Lecture 20: Primal-Dual Interior Point Methods

Lecturer: Banghua Zhu *Scribe: Amy Guo*

In this lecture, we introduce the application of Interior Point (IP) method to Linear Programs (LPs), Conic Quadratic Programs (CQPs), and Semi-definite Programs (SDPs). This scribe is based on the hand-written note by Banghua Zhu. We refer the reader to [1] for more details.

1 Recap of IP Method

For a minimization problem $\min\{c^\top x : x \in \mathcal{X}\}$, we first need to find a barrier function $F(x)$ such that for every sequence $\{x_i\}_{i\in\mathbb{N}} \subseteq \operatorname{int}(\mathcal{X})$ and $x := \lim_{i\to\infty} x_i \in \partial\mathcal{X}$, it holds that $F(x_i) \to \infty$, as $i \to \infty$. Then, we let $F_t(x) = tc^\top x + F(x)$. The central path is defined as: $x_*(t) = \operatorname{argmin}_x F_t(x)$. This transforms the constrained problem to the unconstrained problem.

To solve IP in practice, there are two feasible ways:

1. Solve $x_*(t)$ for large t directly.
2. Start from $x_*(0)$, iteratively increase t to approach $x_*(t)$, i.e., path-following algorithm in Lec 18.

2 Canonical Barrier (CB)

Consider the problem $\min\{c^\top x \;:\; Ax - B \in \mathbb{K}\}$, where

$$\mathbb{K} = \mathbb{S}_+^{k_1} \times \mathbb{S}_+^{k_2} \times \cdots \times \mathbb{S}_+^{k_p} \times \mathbb{L}^{k_{p+1}} \times \cdots \times \mathbb{L}^{k_m} \subset E = \mathbb{S}^{k_1} \times \mathbb{S}^{k_2} \times \cdots \times \mathbb{S}^{k_p} \times \mathbb{R}^{k_{p+1}} \times \cdots \times \mathbb{R}^{k_m}. \tag{1}$$

The set $\mathbb{S}_+^k = \left\{X \in \mathbb{S}^k \mid X \geq 0\right\}$ and $\mathbb{L}^k = \left\{X \in \mathbb{R}^k \mid X_k \geq \sqrt{X_1^2 + \cdots + X_{k-1}^2}\right\}$ refer to the semi-definite cone and Lorentz cone respectively. We remark that the inner product associated with these two cones are defined as: $\langle X_i, Y_i\rangle_{\mathbb{S}^k} = \operatorname{Tr}(X_iY_i)$, $\langle X_i, Y_i\rangle_{\mathbb{R}^k} = X_i^\top Y_i$, and therefore $\langle X, Y\rangle_E = \sum_{i=1}^p \operatorname{Tr}(X_iY_i) + \sum_{i=p+1}^m X_i^\top Y_i$. To find a proper barrier for the cone defined in (1), it suffices to consider the canonical barriers for $\mathbb{S}_+^k$ and $\mathbb{L}^k$ first.

- The canonical barrier for semi-definite cones is defined as $S_k(X) = -\ln\det(X)$. The parameter of logarithmic homogeneity of $S_k(X)$ is $\theta(S_k) = k$, i.e., $S_k(tX) = S_k(X) - \theta(S_k)\ln t = S_k(X) - k\ln t$.
- The canonical barrier for Lorentz cones is defined as $L_k(X) = -\ln(X_k^2 - X_1^2 - \cdots - X_{k-1}^2) = -\ln(X^\top J_k X)$, where $J_k = \begin{pmatrix} -I_{k-1} & \\ & 1 \end{pmatrix}$. The parameter of logarithmic homogeneity associated with $L_k(X)$ is $\theta(L_k) = 2$.

Therefore, for $X \in \mathbb{K}$, define $K(X) = S_{k_1}(X_1) + \cdots + S_{k_p}(X_p) + L_{k_{p+1}}(X_{p+1}) + \cdots + L_{k_m}(X_m)$. Then, $K(X)$ is the canonical barrier for $\mathbb{K}$ with parameter $\theta(K) = \sum_{i\leq p}\theta(S_{k_i}) + \sum_{i=p+1}^m \theta(L_{k_i}) = \sum_{i=1}^p k_i + 2(m-p)$.

Theorem 1 (Properties of CB, Prop 4.3.1 in [1])**.** *The barrier function $K(X) = S_{k_1}(X_1) + \cdots + S_{k_p}(X_p) + L_{k_{p+1}}(X_{p+1}) + \cdots + L_{k_m}(X_m) = \sum_{i=1}^p -\ln\det(X_i) + \sum_{j=p+1}^m -\ln\left(X_j^\top J_{k_j} X_j\right)$ satisfies:*

1. *Barrier property: $K(\cdot)$ is C^∞ strongly convex function, such that $X^i \in int(\mathbb{L})$, $\lim X^i = X \in \partial\mathbb{K} \Rightarrow K(X^i) \to \infty$ as $i \to \infty$.*
2. *Logarithmically homogeneity: $X \in int(\mathbb{K}), t > 0 \Rightarrow K(tX) = K(X) - \theta(K)\ln t$.*
3. *Self-duality: the mapping $X \to -\nabla K(X)$ is a one-to-one mapping from $int(\mathbb{K})$ onto $int(\mathbb{K})$, i.e., $X \in int\mathbb{K}, S = -\nabla K(X) \Leftrightarrow S \in int(\mathbb{K}), X = -\nabla K(S)$.*

3 Primal-Dual Pair and Geometric Form

Consider a conic programming

$$\begin{aligned} &\text{Conic Primal (CP)} \qquad \min_x \left\{c^\top x : Ax - B \in \mathbb{K}\right\} \\ &\text{Conic Dual (CD)} \qquad \max_S \left\{\langle B, S\rangle_E : A^* S = c, S \in \mathbb{K}\right\}, \end{aligned} \tag{2}$$

where $A^* : \langle X, Ax\rangle_E \equiv x^\top A^* X$.

The above CP and CD can be reformulated in geometric form (cf. Section 1.4.4 in [1])

$$\begin{aligned} &\text{Primal (P)} \qquad \min_X \left\{\langle C, X\rangle_E : X \in (\mathcal{L} - B) \cap \mathbb{K}\right\} \\ &\text{Dual (D)} \qquad \max_S \left\{\langle B, S\rangle_E : S \in (\mathcal{L}^\perp + C) \cap \mathbb{K}\right\}, \end{aligned} \tag{3}$$

where $\mathcal{L} = \mathrm{Im}A \equiv \{Ax\}$ and $\mathcal{L}^\perp$ is the orthogonal complement of $\mathcal{L}$. Assume for simplicity that $\mathrm{Ker}A = \{0\}$, (P) and (D) are strictly feasible.

4 Primal-Dual Central Path

Path for (CP):

$$x_*(t) = \mathrm{argmin}_x \left[c^\top x + \frac{1}{t} K(Ax - B)\right]. \tag{4}$$

Path for (P):

$$\begin{aligned} X_*(t) &= Ax_*(t) - B \\ &= \mathrm{argmin}_{X \in (\mathcal{L} - B) \cap \mathrm{int}(\mathbb{K})} \langle C, X\rangle + \frac{1}{t} K(X). \end{aligned} \tag{5}$$

Path for (D):

$$S_*(t) = \mathrm{argmin}_{S \in (L^\perp + c) \cap \text{ int } \mathbb{K}} -t\langle B, S\rangle + K(S). \tag{6}$$

Theorem 2 (Thm 4.4.1, 4.4.2 in [1])**.** *For every $t > 0$,*

$$\begin{aligned} S_*(t) &= -t^{-1} \nabla K\left(X_*(t)\right), \\ X_*(t) &= -t^{-1} \nabla K\left(S_*(t)\right). \end{aligned} \tag{7}$$

$X_*(t)$ is fully characterized by 2 properties:

1. *$X_*(t)$ is strictly primal feasible.*
2. *$-t^{-1} \nabla K\left(X_*(t)\right)$ is strictly dual feasible.*

$S_*(t)$ is fully characterized by 2 properties:

1. *$S_*(t)$ is strictly dual feasible.*
2. *$-t^{-1} \nabla K\left(S_*(t)\right)$ is strictly primal feasible.*

Characterization of central path $(X_*(t), S_*(t))$:

1. *Primal feasibility: $X_*(t)$ strictly primal feasible.*
2. *Dual feasibility: $S_*(t)$ strictly dual feasible.*
3. *Augmented Complementary Slackness (ACS):*

$$S_*(t) + t^{-1} \nabla K\left(X_*(t)\right) = 0 \tag{8}$$

Proposition 3 (Prop 4.4.1 in [1])**.** *It holds that: Duality Gap $(X_*(t), S_*(t)) = t^{-1}\theta(K)$.*

Remark: This proposition is similar to Theorem 1 in Lec 18.

5 Distance to Central Path

We use the metric $\|H\|_Y = \sqrt{\left\langle [\nabla^2 K(Y)]^{-1} H, H \right\rangle}$. Given $Z = (X, S)$, $Z_*(t) = (X_*(t), S_*(t))$, we define there distance as

$$\text{dist}(z, Z_*(t)) \triangleq \|t \cdot S + \nabla K(X)\|_X = \sqrt{\left\langle [\nabla^2 K(X)]^{-1} (tS + \nabla K(X)), tS + \nabla K(X) \right\rangle}. \tag{9}$$

Proposition 4 (Properties of $\text{dist}(\cdot,\cdot)$, Page 283 of [1])**.** *The followings hold:*

1. $\text{dist}(Z, Z_*(t)) = 0$ *iff* $S = -t^{-1}\nabla K(X)$, *which implies* $Z = Z_*(t)$.
2. $\text{dist}(Z, Z_*(t)) \le 1 \;\Rightarrow\;$ *Duality Gap*$(X, S) \le 2\theta(K)/t$.

6 Tracing the Central Path

Given $(\bar{t}, \bar{X}, \bar{S})$ such that $\bar{X} \in \mathcal{L} - B$, $\bar{S} \in \mathcal{L}^+ + C$ and satisfies the system of nonlinear equations $G_{\bar{t}}(\bar{X}, \bar{S}) := \bar{S} + \bar{t}^{-1}\nabla K(\bar{X}) = 0$, we want to find another tuple (t_+, X_+, S_+) such that $t_+ > \bar{t}$, $X_+ \in \mathscr{L} - B$, $S_+ \in \mathcal{L}^+ + C$, and $G_{t_+}(X_t, S_t) = S_+ + t_+^{-1}\nabla K(X_+) = 0$. The idea to find such a tuple is the following:

- Find $t_+ > \bar{t}$.
- Linearize $G_{t_+}(X, S)$ at $(\bar{X}, \bar{S})$.
- Solving the system

$$\begin{cases} G_{t_+}(X, S) + \frac{\partial G_{t_+}(\bar{X}, \bar{X})}{\partial X}(\bar{X} - \bar{X}) + \frac{\partial G_{t_+}(\bar{X}, \bar{S})}{\partial S}(S - \bar{S}) = 0 \\ \Delta X = X - \bar{X} \in \mathcal{L} \\ \Delta S = S - \bar{S} \in \mathcal{L}^\perp \end{cases} \tag{10}$$

  which is equivalent to

$$\begin{cases} A^*[\nabla^2 K(\bar{X})] A \Delta X = -\left[t_+ c + A^* \nabla K(\bar{X})\right] \\ \Delta X = A \Delta X \\ A^* \Delta = 0 \end{cases} \tag{11}$$

  Upon solving the equation, let $(X_+, S_+) = (\bar{X} + \Delta X, \bar{S} + \Delta S)$.

We remark that, $x_+ = \bar{x} - \left[\nabla^2 F_{t_+}(x)\right]^{-1} \nabla F_{t_+}(x)$, where $F_{t_t}(x) = tc^\top x + K(Ax - B)$. The process described above is purely primal.

7 Special Example: IP for SDP

For the special case of semi-definite programs, consider the system of equation $G_t(X, S)$ as defined in Section 6:

$$G_t(X, S) \equiv S + t^{-1}\nabla K(X) = S - t^{-1}X^{-1} = 0 \tag{12}$$

Multiplying both sides by X and rearranging terms, we obtain

$$XS = t^{-1}I, \quad SX = t^{-1}I \Rightarrow XS + SX = 2t^{-1}I \tag{13}$$

Then, we apply Q-scaling. For every $Q \succ 0$, consider $\tilde{X} = QXQ, \tilde{S} = Q^{-1}SQ^{-1}$. With some rearrangements, we get

$$Q^{-1}SXQ = t^{-1}I, \quad QXSQ^{-1} = t^{-1}I \Rightarrow QXSQ^{-1} + Q^{-1}SXQ = 2t^{-1}I \Leftrightarrow \tilde{X}\tilde{S} + \tilde{S}\tilde{X} = 2t^{-1}I \tag{14}$$

Remark We make the following remarks:

1. Q-scaling: $X \to QXQ$ is one-to-one mapping from cone to itself.
2. Q is flexible as it is iteration-dependent. For $\forall \Delta X \in \mathcal{L}$ and $\Delta S \in \mathcal{L}^{\perp}$, we have

$$Q_i \left[\Delta X S_i + X_i \Delta S\right] Q_i^{-1} + Q_i^{-1} \left[\mathrm{S_i}\, \Delta X + \Delta S X_i\right] Q_i = 2t_{i+1}^{-1} I - Q_i X_i S_i Q_i^{-1} - Q_i^{-1} S_i X_i Q_i \tag{15}$$

3. Popular choice of Q:
   - Alizadeh-Haeberly-Overton method: $Q_i = I$.
   - XS-method: $Q_i = S_i^{\frac{1}{2}} \quad \Rightarrow \quad S_i^{\frac{1}{2}} X_i S_i^{\frac{1}{2}} = t^{-1} I$.
   - SX-method: $Q_i = X_i^{-\frac{1}{2}} \quad \Rightarrow \quad X_i^{\frac{1}{2}} S_i X_i^{\frac{1}{2}} = t^{-1} I$.
   - Nesterov-Todd: $Q_i = \left(X_i^{-\frac{1}{2}} \left(X_i^{\frac{1}{2}} S_i X_i^{\frac{1}{2}}\right)^{-\frac{1}{2}} X_i^{\frac{1}{2}} S_i\right)^{\frac{1}{2}} \quad \Rightarrow \quad \tilde{S} = \tilde{X}$.

References

[1] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications.* SIAM, 2001.