School of Mathematics

Interior Point Methods for Linear Programming

Jacek Gondzio
Email: `J.Gondzio@ed.ac.uk`
URL: `http://www.maths.ed.ac.uk/~gondzio`

Outline

- **Part 1: IPM for LP: Motivation**
  - what is wrong with the simplex method?
  - complementarity conditions
  - first order optimality conditions
  - Newton, Lagrange and Fiacco & McCormick
  - central trajectory
  - primal-dual framework
- **Part 2: Polynomial Complexity of IPM**
  - Newton method
  - short step path-following method
  - polynomial complexity proof
- **Final Comments**

Part 1:

IPM for LP: Motivation

Simplex: What's wrong?

A **vertex** is defined by a set of n equations:

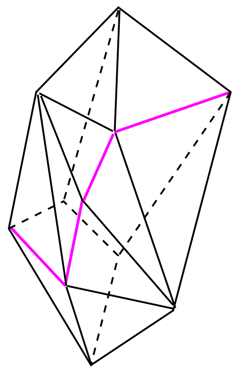
$$\begin{bmatrix} B & N \\ 0 & I_{n-m} \end{bmatrix} \begin{bmatrix} x_B \\ x_N \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}.$$

The LP program with m constraints and n variables ($n \geq m$) may have as many as

$$N_V = \binom{\mathrm{n}}{\mathrm{m}} = \frac{n!}{m!(n-m)!}$$

vertices.

Geometry of the Simplex Method



The simplex method can make a non-polynomial number of iterations to reach the optimality.

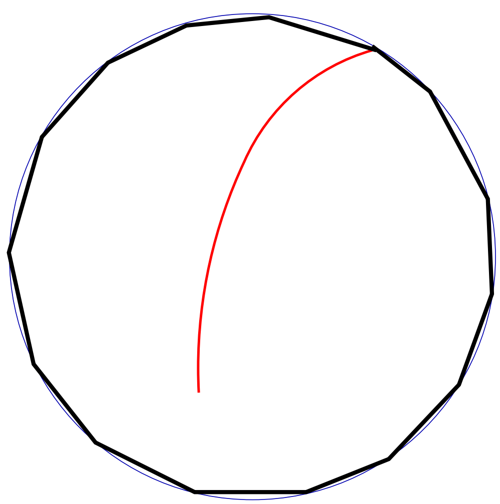
V. Klee and G. Minty gave an example LP the solution of which needs 2^n iterations.

V. Klee and G. Minty,
How good is the simplex algorithm,
in: Inequalities-III, O. Shisha, ed., Acad. Press, 1972, 159–175.

Narendra Karmarkar from AT&T Bell Labs:
"the simplex [method] is complex"

N. Karmarkar: A New Polynomial–time Algorithm for LP, *Combinatorica* 4 (1984) 373–395.

EURO 2016 view of LP



"Elements" of the IPM

What do we need to derive the **Interior Point Method**?

- duality theory:
  Lagrangian function;
  first order optimality conditions.
- logarithmic barriers.
- Newton method.

The following 3 slides remind key facts from the duality theory applied to linear programming.

Duality in LP

Consider a **primal** program

$$\begin{array}{lll} \min & c^T x & \\ \text{s.t.} & Ax & = b, \\ & x \geq 0, & \end{array} \qquad (1)$$

where $c, x \in \mathcal{R}^n, b \in \mathcal{R}^m, A \in \mathcal{R}^{m \times n}$.
With the primal we associate a *dual* program

$$\begin{array}{lll} \max & b^T y & \\ \text{s.t.} & A^T y & \leq c, \\ & y \text{ free}, & \end{array}$$

where $y \in \mathcal{R}^m$. We add *dual slack* $s \in \mathcal{R}^n$, $s \geq 0$, convert inequality $A^T y \leq c$ into an equation $A^T y + s = c$ and get **dual** program

$$\begin{array}{llll} \max & b^T y & & \\ \text{s.t.} & A^T y + & s & = c, \\ & y \text{ free}, & s \geq 0, & \end{array} \qquad (2)$$

where $y \in \mathcal{R}^m$ and $s \in \mathcal{R}^n$.

Let $\mathcal{P}$, $\mathcal{D}$ be the feasible sets of the primal and the dual, resp.:
$\mathcal{P} = \{x \in \mathcal{R}^n \,|\, Ax = b,\ x \geq 0\}$
$\mathcal{D} = \{y \in \mathcal{R}^m, s \in \mathcal{R}^n \,|\, A^T y + s = c,\ s \geq 0\}$.

Let us introduce a convention that
$\inf\limits_{x \in \mathcal{P}} c^T x = +\infty$, if $\mathcal{P} = \emptyset$; $\quad \sup_{y \in \mathcal{D}} b^T y = -\infty$, if $\mathcal{D} = \emptyset$.

Weak Duality Theorem

$$\inf_{x \in \mathcal{P}} c^T x \geq \sup_{y \in \mathcal{D}} b^T y.$$

Strong Duality Theorem

If either $\mathcal{P} \neq \emptyset$ or $\mathcal{D} \neq \emptyset$ then

$$\inf_{x \in \mathcal{P}} c^T x = \sup_{y \in \mathcal{D}} b^T y.$$

If one of problems (1) and (2) is *solvable* then

$$\min_{x \in \mathcal{P}} c^T x = \max_{y \in \mathcal{D}} b^T y.$$

In **IPMs** we shall use the term **interior-point**. Let $\mathcal{P}^0$, $\mathcal{D}^0$ be the **strictly feasible sets** of the primal and the dual, respectively:
$\mathcal{P}^0 = \{x \in \mathcal{R}^n \,|\, Ax = b,\ x > 0\}$
$\mathcal{D}^0 = \{y \in \mathcal{R}^m, s \in \mathcal{R}^n \,|\, A^T y + s = c,\ s > 0\}$.

We shall often refer to the **primal-dual pair**. Hence we define *primal-dual feasible* set $\mathcal{F}$ and *primal-dual strictly feasible* set $\mathcal{F}^0$:
$\mathcal{F} \ = \{(x, y, s) \,|\, Ax = b,\ A^T y + s = c,\ (x, s) \geq 0\}$
$\mathcal{F}^0 = \{(x, y, s) \,|\, Ax = b,\ A^T y + s = c,\ (x, s) > 0\}$.

Primal-Dual Pair of Linear Programs

Primal	Dual
$\begin{array}{lll} \min & c^T x & \\ \text{s.t.} & Ax & = b, \\ & x \geq 0; & \end{array}$	$\begin{array}{lll} \max & b^T y & \\ \text{s.t.} & A^T y + s & = c, \\ & s \geq 0. & \end{array}$

Lagrangian

$$L(x, y) = c^T x - y^T (Ax - b) - s^T x.$$

Optimality Conditions

$$\begin{array}{rcl} Ax & = & b, \\ A^T y + s & = & c, \\ XSe & = & 0, \quad (\text{ i.e., } x_j \cdot s_j = 0 \quad \forall j), \\ (x, s) & \geq & 0, \end{array}$$

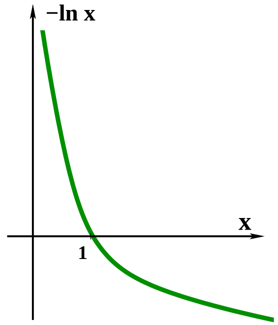
$X = diag\{x_1, \cdots, x_n\}$, $S = diag\{s_1, \cdots, s_n\}$, $e = (1, \cdots, 1) \in \mathcal{R}^n$.

Logarithmic barrier

$-\ln x_j$

"replaces" the inequality

$x_j \geq 0$.



Observe that

$$\min e^{-\sum_{j=1}^n \ln x_j} \iff \max \prod_{j=1}^n x_j$$

The minimization of $-\sum_{j=1}^n \ln x_j$ is equivalent to the maximization of the product of distances from all hyperplanes defining the positive orthant: it prevents all x_j from approaching zero.

Logarithmic barrier

Replace the **primal** LP

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b, \\ & x \geq 0, \end{array}$$

with the **primal barrier program**

$$\begin{array}{ll} \min & c^T x - \mu \sum_{j=1}^n \ln x_j \\ \text{s.t.} & Ax = b. \end{array}$$

Lagrangian: $L(x,y,\mu) = c^T x - y^T (Ax - b) - \mu \sum_{j=1}^n \ln x_j$.

Conditions for a stationary point of the Lagrangian

$$\begin{array}{rcl} \nabla_x L(x,y,\mu) & = & c - A^T y - \mu X^{-1} e = 0 \\ \nabla_y L(x,y,\mu) & = & Ax - b = 0, \end{array}$$

where $X^{-1} = diag\{x_1^{-1}, x_2^{-1}, \cdots, x_n^{-1}\}$.

Let us denote

$$s = \mu X^{-1} e, \quad \text{i.e.} \quad XSe = \mu e.$$

The **First Order Optimality Conditions** are:

$$\begin{array}{rcl} Ax & = & b, \\ A^T y + s & = & c, \\ XSe & = & \mu e, \\ (x,s) & > & 0. \end{array}$$

Complementarity $x_j \cdot s_j = 0 \quad \forall j = 1,2,...,n.$

Simplex Method makes a guess of optimal partition:

For *basic* variables, $s_B = 0$ and

$$(x_B)_j \cdot (s_B)_j = 0 \quad \forall j \in \mathcal{B}.$$

For *non-basic* variables, $x_N = 0$ hence

$$(x_N)_j \cdot (s_N)_j = 0 \quad \forall j \in \mathcal{N}.$$

Interior Point Method uses ε-mathematics:

Replace $x_j \cdot s_j = 0 \quad \forall j = 1,2,...,n$

by $x_j \cdot s_j = \mu \quad \forall j = 1,2,...,n.$

Force convergence $\mu \to 0$.

First Order Optimality Conditions

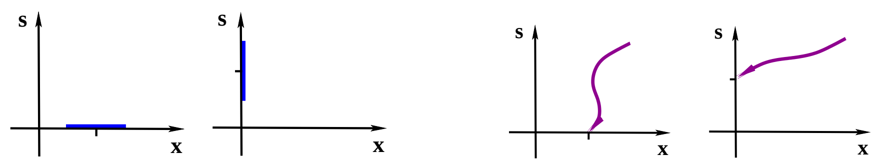
Approaching Optimality:

Simplex Method:

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

Interior Point Method:

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



Basic: x > 0, s = 0 Nonbasic: x = 0, s > 0 "Basic": x > 0, s = 0 "Nonbasic": x = 0, s > 0

Central Trajectory

The first order optimality conditions for the barrier problem

$$\begin{aligned} Ax &= b, \\ A^Ty + s &= c, \\ XSe &= \mu e, \\ (x, s) &\geq 0 \end{aligned}$$

approximate the first order optimality conditions for the LP

$$\begin{aligned} Ax &= b, \\ A^Ty + s &= c, \\ XSe &= 0, \\ (x, s) &\geq 0 \end{aligned}$$

more and more closely as μ goes to zero.

Central Trajectory

Parameter μ controls the distance to optimality.

$$c^Tx - b^Ty = c^Tx - x^TA^Ty = x^T(c - A^Ty) = x^Ts = n\mu.$$

Analytic centre (μ-centre): a (unique) point

$$(x(\mu), y(\mu), s(\mu)), \quad x(\mu) > 0,\ s(\mu) > 0$$

that satisfies FOC.

The path

$$\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$$

is called the **primal-dual central trajectory**.

Newton Method

is used to find a stationary point of the barrier problem.

Recall how to use Newton Method to find a root of a nonlinear equation

$$f(x) = 0.$$

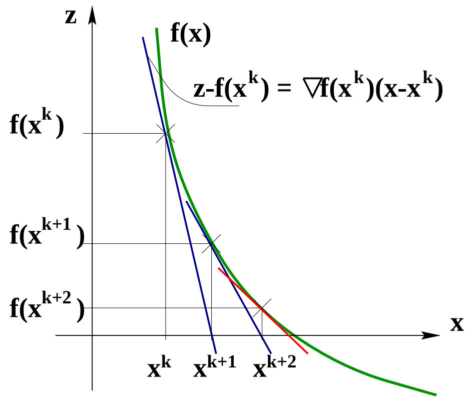
A tangent line

$$z - f(x^k) = \nabla f(x^k) \cdot (x - x^k)$$

is a local approximation of the graph of the function $f(x)$. Substituting $z = 0$ defines a new point

$$x^{k+1} = x^k - (\nabla f(x^k))^{-1} f(x^k).$$

Newton Method



$$x^{k+1} = x^k - (\nabla f(x^k))^{-1} f(x^k).$$

Apply Newton Method to the FOC

The first order optimality conditions for the barrier problem form a large system of nonlinear equations

$$f(x, y, s) = 0,$$

where $f : \mathcal{R}^{2n+m} \mapsto \mathcal{R}^{2n+m}$ is a mapping defined as follows:

$$f(x, y, s) = \begin{bmatrix} Ax - b \\ A^T y + s - c \\ XSe - \mu e \end{bmatrix}.$$

Actually, the first two terms of it are **linear**; only the last one, corresponding to the complementarity condition, is **nonlinear**.

Newton Method (cont'd)

Note that

$$\nabla f(x, y, s) = \begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{bmatrix}.$$

Thus, for a given point (x, y, s) we find the Newton direction $(\Delta x, \Delta y, \Delta s)$ by solving the system of linear equations:

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{bmatrix} \cdot \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} b - Ax \\ c - A^T y - s \\ \mu e - XSe \end{bmatrix}.$$

Interior-Point Framework

The **logarithmic barrier**

$$-\ln x_j$$

"replaces" the inequality

$$x_j \geq 0.$$

We derive the **first order optimality conditions** for the primal barrier problem:

$$\begin{aligned} Ax &= b, \\ A^T y + s &= c, \\ XSe &= \mu e, \end{aligned}$$

and apply **Newton method** to solve this system of (nonlinear) equations.

Actually, we fix the barrier parameter μ and make only **one** (damped) Newton step towards the solution of FOC. We do not solve the current FOC exactly. Instead, we immediately reduce the barrier parameter μ (to ensure progress towards optimality) and repeat the process.

Interior Point Algorithm

Initialize

$k = 0 \qquad (x^0, y^0, s^0) \in \mathcal{F}^0$

$\mu_0 = \frac{1}{n} \cdot (x^0)^T s^0 \qquad \alpha_0 = 0.9995$

Repeat until optimality

$k = k + 1$

$\mu_k = \sigma \mu_{k-1}$, where $\sigma \in (0,1)$

$\Delta = (\Delta x, \Delta y, \Delta s) =$ Newton direction towards μ-centre

Ratio test:

$\alpha_P := \max \{\alpha > 0 : \ x + \alpha \Delta x \geq 0\},$

$\alpha_D := \max \{\alpha > 0 : \ s + \alpha \Delta s \geq 0\}.$

Make step:

$x^{k+1} = x^k + \alpha_0 \alpha_P \Delta x,$

$y^{k+1} = y^k + \alpha_0 \alpha_D \Delta y,$

$s^{k+1} = s^k + \alpha_0 \alpha_D \Delta s.$

Notations

$$X = diag\{x_1, x_2, \cdots, x_n\} = \begin{bmatrix} x_1 & & & \\ & x_2 & & \\ & & \ddots & \\ & & & x_n \end{bmatrix}.$$

$e = (1, 1, \cdots, 1) \in \mathcal{R}^n, \ \ X^{-1} = diag\{x_1^{-1}, x_2^{-1}, \cdots, x_n^{-1}\}.$

An equation $\qquad XSe = \mu e,$

is equivalent to $\qquad x_j s_j = \mu, \quad \forall j = 1, 2, \cdots, n.$

Notations(cont'd)

Primal feasible set $\mathcal{P} = \{x \in \mathcal{R}^n \mid Ax = b, \ x \geq 0\}$.
Primal strictly feasible set $\mathcal{P}^0 = \{x \in \mathcal{R}^n \mid Ax = b, \ x > 0\}$.

Dual feasible set $\mathcal{D} = \{y \in \mathcal{R}^m, s \in \mathcal{R}^n \mid A^T y + s = c, \ s \geq 0\}$.
Dual strictly feasible set $\mathcal{D}^0 = \{y \in \mathcal{R}^m, s \in \mathcal{R}^n \mid A^T y + s = c, \ s > 0\}$.

Primal-dual feasible set
$\mathcal{F} = \{(x, y, s) \mid Ax = b, \ A^T y + s = c, \ (x, s) \geq 0\}$.
Primal-dual strictly feasible set
$\mathcal{F}^0 = \{(x, y, s) \mid Ax = b, \ A^T y + s = c, \ (x, s) > 0\}$.

Interior Point Methods

Marsten, Subramanian, Saltzman, Lustig and Shanno:
"Interior point methods for linear programming:
Just call Newton, Lagrange, and Fiacco and McCormick!",
Interfaces 20 (1990) No 4, pp. 105–116.

- **Fiacco & McCormick (1968)**
  inequality constraints $\longrightarrow$ logarithmic barrier;
  a sequence of unconstrained minimizations
- **Lagrange (1788)**
  equality constraints $\longrightarrow$ multipliers;
- **Newton (1687)**
  solve unconstrained minimization problems;

Part 2:

Path-Following Method: Theory

Path-Following Algorithm

The analysis given in this lecture comes from the book of
Steve Wright: *Primal-Dual Interior-Point Methods*,
SIAM Philadelphia, 1997.

We analyze a **feasible** interior-point algorithm with the following properties:

- all its iterates are feasible and stay in a close neighbourhood of the central path;
- the iterates follow the central path towards optimality;
- systematic (though slow) reduction of duality gap is ensured.

This algorithm is called
the **short-step path-following method**.
Indeed, it makes very slow progress (short-steps) to optimality.

Central Path Neighbourhood

Assume a primal-dual strictly feasible solution $(x, y, s) \in \mathcal{F}^0$ lying in a neighbourhood of the central path is given; namely (x, y, s) satisfies:

$$\begin{aligned} Ax &= b, \\ A^T y + s &= c, \\ XSe &\approx \mu e. \end{aligned}$$

We define a **θ-neighbourhood** of the central path $N_2(\theta)$, a set of primal-dual strictly feasible solutions $(x, y, s) \in \mathcal{F}^0$ that satisfy:

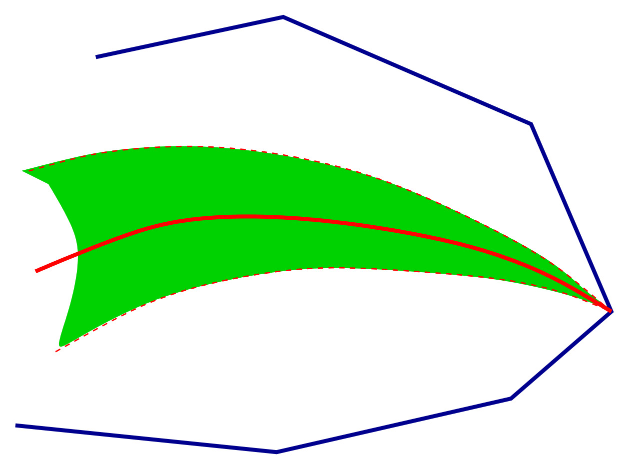
$$\|XSe - \mu e\| \leq \theta \mu,$$

where $\theta \in (0, 1)$ and the barrier μ satisfies:

$$x^T s = n\mu.$$

Hence $N_2(\theta) = \{(x, y, s) \in \mathcal{F}^0 \mid \|XSe - \mu e\| \leq \theta \mu\}$.

Central Path Neighbourhood



$N_2(\theta)$ neighbourhood of the central path

Progress towards optimality

Assume a primal-dual strictly feasible solution $(x, y, s) \in N_2(\theta)$ for some $\theta \in (0,1)$ is given.

Interior point algorithm tries to move from this point to another one that also belongs to a θ-neighbourhood of the central path but corresponds to a smaller μ. The required reduction of μ is small:

$$\mu^{k+1} = \sigma \mu^k, \qquad \text{where} \qquad \sigma = 1 - \beta/\sqrt{n},$$

for some $\beta \in (0,1)$.

This is a **short-step** method:
It makes short steps to optimality.

Progress towards optimality

Given a new μ-centre, interior point algorithm computes Newton direction:

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{bmatrix} \cdot \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \sigma \mu e - XSe \end{bmatrix},$$

and makes step in this direction.

Magic numbers (will be explained later):

$$\theta = 0.1 \quad \text{and} \quad \beta = 0.1.$$

θ controls the proximity to the central path;
β controls the progress to optimality.

How to prove the $\mathcal{O}(\sqrt{n})$ complexity result

We will prove the following:

- full step in Newton direction is feasible;
- the new iterate $(x^{k+1}, y^{k+1}, s^{k+1}) = (x^k, y^k, s^k) + (\Delta x^k, \Delta y^k, \Delta s^k)$ belongs to the θ-neighbourhood of the new μ-centre (with $\mu^{k+1} = \sigma \mu^k$);
- duality gap is reduced $1 - \beta/\sqrt{n}$ times.

$\mathcal{O}(\sqrt{n})$ complexity result

Note that since at one iteration duality gap is reduced $1 - \beta/\sqrt{n}$ times, after $\sqrt{n}$ iterations the reduction achieves:

$$(1 - \beta/\sqrt{n})^{\sqrt{n}} \approx e^{-\beta}.$$

After $C \cdot \sqrt{n}$ iterations, the reduction is $e^{-C\beta}$. For sufficiently large constant C the reduction can thus be arbitrarily large (i.e. the duality gap can become arbitrarily small).

Hence this algorithm has complexity $\mathcal{O}(\sqrt{n})$.

This should be understood as follows:

> *"after the number of iterations proportional to $\sqrt{n}$ the algorithm solves the problem".*

Technical Results

Lemma 1

Newton direction $(\Delta x, \Delta y, \Delta s)$ defined by the equation system

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{bmatrix} \cdot \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \sigma \mu e - XSe \end{bmatrix}, \quad (3)$$

satisfies:

$$\Delta x^T \Delta s = 0.$$

Proof:

From the first two equations in (3) we get

$$A\Delta x = 0 \quad \text{and} \quad \Delta s = -A^T \Delta y.$$

Hence

$$\Delta x^T \Delta s = \Delta x^T \cdot (-A^T \Delta y) = -\Delta y^T \cdot (A \Delta x) = 0.$$

Technical Results (cont'd)

Lemma 2

Let $(\Delta x, \Delta y, \Delta s)$ be the Newton direction that solves the system (3). The new iterate

$$(\bar{x}, \bar{y}, \bar{s}) = (x, y, s) + (\Delta x, \Delta y, \Delta s)$$

satisfies

$$\bar{x}^T \bar{s} = n \bar{\mu},$$

where

$$\bar{\mu} = \sigma \mu.$$

Proof: From the third equation in (3) we get

$$S \Delta x + X \Delta s = -XSe + \sigma \mu e.$$

By summing the n components of this equation we obtain

$$\begin{aligned} e^T(S\Delta x + X\Delta s) = s^T \Delta x + x^T \Delta s &= -e^T XSe + \sigma \mu e^T e \\ &= -x^T s + n\sigma\mu = -x^T s \cdot (1-\sigma). \end{aligned}$$

Thus

$$\begin{aligned} \bar{x}^T \bar{s} &= (x + \Delta x)^T (s + \Delta s) \\ &= x^T s + (s^T \Delta x + x^T \Delta s) + (\Delta x)^T \Delta s \\ &= x^T s + (\sigma - 1) x^T s + 0 = \sigma x^T s, \end{aligned}$$

which is equivalent to: $\quad n \bar{\mu} = \sigma n \mu.$

Reminder: Norms of the vector $x \in \mathcal{R}^n$.

$$\begin{aligned} \|x\| &= (\sum_{j=1}^{n} x_j^2)^{1/2} \\ \|x\|_\infty &= \max_{j \in \{1..n\}} |x_j| \\ \|x\|_1 &= \sum_{j=1}^{n} |x_j| \end{aligned}$$

For any $x \in \mathcal{R}^n$:

$$\begin{aligned} \|x\|_\infty &\leq \|x\|_1 \\ \|x\|_1 &\leq n \cdot \|x\|_\infty \\ \|x\|_\infty &\leq \|x\| \\ \|x\| &\leq \sqrt{n} \cdot \|x\|_\infty \\ \|x\| &\leq \|x\|_1 \\ \|x\|_1 &\leq \sqrt{n} \cdot \|x\| \end{aligned}$$

Reminder: Triangle Inequality

For any vectors p, q and r and for any norm $\|.\|$

$$\|p-q\| \leq \|p-r\| + \|r-q\|.$$

The relation between *algebraic* and *geometric* means.
For any scalars a and b such that $ab \geq 0$:

$$\sqrt{|ab|} \leq \frac{1}{2} \cdot |a+b|.$$

Technical Result (algebra)

Lemma 3 Let u and v be any two vectors in $\mathcal{R}^n$ such that $u^T v \geq 0$. Then

$$\|UVe\| \leq 2^{-3/2}\|u+v\|^2,$$

where $U = diag\{u_1, \cdots, u_n\}$, $V = diag\{v_1, \cdots, v_n\}$.

Proof: Let us partition all products $u_j v_j$ into positive and negative ones:

$$\mathcal{P} = \{j \,|\, u_j v_j \geq 0\} \quad \text{and} \quad \mathcal{M} = \{j \,|\, u_j v_j < 0\} :$$

$$0 \leq u^T v = \sum_{j\in\mathcal{P}} u_j v_j + \sum_{j\in\mathcal{M}} u_j v_j = \sum_{j\in\mathcal{P}} |u_j v_j| - \sum_{j\in\mathcal{M}} |u_j v_j|.$$

Proof (cont'd)

We can now write

$$\begin{aligned}
\|UVe\| &= (\|[u_j v_j]_{j\in\mathcal{P}}\|^2 + \|[u_j v_j]_{j\in\mathcal{M}}\|^2)^{1/2} \\
&\leq (\|[u_j v_j]_{j\in\mathcal{P}}\|_1^2 + \|[u_j v_j]_{j\in\mathcal{M}}\|_1^2)^{1/2} \\
&\leq (2\|[u_j v_j]_{j\in\mathcal{P}}\|_1^2)^{1/2} \\
&\leq \sqrt{2}\|[\frac{1}{4}(u_j+v_j)^2]_{j\in\mathcal{P}}\|_1 \\
&= 2^{-3/2}\sum_{j\in\mathcal{P}}(u_j+v_j)^2 \\
&\leq 2^{-3/2}\sum_{j=1}^{n}(u_j+v_j)^2 \\
&= 2^{-3/2}\|u+v\|^2, \quad \text{as requested.}
\end{aligned}$$

IPM Technical Results (cont'd)

Lemma 4
If $(x, y, s) \in N_2(\theta)$ for some $\theta \in (0, 1)$, then

$$(1-\theta)\mu \leq x_j s_j \leq (1+\theta)\mu \quad \forall j.$$

In other words,

$$\min_{j\in\{1..n\}} x_j s_j \geq (1-\theta)\mu,$$

$$\max_{j\in\{1..n\}} x_j s_j \leq (1+\theta)\mu.$$

Proof:

Since $\|x\|_\infty \le \|x\|$, from the definition of $N_2(\theta)$,

$$N_2(\theta) = \{(x,y,s) \in \mathcal{F}^0 \mid \|XSe - \mu e\| \le \theta\mu\},$$

we conclude

$$\|XSe - \mu e\|_\infty \le \|XSe - \mu e\| \le \theta\mu.$$

Hence

$$|x_j s_j - \mu| \le \theta\mu \quad \forall j,$$

which is equivalent to

$$-\theta\mu \le x_j s_j - \mu \le \theta\mu \quad \forall j.$$

Thus

$$(1-\theta)\mu \le x_j s_j \le (1+\theta)\mu \quad \forall j.$$

IPM Technical Results (cont'd)

Lemma 5
If $(x,y,s) \in N_2(\theta)$ for some $\theta \in (0,1)$, then

$$\|XSe - \sigma\mu e\|^2 \le \theta^2\mu^2 + (1-\sigma)^2\mu^2 n.$$

Proof:
Note first that

$$e^T(XSe - \mu e) = x^T s - \mu e^T e = n\mu - n\mu = 0.$$

Therefore

$$\begin{aligned}
&\|XSe - \sigma\mu e\|^2 \\
&\quad = \|(XSe - \mu e) + (1-\sigma)\mu e\|^2 \\
&\quad = \|XSe - \mu e\|^2 + 2(1-\sigma)\mu e^T(XSe - \mu e) + (1-\sigma)^2\mu^2 e^T e \\
&\quad \le \theta^2\mu^2 + (1-\sigma)^2\mu^2 n.
\end{aligned}$$

IPM Technical Results (cont'd)

Lemma 6
If $(x,y,s) \in N_2(\theta)$ for some $\theta \in (0,1)$, then

$$\|\Delta X \Delta S e\| \le \frac{\theta^2 + n(1-\sigma)^2}{2^{3/2}(1-\theta)}\mu.$$

Proof: 3rd equation in the Newton system gives

$$S\Delta x + X\Delta s = -XSe + \sigma\mu e.$$

Having multiplied it with $(XS)^{-1/2}$, we obtain

$$X^{-1/2}S^{1/2}\Delta x + X^{1/2}S^{-1/2}\Delta s = (XS)^{-1/2}(-XSe + \sigma\mu e).$$

Proof (cont'd)
Define $u = X^{-1/2}S^{1/2}\Delta x$ and $v = X^{1/2}S^{-1/2}\Delta s$ and observe that (by Lemma 1) $u^T v = \Delta x^T \Delta s = 0$. Now apply Lemma 3:

$$\begin{aligned}
\|\Delta X \Delta S e\| &= \|(X^{-1/2}S^{1/2}\Delta X)(X^{1/2}S^{-1/2}\Delta S)e\| \\
&\le 2^{-3/2}\|X^{-1/2}S^{1/2}\Delta x + X^{1/2}S^{-1/2}\Delta s\|^2 \\
&= 2^{-3/2}\|X^{-1/2}S^{-1/2}(-XSe + \sigma\mu e)\|^2 \\
&= 2^{-3/2}\sum_{j=1}^{n}\frac{(-x_j s_j + \sigma\mu)^2}{x_j s_j} \\
&\le 2^{-3/2}\frac{\|XSe - \sigma\mu e\|^2}{\min_j x_j s_j} \\
&\le \frac{\theta^2 + n(1-\sigma)^2}{2^{3/2}(1-\theta)}\mu \qquad \text{(by Lemmas 4 and 5).}
\end{aligned}$$

Magic Numbers

We have previously set two parameters for the short-step path-following method:

$$\theta = 0.1 \quad \text{and} \quad \beta = 0.1.$$

Now it's time to justify this particular choice.

Both θ and β have to be small to make sure that a full step in the Newton direction does not take the new iterate outside the neighbourhood $N_2(\theta)$.

θ controls the proximity to the central path;
β controls the progress to optimality.

Magic numbers choice lemma

Lemma 7 If $\theta = 0.1$ and $\beta = 0.1$, then

$$\frac{\theta^2 + n(1-\sigma)^2}{2^{3/2}(1-\theta)} \leq \sigma\theta.$$

Proof:
Recall that

$$\sigma = 1 - \beta/\sqrt{n}.$$

Hence

$$n(1-\sigma)^2 = \beta^2$$

and for $\beta = 0.1$ (for any $n \geq 1$)

$$\sigma \geq 0.9.$$

Substituting $\theta = 0.1$ and $\beta = 0.1$, we obtain

$$\frac{\theta^2 + n(1-\sigma)^2}{2^{3/2}(1-\theta)} = \frac{0.1^2 + 0.1^2}{2^{3/2} \cdot 0.9} \leq 0.02 \leq 0.9 \cdot 0.1 \leq \sigma\theta.$$

Full Newton step in $N_2(\theta)$

Lemma 8 Suppose $(x, y, s) \in N_2(\theta)$ and $(\Delta x, \Delta y, \Delta s)$ is the Newton direction computed from the system (3). Then the new iterate

$$(\bar{x}, \bar{y}, \bar{s}) = (x, y, s) + (\Delta x, \Delta y, \Delta s)$$

satisfies $(\bar{x}, \bar{y}, \bar{s}) \in N_2(\theta)$, i.e. $\|\bar{X}\bar{S}e - \bar{\mu}e\| \leq \theta\bar{\mu}$.

Proof: From Lemma 2, the new iterate $(\bar{x}, \bar{y}, \bar{s})$ satisfies

$$\bar{x}^T\bar{s} = n\bar{\mu} = n\sigma\mu,$$

so we have to prove that $\|\bar{X}\bar{S}e - \bar{\mu}e\| \leq \theta\bar{\mu}$.
For a given component $j \in \{1..n\}$, we have

$$\begin{aligned} \bar{x}_j\bar{s}_j - \bar{\mu} &= (x_j + \Delta x_j)(s_j + \Delta s_j) - \bar{\mu} \\ &= x_j s_j + (s_j \Delta x_j + x_j \Delta s_j) + \Delta x_j \Delta s_j - \bar{\mu} \\ &= x_j s_j + (-x_j s_j + \sigma\mu) + \Delta x_j \Delta s_j - \sigma\mu \\ &= \Delta x_j \Delta s_j. \end{aligned}$$

Proof (cont'd)

Thus, from Lemmas 6 and 7, we get

$$\begin{aligned} \|\bar{X}\bar{S}e - \bar{\mu}e\| &= \|\Delta X \Delta S e\| \\ &\leq \tfrac{\theta^2 + n(1-\sigma)^2}{2^{3/2}(1-\theta)}\mu \\ &\leq \sigma\theta\mu \\ &= \theta\bar{\mu}. \end{aligned}$$

A property of log function

Lemma 9 For all $\delta > -1$:

$$\ln(1+\delta) \leq \delta.$$

Proof:
Consider the function

$$f(\delta) = \delta - \ln(1+\delta)$$

and its derivative

$$f'(\delta) = 1 - \frac{1}{1+\delta} = \frac{\delta}{1+\delta}.$$

Obviously $f'(\delta) < 0$ for $\delta \in (-1,0)$ and $f'(\delta) > 0$ for $\delta \in (0,\infty)$. Hence $f(.)$ has a *minimum* at $\delta = 0$. We find that $f(\delta = 0) = 0$. Consequently, for any $\delta \in (-1,\infty)$, $f(\delta) \geq 0$, i.e.

$$\delta - \ln(1+\delta) \geq 0.$$

$\mathcal{O}(\sqrt{n})$ Complexity Result

Theorem 10

Given $\epsilon > 0$, suppose that a feasible starting point $(x^0, y^0, s^0) \in N_2(0.1)$ satisfies

$$(x^0)^T s^0 = n\mu^0, \text{ where } \mu^0 \leq 1/\epsilon^\kappa,$$

for some positive constant κ. Then there exists an index K with $K = \mathcal{O}(\sqrt{n}\,\ln(1/\epsilon))$ such that

$$\mu^k \leq \epsilon, \quad \forall k \geq K.$$

$\mathcal{O}(\sqrt{n})$ Complexity Result

Proof: From Lemma 2, $\mu^{k+1} = \sigma\mu^k$. Having taken logarithms of both sides of this equality we obtain

$$\ln \mu^{k+1} = \ln \sigma + \ln \mu^k.$$

By repeatedly applying this formula and using $\mu^0 \leq 1/\epsilon^\kappa$, we get

$$\ln \mu^k = k \ln \sigma + \ln \mu^0 \leq k \ln(1 - \beta/\sqrt{n}) + \kappa \ln(1/\epsilon).$$

From Lemma 9 we have $\ln(1-\beta/\sqrt{n}) \leq -\beta/\sqrt{n}$. Thus

$$\ln \mu^k \leq k(-\beta/\sqrt{n}) + \kappa \ln(1/\epsilon).$$

To satisfy $\mu^k \leq \epsilon$, we need:

$$k(-\beta/\sqrt{n}) + \kappa \ln(1/\epsilon) \leq \ln \epsilon.$$

This inequality holds for any $k \geq K$, where

$$K = \frac{\kappa+1}{\beta} \cdot \sqrt{n} \cdot \ln(1/\epsilon).$$

Polynomial Complexity Result

Main ingredients of the polynomial complexity result for the short-step path-following algorithm:

Stay close to the central path:
all iterates stay in the $N_2(\theta)$ neighbourhood of the central path.

Make (slow) progress towards optimality:
reduce systematically duality gap

$$\mu^{k+1} = \sigma\mu^k,$$

where

$$\sigma = 1 - \beta/\sqrt{n},$$

for some $\beta \in (0,1)$.

Interior Point Methods

Theory: convergence in $\mathcal{O}(\sqrt{n})$ or $\mathcal{O}(n)$ iterations
Practice: convergence in $\mathcal{O}(\log n)$ iterations

Expected number of IPM iterations:

Problem Dimension	LP	QP
1,000	5 - 10	5 - 10
10,000	10 - 15	10 - 15
100,000	15 - 20	10 - 15
1,000,000	20 - 25	15 - 20
10,000,000	25 - 30	15 - 20
100,000,000	30 - 35	20 - 25
1000,000,000	35 - 40	20 - 25

... but one iteration may be expensive!

Reading about IPMs

S. Wright
Primal-Dual Interior-Point Methods, SIAM Philadelphia, 1997.

Gondzio
Interior point methods 25 years later,
European J. of Operational Research 218 (2012) 587–601.
`http://www.maths.ed.ac.uk/~gondzio/reports/ipmXXV.html`

Gondzio and Grothey
Direct solution of linear systems of size 10^9 arising in optimization with interior point methods, in: *Parallel Processing and Applied Mathematics PPAM 2005*, R. Wyrzykowski, J. Dongarra, N. Meyer and J. Wasniewski (eds.), *Lecture Notes in Computer Science*, 3911, Springer-Verlag, Berlin, 2006, pp 513–525.

OOPS: Object-Oriented Parallel Solver

`http://www.maths.ed.ac.uk/~gondzio/parallel/solver.html`

IPMs: Open Questions/Current Research

- iterative methods for indefinite linear systems (very close relation to PDEs)
- preconditioners for iterative methods
- extension to non-convex nonlinear problems

Interior Point Methods:

- Unified view of optimization
  → from LP via QP to NLP
- Predictable behaviour
  → small number of iterations
- Unequalled efficiency
  - competitive for small problems ($n \leq 10^6$)
  - beyond competition for large problems ($n \geq 10^6$)

Use IPMs in your research!