10-725/36-725: Convex Optimization **Fall 2019**

Lecture 16: Oct 21

Lecturer: Lecturer: Ryan Tibshirani *Scribes: Wentai Zhang, Haiguang Liao, Elan Rosenfeld*

16.1 Purturbed KKT conditions, revisited

16.1.1 Barrier versus primal-dual method

Difference between the two methods:

- Both can be motivated in terms of purturbed KKT conditions
- Primal-dual interior-point methods take one Newton step, and move on
- 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

16.1.2 Purturbed KKT conditions

Recall we can motivate barrier method iterates $(x^*(t), u^*(t), v^*(t))$ in terms of the perturbed KKT conditions:

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

Only difference between these and actual KKT conditions for our original problem is second line: these are replaced by: $u_i h_i(x) = 0$ i = 1,..., m, i.e., complementary slackness in actual KKT conditions.

16.1.3 Purturbed KKT as nonlinear system

We can view this as a nonlinear system of equations, written as:

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

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

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

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

We woud apply this to $r(x, u, v) = 0$.

16.1.4 Newton on perturbed KKT, v1

From middle equation (relaxed complimentary slackness), note that $u_i = -1/(th_i(x))$, $i = 1, ..., m$. After eliminating u, we get:

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

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

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

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

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

16.1.5 Newton on perturbed KKT, v2

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

$$r_{dual} = \nabla f(x) + Dh(x)^T u + A^T v$$

$$r_{cent} = -diag(u)h(x) - (1/t)t$$

$$r_{prim} = Ax - b$$

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:

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

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

Some notes:

- In v1, u is eliminated
- In v2, update directions for the primal and dual variables are inexorably linked together
- Also, v2 and v1 leads to different updates (nonequivalent)
- 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
- One complication: in v2, the dual iterates does not necessarily feasible for the original dual problem

16.2 Surrogate duality gap

Recap: For barrier mthod, the duality gap is m/t.

Theorem 16.1 *For primal-dual interior-point method, the surrogate duality gap is:*

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

It is not a bonafide duality gap, since we cannot garantee $r_{prim} = 0$ and $r_{dual} = 0$. In perturbed KKT conditions, this value actually corresponds to $t = m/\eta$.

16.3 Primal-dual interior-point method

Start with $x^{(0)}$ such that $h_i(x^{(0)}) < 0, i = 1, ..., m$ and $u^{(0)} > 0$, $v^{(0)}$. (This makes both primal and dual feasible.) Define $\eta^{(0)} = -h(x^{(0)})^T u^{(0)}$. We fix $\mu > 1$, repeat for $k = 1, 2, 3...$

- Define $t = \mu m/\eta^{(k-1)}$ (make t bigger iteratively)
- Compute primal-dual update direction Δy (consists of $\Delta x, \Delta u and \Delta v$)
- Use backtracking to determine step size s
- Update $y^{(k)} = y^{(k-1)} + s\Delta y$
- Compute $\eta^{(k)} = -h(x^{(k)})^T u^{(k)}$
- Stop if $\eta^{(k)} \leq \epsilon$ and $(||r_{prim}||_2^2 + ||r_{dual}||_2^2) \leq \epsilon$

In this process, we update the value based on line search, which maintains the feasibility: $h_i(x) < 0, u_i > 0, i = 1, ..., m$. The stopping criterion is based on surrogate duality gap.

16.4 Backtracking line search

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

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

that maintains both primal and dual feasibility: $h_i(x) < 0, u_i > 0, i = 1, ..., m$.

A multi-stage backtracking line search for this purpose:

start with largest step size $S_{max} \leq 1$ that makes $u + s\Delta u \geq 0$:

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

Then, with parameters $\alpha, \beta \in (0, 1)$, we set $s = 0.999 s_{max}$, and

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

16.5 Highlight: Standard LP

Recall the standard form LP:

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

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

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

16.5.1 KKT conditions

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

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

The perturbed KKT conditions for the standard form LP are therefore:

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

What do the two interior point methods (barrier method and primal-dual method) do to solve this system?

Barrier method (after eliminating u via substitution):

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

where $\mathbf{1}$ is the all-ones vector. Set $0 = r_{br}(y + \Delta y) \approx r_{br}(y) + Dr_{br}(y)\Delta y$, i.e., solve

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

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:

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

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

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

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

16.5.2 Full Newton

Once backtracking allows for $s = 1$, i.e., we take one full Newton step, primal dual method iterates will be primal and dual feasible from that point onwards. To see this, note that $\Delta x, \Delta u, \Delta v$ are constructed so that

$$\begin{aligned} A^T \Delta v + \Delta u &= -r_{dual} = -(A^T v + u - c) \\ A\Delta x &= -r_{prim} = -(Ax - vb). \end{aligned}$$

Therefore, after one full Newton step, $x^+ = x + \Delta x$, $u^+ = u + \Delta u$, $v^+ = v + \Delta v$, we have

$$\begin{aligned} r^+_{dual} &= A^T v^+ + u^+ - c = 0 \\ r^+_{prim} &= Ax^+ - b = 0, \end{aligned}$$

so our iterates are primal and dual feasible.

See Sections 11.3.2 and 11.7.4 in Boyd & Vandenberghe for an example of solving a standard LP with these two methods. Primal-dual method is faster to converge to high accuracy, yet requires only slightly more iterations.