Lecture 17: October 17

Lecturer: Lecturer: Ryan Tibshirani *Scribes: Scribes: Manjari Das, Leqi Liu, Zhuoran Liu*

17.1 Barrier method: A quick recap

We first provide a quick recap to barrier method introduced in previous lectures, as we will compare it against primal dual method in this note. The barrier method solves an optimization problem

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

where $f, h_i, i = 1, \cdots, m$, 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 a particular $t > 0$, and $f^\star$ be the 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 $m/t \leq \epsilon$.

For fixed $t^{(0)} > 0$, $\mu > 1$, we use Newton's method to compute $x^{(0)} = x^\star(t)$, the solution to barrier problem at $t = t^{(0)}$. Then, for $k = 1, 2, 3, \cdots$,

- 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 $t^{(k+1)} = \mu t$.

17.2 Perturbed KKT conditions

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

$$\begin{aligned}
&\nabla f(x) + \sum_{i=1}^m u_i \nabla h_i(x) + A^T v = 0 \\
&u_i \cdot h_i(x) = -(1/t)1, i = 1, \cdots, m \\
&h_i(x) \leq 0, i = 1, \cdots, m Ax = b \\
&u_i \geq 0, i = 1, \cdots, m
\end{aligned}$$

Observe that the only difference between the perturbed and actual KKT conditions for the original problem is the second line, which is replaced by

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

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

We can combine the perturbed KKT conditions into a nonlinear system of equations, i.e.,

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

where

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

In previous lectures, we mentioned that Newton's method can be viewed as a root-finder for a nonlinear system $F(y) = 0$. In this context, approximating $F(y + \Delta y) \approx F(y) + DF(y)\Delta y$ leades to

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

We apply this result to $r(x, u, v)$ and leads to two versions of perturbed KKT.

17.2.1 Newton on perturbed KKT, v1

This approach starts from the relaxed complementary slackness. We first rearrange the equation and obtain

$$u_i \cdot h_i(x) = -(1/t)1 \implies u_i = -1/(th_i(x)), i = 1, \cdots, m.$$

Plug in this expression and we reduce this nonlinear system to

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

The gradient of the first line of $r(x, v)$ w.r.t x and v is given by

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

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

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

which is precisely the KKT system solved by one iteration of Newton's method for minimizing the barrier problem.

17.2.2 Newton on perturbed KKT, v2

The second version directly apply Newton's root-finding update, without eliminating u as we did in v1. Following notation in the slides, we define

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

denoting the dual, central, and primal residuals at $y = (x, u, v)$. Taking gradient of each entry in $r(x, u, v)$ w.r.t x, u, v respectively, we attain the root finding update $\Delta y = (\Delta x, \Delta u, \Delta v)$

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

where

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

The v2 is precisely the primal-dual interior point method. Some salient points we draw from v1 and v2:

- In primal-dual interior-point method (v2), the update directions for the primal and dual variables are related by the root-finding update matrix. This is not the case in v1.
- v1 and v2 lead to different (nonequivalent) updates.
- One iteration of v1 is equivalent to an inner iteration in the barrier method.
- In v2, the dual iterates are not necessarily feasible for the original dual problem, while v1 is always feasible.
- In practice, v2 is often more efficient than v1, as they can exhibit better-than-linear convergence.

17.3 Primal-dual interior-point method

17.3.1 Surrogate duality gap

In v2, we see that update directions for the primal and dual variables are inexorably linked together and the dual iterates are not necessarily feasible for the original dual problem. To check the optimality of our solution, we need to construct a surrogate duality gap:

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

Remarks:

1. According to the "complementary slackness" condition in the perturbed KKT conditions, we have that $\eta = -h(x)^T u = -\sum_{i=1}^{m} u_i h_i(x) = \frac{m}{t}$. Hence, $t = \frac{m}{\eta}$.
2. If we have feasible points, i.e. $r_{\text{prim}} = 0$ and $r_{\text{dual}} = 0$, then the duality gap is just $-\sum_{i=1}^{m} u_i h_i(x)$.
3. Barrier method doesn't have this problem: since we set $u_i = \frac{1}{t \cdot h_i(x)}$, $i = 1, ..., m$, u_i is dual feasible and the duality gap will be $g(u^*(t), v^*(t)) - f(x^*(t)) = -\frac{m}{t}$ as shown in the last lecture.

17.3.2 Primal-dual interior-point method

- Start with strictly feasible primal point $x^{(0)}$, i.e. $h_i(x^{(0)}) < 0$, $i = 1, ..., m$, and $u^{(0)} > 0$, $v^{(0)}$.
- Define the surrogate duality gap $\eta^{(0)} = -h(x^{(0)})^T u^{(0)}$. Fix $\eta > 1$.
- For $k = 1, 2, 3, ...$
  - Define $t = \frac{\mu m}{\eta^{(k-1)}}$
  - Compute primal-dual update direction $\Delta y = (\Delta x, \Delta u, \Delta v)$
  - Use backtracking to determine step size s
  - Update $y^{(k)} = y^{(k-1)} + s \cdot \Delta y$ [Newton Update]
  - Compute $\eta^{(k)} = -h(x^{(k)})^T u^{(k)}$
  - Stop if $\eta^{(k)} \leq \epsilon$ and $||r_{\text{prim}}||_2^2 + ||r_{\text{dual}}||_2^2 \leq \epsilon$ [Stopping Criterion: surrogate duality gap is small and primal dual residuals are small (an approximation to feasibility).]

17.3.3 Backtracking line search

How do we set the parameter s to ensure that after each update ($x^+ = x+s\Delta x$, $u^+ = u+s\Delta u$, $v^+ = v+s\Delta v$), $h_i(x) < 0$ and $u_i > 0$, $i = 1, ..., m$? We use multi-stage backtracking line search:

- Start with largest step size $s_{\max} \leq 1$ that ensures $u + s\Delta u > 0$: $s_{\max} = \min\{1, \min\{-u_i/\Delta u_i : \Delta u_i < 0\}\}$ [Maintaining $u_i > 0$].
- Then, with parameters $\alpha, \beta \in (0, 1)$, we set $s = .99 s_{\max}$ and update:
  - $s = \beta s$, until $h_i(x^+) < 0$, $i = 1, ..., m$ [Maintaining $h_i(x) < 0$.]
  - $s = \beta s$, until $||r(x^+, u^+, v^+)||_2 \leq (1 - \alpha s)||r(x, u, v)||_2$ [Reducing $||r(x, u, v)||_2$.]

17.3.4 Example: Standard LP

Next, we recall the standard form of linear program

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

for $c \in^n$, $A \in^{m \times n}$, $b \in^m$. The dual of this linear program is

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

The KKT conditions are

$$\begin{aligned} ATv + u &= c && \text{stationarity} \\ x_i u_i &= 0, \ i = 1, \ldots, n && \text{complementary slackness} \\ Ax &= b && \text{primal feasibility} \\ x, u &\geq 0 && \text{primal and dual feasibility} \end{aligned}$$

Points x^* and (u^*, v^*) are optimal solutions for the primal and the dual problem respectively if and only if they solve the KKT conditions above. The interior point method modifies the second condition whereas, the simplex method modifies the fourth condition.

Hence, for the interior piint metnod the KKT conditions are

$$\begin{aligned}
A^T v + u &= c && \text{stationarity} \\
x_i u_i &= 1/t,\ i = 1, \dots, n && \text{complementary slackness} \\
Ax &= b && \text{primal feasibility} \\
x, u &\geq 0 && \text{primal and dual feasibility}
\end{aligned}$$

Barrier method after eliminating u gives, says

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

Then set

$$\begin{aligned}
0 &= r_{br}(y + \Delta y) \\
or,\ 0 &= r_{br}(y) + Dr_{br}(y)\Delta y \\
&\text{Then solve} \\
&\begin{bmatrix} -diag(x)^{-2}/t & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta v \end{bmatrix} = -r_{br}(x, v) \\
&\text{take step } y^+ = y + s\Delta y \text{ with line search } s > 0
\end{aligned}$$

This is repeated until convergence. Then the updated t is μt.

The primal-dual method says

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

Then set

$$\begin{aligned}
0 &= r_{pd}(y + \Delta y) \\
or,\ 0 &= r_{pd}(y) + Dr_{pd}(y)\Delta y \\
&\text{Then solve} \\
&\begin{bmatrix} 0 & I & A^T \\ diag(u) & diag(x) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta u \\ \Delta v \end{bmatrix} = -r_{pd}(x, u, v) \\
&\text{take step } y^+ = y + s\Delta y \text{ with line search } s > 0
\end{aligned}$$

Unlike Barrier method this step is performed only once. The updated t is μt.

17.3.5 Example: barrier versus primal-dual

To compare the performance we look at an example from the book *Convex Optimization* by S. Boyd and L. Vandenberghe. The problem is

$$\min_x \; c^T x$$
$$\text{subject to } \; Ax = 0, \; x \succeq 0$$

$A \in^{100\times 50}$. The the problem is primal and dual feasible and the optimal value is $p^* = 1$. Thus there are 50 variables and 100 equality constraints. The elements of A are i.i.d standard normal. The initial point $x^{(0)}$ has elements which are i.i.d $U[0,1]$ to ensure that $x^{(0)}$ is primal feasible. The cost vector c is constructed by first constructing $z \in^m$ from standard normal and $s \in^n$ from $U[0,1]$ and then setting $c = A^T z + s$. This guarantees dual feasibility. The backtracking parameters used are $\alpha = 0.01$, $\beta = 0.5$. The initial value is t is $t^{(0)} = 1$. The plot for $\mu = 2$, $50 \,\&\, 150$ is shown in Figure 17.3.5 (11.4 from the book). The barrier method with standard LP was repeated by growing n and maintaining $n = 2m$. The convergence is roughly linear for all values of μ.

Next the primal dual interior point method is applied with $\mu = 10$ and $= 0.01$, $\beta = 0.5$. The surrogate gap $\hat{\eta}(x,\lambda) = -f(x)^T\lambda$ (λ is a dual variable) and the norm of the primal dual residuals ($r_{feas} = (\|r_{pri}\|_2^2 + \|r_{dual}\|_2^2)^{1/2}$) are plot against the iteration number in Figure 17.3.5 (11.21 from the book). The convergence is faster compared to the barrier method.

The number of iterations against the number of variables in Figure 17.3.5 (11.8 & 11.23 from the book) for the barrier method with $\mu = 100$ as well as the primal-dual method with $\mu = 10$. The primal-dual method requires only slightly more iterations than the Barrier method however it gives better accuracy.

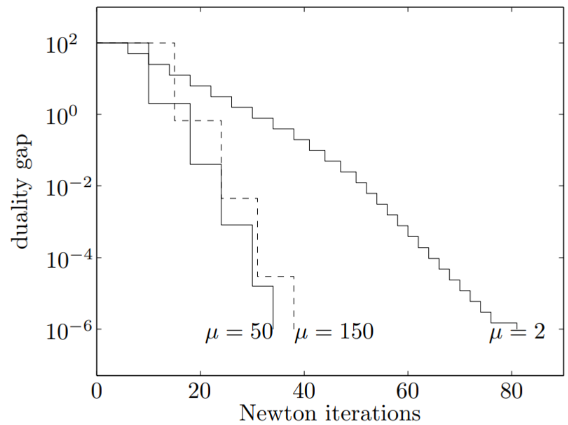

Figure 11.4 Progress of barrier method for a small LP, showing duality gap versus cumulative number of Newton steps. Three plots are shown corresponding to three values of the parameter μ: 2, 50, and 150. In each case, we have approximately linear convergence of duality gap.

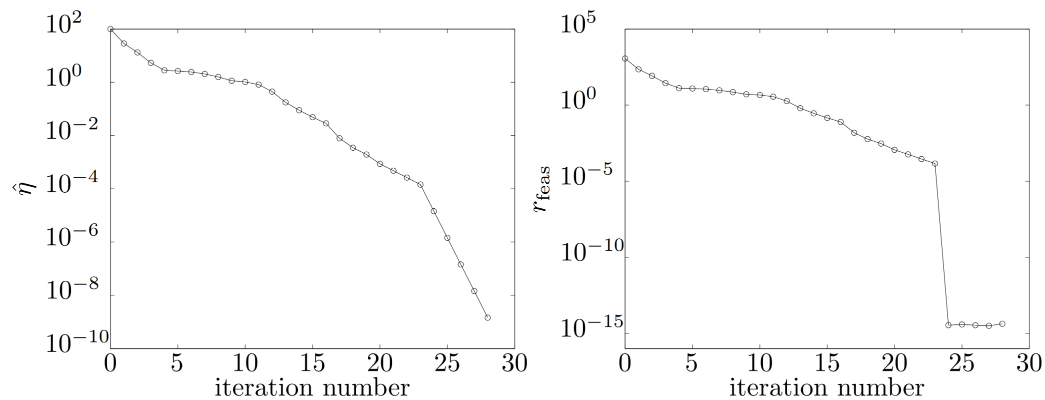

Figure 11.21 Progress of the primal-dual interior-point method for an LP, showing surrogate duality gap $\hat{\eta}$ and the norm of the primal and dual residuals, versus iteration number. The residual converges rapidly to zero within 24 iterations; the surrogate gap also converges to a very small number in about 28 iterations. The primal-dual interior-point method converges faster than the barrier method, especially if high accuracy is required.

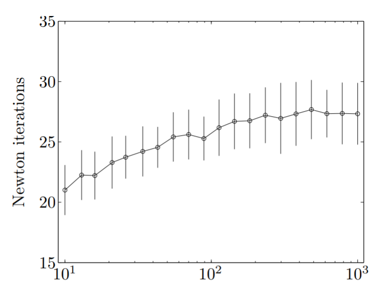

Figure 11.8 Average number of Newton steps required to solve 100 randomly generated LPs of different dimensions, with $n = 2m$. Error bars show standard deviation, around the average value, for each value of m. The growth in the number of Newton steps required, as the problem dimensions range over a 100 : 1 ratio, is very small.

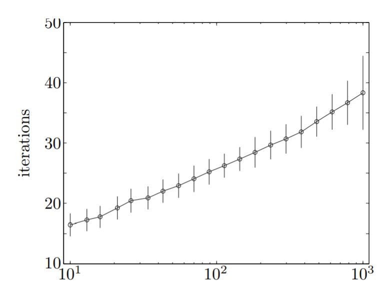

Figure 11.23 Number of iterations required to solve randomly generated standard LPs of different dimensions, with $n = 2m$. Error bars show standard deviation, around the average value, for 100 instances of each dimension. The growth in the number of iterations required, as the problem dimensions range over a 100 : 1 ratio, is approximately logarithmic.

References

[1] S. Boyd and L. Vandenberghe, "Convex optimization," 2004, Chapter 11.

[2] J. Nocedal and S. Wright, "Numerical optimization," 2006, Chapters 14 & 19.

[3] S. Wright, "Primal-dual interior-point methods," 1997, Chapters 5 and 6.

[4] J. Renegar, "A mathematical view of interior-point methods," 2001.