IE 521: Convex Optimization	**Spring 2017, UIUC**

Lecture 22: Interior Point Method for Conic Programs – April 17

Instructor: Niao He *Scribe: Shuanglong Wang*

Courtesy warning: These notes do not necessarily cover everything discussed in the class. Please email TA (swang157@illinois.edu) if you find any typos or mistakes.

In this lecture, we cover the following topics

- IPM for Conic Programs
- Primal Dual path following IPM

22.1 Summary of IPM

Let us first summarize the key concepts discussed in the last few lectures.

- **Self-concordant barrier:** A function $F(x)$ is ν-s.c.b. for a set $X = cl(dom(F))$ if it satisfies for any $x \in dom(F)$, $h \in \mathbf{R}^n$:

$$\begin{aligned} |D^3F(x)[h,h,h]| &\leq 2(D^2F(x)[h,h])^{3/2} \\ |DF(x)[h]| &\leq \sqrt{\nu}(D^2F(x)[h,h])^{1/2} \end{aligned}$$

- **Path-following interior point method:**

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & x \in X \end{aligned} \qquad \Longrightarrow \qquad \min_x \quad tc^T x + F(x)$$

  1. choose (x_0, t_0), where $t_0 > 0$ and x_0 is close to the analytical center $x_F := \arg\min_x F(x)$
  2. do for $k = 0, 1, \ldots$

$$\begin{aligned} t_{k+1} &= t_k\left(1 + \frac{\gamma}{\sqrt{\nu}}\right) \\ x_{k+1} &= x_k - [\nabla^2 F(x_k)]^{-1}[t_{k+1}c + \nabla F(x_k)] \end{aligned}$$

- **Iteration Complexity:** $O(\sqrt{\nu}\log(\frac{\nu}{\epsilon}))$ iterations.

22.2 IPM for Conic Programs

Recall the standard form and dual of conic program:

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & Ax - b \in K \end{aligned} \qquad\qquad \begin{aligned} \max_y \quad & b^T y \\ \text{s.t.} \quad & A^T y = c \\ & y \in K_* \end{aligned}$$

This becomes an LP, SOCP, SDP when $K = \mathbf{R}_+^n, L^n, S_+^n$, respectively.

22.2.1 Self-concordant Barriers for LP, SOCP, SDP

We now introduce the canonical self-concordant barriers for these cones.

Example 1: $F(x) = -\sum_{i=1}^n \ln(x_i)$ is n-s.c.b for the nonnegative orthant $\mathbf{R}_+^n$.

Example 2: $F(x) = -\ln(x_n^2 - x_1^2 - ... - x_{n-1}^2)$ is 2-s.c.b. for the Lorentz cone L^n.

Example 3: $F(X) = -\ln(\det(X)) = -\sum_{i=1}^n \ln(\lambda_i(X))$ is n-s.c.b. for the positive semidefinite cone S_+^n.

The first two examples can be easily verified based on the definition. Let's prove the third case. *Proof:* It suffices to show that given $X \in int(S_+^n)$ and $H \in S^n$, the one-dimensional restriction

$$\phi(t) = F(X + tH) = -\ln(\det(X + tH))$$

is a n-self-concordant barrier. We have

$$\begin{aligned} \phi(t) &= -\ln(\det(X^{1/2}(I + tX^{-1/2}HX^{1/2})X^{1/2})) \\ &= -\ln(\det(I + tX^{-1/2}HX^{-1/2})) - \ln(Det(X)) \\ &= -\sum_{i=1}^n \ln(1 + t\lambda_i(X^{-1/2}HX^{-1/2})) + \phi(0) \end{aligned}$$

Denote $\lambda_i = \lambda_i(X^{-1/2}HX^{-1/2})$, $i = 1, ..., n$, then $\phi(t) = -\sum_{i=1}^m \ln(1 + t\lambda_i) + \phi(0)$. Hence,

$$\begin{aligned} \phi'(0) &= -\sum_{i=1}^n \lambda_i \\ \phi''(0) &= \sum_{i=1}^n \lambda_i^2 \\ \phi'''(0) &= -\sum_{i=1}^n \lambda_i^3 \end{aligned}$$

Note that

$$(\sum_{i=1}^{n}\lambda_i)^2 \leq n\sum_{i=1}^{n}\lambda_i^2 \Longrightarrow |\phi'(0)^2| \leq n\phi''(0)$$

$$|\sum_{i=1}^{n}\lambda_i^3| \leq (\sum_{i=1}^{n}\lambda_i^2)^{3/2} \Longrightarrow |\phi'''(0)| \leq 2[\phi''(0)]^{3/2}$$

Therefore, $\phi(t)$ is a n-self-concordant barrier for any $X \in int(S_+^n)$ and $H \in S^n$. ■

Remark: Indeed, any ν-self-concordant carrier for the cone S_+^n has $\nu \geq n$.

As a result, the function build on the above barriers,

$$\tilde{F}(x) := F(Ax - b)$$

is a self-concordant barrier for $X := \{x : Ax - b \in K\}$.

22.2.2 Linear Program: IPM vs Ellipsoid Method

For instance, consider the linear program $(m > n)$

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & a_j^T x \geq b_j, \quad j = 1, ..., m \end{aligned}$$

The barrier function

$$\tilde{F}(x) = -\sum_{j=1}^{m}\ln(a_j^T x - b_j)$$

is m-self-concordant barrier for the constraint set.

We can compute the gradient and Hessian

$$\nabla\tilde{F}(x) = -\sum_{j=1}^{m}\frac{a_j}{a_j^T x - b_j}$$

$$\nabla^2\tilde{F}(x) = \sum_{j=1}^{m}\frac{a_j a_j^T}{(a_j^T x - b_j)^2}$$

The arithmetic costs of computing $\tilde{F}(x)$, $\nabla\tilde{F}(x)$, $\nabla^2\tilde{F}(x)$ are $O(mn)$, $O(mn)$, $O(mn^2)$, respectively. The arithmetic cost for a Newton step, i,e, solving $[\nabla^2\tilde{F}(x)]y = \nabla\tilde{F}(x)$, is $O(n^3)$. Hence, the overall arithmetic cost of finding a ϵ-solution to the linear program is

$$O(mn^2)O(\sqrt{m}\log(\frac{m}{\epsilon})) = O(m^{3/2}n^2\log(\frac{m}{\epsilon}))$$

In contrast, when applying Ellipsoid method, the overall arithmetic cost is

$$O(mn + n^2)O(n^2\log(\frac{1}{\epsilon})) = O(mn^3\log(\frac{1}{\epsilon}))$$

The interior-point method is more efficient when m is not too large, i.e. $m \leq O(n^2)$

22.3 Primal-Dual Path-Following IPM

When solving conic programs, ideally we would like to develop interior point methods that

- produce primal-dual pairs at each iteration
- handle equality constraints
- require no prior knowledge of a strictly feasible solution
- adjust penalty based on current solution

Key idea: To approximate the KKT conditions.

We focus on the SDP case. Consider the standard form of the primal problem

$$\begin{aligned} \min \quad & \mathrm{Tr}(CX) \\ \text{s.t.} \quad & \mathrm{Tr}(A_i X) = b_i, i = 1, ..., m \\ & X \succcurlyeq 0 \end{aligned} \tag{P}$$

and its dual problem

$$\begin{aligned} \max_{y,Z} \quad & b^T y \\ \text{s.t.} \quad & \sum_{i=1}^{m} y_i A_i + Z = C \\ & Z \succcurlyeq 0 \end{aligned} \tag{D}$$

Assume (P) and (D) are strictly primal-dual feasible, so there is no duality gap.

KKT conditions for (P) and (D):

$$\left.\begin{aligned} X^* \succcurlyeq 0, Z^* \succcurlyeq 0 \\ \forall i = 1, ..., m : \mathrm{Tr}(A_i X^*) = b_i \\ \sum_{i=1}^{m} y_i^* A_i + Z^* = C \end{aligned}\right\} \text{(primal-dual feasibility)}$$

$$X^* Z^* = 0 \quad \text{(complementary slackness)}$$

We now consider the barrier problem of the primal problem

$$\begin{aligned} \min \quad & \mathrm{Tr}(CX) - \mu \ln(\det(X)) \\ \text{s.t.} \quad & \mathrm{Tr}(A_i X) = b_i, \quad i = 1, ..., m \end{aligned} \tag{BP}$$

and that of the dual problem

$$\begin{aligned} \max_y \quad & b^T y + \mu \log(\det(Z)) \\ \text{s.t.} \quad & \sum_{i=1}^m y_i A_i + Z = C \end{aligned} \tag{BD}$$

In fact, these are indeed the Lagrange duals to each other, up to constant.

KKT conditions for (BP) and (BD):

$$\left.\begin{aligned} \forall i = 1, ..., m : \mathrm{Tr}(A_i X^*(\mu)) = b_i \\ \sum_{i=1}^m y_i^*(\mu) A_i + Z^*(\mu) = C \end{aligned}\right\} \text{(primal-dual feasibility)}$$

$$X^*(\mu) Z^*(\mu) = \mu I \quad \text{(complementary slackness)}$$

The duality gap at $(X^*(\mu), y^*(\mu))$ is

$$Tr(CX^*(\mu)) - b^T y^*(\mu) = Tr(Z^*(\mu) X^*(\mu)) = \mu n$$

As $\mu \to 0$, the duality gap is zero, and $(X^*(\mu), y^*(\mu), Z^*(\mu)) \to (X^*, y^*, Z^*)$.

The set $\{(X(\mu), y(\mu), Z(\mu)) : \mu > 0\}$ is called the *primal-dual central path*.

Newton step: Find direction $(\Delta X, \Delta y, \Delta Z)$ by solving the equations:

$$\left\{ \begin{array}{l} Tr(A_i(X + \Delta X)) = b_i, \quad i = 1, ..., m \\ \sum_{i=1}^m (y_i + \Delta y_i) A_i + (Z + \Delta Z) = C \\ (X + \Delta X)(Z + \Delta Z) = \mu I \end{array} \right. \Longrightarrow \left\{ \begin{array}{l} Tr(A_i \Delta X) = 0, \quad i = 1, ..., m \\ \sum_{i=1}^m \Delta y_i A_i + \Delta Z = 0 \\ (X + \Delta X)(Z + \Delta Z) = \mu I \end{array} \right. \tag{$\star$}$$

Basic primal-dual scheme

Initial $(X, y, Z) = (X_0, y_0, Z_0)$ with $X_0 > 0$, $Z_0 > 0$.

At each iteration:

- compute $\mu = \frac{Tr(XZ)}{n}$, $\mu \leftarrow \frac{\mu}{2}$
- compute $(\Delta X, \Delta y, \Delta Z)$ by solving the KKT equations $(\star)$
- update $(X, y, Z) \leftarrow (X + \alpha\Delta X, y + \beta\Delta y, Z + \beta\Delta Z)$ with proper α, β that preserves positivity of (X, Z).

Remark (Approximation of KKT equation). Note that only the last equation in the system of KKT conditions is nonlinear. One cane apply first-order approximation:

$$\mu = (X + \Delta X)(Z + \Delta Z) \approx XZ + \Delta XZ + X\Delta Z$$

and solve the linearized KKT equations.