Course Notes for EE227C (Spring 2018): Convex Optimization and Approximation

Instructor: Moritz Hardt

Email: hardt+ee227c@berkeley.edu

Graduate Instructor: Max Simchowitz

Email: msimchow+ee227c@berkeley.edu

October 15, 2018

26 Primal-dual interior point methods

Previously, we discussed barrier methods and the so-called short-step method for solving constrained LPs, and proved that convergence is guaranteed (albeit slow). Herein, we study a primal-dual interior-point method (the so-called "long-step" path following algorithm), which similarly seeks to approximate points on the central path. Unlike the short-step method, the long-step method considers iterates of primal-dual variables and seeks a more aggressive step size so long as it lies within a neighborhood of the central path.

26.1 Deriving the dual problem

Let $x \in \mathbb{R}^n$ be the decision variable, and $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$

$$\min c^\top x \quad \text{s.t.} \quad Ax = b, x \geqslant 0$$

Here, $\geqslant$ is pointwise.

Observe that we can always write the constraint as

$$\begin{aligned} & \min_{x\geqslant 0} c^\top x + \max_z z^\top(b - Ax) \\ = \ & \min_{x\geqslant 0} \max_z c^\top x + \max_z z^\top(b - Ax) \\ \geqslant \ & \max_z z^\top b - \min_{x\geqslant 0}(c - A^\top z)^\top x \\ = \ & \max_z z^\top b - \infty \cdot \mathbb{I}(A^\top z > c) \end{aligned}$$

Hence, the dual problem is as follows

$$\max b^\top z \quad \text{s.t.} \quad A^\top z \leqslant c$$

This is equivalent to

$$\max b^\top z \quad \text{s.t.} \quad A^\top z + s = c, \quad s \geqslant 0$$

where we introduce the slack variable $s \in \mathbb{R}^n$. If (x, z, s) are only *feasible*, then

$$Ax = b \quad A^\top z + s = c \quad x, s \geqslant 0$$

Moreover, we can compute that for feasible (x, z, s),

$$0 \leqslant \langle x, s\rangle = x^\top(c - A^\top z) = \langle x, c\rangle - \langle Ax, z\rangle = \langle x, c\rangle - \langle b, z\rangle \,.$$

This is a proof of weak duality, namely that for any feasible x and z,

$$\langle x, c\rangle \geqslant \langle b, z\rangle$$

and therefore

$$\langle x^*, c\rangle \geqslant \langle b, z^*\rangle$$

Moreover, if there exists an feasible (x^*, z^*, s^*), with $\langle x^*, s^*\rangle = 0$ then we have

$$\langle x^*, c\rangle = \langle b^*, z\rangle$$

which is strong duality.

Duality is also useful to bound the suboptimality gap, since in fact if (x, z, s) is feasible, then

$$\langle x, s\rangle = \langle x, c\rangle - \langle b, z\rangle \geqslant \langle x, c\rangle - \langle x^*, c\rangle = \langle x - x^*, c\rangle$$

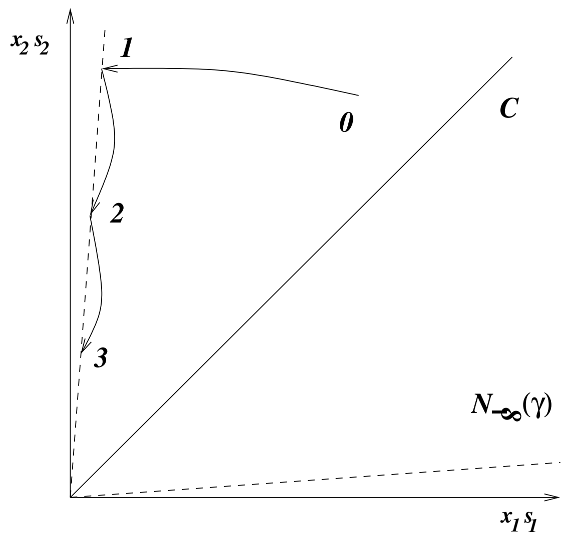

Figure 1: Require iterates to stay within a certain neighborhood of the central path. We want the pairwise products $x_i s_i$ to be not too different for $i = 1, 2, ..., n$.

26.2 Primal-dual iterates along the central path

The above suggests the following approach. Consider iterates (x_k, z_k, s_k), and define

$$\mu_k := \frac{1}{n} \cdot \langle x_k, s_k \rangle = \frac{\langle x_k, c\rangle - \langle b, z_k\rangle}{n} \geqslant \frac{\langle x_k - x^*, c\rangle}{n}$$

Define the strictly feasible set

$$\mathcal{F}^o := \{Ax = b \quad A^\top z + s = c \quad x, s > 0\}$$

Minimizing μ_k thus amounts to minimizing a *bilinear* objective over a linear constraint set. The goal is to generate iterates $(x^{k+1}, z^{k+1}, s^{k+1})$ such that

$$\mu_{k+1} \leqslant (1 - Cn^{-\rho})\mu_k$$

This implies that

$$\langle x_k - x^*, c\rangle \leqslant \epsilon \text{ in } k = \mathcal{O}\left(n^\rho \log(n/\epsilon)\right) \text{steps.}$$

The goal is to find a tuple (x, z, s) such that $\mu \approx 0$. We consider the following approach. Define

$$F_\tau(x,z,s) := \begin{bmatrix} Ax-b \\ A^\top z + s - c \\ x \circ s - \tau \mathbf{1} \end{bmatrix}$$

Then the goal is to approx solve $F_0(x_0, z_0, s_0) = \mathbf{0}$ over $\mathcal{F}^0$. We see that this can be obtained by computing the solutions (x_τ, z_τ, z_τ) to $F_\tau(x,z,s) = \mathbf{0}$. We call the curve $\tau \mapsto (x_\tau, z_\tau, z_\tau)$ the "central path". Note that, on the central path, $x_i s_i = \tau$ for some $\tau > 0$. To ensure we stay close to the central path, we consider

$$\mathcal{N}_{-\infty}(\gamma) \quad := \quad \{(x,z,s) \in \mathcal{F}_0 : \min_i x_i s_i \geqslant \gamma\mu(x,s)\}$$

What we would like to do is take iterates (x_k, z_k, s_k) such that μ_k decreases, and $(x_k, z_k, s_k) \in \mathcal{N}_{-\infty}(\gamma)$ for appropriate constants γ. $\mathcal{N}_{-\infty}(\gamma)$ ensures the nonnegativity contraints. This is portrayed in Figure 1.

Input: Parameters $\gamma \in (0,1)$, $0 < \sigma_{\min} < \sigma_{\max} < 1$, and initialization $(x^0, z^0, t^0) \in \mathcal{N}_{-\infty}(\gamma)$. **for** $t = 0, 1, 2, \ldots$ **do**

Choose $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$;

Run Newton step on $F_{\sigma_k \mu_k}$ (to be defined). Let $(\Delta x^k, \Delta z^k, \Delta s^k)$ denote the Newton step

$$(\Delta x^k, \Delta z^k, \Delta s^k) = -\nabla^2 F_{\tau_k}(w^k)^{-1} \cdot \nabla F_{\tau_k}(w^k),$$
$$\text{where } \tau_k = \sigma_k \mu_k \text{ and } w^k = (x^k, z^k, s^k)\ .$$

Let $\alpha_k \in (0,1]$ be the largest step such that

$$\alpha_k = \max\{\alpha \in (0,1] : (x^k, z^k, s^k) + \alpha(\Delta x^k, \Delta z^k, \Delta s^k) \in \mathcal{N}_\infty(\gamma)\}$$

Set $(x^{k+1}, z^{k+1}, s^{k+1}) \leftarrow (x^k, z^k, s^k) + \alpha_k(\Delta x^k, \Delta z^k, \Delta s^k)$.

end

Algorithm 1: Long-step Path Following method

26.3 Generating Iterates with the Newton Step

The Newton Step for solving fixed point equations $F(w) = 0$. Indeed

$$F(w+d) = F(w) + J(w) \cdot d + o(\|d\|)$$

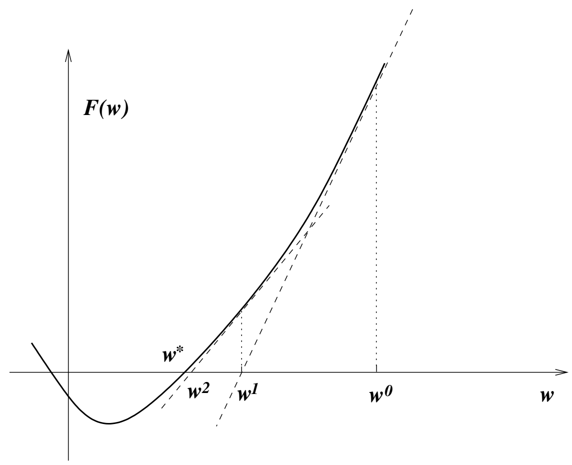

Figure 2: Recall that Newton's method iteratively finds better approximations to the roots (or zeroes) of a real-valued function.

The Newton's method then chooses $w \leftarrow w + d$,

$$J(w)d = -F(w)$$

Which implies that $F(w+d) = o(\|d\|)$ for w sufficiently closed to the fixed point. This gives you the quick converge. Note that, if F is a linear map, then in fact *one Newton step suffices*. This can be seen from the Taylor expansion.

Our function F_{τ_k} is nearly linear, but not quite. Let's compute the Newton Step. We observe that the Jacobian is the linear operator

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^\top & I \\ \mathrm{Diag}(S) & 0 & \mathrm{Diag}(X) \end{bmatrix}$$

Moreover, since $(x^k, z^k, s^k) \in \mathcal{F}^0$, we have that

$$F_{\tau_k}(x^k, z^k, s^k) = \begin{bmatrix} Ax^k - b \\ A^\top z^k + s^k - c \\ x^k \circ s^k - \tau_k \mathbf{1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ x^k \circ s^k - \tau_k \mathbf{1} \end{bmatrix}$$

Let's drop subscripts. Then, on one can verify that the Newton satisfies

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^\top & I \\ \mathrm{Diag}(S) & 0 & \mathrm{Diag}(X) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \\ \Delta s \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -x \circ s + \tau \mathbf{1} \end{bmatrix}$$

Some remarks

1. $A\Delta x = 0$, and that $A^\top \Delta z + \Delta_s = 0$.

2. This implies that $(x^+, z^+, s^+) := (x + \Delta x, z + \Delta z, s + \Delta s)$ satisfies

$$Ax^+ - b = 0 \quad \text{and} \quad A^\top z^+ + s - c = 0$$

3. We also have

$$s \circ \Delta x + x \circ \Delta s = -x \circ s + \tau \mathbf{1}$$

and thus

$$\begin{aligned} x^+ \circ s^+ &= x \circ s + (\circ \Delta x + x \circ \Delta s) + \Delta x \circ \Delta s \\ &= x \circ s - x \circ s + \tau \mathbf{1} + \Delta x \circ \Delta s \end{aligned}$$

4. Thus,

$$F_\tau(x^+ \circ s^+) = \begin{bmatrix} 0 \\ 0 \\ \Delta x \circ \Delta s \end{bmatrix}$$

In other words, if we can argue that $\Delta x \circ \Delta s$ is "negligible", then the Newton step produces an almost exact solution.

A more concrete analysis would be to study the term

$$\begin{aligned} n\mu(x + \alpha \Delta x, s + \alpha \Delta x) &= \langle x + \Delta x, s + \Delta s \rangle \\ &= \langle x, s \rangle + \alpha \left(\langle x, \Delta s \rangle + \langle s, \Delta x \rangle \right) + \alpha^2 \langle \Delta s, \Delta x \rangle \end{aligned}$$

The last term in the above display vanishes, as shown by the above

$$0 = \Delta x^\top (A^\top \Delta z + \Delta s) = (A \Delta x)^\top z + \langle \Delta x, \Delta s \rangle = \langle \Delta x, \Delta \rangle$$

Moreover, since $s \circ \Delta x + x \circ \Delta s = -x \circ s + \tau \mathbf{1}$, we have by summing that

$$\langle x, \Delta s \rangle + \langle s, \Delta x \rangle = -\langle x, x \rangle + \tau n = -(1 - \sigma) \langle x, s \rangle$$

where the last line uses $n\tau = n\sigma\mu = \sigma n \tau = \sigma \langle x, s \rangle$. Hence,

$$n\mu(x + \alpha \Delta x, s + \alpha \Delta x) = n\mu(x, s)(1 - \alpha(1 - \sigma))$$

Hence, if one can show that $(1 - \alpha)\alpha \geqslant C(n) > 0$ for some constant depending on the dimension, then we see that

$$n\mu(x^{k+1}) \leqslant (1 - C(n))^k n\mu(x^0)$$

giving us the rate of decrease. One can then show with more technical work that $\alpha = \Omega(1/n)$ while maintaining the $\mathcal{N}_{-\infty}(\gamma)$ invariant.

References