Supplement, Part II: Path Following Methods

The IMA Summer Course 2016: Semidefinite Programming

Shuzhong Zhang

August 3, 2016

1 Proofs of the Duality Theorems

Write the primal problem as

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & x \in (a + \mathcal{L}) \cap \mathcal{K}, \end{array}$$

and its dual problem

$$\begin{array}{ll} \text{minimize} & a^\top s \\ \text{subject to} & s \in (c + \mathcal{L}^\perp) \cap \mathcal{K}^*. \end{array}$$

Lemma 1. *(Alternatives of strong infeasibility). $\{x \in \mathbf{R}^n \mid Ax = b, x \in \mathcal{K}\}$ is not strongly infeasible $\iff \{y \in \mathbf{R}^m \mid A^\top y \in \mathcal{K}^*, b^\top y < 0\} = \emptyset$.*

Proof. Note that for $s \in \mathcal{L}^\perp$, i.e. $s = A^\top y$ for some $y \in \mathbf{R}^m$, we have $a^\top s = a^\top A^\top y = b^\top y$.

Observe the following chain of equivalent statements: $\{y \in \mathbf{R}^m \mid A^\top y \in \mathcal{K}^*, b^\top y < 0\} = \emptyset \iff$ "for all $s \in \mathcal{L}^\perp \cap \mathcal{K}^*$ it must follow $a^\top s \geq 0$" $\iff a \in (\mathcal{L}^\perp \cap \mathcal{K}^*)^* = \mathrm{cl}\,(\mathcal{K} + \mathcal{L}) \iff$ "$\exists\ x_i \in \mathcal{K}$, $\bar{x}_i \in \mathcal{L}$ such that $\|a - (x_i + \bar{x}_i)\| = \|(a - \bar{x}_i) - x_i\| \to 0$" $\iff (a + \mathcal{L}) \cap \mathcal{K}$ is not strongly infeasible. The theorem is proven. □

Lemma 2. *$(a + \mathcal{L}) \cap \mathrm{int}\, \mathcal{K} \neq \emptyset$ if and only if for any $d \neq 0$ and $d \in \mathcal{L}^\perp \cap \mathcal{K}^*$ it must follow $a^\top d > 0$.*

Proof. Note that $x \in \mathrm{int}\, \mathcal{K} \iff \forall\, 0 \neq y \in \mathcal{K}^*$ it follows $x^\top y > 0$.

'$\Longrightarrow$': Let $x \in (a + \mathcal{L}) \cap \mathrm{int}\, \mathcal{K}$. In particular, write $x = a + z$ with $z \in \mathcal{L}$. Take any $d \neq 0$ and $d \in \mathcal{L}^\perp \cap \mathcal{K}^*$:

$$0 < x^\top d = (a + z)^\top d = a^\top d.$$

'$\Longleftarrow$': If $(a + \mathcal{L}) \cap \mathrm{int}\, \mathcal{K} = \emptyset$, then $a + \mathcal{L}$ and $\mathrm{int}\, \mathcal{K}$ can be separated by a hyperplane. That is, there is $s \neq 0$ such that

$$s^\top x \geq 0 \text{ for all } x \in \mathcal{K} \text{ and } s^\top (a + y) \leq 0 \text{ for all } y \in \mathcal{L}.$$

The last inequality means that $s^\top a \leq 0$ and $s^\top y = 0$ for all $y \in \mathcal{L}$. Thus, we have found a non-zero vector $s \in \mathcal{L}^\perp \cap \mathcal{K}^*$ with $a^\top s \leq 0$, which contradicts the assumption. □

Theorem 1. *If the primal conic program satisfies the Slater condition, and its dual is feasible, then the dual has a non-empty and compact optimal solution set.*

Proof. Clearly, the objective value of the dual problem is bounded below by the weak duality theorem. The only possibility for the dual problem *not* to possess an optimal solution is that the 'sup' is non-attainable; i.e. there is an infinite sequence $s^i = c - A^\top y^i \in (a + \mathcal{L}^\perp) \cap \mathcal{K}^*$ such that

$$a^\top s^i = a^\top c - b^\top y^i \to a^\top c - \text{optimal value of the dual problem},$$

and $\|s^i\| \to \infty$. Without loss of generality, suppose that $s^i/\|s^i\|$ converges to d. Then we have

$$d \in \mathcal{L}^\perp \cap \mathcal{K}^* \text{ and } a^\top d = 0.$$

This contradicts with the primal Slater condition, in light of Lemma 2. Hence the optimal solutions must be bounded. □

Therefore, if a pair of primal-dual conic programs satisfy the Slater condition, then attainable optimal solutions exist for both problems, according to Theorem 1. The only remaining issue in this regard is: how about the strong duality theorem? In other words, do these optimal solutions satisfy the complementarity slackness conditions? The answer is yes.

Theorem 2. *If the primal conic program and its dual conic program both satisfy the Slater condition, then the optimal solution sets for both problems are non-empty and compact. Moreover, the optimal solutions are complementary to each other with zero duality gap.*

Proof. The first part of the theorem is simply Theorem 1, stated both for the primal and for the dual parts separately.

We now focus on the second part. Let x^* and s^* be optimal solutions for the primal and the dual conic optimization problems respectively. Write the primal problem in the parametric form:

$$\begin{array}{lll} (P(a)) & \text{minimize} & c^\top x \\ & \text{subject to} & x \in a + \mathcal{L} \\ & & x \in \mathcal{K}. \end{array}$$

Let the optimal value of $(P(a))$, as a function of the vector a, be $v(a)$. Clearly $v(a)$ is a convex function in its domain, $\mathcal{D}$, whenever $(P(a))$ is feasible. Obviously, $\mathcal{D} \supseteq \mathcal{K}$. Also we see that $v(0) = 0$ due to Lemma 1.

By the Slater condition we know that $a \in \operatorname{int} \mathcal{D}$. Let g be a subgradient for v at a; that is

$$v(z) \geq v(a) + g^\top (z - a) \tag{1}$$

for all $z \in \mathcal{D}$.

Observe that for any $d \in \mathcal{L}$ we have $a + \mathcal{L} = a + d + \mathcal{L}$. Thus, $v(a + d) = v(a)$ for any $d \in \mathcal{L}$. This implies $g \in \mathcal{L}^\perp$. Take any $u \in \mathcal{K}$. Since $x^* + u$ is a feasible solution for $(P(a + u))$, it follows that

$$v(a + u) \leq c^\top (x^* + u) = v(a) + c^\top u.$$

Together with the subgradient inequality (1) this yields

$$(c - g)^\top u \geq 0$$

for any $u \in \mathcal{K}$; hence $c - g \in \mathcal{K}^*$. Thus, $c - g$ is a dual feasible solution, and so

$$a^\top s^* \leq a^\top (c - g).$$

Taking $z = 0$, the subgradient inequality gives us

$$0 = v(0) \geq v(a) + g^\top (0 - a) = c^\top x^* - g^\top a. \tag{2}$$

Combining with (2) we have

$$a^\top c - c^\top x^* - a^\top s^* \geq 0.$$

Hence,

$$0 = (x^* - a)^\top (s^* - c) = (x^*)^\top s^* + a^\top c - c^\top x^* - a^\top s^* \geq (x^*)^\top s^* \geq 0.$$

Thus, $(x^*)^\top s^* = 0$, and the complementarity slackness condition is satisfied. □

2 The Path-Following Methods

A natural method for solving constrained optimization, which has been around for more than a half century, is the so-called *barrier function method*.

The method is designed to solve a generic problem of the type: minimize $f(x)$, subject to $x \in S \subseteq \mathbf{R}^n$, where S is known as the constraint set. Suppose that S is full-dimensional, and there is a barrier function $b(x)$ whose domain is precisely $\operatorname{int} S$, and $b(x) \to +\infty$ as x gets close to the boundary of S. The barrier function approach amounts to consider a sequence of *unconstrained* optimization problems: minimize $f(x) + \mu_k b(x)$, with $\mu_k > 0$ a sequence to be set to converge to 0. The classical theory ensures that if such problems were solved to optimality, then their solutions will eventually converge to the optimal solution set of the original problem, as μ_k converges to 0.

There are immediately a number of questions that one will need to answer, such as: (1) Since the choice of barrier functions can be abundant, what are the good choices, and why? (2) How to control the convergence speed of μ_k to zero? (3) How to solve the unconstrained problem? We aim to answer these questions in our analysis.

Consider

$$\begin{array}{ll} \text{minimize} & c^\top x \\ \text{subject to} & Ax = b \\ & x \in \mathcal{K}. \end{array}$$

The linear equality constraints are easy to deal with, and naturally we only need to consider a barrier function for the conic constraint. Consider a convex barrier function $F(x)$ for $\mathcal{K}$:

- $F(x) < \infty$ for all $x \in \text{int } \mathcal{K}$;
- $F(x^k) \to \infty$ as $x^k \to x$ where x is on the boundary of $\mathcal{K}$.

Example 1. *Examples of barrier functions,*

- *For* $\mathcal{K} = \mathbf{R}^n_+$, *a barrier can be* $-\sum_{i=1}^n \log x_i$, *or* $\sum_{i=1}^n 1/x_i$.
- *For* $\mathcal{K} = \text{SOC}(n+1)$, *it can be* $-\log\left(t - \|x\|^2/t\right)$, *or* $-\log(t - \|x\|)$.
- *For* $\mathcal{K} = \mathcal{S}^{n\times n}_+$, *it can be* $-\log\det X$, *or* $\text{tr}\,(X^{-1})$.
- *Let* f *be a convex function, and let*

$$\mathcal{K} := \text{cl }\left\{(p,q,x) \mid p > 0,\ q - pf(x/p) \geq 0\right\},$$

the function $-\log(q - pf(x/p) - \log p$ *is a barrier function.*

Definition 1. *Let* $\mathcal{K} \subseteq \mathbf{R}^n$ *be a given solid, closed, convex cone, and* F *be a barrier function defined in* $\text{int}\,\mathcal{K}$. *We call* F *to be a self-concordant function if for any* $x \in \text{int}\,\mathcal{K}$ *and any direction* $h \in \mathbf{R}^n$ *the following two properties are satisfied:*

- $|\nabla^3 F(x)[h,h,h]| \leq 2(\nabla^2 F(x)[h,h])^{3/2}$;
- $|\nabla F(x)[h]| \leq \theta(\nabla^2 F(x)[h,h])^{1/2}$.

Such barrier function F *is called a self-concordant barrier function for the cone* $\mathcal{K}$ *with the constant* θ *referred to as the complexity value of the cone with respect to the barrier function* F.

In the above definition,

$$\nabla^k F(x)[\underbrace{h,\cdots,h}_{k}] = \left.\frac{d^k F(x+th)}{dt^k}\right|_{t=0}.$$

In particular,

$$\nabla F(x)[h] = \nabla F(x)^\top h,$$

and

$$\nabla^2 F(x)[h,h] = h^\top \nabla^2 F(x) h.$$

So it is clear that the self-concordance is a line property, just like the convexity is.

As a first example, consider $\mathcal{K} = \mathbf{R}^n_+$. Let $F(x) = -\sum_{i=1}^n \log x_i$. Obviously,

$$\nabla F(x)[h] = -\sum_{i=1}^n \frac{h_i}{x_i},\ \nabla^2 F(x)[h,h] = \sum_{i=1}^n \frac{h_i^2}{x_i^2}$$

and

$$\nabla^3 F(x)[h,h,h] = -2\sum_{i=1}^n \frac{h_i^3}{x_i^3}.$$

Hence,

$$|\nabla^3 F(x)[h,h,h]| \le 2(\nabla^2 F(x)[h,h])^{3/2}$$

and

$$|\nabla F(x)[h]| \le \sqrt{n}(\nabla^2 F(x)[h,h])^{1/2}$$

i.e., $\theta = \sqrt{n}$.

Another example is $\mathcal{K} = \mathcal{S}_+^{n\times n}$, and

$$F(X) = -\log\det X.$$

In that case, for any direction $H \in \mathcal{S}^{n\times n}$ we have

$$\nabla F(X)[H] = -\mathrm{tr}\,(X^{-1}H)$$

and

$$\nabla^2 F(X)[H,H] = \mathrm{tr}\,(X^{-1}H)^2$$

and

$$\nabla^3 F(X)[H,H,H] = -2\mathrm{tr}\,(X^{-1}H)^3.$$

Similar as before, one can easily show the self-concordance with $\theta = \sqrt{n}$.

An important property regarding the trilinear form

$$\nabla^3 F(x)[u,v,w] = \sum_{i,j,k} a_{i,j,k} u_i v_j w_k$$

is that

$$\begin{array}{lll} (T) & \text{maximize} & |\nabla^3 F(x)[u,v,w]| \\ & \text{subject to} & \|u\| = \|v\| = \|w\| = 1 \end{array}$$

attains its optimum at the same solution as for

$$\begin{array}{lll} (T)' & \text{maximize} & |\nabla^3 F(x)[u,u,u]| \\ & \text{subject to} & \|u\| = 1. \end{array}$$

Let us see why it is so. Denote the optimal value of (T) to be B, and optimal value of $(T)'$ to be B'. Let $A_k = [a_{ijk}]_{n\times n}$, $k = 1, ..., n$. By the Cauchy-Schwartz inequality we know that the following problem is equivalent to (T):

$$\begin{array}{lll} (T)'' & \text{maximize} & \|(u^\top A_1 v, \cdots, u^\top A_n v)\| \\ & \text{subject to} & \|u\| = \|v\| = 1. \end{array}$$

Consider the KKT condition at an optimal solution (u, v) for $(T)''$:

$$\sum_{k=1}^{n}(u^\top A_k v)A_k v = \mu u$$

and

$$\sum_{k=1}^{n}(u^\top A_k v)A_k u = \nu v$$

where μ and ν are the respective Lagrangian multipliers.

Since $\|u\| = \|v\| = 1$ we have $\mu = \nu = B^2$, where B is the optimal value for $(T)''$. Now adding these two equations yields

$$\sum_{k=1}^{n}(u^\top A_k v)A_k(u+v) = B^2(u+v).$$

Suppose that $u + v \neq 0$ (otherwise consider $u - v$). Let $w := (u+v)/\|u+v\|$. Then, from the previous equation we have

$$\sum_{k=1}^{n}(u^\top A_k v)w^\top A_k w = B^2.$$

Applying Cauchy-Schwartz once more we have $B^2 \leq BB'$. But obviously, $B \geq B'$, and so $B = B'$. As a consequence, we also know that for self-concordant function F we have

$$|\nabla^3 F(x)[u, v, w]| \leq 2(\nabla^2 F(x)[u, u])^{1/2} \cdot (\nabla^2 F(x)[v, v])^{1/2} \cdot (\nabla^2 F(x)[w, w])^{1/2}.$$

A geometry is associated with a local inner product system. Suppose that $F(x)$ is a strictly convex barrier function for the cone $\mathcal{K}$. Consider

$$\langle u, v\rangle = u^\top \nabla^2 F(x) v.$$

The above inner product is coordinate-free, i.e., if we let $y = A^{-1}x$ then the inner product remains invariant.

To be specific about the locality of the inner product, let us denote

$$\langle u, v\rangle_x := u^\top \nabla^2 F(x) v.$$

The norm induced by the inner product is $\|u\|_x := \sqrt{\langle u, u\rangle_x}$.

We now concentrate on a given self-concordant function, $F(x)$, satisfying $|\nabla^3 F(x)[h, h, h]| \leq 2(\nabla^2 F(x)[h, h])^{3/2}$. The geometry induced by the Hessian matrix of $F(x)$ has nice properties, which will lead us to useful bounds whenever the Newton method is used.

Let x and y be in the domain of F. Consider the function

$$f(t) := F(x + t(y - x))$$

where $t \in [0,1]$. Since $f(t)$ is self-concordant, we have

$$|f'''(t)| \leq 2(f''(t))^{3/2},$$

or,

$$\left|\frac{f'''(t)}{2(f''(t))^{3/2}}\right| = \left|\frac{d}{dt}(f''(t))^{-1/2}\right| \leq 1.$$

Integration yields,

$$\left|\frac{1}{\sqrt{f''(t)}} - \frac{1}{\sqrt{f''(0)}}\right| \leq t.$$

This gives us

$$\frac{\sqrt{f''(0)}}{1+t\sqrt{f''(0)}} \leq \sqrt{f''(t)} \leq \frac{\sqrt{f''(0)}}{1-t\sqrt{f''(0)}}.$$

In other words,

$$\frac{h^\top \nabla^2 F(x) h}{\left(1+t\sqrt{h^\top \nabla^2 F(x) h}\right)^2} \leq h^\top \nabla^2 F(x+th) h \leq \frac{h^\top \nabla^2 F(x) h}{\left(1-t\sqrt{h^\top \nabla^2 F(x) h}\right)^2},$$

where $h = y - x$.

Now we wish to use this property to further prove that the local geometry satisfies the following 'continuity' property

$$\frac{\|u\|_y}{\|u\|_x} \leq \frac{1}{1-\|y-x\|_x}$$

for all $u \neq 0$, whenever $\|y-x\|_x \leq 1$. If this is true then

$$\|y-x\|_y \leq \frac{\|y-x\|_x}{1-\|y-x\|_x},$$

and so

$$\frac{\|u\|_y}{\|u\|_x} \geq \frac{1-2\|y-x\|_x}{1-\|y-x\|_x}.$$

Theorem 3. *Let $F(x)$ be a self-concordant function. Then for any x and y in its domain with $\|y-x\|_x < 1$ it holds that*

$$u^\top \nabla^2 F(y) u \leq \frac{u^\top \nabla^2 F(x) u}{(1-\|y-x\|_x)^2}.$$

Proof. Let $z(t) = (1-t)x + ty$ and $\phi(t) = u^\top \nabla^2 F(z(t)) u$. Then,

$$\phi'(t) = \nabla^3 F(z(t))[y-x, u, u].$$

Hence,

$$|\phi'(t)| \leq 2\|y-x\|_{z(t)} \|u\|_{z(t)}^2 = 2\|y-x\|_{z(t)} \phi(t).$$

Notice that

$$\begin{aligned}\|y-x\|_{z(t)} &= \sqrt{h^\top \nabla^2 F(z(t)) h} \\ &= \sqrt{f''(t)} \\ &\le \frac{\sqrt{f''(0)}}{1-t\sqrt{f''(0)}} \\ &= \frac{\|y-x\|_x}{1-t\|y-x\|_x}.\end{aligned}$$

Again, integrating over t, and it follows that

$$\frac{\phi(1)}{\phi(0)} \le \frac{1}{(1-\|y-x\|_x)^2}.$$

This is precisely

$$u^\top \nabla^2 F(y) u \le \frac{u^\top \nabla^2 F(x) u}{(1-\|y-x\|_x)^2}$$

as stated in the theorem. □

As a consequence of the theorem we have

$$\nabla^2 F(y) \preceq \frac{1}{(1-\|y-x\|_x)^2} \nabla^2 F(x),$$

and so,

$$\nabla^2 F(y) \succeq \left(\frac{1-2\|y-x\|_x}{1-\|y-x\|_x}\right)^2 \nabla^2 F(x).$$

Consequently,

$$(\nabla^2 F(y))^{-1} \preceq \left(\frac{1-\|y-x\|_x}{1-2\|y-x\|_x}\right)^2 (\nabla^2 F(x))^{-1}$$

for any $x, y \in \text{dom } F$, and $\|y-x\|_x < 1/2$.

Also, it follows that

$$\|(\nabla^2 F(x))^{-1} \nabla^2 F(x+t(y-x)) - I\| \le \frac{1}{(1-t\|y-x\|_x)^2} - 1$$

where we use the matrix spectrum norm.

Next let us analyze what happens if we apply Newton's method on a self-concordant function. Let x be in the domain of F. The Newton direction of F at x is

$$n(x) = -(\nabla^2 F(x))^{-1} \nabla F(x).$$

There is a crucial quantity that is of importance to us, $\|n(x)\|_x$, the local norm of the Newton direction.

There are two immediate consequences:

- For any $x \in \text{dom } F$, it follows that $\{y \mid \|y-x\|_x < 1\} \subseteq \text{dom } F$.
- If $\|n(x)\|_x < 1$ then the unit Newton step is feasible, i.e. $x_+ = x + n(x) \in \{y \mid \|y-x\|_x < 1\} \subseteq \text{dom } F$.

By Theorem 3 and its consequence, we have

$$(\nabla^2 F(x_+))^{-1} \preceq \left(\frac{1-\|n(x)\|_x}{1-2\|n(x)\|_x}\right)^2 (\nabla^2 F(x))^{-1}.$$

Therefore,

$$\|n(x_+)\|_{x_+}^2 \leq \left(\frac{1-\|n(x)\|_x}{1-2\|n(x)\|_x}\right)^2 \cdot (\nabla F(x_+))^\top (\nabla^2 F(x))^{-1} \nabla F(x_+).$$

Moreover,

$$\nabla F(x_+) = \nabla F(x) + \int_0^1 \nabla^2 F(x + tn(x)) n(x) dt.$$

This gives,

$$(\nabla^2 F(x))^{-1} \nabla F(x_+) = \int_0^1 \left[(\nabla^2 F(x))^{-1} \nabla^2 F(x+tn(x)) - I\right] n(x) dt.$$

Notice that

$$\begin{aligned}
& (\nabla F(x_+))^\top (\nabla^2 F(x))^{-1} \nabla F(x_+) \\
= \; & \|(\nabla^2 F(x))^{-1} \nabla F(x_+)\|_x^2 \\
\leq \; & \left[\int_0^1 \|(\nabla^2 F(x))^{-1} \nabla^2 F(x+tn(x)) - I\| \cdot \|n(x)\|_x dt\right]^2 \\
\leq \; & \left\{\int_0^1 \left[\frac{1}{(1-t\|n(x)\|_x)^2} - 1\right] \|n(x)\|_x dt\right\}^2 \\
= \; & \frac{(\|n(x)\|_x)^4}{(1-\|n(x)\|_x)^2}.
\end{aligned}$$

Finally, this yields

$$\|n(x_+)\|_{x_+} \leq \frac{\|n(x)\|_x^2}{1-2\|n(x)\|_x}.$$

In particular, if $\|n(x)\|_x \leq 1/4$, then

$$\|n(x_+)\|_{x_+} \leq 1/8.$$

In other words, the local norm of Newton directions converges to zero quadratically if the initial iterate has a local Newton direction less than a half.

As a consequence, we can show that if $\|n(x)\|_x \le 1/4$, and x^* is the true minimum point of F, then

$$\|x^* - x\|_x \le 3/4.$$

So far, we only used the property

$$|\nabla^3 F(x)[h,h,h]| \le 2(\nabla^2 F(x)[h,h])^{3/2}$$

for any $x \in \text{dom } F$, and any direction h.

What if $F(x)$ is a barrier for the cone $\mathcal{K}$, and satisfies

$$|\nabla F(x)[h]| \le \theta(\nabla^2 F(x)[h,h])^{1/2}$$

for any $x \in \text{int}\,\mathcal{K}$ and any direction $h \in \mathbf{R}^n$? Actually, in this case, the barrier function can be used to solve the conic optimization problem denoted as (P) below:

$$\begin{array}{lll} (P) & \text{minimize} & c^\top x \\ & \text{subject to} & x \in a + \mathcal{L} \\ & & x \in \mathcal{K}, \end{array}$$

where $\mathcal{L} = \{x \in \mathbf{R}^n \mid Ax = 0\}$.

Let

$$F_\mu(x) = \frac{1}{\mu} c^\top x + F(x).$$

Obviously, $F_\mu(x)$ is also a barrier function for (P).

Let us suppose that an appropriate initial feasible solution for (P) is available.

For any $0 < \mu' < \mu$, we have

$$\begin{aligned} \nabla^2 F_{\mu'}(x) &= \nabla^2 F_\mu(x) = \nabla^2 F(x) \\ \nabla F_{\mu'}(x) &= \frac{\mu}{\mu'} \nabla F_\mu(x) + \frac{\mu' - \mu}{\mu'} \nabla F(x). \end{aligned}$$

Let

$$n(\mu; x) = -(\nabla^2 F_\mu(x))^{-1} \nabla F_\mu(x),$$

and

$$p(\mu; x) := \|n(\mu; x)\|_x,$$

namely, the local norm of Newton direction of F_μ.

Then we have

$$\begin{aligned} p(\mu';x) &= \left\| \frac{\mu}{\mu'} n(\mu;x) + \frac{\mu'-\mu}{\mu'} n(x) \right\|_x \\ &\le \frac{\mu}{\mu'} \|n(\mu;x)\|_x + \frac{\mu-\mu'}{\mu'} \|n(x)\|_x . \end{aligned}$$

Now, if

$$|\nabla F(x)[h]| \le \theta (\nabla^2 F(x)[h,h])^{1/2}$$

holds for all h, then we may let $h = n(x)$, and this gives

$$\|n(x)\|_x \le \theta.$$

Hence,

$$p(\mu';x) \le \frac{\mu}{\mu'} p(\mu;x) + \frac{\mu-\mu'}{\mu'}\theta.$$

We are now in the position to present a (short-step) version of the central path follow method.

Suppose that we start from a given $\mu^0 > 0$ and x^0 satisfies $p(\mu^0;x^0) \le 1/4$. Let the iteration count be $i = 0$.

Each step of the method now consists of two types of operations: taking a full Newton step, and shifting the target by reducing the parameter μ.

Short-step central path following

Newton step. Let $x^{i+1} = x^i + n(\mu^i;x^i)$.

Target shifting. Let μ^{i+1} be so that $p(\mu^{i+1};x^{i+1}) = 1/4$.

Theorem 4. *It holds that*

$$\frac{\mu^{i+1}}{\mu^i} \le 1 - \frac{1}{2+8\theta}$$

for all $i = 0, 1, \ldots$.

Proof. Since one Newton step brings $p(\mu^i;x^i) = 1/4$ to $p(\mu^i;x^{i+1}) \le 1/8$, we have

$$1/4 \le \frac{\mu}{\mu'} \cdot \frac{1}{8} + \frac{\mu-\mu'}{\mu'}\theta.$$

This can be equivalently written as

$$\frac{\mu'}{\mu} \le 1 - \frac{1/8}{1/4+\theta} = 1 - \frac{1}{2+8\theta}.$$

This proves the desired result. □

As a result we have the following complexity estimation.

Theorem 5. *Suppose that* $\mu^0 = O(1)$. *Then, in* $O(\theta \log \frac{1}{\epsilon})$ *Newton steps we will reach a point* x *with* $\mu < \epsilon$ *and* $p(\mu; x) \leq 1/4$.

The method described above is called *short step path following.* In practical implementation, it is considered too conservative. Below we shall introduce another variant of the method, known as *long step path following.*

Suppose that we start from a given $\mu^0 > 0$ and x^0. Let the iteration count be $i = 0$. Each step of the method also consists of two types of operations: taking a damped Newton step, and shifting the target by reducing the parameter μ.

Long-step central path following

Newton step.

- Let $y = x^i$.
- While $p(\mu^i; y) \geq 1/8$, find the Newton direction $n(\mu^i; y)$ and do line minimization

  $$t := \text{argmin } F_{\mu^i}(y + tn(\mu^i; y)),$$

  and

  $$y := y + tn(\mu^i; y)$$

  and return to while.
- *Update the iterate*: Let $x^{i+1} = y$.

Target shifting. Let $\mu^{i+1} = \mu^i/2$.

By construction, in the *while* loop, it holds that $p(\mu^i; y) \geq 1/8$ and $p(\mu^i; x^{i+1}) < 1/8$.

Theorem 6. *Suppose that* $\mu^0 = O(1)$. *Then, in* $O(\log \frac{1}{\epsilon})$ *number of target shifting we will reach a point* x *with* $\mu < \epsilon$ *and* $p(\mu; x) \leq 1/8$. *Between each target shifting it takes at most* $O(\theta^2)$ *numbers of Newton steps.*

Proof of the theorem.

The key is certainly to estimate the number of Newton steps needed between each time of target shifting.

In particular, we shall estimate the number of Newton steps needed between x^{i+1} and x^{i+2}.

It will be convenient to introduce the following lemma.

Lemma 3. *If $F(x)$ is a convex function satisfying*

$$|\nabla F(x)[h]| \le \theta(\nabla^2 F(x)[h,h])^{1/2}$$

for any $x \in$ dom F and any direction h, then

$$\nabla F(x)^\top (y-x) \le \theta^2$$

for any x and y in dom F.

Proof of the lemma.

The statement is trivial if $\nabla F(x)^\top (y-x) \le 0$.

We consider the case where $\nabla F(x)^\top (y-x) > 0$. Let

$$g(t) = \nabla F(x + t(y-x))^\top (y-x).$$

Then,

$$0 \le g(t)^2 \le \theta^2 g'(t).$$

This means that $g'(t) \ge 0$. Hence $g(t) \ge g(0) > 0$ for all $0 \le t \le 1$.

On the other hand, integration gives us

$$\frac{t}{\theta^2} \le \frac{1}{g(0)} - \frac{1}{g(t)}.$$

Thus,

$$\frac{1}{\theta^2} \le \frac{1}{g(0)} - \frac{1}{g(1)} \le \frac{1}{g(0)}$$

and so $g(0) \le \theta^2$. □

This lemma gives us a handle on how much do we need to reduce the function $F_{\mu^{i+2}}(x)$ once the target is updated.

Let x_j^* be the minimum point of $F_{\mu^j}(x)$. Then,

$$\nabla F_{\mu^j}(x_j^*) = \frac{c}{\mu^j} + \nabla F(x_j^*) = 0.$$

$p(\mu^{i+1}; x^{i+2}) < 1/8 \Longrightarrow \|x^{i+2} - x_{i+1}^*\|_{x^{i+2}} \le 3/4.$

Therefore,

$$\begin{aligned}
& F_{\mu^{i+2}}(x^{i+2}) - F_{\mu^{i+2}}(x^*_{i+2}) \\
= \; & F_{\mu^{i+2}}(x^{i+2}) - F_{\mu^{i+2}}(x^*_{i+1}) + F_{\mu^{i+2}}(x^*_{i+1}) - F_{\mu^{i+2}}(x^*_{i+2}) \\
\leq \; & (\nabla F_{\mu^{i+2}}(x^{i+2}))^\top (x^{i+2} - x^*_{i+1}) + (\nabla F_{\mu^{i+2}}(x^*_{i+1}))^\top (x^*_{i+1} - x^*_{i+2}) \\
= \; & 2 \times (\nabla F_{\mu^{i+1}}(x^{i+2}))^\top (x^{i+2} - x^*_{i+1}) - \nabla F(x^{i+2})^\top (x^{i+2} - x^*_{i+1}) - \nabla F(x^*_{i+1})^\top (x^*_{i+1} - x^*_{i+2}) \\
\leq \; & 2 \times p(\mu^{i+1}; x^{i+2}) \times \frac{3}{4} + 2\theta^2.
\end{aligned}$$

Next, consider one particular step of *Operation 1.*

Our aim now is to prove that, if $\|n(x)\|_x$ is large then the Newton method with line minimization substantially reduces the value of F.

Fix any $x \in \operatorname{dom} F$. Let

$$f(t) = F(x + tn(x)).$$

By the self-concordance, we have

$$|f'''(t)| \leq 2(f''(t))^{3/2}$$

for all $0 < t < 1/\|n(x)\|_x = 1/\sqrt{f''(0)}$.

As we have seen before, integration and re-arrange we get

$$\frac{f''(0)}{(1 + t\sqrt{f''(0)})^2} \leq f''(t) \leq \frac{f''(0)}{(1 - t\sqrt{f''(0)})^2}.$$

Integration once again, we obtain,

$$\frac{tf''(0)}{1 + t\sqrt{f''(0)}} \leq f'(t) - f'(0) \leq \frac{tf''(0)}{1 - t\sqrt{f''(0)}}.$$

Finally, we integrate for the second inequality, and this leads to

$$f(t) \leq f(0) + \left(f'(0) - \sqrt{f''(0)}\right) t - \log\left(1 - t\sqrt{f''(0)}\right)$$

where $f'(0) = -f''(0) = -\|n(x)\|_x^2$.

Notice that $-\log(1 - \delta) \leq \delta + \delta^2$ whenever $0 \leq \delta \leq 1/2$. We have

$$\begin{aligned}
f(t) & \leq f(0) + \left(f'(0) - \sqrt{f''(0)}\right) t - \log\left(1 - t\sqrt{f''(0)}\right) \\
& \leq f(0) - \left(2\|n(x)\|_x + \|n(x)\|_x^2\right) t + \|n(x)\|_x^2 t^2
\end{aligned}$$

where $t \leq 1/(2\|n(x)\|_x)$.

Let us choose $\bar{t} = \max\{1/(2\|n(x)\|_x), 1\}$. Then,

$$\begin{aligned} f(\bar{t}) &\leq f(0) - 2\|n(x)\|_x \bar{t} \\ &\leq f(0) - \max\{2\|n(x)\|_x, 1\} \\ &\leq f(0) - 1/4. \end{aligned}$$

This means that if the local norm of the Newton direction is at least 1/8, then the function value can be reduced by at least 1/4 if one applies Newton's method with line minimization.

This shows that between two target shifting steps we can have at most $O(\theta^2)$ Newton steps.

This completes the proof for the theorem.

The final remaining question is, what good does it do if at each step we are close to the minimum of

$$\begin{array}{ll} \min & \frac{c^\top x}{\mu} + F(x) \\ \text{s.t.} & x \in a + \mathcal{L}? \end{array}$$

Well, the solution $x(\mu)$ itself satisfies

$$\frac{c}{\mu} + \nabla F(x(\mu)) = y(\mu)$$

with $y(\mu) \in \mathcal{L}^\perp$.

This means that

$$s(\mu) = -\mu \nabla F(x(\mu))$$

is a feasible dual slack. Therefore, the duality gap is equal to

$$x(\mu)^\top s(\mu) = \mu(\nabla F(x(\mu)))^\top (0 - x(\mu)) \leq \mu\theta^2$$

where we used Lemma 3.

This shows that if we drive $\mu \to 0$, then the duality gap will converge to zero.

References

[1] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[2] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2001.

[3] J.F. Sturm, *Theory and Algorithms for Semidefinite Programming.* In *High Performance Optimization*, eds. H. Frenk, K. Roos, T. Terlaky, and S. Zhang, Kluwer Academic Publishers, 1999.

[4] Yu. Nesterov and A. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM Studies in Applied Mathematics, 13, 1994.

[5] J. Renegar, *A Mathematical View of Interior Point Methods in Convex Optimization*, MPS-SIAM Series on Optimization, 2001.

[6] Yu. Nesterov, *Introductory Lectures on Convex Optimization*, Kluwer Academic Publishers, 2004.