Lecture 17: CS395T Numerical Optimization for Graphics and AI — Linear Programming (Interior Point Method II

Qixing Huang
The University of Texas at Austin
huangqx@cs.utexas.edu

Disclaimer

This note is adapted from

- Section 14 of *Numerical Optimization* by Jorge Nocedal and Stephen J. Wright. Springer series in operations research and financial engineering. Springer, New York, NY, 2. ed. edition, (2006)

1 Interior Point Methods

This lecture we will work through the proof of a convergence analysis of an interior point method.

The Central Path. The primal-dual feasible set $\mathcal{F}$ and *strictly feasible set* $\mathcal{F}^o$ are defined as follows:

$$\mathcal{F} = \{(\boldsymbol{x}, \lambda, \boldsymbol{s}) | A\boldsymbol{x} = \boldsymbol{b}, A^T\lambda + \boldsymbol{s} = \boldsymbol{c}, (\boldsymbol{x}, \boldsymbol{s}) \geq 0\}, \mathcal{F}^o \quad = \{(\boldsymbol{x}, \lambda, \boldsymbol{s}) | A\boldsymbol{x} = \boldsymbol{b}, A^T\lambda + \boldsymbol{s} = \boldsymbol{c}, (\boldsymbol{x}, \boldsymbol{s}) > 0\}. \tag{1}$$

The central path $\mathcal{C}$ is an arc of strictly feasible points that plays a vital role in primal-dual algorithms. It is parametrized by a scalar $\tau > 0$, and each point $(\boldsymbol{x}^\tau, \lambda^\tau, \boldsymbol{s}^\tau) \in \mathcal{C}$ satisfies the following equations:

$$A^T\lambda + \boldsymbol{s} = \boldsymbol{c}, \tag{2}$$
$$A\boldsymbol{x} = \boldsymbol{b}, \tag{3}$$
$$x_i s_i = \tau, \quad i = 1, 2, \cdots, n, \tag{4}$$
$$(\boldsymbol{x}, \boldsymbol{s}) > 0 \tag{5}$$

These conditions differ from the KKT conditions only in the term τ on the right-hand side of (4). Instead of the complementarity condition (4), we require that the pairwise products $x_i s_i$ have the same (positive) value for all indices i. From (2)-(5), we can define the central path as $\mathcal{C} = \{(\boldsymbol{x}^\tau, \lambda^\tau, \boldsymbol{s}^\tau) | \tau > 0\}$. It can be shown that $(\boldsymbol{x}^\tau, \lambda^\tau, \boldsymbol{s}^\tau$ is defined uniquely for each $\tau > 0$ if and only if $\mathcal{F}^o$ is nonempty. The conditions (2)-(5) are also the optimality conditions for a logarithmic-barrier formulation. By introducing log-barrier terms for the nonnegativity constraints, with barrier parameter $\tau > 0$, we obtain

$$\min \ \boldsymbol{c}^T\boldsymbol{x} - \tau \sum_{i=1}^{n} \log(x_i), \quad \text{subject to} \quad A\boldsymbol{x} = \boldsymbol{b}. \tag{6}$$

Central Path Neighborhoods and Path-Following Methods. Path-following algorithms explicitly restrict the iterates to a neighborhood of the central path $\mathcal{C}$ and follow $\mathcal{C}$ to a solution of the linear program. By preventing the iterates from coming too close to the boundary of the nonnegative orthant, they ensure that it is possible to take a nontrivial step along each search direction. Moreover, by forcing the duality measure μ_k to zero as $k \to \infty$, we ensure that the iterates $(\boldsymbol{x}^k, \lambda^k, \boldsymbol{s}^k)$ come closer and closer to satisfying the KKT conditions.

The two most interesting neighborhoods of $\mathcal{C}$ are

$$\mathcal{N}_2(\theta) = \{(\boldsymbol{x}, \lambda, \boldsymbol{s}) \in \mathcal{F}^o | \|XS\boldsymbol{e} - \mu\boldsymbol{e}\| \leq \theta\mu\}, \tag{7}$$

for some $\theta \in [0, 1)$, and

$$\mathcal{N}_{-\infty}(\gamma) = \{(\boldsymbol{x}, \lambda, \boldsymbol{s}) \in \mathcal{F}^o | x_i s_i \geq \gamma\mu, \quad i = 1, 2, \cdots, n\}, \tag{8}$$

for some $\gamma \in (0, 1]$. (Typical values of the parameters are $\theta = 0.5$ and $\gamma = 10^{-3}$.) If a point lies in $\mathcal{N}_{-\infty}(\gamma)$, each pairwise product $x_i s_i$ must be at least some small multiple γ of their average value μ. This requirement is actually quite modest, and we can make $\mathcal{N}_{-\infty}(\gamma)$ encompass most of the feasible region $\mathcal{F}$ by choosing γ close to zero. The $\mathcal{N}_2(\theta)$ neighborhood is more restrictive, since certain points in $\mathcal{F}^o$ do not belong to $\mathcal{N}_2(\theta)$ no matter how close θ is chosen to its upper bound of 1. By keeping all iterates inside one or other of these neighborhoods, path-following methods reduce all the pairwise products $x_i s_i$ to zero at more or less the same rate.

Long-Step Path-Following. The pesudo-code of the algorithm we want to discuss is given below:

- Given $\gamma, \sigma_{\min}, \sigma_{\max}$ with $\gamma \in (0, 1), 0 < \sigma_{\min} \leq \sigma_{\max} < 1$, and $(\boldsymbol{x}^0, \lambda^0, \boldsymbol{s}^0) \in \mathcal{N}_{-\infty}(\gamma)$;
- **for** $k = 0, 1, 2, \ldots,$
- Choose $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$;
- Solve the following linear system to obtain $(\Delta\boldsymbol{x}^k, \Delta\lambda^k, \Delta\boldsymbol{s}^k)$:

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{bmatrix} \cdot \begin{bmatrix} \Delta\boldsymbol{x}^k \\ \Delta\lambda^k \\ \Delta\boldsymbol{s}^k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -X^k S^k \boldsymbol{e} + \sigma\mu^k \boldsymbol{e} \end{bmatrix}. \tag{9}$$

- Choose α_k as the largest value of α in $[0, 1]$ such that

$$(\boldsymbol{x}^k(\alpha), \lambda^k(\alpha), \boldsymbol{s}^k(\alpha)) \in \mathcal{N}_{-\infty}(\gamma)$$

;

- Set $(\boldsymbol{x}^{k+1}, \lambda^{k+1}, \boldsymbol{s}^{k+1}) = (\boldsymbol{x}^k(\alpha^k), \lambda^k(\alpha^k), \boldsymbol{s}^k(\alpha^k))$.
- **end(for)**

Proof Architecture.

Lemma 1.1. *Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be any two vectors in $\mathbb{R}^n$ with $\boldsymbol{u}^T\boldsymbol{v} \geq 0$. Then*

$$\|UV\boldsymbol{e}\| \leq 2^{-\frac{3}{2}} \|\boldsymbol{u} + \boldsymbol{v}\|_2^2,$$

where

$$U = \mathrm{diag}(u_1, \cdots, u_n), \quad V = \mathrm{diag}(v_1, \cdots, v_n).$$

Lemma 1.2. *If $(\boldsymbol{x}, \lambda, \boldsymbol{s}) \in \mathcal{N}_{-\infty}(\gamma)$, then*

$$\|\Delta X \Delta S\boldsymbol{e}\| \leq 2^{-\frac{3}{2}}(1 + \frac{1}{\gamma})n\mu.$$

Theorem 1.1. *Given the parameters $\gamma, \sigma_{\min}, \sigma_{\max}$, there is a constant δ independent of n such that*

$$\mu_{k+1} \leq (1 - \frac{\delta}{n})\mu_k,$$

for all $k \geq 0$.

Theorem 1.2. *Given $\epsilon \in (0, 1)$ and $\gamma \in (0, 1)$, suppose the starting point satisfies $(\boldsymbol{x}^0, \lambda^0, \boldsymbol{s}^0) \in \mathcal{N}_{-\infty}(\gamma)$. Then there is an index K with $K = O(n \log(1/\epsilon)$ such that*

$$\mu_k \leq \epsilon\mu_0, \qquad \text{for all } k \geq K.$$