

Primal-Dual Interior-Point Methods

I LP 1980's

II SDP 1990's.

Pre-history: Khachiyan's polynomial-time alg for LP 1979
 "Ellipsoid method" — $O(n^6 L)$

$L = \text{bitlength of rational data in } A, b.$

Karmarkar's polynomial-time alg for LP 1984
 "Projective Method":

Much more practical.

Soon realized that this was closely related to
barrier method 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.

$$\begin{array}{ll} \text{I} & \text{LP.} \\ (\text{P}) & \text{min } c^T x \\ & \text{s.t. } Ax = b \\ & \quad x \geq 0 \end{array}$$

Define the barrier function on \mathbb{R}_{++}^n :

$$B(x) = c^T x - \mu \sum_{i=1}^n \log x_i \quad \mu > 0.$$

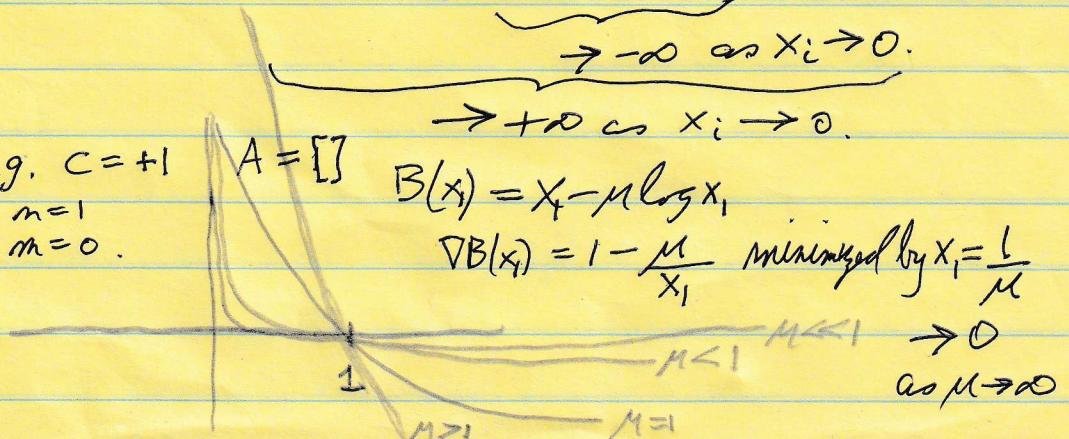
$\underbrace{\quad}_{\rightarrow -\infty \text{ as } x_i \rightarrow 0.}$

$\underbrace{\quad}_{\rightarrow +\infty \text{ as } x_i \rightarrow 0.}$

$$\begin{aligned} \text{e.g. } c &= +1 \\ m &= 1 \\ m &= 0. \end{aligned}$$

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

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



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

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

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

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

KKT:

$$\nabla B(x) = A^T v$$

v = multipliers
(dual vars.)

i.e.

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

Newton's method for unconstrained min of B :

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

but with the constraint $Ax = b$ it is

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

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

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

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

This is equivalent to applying

Newton's method to nonlinear eqns to

the KKT system: $x^{k+1} = x^k + \Delta x, v^{k+1} = v^k + \Delta v$ where

$$F_M'(x^k, v^k) \begin{bmatrix} \Delta x \\ \Delta v \end{bmatrix} = -F_M(x^k, v^k) \text{ where}$$

$$F_M(x, v) = \begin{bmatrix} A^T r + \mu \bar{X}^T e - c \\ Ax - b \end{bmatrix} = 0$$

$(\bar{X} = \text{Diag}(x))$

$$F_M: \mathbb{R}^{m \times m} \longrightarrow \mathbb{R}^{m \times m}$$

To solve the linear system we can use block Gauss:

subtract $-\bar{\mu}^{-1} \bar{A} \bar{X}_k^2 * 1^{\text{st}} \text{ eqn of } (*) \text{ to } 2^{\text{nd}} \text{ eqn:}$

$$(\text{eliminate } \Delta x) \quad \left[\begin{array}{cc} 0 & \bar{\mu}^{-1} \bar{A} \bar{X}_k^2 A^T \\ & \underbrace{\bar{\mu}^{-1} \bar{A} \bar{X}_k^2 A^T}_{\text{tehur compl.}} \end{array} \right] \begin{bmatrix} \Delta x \\ v \end{bmatrix} = b - Ax^k + \bar{\mu}^{-1} \bar{A} \bar{X}_k^2 c - \bar{X}_k^T e$$

$$\xrightarrow{\text{SCHUR COMPLEMENT}} \boxed{\bar{\mu}^{-1} \bar{A} \bar{X}_k^2 A^T v^{k+1} = \text{RHS}}$$

\rightarrow solve for v^{k+1} , then get Δx by subst. into 1^{st} eqn.

One problem: as $\mu \rightarrow 0$, some entries $\{x_{ij}\} \rightarrow 0$ and others don't, or $\bar{A} \bar{X}_k^2 A^T$ 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 zone rule.
See HW.

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

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

$\square \rightarrow \square$
 $m \times m$
no def.
Use Cholesky to

$\text{Diag}(\bar{\mu}(x_i^k)^2)$
 $k = \text{iter count}$
 $i = 1, \dots, n$

Now recall the dual LP:

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

$$\text{or equivalently } A^T y + z = c, \quad z \geq 0$$

Compare with KKT cond on PDZ:

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

$$\text{Write } y = v,$$

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

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

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

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

CENTRAL PATH EQNS.

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

$$\tilde{F}_\mu: \mathbb{R}^{2n+m} \rightarrow \mathbb{R}^{2n+m}$$

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

The central path is $\{(x, y, z) : \tilde{F}_\mu(x, y, z) = 0 \text{ for } \mu > 0\}$

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

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

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

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ 0 & 0 & X_k^{-1} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} c - A^T y^k - z^k \\ b - Ax^k \\ \mu e - X_k^{-1} Z_k^k e \end{bmatrix}$$

$\xrightarrow{\text{Diag}(Z_k)} \quad \xrightarrow{\text{Diag}(x_k)}$

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

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

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

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

where diagonal matrix $D_k = \text{Diag}(x_i^k / z_i^k)$
compared to $\text{Diag}(\bar{\mu}^k (x_i^k)^2)$ 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}_μ (instead
of F_μ) is that as $\mu \rightarrow 0$,

$$\tilde{F}_\mu(x, y, z) = 0 \rightarrow \begin{bmatrix} A^T y + z - c \\ Ax - b \\ X_k^{-1} Z_k^k e \end{bmatrix} = 0$$

KKT + complementarity.

while $F_n(x, \nu)$ is not defined in limit $\mu \rightarrow 0$
 since \tilde{X}^l blows up on central path as $\mu \rightarrow 0$.

After solving Schur compl. eqn for Δy ,
 get $\Delta x, \Delta z$ by substitution.

But what if $x^k + \Delta x \notin \mathbb{R}_+^n$ or $z^k + \Delta z \notin \mathbb{R}_+^m$?

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 $\mu \rightarrow 0$?

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 Resen's book).

FD7

II SDP.

$$\begin{array}{ll} \min_{X \in S^n} & \langle C, X \rangle \\ \text{s.t.} & AX = b \\ & X \succeq 0 \end{array}$$

X no longer diagonal

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

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

Define the barrier function on S_+^n :

$$B(X) = \langle C, X \rangle - \mu \log \det X$$

$\underbrace{}_{\text{determinant.}}$

$\rightarrow -\infty$ as $X \rightarrow \text{bd } S_+^n$.

$\nabla B, \nabla^2 B$: 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. $AX = b$

KKT: $\nabla B(X) = A^T v \quad A^T: \mathbb{R}^m \rightarrow S^n$

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

(the adjoint).

MEANS
 $A^T v = \sum_{i=1}^m v_i A_i$

Again compare with dual SDP:

$\max b^T y$

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

$v = y, \quad Z = \mu X^{-1}, \quad \text{i.e. } XZ = \mu I$

Central Path Eqs

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

$$\tilde{F}_m: S^n \times \mathbb{R}^m \times S^n \rightarrow S^n \times \mathbb{R}^m \times S^n.$$

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}^{m^2} instead of S^n : we make use of the vec operator: $\text{vec}(X) = [x_{11} \dots x_{1n}, x_{21} \dots x_{2n}]^T$ (stack columns of X , ignoring the symmetry) + Kronecker products.

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

Main rule of Kronecker products:

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

so

$$X_k Z_k + (\Delta X) Z_k + X_k (\Delta Z) + O(\|AX\| \|Z\| \|A\|) = \mu I$$

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

So, linearizing \tilde{F}_m get

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

Block Gauss gives Schur complement eqn

$$(A \otimes A^T) \Delta y = \dots$$

$$\begin{aligned} \text{where } E &= (Z_k \otimes I)^{-1} (I \otimes X_k) \\ &= (Z_k^{-1} \otimes I) (I \otimes X_k) \\ &= Z_k^{-1} \otimes X_k \end{aligned}$$

$$\text{Thus we form } M = A \underbrace{E A^T}_B \text{ by } M = AB$$

where j^{th} column of B is

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

$$\text{Thus } M_{ij} = \underset{k \text{ is iter #}}{\text{tr}} A_i X_k A_j Z_k^{-1} \quad i, j = 1, \dots, m$$

Then solve for Δy

Set $\Delta X, \Delta Z$ by substitution in (+).

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.¹
2. "XZ + ZX" method (AHO method)
replace 3rd eqn in def of \tilde{F} by
 $XZ + ZX = 2\mu I$.

Apply Newton in S^m :

$$(X_k + \Delta X)(Z_k + \Delta Z) + (Z_k + \Delta Z)(X_k + \Delta X) = 2\mu I$$

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.

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

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

How let $\mu \rightarrow 0$?

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.

MORE ISSUES