Primal-Dual Interior-Point Methods

Ryan Tibshirani
Convex Optimization 10-725/36-725

Last time: barrier method

Given the problem

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & h_i(x) \le 0, \;\; i = 1, \dots m \\ & Ax = b \end{aligned}$$

where f, h_i, $i = 1, \dots m$ are convex and smooth, we consider

$$\begin{aligned} \min_x \quad & tf(x) + \phi(x) \\ \text{subject to} \quad & Ax = b \end{aligned}$$

where ϕ is the log barrier function

$$\phi(x) = -\sum_{i=1}^{m} \log(-h_i(x))$$

In the barrier method, we solve a sequence of problems

$$\begin{array}{ll} \min_x & tf(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

for increasing values of $t > 0$, until $m/t \leq \epsilon$

In particular, start with $t = t^{(0)} > 0$, and solve the above problem using Newton's method to produce $x^{(0)} = x^\star(t)$. Then, we iterate for $k = 1, 2, 3, \ldots$

- Solve the barrier problem at $t = t^{(k)}$, using Newton's method initialized at $x^{(k-1)}$, to produce $x^{(k)} = x^\star(t)$
- Stop if $m/t \leq \epsilon$
- Else update $t^{(k+1)} = \mu t$, where $\mu > 1$

Outline

Today:

- Motivation from perturbed KKT conditions
- Primal-dual interior-point method
- Backtracking line search
- Highlight on standard form LPs

Barrier method versus primal-dual method

We will cover the primal-dual interior-point method, which solves basically the same problems as the barrier method. These two are pretty similar, but have some key differences

Overview:

- Both can be motivated in terms of perturbed KKT conditions
- Primal-dual interior-point methods take one Newton step, and move on (no separate inner and outer loops)
- Primal-dual interior-point iterates are not necessarily feasible
- Primal-dual interior-point methods can be more efficient, since they can exhibit better than linear convergence
- Primal-dual interior-point methods are less intuitive ...

Back to perturbed KKT conditions

The barrier method iterates at any particular barrier method t can be seen as solving a perturbed version of the KKT conditions

$$\nabla f(x) + \sum_{i=1}^{m} u_i \nabla h_i(x) + A^T v = 0$$
$$u_i h_i(x) = -1/t, \quad i = 1, \ldots m$$
$$h_i(x) \leq 0, \quad i = 1, \ldots m, \quad Ax = b$$
$$u_i \geq 0, \quad i = 1, \ldots m$$

Only difference between these and actual KKT conditions for our original problem is in the second condition: these are replaced by

$$u_i h_i(x) = 0, \quad i = 1, \ldots m$$

i.e., complementary slackness, in actual KKT conditions

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 $u = -1/(th_i(x))$, $i = 1, \dots m$)

Primal-dual interior-point updates are also motivated by a Newton step for solving these nonlinear equations, but without eliminating u. Write it concisely as $r(x, u, v) = 0$, where

$$r(x,u,v) = \begin{pmatrix} \nabla f(x) + Dh(x)^T u + A^T v \\ -\mathrm{diag}(u)h(x) - 1/t \\ Ax - b \end{pmatrix}$$

and

$$h(x) = \begin{pmatrix} h_1(x) \\ \dots \\ h_m(x) \end{pmatrix}, \quad Dh(x) = \begin{bmatrix} \nabla h_1(x)^T \\ \dots \\ \nabla h_m(x)^T \end{bmatrix}$$

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

$$\begin{aligned} r_{\text{dual}} &= \nabla f(x) + Dh(x)^T u + A^T v \\ r_{\text{cent}} &= -\text{diag}(u)h(x) - 1/t \\ r_{\text{prim}} &= Ax - b \end{aligned}$$

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

Now we make our first-order approximation

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

and we want to solve for Δy in the above

I.e., we solve

$$\begin{bmatrix} \nabla^2 f(x) + \sum_{i=1}^m u_i \nabla^2 h_i(x) & Dh(x)^T & A^T \\ -\mathrm{diag}(u) Dh(x) & -\mathrm{diag}(h(x)) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta u \\ \Delta v \end{pmatrix} = -\begin{pmatrix} r_{\mathrm{dual}} \\ r_{\mathrm{cent}} \\ r_{\mathrm{prim}} \end{pmatrix}$$

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)

Surrogate duality gap

Unlike barrier methods, the iterates from primal-dual interior-point method are not necessarily feasible

For barrier method, we have simple duality gap: m/t, since we set $u_i = -1/(th_i(x))$, $i = 1, \ldots m$ and saw this was dual feasible

For primal-dual interior-point method, we construct a surrogate duality gap:

$$\eta = -h(x)^T u = -\sum_{i=1}^{m} u_i h_i(x)$$

This would be a bonafide duality gap if we had feasible points (i.e., if $r_{\text{prim}} = 0$ and $r_{\text{dual}} = 0$), but we don't, so it's not

What value of parameter t does this correspond to in perturbed KKT conditions? This is $t = \eta/m$

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 $\eta^{(0)} = -h(x^{(0)})^T u^{(0)}$. Then for a barrier parameter $\mu > 1$, we repeat for $k = 1, 2, 3 \ldots$

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

Note the stopping criterion checks both the surrogate duality gap, and (approximate) primal and dual feasibility

Backtracking line search

At each step, we need to make sure that we settle on an update $y^+ = y + s\Delta y$, i.e.,

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

that maintains $h_i(x) < 0$, $u_i > 0$, $i = 1, \ldots m$

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

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

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

- Update $s = \beta s$, until $h_i(x^+) < 0$, $i = 1, \ldots m$
- Update $s = \beta s$, until $\|r(x^+, u^+, v^+)\|_2 \leq (1 - \alpha s)\|r(x, u, v)\|_2$

Highlight: standard form LP

Recall the standard form LP:

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

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

$$\begin{aligned} \max_{u,v} \quad & b^T v \\ \text{subject to} \quad & A^T v + u = c \\ & u \geq 0 \end{aligned}$$

(This is not a bad thing to memorize)

Some history

- Dantzig (1940s): the simplex method, still today is one of the most well-known/well-studied algorithms for LPs
- Klee and Minty (1972): pathological LP with n variables and $2n$ constraints, simplex method takes 2^n iterations to solve
- Khachiyan (1979): polynomial-time algorithm for LPs, based on ellipsoid method of Nemirovski and Yudin (1976). Strong in theory, weak in practice
- Karmarkar (1984): interior-point polynomial-time method for LPs. Fairly efficient (US Patent 4,744,026, expired in 2006)
- Renegar (1988): Newton-based interior-point algorithm for LP. Best known theoretical complexity to date
- 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:

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

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):

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

Primal-dual method:

$$
\begin{aligned}
0 &= r_{\text{pd}}(x, u, v) \\
&= \begin{pmatrix} A^T v + u - c \\ \text{diag}(x) u - 1/t \\ Ax - b \end{pmatrix}
\end{aligned}
$$

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

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

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

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

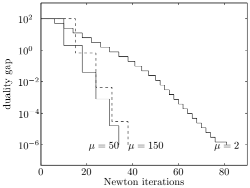
$$\begin{bmatrix} 0 & I & A^T \\ \mathrm{diag}(u) & \mathrm{diag}(x) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta u \\ \Delta v \end{pmatrix} = -r_{\rm pd}(x, u, v)$$

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

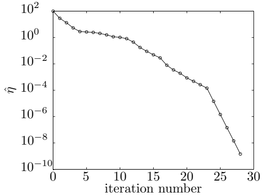
Example: barrier versus primal-dual

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

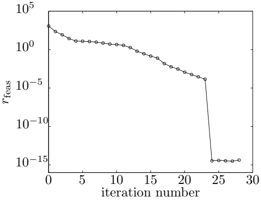
Barrier method uses various values of μ, primal-dual method uses $\mu = 10$. Both use $\alpha = 0.01$, $\beta = 0.5$



Barrier duality gap



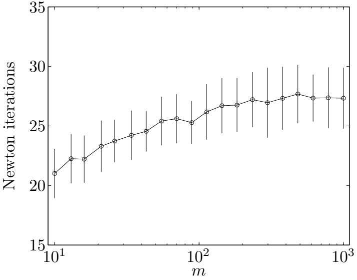
Primal-dual surrogate duality gap



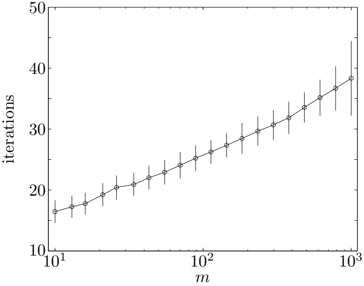
Primal-dual feasibility gap, $r_{\mathrm{feas}} = (\|r_{\mathrm{prim}}\|_2^2 + \|r_{\mathrm{dual}}\|_2^2)^{1/2}$

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 duality gap by 10^4); primal-dual method uses $\mu = 10$, stops when duality gap 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 they it is producing higher accuracy solutions

References and further reading

- S. Boyd and L. Vandenberghe (2004), "Convex optimization," Chapter 11
- J. Nocedal and S. Wright (2006), "Numerical optimization", Chapters 14 and 19
- S. Wright (1997), "Primal-dual interior-point methods," Chapters 5 and 6
- J. Renegar (2001), "A mathematical view of interior-point methods"