Primal-Dual Interior-Point Methods

I LP 1980's

II SDP 1990's.

Pre-history: Khaciyan's polynomial-time alg for LP 1979
"Ellipsoid method" — $O(n^6 L)$
L = bitlength of rational data in A, b.

Karmarkar's polynomial-time alg for LP 1984
"Projective Method":
Much more practical.

Soon realized that this was closely related to barrier methods for NLP: Fiacco + McCormick 1960s.
But at that time no one thought of using these methods for LP: then the simplex method was pre-eminent.
See BV Ch. 11.

I LP.

$$\text{(P)} \quad \min \ c^T x \quad \text{s.t.} \ Ax = b, \ x \geq 0 \qquad A \ \underbrace{\boxed{}}_{}\ m \ (n \text{ columns})$$

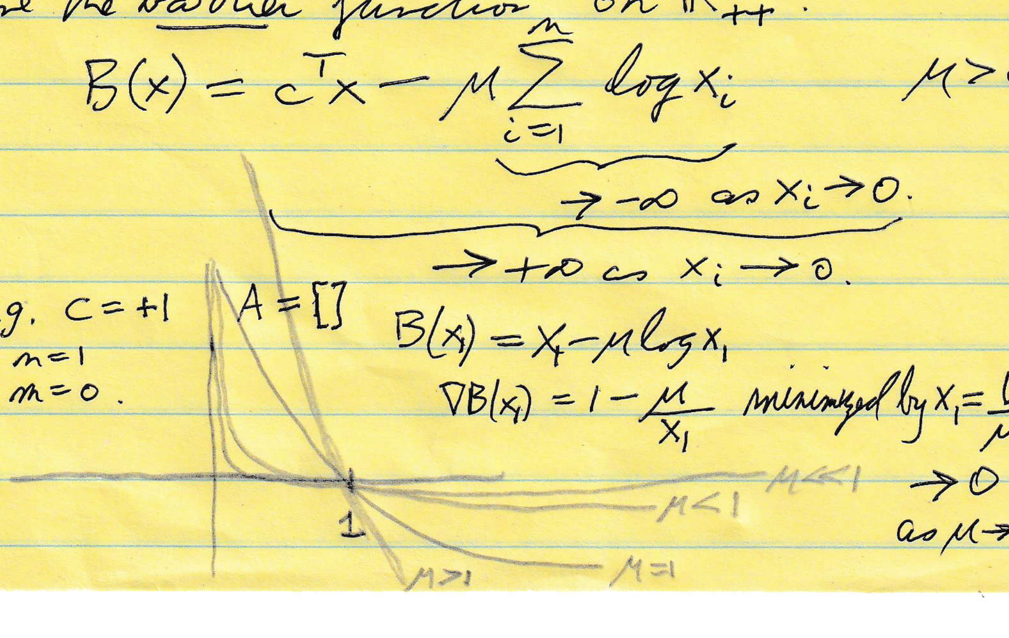
Define the barrier function on $\mathbb{R}^n_{++}$:

$$B(x) = \underbrace{c^T x - \mu \underbrace{\sum_{i=1}^{n} \log x_i}_{\to -\infty \text{ as } x_i \to 0}}_{\to +\infty \text{ as } x_i \to 0} \qquad \mu > 0.$$

e.g. $c = +1$, $A = [\,]$, $n=1$, $m=0$.

$B(x_1) = x_1 - \mu \log x_1$

$\nabla B(x_1) = 1 - \dfrac{\mu}{x_1}$ minimized by $x_1 = \dfrac{1}{\mu}$



$\to 0$ as $\mu \to \infty$

In general $\nabla B(x) = c - \mu \begin{bmatrix} 1/x_1 \\ \vdots \\ 1/x_n \end{bmatrix}$

Let $\underline{\underline{X = \mathrm{Diag}(x)}}$. Then

$$\nabla B(x) = c - \mu X^{-1} e \qquad e = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

Want minimizer of $B(x)$ s.t. $Ax = b$

KKT: $\nabla B(x) = A^T \nu$ $\qquad$ ν = multipliers (dual vars.)

i.e. $c - \mu X^{-1} e = A^T \nu$.

Newton's method for <u>unconstrained min</u> of B:

$$(\nabla^2 B(x^k)) \Delta x + \nabla B(x^k) = 0 \; ; \; x^{k+1} = x^k + \Delta x$$

but with the constraint $Ax = b$ it is

$$\begin{cases} (\nabla^2 B(x^k)) \Delta x + \nabla B(x^k) = A^T(\nu^k + \Delta \nu) \\ A(x^k + \Delta x) = b \end{cases}$$

Note $\nabla^2 B(x) = \mu \begin{bmatrix} 1/x_1^2 & & \\ & \ddots & \\ & & 1/x_n^2 \end{bmatrix} = \mu X^{-2}$.

so need to solve

$$(*) \quad \begin{bmatrix} -\mu X_k^{-2} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \nu \end{bmatrix} = \begin{bmatrix} c - \mu X_k^{-1} e - A^T \nu_k \\ b - A x^k \end{bmatrix}$$

where $X_k = \mathrm{Diag}(x^k)$.

This is equivalent to applying Newton's method for <u>nonlinear eqns</u> to the KKT system: $x^{k+1} = x^k + \Delta x$, $\nu^{k+1} = \nu^k + \Delta \nu$ where

$$F_\mu'(x^k, \nu^k)\begin{bmatrix}\Delta x\\ \Delta\nu\end{bmatrix} = -F_\mu(x^k,\nu^k) \quad \text{where}$$

$$F_\mu(x,\nu) = \begin{bmatrix} A^T\nu + \mu X^{-1}e - c \\ Ax - b\end{bmatrix} = 0 \qquad (X = \underline{\mathrm{Diag}(x)})$$

$$F_\mu : \mathbb{R}^{n+m} \longrightarrow \mathbb{R}^{n+m}.$$

To solve the linear system we can use block Gauss: subtract $-\mu^{-1}AX_k^2$ * 1^{st} eqn of (*) to 2^{nd} eq:

(eliminates Δx) $\quad \begin{bmatrix} 0 & \mu^{-1}\underbrace{AX_k^2A^T}_{\text{schur compl.}}\end{bmatrix}\begin{bmatrix}\Delta x\\ \nu\end{bmatrix} = b - Ax^k + \mu^{-1}AX_k^2c - X_k^{-1}e$

$\nu^k + \Delta\nu$

$m \times m$. po. def.

SCHUR COMPLEMENT $\longrightarrow$ $\mu^{-1}AX_k^2A^T\nu^{k+1} = RHS$

Use Cholesky to $\rightarrow$ Solve for ν^{k+1}, then get Δx by subst. into 1^{st} eqn.

$\mathrm{Diag}(\mu^{-1}(x_i^k)^2)$

k = iter count

$i = 1, \ldots, n$.

One problem: as $\mu \to 0$, some entries $\{x_i\} \to 0$ and others don't, so $AX_k^{-2}A^T$ is very ill conditioned in limit.

Also, what if $x^k + \Delta x \equiv x^{k+1}$ is not > 0? In that case, we set $x^{k+1} > 0$ by some rule. See HW.

This gives us PRIMAL INTERIOR POINT (primal barrier) METHOD. FOR LP.

But there are still many details not specified, including how to reduce $\mu \to 0$.

Now recall the dual LP:

$$\max\ b^T y$$
$$\text{s.t.}\quad A^T y \le c$$

or equivalently $A^T y + z = c$, $z \ge 0$

Compare with KKT cond on PD2:

$$c - \mu X^{-1} e = A^T v$$

Write $y \equiv v$,

$$z = \mu X^{-1} e$$

i.e. $z_i = \mu x_i^{-1}$, i.e. $x_i z_i = \mu$.

As $\mu \to 0$, get $x_i z_i \to 0$ COMPLEMENTARITY.

Now let's write the KKT eqns in (x, y, z)

CENTRAL PATH EQNS.

$$\tilde{F}_\mu(x,y,z) = \begin{bmatrix} A^T y + z - c \\ Ax - b \\ XZe - \mu e \end{bmatrix} = 0 \qquad \tilde{F}_\mu : \mathbb{R}^{2n+m} \to \mathbb{R}^{2n+m}$$

with $X = \mathrm{Diag}(x)$, $Z = \mathrm{Diag}(z)$

The central path is $\left\{ \underbrace{(x,y,z)}_{\in \mathbb{R}^n_+ \times \mathbb{R}^m \times \mathbb{R}^n_+} : F_\mu(x,y,z) = 0 \text{ for } \mu > 0 \right\}$

Idea: follow the central path by computing solutions (x^μ, y^μ, z^μ) and letting $\mu \to 0$.

For fixed μ, let's apply Newton step for nonlinear equations to $F_\mu(x,y,z) = 0$:

$$\tilde{F}_\mu'(x^k, y^k, z^k) \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = -\tilde{F}_\mu(x^k, y^k, z^k).$$

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ Z_k & 0 & X_k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} c - A^T y^k - z^k \\ b - Ax^k \\ \mu e - X_k Z_k e \end{bmatrix}$$

$\mathrm{Diag}(z_k) \rightarrow Z_k$

$\mathrm{Diag}(x_k) \rightarrow X_k$

Solve via block Gauss: subtract $X_k^{-1} * 3^{rd}$ eq from 1^{st} eq.

(eliminates Δz)
$$\begin{bmatrix} -X_k^{-1} Z_k & A^T & 0 \\ A & 0 & 0 \\ \multicolumn{3}{c}{\text{(unchanged, don't need)}} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} c - A^T y^k - \cancel{z^k} - \mu X_k^{-1} e + \cancel{Z_k} e \\ b - Ax^k \\ \text{(don't need)} \end{bmatrix}$$

Now add $A Z_k^{-1} X_k * 1^{st}$ eqn to 2^{nd} eq, to eliminate Δx:

$$(A D_k A^T) \Delta y = \text{RHS}$$

where diagonal matrix $D_k = \mathrm{Diag}(x_i^k / z_i^k)$
compared to $\mathrm{Diag}(\mu^{-1}(x_i^k)^2)$ for primal interior point.

Note: these are the same if (x^k, y^k, z^k) is on the central path! So $A D_k A^T$ is still ill-conditioned.

But one advantage of applying Newton to $\tilde{F}_\mu$ (instead of F_μ) is that as $\mu \to 0$,

$$\tilde{F}_\mu(x, y, z) = 0 \longrightarrow \begin{bmatrix} A^T y + z - c \\ Ax - b \\ XZe \end{bmatrix} = 0$$

KKT + complementarity.

while $F_\mu(x,\nu)$ is not defined in limit $\mu \to 0$
since X^{-1} blows up on central path as $\mu \to 0$.

After solving Schur compl. eqn for Δy,
get Δx, Δz by substitution.
But what if $x^k + \Delta x \notin \mathbb{R}^n_+$ or $z^k + \Delta z \notin \mathbb{R}^n_+$?
As before, need to get $x^{k+1}, z^{k+1} \in \mathbb{R}^n_+$ by
some rule (see HW).

This gives the basic motivation + idea for
the primal-dual interior point method for LP.
But there are still many details, especially, how $\mu \to 0$?

Practical methods: many are based on
Mehrotra's predictor-corrector method -
a heuristic that is very efficient.

But to guarantee a polynomial-time complexity,
need to consider more complicated update
rules.
"Short-step" methods: have best complexity
"Long-step" methods: more work in theoretical
worst case, but more efficient in practice.

Details: Steve Wright's wonderful book.
(Also Renegar's book).

$\boxed{\text{X no longer diagonal}}$

II SDP. $\min_{X \in S^n} \langle C, X\rangle$

s.t. $\mathcal{A}X = b$

$X \succeq 0$

means
$\langle A_1, X\rangle = b_1$
$\vdots$
$\langle A_m, X\rangle = b_m$

where $\mathcal{A}$ is linear operator: $S^n \to \mathbb{R}^m$.

Define the barrier function on S^n_{++}:

$$B(X) = \underbrace{\langle C, X\rangle - \mu \log \det X}_{\to -\infty \text{ as } X \to \text{bd } S^n_+}$$

determinant.

∇B, $\nabla^2 B$: can define as earlier but now $X \in S^n$.

$$\nabla B(X) = C - \mu X^{-1} \in S^n$$

$\nabla \log\det(X)$: see p. 641–642 BV

Want minimizer of $B(X)$ s.t. $\mathcal{A}X = b$

KKT: $\nabla B(X) = \mathcal{A}^T \nu$ $\quad \mathcal{A}^T : \mathbb{R}^m \to S^n$ (the adjoint).

$$C - \mu X^{-1} = \mathcal{A}^T \nu$$

MEANS $\mathcal{A}^T \nu = \sum_{i=1}^m \nu_i A_i$

Again compare with dual SDP:

$\max b^T y$

s.t. $\mathcal{A}^T y + Z = C$, $Z \succeq 0$

$\nu \equiv y$, $\quad Z = \mu X^{-1}$, i.e. $\underline{XZ = \mu I}$

Central Path Eqns $\quad \tilde{\tilde{F}}_\mu(X,y,Z) = \begin{bmatrix} A^T y + Z - C \\ AX - b \\ XZ - \mu I \end{bmatrix} = 0$

$$\tilde{\tilde{F}}_\mu : S^n \times \mathbb{R}^m \times S^n \to S^n \times \mathbb{R}^m \times S^n.$$

Again we'll follow the central path by applying Newton's method. Linearize the 3rd eqn by:

$$(X_k + \Delta X)(Z_k + \Delta Z) = \mu I$$

Let's write this on $\mathbb{R}^{n^2}$ instead of S^n: we make use of the vec operator: $\mathrm{vec}(X) = [x_{11} \ldots x_{n1} x_{21} \ldots \ldots x_{nn}]^T$ (stack columns of X, ignoring the symmetry) + Kronecker products:

$$M \otimes N = \begin{bmatrix} m_{11}N & \ldots & m_{1n}N \\ \vdots & & \\ m_{n1}N & \ldots & m_{nn}N \end{bmatrix}$$

Main rule of Kronecker products:

$$\mathrm{vec}(ABC) = (C^T \otimes A)\,\mathrm{vec}(B).$$

so

$$X_k Z_k + (\Delta X) Z_k + X_k(\Delta Z) + \cancel{O(\|\Delta X\| \|\Delta Z\|)} = \mu I$$

$$(Z_k \otimes I)\,\mathrm{vec}(\Delta X) + (I \otimes X_k)\,\mathrm{vec}(\Delta Z) = \mathrm{vec}(\mu I - X_k Z_k)$$

So, linearizing $\tilde{\tilde{F}}_\mu$ get

PD9

$$(\dagger)\quad \begin{bmatrix} 0 & \mathcal{A}^T & I\otimes I \\ \mathcal{A} & 0 & 0 \\ Z_k\otimes I & 0 & I\otimes X_k \end{bmatrix} \begin{bmatrix} vec(\Delta X) \\ \Delta y \\ vec(\Delta Z) \end{bmatrix} = \begin{bmatrix} vec(C-\mathcal{A}^T y_k - Z_k) \\ b-\mathcal{A}X_k \\ vec(\mu I - X_k Z_k) \end{bmatrix}$$

Block Gauss gives Schur complement eqn

$$(\mathcal{A}\,\mathcal{E}\,\mathcal{A}^T)\,\Delta y = \ldots$$

where
$$\mathcal{E} = (Z_k\otimes I)^{-1}(I\otimes X_k) = (Z_k^{-1}\otimes I)(I\otimes X_k) = Z_k^{-1}\otimes X_k$$

Thus we form $M = \mathcal{A}\underbrace{\mathcal{E}\mathcal{A}^T}_{B}$ by $M = \mathcal{A}B$

where j^{th} column of B is

$$(Z_k^{-1}\otimes X_k)\,vec(A_j) = vec(X_k A_j Z_k^{-1})$$

Thus $M_{ij} = \operatorname{tr} A_i X_k A_j Z_k$

k is iter # $\qquad i,j=1,\ldots,m$

Then solve for Δy

get $\Delta X, \Delta Z$ by substitution in $(\dagger)$.

But there is a problem: ΔX is generally not symmetric, because we applied Newton's method in $\mathbb{R}^{n^2}$, NOT S^n!!!

Note: product of two symmetric matrices is not symmetric in general: $(AB)^T = B^T A^T = BA \neq AB$.

What to do?

1. "XZ" method: replace ΔX by $(\Delta X + (\Delta X)^T)/2$ (HKM method). Easy, not appealing?
2. "XZ+ZX" method (AHO method) replace 3rd eqn in def of $\tilde{\tilde{F}}$ by
$$XZ + ZX = 2\mu I.$$
Apply Newton in S^n:
$$(X_k + \Delta X)(Z_k + \Delta Z) + (Z_k + \Delta Z)(X_k + \Delta X) = 2\mu I$$
Leads to a method with nice properties but more expensive to implement: $M \neq M^T$.
3. "Nesterov-Todd" method (NT method) Beautiful, more complicated.

For years, most codes used XZ method. But now, MOSEK uses NT method.

CVX supports
SeDuMi: NT like
SDPT3: NT or XZ (formerly AHO)
MOSEK: NT

MORE ISSUES

What if $X_k + \Delta X$ or $Z_k + \Delta Z \not\succ 0$?
Need to contract step. See HW.

How let $\mu \to 0$?

And for both LP + SDP, how to ensure eventually get feasible? Will get and stay feasible if ever take unit step $X_k + \Delta X, Z_k + \Delta Z$.

To guarantee eventually get feasible: "homogeneous self dual".

References for more: M.J. Todd, Acta Numerica, 2001.