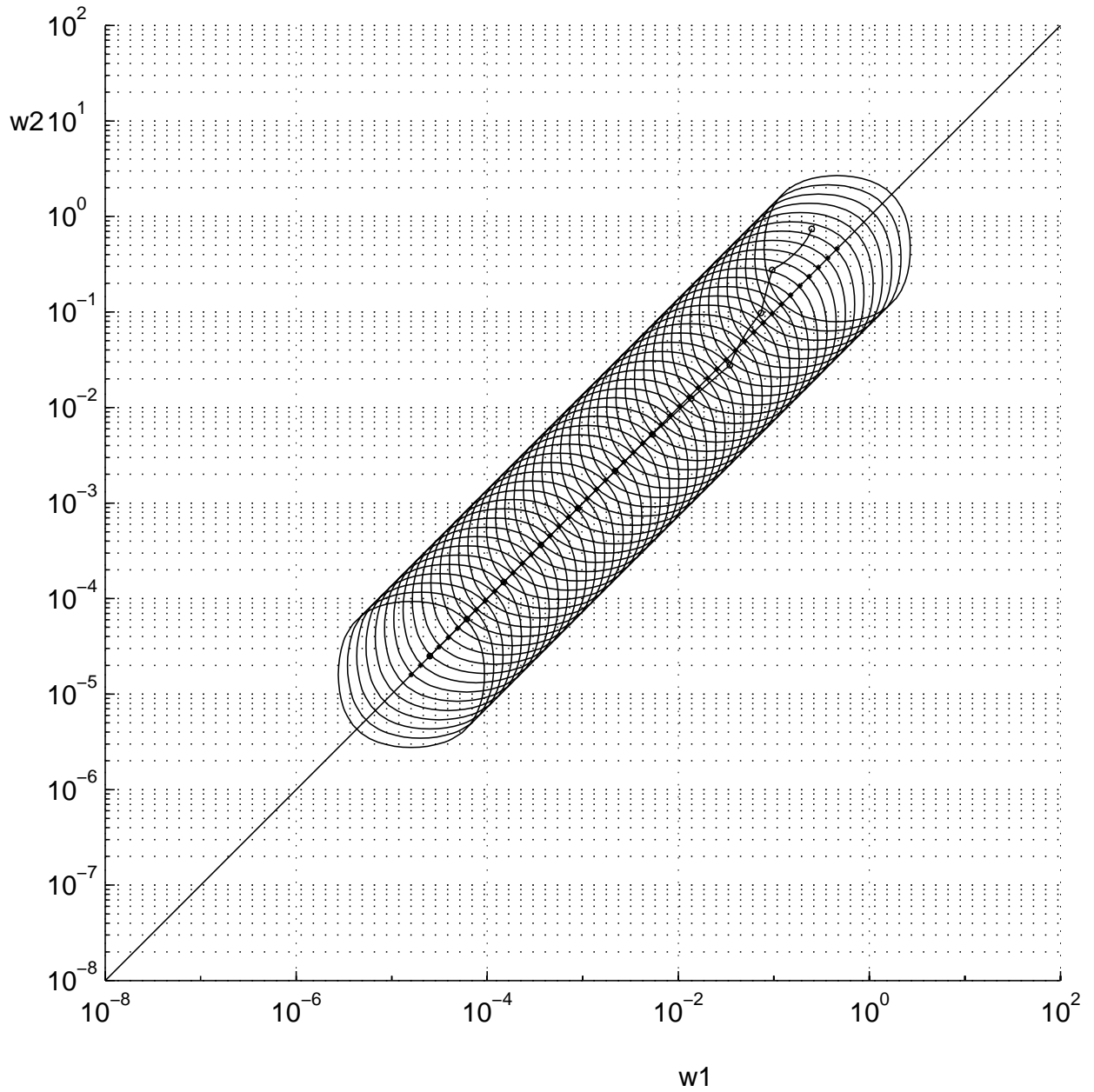
$y := y + \alpha\Delta y$

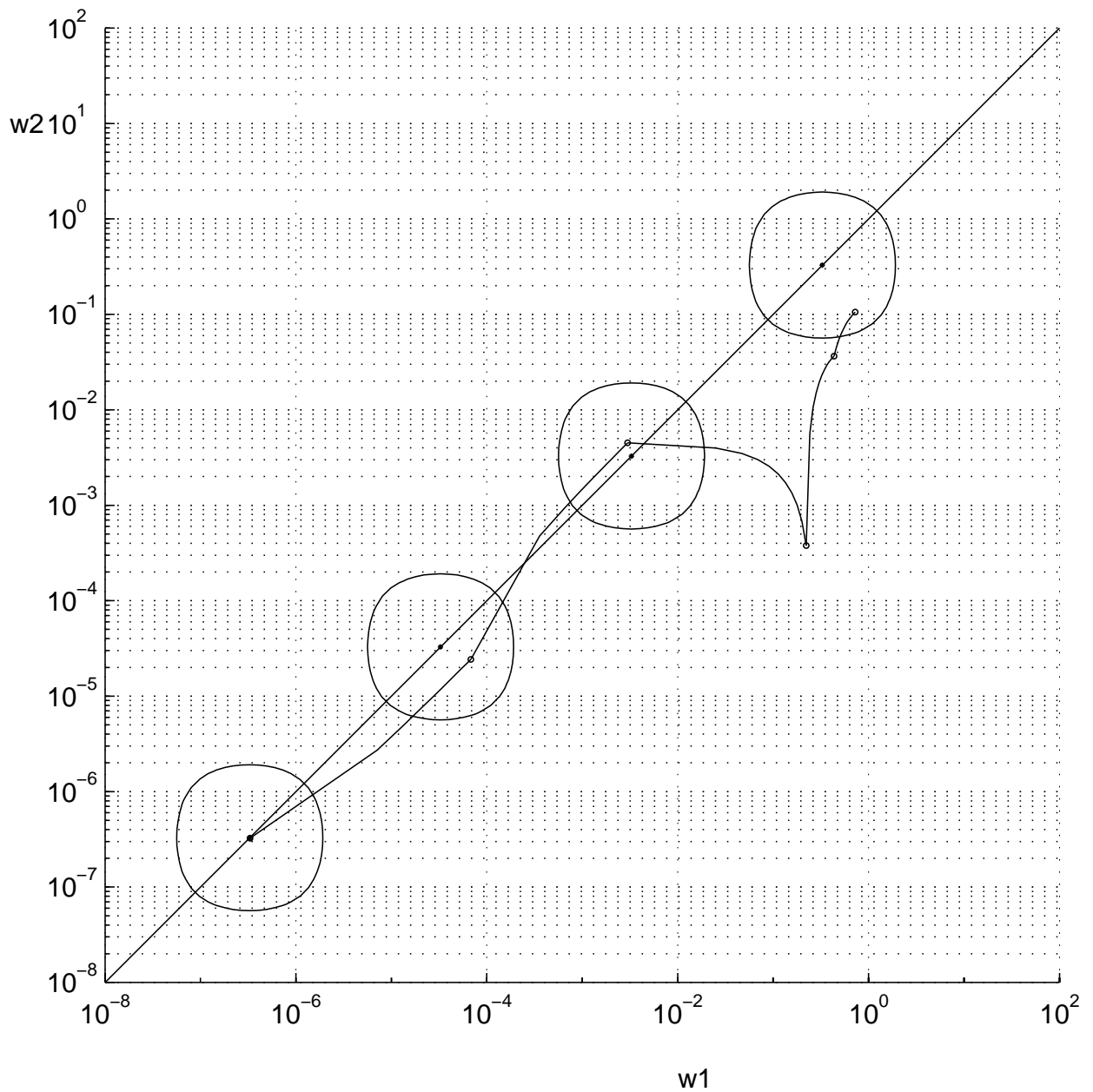
end

end

end

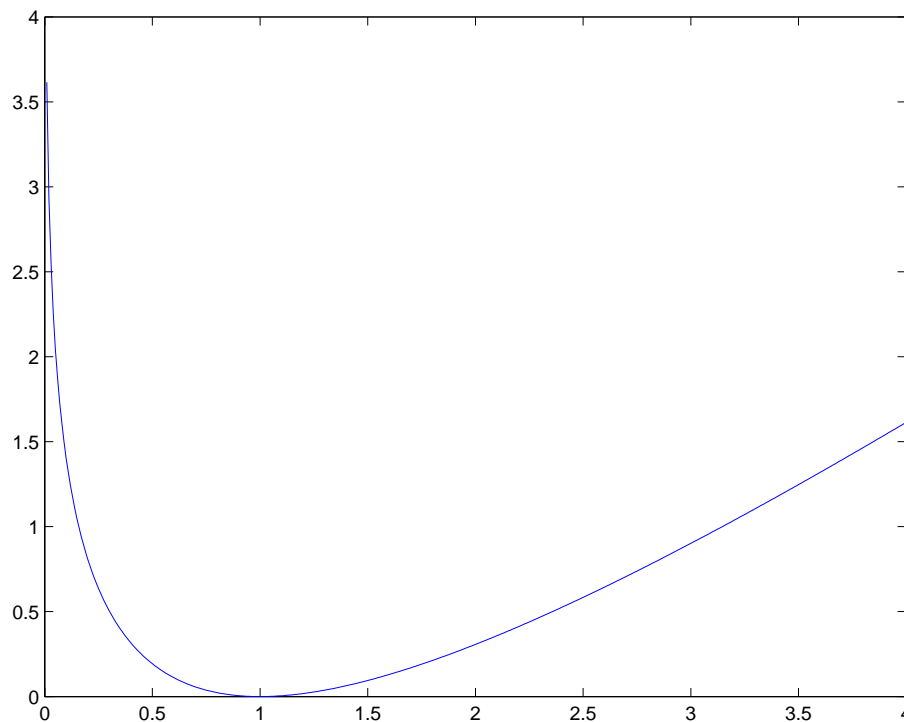






$$\begin{aligned} Ax &= b, \quad x \geq 0, \\ A^T y + s &= c, \quad s \geq 0, \\ xs &= \mu e. \end{aligned}$$

Impose the barrier on the
Central Path conditions: $\frac{x_i s_i}{\mu} = 1!$



A barrier/proximity $\psi(t)$ must satisfy:

$$\begin{aligned} \psi(1) &= 0; & \psi'(1) &= 0 \\ \lim_{t \downarrow 0} \psi(t) &= \infty; & \lim_{t \rightarrow \infty} \psi(t) &= \infty \end{aligned}$$

smooth and convex



The scaled Newton step

$$\begin{aligned}
 v &:= \sqrt{\frac{xs}{\mu}}, & v^{-1} &:= \sqrt{\frac{\mu}{xs}}, \\
 p_x &:= \frac{v\Delta x}{x}, & p_s &:= \frac{v\Delta s}{s}; \\
 \bar{x} &:= \frac{\Delta x}{x}, & \bar{s} &:= \frac{\Delta s}{s}.
 \end{aligned}$$

We can restate the centrality condition as $v = v^{-1} = e$ and rewrite

$$\begin{aligned}
 \bar{A}p_x &= 0, \\
 \bar{A}^T \Delta y + p_s &= 0, \\
 p_x + p_s &= v^{-1} - v.
 \end{aligned}$$

where $\bar{A} = AV^{-1}X$, $V = \text{diag}(v)$, $X = \text{diag}(x)$.

Proximities

$$\begin{aligned}
 \delta_K(xs, \mu) &:= \|v^2 - e\|, \\
 \delta(xs, \mu) &:= \|v - v^{-1}\|, \\
 \Phi(xs, \mu) &:= \sum_{i=1}^n \phi(v_i^2),
 \end{aligned}$$

where $\phi(t) = t - 1 - \log t$, $t > 0$, and

Small and Large update methods.



In 1999 Peng-Roos-Terlaky gave a slight modification of the standard Newton direction:

$$\begin{aligned}\bar{A}p_x &= 0, \\ \bar{A}^T \Delta y + p_s &= 0, \\ p_x + p_s &= v^{-q} - v,\end{aligned}$$

where $\bar{A} = AV^{-1}X$, $V = \text{diag}(v)$ and $q \geq 1$.

Observation: In the special case of $q = 3$, the RHS of the third term above represents the negative gradient of the proximity measure $\frac{1}{2}\delta^2$ in the v -space.

Complexity of large update methods: $O\left(n^{\frac{4}{3+q}} \log \frac{n}{\epsilon}\right)$ iterations where $q \in [1, 3]$

Observation: Also the RHS of the scaled Newton system is the negative gradient of the proximity $\frac{1}{2}\Phi$, thus the primal-dual Newton method is the steepest descent method for minimizing Φ .

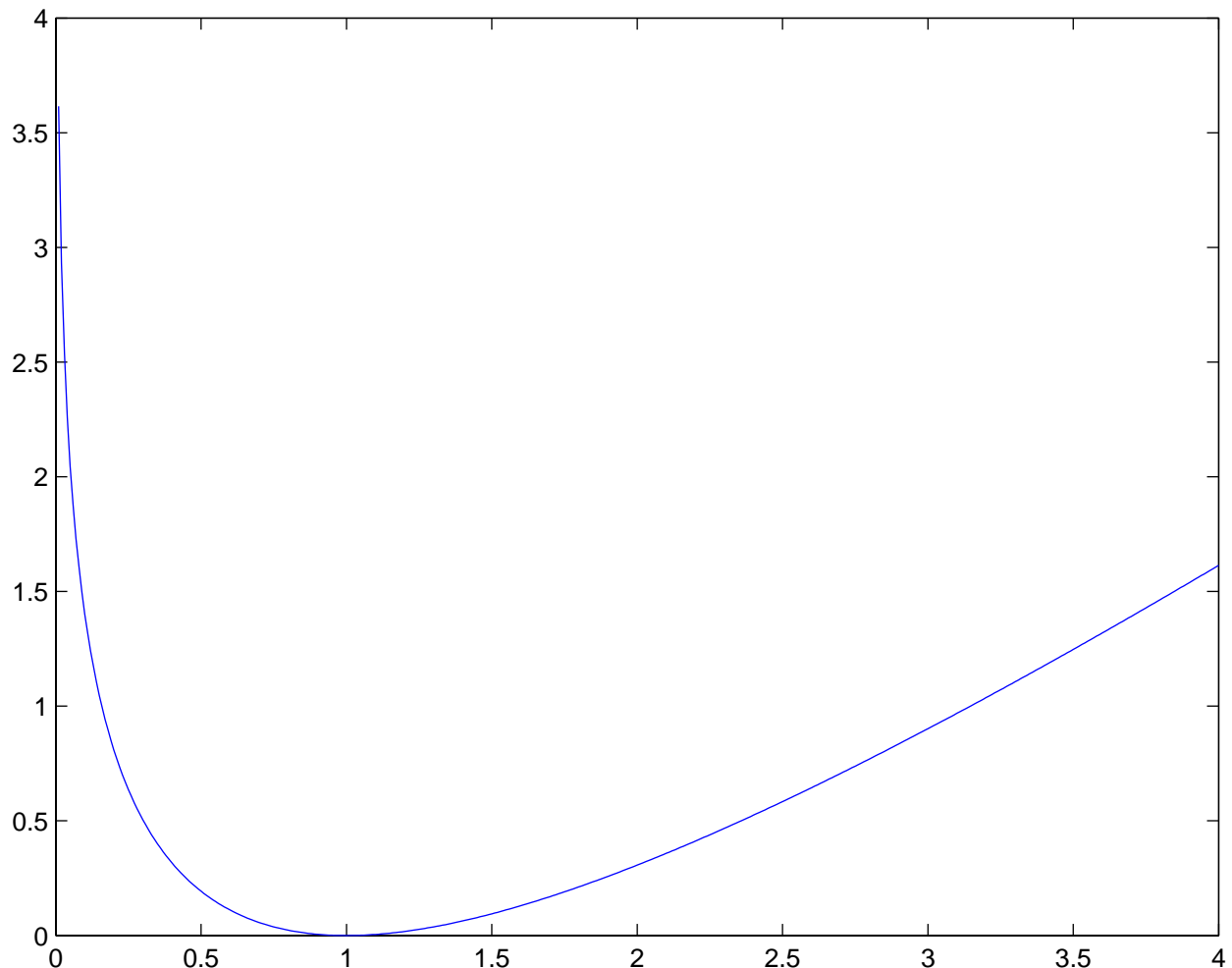
Complexity of large update methods:
 $O\left(n \log \frac{n}{\epsilon}\right)$ iterations.

Aim: Further improve the complexity of large update IPMs. Close the gap between theory and practice!



Impose the barrier on the Central Path

conditions: $\frac{x_i s_i}{\mu} = 1!$



A barrier/proximity $\psi(t)$ must satisfy:

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

$$\lim_{t \downarrow 0} \psi(t) = \infty; \quad \lim_{t \rightarrow \infty} \psi(t) = \infty$$

smooth and convex



Definition 1 A function $\psi(t) : (0, \infty) \rightarrow \mathfrak{R}$ is called *self-regular* if

C.1 $\psi(t)$ is strongly convex w.r.t. $t > 0$;
vanishes at its global minimal point $t = 1$,
i.e., $\psi(1) = \psi'(1) = 0$.

There exist positive constants $\nu_1, \nu_2 > 0$ and $p \geq 1, q \geq 1$ such that $\forall t \in (0, \infty)$

$$\nu_1(t^{p-1} + t^{-1-q}) \leq \psi''(t) \leq \nu_2(t^{p-1} + t^{-1-q})$$

C.2 For any $t_1, t_2 > 0$ and $0 \leq \alpha \leq 1$,

$$\psi(t_1^\alpha t_2^{(1-\alpha)}) \leq \alpha\psi(t_1) + (1-\alpha)\psi(t_2).$$

Terminology: Parameter q is called the **barrier degree** and p the **growth degree** of $\psi(t)$.

Proximity: $\Psi(xs, \mu) := \Psi(v) = \sum_{i=1}^n \psi(v_i)$

The proximity is *self-regular* if $\psi(t)$ is *self-regular*.
We also define $\sigma = \|\nabla\Psi(v)\|$.



Notation: For $p, q \geq 1$ denote

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

Clearly $\Upsilon''(t) = t^{p-1} + t^{-1-q}$, thus this function satisfies condition C.1 with $\nu_1 = \nu_2 = 1$.

For any function $\psi(t)$ satisfying C.1, it holds

$$\begin{aligned} \nu_1 \Upsilon''_{p,q}(t) &\leq \psi''(t) \leq \nu_2 \Upsilon''_{p,q}(t); \\ \nu_1 \Upsilon_{p,q}(t) &\leq \psi(t) \leq \nu_2 \Upsilon_{p,q}(t). \end{aligned}$$

It is also easy to see that

$$\left. \frac{t^{1-q} - 1}{q-1} \right|_{q=1} = \lim_{q \rightarrow 1} \frac{t^{1-q} - 1}{q-1} = -\log t.$$

Hence, in case of $q = 1$, the argument $\frac{t^{1-q}-1}{q-1}$ in the paper is endowed as $-\log t$. Using this notation, one can write

$$\Upsilon_{1,1}(t) := \frac{1}{2}t^2 - \frac{1}{2} - \log t$$

the well-known logarithmic barrier function.

The function $\Upsilon_{1,1}(t)$ satisfies our conditions C.1 and C.2 with $p = q = \nu_1 = \nu_2 = 1$.



Proximity: $\Psi(xs, \mu) := \Psi(v) = \sum_{i=1}^n \psi(v_i)$

Let $p_x, \Delta y, p_s$ denote the solution of the following modified Newton equation system for the scaled Newton system:

$$\begin{aligned}\bar{A}p_x &= 0, \\ \bar{A}^T \Delta y + p_s &= 0, \\ p_x + p_s &= -\nabla \Psi(v),\end{aligned}$$

or, $(\Delta x, \Delta y, \Delta s)$ the solution of its equivalent system

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

The last equation in the above system represents the Newton system for the equation

$$xs = -\mu v(\nabla \Psi(v) - v).$$

Make a damped Newton step:

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



Large Update Primal-Dual Algorithm for Linear Optimization

Input:

A proximity parameter $\tau > \nu_1^{-1}$;
an accuracy parameter $\epsilon > 0$;
a variable damping factor α ;
a fixed barrier update parameter θ , $0 < \theta < 1$;
 (x^0, s^0) and $\mu^0 = 1$ such that $\Psi(x^0, s^0, \mu^0) \leq \tau$.

begin

$x := x^0$; $s := s^0$; $\mu := \mu^0$;

while $n\mu \geq \epsilon$ **do**

begin

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

while $\Psi(x, s, \mu) \geq \tau$ **do**

Solve the system (1) and compute $\Delta x, \Delta y, \Delta s$,

begin

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

$s := s + \alpha\Delta s$;

$y := y + \alpha\Delta y$

end

end

end



- Bound the increase of the proximity after a barrier parameter update.
- Calculate how many updates are needed.
- Estimate the feasible step size.
- Find the step that minimizes the proximity.
- Estimate the decrease of the proximity.
- Bound the number of steps that are needed to come back close to the central path.
- Bound the total number of Newton steps.



Theorem 1 *If $\tau \geq \nu_1^{-1}$, the total number of iterations required by the primal-dual Newton algorithm is not more than*

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

In the special case where $\psi(t) = \Upsilon_{p,q}(t)$, the above bound (with $\nu_1 = \nu_2 = 1$) simplifies to

$$\left\lceil \frac{8q \max(3p + 2, 6q + 4)}{q + 1} (\psi_0(\theta, \tau, n))^{\frac{q+1}{2q}} \right\rceil \left\lceil \frac{1}{\theta} \log \frac{n}{\epsilon} \right\rceil$$

In the very special case $p = 2$, $q = \log n$ an upper bound for the number of iterations in the algorithm as

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



(SDO) and its dual (SDD) in standard form:

$$(SDO) \quad p^* = \inf \{ \text{Tr}(CX) \}$$

$$\text{Tr}(A_i X) = b_i, i = 1, \dots, m; X \succeq 0$$

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

Here C and A_i ($1 \leq i \leq m$) are symmetric matrices, and $b, y \in \mathbb{R}^m$. ‘ $X \succeq 0$ ’ means that X is symmetric positive semidefinite. The matrices A_i are assumed to be linearly independent.

Central path:

$$\begin{cases} \text{Tr}(A_i X) & = & b_i, & i = 1, \dots, m, \\ \sum_{i=1}^m y_i A_i + S & = & C, \\ XS & = & \mu E, & X, S \succeq 0, \end{cases}$$

where E denotes the $n \times n$ identity matrix and $\mu > 0$.

Newton System:

$$\begin{aligned} \text{Tr}(A_i \Delta X) & = 0, & i = 1, \dots, m, \\ - \sum_{i=1}^m \Delta y_i A_i & = \Delta S, \\ X \Delta S + \Delta X S & = \mu E - XS. \end{aligned}$$

Observation: For SDO, the above Newton system might not have a symmetric solution ΔX .



NT-scaling

$$P = X^{\frac{1}{2}} (X^{\frac{1}{2}} S X^{\frac{1}{2}})^{-\frac{1}{2}} X^{\frac{1}{2}} = S^{-\frac{1}{2}} (S^{\frac{1}{2}} X S^{\frac{1}{2}})^{\frac{1}{2}} S^{-\frac{1}{2}},$$

and $D = P^{\frac{1}{2}}$. Matrix D can be used to rescale X and S to the same matrix V :

$$V := \frac{1}{\sqrt{\mu}} D^{-1} X D^{-1} = \frac{1}{\sqrt{\mu}} D S D.$$

Let us further define

$$\begin{aligned} \bar{A}_i &:= D A_i D; & i = 1, \dots, m; \\ D_X &:= \frac{1}{\sqrt{\mu}} D^{-1} \Delta X D^{-1}, & D_S &:= \frac{1}{\sqrt{\mu}} D \Delta S D. \end{aligned}$$

Then the NT search direction can be written as the solution of the following system

$$\begin{cases} \text{Tr}(\bar{A}_i D_X) &= 0, & i = 1, \dots, m, \\ -\sum_{i=1}^m \Delta y_i \bar{A}_i &= D_S \\ D_X + D_S &= V^{-1} - V \end{cases}$$

As in LO, the last equation should be replaced by:

$$D_X + D_S = -\psi'(V)$$



Definition 2 *Let the matrix X be diagonalizable with*

$$X = Q_X^{-1} \text{diag} (\lambda_1(X), \lambda_2(X), \dots, \lambda_n(X)) Q_X,$$

where Q_X is nonsingular and $\psi(t)$ is a function from \mathfrak{R} into itself. The function $\psi(X)$ is defined by

$$\psi(X) = Q_X^{-1} \text{diag} (h(\lambda_1(X)), h(\lambda_2(X)), \dots, h(\lambda_n(X))) Q_X. \quad (1)$$

*If X is symmetric, then Q_X can be orthogonal:
 $Q_X^{-1} = Q_X^T$.*

For any matrix G whose eigenvalues are real^a, let us define

$$\Psi(G) = \sum_{i=1}^n \psi(\lambda_i(G)) = \text{Tr} (\psi(G)), \quad (2)$$

where $\psi(G)$ is defined by (1). The proximity for SDO is defined by

$$\Psi(XS, \mu) := \Psi(V) = \text{Tr} (\psi(V)), \quad (3)$$

where $\psi(G)$ is defined by (1).

^a G might be not symmetric positive definite.

Proposition 2 *Let the matrix functions $\Psi(X), \psi(X)$ be defined by (2) and (1) respectively. If $\psi(t)$ is self-regular, then*

- **[C.3]** $\Psi(X)$ is strongly convex with respect to $X \succ 0$ and vanishes at its global minimal point $X = E$, i.e., $\Psi(E) = 0, \psi'(E) = 0_{n \times n}$. Further there exists two positive constants $\nu_1, \nu_2 > 0$ such that for any $p, q \geq 1$

$$\nu_1(X^{p-1} + X^{-1-q}) \preceq \psi''(X) \preceq \nu_2(X^{p-1} + X^{-1-q}). \quad (4)$$

- **[C. 4]** For any $X_1, X_2 \succ 0$,

$$\Psi \left([X_1^{\frac{1}{2}} X_2 X_1^{\frac{1}{2}}]^{\frac{1}{2}} \right) \leq \frac{1}{2} (\Psi(X_1) + \Psi(X_2)). \quad (5)$$



Large Update Primal-Dual Algorithm for Semidefinite Optimization

Input:

A proximity parameter $\tau > \nu_1^{-1}$;
 an accuracy parameter $\epsilon > 0$;
 a variable damping factor α ;
 a fixed barrier update parameter θ , $0 < \theta < 1$;
 (X^0, S^0) and $\mu^0 = 1$ such that $\Psi(X^0 S^0, \mu^0) \leq \tau$.

begin

$X := X^0$; $S := S^0$; $\mu := \mu^0$;

while $n\mu \geq \epsilon$ **do**

begin

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

while $\Psi(XS, \mu) \geq \tau$ **do**

Solve the Newton system and compute $\Delta X, \Delta y, \Delta Y$,

begin

$X := X + \alpha\Delta X$;

$S := S + \alpha\Delta S$;

$y := y + \alpha\Delta y$

end

end

end



Standard form LO model

$$(P) \quad \min \left\{ c^T x : Ax = b, \quad x \in L^n \right\},$$

$$(D) \quad \max \left\{ b^T y : A^T y + s = c, \quad s \in L^n \right\}.$$

Where L^n is the Lorenz(ice-cream) cone

$L^n := \{x : x_1 \geq \|(x_2, \dots, x_n)^T\| = \|x_{2:n}\|\}$
 or the product of such cones.

OPTIMALITY:

$$\begin{aligned} Ax &= b, \quad x \in L^n, \\ A^T y + s &= c, \quad s \in L^n, \\ x \circ s &= 0. \end{aligned}$$

where $x \circ s := (x^T s, x_1 s_{2:n}^T + s_1 x_{2:n}^T)^T = \text{mat}(x)s$

$$\text{mat}(x) = \begin{pmatrix} x_1 & x_1 & x_1 & \dots & x_1 \\ x_1 & x_2 & 0 & \dots & 0 \\ x_1 & 0 & x_3 & & 0 \\ \vdots & \vdots & & \ddots & 0 \\ x_1 & 0 & 0 & \dots & x_n \end{pmatrix}$$



Central path

$$\begin{aligned} Ax &= b, \quad x \in L^n, \\ A^T y + s &= c, \quad s \in L^n, \\ x \circ s &= \mu e_1. \end{aligned}$$

where $e_1 := (1, 0, \dots, 0)^T$ or the concatenation of these with appropriate dimensions.

Algebra for SOCO

Eigenvalues

$$\lambda_{\max}(x) := x_1 + \|x_{2:n}\|$$

$$\lambda_{\min}(x) := x_1 - \|x_{2:n}\|$$

$$\text{Tr}(x) := \lambda_{\max}(x) + \lambda_{\min}(x) = 2x_1$$

$$\det(x) := \lambda_{\max}(x)\lambda_{\min}(x) = x_1^2 - \|x_{2:n}\|^2$$

Further observe: The circle operation is *commutative*, but not *associative*.



Definition 3 Let $\psi(t) : \mathfrak{R} \rightarrow \mathfrak{R}$. Then

$\psi(x) : \mathfrak{R}^n \rightarrow \mathfrak{R}^n$ is defined as

$$\begin{pmatrix} \frac{1}{2} (\psi(\lambda_{\max}(x)) + \psi(\lambda_{\min}(x))) \\ \frac{\psi(\lambda_{\max}(x)) - \psi(\lambda_{\min}(x))}{2 \|x_{2:n}\|} x_{2:n} \end{pmatrix}, \quad \text{if } x_{2:n} \neq 0;$$

$$(\psi(\lambda_{\max}(x)), 0, \dots, 0)^T, \quad \text{if } x_{2:n} = 0.$$

PROPERTIES:

- If $\psi(t) \geq 0$ for all $t \geq 0$ and $x \in L^n$ then we have $\psi(x) \in L^n$.
- The derivatives can be defined analogously.
- x^p can be defined, e.g if $p = -1$, then $\forall x \in L^n$

$$x^{-1} = \frac{1}{\det(x)} (x_1, -x_2, -x_3, \dots, -x_n)^T$$

Definition 4 The mapping

$\psi(x) : L^n \rightarrow \mathfrak{R}^n$ is Self-Regular iff

the kernel function $\psi(t) : \mathfrak{R} \rightarrow \mathfrak{R}$ is Self-Regular.



The classical Newton step

$$\begin{aligned}A\Delta x &= 0, \\A^T \Delta y + \Delta s &= 0, \\x \circ \Delta s + s \circ p_x &= \mu e_1 - x \circ s.\end{aligned}$$

Use the NT Scaling

$$\begin{aligned}\bar{A}p_x &= 0, \\\bar{A}^T \Delta y + p_s &= 0, \\p_x + p_s &= -\psi'(v).\end{aligned}$$



- **IPMs is still a challenging research field.**
(Both in theory and in implementations.)
- Various new research areas:
 - self-regular function,
 - SOCO, SDO, NLCP
 - implementation, experimental results
 - higher order variants
 - asymptotic convergence / predictor-corrector
 - pure primal variants
 - other proximities, e.g. exponential

