Primal-Dual Interior-Point Methods

Lecturer: Aarti Singh
Co-instructor: Pradeep Ravikumar

Convex Optimization 10-725/36-725

Outline

Today:

- Primal-dual interior-point method
- Special case: linear programming

Barrier method versus primal-dual method

Like the barrier method, primal-dual interior-point methods aim to compute (approximately) points on the central path.

Main differences between primal-dual and barrier methods:

- Both can be motivated by perturbed KKT conditions, but as the name suggests primal-dual methods update both primal and dual variables
- Primal-dual interior-point methods usually take one Newton step per iteration (no additional loop for the centering step).
- Primal-dual interior-point methods are not necessarily feasible.
- Primal-dual interior-point methods are typically more efficient. Under suitable conditions they have better than linear convergence.

Constrained Optimization

Consider the problem

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & Ax = b \\ & g(x) \leq 0 \end{aligned}$$

where the equality constraints are linear.

Lagrangian

$$L(x, u, v) = f(x) + u^\mathsf{T} g(x) + v^\mathsf{T}(Ax - b)$$

KKT conditions

KKT conditions

$$\begin{aligned} \nabla f(x) + \nabla g(x) u + A^\mathsf{T} v &= 0 \\ U g(x) &= 0 \\ Ax &= b \\ u, -g(x) &\geq 0. \end{aligned}$$

Here $U = \mathrm{Diag}(u),\ \nabla g(x) = \begin{bmatrix} \nabla g_1(x) & \cdots & \nabla g_r(x) \end{bmatrix}$

KKT conditions for Barrier problem

Barrier problem

$$\min_x \quad f(x) + \epsilon\phi(x)$$
$$Ax = b$$

where

$$\phi(x) = -\sum_{j=1}^{r} \log(-g_j(x)).$$

KKT conditions for barrier problem

$$\begin{aligned} \nabla f(x) + \nabla g(x)u + A^{\mathsf{T}}v &= 0 \\ Ug(x) &= -\epsilon\mathbf{1} \\ Ax &= b \\ u, -g(x) &> 0. \end{aligned}$$

Same as before, except complementary slackness condition is perturbed.

We didn't cover this, but Newton updates for log barrier problem can be seen as Newton step for solving these nonlinear equations, after eliminating u (i.e., taking $u_j = -\epsilon/g_j(x), j = 1, \ldots, r$).

Primal-dual interior-point updates are also motivated by a Newton step for solving these nonlinear equations, but without eliminating u. Write the KKT conditions as a set of nonlinear equations

$$r(x; u; v) = 0,$$

where

$$r(x, u, v) := \begin{bmatrix} \nabla f(x) + \nabla g(x)u + A^\mathsf{T} v \\ Ug(x) + \epsilon \mathbf{1} \\ Ax - b \end{bmatrix}$$

This is a nonlinear equation is $(x; u; v)$, and hard to solve; so let's linearize, and approximately solve

Let $y = (x; u; v)$ be the current iterate, and $\Delta y = (\Delta x; \Delta u; \Delta v)$ be the update direction. Define

$$\begin{aligned} r_{\text{dual}} &= \nabla f(x) + \nabla g(x)u + A^\top v \\ r_{\text{cent}} &= Ug(x) + \epsilon \mathbf{1} \\ r_{\text{prim}} &= Ax - b \end{aligned}$$

the dual, central, and primal residuals at current $y = (x; u; v)$.

Now we make our first-order approximation

$$0 = r(y + \Delta y) \approx r(y) + \nabla r(y)\Delta y$$

and we want to solve for Δy in the above.

I.e., we solve

$$\begin{bmatrix} \nabla^2 f(x) + \sum_{j=1}^r u_j \nabla^2 g_j(x) & \nabla g(x)^T & A^T \\ U \nabla g(x) & \mathrm{diag}(g(x)) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta u \\ \Delta v \end{pmatrix} = - \begin{pmatrix} r_{\mathrm{dual}} \\ r_{\mathrm{cent}} \\ r_{\mathrm{prim}} \end{pmatrix}$$

Solution $\Delta y = (\Delta x, \Delta u, \Delta v)$ is our primal-dual update direction

Note that the update directions for the primal and dual variables are inexorably linked together

(Also, these are different updates than those from barrier method)

Primal-dual interior-point method

Putting it all together, we now have our primal-dual interior-point method. Start with a strictly feasible point $x^{(0)}$ and $u^{(0)} > 0$, $v^{(0)}$. Define $\eta^{(0)} = -g(x^{(0)})^T u^{(0)}$ and let $\sigma \in (0, 1)$, then we repeat for $k = 1, 2, 3 \ldots$

Primal-dual interior-point method

Putting it all together, we now have our primal-dual interior-point method. Start with a strictly feasible point $x^{(0)}$ and $u^{(0)} > 0$, $v^{(0)}$. Define $\eta^{(0)} = -g(x^{(0)})^T u^{(0)}$ and let $\sigma \in (0, 1)$, then we repeat for $k = 1, 2, 3 \ldots$

- Define $\epsilon = \sigma \eta^{(k-1)}/m$
- Compute primal-dual update direction Δy
- Determine step size s
- Update $y^{(k)} = y^{(k-1)} + s \cdot \Delta y$
- Compute $\eta^{(k)} = -g(x^{(k)})^T u^{(k)}$
- Stop if $\eta^{(k)} \leq \delta$ and $(\|r_{\text{prim}}\|_2^2 + \|r_{\text{dual}}\|_2^2)^{1/2} \leq \delta$

Primal-dual interior-point method

Putting it all together, we now have our primal-dual interior-point method. Start with a strictly feasible point $x^{(0)}$ and $u^{(0)} > 0$, $v^{(0)}$. Define $\eta^{(0)} = -g(x^{(0)})^T u^{(0)}$ and let $\sigma \in (0, 1)$, then we repeat for $k = 1, 2, 3 \ldots$

- Define $\epsilon = \sigma \eta^{(k-1)}/m$
- Compute primal-dual update direction Δy
- Determine step size s
- Update $y^{(k)} = y^{(k-1)} + s \cdot \Delta y$
- Compute $\eta^{(k)} = -g(x^{(k)})^T u^{(k)}$
- Stop if $\eta^{(k)} \leq \delta$ and $(\|r_{\text{prim}}\|_2^2 + \|r_{\text{dual}}\|_2^2)^{1/2} \leq \delta$

Note the stopping criterion checks both the central residual via η, and (approximate) primal and dual feasibility

Backtracking line search

At each step, we need to find s and set

$$x^+ = x + s\Delta x, \quad u^+ = u + s\Delta u, \quad v^+ = v + s\Delta v.$$

Two main goals:

- Maintain $g(x) < 0,\ u > 0$
- Reduce $\|r(x, u, v)\|$

Use a multi-stage backtracking line search for this purpose: start with largest step size $s_{\max} \le 1$ that makes $u + s\Delta u \ge 0$:

$$s_{\max} = \min\left\{1,\ \min\{-u_i/\Delta u_i : \Delta u_i < 0\}\right\}$$

Then, with parameters $\alpha, \beta \in (0, 1)$, we set $s = 0.99 s_{\max}$, and

- Update $s = \beta s$, until $g_j(x^+) < 0$, $j = 1, \ldots r$
- Update $s = \beta s$, until $\|r(x^+, u^+, v^+)\| \le (1 - \alpha s)\|r(x, u, v)\|$

Special case: linear programming

Consider

$$\begin{aligned} \min_x \quad & c^\mathsf{T} x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

for $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$.

Some history:

- Dantzig (1940s): the simplex method, still today is one of the most well-known/well-studied algorithms for LPs
- Karmarkar (1984): interior-point polynomial-time method for LPs. Fairly efficient (US Patent 4,744,026, expired in 2006)
- Modern state-of-the-art LP solvers typically use both simplex and interior-point methods

KKT conditions for standard form LP

The points $x^\star$ and $(u^\star, v^\star)$ are respectively primal and dual optimal LP solutions if and only if they solve:

$$
\begin{aligned}
A^T v + u &= c \\
x_i u_i &= 0, \; i = 1, \ldots, n \\
Ax &= b \\
x, u &\geq 0
\end{aligned}
$$

(Neat fact: the simplex method maintains the first three conditions and aims for the fourth one ... interior-point methods maintain the first and last two, and aim for the second)

The perturbed KKT conditions for standard form LP are hence:

$$\begin{aligned} A^T v + u &= c \\ x_i u_i &= \epsilon, \ i = 1, \ldots, n \\ Ax &= b \\ x, u &\geq 0 \end{aligned}$$

Let's work through the barrier method, and the primal-dual interior point method, to get a sense of these two

Barrier method (after elim u):

$$\begin{aligned} 0 &= r_{\mathrm{br}}(x, v) \\ &= \begin{pmatrix} A^T v + \mathrm{diag}(x)^{-1}\epsilon - c \\ Ax - b \end{pmatrix} \end{aligned}$$

Primal-dual method:

$$\begin{aligned} 0 &= r_{\mathrm{pd}}(x, u, v) \\ &= \begin{pmatrix} A^T v + u - c \\ \mathrm{diag}(x)u - \epsilon \\ Ax - b \end{pmatrix} \end{aligned}$$

Barrier method: $0 = r_{\rm br}(y+\Delta y) \approx r_{\rm br}(y) + \nabla r_{\rm br}(y)\Delta y$, i.e., we solve

$$\begin{bmatrix} -\mathrm{diag}(x)^{-2}\epsilon & A^T \\ A & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta v \end{pmatrix} = -r_{\rm br}(x,v)$$

and take a step $y^+ = y + s\Delta y$ (with line search for $s > 0$), and iterate until convergence. Then update $\epsilon = \sigma\epsilon$

Primal-dual method: $0 = r_{\rm pd}(y+\Delta y) \approx r_{\rm pd}(y) + \nabla r_{\rm pd}(y)\Delta y$, i.e., we solve

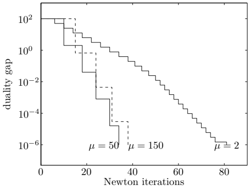
$$\begin{bmatrix} 0 & I & A^T \\ \mathrm{diag}(u) & \mathrm{diag}(x) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta u \\ \Delta v \end{pmatrix} = -r_{\rm pd}(x,u,v)$$

and take a step $y^+ = y + s\Delta y$ (with line search for $s > 0$), but only once. Then update $\epsilon = \sigma\epsilon$

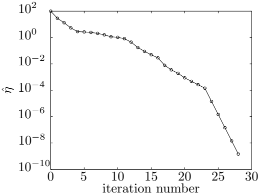
Example: barrier versus primal-dual

Example from B & V 11.3.2 and 11.7.4: standard LP with $n = 50$ variables and $m = 100$ equality constraints

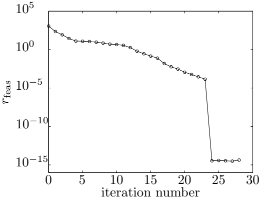
Barrier method uses various values of σ, primal-dual method uses $\sigma = 0.1$. Both use $\alpha = 0.01$, $\beta = 0.5$



Barrier central residual ($\mu = 1/\sigma$)



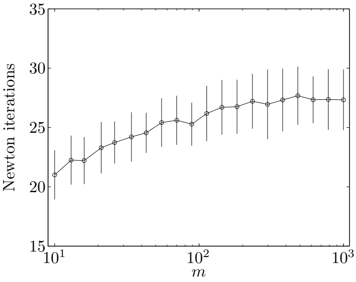
Primal-dual central residual



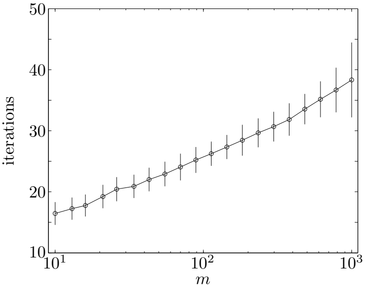
Primal-dual feasibility gap, $r_{\text{feas}} = (\|r_{\text{prim}}\|_2^2 + \|r_{\text{dual}}\|_2^2)^{1/2}$

Can see that primal-dual is faster to converge to high accuracy

Now a sequence of problems with $n = 2m$, and n growing. Barrier method uses $\mu = 100$, runs just two outer loops (decreases central residual by 10^4); primal-dual method uses $\sigma = 0.1$, stops when central residual and feasibility gap are at most 10^{-8}



Barrier method



Primal-dual method

Primal-dual method require only slightly more iterations, despite the fact that it is producing higher accuracy solutions