10-725/36-725: Convex Optimization **Fall 2016**

Lecture 17: October 31

Lecturer: Ryan Tibshirani *Scribes: Yangyi Lu, Tewei Luo, Qinle Ba*

This lecture's notes illustrate some uses of various LaTeX macros. Take a look at this and imitate.

17.1 Last time: Duality Revisited

17.1.1 A brief introduction to the problem

Optimization problem:

$$\begin{aligned} \min_x \quad & f_0(x) \\ \text{subject to} \quad & Ax = b \\ & h(x) \leq 0 \end{aligned}$$

The Lagrangian is as follows:

$$L(x,u,v) = f(x) + u^T h(x) + v^T(Ax - b)$$

By combining the primal and dual variables, we can rewrite the primal problems as:

$$\min_x \max_{u \geq 0, v} L(x,u,v)$$

Dual problems as:

$$\max_{u \geq 0, v} \min_x L(x,u,v)$$

17.1.2 Optimality conditions

Assume $f, h_1, \ldots, h_m$ are convex and differentiable and strong duality holds. The KKT optimality conditions for primal and dual:

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

Where $U = Diag(u)$, $\nabla h(x) = [\nabla h_1(x), \ldots, \nabla h_m(x)]$

17.1.3 Central path equations

Barrier problem

$$\min_x \quad f(x) + \tau\phi(x)$$
$$\text{subject to} \quad Ax = b$$

where

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

The Optimality conditions for barrier problem and its dual is:

$$\begin{aligned}\nabla f(x) + \nabla h(x)u + A^T v &= 0\\ Uh(x) &= -\tau 1\\ Ax &= b\\ u, -h(x) &> 0\end{aligned}$$

There is a useful fact that the solution $(x(\tau), u(\tau), v(\tau))$ has duality gap:

$$f(x(\tau)) - \min_x L(x, u(\tau), v(\tau)) = m\tau$$

17.2 Primal-dual interior-point method

17.2.1 Barrier method v.s. primal-dual method

Commons:

- Both aim to compute (approximately) points on the central path.

Differences:

- 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.

17.2.2 Central path equations and Newton step

Firstly, Primal-dual interior-point method is based on central path equations:

$$\begin{aligned}\nabla f(x) + \nabla h(x)u + A^T v &= 0\\ Uh(x) + \tau 1 &= 0\\ Ax - b &= 0\\ u, -h(x) &> 0.\end{aligned}$$

Then we calculate the Newton step:

$$\begin{bmatrix} \nabla^2 f(x) + \sum_i u_i \nabla^2 h_i(x) & \nabla h(x) & A^T \\ U \nabla h(x)^T & H(x) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \triangle x \\ \triangle u \\ \triangle v \end{bmatrix} = -r(x,u,v)$$

where

$$r(x,u,v) := \begin{bmatrix} \nabla f(x) + \nabla h(x) u + A^T v \\ U h(x) + \tau 1 \\ Ax - b \end{bmatrix}, H(x) = Diag(h(x))$$

17.2.3 Surrogate duality gap, residuals

Define the dual, central, and primal residuals at current (x,u,v) as

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

And given x,u with $h(x) \leq 0$, $u \geq 0$, the surrogate duality gap is

$$-h(x)^T u$$

From the definition of $r_{dual}, r_{cent}, r_{prim}$, we can see that it is a true duality gap when $r_{dual} = r_{prim} = 0$. Observe that (x,u,v) is on the central path if and only if $u > 0, h(x) < 0$ and

$$r(x,u,v) = 0 \quad for \quad \tau = -\frac{h(x)^T u}{m}.$$

17.2.4 Primal-Dual Algorithm

1. Choose $\sigma \in (0,1)$
2. Choose (x^0, u^0, v^0) such that $h(x^0) < 0$, $u^0 > 0$
3. For $k = 0, 1, \ldots$

- Compute Newton step for

$$(x,u,v) = (x^k, u^k, v^k), \tau := \sigma \tau(x^k, u^k)$$

- Choose steplength θ_k and set

$$(x^{k+1}, u^{k+1}, v^{k+1}) := (x^k, u^k, v^k) + \theta_k(\triangle x, \triangle u, \triangle v)$$

Notations here is parallel to the barrier method

$$\tau = \frac{1}{t}, \sigma = \frac{1}{\mu}.$$

17.2.5 Backtracking line search

At each step, we need to find θ and set

$$x^+ = x + \theta \triangle x$$
$$u^+ = u + \theta \triangle u$$
$$v^+ = v + \theta \triangle v.$$

There are two main goals in the primal-dual algorithm:

- Maintain $h(x) < 0$, $u > 0$(strictly feasible points)
- Reduce $||r(x, u, v)||$

In order to satisfy the above two conditions, we use a multi-stage backtracking line search for this purpose: start with largest step size $\theta_{max} \leq 1$ that makes $u + \theta \triangle u \geq 0$:

$$\theta_{max} = min\{1, min\{-u_i/\triangle u_i : \triangle u_i < 0\}\}$$

Then, with parameters $\alpha, \beta \in (0, 1)$, we set $\theta = 0.99\theta_{max}$, the reason we do it is to maintain $u + \theta \triangle u > 0$. And

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

Note: Consider $f(w) = \frac{1}{2}||r(w)||^2$, then the update is also as follows:

$$\triangle w = -r'(w)^{-1} r(w)$$

So $\triangle w$ is a descent direction for f(w), therefore a descent direction for $||r(w)||$, then the second condition will eventually be satisfied.

17.3 Interior-point methods for semidefinite programming

17.3.1 Semidefinite problems

Primal problem:

$$\begin{aligned} \min_X \quad & C \bullet X \\ \text{subject to} \quad & A_i \bullet X = b_i, \ i = 1, ..., m \\ & X \succeq 0. \end{aligned}$$

The corresponding dual problem:

$$\begin{aligned} \max_y \quad & b^T y \\ \text{subject to} \quad & \sum_{y_i=1}^{m} y_i A_i + S = C \\ & S \succeq 0 \end{aligned}$$

where $C \bullet X = \text{trace}(CX)$ in $\mathbb{S}^n$

Recall that if both problems are strictly feasible, strong duality holds and primal and dual solutions are attained.

17.3.2 Optimality conditions for semidefinite programming

Rewrite the semidefinite primal and dual programs:

Let $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ be a linear map,

Primal and dual problem:

$$\begin{array}{ll} \min\limits_{X} & C \bullet X \\ \text{subject to} & \mathcal{A}(X) = b \\ & X \succeq 0 \end{array}$$

$$\begin{array}{ll} \max\limits_{y,S} & b^T y \\ \text{subject to} & \mathcal{A}^*(y) + S = C \\ & S \succeq 0 \end{array}$$

Assume that strong duality holds. X^* and (y^*, S^*) are primal and dual solutions, respectively $\iff$ X^*, y^* and S^* satisfy KKT conditions $\iff$ $(X^* y^* S^*)$ solves the following systems of equations:

$$\begin{aligned} \mathcal{A}^*(y) + S &= C \\ \mathcal{A}(X) &= b \\ XS &= 0 \\ X, S &\succeq 0 \end{aligned}$$

17.3.3 Central path for semidefinite programming

Primal barrier problem

$$\begin{array}{ll} \min\limits_{X} & C \bullet X - \tau log(det(X)) \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

Dual barrier problem

$$\begin{array}{ll} \max\limits_{y,S} & b^T y + \tau log(det(S)) \\ \text{subject to} & \mathcal{A}^*(y) + S = C \end{array}$$

Optimality conditions for primal and dual are as follows:

$$\begin{aligned} \mathcal{A}^*(y) + S &= C \\ \mathcal{A}(X) &= b \\ XS &= \tau I \\ X, S &\succ 0 \end{aligned}$$

17.3.4 Newton step for semidefinite programming

After eliminating dual variable S in the optimality conditions, we have primal central path equations

$$\begin{aligned} \mathcal{A}^*(y) + S &= C \\ \mathcal{A}(X) &= b \\ X &\succ 0 \end{aligned}$$

Newton Steps

$$\begin{aligned} \tau X^{-1} \Delta X X^{-1} + \mathcal{A}^*(\Delta y) &= -(\mathcal{A} + \tau X^{-1} - C) \\ \mathcal{A}(\Delta X) &= -(\mathcal{A}X - b) \end{aligned}$$

Now consider the construction of primal-dual Newton Step.

The first three equations form the following systems of equations

$$\begin{bmatrix} \mathcal{A}^*(y) + S - C \\ \mathcal{A}(X) - b \\ XS \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \tau I \end{bmatrix}, \; X, S \succ 0$$

Formulate the equations as Newton step

$$\begin{bmatrix} 0 & \mathcal{A}^* & I \\ \mathcal{A} & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta X \\ \Delta y \\ \Delta S \end{bmatrix} = - \begin{bmatrix} \mathcal{A}^*(y) + S - C \\ \mathcal{A}(X) - b \\ XS - \tau I \end{bmatrix}$$

We encounter issues of matrix symmetry since product of two symmetric matrices are not necessarily symmetric. Instead of applying Newton step directly, we use Nesterov-Todd direction in this case.

17.3.5 Nesterov-Todd direction for semidefinite problems

Our goal is to linearize $XS - \tau I = 0$, which is achieved by avaraging primal linearization and dual linearization.

Primal linearization: $S - \tau X^{-1} = 0 \rightsquigarrow \tau X^{-1} \Delta X X^{-1} + \Delta S = \tau X^{-1} - S$

Dual linearization: $X - \tau S^{-1} = 0 \rightsquigarrow \tau S^{-1} \Delta S S^{-1} + \Delta X = \tau S^{-1} - X$

Average of the two: $W^{-1}\Delta X W^{-1} + \Delta S = \tau \Delta X^{-1} - S$,provided $WSW = X$

We achieve the above direction by taking W as the geometric mean of X, S:

$W = S^{1/2}(S^{1/2}XS^{1/2})^{1/2}S^{-1/2} = X^{1/2}(X^{1/2}SX^{1/2})^{-1/2}X^{1/2}$

Thus, given $X, S \succeq 0$, define $\tau(X, S) := -X \bullet S/n$,

primal-dual algorithm for semidefinite programming is as follows:

1. Choose $\sigma \in (0, 1)$

2. Choose (X^0, y^0, S^0) $s.t.$ $X^0, S^0 \succeq 0$

3. For k = 0,1,...

1) Compute Nesterov-Todd direction for

$(X, y, S) = (X^k, y^k, S^k)$, $\tau := \sigma\tau(X^k, S^k)$

2) Choose steplength θ_k and set

$(X^{k+1}, y^{k+1}, S^{k+1}) := (X^k, y^k, S^k) + \theta_k(\Delta X, \Delta y, \Delta S)$