Primal-Dual Interior-Point Methods

Ryan Tibshirani
Convex Optimization 10-725

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 twice differentiable, and strong duality holds. 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

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

Let $x^\star(t)$ be a solution to the barrier problem for particular $t > 0$, and $f^\star$ be optimal value in original problem. We can show m/t is a duality gap, so that

$$f(x^\star(t)) - f^\star \leq m/t$$

Motivates the barrier method, where we solve the barrier problem for increasing values of $t > 0$, until duality gap satisfies $m/t \leq \epsilon$

We fix $t^{(0)} > 0$, $\mu > 1$. We use Newton to compute $x^{(0)} = x^\star(t)$, a solution to barrier problem at $t = t^{(0)}$. For $k = 1, 2, 3, \ldots$

- Solve the barrier problem at $t = t^{(k)}$, using Newton initialized at $x^{(k-1)}$, to yield $x^{(k)} = x^\star(t)$
- Stop if $m/t \leq \epsilon$, else update $t^{(k+1)} = \mu t$

Outline

Today:

- Perturbed KKT conditions, revisited
- Primal-dual interior-point method
- Backtracking line search
- Highlight on standard form LPs

Barrier versus primal-dual method

Today we will discuss the primal-dual interior-point method, which solves basically the same problems as the barrier method. What's the difference between these two?

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 are often more efficient, as they can exhibit better than linear convergence
- Primal-dual interior-point methods are less intuitive ...

Perturbed KKT conditions

Barrier method iterates $(x^\star(t), u^\star(t), v^\star(t))$ can be motivated as solving the perturbed KKT conditions:

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

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

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

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

Perturbed KKT as nonlinear system

Can view this as a nonlinear system of equations, written as

$$r(x,u,v) = \begin{pmatrix} \nabla f(x) + Dh(x)^T u + A^T v \\ -\mathrm{diag}(u)h(x) - (1/t)1 \\ Ax - b \end{pmatrix} = 0$$

where

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

Newton's method, recall, is generally a root-finder for a nonlinear system $F(y) = 0$. Approximating $F(y + \Delta y) \approx F(y) + DF(y)\Delta y$ leads to

$$\Delta y = -(DF(y))^{-1} F(y)$$

What happens if we apply this to $r(x,u,v) = 0$ above?

Newton on perturbed KKT, v1

Approach 1: from middle equation (relaxed comp slackness), note that $u_i = -1/(th_i(x))$, $i = 1, \ldots m$. So after eliminating u, we get

$$r(x, v) = \begin{pmatrix} \nabla f(x) + \sum_{i=1}^m \left(-\frac{1}{th_i(x)}\right) \nabla h_i(x) + A^T v \\ Ax - b \end{pmatrix} = 0$$

Thus the Newton root-finding update $(\Delta x, \Delta v)$ is determined by

$$\begin{bmatrix} H_{\text{bar}}(x) & A^T \\ A & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta v \end{pmatrix} = -r(x, v)$$

where $H_{\text{bar}}(x) =$
$\nabla^2 f(x) + \sum_{i=1}^m \frac{1}{th_i(x)^2} \nabla h_i(x) \nabla h_i(x)^T + \sum_{i=1}^m \left(-\frac{1}{th_i(x)}\right) \nabla^2 h_i(x)$

This is just the KKT system solved by one iteration of Newton's method for minimizing the barrier problem

Newton on perturbed KKT, v2

Approach 2: directly apply Newton root-finding update, without eliminating u. Introduce notation

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

called the dual, central, and primal residuals at $y = (x, u, v)$. Now root-finding update $\Delta y = (\Delta x, \Delta u, \Delta v)$ is given by

$$\begin{bmatrix} H_{\text{pd}}(x) & Dh(x)^T & A^T \\ -\text{diag}(u)Dh(x) & -\text{diag}(h(x)) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta u \\ \Delta v \end{pmatrix} = - \begin{pmatrix} r_{\text{dual}} \\ r_{\text{cent}} \\ r_{\text{prim}} \end{pmatrix}$$

where $H_{\text{pd}}(x) = \nabla^2 f(x) + \sum_{i=1}^m u_i \nabla^2 h_i(x)$

Some notes:

- In v2, update directions for the primal and dual variables are inexorably linked together
- Also, v2 and v1 leads to different (nonequivalent) updates
- As we saw, one iteration of v1 is equivalent to inner iteration in the barrier method
- And v2 defines a new method called primal-dual interior-point method, that we will flesh out shortly
- One complication: in v2, the dual iterates are not necessarily feasible for the original dual problem ...

Surrogate duality gap

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 can construct 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., $r_{\text{prim}} = 0$ 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 $t = m/\eta$

Primal-dual interior-point method

Putting it all together, we now have our primal-dual interior-point method. Start with $x^{(0)}$ such that $h_i(x^{(0)}) < 0$, $i = 1, \ldots, m$, and $u^{(0)} > 0$, $v^{(0)}$. Define $\eta^{(0)} = -h(x^{(0)})^T u^{(0)}$. We fix $\mu > 1$, repeat for $k = 1, 2, 3 \ldots$

- Define $t = \mu m / \eta^{(k-1)}$
- Compute primal-dual update direction Δy
- Use backtracking to 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 that we stop based on surrogate duality gap, and approximate feasibility. (Line search maintains $h_i(x) < 0$, $u_i > 0$, $i = 1, \ldots, m$)

Backtracking line search

At each step, must ensure we arrive at $y^+ = y + s\Delta y$, i.e.,

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

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

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

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

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

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

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 complexity ... until Lee and Sidford (2014)
- Modern state-of-the-art LP solvers typically use both simplex and interior-point methods

Highlight: standard 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)

KKT conditions

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:

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

What do our interior point methods do?

Barrier method (after elim u):

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

Primal-dual method:

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

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

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

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: set $0 = r_{\text{pd}}(y + \Delta y) \approx r_{\text{pd}}(y) + Dr_{\text{pd}}(y)\Delta y$, i.e., solve

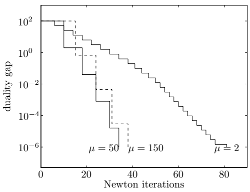
$$\begin{bmatrix} 0 & I & A^T \\ \operatorname{diag}(u) & \operatorname{diag}(x) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta u \\ \Delta v \end{pmatrix} = -r_{\text{pd}}(x, u, v)$$

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

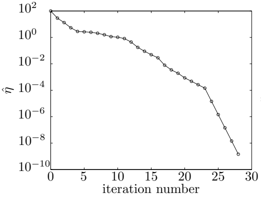
Example: barrier versus primal-dual

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

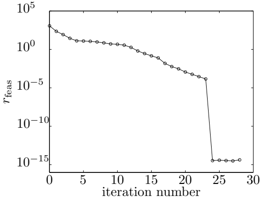
Barrier method uses various values of μ, primal-dual method uses $\mu = 10$. Both use $\alpha = 0.01$, $\beta = 0.5$



Barrier duality gap



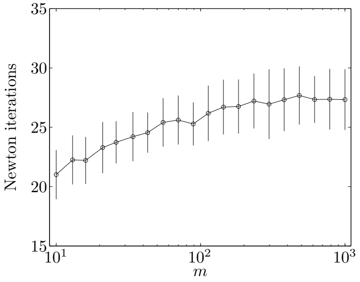
Primal-dual surrogate duality gap



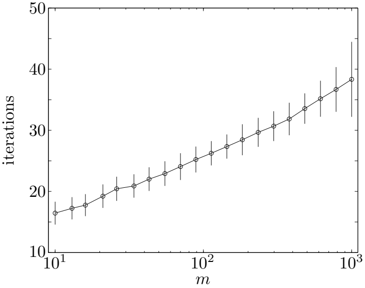
Primal-dual feasibility gap, where $r_{\text{feas}} = (\|r_{\text{prim}}\|_2^2 + \|r_{\text{dual}}\|_2^2)^{1/2}$

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 two outer loops (decreases duality gap by 10^4); primal-dual method uses $\mu = 10$, stops when surrogate duality gap and feasibility gap are at most 10^{-8}



Barrier method



Primal-dual method

Primal-dual method requires only slightly more iterations, despite the fact that it is producing much 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"