Appendix F

Self-Concordant Barrier Functions for Convex Optimization

F.1 Introduction

In this Appendix we present a framework for developing polynomial-time algorithms for the solution of convex optimization problems. The approach is based on an interior-point barrier method. Key to this framework is the assumption that the barrier function is "self concordant," a property that we define below. For linear constraints, the ordinary logarithmic barrier function can be used. It is possible to prove that, for any convex feasible region with the properties we specify below, there exists an appropriate self-concordant barrier function. Thus the results we describe here provide a general theoretical approach for solving convex programming problems.

The interior-point method utilizes Newton's method. We derive a bound on the number of Newton iterations required to determine the optimal function value to within some tolerance. Each Newton iteration involves the solution of the Newton equations, requiring $O(n^3)$ computations. A prescribed step length is used, so there is no line search. Ignoring the computations for evaluating the gradient and Hessian, the algorithm determines the solution to within a specified tolerance in a polynomial number of operations. If polynomial algorithms exist to evaluate the gradient and Hessian, the overall algorithm is polynomial.

Two major results are required to prove that the overall algorithm is a polynomial algorithm. The first states that if Newton's method is applied to a single barrier subproblem, and the initial guess is "close" to the solution, then the number of iterations required to find an approximate solution of this subproblem is bounded. The second states that, if the barrier parameter is not changed too quickly, then an approximate solution of one subproblem will not be "too far" from the solution of the next subproblem.

These two results might at first seem to be obvious. They are not. In the form that we state the convex programming problem, its solution is usually at the boundary of the feasible region. The barrier function is singular on the boundary of the feasible region, and so the Hessian of the barrier function becomes ill-conditioned as the solution is approached. In standard convergence results for Newton's method (see Chapter 11), the rate constants can be shown to depend on the condition number of the Hessian, a number that is tending to $+\infty$ in the case of a barrier function.

To analyze the behavior of Newton's method in this case, we must in some manner take this singularity into account. To do this, it will be useful to define a norm $\|\cdot\|_x$ in terms of the Hessian of the barrier function evaluated at a point x. We will measure "closeness" in terms of this norm.

This norm depends on the Hessian, and so changes as the variables change. If the Hessian changed rapidly, then it might not be possible to use the values of $\|x - x_*\|_x$ to draw conclusions about the convergence of the method. Thus, the rate of change of the Hessian matrix must not be "too great." This reasoning leads us to impose a bound on the third derivatives of the barrier function in terms of the Hessian (see Section F.2.1). This bound is all that is required to prove the first major result corresponding to the behavior of Newton's method on a single barrier subproblem.

To prove the second major result, that is, that the approximate solution of one subproblem will not be too far from the solution of the next subproblem, it is necessary that the values of the barrier functions not change "too quickly" as the barrier parameter changes. To guarantee this, we will impose a bound on the first derivatives of the barrier functions in terms of the Hessian (see Section F.3). By measuring all quantities in terms of the Hessian, we are able to circumvent the difficulties associated with the singularity of the barrier function at the solution.

If the barrier function has these properties, then an interior-point method can be designed so that the optimal solution of a convex programming problem can be found (to within some tolerance) using a polynomial number of Newton iterations.

F.2 Basic Ideas of Self Concordance

Our focus in this section is on the behavior of Newton's method when applied to a single barrier subproblem. We start by analyzing properties of the barrier functions that permit "good" performance of Newton's method.

F.2.1 Self-Concordant Functions

If Newton's method is applied to a quadratic function, then it converges in one iteration. By extension, if the Hessian matrix does not change rapidly, then Newton's method ought to converge quickly. Thus, we might say that the radius of convergence of Newton's method for minimizing a function $F(x)$ is inversely proportional to the "nonlinearity" of F. Newton's method performs "well" if small changes in x lead to small changes in the second derivative of F. Change in the second derivative can be measured using the third derivative. Intuitively, the third derivative should be small relative to the second derivative. The self-concordance property reflects this requirement.

A one-dimensional, convex barrier function $F(x)$ is called *self concordant* if

$$|F'''(x)| \le 2F''(x)^{3/2}$$

for every x in the interior of the function's domain. A simple example is the logarithmic barrier function $F(x) = -\log(x)$ for $x > 0$. Then $F''(x) = 1/x^2$ and $F'''(x) = -2/x^3$, and the inequality above is satisfied.

There is nothing special about the constant 2 in the definition. If instead the function satisfied

$$|F'''(x)| \le CF''(x)^{3/2}$$

for some constant C, we could replace F by the scaled function $\hat{F} = \frac{1}{4}C^2 F$, and then $\hat{F}$ would be self concordant. The number 2 is used in the definition so that the function $F(x) = -\log x$ is self-concordant without any scaling.

To define self-concordancy for a multi-dimensional function, we insist that the third derivative *along each direction* be bounded by the second derivative in that direction via the relation above. We now give a more precise definition.

Let S be a bounded, closed, convex subset of $\Re^n$ with non-empty interior $\operatorname{int} S$. (The assumption that S is bounded is not that important, since we could modify the optimization problem by adding artificial, very large bounds on the variables.) Let $F(x)$ be a convex function defined on the set S, and assume that F has three continuous derivatives. Then F is *self concordant* on S if:

1. (barrier property) $F(x_i) \to \infty$ along every sequence $\{ x_i \} \subset \operatorname{int} S$ converging to a boundary point of S.
2. (differential inequality) F satisfies

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

   for all $x \in \operatorname{int} S$ and all $h \in \Re^n$.

In this definition,

$$\nabla^3 F(x)[h_1,h_2,h_3] \equiv \frac{\partial^3}{\partial t_1 \partial t_2 \partial t_3} F(x + t_1 h_1 + t_2 h_2 + t_3 h_3)\Big|_{t_1=t_2=t_3=0},$$

that is, it is a third-order directional derivative of F. Since

$$\begin{aligned}\nabla^2 F(x)[h,h] &\equiv \frac{\partial^2}{\partial t_1 \partial t_2} F(x + t_1 h + t_2 h)\Big|_{t_1=t_2=0} = h^T \nabla^2 F(x) h,\\ \nabla F(x)[h] &\equiv \frac{\partial}{\partial t_1} F(x + t_1 h)\Big|_{t_1=0} = \nabla F(x)^T h,\end{aligned}$$

these definitions are consistent with the formulas for directional derivatives derived in Section 5 of Chapter 11.

Some examples of self-concordant functions are given in the Problems. One particular case is the logarithmic barrier function

$$F(x) = -\sum_{i=1}^{m} \log(a_i^T x - b_i)$$

that is self concordant on the set $S = \{\, x : a_i^T x - b_i \ge 0,\ i = 1, \ldots, m \,\}$.

For the remainder of this Chapter, we make several assumptions to simplify our discussion. They are not essential; in fact, almost identical results can be proved without these assumptions. We assume that $\nabla^2 F(x)$ is nonsingular for all $x \in \operatorname{int} S$. This allows us to define a norm as follows:

$$\|h\|_x^2 \equiv h^T \nabla^2 F(x) h.$$

(See the Problems.) We also assume that F has a minimizer $x_* \in \operatorname{int} S$. Because F is convex, these assumptions guarantee that x_* is the unique minimizer of F in S.

The following lemmas indicate some basic properties of self-concordant functions. The first shows that the third-order directional derivative can be bounded using the norm that we have defined.

Lemma F.1. *The third-order directional derivative of a self-concordant function satisfies*

$$|\nabla^3 F(x)[h_1, h_2, h_3]| \le 2 \|h_1\|_x \|h_2\|_x \|h_3\|_x .$$

Proof. See the Problems. □

The next lemma bounds how rapidly a self-concordant function $F(x)$ and its Hessian can change if a step is taken whose norm is less than one. The first result is an analog of a Taylor series expansion for a self-concordant function. The second is a bound on how rapidly the norm we have defined can change when x changes.

Lemma F.2. *Let F be self-concordant on S. Let $x \in int S$ and suppose that $\|h\|_x < 1$. Then $x + h \in int S$, and*

$$F(x) + \nabla F(x)^T h + \beta(-\|h\|_x) \le F(x+h) \le F(x) + \nabla F(x)^T h + \beta(\|h\|_x) \quad (F.1)$$

where

$$\beta(s) = -\log(1-s) - s = \frac{s^2}{2} + \frac{s^3}{3} + \frac{s^4}{4} + \cdots .$$

The lower bound in (F.1) is satisfied even if $\|h\|_x \ge 1$. Furthermore, for any $g \in \Re^n$,

$$(1 - \|h\|_x) \|g\|_x \le \|g\|_{x+h} \le (1 - \|h\|_x)^{-1} \|g\|_x . \quad (F.2)$$

Proof. The proof is in three parts. We assume initially that $x+h \in \operatorname{int} S$ and prove (F.1) and (F.2). These two results form parts 1 and 2 of the proof, respectively. In part 3, we prove that $x + h \in \operatorname{int} S$ must hold.

Suppose that $x + h \in \operatorname{int} S$, and let $r = \|h\|_x < 1$. If $r = 0$ then $h = 0$ and the results are trivially true, so we assume that $r > 0$.

Part 1: To prove (F.1), define

$$\phi(t) \equiv h^T \nabla^2 F(x + th) h = \|h\|_{x+th}^2 .$$

We will derive a Taylor series expansion of $\phi(t)^{-1/2}$ and use it to bound $\phi(t)$ in terms of $\phi(0)$. The resulting bound will be used to obtain (F.1).

We first obtain the Taylor series expansion. The function ϕ is continuously differentiable for $0 \le t \le 1$, and

$$\phi(t) > 0, \quad \phi(0) = r^2 < 1, \quad \text{and} \quad |\phi'(t)| = |\nabla^3 F(x+th)[h,h,h]| \le 2\phi(t)^{3/2}.$$

Hence

$$\left| \frac{d}{dt}\left[\phi(t)^{-1/2}\right] \right| = \left| -\tfrac{1}{2}\phi(t)^{-3/2}\phi'(t) \right| \le 1.$$

If we expand $\phi(t)^{-1/2}$ in a Taylor series with remainder, then

$$\phi(t)^{-1/2} = \phi(0)^{-1/2} + t\left[\frac{d}{ds}\phi(s)^{-1/2}\right]_{s=\eta}$$

for some η between 0 and t. Using the bound above we obtain

$$\phi(0)^{-1/2} - t \le \phi(t)^{-1/2} \le \phi(0)^{-1/2} + t$$

for $0 \le t \le 1$. Because $\phi(0) = r^2 < 1$, this can be rearranged to obtain the desired bound on $\phi(t)$:

$$\frac{\phi(0)}{(1+t\phi(0)^{1/2})^2} \le \phi(t) \le \frac{\phi(0)}{(1-t\phi(0)^{1/2})^2}.$$

Thus

$$\frac{r^2}{(1+rt)^2} \le \phi(t) \equiv h^T\nabla^2 F(x+th)h \le \frac{r^2}{(1-rt)^2} \tag{F.3}$$

for $0 \le t \le 1$. (Recall that $r = \|h\|_x$.)

If we integrate (F.3) twice, we get

$$F(x) + \nabla F(x)^T h + \int_0^1 \Big[\int_0^\tau \frac{r^2}{(1+rt)^2}dt\Big]d\tau$$
$$\le F(x+h) \le F(x) + \nabla F(x)^T h + \int_0^1 \Big[\int_0^\tau \frac{r^2}{(1-rt)^2}dt\Big]d\tau.$$

Evaluating the integrals leads to (F.1).

It is straightforward to verify that the derivation of the lower bound remains valid even in the case where $r = \|h\|_x \ge 1$. (See the Problems.) This completes part 1 of the proof.

Part 2: We now prove (F.2). Let $g \in \Re^n$ and define

$$\psi(t) \equiv g^T\nabla^2 F(x+th)g = \|g\|_{x+th}^2 .$$

We will use (F.3) to derive a bound on $\psi'(t)$. This bound will be used to bound $\psi(t)$ in terms of $\psi(0)$, and this will lead directly to (F.2).

The function ψ is a continuously differentiable non-negative function for $0 \le t \le 1$. Lemma F.1 implies that

$$|\psi'(t)| = |\nabla^3 F(x+th)[g,g,h]| \le 2\|g\|_{x+th}^2 \|h\|_{x+th} .$$

Thus using (F.3) we obtain

$$|\psi'(t)| \le 2\psi(t)\phi(t)^{1/2} \le 2\psi(t)\frac{r}{1-rt}$$

for $0 \le t \le 1$. This is the desired bound on $\psi'(t)$.

We now use this to bound $\psi(t)$ in terms of $\psi(0)$. The bound on $\psi'(t)$ implies that

$$\frac{d}{dt}[(1-rt)^2\psi(t)] = (1-rt)^2[\psi'(t) - 2r(1-rt)^{-1}\psi(t)] \le 0$$
$$\frac{d}{dt}[(1-rt)^{-2}\psi(t)] = (1-rt)^{-2}[\psi'(t) + 2r(1-rt)^{-1}\psi(t)] \ge 0$$

and so

$$(1-rt)^2\psi(t) \le \psi(0) \quad \text{and} \quad (1-rt)^{-2}\psi(t) \ge \psi(0).$$

Rearranging yields

$$(1-rt)^2\psi(0) \le \psi(t) \le (1-rt)^{-2}\psi(0).$$

If we substitute into this inequality the definitions of r and ψ, we obtain (F.2). This completes part 2 of the proof.

Part 3: We now prove that if $\|h\|_x < 1$ then $x + h \in \operatorname{int} S$. Suppose by contradiction that $x + h$ is not in $\operatorname{int} S$. Consider now the line segment $x + th$ for $0 \le t \le 1$. Since $\operatorname{int} S$ is open and convex, there is some "boundary" point $y = x + \bar{t}h$ along this line segment such that $y \notin \operatorname{int} S$, but the half-open segment $[x, y) = \{\, z = x + th : 0 \le t < \bar{t}\,\} \subset \operatorname{int} S$. For each point $z \in [x, y)$, the norm $\|z - x\|_x$ is less than 1, and hence from (F.1), $F(z)$ is bounded. But since F has the barrier property, $F(z)$ should go to ∞ as z approaches y. Hence we have a contradiction. This completes the proof. □

F.2.2 The Newton Decrement

Many of our results will be phrased in terms of a quantity called the "Newton decrement." It is defined below. The Newton decrement measures the norm of the Newton direction, but indirectly it indicates how close we are to the solution of a barrier subproblem. For a linear program, the Newton decrement is equivalent to the 2-norm measure of proximity to the barrier trajectory that was defined in Section 10.6 (see the Problems). We will use the Newton decrement in the convergence results, in place of the more traditional measures of convergence, such as $\|x - x_*\|$ and $|F(x) - F(x_*)|$.

Let $x \in \operatorname{int} S$ and let p_N be the Newton direction for F at x. We define the *Newton decrement* of F at x to be

$$\delta = \delta(F, x) = \|p_N\|_x\,.$$

Consider the Taylor series approximation to $F(x + h)$:

$$F(x) + \nabla F(x)^T h + \tfrac{1}{2} h^T \nabla^2 F(x) h = F(x) + \nabla F(x)^T h + \tfrac{1}{2}\|h\|_x^2\,.$$

The Newton direction p_N minimizes this approximation and is the solution to

$$\nabla^2 F(x) p_N = -\nabla F(x).$$

It is straightforward to check that the optimal value of the Taylor series approximation is

$$F(x) - \tfrac{1}{2}\delta(F,x)^2.$$

This indicates why $\delta(F,x)$ is called the Newton decrement.

We have the following lemma.

Lemma F.3. *The Newton decrement satisfies*

$$\delta(F,x) = \max\left\{ \nabla F(x)^T h : \|h\|_x \le 1 \right\}.$$

Proof. See the Problems. □

Many of our results will be expressed in terms of the Newton decrement. We will be able to obtain bounds on $F(x) - F(x_*)$ and $\|x - x_*\|$ in terms of the Newton decrement, and we will also be able to measure the progress at each iteration of the method in terms of the Newton decrement. Thus, statements about the convergence of the method in terms of the Newton decrement will indirectly provide us with information about convergence as measured in the more traditional ways.

The first of these results (a bound on $F(x) - F(x_*)$) is given below. Its proof, which depends on a notion of duality different than that presented in Chapter 14, has not been included here.

Lemma F.4. *Let $x \in \operatorname{int} S$ with $\delta(F,x) < 1$, and let $x_* \in S$ be the minimizer of F. Then*

$$F(x) - F(x_*) \le \beta(\delta(F,x)),$$

where $\beta(\cdot)$ is the function defined in Lemma F.2.

Proof. See the book by Nesterov and Nemirovsky (1993). □

F.2.3 Convergence of the Damped Newton Method

Each barrier subproblem will be solved using a "damped" Newton method, that is, a step is taken along the Newton direction but with a specified step length that is less than one. If we denote the Newton direction at x by p_N, then the method is defined by

$$x_+ = x + \frac{1}{1 + \delta(F,x)} p_N.$$

The reason for including this steplength is that the resulting displacement will always have norm less than one, so that the lemmas of the previous Section apply. As the method converges and the Newton direction approaches zero, the step length approaches the pure Newton step of one. The rest of this Section develops the properties of the damped-Newton method.

The next lemma gives a lower bound on how much the function F will be decreased by a step of the damped Newton method.

Lemma F.5. *If x_+ is the result of the damped Newton iteration, then $x_+ \in \operatorname{int} S$ and*

$$F(x) - F(x_+) \ge \delta(F,x) - \log(1 + \delta(F,x)).$$

Proof. Since $\|p_N\|_x = \delta \equiv \delta(F,x)$, we have that $\|x_+ - x\|_x = \delta/(1+\delta) < 1$ and thus by Lemma F.2, $x_+ \in \operatorname{int} S$. Using Lemma F.2, the definition of $\beta(\cdot)$, and the properties of δ and p_N, we obtain

$$\begin{aligned} F(x_+) &\le F(x) + \frac{1}{1+\delta}\nabla F(x)^T p_N + \beta\left(\frac{1}{1+\delta}\|p_N\|_x\right) \\ &= F(x) - \frac{1}{1+\delta} p_N^T \nabla^2 F(x) p_N + \beta\left(\frac{\delta}{1+\delta}\right) \\ &= F(x) - \frac{\delta^2}{1+\delta} - \log\left(1 - \frac{\delta}{1+\delta}\right) - \frac{\delta}{1+\delta} \\ &= F(x) - \delta + \log(1+\delta). \end{aligned}$$

Thus

$$F(x) - F(x_+) \ge \delta - \log(1+\delta),$$

as desired. □

The lemma shows that the damped Newton step is well-defined, in the sense that the iterates remain in $\operatorname{int} S$. The lemma also shows that

$$F(x) - F(x_+) \ge \delta - \log(1+\delta),$$

where $\delta = \delta(F,x)$. The right-hand side is zero when $\delta = 0$, and is positive and strictly increasing for $\delta > 0$. The result gives a lower bound on how much progress is made at each Newton iteration.

If $\delta(F,x)$ remains large, then Newton's method must decrease the value of $F(x)$ by a nontrivial amount. Since the function is bounded below on S, this cannot go on indefinitely, and so $\delta(F,x)$ must ultimately become small. The next theorem shows that, if $\delta(F,x)$ is small, then the method converges at a "quadratic rate" in terms of the Newton decrement. (This is not the same as the notion of convergence rate defined in Chapter 2; see the Exercises.) Together, these results provide a bound on the number of Newton iterations required to solve the optimization problem to within some tolerance. This argument is made precise in Theorem F.7.

The theorem below proves that $\delta(F,x_+) \le 2\delta(F,x)^2$, regardless of the value of x. This result is primarily of interest to us in the case when δ is small. (When δ is large, Lemma F.5 is more useful.) It also determines a bound on $\|x - x_*\|_{x_*}$ in terms of δ, and thus shows that if δ is small, then the norm of the error is small as well.

Theorem F.6. *If $x \in \text{int}\, S$ then*

$$\delta(F, x_+) \le 2\delta(F, x)^2.$$

Let x_* be the minimizer of F in S, and assume that $\delta(F, x) < 1$. Then

$$\|x - x_*\|_{x_*} \le \frac{\delta(F, x)}{1 - \delta(F, x)}.$$

Proof. The proof is in two parts, proving each of the results in turn.

Part 1: We will prove that $\delta(F, x_+) \le 2\delta(F, x)^2$. Let p_N be the Newton direction at x, $\delta = \delta(F, x)$, and $\alpha = 1/(1+\delta)$. Thus $x_+ = x + \alpha p_N$. For any $h \in \Re^n$ we define

$$\psi(t) = \nabla F(x + tp_N)^T h.$$

This function is twice continuously differentiable for $0 \le t \le \alpha$, with

$$\begin{aligned} \psi'(t) &= p_N^T \nabla^2 F(x + tp_N) h \\ \psi''(t) &= \nabla^3 F(x + tp_N)[h, p_N, p_N]. \end{aligned}$$

By Lemma F.1, Lemma F.2, and the definition of the Newton decrement we have

$$\begin{aligned} |\psi''(t)| &\le 2\,\|h\|_{x+tp_N}\,\|p_N\|_{x+tp_N}^2 \\ &\le 2(1 - t\delta)^{-3}\,\|h\|_x\,\|p_N\|_x^2 \\ &= 2(1 - t\delta)^{-3}\,\|h\|_x\,\delta^2. \end{aligned}$$

If we integrate ψ'' twice, we obtain that

$$\begin{aligned} \nabla F(x_+)^T h \equiv \psi(\alpha) &\le \psi(0) + \alpha\psi'(0) + \|h\|_x \int_0^\alpha \Big[\int_0^t 2(1 - \tau\delta)^{-3}\delta^2 d\tau\Big] dt \\ &= \psi(0) + \alpha\psi'(0) + \frac{\delta^2\alpha^2}{1 - \delta\alpha}\,\|h\|_x \\ &= \nabla F(x)^T h + \alpha p_N^T \nabla^2 F(x) h + \frac{\delta^2\alpha^2}{1 - \delta\alpha}\,\|h\|_x \\ &= (1 - \alpha)\nabla F(x)^T h + \frac{\delta^2\alpha^2}{1 - \delta\alpha}\,\|h\|_x \qquad \text{(definition of } p_N\text{)} \\ &= \frac{\delta}{1+\delta}\nabla F(x)^T h + \frac{\delta^2}{1+\delta}\,\|h\|_x \qquad \text{(definition of } \alpha\text{)} \\ &\le \frac{2\delta^2}{1+\delta}\,\|h\|_x \qquad \text{(by Lemma F.3)} \\ &\le \frac{2\delta^2}{1+\delta}\frac{1}{1 - \alpha\delta}\,\|h\|_{x_+} \qquad \text{(by Lemma F.2)} \\ &= 2\delta^2\,\|h\|_{x_+}. \end{aligned}$$

Since this is true for all h we have

$$\delta(F, x_+) = \max\Big\{\,\nabla F(x_+)^T h : \|h\|_{x_+} \le 1\,\Big\} \le 2\delta^2.$$

This completes Part 1 of the proof.

Part 2: We will prove that $\|x - x_*\|_{x_*} \le \delta(F,x)/1 - \delta(F,x)$. Let $x \in \operatorname{int} S$ be such that $\delta = \delta(F,x) < 1$. From Lemma F.4 we have that

$$F(x) - F(x_*) \le \beta(\delta) = -\log(1-\delta) - \delta.$$

Let $q = \|x - x_*\|_{x_*}$. From Lemma F.2 applied at x_* it follows that

$$F(x) \ge F(x_*) + \beta(-q) = F(x_*) + q - \log(1+q).$$

Combining these two results we get

$$q - \log(1+q) \le -\delta - \log(1-\delta),$$

and it follows that $q \le \delta/(1-\delta)$, which is the desired result. (See the Exercises.) This completes the proof. □

We conclude with a summary theorem, the results of which are direct consequences of the previous results. It provides a bound on the number of Newton iterations required to solve a single barrier subproblem to within a tolerance.

Theorem F.7. *Let S be a bounded, closed, convex subset of $\Re^n$ with non-empty interior, and let $F(x)$ be a convex function that is self concordant on S. Given an initial guess $x_0 \in \operatorname{int} S$, the damped Newton method is defined by the recurrence*

$$x_{i+1} = x_i - \frac{1}{1+\delta(F,x_i)} \nabla^2 F(x_i)^{-1} \nabla F(x_i).$$

Then $x_i \in \operatorname{int} S$ for all i, and

$$F(x_{i+1}) \le F(x_i) - [\delta(F,x_i) - \log(1+\delta(F,x_i))].$$

In particular, if $\delta(F,x_i) \ge 1/4$ then $F(x_i) - F(x_{i+1}) \ge \frac{1}{4} - \log\frac{5}{4} \ge 0.026$. If at some iteration i we have $\delta(F,x_i) \le \frac{1}{4}$ then we are in the region of quadratic convergence of the method, that is, for every $j \ge i$, we have

$$\begin{aligned} \delta(F,x_{j+1}) &\le 2\delta^2(F,x_j) \le \tfrac{1}{2}\delta(F,x_j) \\ F(x_j) - F(x_*) &\le \beta(\delta(F,x_j)) \le \frac{\delta^2(F,x_j)}{2(1-\delta(F,x_j))} \\ \|x_j - x_*\|_{x_*} &\le \frac{\delta(F,x_j)}{1-\delta(F,x_j)}. \end{aligned}$$

The number of Newton steps required to find a point $x \in S$ with $\delta(F,x) \le \kappa < 1$ is bounded by

$$C\left([F(x_0) - F(x_*)] - \log\log\kappa\right)$$

for some constant C.

Proof. See the Exercises. □

Exercises

2.1. Prove that if a one-dimensional convex function satisfies

$$|F'''(x)| \le CF''(x)^{3/2}$$

for some constant C, then the scaled function $\hat{F} = \frac{1}{4}C^2 F$ satisfies

$$|\hat{F}'''(x)| \le C\hat{F}''(x)^{3/2}.$$

2.2. Assume that $\nabla^2 F(x)$ is nonsingular for all $x \in \operatorname{int} S$. Prove that the formula

$$\|h\|_x^2 \equiv h^T \nabla^2 F(x) h$$

defines a norm for all $x \in \operatorname{int} S$.

2.3. Prove Lemma F.1.

2.4. (The next few exercises show how to construct new self-concordant functions from existing ones.) Let $F(x)$ be a self-concordant function on the set $S \subset \Re^n$. Suppose that $x = Ay + b$ where A is an $n \times m$ matrix, b is an n-vector, and $y \in T \subset \Re^m$. Prove that

$$\hat{F}(y) \equiv F(Ay + b)$$

is a self-concordant function on T. What assumptions on the set T are required?

2.5. Let F_i be a self-concordant function on $S_i \subset \Re^n$ for $i = 1, \ldots, m$, and let $\alpha_i \ge 1$ be real numbers. Let S be the intersection of the sets $\{ S_i \}$. Prove that

$$F(x) = \sum_{i=1}^{m} \alpha_i F_i(x)$$

is a self-concordant function on S. What assumptions on the set S are required?

2.6. Let F_i be a self-concordant function on $S_i \subset \Re^{n_i}$ for $i = 1, \ldots, m$. Define

$$S = \{ x = (x_1, \ldots, x_m) : x_i \in S_i \}.$$

Prove that

$$F(x_1, \ldots, x_m) = \sum_{i=1}^{m} F_i(x_i)$$

is a self-concordant function on S. What assumptions on the set S are required?

2.7. In Lemma F.2, prove that the lower bound in (F.1) is valid even in the case where $r = \|h\|_x \ge 1$.

2.8. Prove Lemma F.3 by considering an optimization problem of the form

$$\begin{aligned} \underset{h}{\text{maximize}} \quad & b^T h \\ \text{subject to} \quad & h^T A h - 1 = 0. \end{aligned}$$

for appropriate choices of b and A.

2.9. Complete the proof of Theorem F.6 by proving that

$$q - \log(1+q) \le -\delta - \log(1-\delta)$$

implies $q \le \delta/(1-\delta)$.

2.10. Suppose that the linear programming problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \ge 0 \end{array}$$

is solved by the logarithmic barrier method. Let x be a strictly feasible point, and let μ be the current barrier parameter. Prove that the Newton decrement at x is equal to the 2-norm proximity measure $\delta(x, \mu)$ defined in Section 10.6

2.11. Prove Theorem F.7.

2.12. In Theorem F.7, prove that the sequences

$$\{\, |F(x_j) - F(x_*)| \,\} \quad \text{and} \quad \left\{ \|x_j - x_*\|_{x_*} \right\}$$

are bounded above by sequences that converge quadratically to zero. That is, prove that

$$\begin{aligned} |F(x_j) - F(x_*)| &\le s_j \\ \|x_j - x_*\|_{x_*} &\le t_j \end{aligned}$$

where

$$\lim_{j\to\infty} \frac{s_{j+1}}{s_j^2} < +\infty \quad \text{and} \quad \lim_{j\to\infty} \frac{t_{j+1}}{t_j^2} < +\infty.$$

(Some authors refer to this as "R-quadratic" convergence.)

F.3 The Path-Following Method

In the previous Section we analyzed the behavior of Newton's method when applied to a single barrier subproblem. We now consider the overall interior-point method based on solving a sequence of subproblems. We will require an additional assumption, that is, a bound on the first-derivative of the barrier function.

Let S be a set with the same properties as in the previous Section. A self-concordant function F on S is a *self-concordant barrier function* for S if, for some constant $\nu > 0$,

$$|\nabla F(x)^T h| \le \nu^{1/2} \, \|h\|_x$$

for all $x \in \operatorname{int} S$ and all $h \in \Re^n$. Because of Lemma F.3, the Newton direction for F at a point x must satisfy

$$\|p_N\|_x \le \nu^{1/2}$$

for all $x \in \operatorname{int} S$. (See the Exercises.)

The one-dimensional function $F(x) = -\log x$ is a self-concordant barrier function with $\nu = 1$ for $S = \{\, x : x \ge 0 \,\}$; the n-dimensional function $F(x) =$

$-\sum_{i=1}^{n} \log(x_i)$ is a self-concordant barrier function with $\nu = n$ for $S = \{ x : x \ge 0 \}$; the function $F(x) = -\sum_{i=1}^{m} \log(a_i^T x - b_i)$ is a self-concordant barrier function with $\nu = m$ for the set $S = \left\{ x : a_i^T x \ge b_i,\ i = 1, \ldots, m \right\}$. A self-concordant barrier function exists for any convex set of the form $S = \{ x : g_i(x) \ge 0 \}$; however, evaluating such a barrier function may not be computationally practical.

We will assume that $\nu \ge 1$ throughout this Section. If the above inequality were satisfied for some $\nu < 1$ then it would also be satisfied for $\nu = 1$, so this assumption is not serious. (In fact, there is further justification for making this assumption. It is possible to prove that, if F is a self-concordant barrier function, and $F(x)$ is not constant, then $\nu \ge 1$.)

We require the following technical lemma.

Lemma F.8. *(Semiboundedness) If F is a self-concordant barrier function with parameter $\nu \ge 1$ then*

$$\nabla F(x)^T (y - x) \le \nu$$

for any $x \in \operatorname{int} S$ and $y \in S$.

Proof. See the Exercises. □

The path-following method will be applied to a convex program that has the following *standard form*:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x \in S \end{array} \qquad (\mathcal{P})$$

where $c \ne 0$. This standard form requires that the objective function be a linear function, but otherwise it is unremarkable.

If our optimization problem is written in the form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \ge 0 \end{array}$$

we can easily transform it to standard form. To do this, we introduce a new variable x_{n+1} and consider the problem

$$\begin{array}{ll} \text{minimize} & x_{n+1} \\ \text{subject to} & g(x) \ge 0 \\ & x_{n+1} - f(x) \ge 0. \end{array}$$

It is easy to verify that these two problems are equivalent. The latter problem is in standard form, with $c = (0, \ldots, 0, 1)^T$.

The convex programming problem will be solved using a path-following method of the following form. For $\rho > 0$ we define

$$F_\rho(x) = \rho c^T x + F(x)$$

where F is a self-concordant barrier function for the set S with parameter $\nu \ge 1$, and where the Hessian of F is nonsingular for all $x \in \operatorname{int} S$. Let $x_*(\rho)$ be the minimizer of $F_\rho(x)$ for $x \in \operatorname{int} S$. Our method will generate $x_i \approx x_*(\rho_i)$ for $\rho_i \to +\infty$. The set of minimizers $x_*(\rho)$ forms a path through the feasible region that terminates at the solution x_* to the optimization problem. Note that $F_\rho(x)$ is self-concordant on S. (See the Exrcises.)

Elsewhere in the book, our barrier problems have a parameter multiplying the barrier term that goes to zero. Here, to simplify the exposition, the parameter multiplies the objective and goes to infinity. This just corresponds to multiplying the barrier function by a constant:

$$\rho c^T x + F(x) = \rho(c^T x + \tfrac{1}{\rho} F(x)) = \rho(c^T x + \mu F(x))$$

where $\mu = 1/\rho$. Thus, regardless of which form of barrier function we use, the set of minimizers is the same.

To specify the method, we must specify four things:

- a policy for updating the penalty parameter ρ,
- the optimization method used to minimize F_ρ for a fixed ρ,
- the stopping criteria for the optimization method on a single subproblem,
- the stopping criterion for the path-following method.

These are specified as follows:

- updating ρ: we fix $\gamma > 0$ (the *update parameter*) and set
$$\rho_{i+1} = \left(1 + \frac{\gamma}{\sqrt{\nu}}\right) \rho_i.$$
- the stopping criteria for a subproblem: this is satisfied when the point (ρ_i, x_i) is *close to the path* $x_*(\rho)$, that is, when it satisfies the proximity condition
$$\mathcal{C}_\kappa(\rho, x) : \{\, x \in \operatorname{int} S \,\} \;\&\; \{\, \delta(F_\rho, x) \le \kappa \,\}$$
for some $\kappa \ge 0$.
- optimization method: we update $x_i \to x_{i+1}$ using the damped Newton method:
$$y_{\ell+1} = y_\ell - \frac{1}{1 + \delta(F_{\rho_{i+1}}, y_\ell)} \nabla^2 F(y_\ell)^{-1} \nabla F_{\rho_{i+1}}(y_\ell)$$
with $y_0 = x_i$. This iteration is repeated until (ρ_{i+1}, y_ℓ) satisfies $\mathcal{C}_\kappa(\cdot, \cdot)$, and then we set $x_{i+1} = y_\ell$. (Note that $\nabla^2 F_\rho = \nabla^2 F$ since $F - F_\rho$ is a linear function.)
- the stopping criterion for the path-following method: given some tolerance $\epsilon > 0$, we will terminate the algorithm when we have found an x satisfying $c^T x - z_* \le \epsilon$, where $z_* = c^T x_*$). Since it is unlikely that z_* will be known in advance, we will guarantee this indirectly by ensuring that a bound on $c^T x - z_*$ is at most ϵ. The precise bound is specified in Theorem 17.12.

The only detail left is the initialization, that is, finding an initial pair (ρ_0, x_0) that satisfies the proximity condition. Such a pair can be found using a Phase I procedure (see Section 5.4.1), applying the path-following method to a preliminary optimization problem. (Details are outlined in the Exercises.)

F.3.1 Convergence and Complexity

Our main theorem is the consequence of the following three lemmas. The first lemma determines how rapidly the overall method converges, where each "iteration" corresponds to the approximate solution of a barrier subproblem. It can be used to determine an upper bound on how many barrier subproblems must be solved in order to determine z_* to within some tolerance.

Lemma F.9. *(rate of convergence) Let z_* be the optimal value of $\mathcal{P}$. If a pair (ρ, x) satisfies the proximity condition $\mathcal{C}_\kappa(\cdot,\cdot)$ with $\kappa < 1$ then*

$$c^T x - z_* \le \frac{\chi}{\rho}, \quad \textit{where } \chi = \nu + \frac{\kappa}{1-\kappa}\sqrt{\nu}.$$

In particular, for the path-following method, we have

$$c^T x_i - z_* \le \frac{\chi}{\rho_0}\left[1 + \frac{\gamma}{\sqrt{\nu}}\right]^{-i} \le \frac{\chi}{\rho_0}\exp\left\{-C_\gamma \frac{i}{\sqrt{\nu}}\right\}$$

with positive constant C_γ depending only on γ.

Proof. Let $x_*(\rho)$ be the minimizer of F_ρ. The proof is in two parts. In the first part we determine a bound on $c^T x_*(\rho) - z_*$. If the path-following method solved each barrier subproblem exactly, then this would be sufficient. In the second part, we determine a bound on $c^T x - c^T x_*(\rho)$ corresponding to the approximate solution of a barrier subproblem.

Part I: We will first derive the bound

$$c^T x_*(\rho) - z_* \le \frac{\nu}{\rho}; \tag{F.4}$$

that is, on the trajectory the "error" in the function value is proportional to $1/\rho$. Let x_* be the solution of $\mathcal{P}$. Then from the optimality conditions we have

$$\nabla F_\rho(x_*(\rho)) = \rho c + \nabla F(x_*(\rho)) = 0.$$

Hence

$$\rho(c^T x_*(\rho) - z_*) = \rho(c^T x_*(t) - c^T x_*) = \nabla F(x_*(\rho))^T (x_* - x_*(\rho)).$$

The semi-boundedness property (Lemma 17.8) now gives

$$\nabla F(x_*(\rho))^T (x_* - x_*(\rho)) \le \nu,$$

and (F.4) follows immediately. This completes the first part of the proof.

Part 2: We now prove that $c^T x - z_* \le \chi/\rho$ by deriving a bound on $c^T x - c^T x_*(\rho)$. By the proximity condition and the assumptions of the lemma, $\delta(F_\rho, x) \le \kappa < 1$. From Theorem F.6 we obtain

$$\|x - x_*(\rho)\|_{x_*(\rho)} \le \frac{\kappa}{1-\kappa}.$$

Thus, using Lemma F.3 and the fact that $\rho c = -\nabla F(x_*(\rho))$,

$$\begin{aligned}\rho(c^T x - c^T x_*(\rho)) &= \nabla F(x_*(\rho))^T(x_*(\rho) - x)\\ &\le \|x_*(\rho) - x\|_{x_*(\rho)} \sup\left\{\nabla F(x_*(\rho))^T h : \|h\|_{x_*(\rho)} \le 1\right\}\\ &= \|x_*(\rho) - x\|_{x_*(\rho)}\, \delta(F, x_*(\rho)).\end{aligned}$$

Since F is a self-concordant barrier function with parameter ν, we have

$$\delta(F, x_*(\rho)) \le \sqrt{\nu}.$$

Thus, combining these results we obtain

$$|c^T x - c^T x_*(\rho)| \le \frac{\kappa}{\rho(1-\kappa)}\sqrt{\nu}, \tag{F.5}$$

which, combined with (F.4), gives $c^T x - z_* \le \chi/\rho$. □

The next lemma analyzes what happens when the approximate solution of one barrier subproblem is used as the initial guess for the next subproblem. It determines how "close" the initial guess is to a solution of the new subproblem.

Lemma F.10. *Let ρ and r be two values of the penalty parameter, and let (ρ, x) satisfy the proximity condition $\mathcal{C}_\kappa(\cdot,\cdot)$ for some $\kappa < 1$. Then*

$$F_r(x) - \min_u F_r(u) \le \beta(\kappa) + \frac{\kappa}{1-\kappa}\left|1 - \frac{r}{\rho}\right|\sqrt{\nu} + \nu\beta(1 - r/\rho)$$

where, as before,

$$\beta(s) = -\log(1-s) - s.$$

Proof. The proof is in two parts. In the first part, a bound is obtained for

$$F_r(x_*(\rho)) - F_r(x_*(r)).$$

If the path-following method solved each barrier subproblem exactly, then this would be sufficient. In the second part, this bound is used in the formula

$$F_r(x) - F_r(x_*(r)) = [F_r(x) - F_r(x_*(\rho))] + [F_r(x_*(\rho)) - F_r(x_*(r))]$$

to complete the proof of the lemma.

Part 1: The path of minimizers $x_*(\rho)$ satisfies the equation

$$\rho c + \nabla F(x) = 0. \tag{F.6}$$

Since $\nabla^2 F$ is assumed to be nonsingular in $\operatorname{int} S$, we can use the Implicit Function Theorem (see Section 9 of Appendix B) to show that $x_*(\rho)$ is continuously differentiable. Its derivative can be found by differentiating (F.6) as a function of t (see the Exercises):

$$x_*'(\rho) = -\nabla^2 F(x_*(\rho))^{-1} c.$$

We define

$$\begin{aligned}\phi(r) &= F_r(x_*(\rho)) - F_r(x_*(r)) \\ &= [rc^T x_*(\rho) + F(x_*(\rho))] - [rc^T x_*(r) + F(x_*(r))].\end{aligned}$$

Then, using (F.6)

$$\begin{aligned}\phi'(r) &= c^T x_*(\rho) - c^T x_*(r) - [rc + \nabla F(x_*(r))]^T x_*'(r) \\ &= c^T x_*(\rho) - c^T x_*(r).\end{aligned}$$

Hence

$$\phi(\rho) = \phi'(\rho) = 0.$$

In addition, $\phi'(r)$ is continuously differentiable, with

$$\phi''(r) = -c^T x_*'(r) = c^T \nabla^2 F(x_*(r))^{-1} c.$$

Then (F.6) implies that

$$\begin{aligned}0 \le \phi''(r) &= \frac{1}{r^2} \nabla F(x_*(r))^T \nabla^2 F(x_*(r))^{-1} \nabla F(x_*(r)) \\ &= \frac{1}{r^2} \delta(F, x_*(r))^2 \le \frac{\nu}{r^2}.\end{aligned}$$

If we integrate $\phi''(r)$ twice:

$$\int_\rho^r \int_\rho^s \phi''(y)\, dy\, ds,$$

and use $\phi(\rho) = \phi'(\rho) = 0$ and the bound on $\phi''(r)$, we obtain

$$F_r(x_*(\rho)) - F_r(x_*(r)) = \phi(r) \le \nu\beta \left(1 - \frac{r}{\rho}\right).$$

This completes the first part of the proof.

Part 2: We are now ready to complete the proof of the lemma:

$$\begin{aligned}F_r(x) - \min_u F_r(u) &= F_r(x) - F_r(x_*(r)) \\ &= [F_r(x) - F_r(x_*(\rho))] + [F_r(x_*(\rho)) - F_r(x_*(r)) \\ &= [F_r(x) - F_r(x_*(\rho))] + \phi(r) \\ &= [F_\rho(x) + (\rho - r)c^T x - F_\rho(x_*(\rho)) - (\rho - r)c^T x_*(\rho)] + \phi(r) \\ &= [F_\rho(x) - F_\rho(x_*(\rho))] + (\rho - r)c^T (x - x_*(\rho)) + \phi(r).\end{aligned}$$

Using Lemma F.4 and the inequality $\delta(F_\rho, x) \le \kappa < 1$ we obtain

$$F_\rho(x) - F_\rho(x_*(\rho)) = F_\rho(x) - \min_u F_\rho(u) \le \beta(\delta(F_\rho, x)) \le \beta(\kappa).$$

Combining these results with (F.5) and the bound on $\phi(r)$ from Part 1, gives the desired result. □

The next lemma determines how many Newton iterations are required to obtain an approximate solution to a single barrier subproblem. This result is closely related to Theorem F.7.

Lemma F.11. *(Complexity of a step) The damped Newton iteration is well-defined, that is, it keeps iterates in* $\operatorname{int} S$ *and terminates after finitely many steps. The number of Newton steps until termination does not exceed a certain constant* $N_{\kappa,\gamma}$ *that depends only on* κ *and* γ*.*

Proof. A number of the conclusions follow from Theorem F.7, namely, the damped Newton method keeps the iterates y_ℓ in $\operatorname{int} S$, and ensures that the stopping criteria are satisfied after a finite number of steps. We must prove that the number of Newton steps is bounded by a constant that depends only on κ and γ.

Applying Lemma 17.10 with $x = x_i$, $\rho = \rho_i$, and $r = \rho_{i+1}$ we obtain

$$F_{\rho_{i+1}}(x_i) - \min_u F_{\rho_{i+1}}(u) \le \beta(\kappa) + \frac{\kappa\gamma}{1-\kappa} + \nu\beta\left(-\frac{\gamma}{\sqrt{\nu}}\right).$$

The first two terms only involve κ and γ. We only need worry about the third term:

$$\begin{aligned}\nu\beta\left(-\frac{\gamma}{\sqrt{\nu}}\right) &= \nu[\log(1+\gamma/\sqrt{\nu}) + \gamma/\sqrt{\nu}\,] \\ &\le \nu\gamma/\sqrt{\nu} = \sqrt{\nu}\,\gamma,\end{aligned}$$

since $\nu \ge 1$. Thus, for fixed ν,

$$F_{\rho_{i+1}}(x_i) - \min_u F_{\rho_{i+1}}(u)$$

is bounded in terms of κ and γ, as desired. The bound on the number of Newton steps now follows from Theorem F.7. ∎

The following theorem summarizes the above results. It is the result that we have been working towards. It shows that an interior-point method applied to a convex programming problem can determine the solution to within a specified tolerance in polynomial time.

Theorem F.12. *Suppose that we solve the problem* $\mathcal{P}$ *on a bounded, closed, convex domain* S *using the path-following method associated with a self-concordant barrier function* F *with parameter* $\nu \ge 1$*. Let* $0 < \kappa < 1$ *and* $\gamma > 0$ *be the path tolerance and update parameter, respectively, and assume that* (ρ_0, x_0) *satisfies the proximity condition* $\mathcal{C}_\kappa(\cdot,\cdot)$*. Then*

$$c^T x_i - z_* \le \frac{2\nu}{\rho_0}\left(1 + \frac{\gamma}{\sqrt{\nu}}\right)^{-i}, \quad \text{for } i = 1, 2, \ldots$$

The number of Newton steps required for each iteration $(\rho_i, x_i) \to (\rho_{i+1}, x_{i+1})$ *does not exceed a constant* $N_{\kappa,\gamma}$ *depending only on* κ *and* γ*. In particular, the total* *number of Newton steps required to find an x satisfying $c^T x - z_* \le \epsilon$ is bounded above by*

$$C_{\kappa,\gamma}\sqrt{\nu}\log\left(\frac{2\nu}{\rho_0\epsilon}\right),$$

with constant $C_{\kappa,\gamma}$ depending only on κ and γ.

Proof. The theorem follows from the three previous lemmas. See the Exercises. □

To conclude, we apply these theoretical results to the linear program

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \ge 0. \end{array}$$

The function $F(x) = -\sum_{i=1}^n \log(x_i)$ is a self-concordant barrier function with $\nu = n$ for the set $S = \{\, x : Ax = b,\ x \ge 0 \,\}$. The path-following algorithm in this section is similar to the primal path-following method in Section 10.6. (There are some minor differences. For example, we must project the Newton direction for $F_\rho = \rho c^T x + F(x)$ onto the set $\{\, x : Ax = b \,\}$; this operation affects only the cost of an iteration, not the number of iterations.)

Let L be the length of the input data. Setting $\epsilon = 2^{-2L}$ is sufficient to guarantee that a solution is found. If we assume that the initial penalty parameter ρ is $2^{-O(L)}$, then it follows from Theorem 17.13 that the total number of Newton steps required to solve the problem is $O(\sqrt{n}L)$. This is consistent with the complexity bound derived in Section 10.6.

Exercises

3.1. (The next few exercises show how to construct new self-concordant barrier functions from existing ones.) Let $F(x)$ be a self-concordant barrier function with parameter ν on the set $S \subset \Re^n$. Suppose that $x = Ay + b$ where A is an $n \times m$ matrix, b is an n-vector, and $y \in T \subset \Re^m$. Prove that

$$\hat{F}(y) \equiv F(Ay + b)$$

is a self-concordant barrier function on T with parameter ν. What assumptions on the set T are required?

3.2. Let F_i be a self-concordant barrier function on $S_i \subset \Re^n$ with parameter ν_i for $i = 1, \ldots, m$, and let $\alpha_i \ge 1$ be real numbers. Let S be the intersection of the sets $\{\, S_i \,\}$. Prove that

$$F(x) = \sum_{i=1}^m \alpha_i F_i(x)$$

is a self-concordant barrier function on S with parameter $\sum \alpha_i \nu_i$. What assumptions on the set S are required?

3.3. Let F_i be a self-concordant barrier function on $S_i \subset \Re^{n_i}$ with parameter ν_i for $i = 1, \ldots, m$. Define

$$S = \{\, x = (x_1, \ldots, x_m) : x_i \in S_i \,\}.$$

Prove that

$$F(x_1, \ldots, x_m) = \sum_{i=1}^{m} F_i(x_i)$$

is a self-concordant barrier function on S with parameter $\sum \nu_i$. What assumptions on the set S are required?

3.4. Let F be a self-concordant barrier function on S. Use Lemma F.3 to prove that the Newton direction for F at a point x must satisfy $\|p_N\|_x \le \nu^{1/2}$ for all $x \in \operatorname{int} S$.

3.5. Prove that if $F(x)$ is a self-concordant barrier function on S, then $F_\rho(x)$ is self-concordant on S.

3.6. The goal of this problem is to prove Lemma 17.8. This is accomplished in several stages. If $\nabla F(x)^T(y-x) \le 0$ then the result is obvious. Thus we assume that $\nabla F(x)^T(y-x) > 0$.

(i) Let T be such that

$$\begin{cases} x + t(y-x) \in S, & \text{for } 0 \le t \le T; \\ x + t(y-x) \notin S, & \text{for } t > T. \end{cases}$$

Prove that $T \ge 1$.

(ii) Define $\phi(t) \equiv F(x + t(y-x))$. Prove that

$$|\phi'(t)| \le \sqrt{\nu \phi''(t)}.$$

(iii) Define $\psi(t) \equiv \phi'(t)$. Prove that $\psi(t) > 0$ and that

$$(-\psi^{-1}(t))' = \psi'(t)\psi^{-2}(t) \ge \nu^{-1}.$$

(iv) Use

$$-\psi^{-1}(t) = -\psi(0) + \int_0^\tau (-\psi^{-1}(t))' dt$$

to prove that

$$\psi(t) \ge \frac{\nu \psi(0)}{\nu - t\psi(0)}.$$

(v) Use the fact that $\psi(t)$ is bounded on any sub-interval $[0, \bar{T}]$, for $0 < \bar{T} < T$, to conclude that

$$\nu - \bar{T}\psi(0) > 0 \quad \text{for all } \bar{T} < T.$$

Use this to complete the proof of the lemma.

3.7. Use the Implicit Function Theorem (see Section 9 of Appendix B) to prove that (under the assumptions on the optimization problem made in this Section) the path of minimizers $x_*(\rho)$ is continuously differentiable and its derivative is

$$x'_*(\rho) = -\nabla^2 F(x_*(\rho))^{-1} c.$$

Hint: Differentiate (F.6) as a function of ρ.

3.8. The goal of this problem is to define a Phase-1 procedure to find an initial pair (ρ_0, x_0) for the path-following method. Suppose that some initial guess $\bar{x} \in \operatorname{int} S$ has been specified.

(i) Prove that $\bar{x}$ minimizes the artificial barrier function

$$\bar{F}_\rho(x) \equiv \rho d^T x + F(x)$$

for $\rho = 1$ and $d = -\nabla F(\bar{x})$. Hence $(1, \bar{x})$ can be used as an initial pair for a path-following method based on the artificial barrier function.

(ii) Prove that an initial point for the path-following method for the *original* problem can be obtained by running the method in (a) *in reverse*, that is, for $\rho \to 0$.

3.9. Complete the proof of Theorem 17.12.

3.10. (The next few exercises describe a class of convex programs call semi-definite programming problems.) Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & A(x) \geq 0, \end{array}$$

where

$$A(x) = A_0 + \sum_{i=1}^{m} x_i A_i$$

and where $A_0, A_1, \ldots, A_m$ are $n \times n$ symmetric matrices. The inequality $A(x) \geq 0$ should be interpreted to mean that the matrix $A(x)$ is positive semi-definite. Prove that this problem is a convex programming problem. We refer to this problem as a *semi-definite programming problem*.

3.11. Show that any linear program can be written as a semi-definite programming problem. Hint: Use diagonal matrices $\{ A_i \}$.

3.12. Let

$$A(x) = A_0 + \sum_{i=1}^{m} x_i A_i.$$

Show that the problem of minimizing the maximum eigenvalue of $A(x)$ can be rewritten as a semi-definite program.

3.13. Let

$$A(x) = A_0 + \sum_{i=1}^{m} x_i A_i.$$

Prove that the function $F(x) \equiv -\log\det A(x)$ is a self-concordant barrier function for $S = \{\, x : A(x) \geq 0 \,\}$. (Thus $F(x)$ can be used to definite a path-following method for a semi-definite programming problem.)

F.4 Notes

The results in this Appendix are adapted from the book by Nesterov and Nemirovskii (1993) and from the lecture notes of Nemirovskii (1994).