Lecture 9

Lecturer: Xiangyu Chang *Scribe: Xiangyu Chang*

Edited by: Junbo Hao

1 Interior-Point Method for Nonlinear Programming

Recall: Linear Programming.

$$\begin{aligned} \min_{\mathbf{x}} \ & \mathbf{c}^\top \mathbf{x}, \\ s.t. \ & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \succeq 0. \end{aligned}$$

Why we need the standard form? 1. For designing algorithm uniformly. 2. Any linear programming can be transformed into the standard form. For example, consider the following problem:

$$\begin{aligned} \min_{\mathbf{x}} \ & \mathbf{c}^\top \mathbf{x}, \\ s.t. \ & A\mathbf{x} \preceq \mathbf{b}. \end{aligned}$$

It is equivalent to

$$\begin{aligned} \min_{\mathbf{x}} \ & \mathbf{c}^\top \mathbf{x}, \\ s.t. \ & A\mathbf{x} + \mathbf{s} = \mathbf{b}, \\ & \mathbf{s} \succeq 0. \end{aligned}$$

Then, let $\mathbf{x} = \mathbf{x}^+ - \mathbf{x}^-$, where $\mathbf{x}^+ = \max\{\mathbf{x}, 0\}$ and $\mathbf{x}^- = \max\{-\mathbf{x}, 0\}$. Finally, it has

$$\min_{\mathbf{x}} \; [\mathbf{c}^\top, -\mathbf{x}^\top, 0]^\top \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{x}^- \\ \mathbf{s} \end{bmatrix}$$

$$s.t. \; [A, -A, I] \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{x}^- \\ \mathbf{s} \end{bmatrix} = \mathbf{b},$$

$$\begin{bmatrix} \mathbf{x}^+ \\ \mathbf{x}^- \\ \mathbf{s} \end{bmatrix} \succeq 0.$$

The Lagrangian is

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \mathbf{c}^\top \mathbf{x} + \boldsymbol{\nu}^\top (A\mathbf{x} - \mathbf{b}) - \boldsymbol{\lambda}^\top \mathbf{x}.$$

The KKT conditions of the standard linear programming are

$$\begin{aligned} A^\top \boldsymbol{\nu} + \mathbf{c} &= \boldsymbol{\lambda}, \\ A\mathbf{x} &= \mathbf{b}, \\ \mathbf{x} &\succeq 0, \\ x_i \lambda_i &= 0, \\ \boldsymbol{\lambda} &\succeq 0. \end{aligned}$$

This is equivalent to

$$\begin{aligned} A^\top \boldsymbol{\nu} + \mathbf{c} &= \boldsymbol{\lambda}, \\ A\mathbf{x} &= \mathbf{b}, \\ \mathbf{x} &\succeq 0, \\ \boldsymbol{\lambda} &\succeq 0, \\ \bar{X}\bar{\boldsymbol{\lambda}}\mathbb{1} &= 0, \end{aligned}$$

where $\bar{X} = diag(\mathbf{x}), \bar{\boldsymbol{\lambda}} = diag(\boldsymbol{\lambda})$ and $\mathbb{1} = (1, \ldots, 1)^\top$. Thus, solving the LP is to find the solution of

$$F_0(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{pmatrix} A^\top \boldsymbol{\nu} + \mathbf{c} - \boldsymbol{\lambda} \\ A\mathbf{x} - \mathbf{b} \\ \bar{X}\bar{\boldsymbol{\lambda}}\mathbb{1} \end{pmatrix} = 0,$$

where $\boldsymbol{\lambda} \succeq 0$ and $\boldsymbol{\nu} \succeq 0$. We can use Newton's method with line search to handle this problem.

The conditions $\boldsymbol{\lambda} \succeq 0$ and $\boldsymbol{\nu} \succeq 0$ lead to the significant hurdle of solving $F(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = 0$. How can we overcome this difficulty?

Let us consider

$$\min_{\mathbf{x}} \ \mathbf{c}^\top \mathbf{x} - \mu \sum_i \log x_i,$$
$$s.t. \ \ A\mathbf{x} = \mathbf{b}.$$

The Lagrangian is

$$L_\mu(\mathbf{x}, \boldsymbol{\nu}) = \mathbf{c}^\top \mathbf{x} - \mu \sum_i \log x_i + \boldsymbol{\nu}^\top (A\mathbf{x} - \mathbf{b}).$$

We compute

$$\frac{\partial L_\mu(\mathbf{x}, \boldsymbol{\nu})}{\partial x_i} = c_i - \mu / x_i + A_i^\top \boldsymbol{\nu}.$$

If we further assume that $\mu / x_i = \lambda_i$, then the KKT conditions of the standard linear programming are

$$A^\top \boldsymbol{\nu} + \mathbf{c} = \boldsymbol{\lambda},$$
$$A\mathbf{x} = \mathbf{b},$$
$$\bar{X}\bar{\boldsymbol{\lambda}}\mathbb{1} = \mu\mathbb{1},$$
$$\mathbf{x} \succ 0.$$

Thus, solving the LP is to find the solution of

$$F_\mu(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{pmatrix} A^\top \boldsymbol{\nu} + \mathbf{c} - \boldsymbol{\lambda} \\ A\mathbf{x} - \mathbf{b} \\ \bar{X}\bar{\boldsymbol{\lambda}}\mathbb{1} - \mu\mathbb{1} \end{pmatrix} = 0.$$

Solving $F_\mu(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = 0$ to obtain $(\mathbf{x}(\mu), \boldsymbol{\lambda}(\mu), \boldsymbol{\nu}(\mu))$, then let $\mu \to 0$.

Quadratic Programming: we consider

$$\min_{\mathbf{x}} \ \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|^2,$$
$$s.t. \ \ C\mathbf{x} \preceq \mathbf{d}.$$

Using slack variables (barrier method) is to obtain the equivalent problem

$$\min_{\mathbf{x}} \ \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|^2,$$
$$s.t. \ \ C\mathbf{x} + \mathbf{s} = \mathbf{d},$$
$$\mathbf{s} \succeq 0.$$

This is equivalent to

$$\min_{\mathbf{x}} \ \frac{1}{2}\|A\mathbf{x}-\mathbf{b}\|^2 - \mu\sum_i \log s_i,$$
$$s.t. \ \ C\mathbf{x}+\mathbf{s}=\mathbf{d}.$$

The Lagrangian function is

$$L_\mu(\mathbf{x},\mathbf{s},\boldsymbol{\nu}) = \frac{1}{2}\|A\mathbf{x}-\mathbf{b}\|^2 - \mu\sum_i \log s_i + \boldsymbol{\nu}^\top(C\mathbf{x}+\mathbf{s}-\mathbf{d}).$$

Thus, its KKT conditions are

$$A^\top A\mathbf{x} - A^\top\mathbf{b} + C^\top\boldsymbol{\nu} = 0,$$
$$C\mathbf{x}+\mathbf{s}-\mathbf{d} = 0$$
$$\bar{V}\bar{S}\mathbb{1} = \mu\mathbb{1},$$

where $\bar{V} = diag(\boldsymbol{\nu}), \bar{S} = diag(\mathbf{s})$ and $\mathbb{1} = (1,\ldots,1)^\top$.

Let

$$F_\mu(\mathbf{x},\mathbf{s},\boldsymbol{\nu}) = \begin{pmatrix} A^\top A\mathbf{x} - A^\top\mathbf{b} + C^\top\boldsymbol{\nu} \\ C\mathbf{x}+\mathbf{s}-\mathbf{d} \\ \bar{V}\bar{S}\mathbb{1} - \mu\mathbb{1} \end{pmatrix}.$$

Solving QP is to find the solution of $F_\mu(\mathbf{x},\mathbf{s},\boldsymbol{\nu}) = 0$. And the real KKT system is $F_0(\mathbf{x},\mathbf{s},\boldsymbol{\nu}) = 0$.

We summarize Algorithm 1 for solving $F_0(\mathbf{x},\mathbf{s},\boldsymbol{\nu}) = 0$ approximately.

For QP, Eq.(1) is a linear system. For example, given μ, and we can compute that

$$\nabla F_\mu(\mathbf{s},\boldsymbol{\nu},\mathbf{x}) = \begin{pmatrix} I & 0 & C \\ \bar{V} & \bar{S} & 0 \\ 0 & C^\top & A^\top A \end{pmatrix}.$$

Denote that $\mathbf{r}_1 = C\mathbf{x}+\mathbf{s}-\mathbf{d}$, $\mathbf{r}_2 = \bar{V}\bar{S}\mathbb{1} - \mu\mathbb{1}$,$\mathbf{r}_3 = A^\top A\mathbf{x} - A^\top\mathbf{b} + C^\top\boldsymbol{\nu}$, then Eq.(1) is

$$\begin{pmatrix} I & 0 & C \\ \bar{V} & \bar{S} & 0 \\ 0 & C^\top & A^\top A \end{pmatrix}\begin{pmatrix} \Delta\mathbf{s} \\ \Delta\boldsymbol{\nu} \\ \Delta\mathbf{x} \end{pmatrix} = \begin{pmatrix} -\mathbf{r}_1 \\ -\mathbf{r}_2 \\ -\mathbf{r}_3 \end{pmatrix}.$$

Using Gaussian elimination method to solve the linear system as the following three steps.

- Step 1: $R_2 \leftarrow R_2 - \bar{V}R_1$, that is

$$\begin{pmatrix} I & 0 & C \\ 0 & \bar{S} & -\bar{V}C \\ 0 & C^\top & A^\top A \end{pmatrix}\begin{pmatrix} \Delta\mathbf{s} \\ \Delta\boldsymbol{\nu} \\ \Delta\mathbf{x} \end{pmatrix} = \begin{pmatrix} -\mathbf{r}_1 \\ \bar{V}\mathbf{r}_1 - \mathbf{r}_2 \\ -\mathbf{r}_3 \end{pmatrix}.$$

Algorithm 1 Interior Point Method for QP

1: **Input:** Given a initial starting point $\mathbf{x}^0, \mathbf{s}^0, \boldsymbol{\nu}^0, \mu^0 = 1, \epsilon$, and $t = 0$

2: **while** $\|F_{\mu^t}(\mathbf{x}^t, \mathbf{s}^t, \boldsymbol{\nu}^t)\| \geqslant \epsilon$ **do**

3: Get an update direction $\Delta\mathbf{s}, \Delta\boldsymbol{\nu}, \Delta\mathbf{x}$ that satisfies

$$\nabla F_{\mu^t}(\mathbf{x}^t, \mathbf{s}^t, \boldsymbol{\nu}^t) \begin{pmatrix} \Delta\mathbf{s} \\ \Delta\boldsymbol{\nu} \\ \Delta\mathbf{x} \end{pmatrix} = -F_{\mu^t}(\mathbf{x}^t, \mathbf{s}^t, \boldsymbol{\nu}^t). \tag{1}$$

4: Update

$$\begin{pmatrix} \mathbf{s}^{t+1} \\ \boldsymbol{\nu}^{t+1} \\ \mathbf{x}^{t+1} \end{pmatrix} = \begin{pmatrix} \mathbf{s}^{t} \\ \boldsymbol{\nu}^{t} \\ \mathbf{x}^{t} \end{pmatrix} + \alpha \begin{pmatrix} \Delta\mathbf{s} \\ \Delta\boldsymbol{\nu} \\ \Delta\mathbf{x} \end{pmatrix},$$

where α is chosen by the line search method and ensure that

$$\|F_{\mu^t}(\mathbf{x}^{t+1}, \mathbf{s}^{t+1}, \boldsymbol{\nu}^{t+1}) \leqslant 0.99\|F_{\mu^t}(\mathbf{x}^t, \mathbf{s}^t, \boldsymbol{\nu}^t)\|, \tag{2}$$

$$\mathbf{s}^{t+1} \succeq 0, \tag{3}$$

$$\boldsymbol{\nu}^{t+1} \succeq 0. \tag{4}$$

5: Update $\mu^{t+1} = \frac{0.1}{n}\left\langle \mathbf{s}^{t+1}, \boldsymbol{\nu}^{t+1} \right\rangle$ (this is also called "duality measure").

6: $t := t + 1$.

7: **end while**

8: **Output:** $(\mathbf{x}^T, \mathbf{s}^T, \boldsymbol{\nu}^T)$.

- Step 2:

  $R_3 \leftarrow R_3 - C^\top \bar{S}^{-1} R_2$, that is

$$\begin{pmatrix} I & 0 & C \\ 0 & \bar{S} & -\bar{V}C \\ 0 & 0 & A^\top A + C^\top \bar{S}^{-1}\bar{V}C \end{pmatrix} \begin{pmatrix} \Delta\mathbf{s} \\ \Delta\boldsymbol{\nu} \\ \Delta\mathbf{x} \end{pmatrix} = \begin{pmatrix} -\mathbf{r}_1 \\ \bar{V}\mathbf{r}_1 - \mathbf{r}_2 \\ -\mathbf{r}_3 - C^\top \bar{S}^{-1}(\bar{V}\mathbf{r}_1 - \mathbf{r}_2) \end{pmatrix}.$$

- Step 3:

$$\Delta\mathbf{x} = (A^\top A + C^\top \bar{S}^{-1}\bar{V}C)^{-1}(-\mathbf{r}_3 - C^\top \bar{S}^{-1}(\bar{V}\mathbf{r}_1 - \mathbf{r}_2)).$$

General Case:

$$\min_{\mathbf{x}} \ f(\mathbf{x}),$$
$$s.t. \ f_i(\mathbf{x}) \leqslant 0.$$

This is equivalent to

$$
\begin{aligned}
\min_{\mathbf{x}} \; & f(\mathbf{x}) - \mu \sum_i \log s_i, \\
s.t. \; & f_i(\mathbf{x}) + s_i = 0.
\end{aligned}
$$

The Lagrangian function is

$$L_\mu(\mathbf{x}, \mathbf{s}, \boldsymbol{\nu}) = f(\mathbf{x}) - \mu \sum_i \log s_i + \sum_i \nu_i (f_i(\mathbf{x}) + s_i).$$

Then the KKT system is

$$
\begin{aligned}
\nabla f(\mathbf{x}) + \sum_i \nu_i \nabla f_i(\mathbf{x}) &= 0, \\
\mathbf{s} + F(\mathbf{x}) &= 0 \\
\bar{V}\bar{S}\mathbb{1} - \mu\mathbb{1} &= 0,
\end{aligned}
$$

where $\bar{V} = diag(\boldsymbol{\nu}), \bar{S} = diag(\mathbf{s}), F(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))^\top$ and $\mathbb{1} = (1, \ldots, 1)^\top$.

Let

$$G_\mu(\mathbf{x}, \mathbf{s}, \boldsymbol{\nu}) = \begin{pmatrix} f(\mathbf{x}) + \sum_i \nu_i \nabla f_i(\mathbf{x}) \\ \mathbf{s} + F(\mathbf{x}) \\ \bar{V}\bar{S}\mathbb{1} - \mu\mathbb{1} \end{pmatrix}.$$

Solving the general optimization problem is to find the solution of $G_\mu(\mathbf{x}, \mathbf{s}, \boldsymbol{\nu}) = 0$. And the real KKT system is $G_0(\mathbf{x}, \mathbf{s}, \boldsymbol{\nu}) = 0$. The similar algorithm with Algorithm 1 could be designed.