Primal–Dual Interior Point Methods

Key idea: follow central path WITHOUT solving repeated minimization problems.

Developed extensively in the late 80s + early 90s for LP, + extended to SDP and SOCP in the mid 90s.

Text continues to use the general convex programming framework, but we'll switch to addressing LP and SDP specifically — our derivation is quite different from ~~that in the BV text~~.

(I). LP. $\min \; c^T x$

$$\text{s.t.} \quad Ax = b \qquad A \; \underset{m \times n}{\square}$$

$$x \geq 0$$

Pt on central path is characterized by $\nabla B(x) = A^T \nu$

where $$B(x) = \cancel{t}\, c^T x - \frac{1}{\cancel{t}} \sum_{i=1}^{n} \log x_i$$

BAR. PAR.

$$\nabla B(x) = \cancel{t}\, c - \mu \begin{bmatrix} 1/x_1 \\ \vdots \\ 1/x_n \end{bmatrix} \quad \text{where } \mu = \frac{1}{t} \to 0.$$

$$= \cancel{t}\, c - \mu X^{-1} e \quad \text{where } X = \underbrace{\begin{bmatrix} x_1 & & \\ & \ddots & \\ & & x_n \end{bmatrix}}_{\operatorname{diag}(x)}, \; e = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

∴ need

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

A.K.A. ν (BV), λ (N+W)

Recall LP dual: $\max \; b^T y$

$$\text{s.t.} \quad A^T y + z = c \qquad z \geq 0$$

A.K.A. λ, s

"dual slack".

PD2

Thus $\mu X^{-1} e = z$, i.e. $\frac{\mu}{x_i} = z_i$ $\quad i = 1, \dots, n$

just a special case of what we wrote last week,

$$\lambda_i = -\frac{1}{t f_i(x)} \qquad \left(\begin{array}{l} f_i(x) \text{ is now } -x_i \\ \lambda_i \quad \text{"} \quad \text{"} \quad z_i \\ t \quad \text{"} \quad \text{"} \quad 1/\mu \end{array} \right)$$

Thus pts on the central path are characterized by

$$\left.\begin{array}{r} A^T y + \mu X^{-1} e = c \\ Ax = b \end{array}\right\} \begin{array}{l} m+n \text{ eqns in} \\ m+n \text{ vars.} \end{array}$$

Let's write

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

& consider apply Newton's method for NONLINEAR EQNS to solve $F(x, y) = 0$ and ~~[illegible]~~

$$F'(x_k, y_k) \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = -F(x_k, y_k).$$

$$\begin{bmatrix} -\mu X_k^{-2} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} c - A^T y_k - \mu X_k^{-1} e \\ b - A x_k \end{bmatrix}$$

Looks just like Newton for minimizing $B(x)$ s.t. $Ax = b$
-INDEED IT IS, as $\nabla^2 B(x) = \mu X^{-2}$ $(\succ 0)$.

To see equivalence, ~~[illegible]~~ change sign of 1ST eqn:
RHS is $+\nabla B(x_k) - A^T y_k$.
NO SURPRISE: we are applying Newton to $F(x, y) = 0$,
and 1ST block of F is $-\nabla B(x) + A^T y$.

Recall block Gauss: add $A\left(\frac{1}{\mu} X_k^2\right) * 1^{ST}$ eqn to 2nd, get

$$\begin{bmatrix} 0 & \underbrace{\frac{1}{\mu} A X_k^2 A^T}_{\text{SCHUR COMPL. MTX.}} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \text{some rhs.}$$

PD3

New Idea: REWRITE THE EQNS AS

$$A^T y + z = c \quad \text{dual feas.}$$
$$Ax = b \quad \text{primal feas}$$
$$\text{diag}(x)\,\text{diag}(z)e = XZe = \mu e \quad \text{centrality} \xrightarrow[\mu \to 0]{} XZ = 0 \text{ compl.}$$

equiv, $Xz = \mu e$, i.e. $x_i z_i = \mu$ $i = 1, \dots, n$.

Now

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

Newton: $F'(x_k, y_k, z_k) \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = -F(x_k, y_k, z_k)$

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

Block ~~Gauss~~ Gauss: subtract X_k^{-1} ~~A~~ $*$ 3rd row from 1st row

$$\begin{bmatrix} -X_k^{-1} Z_k & A^T & \cancel{0} \\ A & 0 & \cancel{0} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \cancel{\Delta z} \end{bmatrix} = \dots$$

Block Gauss again: ~~subtract~~ ~~add~~ add $A Z_k^{-1} X_k * 1^{ST}$ row ~~from~~ to 2^{nd} row

$$A D_k A^T \Delta y = \text{some r.h.s.}$$

where now $D_k = Z_k^{-1} X_k$ instead of $\frac{1}{\mu} X_k^2$.

SAME ON CENTRAL PATH, NOT OTHERWISE.

THEN GET Δx, Δz BY SUBSTITUTION.

NOTE: UNLIKE BEFORE, F is defined at the end pt of the central path, the limit as $\mu \to 0$.

Furthermore,

PD4

turns out that $F'(x^*, y^*, z^*)$ is NONSINGULAR if LP is NONDEGENERATE, i.e. has a UNIQUE sol'n (x^*, y^*, z^*), which must be the limit of the central path as $\mu \to 0$, and ~~and must~~ satisfies strict complementarity: only one of $x_i^*, z_i^* = 0$ for each i: thus NO ROW of $[Z^* \;\; 0 \;\; X^*]$ is 0.

But how do we

(a) make sure $\{x_k\}$, $\{z_k\}$ stay > 0

(b) let $\mu \to 0$.

simplest:

$$x_{k+1} = x_k + \alpha_k \Delta x$$

where $\alpha_k = \min(1, \tau \hat{\alpha})$

where τ is close to 1, e.g. 0.999

$$\hat{\alpha} = \max\{\alpha : x_k + \alpha \Delta x \geq 0\}$$

"ratio test".

$$y_{k+1} = y_k + \beta_k \Delta y$$

$$z_{k+1} = z_k + \beta_k \Delta z$$

where $\beta_k = \min(1, \tau \hat{\beta})$

$$\hat{\beta} = \max\{\beta : z_k + \beta \Delta z \geq 0\}$$

simplest: $\mu = \sigma \dfrac{x_k^T z_k}{n}$. $\quad \sigma < 1$

e.g. $\sigma = 1/4$.

KEY POINT: NO INNER ITERATION!

These choices work very well in practice.

Mehrotra's predictor–corrector: greatly speeds things up.

CAN START INFEASIBLE – BECOME FEAS IF EVER TAKE $\alpha_k = 1, \beta_k = 1$

Methods with complexity results

"short-step" method: best complexity, v. slow (in practice)

"long-step" methods: more practical

use the $N_{-\infty}(\gamma)$ region to keep $x_i z_i \geq \gamma(x^T z/n)$ ~~[illegible]~~ TOO SMALL.

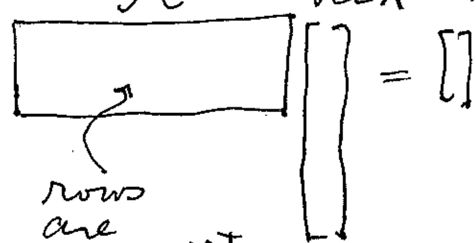
(II) SDP. min $C\cdot X \leftarrow \mathrm{tr}\, CX = \sum c_{ij}x_{ij}$

s.t. $A_i \cdot X = b_i$

$X \succeq 0$

$X \in S^n$ NO LONGER DIAGONAL!

$\mathcal{A}\ \mathrm{vec}X = b$



rows are $(\mathrm{vec}(A_i))^T$

$$B(X) = C\cdot X - \mu \log\det X$$

$$\nabla_X B(X) = C - \mu X^{-1} \quad (\text{like } X, \text{ this is } \in S^n)$$

Condition for X on central path is

$$C - \mu X^{-1} = \sum_{i=1}^{m} y_i A_i$$

Recall dual: $\max\ b^T y$

s.t. $\left(\sum y_i A_i\right) + Z = C$

$Z \succeq 0$

thus need

$$\mu X^{-1} = Z.$$

Let $F(X,y) = \begin{bmatrix} \mathrm{vec}\{(\sum_{i=1}^{m} y_i A_i) + \mu X^{-1} - C\} \\ \mathcal{A}\ \mathrm{vec}X - b \end{bmatrix}$

a matrix eqn and m scalar eqns

in a matrix var and m " vars.

As earlier, Newton for $F(X,y) = 0$ is equivalent to Newton for minimizing $B(X)$ s.t. $\mathcal{A}\ \mathrm{vec}X = b$

Get $\begin{bmatrix} -\mu X_k^{-1} \otimes X_k^{-1} & \mathcal{A}^T \\ \mathcal{A} & 0 \end{bmatrix} \begin{bmatrix} \mathrm{vec}\Delta X \\ \Delta y \end{bmatrix} = \ldots$

k is iteration index — unlike i, j

$\begin{bmatrix} m_{11}N & \cdots & m_{1n}N \\ \vdots & & \\ m_{n1}N & \cdots & m_{nn}N \end{bmatrix}$

$\downarrow$

$M\otimes N$ is an $n^2\times n^2$ matrix representing the linear op def by $K \to NKM^T$

Here the $n^2\times n^2$ Kronecker product $X_k^{-1}\otimes X_k^{-1}$ is the rep'n of the Hessian of B on $\mathbb{R}^{n^2}$.

Key thing to know about Kronecker products

$$\boxed{\mathrm{vec}(ABC) = (C^T\otimes A)\,\mathrm{vec}\,B} \quad A,B,C\in\mathbb{R}^{n\times n}$$

omit (The Hessian of B on $\mathbb{R}^{n\times n}$ is a linear operator defined by

$$\nabla^2_{XX} B(X)(U,V) = \mathrm{tr}\, UX^{-1}VX^{-1}.$$

Likewise $\mathrm{vec}(U)^T(X^{-1}\otimes X^{-1})\mathrm{vec}V = \mathrm{tr}\,U(X^{-1}VX^{-1})$.

PRIMAL DUAL EQNS.

$$F(X,y,Z)=0 \quad \begin{cases} \{A_i\cdot X = b_i\},\ i=1,\dots,m & (A\,\mathrm{vec}X = b) \\ \sum y_iA_i + Z = C & (A^Ty + \mathrm{vec}Z = \mathrm{vec}C) \\ XZ = \mu I & \text{centrality} \end{cases}$$

2 matrix eqns + m scalar ~~vars~~ eqns in
2 " vars " " " vars.

As $\mu\to 0$, centrality eqn $\to$ $XZ=0$.

Observe: opt conds $X\succeq 0, Z\succeq 0, \mathrm{tr}XZ = 0$ indeed $\Rightarrow XZ=0$

$X=LL^T \quad Z=MM^T$

$\mathrm{tr}XZ = \mathrm{tr}(L^TM)(M^TL) = 0$

so $\|L^TM\| = 0$

so $XZ=0$.

Transposing, $ZX=0$ too

Thus X, Z commute + share common o.n. set of e-vectors.

$$Q^TXQ = \mathrm{Diag}(\xi_i)$$
$$Q^TZQ = \mathrm{Diag}(\omega_i)$$

satisfying $\xi_i\omega_i = 0, \quad i=1,\dots,n$

Eigenvalue Complementarity!

apply Newton to $F(X,y,Z)=0$ on $\mathbb{R}^{n\times n}\times\mathbb{R}^m\times\mathbb{R}^{n\times n}$

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

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

PD7

Represent this on $\mathbb{R}^{n^2}$ by

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

Thus get

$$\begin{bmatrix} 0 & \mathcal{A}^T & I \\ \mathcal{A} & 0 & 0 \\ \mathcal{E} & 0 & \mathcal{F} \end{bmatrix} \begin{bmatrix} \mathrm{vec}\,\Delta X \\ \Delta y \\ \mathrm{vec}\,\Delta Z \end{bmatrix} = \ldots$$

where $\mathcal{E} = Z_k \otimes I$ $\qquad \mathcal{A} = \begin{bmatrix} (\mathrm{vec}\,A_1)^T \\ \vdots \\ (\mathrm{vec}\,A_m)^T \end{bmatrix}$

$\mathcal{F} = I \otimes X_k$

Block Gauss gives

$\mathcal{M}$ is SCHUR COMPLEMENT MATRIX

$$\underbrace{\mathcal{A}\,\mathcal{E}^{-1}\mathcal{F}\,\mathcal{A}^T}_{\mathcal{M} =}\,\Delta y = \ldots$$

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

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

$$= Z_k^{-1} \otimes X_k$$

Thus we form $\mathcal{M}$ by $\mathcal{M} = \mathcal{A}\mathcal{B}$ where the j^{th} col (NOT k^{th}!) of $\mathcal{B}$ is $(Z_k^{-1} \otimes X_k)\,\mathrm{vec}(A_j)$

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

Thus $\mathcal{M}_{ij} = \mathrm{tr}\, A_i X_k A_j Z_k$

Don't confuse iteration index k with component index j

Unfortunately in (2.16) in paper we use k for j.

Then solve for Δy

Get ΔX, ΔZ by subst of formula for Δy.

PD8

Problem: ΔX is generally NOT SYMMETRIC, because we applied Newton's method in $\mathbb{R}^{n^2}$, NOT in S^n.
Note: product of 2 symmetric matrices is generally not symmetric.

What to do?
1. "XZ" method: replace ΔX by $\frac{1}{2}(\Delta X + (\Delta X)^T)$ "H..K..M"
2. "XZ+ZX" method: "AHO"
replace 3rd eqn in $F(X,y,Z)$ by $XZ+ZX = 2\mu I$.
Apply Newton in S^n:
$$(X_k + \Delta X)(Z_k + \Delta Z) + (Z_k + \Delta Z)(X_k + \Delta X) = 2\mu I$$
linearize:
$$\underbrace{X_k(\Delta Z) + (\Delta Z)X_k}_{2\text{ "}\Delta Z\text{" terms}} + \underbrace{(\Delta X)Z_k + Z_k(\Delta X)}_{2\text{ "}\Delta X\text{" terms}} = 2\mu I - X_k Z_k - Z_k X_k$$

Leads to a method with very nice properties but more expensive to implement.
Alizadeh, Haeberly, Overton 1998.

3. Nesterov–Todd Method.
Also has v nice properties but more expensive to implement.

So, "XZ" has become the standard.

PD9

... Then $X_{k+1} = X_k + \alpha_k \Delta X$,
$y_{k+1} = y_k + \beta_k \Delta y$, $Z_{k+1} = Z_k + \beta_k \Delta Z$

where $\alpha_k = \min(1, \tau\hat{\alpha})$

$$\hat{\alpha} = \max\{\alpha : X_k + \alpha(\Delta X) \succeq 0\}$$

$$\beta_k = \min(1, \tau\hat{\beta})$$

$$\hat{\beta} = \max\{\beta : Z_k + \beta(\Delta Z) \succeq 0\}$$

How? Let $X_k = LL^T$

Need $I + \alpha L^{-1}(\Delta X)L^{-T} \succeq 0$

$$1 + \alpha\lambda_{\min}(L^{-1}(\Delta X)L^{-T}) \geq 0$$

$$1 - \alpha\lambda_{\max}(-L^{-1}(\Delta X)L^{-T}) \geq 0$$

$$\alpha\lambda_{\max}(-L^{-1}(\Delta X)L^{-T}) \leq 1$$

$$\hat{\alpha} = 1/\lambda_{\max}(-L^{-1}(\Delta X)L^{-T}).$$

(Reduce μ as before).

As in XZ method, if ever take step $\alpha_k = 1$, get primal feasible, & " " " " $\beta_k = 1$, get dual feasible (+ stay feasible)

Turns out can take τ much closer to 1, e.g. 0.999 with XZ+ZX method that with XZ method.

~~On~~ when started on central path, all methods generate same next iterate (but it won't be on central path).

M same for each method.