Interior-Point Theory for Convex Optimization

Robert M. Freund

May, 2014

©2014 Massachusetts Institute of Technology. All rights reserved.
1 Background

The material presented herein is based on the following two research texts:

*Interior-Point Polynomial Algorithms in Convex Programming* by Yurii Nesterov and Arkadii Nemirovskii, SIAM 1994, and


2 Barrier Scheme for Solving Convex Optimization

Our problem of interest is

\[ P : \ \text{minimize}_{x} \ c^T x \]

\[ \text{s.t.} \quad x \in S , \]

where \( S \) is some closed convex set, and denote the optimal objective value by \( V^* \). Let \( f(\cdot) \) be a *barrier function* for \( S \), namely \( f(\cdot) \) satisfies:

(a) \( f(\cdot) \) is strictly convex on its domain \( D_f := \text{int}S \), and

(b) \( f(x) \to \infty \) as \( x \to \partial S \).

The idea of the barrier method is to dissuade the algorithm from computing points too close to \( \partial S \), effectively eliminating the complicating factors of dealing with \( \partial S \). For every value of \( \mu > 0 \) we create the barrier problem:

\[ P_\mu : \ \text{minimize}_{x} \ \mu c^T x + f(x) \]

\[ \text{s.t.} \quad x \in D_f . \]

Note that \( P_\mu \) is effectively unconstrained, since the boundary of the feasible region will never be encountered. The solution of \( P_\mu \) is denoted \( z(\mu) \):
\[ z(\mu) := \arg \min_{x} \{ \mu c^T x + f(x) : x \in D_f \} . \]

Intuitively, as \( \mu \to \infty \), the impact of the barrier function on the solution of \( P_\mu \) should become less and less, so we should have \( c^T z(\mu) \to V^* \) as \( \mu \to \infty \). Presuming this is the case, the barrier scheme tries to use Newton’s method to solve for approximate solutions \( x^i \) of \( P_{\mu_i} \) for an increasing sequence of values of \( \mu_i \to \infty \).

In order to be more specific about how the barrier scheme might work, let us assume that at each iteration we have some value \( x \in D_f \) that is an approximate solution of \( P_\mu \) for a given value \( \mu > 0 \). We will, of course, need a way to define “is an approximate solution of \( P_\mu \)” that will be developed later. We then will increase the barrier parameter \( \mu \) by a multiplicative factor \( \alpha > 1 \):

\[ \hat{\mu} \leftarrow \alpha \mu . \]

Then we will take a Newton step at \( x \) for the problem \( P_{\hat{\mu}} \) to obtain a new point \( \hat{x} \) that we would like to then be an approximate solution of \( P_{\hat{\mu}} \). If so, we can continue the scheme inductively.

We typically use \( g(\cdot) \) and \( H(\cdot) \) to denote the gradient and Hessian of \( f(\cdot) \). Note that the Newton iterate for \( P_{\hat{\mu}} \) has the formula:

\[ \hat{x} \leftarrow x - H(x)^{-1}(\hat{\mu} c + g(x)) . \]

The general algorithmic scheme is presented in Algorithm 1.

### 3 Some Plain Facts

Let \( f(\cdot) : \mathbb{R}^n \to \mathbb{R} \) be a twice-differentiable function. We typically use \( g(\cdot) \) and \( H(\cdot) \) denote the gradient and Hessian of \( f(\cdot) \).

Here are four facts about integrals and derivatives:

**Fact 3.1**

\[ g(y) = g(x) + \int_{0}^{1} H(x + t(y - x))(y - x)dt \]
Algorithm 1 General Barrier Scheme

Initialize.

Initialize with $\mu_0 > 0$, $x^0 \in D_f$ that is “an approximate solution of $P_{\mu_0}$,” $i \leftarrow 0$.

Define $\alpha > 1$.

At iteration $i$:

1. Current values.
   
   \[
   \mu \leftarrow \mu_i \quad \quad \quad x \leftarrow x^i
   \]

2. Increase $\mu$ and take Newton step.
   
   \[
   \hat{\mu} \leftarrow \alpha \mu \quad \quad \quad \hat{x} \leftarrow x - H(x)^{-1}(\hat{\mu}c + g(x))
   \]

3. Update values.
   
   \[
   \mu_{i+1} \leftarrow \hat{\mu} \quad \quad \quad x^{i+1} \leftarrow \hat{x}
   \]
**Fact 3.2** Let \( h(t) := f(x + tv) \). Then

(i) \( h'(t) = g(x + tv)^Tv \), and

(ii) \( h''(t) = v^TH(x + tv)v \).

**Fact 3.3**

\[
f(y) = f(x) + g(x)^T(y - x) + \frac{1}{2}(y - x)^TH(x)(y - x) \]
\[
+ \int_0^1 \int_0^t (y - x)^T[H(x + s(y - x)) - H(x)](y - x) \, ds \, dt
\]

**Fact 3.4**

\[
\int_0^r \left( \frac{1}{(1-at)^2} - 1 \right) \, dt = \frac{ar^2}{1-ar}
\]

This follows by observing that \( \int \frac{1}{(1-at)^2} dt = \frac{1}{a(1-at)} \).

We also present five additional facts that we will need in our analyses.

**Fact 3.5** Suppose \( f(\cdot) \) is a convex function on \( \mathbb{R}^n \), and \( S \subset \mathbb{R}^n \) is a compact convex set, and suppose \( x \in \text{int}S \) satisfies \( f(x) \leq f(y) \) for all \( y \in \partial S \). Then \( f(\cdot) \) attains its global minimizer on \( S \).

**Fact 3.6** Let \( \|v\| := \sqrt{v^Tv} \) be the Euclidean norm. Let \( \lambda_1 \leq \ldots \leq \lambda_n \) be the ordered eigenvalues of the symmetric matrix \( M \), and define \( \|M\| := \max\{\|Mx\| : \|x\| \leq 1\} \). Then \( \|M\| = \max_i \{|\lambda_i|\} = \max\{|\lambda_n|, |\lambda_1|\} \).

**Fact 3.7** Suppose \( A, B \) are symmetric and \( A + B = \theta I \) for some \( \theta \in \mathbb{R} \). Then \( AB = BA \). Furthermore, if \( A \succeq 0, B \succeq 0 \), then \( A^\alpha B^\beta = B^\beta A^\alpha \) for all \( \alpha, \beta \geq 0 \).

To see why this is true, decompose \( A = PDPT \) where \( P \) is orthonormal \((P^T = P^{-1})\) and \( D \) is diagonal. Then \( B = P(\theta I - D)P^T \), whereby \( A^\alpha B^\beta = PD^\alpha P^T P(\theta I - D)^\beta P^T = PD^\alpha(\theta I - D)^\beta P^T = P(\theta I - D)^\beta D^\alpha P^T = P(\theta I - D)^\beta P^T PD^\alpha P^T = B^\beta A^\alpha \).
Fact 3.8 Suppose $\lambda_n \geq \ldots \geq \lambda_1 > 0$. Then
\[
\max_i \{|\lambda_i - 1|\} \leq \max \{\lambda_n - 1, 1/\lambda_1 - 1\}.
\]

Fact 3.9 Suppose $a, b, c, d > 0$. Then
\[
\min \left\{ \frac{a}{b}, \frac{c}{d} \right\} \leq \frac{a + c}{b + d} \leq \max \left\{ \frac{a}{b}, \frac{c}{d} \right\}.
\]

4 Self-Concordant Functions and Properties

Let $f(\cdot)$ be a strictly convex twice-differentiable function defined on the open set $D_f := \text{domain } f(\cdot)$ and let $\bar{D}_f := \text{cl } D_f$. Consider $x \in D_f$. We will often abbreviate $H_x := H(x)$ for the Hessian at $x$. Here we assume that $H_x > 0$, whereby $H_x$ can be used to define the norm
\[
\|v\|_x := \sqrt{v^T H_x v}
\]
which is the “local norm” at $x$. Notice that
\[
\|v\|_x = \sqrt{v^T H_x v} = \|H_{1/2}^x v\|,
\]
where $\|w\| = \sqrt{w^T w}$ is the standard Euclidean ($L_2$) norm. Let
\[
B_x(x, 1) := \{y : \|y - x\|_x < 1\}.
\]
This is called the open Dikin ball at $x$ after the Russian mathematician I.I.Dikin.

Definition 4.1 $f(\cdot)$ is said to be (strongly nondegenerate) self-concordant if for all $x \in D_f$ we have $B_x(x, 1) \subset D_f$, and for all $y \in B_x(x, 1)$ we have:
\[
1 - \|y - x\|_x \leq \frac{\|v\|_y}{\|v\|_x} \leq \frac{1}{1 - \|y - x\|_x}
\]
for all $v \neq 0$.

Let $\mathcal{SC}$ denote the class of all such functions.
Remark 1 The following are the most-used self-concordant functions:

(i)  \( f(x) = -\ln(x) \) for \( x \in D_f = \{ x \in \mathbb{R} : x > 0 \} \),

(ii) \( f(X) = -\ln \det(X) \) for \( X \in D_f = \{ X \in S^{k \times k} : X > 0 \} \), and

(iii) \( f(x) = -\ln(x_1^2 - \sum_{j=2}^{n} x_j^2) \) for \( x \in D_f := \{ x : \| (x_2, \ldots, x_n) \| < x_1 \} \).

Before showing that these functions are self-concordant, let us see how we can combine self-concordant functions to obtain other self-concordant functions.

Proposition 4.1 (self-concordance under addition/intersection) Suppose that \( f_i(\cdot) \in SC \) with domain \( D_i := D_{f_i} \) for \( i = 1, 2 \), and suppose that \( D := D_1 \cap D_2 \neq \emptyset \). Define \( f(\cdot) = f_1(\cdot) + f_2(\cdot) \). Then \( D_f = D \) and \( f(\cdot) \in SC \).

Proof: Consider \( x \in D = D_1 \cap D_2 \). Let \( B^k_x(c,r) \) denote the Dikin ball centered at \( c \) with radius \( r \) defined by \( f_i(\cdot) \) and let \( \| \cdot \|_{x,i} \) denote the norm induced at \( x \) using the Hessian \( H_i(x) \) of \( f_i(x) \) for \( i = 1, 2 \). Then since \( x \in D_i \) we have \( B^k_x(c,1) \subset D_i \) and \( \| v \|_x^2 = \| v \|_{x,1}^2 + \| v \|_{x,2}^2 \) because \( H(x) = H_1(x) + H_2(x) \). Therefore if \( \| y - x \|_x < 1 \) it follows that \( \| y - x \|_{x,1} < 1 \) and \( \| y - x \|_{x,2} < 1 \), whereby \( y \in B^k_x(c,1) \subset D_i \) for \( i = 1, 2 \), and hence \( y \in D_1 \cap D_2 = D \). Also, for any \( v \neq 0 \), using Fact 3.9 we have

\[
\frac{\| v \|_x^2}{\| v \|_x^2} = \frac{\| v \|_{x,1}^2 + \| v \|_{x,2}^2}{\| v \|_{x,1}^2 + \| v \|_{x,2}^2} \leq \max \left\{ \frac{\| v \|_{x,1}^2}{\| v \|_{x,2}^2}, \frac{\| v \|_{x,2}^2}{\| v \|_{x,1}^2} \right\}
\]

\[
\leq \max \left\{ \left( \frac{1}{1-\| y-x \|_{x,1}} \right)^2, \left( \frac{1}{1-\| y-x \|_{x,2}} \right)^2 \right\}
\]

\[
\leq \left( \frac{1}{1-\| y-x \|_x} \right)^2.
\]

The virtually identical argument can also be applied to prove the “\( \geq \)” inequality of the definition of self-concordance by replacing “max” by “min” above and applying the other inequality of Fact 3.9. \( \blacksquare \)
Proposition 4.2 (self-concordance under affine transformation) Let $A \in \mathbb{R}^{m \times n}$ satisfy $\text{rank} A = n \leq m$. Suppose that $f(\cdot) \in \mathcal{SC}$ with domain $D_f \subset \mathbb{R}^m$ and define $\hat{f}(\cdot)$ by $\hat{f}(x) = f(Ax - b)$. Then $\hat{f}(\cdot) \in \mathcal{SC}$ with domain $\hat{D} := \{ x : Ax - b \in D_f \}$.

**Proof:** Consider $x \in \hat{D}$ and $s = Ax - b$. Letting $g(s)$ and $H(s)$ denote the gradient and Hessian of $f(s)$ and $\hat{g}(x)$ and $\hat{H}(x)$ the gradient and Hessian of $\hat{f}(x)$, we have $\hat{g}(x) = A^T g(s)$ and $\hat{H}(x) = A^T H(s) A$. Suppose that $\| y - x \|_x < 1$. Then defining $t := Ay - b$ we have $1 > \| y - x \|_x = \sqrt{(AT - x^T AT)H(s)(Ay - Ax) = \| t - s \|_s}$, whereby $t \in D_f$ and so $y \in \hat{D}$. Therefore $B_y(x, 1) \subset \hat{D}$. Also, for any $v \neq 0$, we have

$$\frac{\| v \|_y}{\| v \|_x} = \frac{\sqrt{v^T AT H(t) Av}}{\sqrt{v^T AT H(s) Av}} = \frac{\| Av \|_t}{\| Av \|_s} \leq \frac{1}{1 - \| s - t \|_s} = \frac{1}{1 - \| y - x \|_x}.$$ 

The exact same argument can also be applied to prove the “≥” inequality of the definition of self-concordance.

Proposition 4.3 The three functions defined in Remark 1 are self-concordant.

**Proof:** We will prove that $f(X) := -\ln \det(X)$ is self-concordant on its domain $\{ X \in S^{k \times k} : X \succ 0 \}$. When $k = 1$, this is the logarithmic barrier function. Although it is true, we will not prove that $f(x) = -\ln(x_1^2 - \sum_{j=2}^n x_j^2)$ is a self-concordant barrier for the interior of the second-order cone $Q^n := \{ x : \|(x_2, \ldots, x_n)\| \leq x_1 \}$, as this proof is arithmetically uninspiring.

Before we get started with the proof, we have to expand and amend our notation a bit. If we consider two vectors $x, y$ in the $n$-dimensional space $V$ (which we typically identify with $\mathbb{R}^n$), we use the standard inner product $x^T y$ to define the standard norm $\| v \| := \sqrt{v^T v}$. When we consider the vector space $V$ to be $S^{k \times k}$ (the set of symmetric matrices of order $k$), we need to define the standard inner product of two symmetric $k \times k$ matrices $X$ and $Y$ and also to define the standard norm on $V = S^{k \times k}$. The standard inner product in this case is the trace inner product:

$$X \bullet Y := \text{Tr}(XY) = \sum_{i=1}^n (XY)_{ii} = \sum_{i=1}^k \sum_{j=1}^k X_{ij} Y_{ij}.$$ 

8
The trace inner product has the following properties which are elementary to establish:

1. \( A \bullet BC = AB \bullet C \)

2. \( \text{Tr}(AB) = \text{Tr}(BA) \), whereby \( A \bullet B = B \bullet A \)

The trace inner product is used to define the standard norm:

\[
\|X\| := \sqrt{X \bullet X}.
\]

This norm is also called the Frobenius norm of a matrix \( X \). Notice that the Frobenius norm of \( X \) is not the operator norm of \( X \). For the remainder of this proof we use the trace inner product and the Frobenius norm, and the reader should ignore any mental mention of operator norms in this proof.

To prove \( f(X) := -\ln \det(X) \) is self-concordant, let \( X \succ 0 \) be given, and let \( Y \in B_X(X,1) \) and \( V \in S^{k \times k} \) be given. We need to verify three statements:

1. \( Y \succ 0 \),

2. \( \frac{\|V\|_Y}{\|V\|_X} \leq \frac{1}{1 - \|Y - X\|_X} \), and

3. \( \frac{\|V\|_Y}{\|V\|_X} \geq 1 - \|Y - X\|_X \)

To get started, direct expansion yields the following second-order expansion of \( f(X) \):

\[
f(X + \Delta X) \approx f(X) - X^{-1} \bullet \Delta X + \frac{1}{2} \Delta X \bullet X^{-1} \Delta XX^{-1}
\]

and indeed it is easy to derive:

• \( g(X) = -X^{-1} \) and

• \( H(X) \Delta X = X^{-1} \Delta XX^{-1} \)
It therefore follows that
\[
\|\Delta X\|_X = \sqrt{\text{Tr}(\Delta X X^{-1} \Delta X X^{-1})} \\
= \sqrt{\text{Tr}(X^{-\frac{1}{2}} \Delta X X^{-\frac{1}{2}} X^{-\frac{1}{2}} \Delta X X^{-\frac{1}{2}})} \\
= \sqrt{\text{Tr}([X^{-\frac{1}{2}} \Delta X X^{-\frac{1}{2}}]^2)}.
\] (1)

Now define two auxiliary matrices:

\[ F := X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \quad \text{and} \quad S := X^{-\frac{1}{2}} V X^{-\frac{1}{2}}. \]

Note that
\[
\|S\| = \sqrt{\text{Tr}(X^{-\frac{1}{2}} V X^{-\frac{1}{2}} X^{-\frac{1}{2}} V X^{-\frac{1}{2}})} = \|V\|_X. \] (2)

Furthermore let us write \( F = QDQ^T \) where \( Q \) is orthonormal \( (Q^T = Q^{-1}) \) and the diagonal matrix \( D \) is comprised of the eigenvalues of \( F \), and let \( \lambda \) denote the vector of eigenvalues, with minima and maxima \( \lambda_{\min} \) and \( \lambda_{\max} \).

To prove item (1.) above, we observe:
\[
1 > \|Y - X\|_X^2 = \text{Tr}(X^{-\frac{1}{2}} (Y - X) X^{-\frac{1}{2}} X^{-\frac{1}{2}} (Y - X) X^{-\frac{1}{2}}) \\
= \text{Tr}(F - I)(F - I) \\
= \text{Tr}(Q(D - I)Q^T Q(D - I)Q^T) \\
= \text{Tr}((D - I)(D - I)) \\
= \sum_{j=1}^k (\lambda_j - 1)^2 \\
= \|\lambda - e\|_2^2
\] (3)

where \( e = (1, \ldots, 1) \). Since the last quantity above is less than 1, it follows that \( \lambda > 0 \) and hence \( F > 0 \) and therefore \( Y > 0 \), establishing (1.). In order to establish (2.) and (3.) we will need the following
\[
\|F^{-\frac{1}{2}} S F^{-\frac{1}{2}}\| \leq \frac{1}{\lambda_{\min}} \|S\| \quad \text{and} \quad \|F^{-\frac{1}{2}} S F^{-\frac{1}{2}}\| \geq \frac{1}{\lambda_{\max}} \|S\|. \] (4)
To prove (4), we proceed as follows:

\[ \|F^{-\frac{1}{2}}SF^{-\frac{1}{2}}\| = \sqrt{\text{Tr}(QD^{-\frac{1}{2}}QTD^{-\frac{1}{2}}QTD^{-\frac{1}{2}}QTSD^{-\frac{1}{2}}Q)} \]

\[ = \sqrt{\text{Tr}(D^{-1}QTD^{-1}QTSD)} \]

\[ \leq \frac{1}{\sqrt{\lambda_{\text{min}}}} \sqrt{\text{Tr}(QTD^{-1}QTSQD)} \]

\[ = \frac{1}{\sqrt{\lambda_{\text{min}}}} \sqrt{\text{Tr}(D^{-1}QTDQS)} \]

\[ = \frac{1}{\lambda_{\text{min}}} \sqrt{\text{Tr}(SS)} = \frac{1}{\lambda_{\text{min}}} \|S\| \cdot \]

The other inequality of (4) follows by substituting \(\lambda_{\text{max}}\) for \(\lambda_{\text{min}}\) and switching \(\geq\) for \(\leq\) in the above chain of equalities and inequalities. We now have:

\[ \|V\|_X^2 = \text{Tr}(VY^{-1}VY^{-1}) \]

\[ = \text{Tr}(X^{-\frac{1}{2}}VX^{-\frac{1}{2}}X^{\frac{1}{2}}X^{-\frac{1}{2}}Y^{-1}X^{\frac{1}{2}}X^{-\frac{1}{2}}VX^{-\frac{1}{2}}X^{\frac{1}{2}}Y^{-1}X^{\frac{1}{2}}) \]

\[ = \text{Tr}(SF^{-1}SF^{-1}) \]

\[ = \text{Tr}(F^{-\frac{1}{2}}SF^{-\frac{1}{2}}F^{-\frac{1}{2}}SF^{-\frac{1}{2}}) \]

\[ = \|F^{-\frac{1}{2}}SF^{-\frac{1}{2}}\|^2 \leq \frac{1}{\lambda_{\text{min}}} \|S\|^2 = \frac{1}{\lambda_{\text{min}}} \|V\|_X^2 \]

where the last inequality follows from (4) and the last equality from (2). Therefore

\[ \frac{\|V\|_Y}{\|V\|_X} \leq \frac{1}{\lambda_{\text{min}}} \leq \frac{1}{1 - |1 - \lambda_{\text{max}}|} \leq \frac{1}{1 - \|e - \lambda\|_2} = \frac{1}{1 - \|Y - X\|_X} \]

where the last equality is from (3). This proves (2). To prove (3), use the same equalities as above and the second inequality of (4) to obtain:

\[ \|V\|_Y^2 = \|F^{-\frac{1}{2}}SF^{-\frac{1}{2}}\|^2 \geq \frac{1}{\lambda_{\text{max}}} \|V\|_X^2 \]
and therefore \( \| V \|_Y \geq \frac{1}{\lambda_{\max}} \). If \( \lambda_{\max} \leq 1 \) it follows directly that \( \| V \|_X \geq \frac{1}{\lambda_{\max}} \geq 1 \geq 1 - \| Y - X \|_X \), while if \( \lambda_{\max} > 1 \) we have:

\[
\| Y - X \|_X = \| \lambda - e \|_2 \geq \lambda_{\max} - 1 ,
\]

from which it follows that

\[
\lambda_{\max} \| Y - X \|_X \geq \| Y - X \|_X \geq \lambda_{\max} - 1
\]

and so

\[
\lambda_{\max} (1 - \| Y - X \|_X) \leq 1 .
\]

From this it then follows that \( \| V \|_Y \| V \|_X \geq \| Y - X \|_X \geq \lambda_{\max} - 1 \) and so

\[
\lambda_{\max} (1 - \| Y - X \|_X) \leq 1 .
\]

Our next result is rather technical, as it shows further properties of changes in Hessian matrices under self-concordance. Recall from Fact 3.6 that the operator norm of a matrix \( M \) is defined using the Euclidean norm, namely

\[
\| M \| := \max \{ \| Mx \| : \| x \| \leq 1 \} .
\]

**Lemma 4.1** Suppose that \( f(\cdot) \in SC \) and \( x \in Df \). If \( \| y - x \| < 1 \), then

1. \( \| H_x^{-\frac{1}{2}} H_y H_x^{-\frac{1}{2}} \| \leq \left( \frac{1}{1 - \| y - x \|} \right)^2 \),
2. \( \| H_x^\frac{1}{2} H_y^{-1} H_x^\frac{1}{2} \| \leq \left( \frac{1}{1 - \| y - x \|} \right)^2 \),
3. \( \| I - H_x^{-\frac{1}{2}} H_y H_x^{-\frac{1}{2}} \| \leq \left( \frac{1}{1 - \| y - x \|} \right)^2 - 1 \), and
4. \( \| I - H_x^\frac{1}{2} H_y^{-1} H_x^\frac{1}{2} \| \leq \left( \frac{1}{1 - \| y - x \|} \right)^2 - 1 \).

**Proof:** Let \( Q := H_x^{-\frac{1}{2}} H_y H_x^{-\frac{1}{2}} \), and observe that \( Q \succ 0 \) with eigenvalues \( \lambda_n \geq \ldots \geq \lambda_1 > 0 \). From Fact 3.6 we have

\[
\sqrt{\| Q \|} = \sqrt{\lambda_n} = \max_w \sqrt{w^T Q w} = \max_v \sqrt{v^T H_y v} \leq \max_v \| v \|_y \leq \frac{1}{1 - \| y - x \|}.
\]
(where the third equality uses the substitution $v = H_x^{-\frac{1}{2}}w$) and squaring yields the first assertion. Similarly, we have

$$\frac{1}{\sqrt{\|Q^{-1}\|}} = \sqrt{\lambda_1} = \min_w \sqrt{\frac{w^TQw}{w^Tw}} = \min_v \sqrt{\frac{v^TH_xv}{v^Tv}} = \min_v \|v\|_y \geq 1 - \|y-x\|_x$$

(where the third equality again uses the substitution $v = H_x^{-\frac{1}{2}}w$) and squaring and rearranging yields the second assertion. Next observe

$$\|I - Q\| = \max_i \{|\lambda_i - 1|\} \leq \max\{\lambda_n - 1, 1/\lambda_1 - 1\} \leq \left(\frac{1}{1 - \|y - x\|_x}\right)^2 - 1$$

where the first inequality is from Fact 3.8 and the second inequality follows from the two equation streams above, thus showing the third assertion of the lemma. Finally, we have

$$\|I - Q^{-1}\| = \max_i \{1/\lambda_i - 1\} \leq \max\{1/\lambda_1 - 1, \lambda_n - 1\} \leq \left(\frac{1}{1 - \|y - x\|_x}\right)^2 - 1$$

where the first inequality is from Fact 3.8 and the second inequality follows from the two equation streams above, thus showing the fourth assertion of the lemma.

The next result states that we can bound the error in the quadratic approximation of $f(\cdot)$ inside the Dikin ball $B_x(x, 1)$.

**Proposition 4.4** Suppose that $f(\cdot) \in \mathcal{SC}$ and $x \in D_f$. If $\|y - x\|_x < 1$, then

$$\left| f(y) - \left[ f(x) + g(x)^T(y - x) + \frac{1}{2}(y - x)^T H_x(y - x) \right] \right| \leq \frac{\|y - x\|_x^3}{3(1 - \|y - x\|_x)}.$$
Proof: Let $L$ denote the left-hand side of the inequality to be proved. From Fact 3.3 we have

\[
L = \left| \int_0^1 \int_0^t (y - x)^T \left[ H(x + s(y - x)) - H(x) \right] (y - x) \, ds \, dt \right|
\]

\[
= \left| \int_0^1 \int_0^t (y - x)^T H_x^{1/2} \left[ H_x^{1/2} H(x + s(y - x)) H_x^{1/2} - I \right] H_x^{1/2} (y - x) \, ds \, dt \right|
\]

\[
\leq \|y - x\|_x^2 \left| \int_0^1 \int_0^t \left( \frac{1}{1 - s \|y - x\|_x} \right)^2 - 1 \, ds \, dt \right| \tag{from Lemma 4.1}
\]

\[
= \|y - x\|_x^2 \int_0^1 \frac{\|y - x\|_x t^2}{1 - t \|y - x\|_x} \, dt \tag{from Fact 3.4}
\]

\[
\leq \frac{\|y - x\|_x^3}{1 - \|y - x\|_x} \int_0^1 t^2 \, dt = \frac{\|y - x\|_x^3}{3(1 - \|y - x\|_x)} .
\]

Recall Newton’s method to minimize $f(\cdot)$. At $x \in D_f$ we compute the Newton step:

\[
n(x) := -H(x)^{-1} g(x)
\]

and compute the Newton iterate:

\[
x_+ := x + n(x) = x - H(x)^{-1} g(x) .
\]

When $f(\cdot) \in SC$, Newton’s method has some very wonderful properties as we now show.

**Theorem 4.1** Suppose that $f(\cdot) \in SC$ and $x \in D_f$. If $\|n(x)\|_x < 1$, then

\[
\|n(x_+)\|_{x+} \leq \left( \frac{\|n(x)\|_x}{1 - \|n(x)\|_x} \right)^2 .
\]
Proof: We will prove this by proving the following two results which together establish the result:

(a) \[ \|n(x_+)\|_x \leq \frac{\|H_x^{-\frac{1}{2}}g(x_+)\|}{1 - \|n(x)_x\|}, \]

(b) \[ \|H_x^{-\frac{1}{2}}g(x_+)\| \leq \frac{\|n(x)_x\|^2}{1 - \|n(x)_x\|}. \]

First we prove (a):

\[ \|n(x_+)\|_x^2 = g(x_+)^TH^{-1}(x_+)H(x_+)H^{-1}(x_+)g(x_+) \]

\[ = g(x_+)^TH_x^{-\frac{1}{2}}H_x^{\frac{1}{2}}H^{-1}(x_+)H_x^{\frac{1}{2}}H_x^{-\frac{1}{2}}g(x_+) \]

\[ \leq \|H_x^{\frac{1}{2}}H^{-1}(x_+)H_x^{\frac{1}{2}}\| \|H_x^{-\frac{1}{2}}g(x_+)\|^2 \]

\[ \leq \left( \frac{1}{1 - \|n(x)_x\|} \right)^2 \|H_x^{-\frac{1}{2}}g(x_+)\|^2 \] (from Lemma 4.1)

\[ = \left( \frac{1}{1 - \|n(x)_x\|} \right)^2 \|H_x^{-\frac{1}{2}}g(x_+)\|^2, \]

which proves (a). To prove (b), observe first that

\[ g(x_+) = g(x_+) - g(x) + g(x) \]

\[ = g(x_+) - g(x) - H_xn(x) \]

\[ = \int_0^1 H(x + t(x_+ - x))(x_+ - x)dt - H_xn(x) \] (from Fact 3.1)

\[ = \int_0^1 [H(x + tn(x)) - H_x]n(x)dt \]

\[ = \int_0^1 [H(x + tn(x)) - H_x]H_x^{-\frac{1}{2}}H_x^{\frac{1}{2}}n(x)dt . \]

Therefore

\[ H_x^{-\frac{1}{2}}g(x_+) = \int_0^1 \left[ H_x^{-\frac{1}{2}}H(x + tn(x))H_x^{-\frac{1}{2}} - I \right]H_x^{\frac{1}{2}}n(x)dt \]
which then implies
\[ \|H_x^{-\frac{1}{2}} g(x_+)\| \leq \int_0^1 \|H_x^{-\frac{1}{2}} H(x + tn(x))H_x^{-\frac{1}{2}} - I\| \|H_x^{\frac{1}{2}} n(x)\| dt \]
\[ \leq \|H_x^{\frac{1}{2}} n(x)\| \int_0^1 \left( \frac{1}{1-t\|n(x)\|_x} \right)^2 - 1 \, dt \quad \text{(from Lemma 4.1)} \]
\[ = \|n(x)\|_x \|n(x)\|_x \quad \text{(from Fact 3.4)} \]
which proves (b).

**Theorem 4.2** Suppose that \( f(\cdot) \in SC \) and \( x \in D_f \). If \( \|n(x)\|_x \leq \frac{1}{4} \), then \( f(\cdot) \) has a minimizer \( z \), and
\[ \|z - x_+\|_x \leq \frac{3\|n(x)\|_x^2}{(1 - \|n(x)\|_x^3}. \]

**Proof:** First suppose that \( \|n(x)\|_x \leq 1/9 \), and define \( Q_x(y) := f(x) + g(x)^T(y - x) + \frac{1}{2}(y - x)^T H_x(y - x) \). Let \( y \) satisfy \( \|y - x\|_x \leq 1/3 \). Then from Proposition 4.4 we have
\[ |f(y) - Q_x(y)| \leq \frac{\|y - x\|_x^2}{3(1 - 1/3)} \leq \frac{\|y - x\|_x^2}{9(2/3)} = \frac{\|y - x\|_x^2}{6}, \]
and therefore
\[ f(y) \geq f(x) + g(x)^T H_x^{-1} H_x^{\frac{1}{2}} y - x + \frac{1}{2}\|y - x\|_x^2 - \frac{1}{6}\|y - x\|_x^2 \]
\[ \geq f(x) - \|n(x)\|_x \|y - x\|_x + \frac{1}{3}\|y - x\|_x^2 \]
\[ = f(x) + \frac{4}{3}\|y - x\|_x (-3\|n(x)\|_x + \|y - x\|_x). \]
Now if \( y \in \partial S := \partial \{y : \|y - x\|_x \leq 3\|n(x)\|_x\} \), it follows that \( f(y) \geq f(x) \). So, by Fact 3.5, \( f(\cdot) \) has a global minimizer \( z \in S \), and so \( \|z - x\|_x \leq 3\|n(x)\|_x \).

Now suppose that \( \|n(x)\|_x \leq 1/4 \). From Theorem 4.1 we have
\[ \|n(x_+\|_x \leq \left( \frac{1/4}{1 - 1/4} \right)^2 = 1/9, \]
so \( f(\cdot) \) has a global minimizer \( z \) and \( \|z - x_+\|_{x_+} \leq 3\|n(x_+)\|_{x_+} \). Therefore

\[
\|z - x_+\|_{x_+} \leq \frac{\|z - x_+\|_{x_+}}{1 - \|x - x_+\|_{x_+}} \quad \text{(from Definition 4.1)}
\]

\[
= \frac{\|z - x_+\|_{x_+}}{1 - \|n(x)\|_{x}}
\]

\[
\leq \frac{3\|n(x_+)\|_{x_+}}{1 - \|n(x)\|_{x}}
\]

\[
\leq \frac{3\|n(x)\|_{x}^2}{(1 - \|n(x)\|_{x})^3} \quad \text{(from Theorem 4.1)}
\]

Last of all in this section, we present a more traditional result about the convergence of Newton’s method for a self-concordant function. This result is not used elsewhere in our development, and is only included to relate the results herein to more traditional theory of Newton’s method. The proof of this result is left as a somewhat-challenging exercise.

**Theorem 4.3** Suppose that \( f(\cdot) \in SC \) and \( x \in D_f \) and \( f(\cdot) \) has a minimizer \( z \). If \( \|x - z\|_z < \frac{1}{4} \), then

\[
\|x_+ - z\|_z < 4\|x - z\|_z^2 .
\]

## 5 Self-Concordant Barriers

We begin with another definition.

**Definition 5.1** \( f(\cdot) \) is a \( \psi \)-(strongly nondegenerate self-concordant)-barrier if \( f(\cdot) \in SC \) and

\[
\psi = \psi_f := \max_{x \in D_f} \|n(x)\|_x^2 < \infty .
\]
Note that \( \|n(x)\|_2^2 = (-g(x)^T H(x)^{-1} H(x) H(x)^{-1} (-g(x)) = g(x)^T H(x)^{-1} g(x) \), so we can equivalently define
\[
\vartheta_f := \max_{x \in D_f} g(x)^T H(x)^{-1} g(x)
\]
or
\[
\vartheta_f := \max_{x \in D_f} n(x)^T H(x) n(x).
\]
The quantity \( \vartheta_f \) is called the complexity value of the barrier \( f(\cdot) \).

Let \( SCB \) denote the class of all such functions. The following property is very important.

**Theorem 5.1** Suppose that \( f(\cdot) \in SCB \) and \( x, y \in D_f \). Then
\[
g(x)^T (y - x) < \vartheta_f.
\]

**Proof:** Define \( \phi(t) := f(x + t(y - x)) \), whereby \( \phi'(t) = g(x + t(y - x))^T (y - x) \) and \( \phi''(t) = (y - x)^T H(x + t(y - x))(y - x) \). We want to prove that \( \phi'(0) < \vartheta_f \).
If \( \phi'(0) \leq 0 \) there is nothing further to prove, so we can assume that \( \phi'(0) > 0 \) whereby from convexity it also follows that \( \phi'(t) > 0 \) for all \( t \in [0, 1] \). (Notice that \( \phi(1) = f(y) \) and so \( t = 1 \) is in the domain of \( \phi(\cdot) \).) Let \( t \in [0, 1] \) be given and let \( v = x + t(y - x) \). Then
\[
\phi'(t) = g(v)^T (y - x)
\]
\[
= g(v)^T H_v^{-1} H_v^{1/2} H_v^{1/2} (y - x)
\]
\[
= -n(v)^T H_v^{1/2} H_v^{1/2} (y - x)
\]
\[
\leq \|H_v^{1/2} n(v)\|\|H_v^{1/2} (y - x)\|
\]
\[
= \|n(v)\|_v \|y - x\|_v \leq \sqrt{\vartheta_f} \|y - x\|_v.
\]
Also \( \phi''(t) = (y - x) H_v (y - x) = \|y - x\|_v^2 \), whereby
\[
\frac{\phi''(t)}{\phi'(t)^2} \geq \frac{\|y - x\|_v^2}{\vartheta_f \|y - x\|_v^2} = \frac{1}{\vartheta_f}.
\]
It follows that
\[
\frac{1}{\vartheta_f} \leq \int_0^1 \frac{\phi''(t)}{\phi(t)^2} \, dt = \left. -\frac{1}{\phi'(t)} \right|_0^1 = \frac{1}{\phi'(0)} - \frac{1}{\phi'(1)},
\]
and we have
\[
\frac{1}{\phi'(0)} \geq \frac{1}{\phi'(1)} + \frac{1}{\vartheta_f},
\]
which proves the result. \(\Box\)

The next two results show how the complexity value \(\vartheta\) behaves under addition/intersection and affine transformation.

**Theorem 5.2 (self-concordant barriers under addition/intersection)**

Suppose that \(f_i(\cdot) \in \mathcal{SCB}\) with domain \(D_i := D_{f_i}\) and complexity values \(\vartheta_i := \vartheta_{f_i}\) for \(i = 1, 2\), and suppose that \(D := D_1 \cap D_2 \neq \emptyset\). Define \(f(\cdot) = f_1(\cdot) + f_2(\cdot)\). Then \(f(\cdot) \in \mathcal{SCB}\) with domain \(D\), and \(\vartheta_f \leq \vartheta_1 + \vartheta_2\).

**Proof:** Fix \(x \in D\) and let \(g, g_1, g_2\) and \(H, H_1, H_2\) denote the gradients and Hessians of \(f(\cdot), f_1(\cdot), f_2(\cdot)\) at \(x\), whereby \(g = g_1 + g_2\) and \(H = H_1 + H_2\). Define \(A_i = H^{-\frac{1}{2}} H_i H^{-\frac{1}{2}}\) for \(i = 1, 2\). Then \(A_i \succ 0\), \(A_1 + A_2 = I\), so \(A_1, A_2\) commute and \(A_1^\frac{1}{2}, A_2^\frac{1}{2}\) commute, from Fact 3.7. Also define \(u_i = A_i^{-\frac{1}{2}} H^{-\frac{1}{2}} g_i\)
for $i = 1, 2$. We have
\[ g^T H^{-1} g = g_1^T H^{-1} g_1 + g_2^T H^{-1} g_2 + 2g_1^T H^{-1} g_2 \]
\[ = u_1^T A_1 u_1 + u_2^T A_2 u_2 + 2u_1^T A_1^\frac{1}{2} A_2^\frac{1}{2} u_2 + 2u_2^T A_1^\frac{1}{2} A_2^\frac{1}{2} u_1 \]
\[ = u_1^T [I - A_2] u_1 + u_2^T [I - A_1] u_2 + 2u_1^T A_3^\frac{1}{2} A_2^\frac{1}{2} u_2 \]
\[ = u_1^T u_1 + u_2^T u_2 - \left[ u_1^T A_2 u_1 + u_2^T A_1 u_2 - 2u_1^T A_2^\frac{1}{2} A_1^\frac{1}{2} u_2 \right] \]
\[ = g_1^T H^{-\frac{1}{2}} A_1^{-1} H^{-\frac{1}{2}} g_1 + g_2^T H^{-\frac{1}{2}} A_2^{-1} H^{-\frac{1}{2}} g_2 - \frac{1}{2} g_2^T H^{-\frac{1}{2}} A_2^{-1} H^{-\frac{1}{2}} g_2 \]
\[ \leq g_1^T H^{-\frac{1}{2}} H_1^\frac{1}{2} H_2^\frac{1}{2} H^{-\frac{1}{2}} g_1 + g_2^T H^{-\frac{1}{2}} H_2^\frac{1}{2} H_1^\frac{1}{2} H^{-\frac{1}{2}} g_2 \]
\[ \leq \vartheta_1 + \vartheta_2 \]

thereby showing that $\vartheta_f \leq \vartheta_1 + \vartheta_2$. \hfill \[\blacksquare\]

**Theorem 5.3 (self-concordant barriers under affine transformation)**

Let $A \in \mathbb{R}^{m \times n}$ satisfy $\text{rank}A = n \leq m$. Suppose that $f(\cdot) \in \text{SCB}$ with complexity value $\vartheta_f$, with domain $D_f \subseteq \mathbb{R}^m$ and define $f(\cdot)$ by $f(x) = f(Ax - b)$. Then $\hat{f}(\cdot) \in \text{SCB}$ and $\vartheta_f \leq \vartheta_f$.

**Proof:** Fix $x \in \hat{D}$ and define $s = Ax - b$. Letting $g$ and $H$ denote the gradient and Hessian of $f(s)$ at $s$ and $\hat{g}$ and $\hat{H}$ the gradient and Hessian of $\hat{f}(x)$ at $x$, we have $\hat{g} = A^T g$ and $\hat{H} = A^T H A$. Then
\[ g^T \hat{H}^{-1} \hat{g} = g^T A^T H A^{-1} A^T g = g^T H^{-\frac{1}{2}} H_1^\frac{1}{2} H_2^\frac{1}{2} H^{-\frac{1}{2}} A^T H_1^\frac{1}{2} H_2^\frac{1}{2} A^{-1} A^T H_1^\frac{1}{2} H_2^\frac{1}{2} g \]
\[ \leq g^T H^{-\frac{1}{2}} H_2^\frac{1}{2} H^{-\frac{1}{2}} g = g^T H^{-1} g \leq \vartheta_f \]

since the matrix $H_2^\frac{1}{2} A^T H_1^\frac{1}{2} H_2^\frac{1}{2} A^{-1} A^T H_1^\frac{1}{2} H_2^\frac{1}{2}$ is a projection matrix. \hfill \[\blacksquare\]

**Remark 2** The complexity values of the three most-used barriers are as follows:
1. \( \varphi_f = 1 \) for the barrier \( f(x) = -\ln(x) \) defined on \( D_f = \{ x : x > 0 \} \)

2. \( \varphi_f = k \) for the barrier \( f(X) = -\ln \det(X) \) defined on \( D_f = \{ X \in S^{k \times k} : X > 0 \} \)

3. \( \varphi_f = 2 \) for the barrier \( f(x) = -\ln(x_1^2 - \sum_{j=2}^{n} x_j^2) \) defined on \( D_f = \{ x : \| (x_2, \ldots, x_n) \| \leq x_1 \} \)

**Proof:** Item (1.) follows from item (2.) so we first prove (2.). Recall the use of the trace inner product \( X \cdot Y = \text{Tr}(XY) \) and the Frobenius norm \( \|X\| := \sqrt{X \cdot X} \) as discussed early in the proof of Proposition 4.3. Also recall that for \( f(X) = -\ln \det(X) \) we have \( g(X) = -X^{-1} \) and \( H(X) \Delta X = X^{-1} \Delta XX^{-1} \). Therefore the Newton step at \( X \), denoted by \( \mathbf{n}(X) \), is the solution of the following equation:

\[
X^{-1}[\mathbf{n}(X)]X^{-1} = X^{-1}
\]

and it follows that \( \mathbf{n}(X) = X \). Therefore, using (1) we have

\[
\|\mathbf{n}(X)\|_X^2 = n(X) \cdot X^{-1} n(X) X^{-1} = X \cdot X^{-1} X^{-1} = X \cdot X^{-1} = \text{Tr}(XX^{-1}) = \text{Tr}(I) = k
\]

and therefore \( \varphi_f = \max_{X \succ 0} \| n(X) \|^2_X = k \), which proves (2.) and hence (1.).

In order to prove (3.) we amend our notation a bit, letting \( Q^n = \{ (t, x) \in \mathbb{R}^1 \times \mathbb{R}^{n-1} : \| x \| \leq t \} \). For \( (t, x) \in \text{int}Q^n \) we have \( \| x \| < t \) and mechanically we can derive:

\[
g(t, x) = \begin{pmatrix} -2t \frac{t^2 - x^T x}{t^2 - x^T x} \\ \frac{2t^2 + 2x^T x}{(t^2 - x^T x)^2} - 4tx^T \\ \frac{2(t^2 - x^T x)I + 4xx^T}{(t^2 - x^T x)^2} \end{pmatrix} \quad H(t, x) = \begin{pmatrix} -4tx^T \\ \frac{2(t^2 - x^T x)I + 4xx^T}{(t^2 - x^T x)^2} \end{pmatrix}
\]

and the Hessian inverse is given by

\[
H(t, x)^{-1} = \begin{pmatrix} \frac{t^2 + x^T x}{2} & \frac{tx^T}{2} \\ \frac{tx^T}{2} & \frac{t^2 - x^T x}{2} - I + xx^T \end{pmatrix}
\]

Directly plugging in yields

\[
g(t, x)^T H^{-1}(t, x) g(t, x) = 2
\]
from which it follows that \( \vartheta_f = \max_{(t,x) \in \mathbb{Q}^n} g(x)^T H(x)^{-1} g(x) = 2. \)

Finally, we present a result that shows that “one half of the Dikin ball” approximates the shape of a relevant portion of \( D_f \) to within a factor of \( \vartheta_f \).

**Proposition 5.1** Suppose that \( f(\cdot) \in \mathcal{SCB} \) with complexity value \( \vartheta_f \), let \( x \in D_f \), and define the following three sets:

\[
\begin{align*}
(a) \quad S_1 & := \{ y : \|y - x\|_x < 1, \ g(x)^T (y - x) \geq 0 \} , \\
(b) \quad S_2 & := \{ y : y \in D_f, \ g(x)^T (y - x) \geq 0 \} , \text{ and} \\
(c) \quad S_3 & := \{ y : \frac{\|y - x\|_x}{4^{\vartheta_f} + 1} < 1, \ g(x)^T (y - x) \geq 0 \} .
\end{align*}
\]

Then \( S_1 \subset S_2 \subset S_3 \).

The proof of Proposition 5.1 is not more involved than other results herein. It is not included because it is not necessary for subsequent results. It is stated here to show the power and beauty of self-concordant barriers.

**Corollary 5.1** If \( z \) is the minimizer of \( f(\cdot) \), then
\[
B_z(z,1) \subset D_f \subset B_z(z,4^{\vartheta_f} + 1) .
\]

### 6 The Barrier Method and its Analysis

Our original problem of interest is
\[
P : \quad \text{minimize}_{x} \quad c^T x \\
\text{s.t.} \quad x \in S ,
\]
whose optimal objective value we denote by \( V^* \). Let \( f(\cdot) \) be a self-concordant barrier on \( D_f = \text{int}S \). For every \( \mu > 0 \) we create the barrier problem:
\( P_\mu : \text{ minimize}_x \; \mu c^T x + f(x) \)

s.t. \( x \in D_f \).

The solution of this problem for each \( \mu \) is denoted \( z(\mu) \):

\[
z(\mu) := \arg \min_x \{ \mu c^T x + f(x) : x \in D_f \}.
\]

Intuitively, as \( \mu \to \infty \), the impact of the barrier function on the solution of \( P_\mu \) should become less and less, so we should have \( c^T z(\mu) \to V^* \) as \( \mu \to \infty \). Presuming this is the case, the barrier scheme will use Newton’s method to solve for approximate solutions \( x^i \) of \( P_{\mu_i} \) for an increasing sequence of values of \( \mu_i \to \infty \). In order to be more specific about how the barrier scheme might work, let us assume that at each iteration we have some value \( x \in D_f \) that is an approximate solution of \( P_\mu \) for a given value \( \mu > 0 \). (We will define “an approximate solution of \( P_\mu \)” shortly.) We then will increase the barrier parameter \( \mu \) by a multiplicative factor \( \alpha > 1 \):

\[
\hat{\mu} \leftarrow \alpha \mu.
\]

Then we will take a Newton step at \( x \) for the problem \( P_{\hat{\mu}} \) to obtain a new point \( \hat{x} \) that we would like to then be an approximate solution of \( P_{\hat{\mu}} \). If so, we can continue the scheme inductively.

Let \( x \in D_f \) be given, and let us compute the Newton step for \( P_\mu \) at \( x \). The objective function of \( P_\mu \) is

\[
h_\mu(x) := \mu c^T x + f(x),
\]

whereby we have:

(a) \( \nabla h_\mu(x) = \mu c + g(x) \) and

(b) \( \nabla^2 h_\mu(x) = H(x) = H_x \).
Therefore the Newton step for $h_\mu(\cdot)$ at $x$ is:

$$n_\mu(x) := -H_x^{-1}(\mu c + g(x)) = n(x) - \mu H_x^{-1}c$$

and the new iterate is:

$$\hat{x} := x + n_\mu(x) = x - H_x^{-1}(\mu c + g(x)).$$

**Remark 3** Notice that $h_\mu(\cdot) \in SC$, since membership in $SC$ has only to do with Hessians, and $h_\mu(\cdot)$ and $f(\cdot)$ have the same Hessian. However, membership in $SCB$ depends also on gradients, and $h_\mu(x)$ and $f(x)$ have different gradients, whereby it will be typically true that $h_\mu(\cdot) \notin SCB$ (unless $c = 0$).

We now define what we mean for $y$ to be an “approximate solution” of $P_\mu$.

**Definition 6.1** Let $\gamma \in [0,1)$ be given. We say that $y \in D_f$ is a $\gamma$-approximate solution of $P_\mu$ if

$$\|n_\mu(y)\|_y \leq \gamma.$$ 

Notice that this definition is equivalent to

$$\|n(y) - \mu H_y^{-1}c\|_y \leq \gamma.$$ 

Essentially, the definition states that $y$ is a $\gamma$-approximate solution of $P_\mu$ if the Newton step for $P_\mu$ at $y$ is small (measured using the local norm at $y$).

The following theorem gives an explicit optimality gap bound for $y$ if $y$ is a $\gamma = 1/4$-approximate solution of $P_\mu$.

**Theorem 6.1** Suppose $\gamma \leq \frac{1}{4}$, and $y \in D_f$ is a $\gamma$-approximate solution of $P_\mu$. Then

$$c^Ty \leq V^* + \frac{\vartheta_f}{\mu} \left( \frac{1}{1 - \delta} \right)$$

where

$$\delta = \gamma + \frac{3\gamma^2}{(1 - \gamma)^2}. $$

24
**Proof:** From Theorem 4.2 we know that $z(\mu)$ exists and furthermore

$$
\|y - z(\mu)\|_y = \|y + n_\mu(y) - z(\mu) - n_\mu(y)\|_y
$$

$$
\leq \|y_+ - z(\mu)\|_y + \|n_\mu(y)\|_y
$$

$$
\leq \frac{3\gamma^2}{(1-\gamma)^2} + \gamma = \delta .
$$

From basic first-order optimality conditions we know that $z(\mu)$ satisfies

$$
\mu c + g(z(\mu)) = 0
$$

and from Theorem 5.1 we have

$$
-\mu c^T (w - z(\mu)) = g(z(\mu))^T (w - z(\mu)) < \vartheta_f \quad \text{for all } w \in D_f .
$$

Rearranging we have

$$
c^T w + \frac{\vartheta_f}{\mu} > c^T z(\mu) \quad \text{for all } w \in D_f ,
$$

whereby $V^* + \frac{\vartheta_f}{\mu} \geq c^T z(\mu)$. Now for notational convenience denote $z := z(\mu)$.
and observe
\[ c^T y = c^T z + c^T (y - z) \]
\[ \leq V^* + \frac{\varphi_f}{\mu} + c^T H_z^{-\frac{1}{2}} H_z^{\frac{1}{2}} (y - z) \]
\[ \leq V^* + \frac{\varphi_f}{\mu} + \|H_z^{-\frac{1}{2}} c\| \| (y - z) \|_z \]
\[ \leq V^* + \frac{\varphi_f}{\mu} + \left( \sqrt{c^T H_z^{-1} c} \right) \frac{\| (y - z) \|_y}{1 - \| (y - z) \|_y} \]
\[ \leq V^* + \frac{\varphi_f}{\mu} + \left( \sqrt{\frac{(g(z)/\mu)^T H_z^{-1} (g(z)/\mu)}{\mu}} \right) \frac{\delta}{1 - \delta} \]
\[ = V^* + \frac{\varphi_f}{\mu} + \frac{\sqrt{\varphi_f}}{\mu} \frac{\delta}{1 - \delta} \]
\[ \leq V^* + \frac{\varphi_f}{\mu} \left( 1 + \frac{\delta}{1 - \delta} \right) \]
\[ = V^* + \frac{\varphi_f}{\mu(1 - \delta)}. \]

The last inequality above follows from the fact (which we will not prove) that \( \varphi_f \geq 1 \) for any \( f(\cdot) \in S\mathcal{C}E. \]

Note that with \( \gamma = 1/9 \) we have \( 1/(1 - \delta) \leq 6/5 \), whereby \( c^T y \leq V^* + 1.2 \varphi_f/\mu. \]

**Theorem 6.2** Let \( \beta := \frac{1}{4}, \gamma := \frac{1}{5}, \) and \( \alpha := \frac{\sqrt{\varphi_f + \beta}}{\sqrt{\varphi_f + \gamma}}. \) Suppose \( x \) is a \( \gamma \)-approximate solution of \( P_\mu \). Define \( \tilde{\mu} := \alpha \mu, \) and let \( \tilde{x} \) be the Newton iterate for \( P_{\tilde{\mu}} \) at \( x \), namely
\[ \tilde{x} := x - H(x)^{-1} (\tilde{\mu} c + g(x)) \].

Then
Algorithm 2 Barrier Method

Initialize.

Define $\gamma := 1/9$, $\beta := 1/4$, $\alpha := \frac{\sqrt{3}+\beta}{\sqrt{3}+\gamma}$.

Initialize with $\mu_0 > 0$, $x^0 \in D_f$ that is a $\gamma$-approximate solution of $P_{\mu_0}$. $i \leftarrow 0$.

At iteration $i$:

1. Current values.

   $\mu \leftarrow \mu_i$
   
   $x \leftarrow x^i$

2. Increase $\mu$ and take Newton step.

   $\hat{\mu} \leftarrow \alpha \mu$
   
   $\hat{x} \leftarrow x - H(x)^{-1}(\hat{\mu}c + g(x))$

3. Update values.

   $\mu_{i+1} \leftarrow \hat{\mu}$
   
   $x^{i+1} \leftarrow \hat{x}$
1. \( x \) is a \( \beta \)-approximate solution of \( P^\mu \), and

2. \( \hat{x} \) is a \( \gamma \)-approximate solution of \( P^\mu \).

**Proof:** To prove (1.) we have:

\[
\|n^\mu(x)\|_x = \|n_{\alpha\mu}(x)\|_x = \|H(x)^{-1}(\alpha\mu c + g(x))\|_x
\leq \alpha \|H(x)^{-1}(\mu c + g(x))\|_x + (1 - \alpha)\|H(x)^{-1}g(x)\|_x
\leq \alpha \gamma + (1 - \alpha)\|n(x)\|_x
\leq \alpha \gamma + (1 - \alpha)\sqrt{\mathcal{F}} = \beta.
\]

To prove (2.) we invoke Theorem 4.1:

\[
\|n^\mu(\hat{x})\|_{\hat{x}} \leq \frac{\|n^\mu(x)\|_x^2}{(1 - \|n^\mu(x)\|_x^2)^2} \leq \frac{\beta^2}{(1 - \beta)^2} = \frac{1}{9} = \gamma.
\]

Applying Theorem 6.2 inductively we obtain the basic barrier method for self-concordant barriers as presented in Algorithm 2. The complexity of this scheme is presented below.

**Theorem 6.3** Let \( \varepsilon > 0 \) be the desired optimality tolerance, and define

\[
J := \left\lceil 9\sqrt{\delta} \ln \left( \frac{6\delta}{5\mu_0\varepsilon} \right) \right\rceil
\]

Then by iteration \( J \) of the barrier method the current iterate \( x \) satisfies \( c^T x \leq V^* + \varepsilon \).

**Proof:** With the given values of \( \gamma, \beta, \alpha \) we have \( \delta := \gamma + \frac{3\gamma^2}{(1-\gamma)^3} \) satisfies \( \frac{1}{1-\delta} \leq 6/5 \) and

\[
1 - \frac{1}{\alpha} = \frac{1}{7.2\sqrt{\theta} + 1.8} \geq \frac{1}{9\sqrt{\theta}}.
\]

28
After $J$ iterations the current iterate $x$ is a $\gamma$-approximate solution of $P_\mu$ where $\mu = \alpha^J \mu_0$. Therefore

$$\ln \mu_0 - \ln \mu = J \ln \left( \frac{1}{\alpha} \right) \leq J \left( \frac{1}{\alpha} - 1 \right).$$

(Note that the inequality above follows from the fact that $\ln t \leq t - 1$ for $t > 0$, which itself is a consequence of the concavity of $\ln(\cdot)$.) Therefore

$$\ln \mu \geq \ln \mu_0 + J \left( 1 - \frac{1}{\alpha} \right) \geq \ln \mu_0 + J \frac{\mu_0}{9\sqrt{\vartheta}} \geq \ln \mu_0 + \ln \left( \frac{6\vartheta}{5\mu_0\varepsilon} \right) \geq \ln \left( \frac{\vartheta}{(1 - \delta)\varepsilon} \right).$$

Therefore $\mu \geq \frac{\vartheta}{(1 - \delta)^2}$, and from Theorem 6.1 we have

$$e^T x \leq V^* + \frac{\vartheta}{\mu(1 - \delta)} \leq V^* + \varepsilon.$$

\section{Remarks and other Matters}

\subsection{Nesterov-Nemirovskii definition of self-concordance}

The original definition of self-concordance due to Nesterov and Nemirovskii is different than that presented here in Definition 4.1. Let $f(\cdot)$ be a function defined on an open set $D_f \subset \mathbb{R}^n$. For every $x \in D_f$ and every vector $h \in \mathbb{R}^n$ define the univariate function

$$\phi_{x,h}(\alpha) := f(x + \alpha h).$$

Then Nesterov and Nemirovskii’s definition of self-concordance is as follows.

\textbf{Definition 7.1} The function $f(\cdot)$ defined on the domain $D_f \subset \mathbb{R}^n$ is (strongly nondegenerate) self-concordant if:

(a) $H(x) \succ 0$ for all $x \in D_f$,

(b) $f(x) \to \infty$ as $x \to \partial D_f$, and
(c) for all \( x \in D_f \) and \( h \in \mathbb{R}^n \), \( \phi_{x,h} (\cdot) \) satisfies 
\[
|\phi''_{x,h} (0)| \leq 2 \left[ \phi'''_{x,h} (0) \right]^\frac{3}{2}.
\]

Definition 7.1 roughly states that the third derivative of \( f(\cdot) \) is bounded by an appropriate function of the second derivative. This definition assumes \( f(\cdot) \) is three-times differentiable, unlike Definition 4.1. It turns out that when \( f(\cdot) \) is three-times differentiable, then Definition 7.1 implies the properties of Definition 4.1 and vice versa, thus the two definitions are equivalent. We prefer Definition 4.1 because proofs of later properties are far more easy to establish. The key advantage of Definition 7.1 is that it leads to an easier way to validate the self-concordance of most self-concordant functions, especially \( -\ln \det(X) \) and \( -\ln(t^2 - x^T x) \) for the semidefinite and second-order cones, respectively.

### 7.2 Universal Barrier

One natural question is: given a convex set \( S \subset \mathbb{R}^n \) with nonempty interior, does there exist a self-concordant barrier \( f_S(\cdot) \) whose domain is \( D_f = \text{int}S \)? Furthermore, if the answer is yes, does there exist a barrier whose complexity value is, say, \( O(n) \)? It turns out that the answers to these two questions are both “yes,” but the proofs are incredibly difficult.

Let \( S \subset \mathbb{R}^n \) be a closed convex set with nonempty interior. For each point \( x \in \text{int}S \), define the “local polar of \( S \) relative to \( x \)”: 
\[
P_S(x) := \{ w \in \mathbb{R}^n : w^T(y - x) \leq 1 \text{ for all } y \in S \}.
\]

Basic geometry suggests that as \( x \to \partial S \), the volume of \( P_S(x) \) should approach \( \infty \). Now define the function:
\[
f_S(x) := \ln (\text{volume} (P_S(x))) .
\]

It turns out that \( f_S(\cdot) \) is a self-concordant function, but this is quite hard to prove. Furthermore, there is a universal constant \( \bar{C} \) such that for any dimension \( n \) and any closed convex set \( S \) with nonempty interior, it is true that
\[
\vartheta_{f_S(\cdot)} \leq \bar{C} \cdot n.
\]
This is a remarkable and very deep result. However, it is computationally useless, since the function values of $f_S(\cdot)$ are not only extremely difficult to compute, but gradients and Hessian information is also extremely difficult to compute.

7.3 Other Matters

(i) getting started,
(ii) other formats for convex optimization,
(iii) $\vartheta$-logarithmic homogeneous barriers for cones,
(iv) primal-dual methods,
(v) computational practice, and
(vi) other self-concordant functions and self-concordant calculus.

7.4 Exercises

1. Prove that if $\|n(x)\|_x > 1/4$, then upon setting

$$y = x + \frac{1}{5\|n(x)\|_x} n(x),$$

we have $f(y) \leq f(x) - 1/37.5$.

2. Prove Theorem 4.3. To get started, observe that $x_+ - z = x - z - H_x^{-1}(g(x) - g(z))$ (since $z$ is a minimizer and hence $g(z) = 0$), and then use Fact 3.1 to show that

$$x_+ - z = H_x^{\frac{1}{2}} \int_0^1 H_x^{\frac{1}{2}} \left[ H_x - H(x + t(z - x)) \right] H_x^{-\frac{1}{2}} H_x^{\frac{1}{2}} (x - z) dt.$$

Next multiply each side by $H_x^{\frac{1}{2}}$ and take norms, and then invoke Lemma 4.1. This should help get you started.
3. Use Theorem 4.3 to show that under the hypothesis of the theorem, the sequence of Newton iterates starting with $x, x^1 = x_+, x^2 = (x^1)_+, \ldots$, satisfies
\[ \|x^i - z\|_z < \frac{1}{4} (4\|x - z\|_z)^{2^i}. \]

4. Complete the proof Proposition 4.1 by proving the “$\geq$” inequality in the definition of self-concordance using the suggestions at the end of the proof earlier in the text.

5. Complete the proof Proposition 4.2 by proving the “$\geq$” inequality in the definition of self-concordance using the suggestions at the end of the proof earlier in the text.

6. The proof of Proposition 4.3 uses the inequalities presented in (4), but only the left-most inequality is proved in the text herein. Prove the right-most inequality by substituting $\lambda_{\text{max}}$ for $\lambda_{\text{min}}$ and switching $\geq$ for $\leq$ in the chain of equalities and inequalities in the text.