Primal-dual interior-point methods part II

Javier Peña (guest lecturer)
Convex Optimization 10-725/36-725

Last time: primal-dual IPM for linear programming

Consider the primal-dual linear programming pair

$$\begin{array}{llcll} \min & c^\mathsf{T} x & \qquad & \max & b^\mathsf{T} y \\ & Ax = b & & & A^\mathsf{T} y + s = c \\ & x \geq 0 & & & s \geq 0. \end{array}$$

Notation:

- $\mathcal{F}^0 := \{(x, y, s) : Ax = b,\ A^\mathsf{T} y + s = c,\ x, s > 0\}.$
- Given $x, s \in \mathbb{R}^n_+$, $\mu(x, s) := \frac{x^\mathsf{T} s}{n}$
- $\mathcal{N}_2(\theta) := \{(x, y, s) \in \mathcal{F}^0 : \|XS\mathbf{1} - \mu(x, s)\mathbf{1}\|_2 \leq \theta\mu(x, s)\}$

Newton step equations:

$$\begin{bmatrix} 0 & A^\mathsf{T} & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \tau\mathbf{1} - XS\mathbf{1} \end{bmatrix}.$$

Short-step path following algorithm

Algorithm SPF

1. Let $\theta, \delta \in (0,1)$ be such that $\frac{\theta^2+\delta^2}{2^{3/2}(1-\theta)} \leq \left(1 - \frac{\delta}{\sqrt{n}}\right)\theta$.
2. Let $(x^0, y^0, s^0) \in \mathcal{N}_2(\theta)$.
3. For $k = 0, 1, \ldots$
   - Compute Newton step for
     $$(x, y, s) = (x^k, y^k, s^k),\ \tau = \left(1 - \frac{\delta}{\sqrt{n}}\right)\mu(x, s).$$
   - Set $(x^{k+1}, y^{k+1}, s^{k+1}) := (x^k, y^k, s^k) + (\Delta x, \Delta y, \Delta s)$.

Theorem

The sequence generated by Algorithm SPF satisfies
$(x^k, y^k, s^k) \in \mathcal{N}_2(\theta)$ *and* $\mu(x^{k+1}, s^{k+1}) = \left(1 - \frac{\delta}{\sqrt{n}}\right)\mu(x^k, s^k)$

Infeasible interior-point algorithm

Given (x, y, s), let $r_b := Ax - b, r_c := A^\mathsf{T} y + s - c$.

Assume (x^0, y^0, s^0) with $x^0, s^0 > 0$ is given.

$$\mathcal{N}_{-\infty}(\gamma, \beta) := \{(x, y, s) : \|(r_b, r_c)\| \le [\|(r_b^0, r_c^0)\|/\mu^0]\beta\mu, \\ x, s > 0, x_i s_i \ge \gamma\mu, i = 1, \dots, n\}$$

for $\gamma \in (0, 1),\ \beta \ge 1$.

Newton step equations:

$$\begin{bmatrix} 0 & A^\mathsf{T} & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = - \begin{bmatrix} r_c \\ r_b \\ XS\mathbf{1} - \tau\mathbf{1} \end{bmatrix}.$$

Algorithm IPF

1. Choose $\gamma \in (0,1)$, $0 < \sigma_{\min} < \sigma_{\max} < 0.5$, and $\beta \geq 1$.
2. Choose (x^0, y^0, s^0) with $x^0, s^0 > 0$
3. For $k = 0, 1, \ldots$
   - Choose $\sigma \in [\sigma_{\min}, \sigma_{\max}]$
   - Compute Newton step for

$$(x, y, s) = (x^k, y^k, s^k), \; \tau = \sigma\mu(x^k, s^k)$$

   - Choose α_k as the largest $\alpha \in [0,1]$ such that

$$(x^k, y^k, s^k) + \alpha(\Delta x, \Delta y, \Delta s) \in \mathcal{N}_{-\infty}(\gamma, \beta)$$

$$\mu(x^k + \alpha\Delta x, s^k + \alpha\Delta s) \leq (1 - 0.01\alpha)\mu(x_k, s_k)$$

   - Set $(x^{k+1}, y^{k+1}, s^{k+1}) := (x^k, y^k, s^k) + \alpha_k(\Delta x, \Delta y, \Delta s)$

Theorem

If primal and dual are feasible then the sequence generated by Algorithm IPF satisfies $\mu_k := \mu(x^k, y^k, s^k) \to 0$ linearly.

Outline

Today:

- Recap of semidefinite programming and duality
- The central path
- Primal-dual methods for SDP
- Self-scaled (symmetric) conic programming
- Solvers: `SeDuMi`, `SDPT3`

Semidefinite program in standard form

Problem of the form

$$\begin{array}{ll} \min\limits_{X} & C \bullet X \\ \text{subject to} & \mathcal{A}(X) = b \\ & X \succeq 0, \end{array}$$

where $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ linear map, $b \in \mathbb{R}^m, C \in \mathbb{S}^n$.

Any semidefinite program can be rewritten in standard form.

Standard assumption: $\mathcal{A}$ is surjective.

Semidefinite programming duality

The dual of the above problem is

$$\max_{y} \quad b^{\mathsf{T}} y \qquad \text{subject to} \quad \mathcal{A}^*(y) \preceq C.$$

Or equivalently

$$\begin{aligned} \max_{y,S} \quad & b^{\mathsf{T}} y \\ \text{subject to} \quad & \mathcal{A}^*(y) + S = C \\ & S \succeq 0. \end{aligned}$$

Throughout the sequel refer to the SDP from the previous slide as the primal problem and to the above SDP as the dual problem.

Semidefinite programming duality

Theorem (Weak duality)

Assume X is primal feasible and y is dual feasible. Then

$$b^{\mathsf{T}} y \leq C \bullet X.$$

Theorem (Strong duality)

Assume both primal and dual problems are strictly feasible. Then their optimal values are the same and they are attained.

Strong duality does not always hold

Examples

$$\begin{aligned} \min \quad & 2x_{12} \\ & \begin{bmatrix} 0 & x_{12} \\ x_{12} & x_{22} \end{bmatrix} \succeq 0. \end{aligned}$$

$$\begin{aligned} \min \quad & x_{11} \\ & \begin{bmatrix} x_{11} & 1 \\ 1 & x_{22} \end{bmatrix} \succeq 0. \end{aligned}$$

$$\begin{aligned} \min \quad & ax_{22} \\ & \begin{bmatrix} 0 & x_{12} & 1 - x_{22} \\ x_{12} & x_{22} & x_{23} \\ 1 - x_{22} & x_{23} & x_{33} \end{bmatrix} \succeq 0, \quad \text{for } \; a > 0. \end{aligned}$$

Optimality conditions

Assume strong duality holds. Then the points X^* and (y^*, S^*) are respectively primal and dual optimal solutions if and only if (X^*, y^*, S^*) solves

$$\begin{aligned} \mathcal{A}(X) &= b \\ \mathcal{A}^*(y) + S &= C \\ XS &= 0 \\ X, S &\succeq 0. \end{aligned}$$

Interior-point methods: Maintain first two and the fourth conditions and aim for the third one.

Historical remark

IPM for SDP developed independently by Nesterov & Nemirovski and Alizadeh in the late 1980s. The topic had a massive burst of research in the 1990s.

Barrier method for primal and dual problems

Pick $\tau > 0$. Approximate the primal SDP with

$$\begin{array}{ll} \min\limits_{X} & C \bullet X - \tau \log(\det X) \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

and the dual SDP with

$$\begin{array}{ll} \max\limits_{y,S} & b^{\mathsf{T}} y + \tau \log(\det S) \\ \text{subject to} & \mathcal{A}^*(y) + S = C. \end{array}$$

Neat fact:
The above two problems are, modulo a constant, Lagrangian duals of each other.

Primal-dual central path

Assume the primal and dual problems are strictly feasible. The primal-dual central path is the set

$$\{(X(\tau), y(\tau), S(\tau)) : \tau > 0\}$$

where $X(\tau)$, and $(y(\tau), S(\tau))$ solve the above pair of barrier problems. Equivalently, $(X(\tau), y(\tau), S(\tau))$ is the solution to

$$\begin{aligned} \mathcal{A}(X) &= b \\ \mathcal{A}^*(y) + S &= C \\ XS &= \tau I \\ X, S &\succ 0. \end{aligned}$$

Path following interior-point methods

Main idea:
Generate $(X^k, y^k, S^k) \approx (x(\tau^k), y(\tau^k), s(\tau^k))$ for $\tau^k \downarrow 0$.

Two main issues:

- Measure of proximity to the central path
- Update: Newton-like step

Notation:

$$\mathcal{F}^0 := \{(X, y, S) : \mathcal{A}X = b,\ \mathcal{A}^* y + S = C,\ X, S \succ 0\}.$$

Local neighborhood of the central path

Given $X, S \succeq 0$, let

$$\mu(X,S) := \frac{X \bullet S}{n}$$

and

$$\begin{aligned} d_F(X,S)) &:= \|\lambda(XS) - \mu(X,S)\mathbf{1}\|_2 \\ &= \|X^{1/2}SX^{1/2} - \mu(X,S)I\|_F \\ &= \|S^{1/2}XS^{1/2} - \mu(X,S)I\|_F. \end{aligned}$$

Given $\theta \in (0,1)$ define the local neighborhood $\mathcal{N}_F(\theta)$ as

$$\mathcal{N}_F(\theta) := \{(X,y,S) \in \mathcal{F}^0 : d_F(X,S) \leq \theta\mu(X,S)\}.$$

Newton step

Recall: $(X(\tau), y(\tau), S(\tau))$ solution to

$$\begin{bmatrix} \mathcal{A}^*(y) + S - C \\ \mathcal{A}(X) - b \\ XS \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \tau I \end{bmatrix}, \quad X, S \succ 0.$$

Natural Newton step:

$$\begin{bmatrix} 0 & \mathcal{A}^* & I \\ \mathcal{A} & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta X \\ \Delta y \\ \Delta S \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \tau I - XS \end{bmatrix}.$$

But we run into issues of symmetry...

Nesterov-Todd direction

Crux of Newton's method: Given trial solution x to

$$G(x) = 0$$

update to $x^+ = x + \Delta x$ by solving

$$G(x) + G'(x)\Delta x = 0 \Leftrightarrow G'(x)\Delta x = -G(x)$$

We want to linearize

$$XS - \tau I = 0.$$

Primal linearization:

$$S - \tau X^{-1} = 0 \rightsquigarrow \tau X^{-1}\Delta X X^{-1} + \Delta S = \tau X^{-1} - S.$$

Dual linearization:

$$X - \tau S^{-1} = 0 \rightsquigarrow \Delta X + \tau S^{-1}\Delta S S^{-1} = \tau S^{-1} - X.$$

Nesterov-Todd direction

Proper primal-dual linearization: average of previous two

$$W^{-1}\Delta X W^{-1} + \Delta S = \tau X^{-1} - S$$

or equivalently

$$\Delta X + W \Delta S W = \tau S^{-1} - X$$

provided

$$WSW = X.$$

Achieve the above by taking W as the geometric mean of X, S:

$$\begin{aligned} W &= S^{-1/2}(S^{1/2} X S^{1/2})^{1/2} S^{-1/2} \\ &= X^{1/2}(X^{1/2} S X^{1/2})^{-1/2} X^{1/2} \end{aligned}$$

Short-step path following algorithm

Algorithm SPF

1. Let $\theta, \delta \in (0,1)$ be such that

$$\frac{7(\theta^2 + \delta^2)}{1-\theta} \leq \left(1 - \frac{\delta}{\sqrt{n}}\right)\theta, \quad \frac{2\sqrt{2}\theta}{1-\theta} \leq 1.$$

2. Let $(X^0, y^0, S^0) \in \mathcal{N}_F(\theta)$.
3. For $k = 0, 1, \ldots$
   - Compute Nesterov-Todd direction for

$$(X, y, S) = (X^k, y^k, S^k),\ \tau = \left(1 - \frac{\delta}{\sqrt{n}}\right)\mu(X, S).$$

   - Set $(X^{k+1}, y^{k+1}, S^{k+1}) := (X^k, y^k, S^k) + (\Delta X, \Delta y, \Delta S)$.

Theorem

The sequence generated by Algorithm SPF satisfies

$$(X^k, y^k, S^k) \in \mathcal{N}_F(\theta),$$

and

$$\mu(X^{k+1}, S^{k+1}) = \left(1 - \frac{\delta}{\sqrt{n}}\right) \mu(X^k, S^k).$$

Corollary

In $\mathcal{O}\left(\sqrt{n} \log\left(\frac{n\mu(X^0, S^0)}{\epsilon}\right)\right)$ *the algorithm yields* $(X^k, y^k, S^k) \in \mathcal{F}^0$ *such that*

$$C \bullet X_k - b^\mathsf{T} y_k \leq \epsilon.$$

Have also “long-step”, and “infeasible” algorithms (as in LP).

Self-scaled cones

Extend IPM machinery to more general conic programming

$$\begin{array}{ll} \min\limits_{x} & c^{\mathsf{T}} x \\ \text{subject to} & Ax = b \\ & x \in K. \end{array}$$

Technical ingredients:

- $F : \mathsf{int}(K) \to \mathbb{R}$ is *logarithmically homogeneous* if for all $x \in \mathsf{int}(K),\ t > 0$

$$F(tx) = F(x) - \nu \log t$$

- A LHB $F : \mathsf{int}(K) \to \mathbb{R}$ is *self-scaled* if for all $x, w \in \mathsf{int}(K)$

$$F''(w)x \in \mathsf{int}(K^*) \ \text{ and } \ F^*(-F(w)x) = F(x) - 2F(w) - \nu.$$

Self-scaled cones

A convex cone $K \subseteq \mathbb{R}^n$ is self-scaled if there exists a self-scaled LHB $F : \mathrm{int}(K) \to \mathbb{R}$.

Examples

- $K = \mathbb{R}^n_+,\ F(x) = -\sum_{j=1}^n \log x_j.$
- $K = \mathbb{S}^n_+,\ F(X) = -\log\det X.$
- $K = \mathcal{Q}_n,\ F(x) = -\log(x_0^2 - \|\bar{x}\|^2).$

  Recall $\mathcal{Q}_n := \left\{ x = \begin{bmatrix} x_0 \\ \bar{x} \end{bmatrix} \in \mathbb{R}^n : x_0 \geq \|\bar{x}\| \right\}.$
- Any cartesian product of the above.

Self-scaled cones are the same as *symmetric* cones, a class of convex cones studied in harmonic analysis.

Self-scaled cones

Theorem

Assume K is a self-scaled cone with self-scaled barrier F. Then $K = K^*$ and for all $x, s \in$ int(K) there exists a unique scaling point $w \in$ int(K) such that

$$F''(w)x = s.$$

Nice symmetry:

$$F''(w)x = s \Leftrightarrow F''(-F'(w))s = x.$$

Self-scaled conic programming

Consider

$$\begin{array}{ll} \min\limits_{x} & c^{\mathsf{T}} x \\ \text{subject to} & Ax = b \\ & x \in K, \end{array}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m, c \in \mathbb{R}^n$ and K is a self-scaled cone.

Dual

$$\begin{array}{ll} \max\limits_{y,s} & b^{\mathsf{T}} y \\ \text{subject to} & A^{\mathsf{T}} y + s = c \\ & s \in K. \end{array}$$

Optimality conditions

Assume strong duality holds. (For instance, both primal and dual are strictly feasible.)
The points x^* and (y^*, s^*) are respectively primal and dual optimal solutions if and only if (x^*, y^*, s^*) solves

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

Central path

Assume $F : \text{int}(K) \to \mathbb{R}$ is a self-scaled LHB for K.

Central path

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

where $(x(\tau), y(\tau), s(\tau))$ solves

$$\begin{aligned} Ax &= b \\ A^\mathsf{T} y + s &= c \\ \tau F'(x) + s &= 0 \\ x, s &\in \text{int}(K). \end{aligned}$$

Nesterov-Todd direction

As before, need to linearize

$$\tau F'(x) + s = 0.$$

Primal-dual linearization

$$F''(w)\Delta x + \Delta s = -\tau F'(x) - s$$

where w is the scaling point of x, s.

Nesterov-Todd equations

$$\begin{bmatrix} 0 & A^\mathsf{T} & I \\ A & 0 & 0 \\ F''(w) & 0 & I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = - \begin{bmatrix} A^\mathsf{T} y + s - c \\ Ax - b \\ \tau F'(x) + s \end{bmatrix}.$$

Local neighborhood of the central path

For $x \in \text{int}(K)$ let

$$\|v\|_x := \left(v^{\mathsf{T}} F''(x) v\right)^{1/2}.$$

Given $x, s \in \text{int}(K)$, let

$$\mu(x,s) := \frac{x^{\mathsf{T}} s}{\nu}.$$

Given $\theta \in (0,1)$ define the local neighborhood $\mathcal{N}(\theta)$ as

$$\mathcal{N}(\theta) := \{(x,y,s) \in \mathcal{F}^0 : \|s + \mu(x,s)F'(x)\|_{-F'(x)} \leq \theta\mu(x,s)\}.$$

Previous IPM machinery extends.

Conic programming solvers

When we mix LP/SOCP/SDP it is convenient to convert matrices into vectors

`vec`: $\mathbb{R}^{n\times n} \to \mathbb{R}^{n^2}$ is the mapping

$$X \mapsto \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} & X_{21} & X_{22} & \cdots & X_{nn} \end{bmatrix}^{\mathsf{T}}$$

`mat`: $\mathbb{R}^{n^2} \to \mathbb{R}^{n\times n}$ is the inverse mapping.

Related mapping `svec`: $\mathbb{S}^n \to \mathbb{R}^{n(n+1)/2}$

$$X \mapsto [X_{11} \;\; \sqrt{2}X_{12} \;\; \cdots \;\; \sqrt{2}X_{1n} \;\; X_{22} \;\; \sqrt{2}X_{23} \;\; \cdots \;\; \sqrt{2}X_{n-1,n} \;\; X_{nn}]^{\mathsf{T}}$$

Notice: For $X, S \in \mathbb{S}^n$

$$X \bullet S = \mathtt{vec}(X)^{\mathsf{T}}\mathtt{vec}(S) = \mathtt{svec}(X)^{\mathsf{T}}\mathtt{svec}(S).$$

Conic programming solvers

`SeDuMi`: Developed by late J. Sturm. Freely available from :

`http://sedumi.ie.lehigh.edu`

Matlab-based, syntax:

```
> [x,y,info] = sedumi(A,b,c,K) ;
```

This solves

$$\begin{array}{ll} \min_x & c^\mathsf{T} x \\ & Ax = b \\ & x \in K \end{array} \qquad \begin{array}{ll} \max_{y,s} & b^\mathsf{T} y \\ & A^\mathsf{T} y + s = c \\ & s \in K^*. \end{array}$$

Normal termination gives either a primal-dual optimal solution, or a certificate of infeasibility.

In matlab environment `A` is an $m \times n$ matrix, `c,x` are n-vectors, and `b,y` are m-vectors.

`K` is a structure that describes K:

K.f is the number of free components.

K.l is the number of non-negative components.

K.q lists the dimensions of second-order constraints.

K.s lists the dimensions of SDP constraints.

`SDPT3`: Developed by M. Todd, K. Toh, and R. Tütüncü.

Freely available from

`http://www.math.nus.edu.sg/~mattohkc/sdpt3.html`

Matlab-based as well. Syntax:

```
> [obj,X,y,S] = sqlp(blk,A,C,b) ;
```

`blk` describes the blocks (LP/SOCP/SDP) in K.

It works with `svec` instead of `vec`.

References and further reading

- J. Renegar (2001), "A Mathematical View of Interior-Point Methods."
- R. Tütüncü, K. Toh, and M. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," Mathematical Programming 95 (2003), pp. 189–217.
- J. Sturm, "Implementation of Interior Point Methods for Mixed Semidefinite and Second Order Cone Optimization Problems," Optimization Methods and Software 17 (2002), pp. 1105–1154.