School of Mathematics

Interior-Point Methods
Why are they so efficient?

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

Sorry, you expected a different talk

The Return of Filter

by Sven Leyffer

... Filter Will Return

Outline

- Optimality Conditions for LP
- Simplex Method vs Interior Point Method
- IPM Framework: LP, QP, NLP, SDP
- Features of Logarithmic Function (Selfconcordant Barrier)
- From Sparse to Block-Sparse Problems
- Interaction with Differential Equations Techniques
  - Linear Algebra: Saddle Point Problem or KKT System
  - Predictor-Corrector Method
  - Krylov Subspace Correctors
- Conclusions

Primal-Dual Pair of Linear Programs

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

Lagrangian

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

Optimality Conditions

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

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

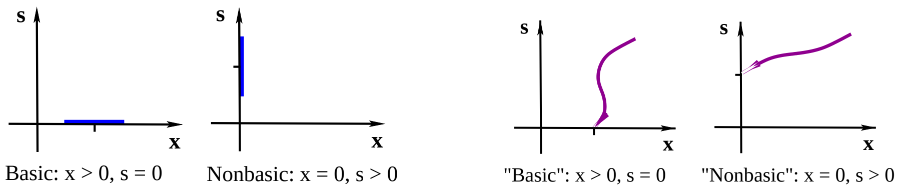
First Order Optimality Conditions

Simplex Method:

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

Interior Point Method:

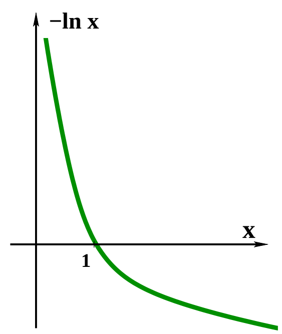
$$\begin{aligned} Ax &= b \\ A^T y + s &= c \\ XSe &= \mu e \\ x,\ s &\geq 0. \end{aligned}$$



Theory: IPMs converge in $\mathcal{O}(\sqrt{n})$ or $\mathcal{O}(n)$ iterations
Practice: IPMs converge in $\mathcal{O}(\log n)$ iterations
... but one iteration may be expensive!

Logarithmic barrier $-\ln x_j$

"replaces" the inequality $x_j \geq 0$.



Observe that

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

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.

Complementarity

Recall that the **Simplex Method** works with a partitioned formulation:

LP constraint matrix	$A = [B, N]$,	B is nonsingular
primal variables	$x = (x_B, x_N)$,	
reduced costs	$s = (s_B, s_N)$.	

The simplex method maintains the complementarity of primal and dual solution

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

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

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

What's wrong with the Simplex Method?

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 linear program with m constraints and n variables ($n \geq m$) has at most

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

vertices and the simplex method can make a non-polynomial number of iteratio to reach the optimality.

V. Klee and G. Minty's example LP: simplex method needs 2^n iterations
How good is the simplex algorithm,
in: Inequalities-III, O. Shisha, ed., Academic Press, 1972, 159–175.

Central Trajectory

Parameter μ controls the distance to optimality.

$$c^T x - b^T y = c^T x - x^T A^T y = x^T (c - A^T y) = x^T s = n\mu.$$

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

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

that satisfies the **first order optimality conditions**.

The path

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

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

Use Logarithmic Barrier

Primal Problem

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

Dual Problem

$$\begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y + s = c, \\ & s \geq 0. \end{array}$$

Primal Barrier Problem

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

Dual Barrier Problem

$$\begin{array}{ll} \max & b^T y + \sum_{j=1}^{n} \ln s_j \\ \text{s.t.} & A^T y + s = c, \end{array}$$

Newton Method

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 an application 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*.

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:

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

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

Stationarity: $\nabla_x L(x, y, \mu) = c - A^T y - \mu X^{-1} e = 0$

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

The **First Order Optimality Conditions** are:

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

IPM for QP

$$\begin{array}{ll} \min & c^Tx + \frac{1}{2}x^T\mathbf{Q}\,x \\ \text{s.t.} & Ax = b, \\ & x \geq 0. \end{array} \quad \rightarrow \quad \begin{array}{ll} \min & c^Tx + \frac{1}{2}x^T\mathbf{Q}\,x - \mu\sum_{j=1}^n \ln x_j \\ \text{s.t.} & Ax = b, \end{array}$$

The first order conditions (for the barrier problem)

$$\begin{array}{rcl} Ax & = & b, \\ A^Ty + s - \mathbf{Q}x & = & c, \\ XSe & = & \mu e. \end{array}$$

Newton direction

$$\begin{bmatrix} A & 0 & 0 \\ -\mathbf{Q} & A^T & I \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} \xi_p \\ \xi_d \\ \xi_\mu \end{bmatrix} = \begin{bmatrix} b - Ax \\ c - A^Ty - s + \mathbf{Q}x \\ \mu e - XSe \end{bmatrix}.$$

Augmented system

$$\begin{bmatrix} -\mathbf{Q} - \Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} \xi_d - X^{-1}\xi_\mu \\ \xi_p \end{bmatrix}.$$

IPM for NLP

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) + z = 0 \\ & z \geq 0. \end{array} \quad \rightarrow \quad \begin{array}{ll} \min & f(x) - \mu\sum_{i=1}^m \ln z_i \\ \text{s.t.} & g(x) + z = 0, \end{array}$$

Lagrangian: $L(x,y,z,\mu) = f(x) + y^T(g(x)+z) - \mu\sum_{i=1}^m \ln z_i.$

The first order conditions (for the barrier problem)

$$\begin{array}{rcl} \nabla f(x) + \nabla g(x)^T y & = & 0, \\ g(x) + z & = & 0, \\ YZe & = & \mu e. \end{array}$$

Newton direction

$$\begin{bmatrix} Q(x,y) & A(x)^T & 0 \\ A(x) & 0 & I \\ 0 & Z & Y \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} -\nabla f(x) - A(x)^Ty \\ -g(x) - z \\ \mu e - YZe \end{bmatrix}.$$

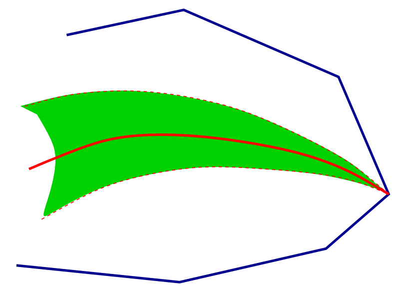
Augmented system

$$\begin{bmatrix} Q(x,y) & A(x)^T \\ A(x) & -ZY^{-1} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} -\nabla f(x) - A(x)^Ty \\ -g(x) - \mu Y^{-1}e \end{bmatrix} \quad \text{where} \quad \begin{array}{rcl} A(x) & = & \nabla g \\ Q(x,y) & = & \nabla^2_{xx}L \end{array}$$

Follow the Central Path

$$\begin{array}{rcl} Ax & = & b, \\ A^Ty + s & = & c, \\ XSe & \approx & \mu e, \end{array} \qquad \text{i.e.} \quad \|XSe - \mu e\| \leq \theta\mu,$$

where $\theta \in (0,1)$ and the barrier μ satisfies $x^Ts = n\mu$.



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

Progress to optimality

Reduce the barrier: $\mu^{k+1} = \sigma\mu^k$, where $\sigma = 1 - \beta/\sqrt{n}$ for some $\beta \in (0,1)$.
Compute 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 make step.

At the new iterate $(x^{k+1}, y^{k+1}, s^{k+1}) = (x^k, y^k, s^k) + (\Delta x^k, \Delta y^k, \Delta s^k)$
duality gap is reduced $1 - \beta/\sqrt{n}$ times.

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

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

After $C \cdot \sqrt{n}$ iterations, the reduction is $e^{-C\beta}$.
For sufficiently large constant C the duality gap becomes arbitrarily small.
Hence this algorithm has complexity $\mathcal{O}(\sqrt{n})$.

Primal-Dual Pair of SDPs

Primal

$$\begin{array}{ll} \min & C \bullet X \\ \text{s.t.} & A_i \bullet X = b_i, \; i = 1..m \\ & X \succeq 0; \end{array}$$

Dual

$$\begin{array}{ll} \max & b^T y \\ \text{s.t.} & \sum_{i=1}^m y_i A_i + S = C, \\ & S \succeq 0, \end{array}$$

where $A_i \in \mathcal{SR}^{n \times n}$, $b \in \mathcal{R}^m$, $C \in \mathcal{SR}^{n \times n}$ are given;
and $X, S \in \mathcal{SR}^{n \times n}$, $y \in \mathcal{R}^m$ are the variables.

SDP Example: stabilizing a differential equation

Let $A(x) = A_0 + x_1 A_1 + \ldots + x_k A_k$, where $A_i \in \mathcal{R}^{n \times n}$ and $A_i = A_i^T$.
Choose $x \in \mathcal{R}^k$ to minimize the maximum eigenvalue of $A(x)$.
Observe that $\lambda_{max}(A(x)) \le t$ if and only if $tI - A(x) \succeq 0$.
So we get the SDP in the *dual* form:

$$\begin{array}{ll} \max & -t \\ \text{s.t.} & tI - A(x) \succeq 0, \end{array}$$

where the variable is $y := (t, x)$.

Logarithmic Barrier Function for the cone $\mathcal{SR}_+^{n \times n}$ of positive definite matrices, $f : \mathcal{SR}_+^{n \times n} \mapsto \mathcal{R}$

$$f(X) = \begin{cases} -\ln \det X & \text{if } X \succ 0 \\ +\infty & \text{otherwise.} \end{cases}$$

LP: Replace $x \ge 0$ with $-\mu \sum_{j=1}^n \ln x_j$.

SDP: Replace $X \succeq 0$ with $-\mu \sum_{j=1}^n \ln \lambda_j = -\mu \ln(\prod_{j=1}^n \lambda_j)$.

Nesterov and Nemirovskii, *Interior Point Polynomial Algorithms in Convex Programming: Theory and Applications*, SIAM, Philadelphia, 1994.

IPM for SDP

$$\begin{array}{ll} \min & C \bullet X \\ \text{s.t.} & \mathcal{A}X = b \\ & X \succeq 0. \end{array} \quad \rightarrow \quad \begin{array}{ll} \min & C \bullet X + \mu f(X) \\ \text{s.t.} & \mathcal{A}X = b \end{array}$$

where $\mathcal{A}X = (A_i \bullet X)_{i=1}^m \in \mathcal{R}^m$ and $\mathcal{A}^* y = \sum_{i=1}^m y_i A_i$.

Self-concordant Barrier

Def: Let $C \in \mathcal{R}^n$ be an open nonempty convex set.
Let $f : C \mapsto \mathcal{R}$ be a three times continuously differentiable convex function.
A function f is called **self-concordant** if there exists a constant $p > 0$ such tha

$$|\nabla^3 f(x)[h,h,h]| \le 2p^{-1/2} (\nabla^2 f(x)[h,h])^{3/2},$$

$\forall x \in C$, $\forall h : x + h \in C$. (We then say that f is p-self-concordant).

Note that a self-concordant function is always well approximated by the quadrat model because the error of such an approximation can be bounded by the 3/ power of $\nabla^2 f(x)[h,h]$.

Lemma The barrier function $-\log x$ is self-concordant on $\mathcal{R}_+$.
Proof: Compute: $f'(x) = -x^{-1}$, $f''(x) = x^{-2}$ and $f'''(x) = -2x^{-3}$
and check that the self-concordance condition is satisfied for $p = 1$.

Use self-concordant barriers in optimization

Linear Matrix Inequalities

Def. A matrix $H \in \mathcal{R}^{n \times n}$ is positive semidefinite if $x^T H x \ge 0$ for any $x \ne 0$.
We write $H \succeq 0$.
Def. A matrix $H \in \mathcal{R}^{n \times n}$ is positive definite if $x^T H x > 0$ for any $x \ne 0$.
We write $H \succ 0$.

We denote with $\mathcal{SR}^{n \times n}$ the set of symmetric positive semidefinite matrices.
We denote with $\mathcal{SR}_+^{n \times n}$ the set of symmetric positive definite matrices.

Let $U, V \in \mathcal{SR}^{n \times n}$.
Define the inner product between U and V as $U \bullet V = \mathrm{trace}(U^T V)$,
where $\mathrm{trace}(H) = \sum_{i=1}^n h_{ii}$.
The associated Frobenius norm writes $\|U\|_F = (U \bullet U)^{1/2}$ (or just $\|U\|$).

Def. *Linear Matrix Inequalities*
Let $U, V \in \mathcal{SR}^{n \times n}$.
Write $U \succeq V$ iff $U - V \succeq 0$ (write $U \succ V$ iff $U - V \succ 0$).
Write $U \preceq V$ iff $U - V \preceq 0$ (write $U \prec V$ iff $U - V \prec 0$).

Direct Methods: Symmetric LDL^T Factorization

Indefinite	**Quasidefinite**	**Positive Definite**
$H = \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix}$	$H = \begin{bmatrix} Q & A^T \\ A & -R \end{bmatrix}$	$H = AQ^{-1}A^T$
2×2 pivots needed	1×1 pivots (any sign)	1×1 pivots (positive)
$\begin{bmatrix} 0 & a \\ a & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & a \\ a & d \end{bmatrix}$	strongly factorizable	easy

Vanderbei, *SIOPT* (1995): Symmetric QDFM's are strongly factorizable.
For any quasidefinite matrix there exists a **Cholesky-like** factorization

$$\bar{H} = LDL^T,$$

where D is **diagonal** but **not positive definite**:
D has n negative pivots and m positive pivots.

Minimum Degree Ordering

Sparse Matrix / **Pivot h_{11}** / **Pivot h_{22}**

$$H = \begin{bmatrix} x & & x & x & x & \\ & x & & & x & \\ x & & x & & & x \\ x & & & x & & x \\ x & x & & & x & \\ & & x & x & & x \end{bmatrix} \quad \begin{bmatrix} \mathbf{p} & & \mathbf{x} & \mathbf{x} & \mathbf{x} & \\ & x & & & x & \\ \mathbf{x} & & x & \mathbf{f} & \mathbf{f} & x \\ \mathbf{x} & & \mathbf{f} & x & \mathbf{f} & x \\ \mathbf{x} & x & \mathbf{f} & \mathbf{f} & x & \\ & & x & x & & x \end{bmatrix} \quad \begin{bmatrix} x & & x & x & x & \\ & \mathbf{p} & & & \mathbf{x} & \\ x & & x & & & x \\ x & & & x & & x \\ x & \mathbf{x} & & & x & \\ & & x & x & & x \end{bmatrix}$$

Minimum degree ordering:
choose a diagonal element corresponding to a row with the *min* number of nonzeros.
Permute rows and columns of H accordingly.

IPM for SDP

Lagrangian: $L(X,y,S) = C \bullet X + \mu f(X) - y^T(\mathcal{A}X - b)$,

The first order conditions (for the barrier problem)

$$C + \mu f'(X) - \mathcal{A}^* y = 0.$$

Use $f(X) = -\ln\det(X)$ and $f'(X) = -X^{-1}$.
Therefore the FOC become:

$$C + \mu X^{-1} - \mathcal{A}^* y = 0.$$

Denote $S = \mu X^{-1}$, i.e., $XS = \mu I$.
X is positive definite matrix hence its inverse is also positive definite.
The FOC now become:

$$\begin{aligned} \mathcal{A}X &= b, \\ \mathcal{A}^* y + S &= C, \\ XS &= \mu I, \end{aligned}$$

with $X \succ 0$ and $S \succ 0$.

Optimality Conditions:

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

Newton Direction:

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} \xi_p \\ \xi_d \\ \xi_\mu \end{bmatrix}.$$

Linear Algebra involves an (ill-conditioned) scaling matrix $\Theta = XS^{-1}$.

Augmented System vs Normal Equations

LP

$$\begin{bmatrix} \Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f \\ d \end{bmatrix}$$

$$(A\Theta A^T)\Delta y = g$$

QP

$$\begin{bmatrix} Q+\Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f \\ d \end{bmatrix}$$

$$(A(Q+\Theta^{-1})^{-1}A^T)\Delta y = g$$

NLP

$$\begin{bmatrix} Q(x,y) & A(x)^T \\ A(x) & -ZY^{-1} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f \\ d \end{bmatrix}$$

$$(AQ^{-1}A^T + ZY^{-1})\Delta y = g$$

Predictor-Corrector

Mehrotra, *SIOPT* 2 (1992) pp. 575-601.

Split the right hand side of the Newton direction:

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

and compute two steps:

- put $\mu = 0$ to compute the **predictor** Δ_p,
- "guess" μ and compute the **corrector** Δ_c.

Then combine two components:

$$\Delta = \Delta_p + \Delta_c.$$

Mehrotra's Predictor-Corrector

The third equation in Affine-Scaling Direction satisfies:

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

If a full step in this direction is made, then the new complementarity product is

$$\begin{aligned} \bar{X}\bar{S}e &= (X + \Delta X)^T (S + \Delta S)e \\ &= XSe + (S\Delta x + X\Delta s) + \Delta X \Delta S e \\ &= XSe - XSe + \Delta X \Delta S e \\ &= \Delta X \Delta S e. \end{aligned}$$

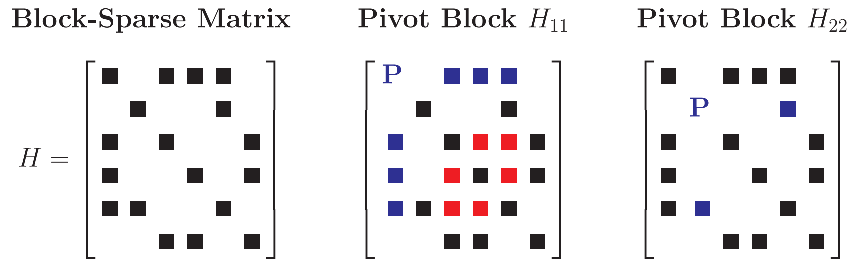
Hence the second component of direction comes from:

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

C. Cartis → new insight into Mehrotra's P-C method.

From Sparsity to Block-Sparsity:

Apply minimum degree ordering to **(sparse) blocks**:



Object-Oriented Parallel Solver → problems of size $\mathbf{10^6, 10^7, 10^8, 10^9}$,
G. & Sarkissian, *MP* 96 (2003) 561-584.
G. & Grothey, *SIOPT* 13 (2003) 842-864.
G. & Grothey, *AOR* (to appear).

Talk of **Andreas Grothey** later today: "How to solve QPs with 10^9 variables

Iterative Methods (with Indefinite Preconditioners)

LP $\begin{bmatrix} \Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f \\ d \end{bmatrix}$

QP $\begin{bmatrix} Q + \Theta^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f \\ d \end{bmatrix}$

NLP $\begin{bmatrix} Q(x,y) & A(x)^T \\ A(x) & -ZY^{-1} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f \\ d \end{bmatrix}$

It is important to keep Θ^{-1} in the preconditioner. Θ is ill-conditioned:

For **"basic"** variables: $\Theta_j = x_j/s_j \rightarrow \infty$ $\Theta_j^{-1} \rightarrow 0$;
For **"non-basic"** variables: $\Theta_j = x_j/s_j \rightarrow 0$ $\Theta_j^{-1} \rightarrow \infty$.

see my talk at SIAM Conference on Optimization, Stockholm, May 2005.

Optimization: KKT System

$$\begin{bmatrix} Q + \Theta_1^{-1} & A^T \\ A & -\Theta_2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f \\ d \end{bmatrix}$$

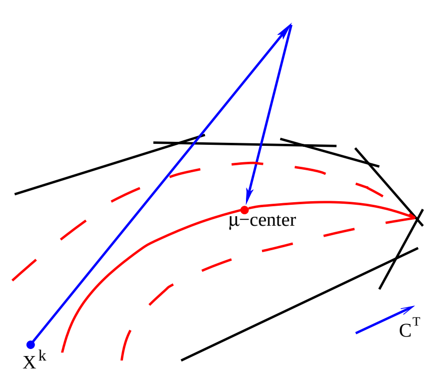
PDE: Saddle Point Problem

$$\begin{bmatrix} H & A^T \\ A & -C \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} f \\ d \end{bmatrix}$$

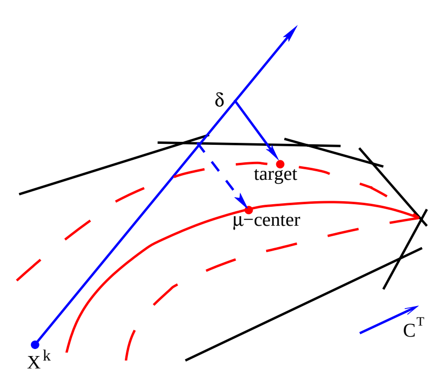
Benzi, Golub & Liesen, "Numerical Solution of Saddle Point Problems", *Acta Numerica* 2005 (to appear).

Interpretation of Correctors

Mehrotra's Predictor-Corrector



Multiple Centrality Correctors



Krylov Subspace Directions

Mehrotra and Li, *SIOPT* (2005) (to appear).

At iteration k set $\mu = 0$ and solve

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} b - Ax^k \\ c - A^T y^k - s^k \\ \mu e - X^k S^k e \end{bmatrix} = \begin{bmatrix} \xi_p \\ \xi_d \\ \xi_\mu \end{bmatrix}, \quad \text{i.e.} \quad H_k \Delta = \xi.$$

Compute:

$$\begin{aligned} \bar{x} &= x^k + \alpha_P \Delta x \\ \bar{y} &= y^k + \alpha_D \Delta y \\ \bar{s} &= s^k + \alpha_D \Delta s \end{aligned} \quad \text{and} \quad \bar{\xi} = \begin{bmatrix} \bar{\xi}_p \\ \bar{\xi}_d \\ \bar{\xi}_\mu \end{bmatrix} = \begin{bmatrix} b - A\bar{x} \\ c - A^T \bar{y} - \bar{s} \\ \mu e - \bar{X}\bar{S}e \end{bmatrix}.$$

At the trial point $(\bar{x}, \bar{y}, \bar{s})$ IPM would have to solve:

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ \bar{S} & 0 & \bar{X} \end{bmatrix} \begin{bmatrix} \Delta \bar{x} \\ \Delta \bar{y} \\ \Delta \bar{s} \end{bmatrix} = \begin{bmatrix} b - A\bar{x} \\ c - A^T \bar{y} - \bar{s} \\ \mu e - \bar{X}\bar{S}e \end{bmatrix}, \quad \text{i.e.} \quad \bar{H}\bar{\Delta} = \bar{\xi}.$$

Multiple Centrality Correctors

G. *COAP* 6 (1996) pp. 137–156.

Compute Newton's direction

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

combining it from two components:

$$\Delta = \Delta_p + \Delta_c.$$

Assume that a **predictor direction** is given and feasible stepsizes α_P and α
are determined.

We look for a **centrality corrector** such that larger steps will be made in a ne
composite direction $\Delta = \Delta_p + \Delta_c$.

We want to enlarge the stepsizes to

$$\tilde{\alpha}_P = \min(\alpha_P + \delta, 1) \quad \text{and} \quad \tilde{\alpha}_D = \min(\alpha_D + \delta, 1),$$

respectively.

Multiple Centrality Correctors

Compute a **trial point**

$$\begin{aligned} \tilde{x} &= x + \tilde{\alpha}_P \Delta_p x, \\ \tilde{s} &= s + \tilde{\alpha}_D \Delta_p s. \end{aligned}$$

and the corresponding complementarity products

$$\tilde{v} = \tilde{X}\tilde{S}e \in \mathcal{R}^n.$$

Correct only the **outliers**:

move small products $(\tilde{x}_j \tilde{s}_j \leq \gamma_{min}\mu)$ to $\gamma_{min}\mu$;

move large products $(\tilde{x}_j \tilde{s}_j \geq \gamma_{max}\mu)$ to $\gamma_{max}\mu$.

Technique used by

BPMPD, Cplex, HOPDM, OOPS, OOQP, PCx, XPress.

New results obtained recently:

M. Colombo → significant computational improvements (talk later today

Define **Krylov Subspace** for $\bar{H}\bar{\Delta} = \bar{\xi}$. Precondition $\bar{H}$ with H_k.

$$K_j(H_k, \bar{H}, \bar{\xi}\,) := \mathrm{span}\{\xi_H, G\xi_H, G^2\xi_H, \ldots, G^j\xi_H\},$$

where $\xi_H = H_k^{-1}\bar{\xi}$ and $G = I - H_k^{-1}\bar{H}$.
Observe that

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \end{bmatrix} H_k^{-1} \begin{bmatrix} \bar{\xi}_p \\ \bar{\xi}_d \\ \bar{\xi}_\mu \end{bmatrix} = \begin{bmatrix} \bar{\xi}_p \\ \bar{\xi}_d \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \end{bmatrix} G^i H_k^{-1} \begin{bmatrix} \bar{\xi}_p \\ \bar{\xi}_d \\ \bar{\xi}_\mu \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Idea: Use the affine-scaling direction Δ_{aff},
the first j directions from $K_j(H_k, \bar{H}, \bar{\xi}\,)$ that is $\Delta^0, \Delta^1, \ldots, \Delta^j$,
and the centering direction Δ_{cen}
and combine them:

$$\Delta = \Delta_{aff} + \sum_{i=0}^{j} \rho_i \Delta^i + \rho_{cen}\Delta_{cen}.$$

Choose the scalars ρ_i and ρ_{cen} to satisfy the following objectives:

- reduce duality (complementarity) gap,
- produce well-centered point.

Conclusions:

- Interior Point Methods are the key optimization technique.
- The theory of IPMs is well understood.
- IPMs demonstrate spectacular efficiency.
- Today IPMs can solve problems of dimension $\mathbf{10^9}$.

Numerical analysis keeps inspiring optimization field

A development of new preconditioners for IPMs is a challenge.

A plea to:

David Griffiths and Alistair Watson

The community needs the Dundee NA Conference!

Thank you for your attention!