Chapter 5
Polynomial-Time Interior-Point Methods

In this section, we present the problem classes and complexity bounds of polynomial-time interior-point methods. These methods are based on the notion of a self-concordant function. It appears that such a function can be easily minimized by the Newton's Method. On the other hand, an important subclass of these functions, the self-concordant barriers, can be used in the framework of path-following schemes. Moreover, it can be proved that we can follow the corresponding central path with polynomial-time complexity. The size of the steps in the penalty coefficient of the central path depends on the corresponding barrier parameter. It appears that for any convex set there exists a self-concordant barrier with parameter proportional to the dimension of the space of variables. On the other hand, for any convex set with explicit structure, such a barrier with a reasonable value of parameter can be constructed by simple combination rules. We present applications of this technique to Linear and Quadratic Optimization, Linear Matrix Inequalities and other optimization problems.

5.1 Self-concordant Functions

(Do we really have a Black Box? What does the Newton method actually do? Definition of self-concordant functions; Main properties; The Implicit Function Theorem; Minimizing self-concordant functions; Relations with the standard second-order methods.)

5.1.1 The Black Box Concept in Convex Optimization

In this chapter, we are going to present the main ideas underlying the modern polynomial-time interior-point methods in Nonlinear Optimization. In order to start, let us look first at the traditional formulation of a minimization problem.

Suppose we want to solve a minimization problem in the following form:

$$\min_{x \in \mathbb{R}^n} \{ f_0(x) : \ f_j(x) \leq 0, \ j = 1 \ldots m \}.$$

We assume that the functional components of this problem are convex. Note that all standard convex optimization schemes for solving this problem are based on the Black-Box concept. This means that we assume our problem to be equipped with an oracle, which provides us with some information on the functional components of the problem at some test point x. This oracle is local: If we change the shape of the component far enough from the test point, the answer of the oracle does not change. These answers comprise the only information available for numerical methods.[1]

However, looking carefully at the above situation, we can discover a certain contradiction. Indeed, in order to apply the convex optimization methods, we need to be *sure* that our functional components are convex. However, we can check convexity only by analyzing the *structure* of these functions[2]: If our function is obtained from the *basic* convex functions by *convex* operations (summation, maximum, etc.), we conclude that it is convex.

Thus, the functional components of the problem are not in the Black Box at the moment we are checking their convexity and choose the minimization scheme. However, we lock them in the Black Box for numerical methods. This is the main conceptual contradiction of the standard Convex Optimization theory.[3]

The above observation gives us hope that the structure of the problem could be used to improve performance of convex minimization schemes. Unfortunately, structure is a very fuzzy notion, which is quite difficult to formalize. One possible way to describe the structure is to fix the *analytical type* of functional components. For example, we can consider the problems with linear functions $f_j(\cdot)$ only. This works, but note that this approach is very fragile: If we introduce in our problem just a single functional component of different type, we get another problem class and all the theory must be redone from scratch.

Alternatively, it is clear that having the structure at hand, we can play with the *analytical form* of the problem. We can rewrite the problem in many equivalent forms using nontrivial transformations of variables or constraints, introducing additional variables, etc. However, this would serve no purpose without realizing the final goal of such transformations. So, let us try to find such a goal.

At this moment, it is better to look at classical examples. In many situations, the sequential reformulations of the initial problem can be seen as a part of the numerical method. We start from a complicated problem $\mathscr{P}$ and, step by step, simplify its structure up to the moment we get a trivial problem (or, a problem

---

[1] We have already discussed this concept and the corresponding methods in Part I of the book.

[2] A numerical verification of convexity is a hopeless computational task.

[3] Nevertheless, the conclusions of the theory concerning the oracle-based minimization schemes remain valid, of course, for the methods which are *designed* in accordance with the Black-Box principles.

which we know how to solve):

$$\mathcal{P} \longrightarrow \ldots \longrightarrow (f^*, x^*).$$

Let us look at the standard approach for solving the system of linear equations, namely,

$$Ax = b.$$

We can proceed as follows:

1. Check that matrix A is symmetric and positive definite. Sometimes this is clear from its origin.
2. Compute the Cholesky factorization of the matrix:

$$A = LL^T,$$

where L is a lower-triangular matrix. Form two auxiliary systems

$$Ly = b, \quad L^T x = y.$$

3. Solve the auxiliary systems.

This process can be seen as a sequence of equivalent transformations of the initial problem.

Imagine for a moment that we do not know how to solve the systems of linear equation. In order to discover the above technology, we should perform the following steps:

1. Find a class of problems which can be solved very efficiently (linear systems with triangular matrices in our example).
2. Describe the transformation rules for converting our initial problem into the desired form.
3. Describe the class of problems for which these transformation rules are applicable.

We are ready to explain how it works in Convex Optimization. First of all, we need to find a *basic* numerical scheme and problem formulation at which this scheme is very efficient. We will see that for our goals the most appropriate candidate is the *Newton's method* (see Sect. 1.2.4 and Chap. 4) as applied in the framework of *Sequential Unconstrained Minimization* (see Sect. 1.3.3).

In the next section, we will analyze some drawbacks of the standard theory on the Newton's method. From this analysis, we derive a family of very special convex functions, so-called *self-concordant functions* and *self-concordant barriers*, which can be efficiently minimized by the Newton's method. We use these objects in the description of a transformed version of the initial problem. In the sequel, we refer to this description as a *barrier model* of our problem. This model will replace

the standard functional model of the optimization problem used in the previous chapters.

5.1.2   What Does the Newton's Method Actually Do?

Let us look at the standard result on the local convergence of Newton's method (we have proved it as Theorem 1.2.5). We need to find an unconstrained local minimum x^* of the twice differentiable function $f(\cdot)$:

$$\min_{x \in \mathbb{R}^n} f(x), \tag{5.1.1}$$

For the moment, all the norms we use are standard Euclidean. Assume that:

- $\nabla^2 f(x^*) \succeq \mu I_n$ with some constant $\mu > 0$,
- $\| \nabla^2 f(x) - \nabla^2 f(y) \| \leq M \| x - y \|$ for all x and $y \in \mathbb{R}^n$.

Assume also that the starting point of the Newton process x_0 is close enough to x^*:

$$\| x_0 - x^* \| < \bar{r} = \tfrac{2\mu}{3M}. \tag{5.1.2}$$

Then we can prove (see Theorem 1.2.5) that the sequence

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k), \quad k \geq 0, \tag{5.1.3}$$

is well defined. Moreover, $\| x_k - x^* \| < \bar{r}$ for all $k \geq 0$ and the Newton's method (5.1.3) converges quadratically:

$$\| x_{k+1} - x^* \| \leq \tfrac{M \| x_k - x^* \|^2}{2(\mu - M \| x_k - x^* \|)}.$$

What is wrong with this result? Note that the description of the *region of quadratic convergence* (5.1.2) for this method is given in terms of the *standard inner product*

$$\langle x, y \rangle = \sum_{i=1}^{n} x^{(i)} y^{(i)}, \quad x, y \in \mathbb{R}^n.$$

If we choose a new basis in $\mathbb{R}^n$, then all objects in our description change: the metric, the Hessians, the bounds μ and M. However, let us see what happens in this situation with the Newton process. Namely, let B be a nondegenerate $(n \times n)$-matrix. Consider the function

$$\phi(y) = f(By), \quad y \in \mathbb{R}^n.$$

The following result is very important for understanding the nature of the Newton's method.

Lemma 5.1.1 *Let the sequence $\{x_k\}$ be generated by the Newton's method as applied to the function f:*

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k), \quad k \geq 0.$$

Consider the sequence $\{y_k\}$, generated by the Newton's method for the function ϕ:

$$y_{k+1} = y_k - [\nabla^2 \phi(y_k)]^{-1} \nabla \phi(y_k), \quad k \geq 0,$$

with $y_0 = B^{-1} x_0$. Then $y_k = B^{-1} x_k$ for all $k \geq 0$.

Proof Let $y_k = B^{-1} x_k$ for some $k \geq 0$. Then

$$y_{k+1} = y_k - [\nabla^2 \phi(y_k)]^{-1} \nabla \phi(y_k) = y_k - [B^T \nabla^2 f(By_k)B]^{-1} B^T \nabla f(By_k)$$

$$= B^{-1} x_k - B^{-1} [\nabla^2 f(x_k)]^{-1} \nabla f(x_k) = B^{-1} x_{k+1}. \qquad \square$$

Thus, the Newton's method is *affine invariant* with respect to affine transformations of variables. Therefore, its actual region of quadratic convergence *does not depend* on a particular choice of the basis. It depends only on the local topological structure of the function $f(\cdot)$.

Let us try to understand what was wrong in our assumptions. The main assumption is related to the Lipschitz continuity of the Hessians:

$$\| \nabla^2 f(x) - \nabla^2 f(y) \| \leq M \| x - y \|, \quad \forall x, y \in \mathbb{R}^n.$$

Let us assume that $f \in C^3(\mathbb{R}^n)$. Define

$$f'''(x)[u] = \lim_{\alpha \to 0} \tfrac{1}{\alpha} [\nabla^2 f(x + \alpha u) - \nabla^2 f(x)] \equiv D^3 f(x)[h].$$

The object in the right-hand side of this equality (and, consequently, in its left-hand side) is an $(n \times n)$-matrix. Thus, our assumption is equivalent to the condition

$$\| f'''(x)[u] \| \leq M \| u \|.$$

This means that at any point $x \in \mathbb{R}^n$, we have

$$\langle f'''(x)[u]v, v \rangle \equiv D^3 f(x)[u, v, v] \leq M \| u \| \cdot \| v \|^2 \quad \forall u, v \in \mathbb{R}^n.$$

Note that the value in the left-hand side of this inequality is invariant with respect to affine transformations of variables (since this is just a third directional derivative along direction u and twice along direction v). However, its right-hand side does

depend on the choice of coordinates. Therefore, the most natural way to improve our situation consists in finding an affine-invariant replacement for the standard Euclidean norm $\| \cdot \|$. The most natural candidate for such a replacement is quite evident: This is the norm defined by the Hessian $\nabla^2 f(x)$ itself, namely,

$$\| u \|^2_{\nabla^2 f(x)} = \langle \nabla^2 f(x) u, u \rangle \equiv D^2 f(x)[h, h].$$

This choice results in the definition of a *self-concordant function*.

5.1.3 Definition of Self-concordant Functions

Since we are going to work with affine-invariant objects, it is natural to get rid of coordinate representations and denote by $\mathbb{E}$ a real vector space for our variables, and by $\mathbb{E}^*$ the dual space (see Sect. 4.2.1).

Let us consider a *closed convex* function $f(\cdot) \in C^3(\text{dom } f)$ with *open* domain. By fixing a point $x \in \text{dom } f$ and direction $u \in \mathbb{E}$, we define a function

$$\phi(x; t) = f(x + tu),$$

dependent on the variable $t \in \text{dom } \phi(x; \cdot) \subseteq \mathbb{R}$. Define

$$Df(x)[u] = \phi'(x; 0) = \langle \nabla f(x), u \rangle,$$

$$D^2 f(x)[u, u] = \phi''(x; 0) = \langle \nabla^2 f(x) u, u \rangle = \| u \|^2_{\nabla^2 f(x)},$$

$$D^3 f(x)[u, u, u] = \phi'''(x; 0) = \langle D^3 f(x)[u] u, u \rangle.$$

Definition 5.1.1 A function f is called *self-concordant* if there exists a constant $M_f \geq 0$ such that the inequality

$$|D^3 f(x)[u, u, u]| \leq 2M_f \| u \|^3_{\nabla^2 f(x)} \tag{5.1.4}$$

holds for all $x \in \text{dom } f$ and $u \in \mathbb{E}$. If $M_f = 1$, the function is called *standard self-concordant*.

Note that we are going to use these functions to construct a barrier model of our problem. Our main hope is that they can be easily minimized by the Newton's method.

Let us point out an equivalent definition of self-concordant functions.

Lemma 5.1.2 *A function f is self-concordant if and only if for any $x \in \text{dom } f$ and any triple of directions $u_1, u_2, u_3 \in \mathbb{E}$ we have*

$$| D^3 f(x)[u_1, u_2, u_3] | \leq 2M_f \prod_{i=1}^{3} \| u_i \|_{\nabla^2 f(x)} . \tag{5.1.5}$$

We accept this statement without proof since it needs some special facts from the theory of tri-linear symmetric forms. For the same reason, we accept without proof the following corollary.

Corollary 5.1.1 *A function f is self-concordant if and only if for any $x \in \text{dom } f$ and any direction $u \in \mathbb{R}^n$ we have*

$$D^3 f(x)[u] \preceq 2M_f \|u\|_{\nabla^2 f(x)} \nabla^2 f(x). \tag{5.1.6}$$

In what follows, we often use Definition 5.1.1 in order to prove that some f is self-concordant. In contrast, Lemma 5.1.2 is useful for establishing different properties of self-concordant functions.

Let us consider several examples.

Example 5.1.1

1. *Linear function.* Consider the function

$$f(x) = \alpha + \langle a, x \rangle, \quad \text{dom } f = \mathbb{E}.$$

   Then

$$\nabla f(x) = a, \quad \nabla^2 f(x) = 0, \quad \nabla^3 f(x) = 0,$$

   and we conclude that $M_f = 0$.
2. *Convex quadratic function.* Consider the function

$$f(x) = \alpha + \langle a, x \rangle + \frac{1}{2} \langle Ax, x \rangle, \quad \text{dom } f = \mathbb{E},$$

   where $A = A^* \succeq 0$. Then

$$\nabla f(x) = a + Ax, \quad \nabla^2 f(x) = A, \quad \nabla^3 f(x) = 0,$$

   and we conclude that $M_f = 0$.
3. *Logarithmic barrier for a ray.* Consider a univariate function

$$f(x) = -\ln x, \quad \text{dom } f = \{x \in \mathbb{R} \mid x > 0\}.$$

Then

$$f'(x) = -\tfrac{1}{x}, \quad f''(x) = \tfrac{1}{x^2}, \quad f'''(x) = -\tfrac{2}{x^3}.$$

Therefore, $f(\cdot)$ is self-concordant with $M_f = 1$.

4. *Logarithmic barrier for an ellipsoid.* Let $A = A^* \succeq 0$. Consider the *concave* function

$$\phi(x) = \alpha + \langle a, x \rangle - \frac{1}{2} \langle Ax, x \rangle.$$

Define $f(x) = -\ln \phi(x)$, with $\mathrm{dom}\, f = \{x \in \mathbb{E} : \phi(x) > 0\}$. Then

$$Df(x)[u] = -\tfrac{1}{\phi(x)}[\langle a, u \rangle - \langle Ax, u \rangle],$$

$$D^2 f(x)[u, u] = \tfrac{1}{\phi^2(x)}[\langle a, u \rangle - \langle Ax, u \rangle]^2 + \tfrac{1}{\phi(x)}\langle Au, u \rangle,$$

$$D^3 f(x)[u, u, u] = -\tfrac{2}{\phi^3(x)}[\langle a, u \rangle - \langle Ax, u \rangle]^3$$

$$-\tfrac{3}{\phi^2(x)}[\langle a, u \rangle - \langle Ax, u \rangle]\langle Au, u \rangle.$$

Let $\omega_1 = Df(x)[u]$ and $\omega_2 = \tfrac{1}{\phi(x)}\langle Au, u \rangle$. Then

$$D^2 f(x)[u, u] = \omega_1^2 + \omega_2 \geq 0,$$

$$|D^3 f(x)[u, u, u]| = |2\omega_1^3 + 3\omega_1\omega_2|.$$

The only nontrivial case is $\omega_1 \neq 0$. Let $\xi = \omega_2/\omega_1^2$. Then

$$\frac{|D^3 f(x)[u,u,u]|}{(D^2 f(x)[u,u])^{3/2}} \leq \frac{2|\omega_1|^3 + 3|\omega_1|\omega_2}{(\omega_1^2 + \omega_2)^{3/2}} = \frac{2(1 + \frac{3}{2}\xi)}{(1+\xi)^{3/2}} \leq 2,$$

where the last inequality follows from the convexity of the function $(1 + \xi)^{3/2}$ for $\xi \geq -1$. Thus, the function f is self-concordant and $M_f = 1$.

5. It is easy to verify that none of the following univariate functions is self-concordant:

$$f(x) = e^x, \quad f(x) = \tfrac{1}{x^p}, \ x > 0, \ p > 0, \quad f(x) = |x|^p, \ p > 2.$$

However the function $f_p(x) = \tfrac{1}{2}x^2 + \tfrac{1}{px^p} - \tfrac{1}{p}$ with $p > 0$ is self-concordant for $x > 0$. Let us prove this. Indeed,

$$f_p'(x) = x - \tfrac{1}{x^{p+1}}, \quad f_p''(x) = 1 + \tfrac{p+1}{x^{p+2}} \geq 1, \quad f_p'''(x) = -\tfrac{(p+1)(p+2)}{x^{p+3}}.$$

If $x \geq 1$, then

$$|f_p'''(x)| = \frac{(p+1)(p+2)}{x^{p+2}} \leq (p+2)f_p''(x) \leq (p+2)[f_p''(x)]^{3/2}.$$

If $x \in (0, 1]$, then

$$|f_p'''(x)| = \frac{(p+1)(p+2)}{x^{p+3}} \leq (p+1)(p+2)\left(\frac{1}{x^{p+2}}\right)^{3/2}$$

$$\leq (p+1)(p+2)\left(\frac{f_p''(x)}{p+1}\right)^{3/2}.$$

Thus, we can take $M_{f_p} = \max\left\{1 + \frac{p}{2}, \frac{p+2}{2\sqrt{p+1}}\right\} = 1 + \frac{p}{2}$. Note that the function f_p is well defined as $p \to 0$. Indeed,

$$\lim_{p \to 0} f_p(x) = \tfrac{1}{2}x^2 + \lim_{p \to 0} \tfrac{1}{p}\left[e^{p \ln \frac{1}{x}} - 1\right] = \tfrac{1}{2}x^2 - \ln x.$$

6. Let $f \in C_{L_3}^{3,2}(\mathbb{R}^n)$. Assume that it is strongly convex on $\mathbb{R}^n$ with convexity parameter $\sigma_2(f)$. Then, for any $x \in \mathbb{R}^n$ and direction $u \in \mathbb{R}^n$ we have

$$D^3 f(x)[u] \preceq L_3\|u\| I_n \overset{(2.1.28)}{\preceq} L_3 \left(\frac{1}{\sigma_2(f)}\|u\|_{\nabla^2 f(x)}^2\right)^{1/2} \frac{1}{\sigma_2(f)}\nabla^2 f(x).$$

Thus, in view of Corollary 5.1.1, we can take $M_f = \frac{L_3}{2\sigma_2^{3/2}(f)}$. □

Let us now look at the main properties of self-concordant functions.

Theorem 5.1.1 *Let functions f_i be self-concordant with constants M_i, $i = 1, 2$, and let α, $\beta > 0$. Then the function $f(x) = \alpha f_1(x) + \beta f_2(x)$ is self-concordant with constant*

$$M_f = \max\left\{\tfrac{1}{\sqrt{\alpha}}M_1, \tfrac{1}{\sqrt{\beta}}M_2\right\}$$

and $\text{dom } f = \text{dom } f_1 \bigcap \text{dom } f_2$.

Proof In view of Theorem 3.1.5, f is a closed convex function. Let us fix some $x \in \text{dom } f$ and $u \in \mathbb{E}$. Then

$$|D^3 f_i(x)[u, u, u]| \leq 2M_i\left[D^2 f_i(x)[u, u]\right]^{3/2}, \quad i = 1, 2.$$

Let $\omega_i = D^2 f_i(x)[u, u] \geq 0$. Then

$$\frac{|D^3 f(x)[u,u,u]|}{[D^2 f(x)[u,u]]^{3/2}} \leq \frac{\alpha|D^3 f_1(x)[u,u,u]|+\beta|D^3 f_2(x)[u,u,u]|}{[\alpha D^2 f_1(x)[u,u]+\beta D^2 f_2(x)[u,u]]^{3/2}} \leq \frac{\alpha M_1\omega_1^{3/2}+\beta M_2\omega_2^{3/2}}{[\alpha\omega_1+\beta\omega_2]^{3/2}}.$$
$$(5.1.7)$$

The right-hand side of this inequality does not change when we replace (ω_1, ω_2) by $(t\omega_1, t\omega_2)$ with $t > 0$. Therefore, we can assume that

$$\alpha\omega_1 + \beta\omega_2 = 1.$$

Let $\xi = \alpha\omega_1$. Then the right-hand side of inequality (5.1.7) becomes equal to

$$\frac{M_1}{\sqrt{\alpha}}\xi^{3/2} + \frac{M_2}{\sqrt{\beta}}(1 - \xi)^{3/2}, \quad \xi \in [0, 1].$$

This function is convex in ξ. Therefore it attains its maximum at the end points of the interval (see Corollary 3.1.1). □

Corollary 5.1.2 *Let a function f be self-concordant with some constant M_f. If $A = A^* \succeq 0$, then the function*

$$\phi(x) = \alpha + \langle a, x \rangle + \frac{1}{2}\langle Ax, x \rangle + f(x)$$

is also self-concordant with constant $M_\phi = M_f$.

Proof We have seen that any convex quadratic function is self-concordant with zero constant. □

Corollary 5.1.3 *Let a function f be self-concordant with some constant M_f and $\alpha > 0$. Then the function $\phi(x) = \alpha f(x)$ is also self-concordant with constant $M_\phi = \frac{1}{\sqrt{\alpha}}M_f$.* □

Let us now prove that self-concordance is an affine-invariant property.

Theorem 5.1.2 *Let $\mathscr{A}(x) = Ax + b: \mathbb{E} \to \mathbb{E}_1$ be a linear operator. Assume that a function $f(\cdot)$ is self-concordant with constant M_f. Then the function*

$$\phi(x) = f(\mathscr{A}(x))$$

is also self-concordant and $M_\phi = M_f$.

Proof The function $\phi(\cdot)$ is closed and convex in view of Theorem 3.1.6. Let us fix some $x \in \operatorname{dom}\phi = \{x : \mathscr{A}(x) \in \operatorname{dom} f\}$ and $u \in \mathbb{E}$. Define $y = \mathscr{A}(x)$, $v = Au$. Then

$$D\phi(x)[u] = \langle \nabla f(\mathscr{A}(x)), Au \rangle = \langle \nabla f(y), v \rangle,$$

$$D^2\phi(x)[u, u] = \langle \nabla^2 f(\mathscr{A}(x))Au, Au \rangle = \langle \nabla^2 f(y)v, v \rangle,$$

$$D^3\phi(x)[u, u, u] = D^3 f(\mathscr{A}(x))[Au, Au, Au] = D^3 f(y)[v, v, v].$$

Therefore,

$$| D^3\phi(x)[u, u, u] | = | D^3 f(y)[v, v, v] | \le 2M_f \langle \nabla^2 f(y)v, v \rangle^{3/2}$$
$$= 2M_f (D^2\phi(x)[u, u])^{3/2}. \qquad \square$$

Finally, let us describe the behavior of a self-concordant function near the boundary of its domain.

Theorem 5.1.3 *Let f be a self-concordant function. Then for any $\bar{x} \in \partial(dom\, f)$ and any sequence*

$$\{x_k\} \subset dom\, f : \quad x_k \to \bar{x}$$

we have $f(x_k) \to +\infty$.

Proof Since f is a closed convex function with open domain, this statement follows from Item 2 of Theorem 3.1.4. $\square$

Thus, f is a *barrier function* for $\mathrm{cl}\,(dom\, f)$ (see Sect. 1.3.3). Finally, let us establish the self-concordance of a logarithmic barrier for the level set of self-concordant function.

Theorem 5.1.4 *Let a function f be self-concordant with constant M_f and $f(x) \ge f^*$ for all $x \in dom\, f$. For arbitrary $\beta > f^*$, consider the function*

$$\phi(x) = -\ln(\beta - f(x)).$$

Then

1. *ϕ is well defined on $dom\,\phi = \{x \in dom\, f : f(x) < \beta\}$.*
2. *For any $x \in dom\,\phi$ and $h \in \mathbb{E}$ we have*

$$\langle \nabla^2\phi(x)h, h \rangle \ge \langle \nabla\phi(x), h \rangle^2. \qquad (5.1.8)$$

3. *ϕ is self-concordant with constant $M_\phi = \sqrt{1 + M_f^2(\beta - f^*)}$.*

Proof Let us fix $x \in dom\,\phi$ and $h \in \mathbb{E}$. Consider the function $\psi(\tau) = \phi(x + \tau h)$. Define $\omega = \beta - f(x)$. Then

$$\psi'(0) = \frac{1}{\omega}\langle \nabla f(x), h \rangle, \quad \psi''(0) = \frac{1}{\omega}\langle \nabla^2 f(x)h, h \rangle + \frac{1}{\omega^2}\langle \nabla f(x), h \rangle^2,$$

$$\psi'''(0) = \frac{1}{\omega}D^3 f(x)[h, h, h] + \frac{3}{\omega^2}\langle \nabla^2 f(x)h, h \rangle\langle \nabla f(x), h \rangle + \frac{2}{\omega^3}\langle \nabla f(x), h \rangle^3.$$

Thus, $\psi''(0) \ge (\psi'(0))^2$, and this is inequality (5.1.8).

Further, we need to bound $\psi'''(0)$ from above by $\psi''(0)^{3/2}$. Since f is self-concordant, we have

$$\psi'''(0) \overset{(5.1.4)}{\leq} \frac{2M_f}{\omega}\langle \nabla^2 f(x)h, h\rangle^{3/2} + \frac{3}{\omega^2}\langle \nabla^2 f(x)h, h\rangle\langle \nabla f(x), h\rangle$$

$$+\frac{2}{\omega^3}\langle \nabla f(x), h\rangle^3.$$

The right-hand side of this inequality is homogeneous in h of degree three. Therefore, let us find an upper bound for it assuming that $\psi''(0) = 1$. Defining

$$\tau = \left(\frac{1}{\omega}\langle \nabla^2 f(x)h, h\rangle\right)^{1/2}, \quad \xi = \frac{1}{\omega}\langle \nabla f(x), h\rangle,$$

we come to the following maximization problem:

$$\max_{\tau,\xi\in\mathbb{R}}\left\{2\hat{\omega}^{1/2}\tau^3 + 3\tau^2\xi + 2\xi^3 : \tau^2 + \xi^2 = 1\right\},$$

where $\hat{\omega} = M_f^2\omega$. Note that the optimal values of τ and ξ in this problem are non-negative. Therefore, in view of the equality constraint, we can rewrite the objective function as follows.

$$2\hat{\omega}^{1/2}\tau^3 + 3\tau^2\xi + 2\xi^3 = 2\hat{\omega}^{1/2}\tau^3 + \tau^2\xi + 2\xi(\tau^2 + \xi^2) = 2\hat{\omega}^{1/2}\tau^3 + (\tau^2 + 2)\xi$$

$$= 2\hat{\omega}^{1/2}\tau^3 + (\tau^2 + 2)\sqrt{1 - \tau^2}.$$

The first-order optimality condition for this univariate function can be written as follows:

$$0 = 6\hat{\omega}^{1/2}\tau^2 + 2\tau\sqrt{1 - \tau^2} - (\tau^2 + 2)\frac{\tau}{\sqrt{1-\tau^2}} = 6\hat{\omega}^{1/2}\tau^2 - \frac{3\tau^3}{\sqrt{1-\tau^2}}.$$

Thus, the optimal value τ_* satisfies equation $2\hat{\omega}^{1/2} = \frac{\tau_*}{\sqrt{1-\tau_*^2}}$. Hence, $\tau_* = \sqrt{\frac{4\hat{\omega}}{1+4\hat{\omega}}}$. Substituting this value into the objective function, we come to the following bound:

$$2\hat{\omega}^{1/2}\left(\frac{4\hat{\omega}}{1+4\hat{\omega}}\right)^{3/2} + \frac{2+12\hat{\omega}}{(1+4\hat{\omega})^{3/2}} = \frac{2+12\hat{\omega}+16\hat{\omega}^2}{(1+4\hat{\omega})^{3/2}} = 2\frac{1+2\hat{\omega}}{(1+4\hat{\omega})^{1/2}} \leq 2\sqrt{1+\hat{\omega}}.$$

It remains to note that $\hat{\omega} \leq M_f^2(\beta - f^*)$. $\square$

5.1.4   Main Inequalities

Let f be a self-concordant function. Define

$$\| h \|_x = \langle \nabla^2 f(x)h, h \rangle^{1/2}.$$

We call $\| h \|_x$ the *(primal) local norm* of direction h with respect to x. Let us fix a point $x \in \mathrm{dom}\, f$ and a direction $h \in \mathbb{E}$ such that $\langle \nabla^2 f(x)h, h \rangle > 0$. Consider the univariate function

$$\phi(t) = \frac{1}{\langle \nabla^2 f(x+th)h, h \rangle^{1/2}}.$$

In view of the continuity of the second derivative of the function f, $0 \in \mathrm{int}\,(\mathrm{dom}\, \phi)$.

Lemma 5.1.3 *For all feasible t, we have* $| \phi'(t) | \le M_f$.

Proof Indeed,

$$\phi'(t) = -\frac{D^3 f(x+th)[h,h,h]}{2\langle \nabla^2 f(x+tu)h, h \rangle^{3/2}}.$$

Therefore $| \phi'(t) | \le M_f$ in view of Definition 5.1.1.   □

Corollary 5.1.4 *The domain of function* $\phi(\cdot)$ *contains the interval*

$$I_x = \left(-\frac{1}{M_f}\phi(0),\ \frac{1}{M_f}\phi(0) \right).$$

Proof Indeed, in view of Lemma 5.1.3, the values $\langle \nabla^2 f(x+\tau h)h, h \rangle$ are positive at any subinterval of I_x and $\phi(t) \ge \phi(0) - M_f | t |$. Moreover, since $f(x+th) \to \infty$ as the points $x+th$ approach the boundary of dom f (see Theorem 5.1.3), the cannot intersect the boundary as $t \in I_x$.   □

Let us consider the following ellipsoids:

$$W^0(x; r) = \{y \in \mathbb{E} \mid \| y - x \|_x < r\},$$

$$W(x; r) = \mathrm{cl}\left(W^0(x; r) \right) = \{y \in \mathbb{E} \mid \| y - x \|_x \le r\}.$$

This set is called the *Dikin ellipsoid* of the function f at x.

Theorem 5.1.5 *1. For any* $x \in \mathrm{dom}\, f$, *we have* $W^0(x; \frac{1}{M_f}) \subseteq \mathrm{dom}\, f$.
2. For all $x, y \in \mathrm{dom}\, f$, *the following inequality holds:*

$$\| y - x \|_y \ge \frac{\|y-x\|_x}{1+M_f\|y-x\|_x}. \tag{5.1.9}$$

3. If $\| y - x \|_x < \frac{1}{M_f}$, *then*

$$\| y - x \|_y \leq \frac{\|y-x\|_x}{1-M_f\|y-x\|_x}. \tag{5.1.10}$$

Proof 1. Let us choose in $\mathbb{E}$ a Euclidean norm $\| \cdot \|$ and small $\epsilon > 0$. Consider the function $f_\epsilon(x) = f(x) + \frac{1}{2}\epsilon\|x\|^2$. In view of Corollary 5.1.2, it is self-concordant with constant M_f. Moreover, for any $h \in \mathbb{E}$ we have $\langle \nabla^2 f_\epsilon(x)h, h \rangle > 0$. Therefore, in view of Corollary 5.1.4, dom $f_\epsilon \equiv$ dom f contains the set

$$\left\{ y = x + th \mid t^2(\| h \|_x^2 + \epsilon\|h\|^2) < \frac{1}{M_f^2} \right\}$$

(since $\phi(0) = 1/\langle \nabla^2 f_\epsilon(x)h, h \rangle^{1/2}$). Since ϵ can be arbitrarily small, this means that dom f contains $W^0(x; \frac{1}{M_f})$.

2. Let us choose $h = y - x$. Assume for a moment that $\|h\|_x > 0$. Then

$$\phi(1) = \frac{1}{\|y-x\|_y}, \quad \phi(0) = \frac{1}{\|y-x\|_x},$$

and $\phi(1) \leq \phi(0) + M_f$ in view of Lemma 5.1.3. This is inequality (5.1.9).

3. If $\| y - x \|_x < \frac{1}{M_f}$, then $\phi(0) > M_f$, and in view of Lemma 5.1.3 $\phi(1) \geq \phi(0) - M_f$. This is inequality (5.1.10).

In the case when $\|h\|_x = 0$, both items can be justified by the trick used in the proof of Item 1. $\quad\square$

The next statement demonstrates that some local properties of self-concordant functions reflect somehow the global properties of its domain.

Theorem 5.1.6 *Let a function f be self-concordant and dom f contains no straight lines. Then the Hessian $\nabla^2 f(x)$ is nondegenerate at all points $x \in$ dom f.*

Proof Assume that $\langle \nabla^2 f(\bar{x})h, h \rangle = 0$ for some $\bar{x} \in$ dom f and direction $h \in \mathbb{E}$, $h \neq 0$. Then, all points of the line $\{x = \bar{x} + \tau h, \ \tau \in \mathbb{R}\}$ belong to the ellipsoid $W^0(x; \frac{1}{M_f})$. However, in view of Item 1 of Theorem 5.1.5, this ellipsoid belongs to dom f. This contradicts the conditions of the theorem. $\quad\square$

Theorem 5.1.7 *Let $x \in$ dom f. Then for any $y \in W^0(x; \frac{1}{M_f})$ we have*

$$(1 - M_f r)^2 \, \nabla^2 f(x) \preceq \nabla^2 f(y) \preceq \frac{1}{(1-M_f r)^2}\nabla^2 f(x), \tag{5.1.11}$$

where $r = \| y - x \|_x$.

Proof Let us fix an arbitrary direction $h \in \mathbb{E}$, $h \neq 0$. Consider the function

$$\psi(t) = \langle \nabla^2 f(x + t(y - x))h, h \rangle, \quad t \in [0, 1].$$

Define $y_t = x + t(y - x)$ and $r = \| y - x \|_x$. Then, in view of Lemma 5.1.2 and inequality (5.1.10), we have

$$| \psi'(t) | = | D^3 f(y_t)[y - x, h, h] | \leq 2M_f \| y - x \|_{y_t} \| h \|_{y_t}^2$$

$$= \tfrac{2M_f}{t} \| y_t - x \|_{y_t} \psi(t) \leq \tfrac{2M_f}{t} \cdot \tfrac{\|y_t - x\|_x}{1 - M_f \|y_t - x\|_x} \cdot \psi(t)$$

$$= \tfrac{2M_f r}{1 - tM_f r} \psi(t).$$

If $\| y - x \|_x = 0$, then $\psi(t) = \psi(0)$, $t \in [0, 1]$, and therefore

$$(1 - M_f r)^2 \psi(0) \leq \psi(t) \leq \tfrac{1}{(1 - M_f r)^2} \psi(0). \tag{5.1.12}$$

If $r > 0$, then $2(\ln(1 - tM_f r))' \leq (\ln \psi(t))' \leq -2(\ln(1 - tM_f r))'$ for all $t \in [0, 1]$. Integrating these inequalities in $t \in [0, 1]$, we get again (5.1.12), which is equivalent to (5.1.11) since h was chosen arbitrarily. $\square$

Corollary 5.1.5 *Let $x \in dom\, f$ and $r = \| y - x \|_x < \tfrac{1}{M_f}$. Then we can bound the operator*

$$G = \int_0^1 \nabla^2 f(x + \tau(y - x)) d\tau$$

as follows:

$$\left(1 - M_f r + \tfrac{1}{3} M_f^2 r^2\right) \nabla^2 f(x) \preceq G \preceq \tfrac{1}{1 - M_f r} \nabla^2 f(x).$$

Proof Indeed, in view of Theorem 5.1.7 we have

$$G = \int_0^1 \nabla^2 f(x + \tau(y - x)) d\tau \succeq \nabla^2 f(x) \cdot \int_0^1 (1 - \tau M_f r)^2 d\tau$$

$$= (1 - M_f r + \tfrac{1}{3} M_f^2 r^2) \nabla^2 f(x),$$

and $G \preceq \nabla^2 f(x) \cdot \int_0^1 \tfrac{d\tau}{(1 - \tau M_f r)^2} = \tfrac{1}{1 - M_f r} \nabla^2 f(x)$. $\square$

Remark 5.1.1 The statement of Corollary 5.1.5 remains valid for $r = \| y - x \|_y$.

Let us now recall the most important facts we have already proved.

- At any point $x \in \operatorname{dom} f$ we can define an *ellipsoid*

$$W^0 \left(x; \tfrac{1}{M_f} \right) = \left\{ x \in \mathbb{E} \mid \langle \nabla^2 f(x)(y - x), y - x) \rangle < \tfrac{1}{M_f^2} \right\},$$

  belonging to dom f.
- Inside the ellipsoid $W(x; r)$ with $r \in [0, \tfrac{1}{M_f})$ the function f is almost quadratic:

$$(1 - M_f r)^2 \nabla^2 f(x) \preceq \nabla^2 f(y) \preceq \tfrac{1}{(1 - M_f r)^2} \nabla^2 f(x)$$

  for all $y \in W(x; r)$. Choosing r small enough, we can make the quality of quadratic approximation acceptable for our goals.

These two facts form the basis for all consequent results.

Let us now prove several inequalities related to the divergence of the value of a self-concordant function with respect to its linear approximation.

Theorem 5.1.8 *For any x, $y \in \operatorname{dom} f$, we have*

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \tfrac{\|y - x\|_x^2}{1 + M_f \|y - x\|_x}, \tag{5.1.13}$$

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{1}{M_f^2} \omega(M_f \| y - x \|_x), \tag{5.1.14}$$

where $\omega(t) = t - \ln(1 + t)$.

Proof Let $y_\tau = x + \tau(y - x)$, $\tau \in [0, 1]$, and $r = \| y - x \|_x$. Then, in view of (5.1.9) we have

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle = \int_0^1 \langle \nabla^2 f(y_\tau)(y - x), y - x \rangle d\tau$$

$$= \int_0^1 \tfrac{1}{\tau^2} \| y_\tau - x \|_{y_\tau}^2 \, d\tau$$

$$\geq \int_0^1 \tfrac{r^2}{(1 + \tau M_f r)^2} d\tau \; = \; \tfrac{r}{M_f} \int_0^{M_f r} \tfrac{1}{(1 + t)^2} dt \; = \; \tfrac{r^2}{1 + M_f r}.$$

Further, using (5.1.13), we obtain

$$f(y) - f(x) - \langle \nabla f(x), y - x \rangle = \int_0^1 \langle \nabla f(y_\tau) - \nabla f(x), y - x \rangle d\tau$$

$$= \int_0^1 \frac{1}{\tau} \langle \nabla f(y_\tau) - \nabla f(x), y_\tau - x \rangle d\tau \geq \int_0^1 \frac{\|y_\tau - x\|_x^2}{\tau(1 + M_f \|y_\tau - x\|_x)} d\tau = \int_0^1 \frac{\tau r^2}{1 + \tau M_f r} d\tau$$

$$= \frac{1}{M_f^2} \int_0^{M_f r} \frac{t \, dt}{1 + t} = \frac{1}{M_f^2} \omega(M_f r). \qquad \square$$

Theorem 5.1.9 *Let $x \in \operatorname{dom} f$ and $\| y - x \|_x < \frac{1}{M_f}$. Then*

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \leq \frac{\|y - x\|_x^2}{1 - M_f \|y - x\|_x}, \tag{5.1.15}$$

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{M_f^2} \omega_*(M_f \| y - x \|_x), \tag{5.1.16}$$

where $\omega_*(t) = -t - \ln(1 - t)$.

Proof Let $y_\tau = x + \tau(y - x)$, $\tau \in [0, 1]$, and $r = \| y - x \|_x$. Since $\| y_\tau - x \| < \frac{1}{M_f}$, in view of (5.1.10) we have

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle = \int_0^1 \langle \nabla^2 f(y_\tau)(y - x), y - x \rangle d\tau$$

$$= \int_0^1 \frac{1}{\tau^2} \| y_\tau - x \|_{y_\tau}^2 \, d\tau$$

$$\leq \int_0^1 \frac{r^2}{(1 - \tau M_f r)^2} d\tau = \frac{r}{M_f} \int_0^{M_f r} \frac{1}{(1 - t)^2} dt = \frac{r^2}{1 - M_f r}.$$

Further, using (5.1.15), we obtain

$$f(y) - f(x) - \langle \nabla f(x), y - x \rangle = \int_0^1 \langle \nabla f(y_\tau) - \nabla f(x), y - x \rangle d\tau$$

$$= \int_0^1 \frac{1}{\tau} \langle \nabla f(y_\tau) - \nabla f(x), y_\tau - x \rangle d\tau \leq \int_0^1 \frac{\|y_\tau - x\|_x^2}{\tau(1 - M_f \|y_\tau - x\|_x)} d\tau = \int_0^1 \frac{\tau r^2}{1 - \tau M_f r} d\tau$$

$$= \frac{1}{M_f^2} \int_0^{M_f r} \frac{t \, dt}{1 - t} = \frac{1}{M_f^2} \omega_*(M_f r). \qquad \square$$

Theorem 5.1.10 *Inequalities (5.1.9), (5.1.10), (5.1.13), (5.1.14), (5.1.15) and (5.1.16) are necessary and sufficient characteristics of self-concordant functions.*

Proof We have already justified two sequences of implications:

$$\text{Definition } 5.1.1 \Rightarrow (5.1.9) \Rightarrow (5.1.13) \Rightarrow (5.1.14),$$

$$\text{Definition } 5.1.1 \Rightarrow (5.1.10) \Rightarrow (5.1.15) \Rightarrow (5.1.16).$$

Let us prove the implication $(5.1.14) \Rightarrow$ Definition 5.1.1. Let $x \in \text{dom} f$ and $x - \alpha u \in \text{dom} f$ for $\alpha \in [0, \epsilon)$. Consider the function

$$\psi(\alpha) = f(x - \alpha u), \quad \alpha \in [0, \epsilon).$$

Let $r = \|u\|_x \equiv [\psi''(0)]^{1/2}$. Assuming that (5.1.14) holds for all x and y from dom f, we have

$$\psi(\alpha) - \psi(0) - \psi'(0)\alpha - \tfrac{1}{2}\psi''(0)\alpha^2 \geq \tfrac{1}{M_f^2}\omega(\alpha M_f r) - \tfrac{1}{2}\alpha^2 r^2.$$

Therefore

$$\tfrac{1}{6}\psi'''(0) = \lim_{\alpha \downarrow 0} \tfrac{1}{\alpha^3}\left[\psi(\alpha) - \psi(0) - \psi'(0)\alpha - \tfrac{1}{2}\psi''(0)\alpha^2\right]$$

$$\geq \lim_{\alpha \downarrow 0} \tfrac{1}{\alpha^3}\left[\tfrac{1}{M_f^2}\omega(\alpha M_f r) - \tfrac{1}{2}\alpha^2 r^2\right] = \lim_{\alpha \downarrow 0} \tfrac{r}{3\alpha^2}\left[\tfrac{1}{M_f}\omega'(\alpha M_f r) - \alpha r\right]$$

$$= \lim_{\alpha \downarrow 0} \tfrac{r}{3\alpha^2}\left[\tfrac{\alpha r}{1 + \alpha M_f r} - \alpha r\right] = -\tfrac{1}{3}M_f r^3.$$

Therefore, $D^3 f(x)[u, u, u] = -\psi'''(0) \leq 2M_f[\psi''(0)]^{3/2}$ and this is Definition 5.1.1. Implication $(5.1.16) \Rightarrow$ Definition 5.1.1 can be proved in a similar way. □

Sometimes Theorem 5.1.10 is convenient for establishing self-concordance of certain functions. Let us demonstrate this with an *Implicit Function Theorem*.

Let us assume that $\mathbb{E} = \mathbb{E}_1 \times \mathbb{E}_2$. Thus, we have a corresponding partition of variable $z = (x, y) \in \mathbb{E}$. Let Φ be a self-concordant function with dom $\Phi \subseteq \mathbb{E}$. Consider the following implicit function:

$$f(x) = \min_{y}\{\Phi(x, y) : (x, y) \in \text{dom} \Phi\}. \tag{5.1.17}$$

In order to simplify the situation, let us assume that for any x such that the set $Q(x) = \{y : (x, y) \in \operatorname{dom} \Phi\}$ is nonempty, it does not contain a straight line. Then simple conditions, like boundedness of Φ from below, guarantee existence of the unique solution $y(x)$ of the optimization problem in (5.1.17) (see Sect. 5.1.5).

Anyway, let us assume existence of point $y(x)$. Then it is characterized by the first-order optimality condition:

$$\nabla_y \Phi(x, y(x)) = 0. \tag{5.1.18}$$

Moreover, by Theorem 3.1.25 and Lemma 3.1.10, we have

$$\nabla f(x) = \nabla_x \Phi(x, y(x)). \tag{5.1.19}$$

Let us compute the Hessian of the function f. Differentiating equation (5.1.18) along direction $h \in \mathbb{E}_1$, we get

$$\nabla_{yx}^2 \Phi(x, y(x))h + \nabla_{yy}^2 \Phi(x, y(x))y'(x)h = 0.$$

Therefore, by differentiating equality (5.1.19) along direction h, we obtain

$$\nabla^2 f(x)h = \nabla_{xx}^2 \Phi(x, y(x))h + \nabla_{xy}^2 \Phi(x, y(x))y'(x)h$$

$$= \nabla_{xx}^2 \Phi(x, y(x))h - \nabla_{xy}^2 \Phi(x, y(x))[\nabla_{yy}^2 \Phi(x, y(x))]^{-1}\nabla_{yx}^2 \Phi(x, y(x))h. \tag{5.1.20}$$

Theorem 5.1.11 *Let Φ be a self-concordant function. Then the function f defined by (5.1.17) is also self-concordant with constant M_Φ.*

Proof Let us fix $\bar{x} \in \operatorname{dom} f$. Define $\bar{z} = (\bar{x}, y(\bar{x}))$ and let $x \in \operatorname{dom} f$. Then with $z = (x, y)$, we have

$$f(x) = \min_{y \in Q(x)} \Phi(x, y)$$

$$\overset{(5.1.14)}{\geq} \min_{y \in Q(x)} \left\{ \Phi(\bar{x}, y(\bar{x})) + \langle \nabla \Phi(\bar{x}, y(\bar{x})), z - \bar{z} \rangle + \frac{1}{M_f^2}\omega(M_f\|z - \bar{z}\|_{\bar{z}}) \right\}$$

$$\overset{(5.1.19)}{=} f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle_{E_1} + \frac{1}{M_f^2}\omega\left(M_f \min_{y \in Q(x)} \|z - \bar{z}\|_{\bar{z}} \right).$$

It remains to compute the minimum in the last line. Let $h = x - \bar{x}$. Then

$$\min_{y \in Q(x)} \|z - \bar{z}\|_{\bar{z}}^2 = \langle \nabla^2_{xx} \Phi(\bar{z})h, h \rangle_{\mathbb{E}_1}$$

$$+ \min_{y \in Q(x)} \left\{ 2\langle \nabla^2_{xy} \Phi(\bar{z})(y - \bar{y}), h \rangle_{\mathbb{E}_1} + \langle \nabla^2_{yy} \Phi(\bar{z})(y - \bar{y}), y - \bar{y} \rangle_{\mathbb{E}_2} \right\}$$

$$\geq \langle \nabla^2_{xx} \Phi(\bar{z})h, h \rangle_{\mathbb{E}_1} + \min_{\delta \in \mathbb{E}_2} \left\{ 2\langle \nabla^2_{xy} \Phi(\bar{z})\delta, h \rangle_{\mathbb{E}_1} + \langle \nabla^2_{yy} \Phi(\bar{z})\delta, \delta \rangle_{\mathbb{E}_2} \right\}$$

$$= \langle \nabla^2_{xx} \Phi(\bar{z})h, h \rangle_{\mathbb{E}_1} - \langle [\nabla^2_{yy} \Phi(\bar{z})]^{-1} \nabla^2_{yx} \Phi(\bar{z})h, \nabla^2_{yx} \Phi(\bar{z})h \rangle_{\mathbb{E}_1}$$

$$\overset{(5.1.20)}{=} \langle \nabla^2 f(\bar{x})h, h \rangle.$$

It remains to apply Theorem 5.1.10. $\quad\square$

Let us prove two more inequalities. From now on, we assume that dom f contains no straight lines. In this case, in view of Theorem 5.1.6, all Hessians $\nabla^2 f(x)$ with $x \in \text{dom } f$ are nondegenerate. Denote by

$$\| g \|_x^* = \langle g, [\nabla^2 f(x)]^{-1} g \rangle^{1/2}, \quad g \in \mathbb{E}^*,$$

the *dual local norm*. Clearly, $| \langle g, h \rangle | \leq \| g \|_x^* \cdot \| h \|_x$.

Theorem 5.1.12 *For any x and y from dom f we have*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{1}{M_f^2} \omega(M_f \|\nabla f(y) - \nabla f(x)\|_y^*). \quad (5.1.21)$$

If in addition $\|\nabla f(y) - \nabla f(x)\|_y^* < \tfrac{1}{M_f}$, then

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{1}{M_f^2} \omega_*(M_f \|\nabla f(y) - \nabla f(x)\|_y^*). \quad (5.1.22)$$

Proof Let us fix arbitrary points x and y from dom f. Consider the function

$$\phi(z) = f(z) - \langle \nabla f(x), z \rangle, \quad z \in \text{dom } f.$$

Note that this function is self-concordant and $\nabla \phi(x) = 0$. Therefore, using inequality (5.1.16), we get

$$f(x) - \langle \nabla f(x), x \rangle = \phi(x) = \min_{z \in \text{dom } f} \phi(z)$$

$$\leq \min_z \left\{ \phi(y) + \langle \nabla \phi(y), z - y \rangle + \tfrac{1}{M_f^2} \omega_*(M_f \|z - y\|_y) : \|z - y\|_y < \tfrac{1}{M_f} \right\}$$

$$= \min_{0 \le \tau < 1} \left\{ \phi(y) - \tfrac{\tau}{M_f} \|\nabla \phi(y)\|_y^* + \tfrac{1}{M_f^2} \omega_*(\tau) \right\} = \phi(y) - \tfrac{1}{M_f^2} \omega(M_f \|\nabla \phi(y)\|_y^*)$$

$$= f(y) - \langle \nabla f(x), y \rangle - \tfrac{1}{M_f^2} \omega(M_f \|\nabla f(y) - \nabla f(x)\|_y^*),$$

and this is inequality (5.1.21). In order to prove inequality (5.1.22), we use a similar reasoning based on inequality (5.1.14). □

All theorems above are written in terms of two auxiliary univariate functions,

$$\omega(t) = t - \ln(1 + t), \quad \omega_*(\tau) = -\tau - \ln(1 - \tau).$$

Note that

$$\omega'(t) = \tfrac{t}{1+t} \ge 0, \quad \omega''(t) = \tfrac{1}{(1+t)^2} > 0,$$

$$\omega_*'(\tau) = \tfrac{\tau}{1-\tau} \ge 0, \quad \omega_*''(\tau) = \tfrac{1}{(1-\tau)^2} > 0.$$

Therefore, $\omega(\cdot)$ and $\omega_*(\cdot)$ are convex functions. In what follows, we often use different relations between these objects. Let us provide them with a formal justification.

Lemma 5.1.4 *For any $t \ge 0$ and $\tau \in [0, 1)$, we have*

$$\omega'(\omega_*'(\tau)) = \tau, \quad \omega_*'(\omega'(t)) = t,$$

$$\omega(t) = \max_{0 \le \xi < 1} [\xi t - \omega_*(\xi)], \quad \omega_*(\tau) = \max_{\xi \ge 0} [\xi \tau - \omega(\xi)],$$

$$\omega(t) + \omega_*(\tau) \ge \tau t,$$

$$\omega_*(\tau) = \tau \omega_*'(\tau) - \omega(\omega_*'(\tau)), \quad \omega(t) = t \omega'(t) - \omega_*(\omega'(t)).$$

We leave the proof of this lemma as an exercise for the reader. Note that the main reason for the above relations is that functions $\omega(t)$ and $\omega_*(t)$ are *Fenchel conjugate* (see definition (3.1.27)).

Functions $\omega(\cdot)$ and $\omega_*(\cdot)$ will often be used for estimating the rate of growth of self-concordant functions. Sometimes, it is more convenient to replace them by appropriate lower and upper bounds.

Lemma 5.1.5 *For any $t \ge 0$ we have*

$$\tfrac{t^2}{2(1+t)} \le \tfrac{t^2}{2\left(1+\frac{2}{3}t\right)} \le \omega(t) \le \tfrac{t^2}{2+t}, \tag{5.1.23}$$

and for $t \in [0, 1)$,

$$\frac{t^2}{2-t} \leq \omega_*(t) \leq \frac{t^2}{2(1-t)}. \tag{5.1.24}$$

Proof Let $\psi_1(t) = \frac{t^2}{2\left(1+\frac{2}{3}t\right)}$. Note that $\psi_1(0) = \omega(0) = 0$. At the same time,

$$\psi_1'(t) = \frac{t}{1+\frac{2}{3}t} - \frac{t^2}{3\left(1+\frac{2}{3}t\right)^2} = \frac{t(3+t)}{3\left(1+\frac{2}{3}t\right)^2} \leq \frac{t}{1+t} = \omega'(t).$$

Similarly, for $\psi_2(t) = \frac{t^2}{2+t}$, we have

$$\psi_2'(t) = \frac{2t}{2+t} - \frac{t^2}{(2+t)^2} = \frac{4t+t^2}{(2+t)^2} \geq \frac{t}{1+t} = \omega'(t).$$

For the second inequality, let $\psi_3(t) = \frac{t^2}{2-t}$ and $\psi_4(t) = \frac{t^2}{2(1-t)}$. Then

$$\psi_3'(t) = \frac{2t}{2-t} + \frac{t^2}{(2-t)^2} = \frac{4t-t^2}{(2-t)^2} \leq \frac{t}{1-t},$$

$$\psi_4'(t) = \frac{t}{1-t} + \frac{t^2}{2(1-t)^2} = \frac{2t-t^2}{2(1-t)^2} \geq \frac{t}{1-t}.$$

Since $\frac{t}{1-t} = \omega_*'(t)$ and $\omega_*(0) = \psi_3(0) = \psi_4(0) = 0$, we get (5.1.24) by integration. □

5.1.5 Self-Concordance and Fenchel Duality

Let us start with some preliminary results. Consider the following minimization problem:

$$\min\{f(x) \mid x \in \text{dom } f\}, \tag{5.1.25}$$

where we assume that f is self-concordant and all Hessians $\nabla^2 f(x)$, $x \in \text{dom } f$, are positive definite. In view of Theorem 5.1.6, this can be derived from the fact that dom f contains no straight lines. Or, we can assume that f is strongly convex.

Define

$$\lambda_f(x) = \langle \nabla f(x), [\nabla^2 f(x)]^{-1} \nabla f(x) \rangle^{1/2}.$$

We call $\lambda_f(x) = \| \nabla f(x) \|_x^*$ the *local norm of the gradient* $\nabla f(x)$.[4]

---

[4]Sometimes $\lambda_f(x)$ is called the *Newton decrement* of the function f at x.

The next theorem provides us with a sufficient condition for existence of solution of problem (5.1.25).

Theorem 5.1.13 *Let $\lambda_f(x) < \frac{1}{M_f}$ for some $x \in \operatorname{dom} f$. Then there exists a unique solution x_f^* of problem (5.1.25) and*

$$f(x) - f(x_f^*) \leq \frac{1}{M_f^2} \omega_*(M_f \lambda_f(x)). \tag{5.1.26}$$

Proof Indeed, in view of (5.1.14), for any $y \in \operatorname{dom} f$ we have

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{M_f^2} \omega(M_f \parallel y - x \parallel_x)$$

$$\geq f(x) - \lambda_f(x) \cdot \parallel y - x \parallel_x + \frac{1}{M_f^2} \omega(M_f \parallel y - x \parallel_x)$$

$$= f(x) + \left(\frac{1}{M_f} - \lambda_f(x)\right) \|y - x\|_x - \frac{1}{M_f^2} \ln(1 + M_f \|y - x\|_x).$$

Thus, the level set $\mathscr{L}_f(f(x))$ is bounded and therefore x_f^* exists. It is unique since in view of (5.1.14), for all $y \in \operatorname{dom} f$ we have

$$f(y) \geq f(x_f^*) + \frac{1}{M_f^2} \omega(M_f \parallel y - x_f^* \parallel_{x_f^*}).$$

Finally, taking in (5.1.22) $x = x^*$ and $y = x$, we get inequality (5.1.26). $\quad\square$

Thus, we have proved that a local condition $\lambda_f(x) < \frac{1}{M_f}$ provides us with some global information on the function f, namely, the existence of the minimum x_f^*. Note that the result of Theorem 5.1.13 cannot be strengthened.

Example 5.1.2 Let us fix some $\epsilon > 0$. Consider a function of one variable

$$f_\epsilon(x) = \epsilon x - \ln x, \quad x > 0.$$

This function is self-concordant in view of Example 5.1.1 and Corollary 5.1.2. Note that

$$\nabla f_\epsilon(x) = \epsilon - \frac{1}{x}, \quad \nabla^2 f_\epsilon = \frac{1}{x^2}.$$

Therefore $\lambda_{f_\epsilon}(x) = \mid 1 - \epsilon x \mid$. Thus, for $\epsilon = 0$ we have $\lambda_{f_0}(x) = 1$ for any $x > 0$. Note that the function f_0 is not bounded below.

If $\epsilon > 0$, then $x_{f_\epsilon}^* = \frac{1}{\epsilon}$. However, we can guarantee the existence of this point by collecting information at the point $x = 1$ even if ϵ is arbitrary small. $\quad\square$

Theorem 5.1.13 has several important consequences. One of them is called the *Theorem on Recession Direction*. Note that for its validity, we *do not need* the assumption that all Hessians of the function f are positive definite.

Theorem 5.1.14 *Let $h \in \mathbb{E}$ be a recession direction of the self-concordant function f: for any $x \in$ dom f we have*

$$\langle \nabla f(x), h \rangle \leq 0,$$

and there exists a $\tau = \tau(x)$ such that $x - \tau h \in \partial$ dom f. Then

$$\langle \nabla^2 f(x)h, h \rangle^{1/2} \leq M_f \langle -\nabla f(x), h \rangle, \quad x \in \text{dom } f. \tag{5.1.27}$$

Proof Let us fix an arbitrary $x \in$ dom f. Consider a univariate function $\phi(\tau) = f(x + \tau h)$. This function is self-concordant and $0 \in$ dom ϕ. As dom ϕ contains no straight line, by Theorem 5.1.6, $\phi''(\tau) > 0$ for all $\tau \in$ dom ϕ. Therefore, we must have

$$\lambda_\phi^2(0) \equiv \frac{\langle \nabla f(x), h \rangle^2}{\langle \nabla^2 f(x)h, h \rangle} \geq \frac{1}{M_f^2}$$

since otherwise, by Theorem 5.1.13, the minimum of $\phi(\cdot)$ exists. Thus,

$$\langle \nabla f(x), h \rangle^2 \geq \frac{1}{M_f^2} \langle \nabla^2 f(x)h, h \rangle,$$

and we get (5.1.27) taking into account the sign of the first derivative. □

Let us consider now the scheme of the *Damped Newton's method*.

Damped Newton's method

---

0. Choose $x_0 \in$ dom f.
1. Iterate $x_{k+1} = x_k - \frac{1}{1+M_f \lambda_f(x_k)}[\nabla^2 f(x_k)]^{-1} \nabla f(x_k), \; k \geq 0.$

(5.1.28)

Theorem 5.1.15 *For any $k \geq 0$, we have*

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{M_f^2} \omega(M_f \lambda_f(x_k)). \tag{5.1.29}$$

Proof Let $\lambda = \lambda_f(x_k)$. Then $\| x_{k+1} - x_k \|_{x_k} = \frac{\lambda}{1+M_f\lambda} = \frac{1}{M_f}\omega'(M_f\lambda)$. Therefore, in view of (5.1.16) and Lemma 5.1.4, we have

$$f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{1}{M_f^2}\omega_*(M_f \| x_{k+1} - x_k \|_x)$$

$$= f(x_k) - \frac{\lambda^2}{1+M_f\lambda} + \frac{1}{M_f^2}\omega_*(\omega'(M_f\lambda))$$

$$= f(x_k) - \frac{\lambda}{M_f}\omega'(M_f\lambda) + \frac{1}{M_f^2}\omega_*(\omega'(M_f\lambda))$$

$$= f(x_k) - \frac{1}{M_f^2}\omega(M_f\lambda). \qquad \square$$

Thus, for all $x \in \text{dom } f$ with $\lambda_f(x) \geq \beta > 0$, one step of the damped Newton's Method decreases the value of the function $f(\cdot)$ at least by a constant $\frac{1}{M_f^2}\omega(M_f\beta) > 0$. Note that the result of Theorem 5.1.15 is *global*. In Sect. 5.2 it will be used to obtain a global efficiency bound of the process. However, now we employ it to prove an existence theorem. Recall that we assume that dom f contains no straight line.

Theorem 5.1.16 *Let a self-concordant function f be bounded below. Then it attains its minimum at a single point.*

Proof Indeed, assume that $f(x) \geq f^*$ for all $x \in \text{dom } f$. Let us start the process (5.1.28) from some $x_0 \in \text{dom } f$. If the number of steps of this method exceeds $M_f^2(f(x_0) - f^*)/\omega(1)$, then in view of (5.1.29) we must get a point x_k with $\lambda_f(x_k) < \frac{1}{M_f}$. However, by Theorem 5.1.13 this implies the existence of a point x_f^*. It is unique since all Hessians of the function f are nondegenerate. $\square$

Now we can introduce the *Fenchel dual* of a self-concordant function f (sometimes called a *conjugate function*, or *dual function* of f). For $s \in \mathbb{E}^*$, the value of this function is defined as follows:

$$f_*(s) = \sup_{x \in \text{dom } f} [\langle s, x \rangle - f(x)]. \tag{5.1.30}$$

Clearly, dom $f_* = \{s \in \mathbb{E}^* : f(x) - \langle s, x \rangle \text{ is bounded below on dom } f\}$.

Lemma 5.1.6 *The function f_* is a closed convex function with nonempty open domain. Moreover, dom $f_* = \{\nabla f(x) : x \in \text{dom } f\}$.*

Proof Indeed, for any $\bar{x} \in \text{dom } f$, we have $\nabla f(\bar{x}) \in \text{dom } f_*$. On the other hand, if $s \in \text{dom } f_*$, then $f(x) - \langle s, x \rangle$ is below bounded. Hence, by Theorem 5.1.16 and the first-order optimality condition, there exists an $x \in \text{dom } f$ such that $s = \nabla f(x)$.

Further, the epigraph of the function f_* is an intersection of half-spaces

$$\{(s, \tau) \in \mathbb{E}^* \times \mathbb{R} : \tau \geq \langle s, x \rangle - f(x)\}, \quad x \in \mathrm{dom}\, f,$$

which are closed and convex. Therefore, the epigraph of f_* is also closed and convex.

Suppose for s_1 and s_2 from $\mathrm{dom}\, f_*$ we have

$$f(x) - \langle s_1, x \rangle \geq f_1^*, \quad f(x) - \langle s_2, x \rangle \geq f_2^*$$

for all $x \in \mathrm{dom}\, f$. Then, for any $\alpha \in [0, 1]$

$$f(x) - \langle \alpha s_1 + (1 - \alpha)s_2, x \rangle = \alpha(f(x) - \langle s_1, x \rangle) + (1 - \alpha)(f(x) - \langle s_2, x \rangle)$$

$$\geq \alpha f_1^* + (1 - \alpha) f_2^*, \quad x \in \mathrm{dom}\, f.$$

Thus, $\alpha s_1 + (1 - \alpha)s_2 \in \mathrm{dom}\, f_*$.

Finally, let $s \in \mathrm{dom}\, f_*$. Denote by $x(s) \in \mathrm{dom}\, f$ the unique solution of the equation

$$s = \nabla f(x(s)).$$

Let $\delta \in \mathbb{E}^*$ be small enough: $\|\delta\|_{x(s)}^* < \frac{1}{M_f}$. Consider the function

$$f_\delta(x) = f(x) - \langle s + \delta, x \rangle.$$

Then $\nabla f_\delta(x(s)) = \nabla f(x(s)) - s - \delta = -\delta$. Therefore, $\lambda_{f_\delta}(x(s)) = \|\delta\|_{x(s)}^* < \frac{1}{M_f}$. Thus, in view of Theorem 5.1.13 the function f_δ attains its minimum. Consequently, $s + \delta \in \mathrm{dom}\, f_*$, and we conclude that s is an interior point of $\mathrm{dom}\, f_*$. $\square$

Example 5.1.3 Note that in general, the structure of the set $\{\nabla f(x) : x \in \mathrm{dom}\, f\}$ can be quite complicated. Consider the function

$$f(x) = \frac{1}{x^{(1)}} \left(x^{(2)}\right)^2, \quad \mathrm{dom}\, f = \{x \in \mathbb{R}^2 : x^{(1)} > 0\} \bigcup \{0\}, \quad f(0) = 0.$$

In Example 3.1.2(5) we have seen that this is a closed convex function. However,

$$\nabla f(x) = \left(-\left(\frac{x^{(2)}}{x^{(1)}}\right)^2, 2\frac{x^{(2)}}{x^{(1)}}\right), \quad x \neq 0, \quad \nabla f(0) = 0.$$

Thus, $\{\nabla f(x) : x \in \mathrm{dom}\, f\} = \{g \in \mathbb{R}^2 : g^{(1)} = -\frac{1}{2}(g^{(2)})^2\}$. $\square$

Let us now look at the derivatives of the function f_*. Since f is self-concordant, for any $s \in \mathrm{dom}\, f_*$, the supremum in (5.1.30) is attained (see Theorem 5.1.16).

Define

$$x(s) = \arg \max_{x \in \text{dom } f} [\langle s, x \rangle - f(x)].$$

Thus,

$$\nabla f(x(s)) = s. \tag{5.1.31}$$

In view of Lemma 3.1.14, we have $x(s) \in \partial f_*(s)$. On the other hand, for s_1 and s_2 from dom f_* we have

$$\frac{\|x(s_1) - x(s_2)\|_{x(s_1)}^2}{1 + M_f \|x(s_1) - x(s_2)\|_{x(s_1)}} \overset{(5.1.13)}{\leq} \langle \nabla f(x(s_1)) - \nabla f(x(s_2)), x(s_1) - x(s_2) \rangle$$

$$\overset{(5.1.31)}{=} \langle s_1 - s_2, x(s_1) - x(s_2) \rangle$$

$$\leq \|s_1 - s_2\|_{x(s_1)}^* \|x(s_1) - x(s_2)\|_{x(s_1)}.$$

Thus, $x(s)$ is a continuous function of s and by Lemma 3.1.10 we conclude that

$$\nabla f_*(s) = x(s). \tag{5.1.32}$$

Let us differentiate identities (5.1.31) and (5.1.32) along direction $h \in \mathbb{E}^*$:

$$\nabla^2 f(x(s)) x'(s) h = h, \quad \nabla^2 f_*(s) h = x'(s) h.$$

Thus,

$$\nabla^2 f_*(s) = [\nabla^2 f(x(s))]^{-1}, \quad s \in \text{dom } f_*. \tag{5.1.33}$$

In other words, if $s = \nabla f(x)$, then

$$\nabla^2 f_*(s) = [\nabla^2 f(x)]^{-1}, \quad x \in \text{dom } f. \tag{5.1.34}$$

Let us compute the third derivative of the dual function f_* along direction $h \in \mathbb{E}^*$ using the representation (5.1.33).

$$D^3 f_*(s)[h] = \lim_{\alpha \to 0} \frac{1}{\alpha} \left([\nabla^2 f(x(s + \alpha h))]^{-1} - [\nabla^2 f(x(s))]^{-1} \right)$$

$$= \lim_{\alpha \to 0} \frac{1}{\alpha} [\nabla^2 f(x(s))]^{-1} \left(\nabla^2 f(x(s)) - \nabla^2 f(x(s + \alpha h)) \right) [\nabla^2 f(x(s + \alpha h))]^{-1}$$

$$= -[\nabla^2 f(x(s))]^{-1} D^3 f(x(s))[x'(s)h][\nabla^2 f(x(s))]^{-1}.$$

Thus, we have proved the following representation:

$$D^3 f_*(s)[h] = \nabla^2 f_*(s) D^3 f(x(s)) \left[-\nabla^2 f_*(s)h \right] \nabla^2 f_*(s), \tag{5.1.35}$$

which is valid for all $s \in \operatorname{dom} f_*$ and $h \in \mathbb{E}^*$. Now we can prove our main statement.

Theorem 5.1.17 *The function f_* is self-concordant with $M_{f_*} = M_f$.*

Proof Indeed, in view of Lemma 5.1.6, f_* is a closed convex function with open domain. Further, for any $s \in \operatorname{dom} f_*$ and $h \in \mathbb{E}^*$ we have

$$\|\nabla^2 f_*(s)h\|_{x(s)}^2 \overset{(5.1.33)}{=} \langle h, \nabla^2 f_*(s)h \rangle \overset{\text{def}}{=} r^2.$$

Therefore, in view of (5.1.35),

$$D^3 f_*(s)[h] \overset{(5.1.6)}{\preceq} 2M_f r \, \nabla^2 f_*(s) \, \nabla^2 f(x(s)) \, \nabla^2 f_*(s) \overset{(5.1.33)}{=} 2M_f r \, \nabla^2 f_*(s).$$

It remains to use Corollary 5.1.1. □

As an example of application of Theorem 5.1.17, let us prove the following result.

Lemma 5.1.7 *Let $x, y \in \operatorname{dom} f$ and $d = \|\nabla f(x) - \nabla f(y)\|_x^* < \frac{1}{M_f}$. Then*

$$(1 - M_f d)^2 \nabla^2 f(x) \preceq \nabla^2 f(y) \preceq \frac{1}{(1-M_f d)^2} \nabla^2 f(x). \tag{5.1.36}$$

Proof Let $u = \nabla f(x)$ and $v = \nabla f(y)$. In view of Lemma 5.1.6, both points belong to $\operatorname{dom} f_*$. Note that

$$d^2 = (\|\nabla f(x) - \nabla f(y)\|_x^*)^2 = \langle u - v, \nabla^2 f_*(u)(u - v) \rangle.$$

Since f_* is self-concordant with constant M_f, by Theorem 5.1.7 we have

$$(1 - M_f d)^2 \nabla^2 f_*(u) \preceq \nabla^2 f_*(v) \preceq \frac{1}{(1-M_f d)^2} \nabla^2 f_*(u).$$

In view of (5.1.33), this is exactly (5.1.36). □

Remark 5.1.2 Some results on self-concordant functions have a more natural *dual* interpretation. Let us look at the statement of Theorem 5.1.13. Since the function f_* is self-concordant, for any $\bar{s} \in \operatorname{dom} f_*$ the ellipsoid

$$W_*^0(\bar{s}) = \left\{ s \in \mathbb{E}^* : \ \langle s - \bar{s}, \nabla^2 f_*(\bar{s})(s - \bar{s}) \rangle < \frac{1}{M_f^2} \right\}$$

belongs to dom f_*. Note that for $\bar{s} = \nabla f(x)$, in view of (5.1.33), condition $\lambda_f(x) < \frac{1}{M_f}$ is equivalent to

$$\langle \bar{s}, \nabla^2 f_*(\bar{s})\bar{s} \rangle < \frac{1}{M_f^2}.$$

This guarantees that $0 \in W_*^0(\bar{s})$. Consequently, $0 \in \text{dom } f_*$ and consequently the function f_* is below bounded. $\square$

5.2 Minimizing Self-concordant Functions

(Local convergence of different variants of Newton's Method; Path-following method; Minimization of strongly convex functions.)

5.2.1 Local Convergence of Newton's Methods

In this section, we are going to study the complexity of solving the problem (5.1.25) by different optimization strategies. Let us look first at different variants of Newton's Method.

Variants of Newton's Method

---

0. Choose $x_0 \in \text{dom } f$.
1. For $k \geq 0$, iterate

$$x_{k+1} = x_k - \frac{1}{1+\xi_k}[\nabla^2 f(x_k)]^{-1}\nabla f(x_k), \qquad (5.2.1)$$

where ξ_k is chosen in one of the following ways:

(A) $\xi_k = 0$ (this is the *Standard* Newton's Method),

(B) $\xi_k = M_f \lambda_k$ (this is the *Damped* Newton's Method (5.1.28)),

(C) $\xi_k = \frac{M_f^2 \lambda_k^2}{1+M_f \lambda_k}$ (this is the *Intermediate* Newton's Method),

where $\lambda_k = \lambda_f(x_k)$.

We call method $(5.2.1)_C$ intermediate since for big λ_k it is close to variant B, and for small values of λ_k it is very close to variant A. However, note that its step size is always bigger than the step size of variant B, which was obtained

by minimizing an upper bound for the self-concordant function (see the proof of Theorem 5.1.15). Nevertheless, method $(5.2.1)_C$ ensures a monotone decrease of the value of objective function in problem (5.1.25).

Lemma 5.2.1 *Let points* $\{x_k\}_{k\geq 0}$ *be generated by method* $(5.2.1)_C$. *Then, for any* $k \geq 0$ *we have*

$$f(x_k) - f(x_{k+1}) \geq \frac{\lambda_k^2}{2(1+M_f\lambda_k+M_f^2\lambda_k^2)} + \frac{M_f\lambda_k^3}{2(1+M_f\lambda_k)(3+2M_f\lambda_k)}. \qquad (5.2.2)$$

Proof Indeed, in view of inequality (5.1.16), we have

$$f(x_{k+1}) \leq f(x_k) - \frac{\lambda_k^2}{1+\xi_k} + \frac{1}{M_f^2}\omega_*\left(\frac{M_f\lambda_k}{1+\xi_k}\right)$$

$$= f(x_k) - \frac{\lambda_k^2(1+M_f\lambda_k)}{1+M_f\lambda_k+M_f^2\lambda_k^2} + \frac{1}{M_f^2}\left[-\frac{M_f\lambda_k(1+M_f\lambda_k)}{1+M_f\lambda_k+M_f^2\lambda_k^2} + \ln\left(1 + M_f\lambda_k + M_f^2\lambda_k^2\right)\right].$$

Defining $\tau_k = M_f\lambda_k$, we have

$$\frac{\tau_k(1+\tau_k)^2}{1+\tau_k+\tau_k^2} - \ln\left(1 + \tau_k + \tau_k^2\right) = \frac{\tau_k(1+\tau_k)^2}{1+\tau_k+\tau_k^2} - \tau_k + \omega(\tau_k) - \ln\left(1 + \frac{\tau_k^2}{1+\tau_k}\right)$$

$$\overset{(5.1.23)}{\geq} \frac{\tau_k^2}{1+\tau_k+\tau_k^2} + \frac{\tau_k^2}{2\left(1+\frac{2}{3}\tau_k\right)} - \ln\left(1 + \frac{\tau_k^2}{1+\tau_k}\right) = \frac{\tau_k^2}{2\left(1+\frac{2}{3}\tau_k\right)} - \xi_k + \frac{\xi_k}{1+\xi_k} + \omega(\xi_k).$$

It remains to note that

$$\frac{\tau_k^2}{2\left(1+\frac{2}{3}\tau_k\right)} - \frac{1}{2}\xi_k = \frac{\tau_k^2}{2}\left(\frac{1}{1+\frac{2}{3}\tau_k} - \frac{1}{1+\tau_k}\right) = \frac{\tau_k^3}{2(1+\tau_k)(3+2\tau_k)},$$

and $-\frac{\xi_k}{2} + \omega(\xi_k) \overset{(5.1.23)}{\geq} -\frac{\xi_k}{2} + \frac{\xi_k^2}{2(1+\xi_k)} = -\frac{\xi_k}{2(1+\xi_k)}$. $\square$

Let us describe now the *local* convergence of different variants of the Newton's Method. Note that we can measure the convergence of these schemes in four different ways. We can estimate the rate of convergence for the *functional gap* $f(x_k) - f(x_f^*)$, or for the local norm of the gradient $\lambda_f(x_k) = \| \nabla f(x_k) \|_{x_k}^*$, or for the *local distance to the minimum* $\| x_k - x_f^* \|_{x_k}$. Finally, we can look at the distance to the minimum in a fixed metric

$$r_*(x_k) = \|x_k - x_f^*\|_{x_f^*},$$

defined by the minimum itself. Let us prove that locally all these measures are equivalent.

Theorem 5.2.1 *Let $\lambda_f(x) < \frac{1}{M_f}$. Then*

$$\omega(M_f\lambda_f(x)) \leq M_f^2(f(x) - f(x_f^*)) \leq \omega_*(M_f\lambda_f(x)), \qquad (5.2.3)$$

$$\omega'(M_f\lambda_f(x)) \leq M_f \parallel x - x_f^* \parallel_x \leq \omega_*'(M_f\lambda_f(x)), \qquad (5.2.4)$$

$$\omega(M_f r_*(x)) \leq M_f^2(f(x) - f(x_f^*)) \leq \omega_*(M_f r_*(x)), \qquad (5.2.5)$$

where the last inequality is valid for $r_*(x) < \frac{1}{M_f}$.

Proof Let $r = \parallel x - x_f^* \parallel_x$ and $\lambda = \lambda_f(x)$. Inequalities (5.2.3) follow from Theorem 5.1.12. Further, in view of (5.1.13), we have

$$\frac{r^2}{1+M_f r} \leq \langle \nabla f(x), x - x_f^* \rangle \leq \lambda r.$$

Applying the function $\omega_*'(\cdot)$ to both sides of inequality $\frac{M_f r}{1+M_f r} \leq M_f \lambda$, we get the right-hand side of inequality (5.2.4). If $r \geq \frac{1}{M_f}$, then the left-hand side of this inequality is trivial. Suppose that $r < \frac{1}{M_f}$. Then $\nabla f(x) = G(x - x_f^*)$ with

$$G = \int\limits_0^1 \nabla^2 f(x_f^* + \tau(x - x_f^*))d\tau > 0,$$

and $\lambda_f^2(x) = \langle G[\nabla^2 f(x)]^{-1}G(x - x_f^*), x - x_f^* \rangle$. Let us introduce in $\mathbb{E}$ a canonical basis. Then all self-adjoint operators from $\mathbb{E}$ to $\mathbb{E}^*$ can be represented by symmetric matrices (we do not change the existing notation). Define

$$H = \nabla^2 f(x), \quad S = H^{-1/2}GH^{-1}GH^{-1/2} = \left(H^{-1/2}GH^{-1/2}\right)^2 \stackrel{\text{def}}{=} P^2 \succ 0.$$

Then $\|H^{1/2}(x - x_f^*)\|_2 = \|x - x_f^*\|_x = r$, where $\|\cdot\|_2$ is the standard Euclidean norm, and

$$\lambda_f(x) = \langle H^{1/2}SH^{1/2}(x-x^*), x-x^* \rangle^{1/2} \leq \parallel P \parallel_2 \|H^{1/2}(x-x^*)\|_2 = \parallel P \parallel_2 r.$$

In view of Corollary 5.1.5 (see Remark 5.1.1), we have

$$G \preceq \frac{1}{1-M_f r}H.$$

Therefore, $\parallel P \parallel_2 \leq \frac{1}{1-M_f r}$ and we conclude that

$$M_f\lambda_f(x) \leq \frac{M_f r}{1-M_f r} = \omega_*'(M_f r).$$

Applying the function $\omega'(\cdot)$ to both sides of this inequality, we get the remaining part of (5.2.4). Finally, inequalities (5.2.5) follow from (5.1.14) and (5.1.16). □

We are going to estimate the local rate of convergence of different variants of the Newton's method (5.2.1) in terms of $\lambda_f(\cdot)$, the local norm of the gradient.

Theorem 5.2.2 *Let $x \in \text{dom } f$ and $\lambda = \lambda_f(x)$.*

1. If $\lambda < \frac{1}{M_f}$ and the point x_+ is generated by variant A of method (5.2.1), then $x_+ \in \text{dom } f$ and

$$\lambda_f(x_+) \le \frac{M_f \lambda^2}{(1 - M_f \lambda)^2}. \tag{5.2.6}$$

2. If point x_+ is generated by variant B of method (5.2.1), then $x_+ \in \text{dom } f$ and

$$\lambda_f(x_+) \le M_f \lambda^2 \left(1 + \frac{1}{1 + M_f \lambda}\right). \tag{5.2.7}$$

3. If $M_f \lambda + M_f^2 \lambda^2 + M_f^3 \lambda^3 \le 1$ and point x_+ is generated by method (5.2.1)$_C$, then $x_+ \in \text{dom } f$ and

$$\lambda_f(x_+) \le M_f \lambda^2 \left(1 + M_f \lambda + \frac{M_f \lambda}{1 + M_f \lambda + M_f^2 \lambda^2}\right) \le M_f \lambda^2 \left(1 + 2M_f \lambda\right). \tag{5.2.8}$$

Proof Let $h = x_+ - x$, $\lambda = \lambda_f(x)$, and $r = \|h\|_x$. Then $r = \frac{\lambda}{1 + \xi}$. Note that for all variants of method (5.2.1), we have $M_f \lambda < 1 + \xi$. Therefore, in all cases, $M_f r < 1$ and $x_+ \in \text{dom } f$ (see Theorem 5.1.5). Hence, in view of Theorem 5.1.7 we have

$$\lambda_f(x_+) = \langle \nabla f(x_+), [\nabla^2 f(x_+)]^{-1} \nabla f(x_+) \rangle^{1/2} \le \frac{1}{1 - M_f r} \| \nabla f(x_+) \|_x^*.$$

Further, by (5.2.1)

$$\nabla f(x_+) = \nabla f(x) + \int_0^1 \nabla^2 f(x + \tau h) h d\tau = Gh,$$

where $G = \int_0^1 [\nabla^2 f(x + \tau h) - (1 + \xi)\nabla^2 f(x)]d\tau$. As in the proof of Theorem 5.2.1, let us pass to matrices. Define

$$H = \nabla^2 f(x), \quad S = H^{-1/2} G H^{-1} G H^{-1/2} \stackrel{\text{def}}{=} P^2,$$

where $P = H^{-1/2} G H^{-1/2}$. Then $\|H^{1/2} h\|_2 = \|h\|_x = r$, and

$$\| \nabla f(x_+) \|_x^* = \langle Gh, H^{-1} Gh \rangle^{1/2} = \langle H^{1/2} S H^{1/2} h, h \rangle^{1/2} \le \| P \|_2 r.$$

In view of Corollary 5.1.5,

$$\left(-\xi - M_f r + \tfrac{1}{3} M_f^2 r^2\right) H \preceq G \preceq \left(\tfrac{1}{1-M_f r} - (1+\xi)\right) H.$$

Therefore, $\| P \|_2 \leq \max\left\{\frac{M_f r}{1-M_f r} - \xi, M_f r + \xi\right\}.$

For the variant A, $\xi = 0$. Thus, $r = \lambda$ and we get $\|P\|_2 \leq \frac{M_f \lambda}{1-M_f \lambda}.$ Therefore,

$$\lambda_f(x_+) \leq \tfrac{\lambda}{1-M_f\lambda} \| P \|_2 \leq \tfrac{M_f\lambda^2}{(1-M_f\lambda)^2}.$$

For the variant B, $\xi = M_f \lambda$. Therefore, $r = \frac{\lambda}{1+M_f\lambda}$, and we get $\|P\|_2 \leq M_f\lambda + \frac{M_f\lambda}{1+M_f\lambda}$. Consequently,

$$\lambda_f(x_+) \leq \tfrac{r}{1-M_f r} \| P \|_2 \leq M_f\lambda^2 \left(1 + \tfrac{1}{1+M_f\lambda}\right).$$

Finally, for variant C, $\xi = \frac{M_f^2\lambda^2}{1+M_f\lambda}$. Then, $r = \frac{\lambda(1+M_f\lambda)}{1+M_f\lambda+M_f^2\lambda^2}$, and we have

$$\frac{M_f r}{1-M_f r} - M_f r - \xi = \frac{M_f^2 r^2}{1-M_f r} - \xi = \frac{M_f^2\lambda^2(1+M_f\lambda)^2}{1+M_f\lambda+M_f^2\lambda^2} - \frac{M_f^2\lambda^2}{1+M_f\lambda}$$

$$= \frac{M_f^2\lambda^2(2M_f\lambda+2M_f^2\lambda^2+M_f^3\lambda^3)}{(1+M_f\lambda+M_f^2\lambda^2)(1+M_f\lambda)} = \frac{\xi(2M_f\lambda+2M_f^2\lambda^2+M_f^3\lambda^3)}{1+M_f\lambda+M_f^2\lambda^2} \leq \xi$$

in view of the condition of this item. Hence

$$\lambda_f(x_+) \leq \tfrac{r}{1-M_f r} \| P \|_2 \leq \tfrac{r}{1-M_f r}(M_f r + \xi)$$

$$= \tfrac{\lambda(1+M_f\lambda)}{1+M_f\lambda+M_f^2\lambda^2}(1 + M_f\lambda + M_f^2\lambda^2)\left(\tfrac{M_f\lambda(1+M_f\lambda)}{1+M_f\lambda+M_f^2\lambda^2} + \tfrac{M_f^2\lambda^2}{1+M_f\lambda}\right)$$

$$= M_f\lambda^2 \left(\tfrac{(1+M_f\lambda)^2}{1+M_f\lambda+M_f^2\lambda^2} + M_f\lambda\right)$$

$$= M_f\lambda^2 \left(1 + M_f\lambda + \tfrac{M_f\lambda}{1+M_f\lambda+M_f^2\lambda^2}\right). \qquad \square$$

Among all variants of the rate of convergence, described in Theorem 5.2.2, the estimate (5.2.8) looks more attractive. It provides us with the following description

of the region of quadratic convergence for method $(5.2.1)_C$:

$$\mathscr{Q}_f \stackrel{\text{def}}{=} \left\{ x \in \text{dom } f : \; \lambda_f(x) < \tfrac{1}{2M_f} \right\}. \tag{5.2.9}$$

In this case, we can guarantee that $\lambda_f(x_+) < \lambda_f(x)$, and then the quadratic convergence starts (see $(5.2.8)$). Thus, our results lead to the following strategy for solving the initial problem $(5.1.25)$.

- *First stage*: $\lambda_f(x_k) \geq \tfrac{1}{2M_f}$. At this stage we apply the Damped Newton's Method $(5.1.28)$. At each iteration of this method, we have

$$f(x_{k+1}) \leq f(x_k) - \tfrac{1}{M_f^2}\omega(\tfrac{1}{2}).$$

  Thus, the number of steps of this stage is bounded as follows:

$$N \leq M_f^2[f(x_0) - f(x_f^*)]/\omega(\tfrac{1}{2}). \tag{5.2.10}$$

- *Second stage*: $\lambda_f(x_k) < \tfrac{1}{2M_f}$. At this stage, we apply method $(5.2.1)_C$. This process converges quadratically:

$$\lambda_f(x_{k+1}) \leq M_f \lambda_f^2(x_k)(1 + 2M_f \lambda_f(x_k)) < \lambda_f(x_k).$$

Since the quadratic convergence is very fast, the main efforts in the above strategy are spent at the first stage. The estimate $(5.2.10)$ shows that the length of this stage is $O(\Delta_f(x_0))$, where

$$\Delta_f(x_0) \stackrel{\text{def}}{=} M_f^2[f(x_0) - f(x_f^*)]. \tag{5.2.11}$$

Is it possible to reach the region of quadratic convergence in a faster way? In order to answer this question, let us consider an alternative way to solve the problem $(5.1.25)$, based on a *path-following scheme*. In Sect. 5.3 we will see how we can use this idea for solving a *constrained* minimization problem.

5.2.2 Path-Following Scheme

Assume that we have $y_0 \in \text{dom } f$. Let us define an *auxiliary central path*

$$y(t) = \arg \min_{y \in \text{dom } f} \left[\psi(t; y) \stackrel{\text{def}}{=} f(y) - t\langle \nabla f(y_0), y \rangle \right], \quad t \in [0, 1]. \tag{5.2.12}$$

This minimization problem corresponds to computation of the value of the dual function $-f_*(s)$ with $s = t\nabla f(y_0)$ (see $(5.1.30)$). Note that $\nabla f(y_0) \in \text{dom } f_*$ and

the origin in the dual space also belongs to dom f_* since the problem (5.1.25) is solvable. Therefore, in view of Lemma 5.1.6,

$$t\nabla f(y_0) \in \text{dom } f_*, \quad 0 \le t \le 1,$$

and trajectory (5.2.12) is well defined.

We are going to follow the auxiliary central path with parameter t changing from one to zero by updating points satisfying the *approximate centering condition*

$$\lambda_{\psi(t;\cdot)}(y) \stackrel{\text{def}}{=} \|\nabla f(y) - t\nabla f(y_0)\|_y^* \le \tfrac{\beta}{M_f}, \tag{5.2.13}$$

where the *centering parameter* β is small enough. Note that the function $\psi(t; \cdot)$ is self-concordant with constant M_f and domain dom f (see Corollary 5.1.2).

Consider the following iterate:

$$(t_+, y_+) = \mathscr{P}_\gamma(t, y) \equiv \begin{cases} t_+ = t - \frac{\gamma}{M_f \|\nabla f(y_0)\|_y^*}, \\[2mm] y_+ = y - \frac{[\nabla^2 f(y)]^{-1}(\nabla f(y) - t_+ \nabla f(y_0))}{1+\xi}, \end{cases} \tag{5.2.14}$$

where $\xi = \frac{M_f^2 \lambda^2}{1+M_f \lambda}$ and $\lambda = \lambda_{\psi(t;\cdot)}(y)$ (this is one iteration of method (5.2.1)$_C$). For future use, we allow the parameter γ in (5.2.14) to be both positive or negative.

Lemma 5.2.2 *Let the pair (t, y) satisfy (5.2.13) with $\beta = \tau^2(1 + \tau + \frac{\tau}{1+\tau+\tau^2})$, where $\tau \le \frac{1}{2}$. Then the pair (t_+, y_+) satisfies the same condition for γ small enough, namely*

$$|\gamma| \le \tau - \tau^2\left(1 + \tau + \tfrac{\tau}{1+\tau+\tau^2}\right). \tag{5.2.15}$$

Proof Let $\lambda = \|\nabla f(y) - t\nabla f(y_0)\|_y^* \le \frac{\beta}{M_f}$, $\lambda_1 = \|\nabla f(y) - t_+\nabla f(y_0)\|_y^*$, and $\lambda_+ = \|\nabla f(y) - t_+\nabla f(y_0)\|_{y_+}^*$. Then $\lambda_1 \le \lambda + \frac{|\gamma|}{M_f} \le \frac{1}{M_f}(\beta + |\gamma|) \stackrel{(5.2.15)}{\le} \frac{\tau}{M_f}$. Hence,

$$\lambda_+ \stackrel{(5.2.8)}{\le} \tfrac{\tau^2}{M_f}\left(1 + \tau + \tfrac{\tau}{1+\tau+\tau^2}\right) = \tfrac{\beta}{M_f}. \qquad \square$$

Let us derive from this fact a complexity bound of the path-following scheme as applied to problem (5.1.25).

Theorem 5.2.3 *Consider the following process:*

$$t_0 = 1, \ y_0 \in dom\ f, \quad (t_{k+1}, y_{k+1}) = \mathscr{P}_\gamma(t_k, y_k), \quad k \geq 0, \tag{5.2.16}$$

where $\gamma = \gamma(\tau) = \tau - \beta$, $\beta = \beta(\tau) = \tau^2 \left(1 + \tau + \frac{\tau}{1+\tau+\tau^2}\right)$, *and* $\tau \leq 0.23$. *Then*

$$\lambda_k \overset{def}{=} \|\nabla f(y_k) - t_k \nabla f(y_0)\|_{y_k}^* \leq \frac{\beta}{M_f}, \quad k \geq 0. \tag{5.2.17}$$

Assume that $\lambda_f(y_k) \geq \frac{1}{2M_f}$ *for all* $k = 0, \dots, N$. *Then*

$$t_N \leq \exp\left\{-\frac{\gamma \varkappa(\tau)N^2}{\Delta_f(x_0)}\right\}, \tag{5.2.18}$$

where $\varkappa(\tau) = \frac{(\tau - 3\beta)(1+\beta)}{2(1+\beta+\beta^2)}$.

Proof Since $\lambda_0 = 0 < \frac{\beta}{M_f}$, by Lemma 5.2.2 we prove that inequality (5.2.17) is valid for all $k \geq 0$. Let $c = -\nabla f(y_0)$. Note that

$$y_k - y_{k+1} \overset{(5.2.14)}{=} \frac{1}{1+\xi_k}[\nabla^2 f(y_k)]^{-1}\left(t_k c + \nabla f(y_k) - \frac{\gamma c}{M_f \|c\|_{y_k}^*}\right), \tag{5.2.19}$$

where $\xi_k = \frac{M_f^2 \lambda_k^2}{1+M_f \lambda_k}$. Therefore,

$$r_k \overset{def}{=} \|y_k - y_{k+1}\|_{y_k} \leq \frac{\lambda_k}{1+\xi_k} + \frac{\gamma}{M_f(1+\xi_k)} = \frac{\gamma + M_f \lambda_k}{M_f(1+\xi_k)} \overset{(5.2.17)}{\leq} \frac{\tau}{M_f}. \tag{5.2.20}$$

Further,

$$t_{k+1} \overset{(5.2.14)}{=} t_k - \frac{\gamma}{M_f \|c\|_{y_k}^*} = t_k\left(1 - \frac{\gamma}{M_f t_k \|c\|_{y_k}^*}\right) \leq t_k \exp\left\{-\frac{\gamma}{M_f t_k \|c\|_{y_k}^*}\right\}.$$

Thus, $t_N \leq \exp\left\{-\frac{\gamma}{M_f} S_N\right\}$, where $S_N = \sum_{k=0}^{N} \frac{1}{t_k \|c\|_{y_k}^*}$. Let us estimate this value from below.

Since $\frac{\beta^2}{M_f^2} \overset{(5.2.17)}{\geq} \lambda_f^2(y_k) + 2t_k\langle\nabla f(y_k), [\nabla^2 f(y_k)]^{-1}c\rangle + t_k^2(\|c\|_{y_k}^*)^2$, we have

$$-\langle\nabla f(y_k), [\nabla^2 f(y_k)]^{-1}c\rangle \geq \frac{1}{2t_k}\left[\lambda_f^2(y_k) + t_k^2(\|c\|_{y_k}^*)^2 - \frac{\beta^2}{M_f^2}\right]. \tag{5.2.21}$$

Therefore,

$$f(y_k) - f(y_{k+1}) \overset{(5.1.16)}{\geq} \langle \nabla f(y_k), y_k - y_{k+1} \rangle - \frac{1}{M_f^2}\omega_*(M_f r_k)$$

$$\overset{(5.2.19)}{=} \frac{1}{1+\xi_k}\langle \nabla f(y_k), [\nabla^2 f(y_k)]^{-1}\left(t_k c + \nabla f(y_k) - \frac{\gamma c}{M_f\|c\|_{y_k}^*}\right)\rangle - \frac{1}{M_f^2}\omega_*(M_f r_k)$$

$$= \frac{\lambda_k^2}{1+\xi_k} - \frac{t_k}{1+\xi_k}\langle c, [\nabla^2 f(y_k)]^{-1}(t_k c + \nabla f(y_k))\rangle$$

$$+ \frac{1}{1+\xi_k}\langle \nabla f(y_k), [\nabla^2 f(y_k)]^{-1}\left(\frac{-\gamma c}{M_f\|c\|_{y_k}^*}\right)\rangle - \frac{1}{M_f^2}\omega_*(M_f r_k)$$

$$\overset{(5.2.17)}{\geq} \frac{\lambda_k^2 - t_k\|c\|_{y_k}^*\lambda_k}{1+\xi_k} - \frac{\gamma}{M_f\|c\|_{y_k}^*(1+\xi_k)}\langle \nabla f(y_k), [\nabla^2 f(y_k)]^{-1}c\rangle - \frac{1}{M_f^2}\omega_*(M_f r_k)$$

$$\overset{(5.2.21)}{\geq} \frac{\lambda_k^2 - t_k\|c\|_{y_k}^*\lambda_k}{1+\xi_k} + \frac{\gamma}{2M_f t_k\|c\|_{y_k}^*(1+\xi_k)}\left[\lambda_f^2(y_k) + t_k^2(\|c\|_{y_k}^*)^2 - \frac{\beta^2}{M_f^2}\right]$$

$$- \frac{1}{M_f^2}\omega_*(M_f r_k)$$

$$\overset{(5.2.20)}{\geq} \frac{\gamma - 2M_f\lambda_k}{2M_f(1+\xi_k)}t_k\|c\|_{y_k}^* + \rho_k,$$

where $\rho_k = \frac{\gamma}{2M_f t_k\|c\|_{y_k}^*(1+\xi_k)}\left[\lambda_f^2(y_k) - \frac{\beta^2}{M_f^2}\right] - \frac{1}{M_f^2}\omega_*(\tau)$.

Our next goal is to show that $\rho_k \geq 0$. Note that $t_k\|c\|_{y_k}^* \overset{(5.2.17)}{\leq} \lambda_f(y_k) + \frac{\beta}{M_f}$. Since $\lambda_f(y_k) \geq \frac{1}{2M_f}$, we have

$$\rho_k \geq \frac{\gamma}{2M_f(1+\xi_k)}\left[\lambda_f(y_k) - \frac{\beta}{M_f}\right] - \frac{1}{M_f^2}\omega_*(\tau) \geq \frac{\gamma(1-2\beta)}{4M_f^2(1+\xi_k)} - \frac{1}{M_f^2}\omega_*(\tau)$$

$$\overset{(5.2.17)}{\geq} \frac{1}{M_f^2}\left[\frac{\gamma(1-2\beta)(1+\beta)}{4(1+\beta+\beta^2)} - \omega_*(\tau)\right].$$

Note that $\gamma = O(\tau)$, $\beta = O(\tau^2)$, and $\omega_*(\tau) = O(\tau^2)$. Therefore, for τ small enough we have $\rho_k \geq 0$. By numerical evaluation, it is easy to check that this can be achieved by taking $\tau \leq 0.23$.

Further,

$$\frac{\gamma - 2M_f\lambda_k}{2(1+\xi_k)} \overset{(5.2.17)}{\geq} \frac{(\gamma-2\beta)(1+\beta)}{2(1+\beta+\beta^2)} = \frac{(\tau-3\beta)(1+\beta)}{2(1+\beta+\beta^2)} \overset{\text{def}}{=} \varkappa(\tau).$$

Again, it is easy to check that $\varkappa(\tau) > 0$ for $\tau \in (0, 0.23]$. Thus, we have proved that $f(y_k) - f(y_{k+1}) \geq \frac{\varkappa(\tau)}{M_f} t_k \|c\|_{y_k}$. Therefore,

$$S_N \geq \sum_{k=0}^{N} \frac{\varkappa(\tau)}{M_f(f(y_k)-f(y_{k+1}))} \geq \frac{\varkappa(\tau)\Lambda^*(N)}{M_f(f(y_0)-f(y_{N+1}))},$$

where $\Lambda^*(N) = \min_{\lambda \in \mathbb{R}_+^{N+1}} \left\{ \sum_{i=1}^{N+1} \frac{1}{\lambda^{(i)}} : \sum_{i=1}^{N+1} \lambda^{(i)} = 1 \right\} = (N+1)^2.$ □

Let us estimate now the number of iterations, which are necessary for method (5.2.16) to enter the region of quadratic convergence $\mathcal{Q}_f$. Define

$$D = \max_{x,y \in \text{dom} f} \{ \|x - y\|_{y_0} : f(x) \leq f(y_0), \ f(y) \leq f(y_0) \}.$$

Theorem 5.2.4 *Let the sequence $\{y_k\}_{k \geq 0}$ be generated by method (5.2.16). Then for all*

$$N \geq \left[\frac{\Delta_f(x_0)}{\gamma \varkappa(\tau)} \ln \left(\frac{M_f D \omega^{-1}(\Delta_f(x_0))}{\omega(\frac{(1-\beta)(1-2\beta)}{2})} \right) \right]^{1/2} \tag{5.2.22}$$

and we have $y_N \in \mathcal{Q}_f$.

Proof Indeed,

$$f(y(t_k)) - f^* \leq \langle \nabla f(y(t_k)), y(t_k) - x^* \rangle \overset{(5.2.12)}{=} t_k \langle \nabla f(y_0), y(t_k) - x^* \rangle$$

$$\leq t_k \lambda_f(y_0) D.$$

Note that $\omega(M_f \lambda_f(y_0)) \overset{(5.1.29)}{\leq} M_f^2(f(y_0) - f^*) = \Delta_f(y_0)$. Thus,

$$\frac{1}{M_f^2} \omega(M_f \lambda_f(y(t_k))) \overset{(5.1.29)}{\leq} f(y(t_k)) - f^* \leq \frac{t_k}{M_f} \omega^{-1}(\Delta_f(y_0)) D.$$

Since $\|\nabla f(y_k) - \nabla f(y(t_k))\|_{y_k}^* \overset{(5.2.12)}{=} \|\nabla f(y_k) - t_k \nabla f(y_0)\|_{y_k}^* \leq \frac{\beta}{M_f}$, we have

$$\lambda_f(y_k) \overset{(5.2.17)}{\leq} t_k \|\nabla f(y_0)\|_{y_k}^* + \frac{\beta}{M_f} = \langle \nabla f(y(t_k)), [\nabla^2 f(y_k)]^{-1} \nabla f(y(t_k)) \rangle^{1/2}$$

$$+ \frac{\beta}{M_f}$$

$$\overset{(5.1.36)}{\leq} \frac{1}{1-\beta} \lambda_f(y(t_k)) + \frac{\beta}{M_f}.$$

Thus, inclusion $y_k \in \mathcal{Q}_f$ is ensured by the inequality $\lambda_f(y(t_k)) \leq \frac{(1-\beta)(1-2\beta)}{2M_f}$. Consequently, we need to ensure the inequality

$$\frac{t_k}{M_f}\omega^{-1}(\Delta_f(x_0))D \leq \frac{1}{M_f^2}\omega\left(\frac{(1-\beta)(1-2\beta)}{2}\right).$$

It remains to use the estimate (5.2.18).   □

As we can see from the estimate (5.2.22), up to a logarithmic factor, the number of iterations of the path-following scheme is proportional to $\Delta_f^{1/2}(y_0)$. This is much better than the guarantee (5.2.10) obtained for the Damped Newton's Method (5.1.28). However, as we will see in Sect. 5.2.3, for some special subclasses of self-concordant functions the performance estimate (5.2.22) can be significantly improved.

From the practical point of view, reasonable values of parameters for path-following scheme (5.2.16) correspond to $\tau = 0.15$. In this case, $\left[\frac{1}{\gamma(\tau)\varkappa(\tau)}\right]^{1/2} \leq 16.1$.

Remark 5.2.1 The dual interpretation of the central path (5.2.12) is quite straightforward: it is just a straight line. We follow the primal image of the *dual* central path

$$s(t) = t\nabla f(y_0) \in \operatorname{dom} f_*, \quad 0 \leq t \leq 1,$$

by generating points $s_k = \nabla f(y_k)$ in a small neighborhood of this trajectory:

$$\langle s_k - s(t_k), \nabla^2 f_*(s_k)(s_k - s(t_k))\rangle \overset{(5.2.13)}{\leq} \frac{\beta^2}{M_f^2}. \qquad □$$

5.2.3   Minimizing Strongly Convex Functions

Let $B = B^* \succ 0$ map $\mathbb{E}$ to $\mathbb{E}^*$. Define the Euclidean metric

$$\|x\|^2 = \langle Bx, x\rangle^{1/2}, \quad x \in \mathbb{E}.$$

In this section, we consider the following minimization problem

$$\min_{x \in \mathbb{E}} f(x), \tag{5.2.23}$$

where f is a strongly convex function:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \tfrac{1}{2}\sigma_2(f)\|y - x\|^2, \quad x, y \in \mathbb{E}, \tag{5.2.24}$$

where $\sigma_2(f) > 0$. We also assume that the function f belongs to $\mathbb{C}^3(\mathbb{E})$ and its Hessian is Lipschitz continuous:

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \le L_3(f)\|x - y\|, \quad x, y \in \mathbb{E}. \tag{5.2.25}$$

As we have seen in Example 5.1.1 (6), this function is self-concordant on $\mathbb{E}$ with the constant

$$M_f = \frac{L_3(f)}{2\sigma_2^{3/2}(f)}. \tag{5.2.26}$$

Thus, problem (5.2.23) can be solved by methods (5.1.28) and (5.2.16). The corresponding complexity bounds can be given in terms of the complexity measure

$$\Delta_f(x_0) = \frac{L_3(f)}{2\sigma_2^{3/2}(f)}(f(x_0) - f^*).$$

As we have seen, the first method needs $O(\Delta_f(x_0))$ iterations. The complexity bound for the second scheme is of the order $\tilde{O}(\Delta_f^{1/2}(x_0))$, where $\tilde{O}(\cdot)$ denotes the hidden logarithmic factors. Let us show that for our particular subclass of self-concordant functions these bounds can be significantly improved.

We will do this by the second-order methods based on cubic regularization of the Newton's Method (see Sect. 4.2). In view of (4.2.60), the region of quadratic convergence of the Cubic Newton's Method (4.2.33) in terms of function value is defined as

$$\mathbb{Q}_f = \left\{ x \in \mathbb{E} : \ f(x) - f^* \le \frac{\sigma_2^3(f)}{2L_3^2(f)} = \frac{1}{8M_f^2} \right\}.$$

Let us check how many iterations we need to enter this region by different schemes based on the cubic Newton step.

Assume our method has the following rate of convergence:

$$f(x_k) - f^* \le \frac{cL_3(f)D^3}{k^p},$$

where c is an absolute constant, $p > 0$, and $D = \max\limits_{x \in \mathbb{E}}\{\|x - x^*\| : \ f(x) \le f(x^0)\}$. Since f is strongly convex, for all x with $f(x) \le f(x_0)$ we have

$$\tfrac{1}{2}\sigma_2(f)\|x - x^*\|^2 \overset{(5.2.24)}{\le} f(x) - f^* \le f(x_0) - f^*.$$

Therefore,

$$f(x_k) - f^* \;\leq\; \frac{cL_3(f)}{k^p}\left(\frac{2}{\sigma_2(f)}(f(x_0) - f^*)\right)^{3/2}$$

$$\stackrel{(5.2.26)}{=} \frac{2^{5/2}cM_f}{k^p}(f(x_0) - f^*)^{3/2}. \tag{5.2.27}$$

Thus, we need $O\left(\left[M_f^3(f(x_0) - f^*)^{3/2}\right]^{1/p}\right) = O\left(\Delta_f^{\frac{3}{2p}}(x_0)\right)$ iterations to enter the region of quadratic convergence $\mathbb{Q}_f$. For the Cubic Newton's method (4.2.33) we have $p = 2$. Thus, it ensures complexity $O(\Delta_f^{3/4}(x_0))$. For the accelerated Cubic Newton's method (4.2.46) we have $p = 3$. Thus, it needs $O(\Delta^{1/2}(x_0))$ iterations (which is slightly better than (5.2.22)). However, note that for these methods there exists a powerful acceleration tool based on a *restarting procedure*.

Let us define k_p as the first integer for which the right-hand side of inequality (5.2.27) is smaller than $\frac{1}{2}(f(x_0) - f^*)$:

$$\frac{2^{5/2}cM_f}{k^p}(f(x_0) - f^*)^{3/2} \leq \tfrac{1}{2}(f(x_0) - f^*).$$

Clearly $k_p = O\left(\left[M_f(f(x_0) - f^*)^{1/2}\right]^{1/p}\right) = O\left(\Delta_f^{\frac{1}{2p}}(x_0)\right)$. This value can be used in the following multi-stage scheme.

Multi-stage Acceleration Scheme

Set $y_0 = x_0$

At the kth stage ($k \geq 1$) the method starts from the point y_{k-1}.

After $t_k = \left\lceil \frac{k_p}{2^{(k-1)/(2p)}} \right\rceil$ steps it generates the output y_k.

The method stops when $y_k \in \mathbb{Q}_f$.

(5.2.28)

Theorem 5.2.5 *The total number of stages T in the optimizations strategy (5.2.28) satisfies the inequality*

$$T \leq 4 + \log_2 \Delta_f(x_0). \tag{5.2.29}$$

The total number of lower-level iterations N in this scheme does not exceed

$$4 + \log_2 \Delta_f(x_0) + \frac{2^{1/(2p)}}{2^{1/(2p)}-1}k_p.$$

Proof Let us prove by induction that $f(y_k) - f^* \leq (\frac{1}{2})^k (f(y_0) - f^*)$. For $k = 0$ this is true. Assume that this is also true for some $k \geq 0$. Note that $t_{k+1}^p \geq (\frac{1}{2})^{k/2} k_p^p$. Therefore,

$$\frac{f(y_{k+1}) - f^*}{f(y_k) - f^*} \leq \frac{2^{5/2} c M_f}{t_{k+1}^p} (f(y_k) - f^*)^{1/2} \leq \frac{k_p^p (f(y_k) - f^*)^{1/2}}{2 t_{k+1}^p (f(x_0) - f^*)^{1/2}}$$

$$\leq \frac{1}{2} \left[\frac{2^k (f(y_k) - f^*)}{f(x_0) - f^*} \right]^{1/2} \leq \frac{1}{2}.$$

Thus, the total number of stages satisfies inequality $\left(\frac{1}{2}\right)^{T-1} (f(x_0) - f^*) \geq \frac{1}{8 M_f^2}$. Finally,

$$N = \sum_{k=1}^{T} t_k \leq T + k_p \sum_{k=0}^{T-1} \left(\frac{1}{2}\right)^{\frac{k}{2p}} \leq T + k_p \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^{\frac{k}{2p}}$$

$$= T + \frac{k_p}{1 - \left(\frac{1}{2}\right)^{1/(2p)}}. \qquad\qquad \Box$$

Applying Theorem 5.2.5 to different second-order methods based on Cubic Regularization, we get the following complexity bounds.

- **Cubic Newton's Method (4.2.33).** For this method $p = 2$. Therefore, the complexity bound of this scheme, used in the framework of multi-stage method (5.2.28), is of the order

$$O\left(\Delta_f^{1/4}(x_0)\right).$$

In fact, this method does not need a restarting strategy. Thus, Theorem 5.2.5 provides the Cubic Newton method with a better way of estimating its rate of convergence.

- **Accelerated Newton's Method (4.2.46).** For this method $p = 3$. Hence, the complexity bound of the corresponding multi-stage scheme (5.2.28) becomes

$$O\left(\Delta^{1/6}(x_0)\right).$$

- **Optimal second-order method** (see Sect. 4.3.2). For this method $p = 3.5$. Therefore, the corresponding complexity bound is

$$\tilde{O}\left(\Delta^{1/7}(x_0)\right).$$

However, note that this method includes an expensive line-search procedure. Consequently, its practical efficiency should be worse that the efficiency of the

method from the previous item. Note that the theoretical gap in the complexity estimates of these methods is negligibly small, of the order of $O\left(\Delta_f^{1/42}(x_0)\right)$. For all reasonable values of the complexity measure $\Delta_f(x_0)$, feasible for modern computers, it should be much smaller than the logarithmic factors coming from the line search.

5.3 Self-concordant Barriers

(Motivation; Definition of self-concordant barriers; Barriers related to self-concordant functions; The implicit barrier theorem; Main properties; Standard minimization problems; The central path; The path-following method; How to initialize the process? Problems with functional constraints.)

5.3.1 Motivation

In the previous section, we have seen that the Newton's Method is very efficient in minimizing *self-concordant* functions. Such a function is always a barrier for its domain. Let us check what can be proved about the Sequential Unconstrained Minimization approach (Sect. 1.3.3) based on these barriers. From now on, we are always working with *standard* self-concordant functions, which means that

$$M_f = 1. \tag{5.3.1}$$

In what follows, we deal with constrained minimization problems of a special type. Let Dom $f = $ cl (dom f).

Definition 5.3.1 A constrained minimization problem is called *standard* if it has the following form:

$$\min\{\langle c, x \rangle \mid x \in Q\}, \tag{5.3.2}$$

where Q is a closed convex set. It is also assumed that we know a *standard* self-concordant function f such that Dom $f = Q$.

Note that the assumption $M_f = 1$ is not binding since otherwise we can multiply f by an appropriate constant (see Corollary 5.1.3).

Let us introduce a parametric family of penalty functions

$$f(t; x) = t\langle c, x \rangle + f(x)$$

with $t \geq 0$. Note that $f(t; x)$ is self-concordant in x (see Corollary 5.1.2). Define

$$x^*(t) = \arg \min_{x \in \text{dom} f} f(t; x).$$

This trajectory is called the *central path* of problem (5.3.2). We can expect that $x^*(t) \to x^*$ as $t \to \infty$ (see Sect. 1.3.3). Therefore, it should be a good idea to keep our test points close to this trajectory.

Recall that the Newton's Methods, as applied to the minimization of the function $f(t; \cdot)$, have local quadratic convergence (Theorem 5.2.2). Our subsequent analysis is based on the Intermediate Newton Method $(5.2.1)_C$, which has the following region of quadratic convergence:

$$\lambda_{f(t; \cdot)}(x) \leq \beta < \tfrac{1}{2}.$$

Let us study our possibilities to move forward in t, assuming that we know exactly $x = x^*(t)$ for some $t > 0$.

Thus, we are going to increase t:

$$t_+ = t + \Delta, \quad \Delta > 0.$$

However, we need to keep x in the region of quadratic convergence of Newton's Method for the function $f(t + \Delta; \cdot)$:

$$\lambda_{f(t+\Delta; \cdot)}(x) \leq \beta < \tfrac{1}{2}.$$

Note that the update $t \to t_+$ does not change the Hessian of the barrier function:

$$\nabla^2 f(t + \Delta; x) = \nabla^2 f(t; x).$$

Therefore, it is easy to estimate how big the step Δ can be. Indeed, the first-order optimality condition (1.2.4) provides us with the following *central path equation:*

$$tc + \nabla f(x^*(t)) = 0. \tag{5.3.3}$$

Since $tc + \nabla f(x) = 0$, we obtain

$$\lambda_{f(t+\Delta; \cdot)}(x) = \| t_+c + \nabla f(x) \|_x^* \stackrel{(5.3.3)}{=} \Delta \| c \|_x^* = \tfrac{\Delta}{t} \| \nabla f(x) \|_x^* \leq \beta.$$

Hence, if we want to increase t at some *linear rate*, we need to assume that the value

$$\lambda_f^2(x) = (\| \nabla f(x) \|_x^*)^2 \equiv \langle \nabla f(x), [\nabla^2 f(x)]^{-1} \nabla f(x) \rangle$$

is *uniformly bounded* on dom f. Without this assumption, we can have only a sublinear rate of convergence of the process (see Sect. 5.2.2).

Thus, we come to a definition of a *self-concordant barrier*.

5.3.2 Definition of a Self-concordant Barrier

Definition 5.3.2 Let $F(\cdot)$ be a standard self-concordant function. We call it a ν-*self-concordant barrier* for the set Dom F, if

$$\sup_{u \in \mathbb{E}} [2\langle \nabla F(x), u \rangle - \langle \nabla^2 F(x)u, u \rangle] \leq \nu \qquad (5.3.4)$$

for all $x \in \text{dom } F$. The value ν is called the *parameter* of the barrier.

Note that we do not assume $\nabla^2 F(x)$ to be nondegenerate. However, if this is the case, then inequality (5.3.4) is equivalent to the following:

$$\langle \nabla F(x), [\nabla^2 F(x)]^{-1} \nabla F(x) \rangle \leq \nu. \qquad (5.3.5)$$

We will also use another equivalent form of inequality (5.3.4):

$$\langle \nabla F(x), u \rangle^2 \leq \nu \langle \nabla^2 F(x)u, u \rangle \quad \forall u \in \mathbb{E}. \qquad (5.3.6)$$

(To see this for u with $\langle \nabla^2 F(x)u, u \rangle > 0$, replace u in (5.3.4) by τu and find the maximum of the left-hand side in τ.) Note that the condition (5.3.6) can be rewritten in matrix notation:

$$\nabla^2 F(x) \succeq \tfrac{1}{\nu} \nabla F(x) \nabla F(x)^T. \qquad (5.3.7)$$

Lemma 5.3.1 *Let F be a ν-self-concordant barrier. Then for any $p \geq \nu$ the function $\xi_p(x) = \exp\left\{-\frac{1}{p}F(x)\right\}$ is concave on dom F. On the other hand, if function $\xi_\nu(\cdot)$ is concave on dom F, then F is a self-concordant barrier.*

Proof Indeed, for any $x \in \text{dom } F$ and $h \in \mathbb{E}$, we have

$$\langle \nabla \xi_p(x), h \rangle = -\tfrac{1}{p} \langle \nabla F(x), h \rangle \xi_p(x),$$

$$\langle \nabla^2 \xi_p(x)h, h \rangle = \tfrac{1}{p^2} \langle \nabla F(x), h \rangle^2 \xi_p(x) - \tfrac{1}{p} \langle \nabla^2 F(x)h, h \rangle \xi_p(x).$$

It remains to use definition (5.3.6). □

Note that condition (5.3.5) has interesting dual interpretation. In view of relation (5.1.34), definition (5.3.5) is equivalent to the following condition:

$$\langle s, \nabla^2 F_*(s)s \rangle \leq \nu, \quad s \in \text{dom } F_*. \qquad (5.3.8)$$

In other words, at any feasible s, the distance to the origin is proportional to the size of the unit Dikin ellipsoid, which describes an ellipsoidal neighborhood in dom f_* with similar Hessians.

Let us now check which self-concordant functions presented in Example 5.1.1 are also self-concordant barriers.

Example 5.3.1

1. *Linear function*: $f(x) = \alpha + \langle a, x \rangle$, dom $f = \mathbb{E}$. Clearly, for $a \neq 0$ this function is not a self-concordant barrier since $\nabla^2 F(x) = 0$.
2. *Convex quadratic function.* Let $A = A^T \succ 0$. Consider the function

$$f(x) = \alpha + \langle a, x \rangle + \frac{1}{2} \langle Ax, x \rangle, \quad \text{dom } f = \mathbb{R}^n.$$

Then $\nabla f(x) = a + Ax$ and $\nabla^2 f(x) = A$. Therefore,

$$\langle [\nabla^2 f(x)]^{-1} \nabla f(x), \nabla f(x) \rangle = \langle A^{-1}(Ax + a), Ax + a \rangle$$

$$= \langle Ax, x \rangle + 2\langle a, x \rangle + \langle A^{-1}a, a \rangle.$$

Clearly, this value is unbounded from above on $\mathbb{R}^n$. Thus, a quadratic function is not a self-concordant barrier.
3. *Logarithmic barrier for a ray.* Consider the following function of one variable:

$$F(x) = -\ln x, \quad \text{dom } F = \{x \in \mathbb{R} \mid x > 0\}.$$

Then $\nabla F(x) = -\frac{1}{x}$ and $\nabla^2 F(x) = \frac{1}{x^2} > 0$. Therefore

$$\frac{(\nabla F(x))^2}{\nabla^2 F(x)} = \frac{1}{x^2} \cdot x^2 = 1.$$

Thus, $F(\cdot)$ is a ν-self-concordant barrier for the set $\{x \geq 0\}$ with $\nu = 1$.
4. *Logarithmic barrier for a second-order region.* Let $A = A^T \succeq 0$. Consider the *concave* quadratic function

$$\phi(x) = \alpha + \langle a, x \rangle - \frac{1}{2} \langle Ax, x \rangle.$$

Define $F(x) = -\ln \phi(x)$, dom $f = \{x \in \mathbb{R}^n \mid \phi(x) > 0\}$. Then

$$\langle \nabla F(x), u \rangle = -\frac{1}{\phi(x)}[\langle a, u \rangle - \langle Ax, u \rangle],$$

$$\langle \nabla^2 F(x)u, u \rangle = \frac{1}{\phi^2(x)}[\langle a, u \rangle - \langle Ax, u \rangle]^2 + \frac{1}{\phi(x)} \langle Au, u \rangle.$$

Let $\omega_1 = \langle \nabla F(x), u \rangle$ and $\omega_2 = \frac{1}{\phi(x)} \langle Au, u \rangle$. Then

$$\langle \nabla^2 F(x)u, u \rangle = \omega_1^2 + \omega_2 \geq \omega_1^2.$$

Therefore $2\langle \nabla F(x), u \rangle - \langle \nabla^2 F(x)u, u \rangle \leq 2\omega_1 - \omega_1^2 \leq 1$. Thus, $F(\cdot)$ is a ν-self-concordant barrier with $\nu = 1$. $\quad\square$

Let us now check the results of some simple operations with self-concordant barriers.

Theorem 5.3.1 *Let $F(\cdot)$ be a self-concordant barrier. Then the function $\langle c, x \rangle + F(x)$ is a standard self-concordant function on $\mathrm{dom}\, F$.*

Proof Since $F(\cdot)$ is a self-concordant function, we just apply Corollary 5.1.2. $\quad\square$

Note that this property is important for path-following schemes.

Theorem 5.3.2 *Let F_i be ν_i-self-concordant barriers, $i = 1, 2$. Then the function*

$$F(x) = F_1(x) + F_2(x)$$

is a self-concordant barrier for a convex set $\mathrm{Dom}\, F = \mathrm{Dom}\, F_1 \bigcap \mathrm{Dom}\, F_2$ with the parameter $\nu = \nu_1 + \nu_2$.

Proof In view of Theorem 5.1.1, F is a standard self-concordant function. Let us fix $x \in \mathrm{dom}\, F$. Then

$$\max_{u \in \mathbb{R}^n} [2\langle \nabla F(x), u \rangle - \langle \nabla^2 F(x)u, u \rangle]$$

$$= \max_{u \in \mathbb{R}^n} [2\langle \nabla F_1(x), u \rangle - \langle \nabla^2 F_1(x)u, u \rangle + 2\langle \nabla F_2(x), u \rangle - \langle \nabla^2 F_2(x)u, u \rangle]$$

$$\leq \max_{u \in \mathbb{R}^n} [2\langle \nabla F_1(x), u \rangle - \langle \nabla^2 F_1(x)u, u \rangle] + \max_{u \in \mathbb{R}^n} [2\langle \nabla F_2(x), u \rangle - \langle \nabla^2 F_2(x)u, u \rangle]$$

$$\leq \nu_1 + \nu_2. \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$$

It is easy to see that the value of the parameter of a self-concordant barrier is invariant with respect to an affine transformation of variables.

Theorem 5.3.3 *Let $\mathscr{A}(x) = Ax + b$ be a linear operator, $\mathscr{A} : \mathbb{E} \to \mathbb{E}_1$. Assume that function F is a ν-self-concordant barrier with $\mathrm{Dom}\, F \subset \mathbb{E}_1$. Then the function*

$$\Phi(x) = F(\mathscr{A}(x))$$

is a ν-self-concordant barrier for the set $\mathrm{Dom}\, \Phi = \{x \in \mathbb{E} : \mathscr{A}(x) \in \mathrm{Dom}\, F\}$.

Proof The function $\Phi(\cdot)$ is a standard self-concordant function in view of Theorem 5.1.2. Let us fix $x \in \mathrm{dom}\, \Phi$. Then $y = \mathscr{A}(x) \in \mathrm{dom}\, F$. Note that for any $u \in \mathbb{E}$ we have

$$\langle \nabla \Phi(x), u \rangle = \langle \nabla F(y), Au \rangle, \quad \langle \nabla^2 \Phi(x)u, u \rangle = \langle \nabla^2 F(y)Au, Au \rangle.$$

Therefore

$$\max_{u \in \mathbb{E}} [2\langle \nabla \Phi(x), u \rangle - \langle \nabla^2 \Phi(x)u, u \rangle] = \max_{u \in \mathbb{E}} [2\langle \nabla F(y), Au \rangle - \langle \nabla^2 F(y)Au, Au \rangle]$$

$$\leq \max_{w \in \mathbb{E}_1} [2\langle \nabla F(y), w \rangle - \langle \nabla^2 F(y)w, w \rangle] \leq \nu. \qquad \square$$

To conclude this section, let us show how to construct self-concordant barriers for the level sets of self-concordant functions and for the epigraphs of self-concordant barriers.

Theorem 5.3.4 *Let the function f be self-concordant with constant $M_f \geq 0$. Suppose that the set*

$$\mathcal{L}(\beta) = \{x \in dom\, f : \ f(x) \leq \beta\}$$

has nonempty interior and $f(x) \geq f^*$ for all $x \in dom\, f$. Then the function

$$F(x) = -\nu \ln(\beta - f(x))$$

with any $\nu \geq 1 + M_f^2(\beta - f^*)$ is a ν-self-concordant barrier for the level set $\mathcal{L}(\beta)$.

Proof Let $\phi(x) = -\ln(\beta - f(x))$. In view of Theorem 5.1.4 and Corollary 5.1.3, the function $F(x) = \nu\phi(x)$ is a standard self-concordant function on $dom\, f$. On the other hand, for any $h \in \mathbb{E}$ we have

$$\langle \nabla F(x), h \rangle^2 = \nu^2 \langle \nabla \phi(x), h \rangle^2 \overset{(5.1.8)}{\leq} \nu^2 \langle \nabla^2 \phi(x)h, h \rangle = \nu \langle \nabla^2 F(x)h, h \rangle.$$

Thus, by definition (5.3.6), F is a ν-self-concordant barrier for $\mathcal{L}(\beta)$. $\square$

Theorem 5.3.5 *Let f be a ν-self-concordant barrier. Then the function*

$$F(x, t) = f(x) - \ln(t - f(x))$$

is a $(\nu + 1)$-self-concordant barrier for the epigraph

$$\mathcal{E}_f = \{(x, t) \in dom\, f \times \mathbb{R} : \ t \geq f(x)\}.$$

Proof Let us fix a direction $h \in \mathbb{E}$ and $\delta \in \mathbb{R}$. Consider the function

$$\phi(\tau) = F(x + \tau h, t + \tau \delta) = f(x + \tau h) - \ln(t + \tau \delta - f(x + \tau h)).$$

Let $\omega = t - f(x)$ and $\hat{\omega} = 1 + \frac{1}{\omega}$. Then

$$\phi'(0) = \langle \nabla f(x), h \rangle + \frac{1}{\omega}(\langle \nabla f(x), h \rangle - \delta),$$

$$\phi''(0) = \langle \nabla^2 f(x)h, h \rangle + \frac{1}{\omega^2}(\langle \nabla f(x), h \rangle - \delta)^2 + \frac{1}{\omega}\langle \nabla^2 f(x)h, h \rangle$$

$$= \hat{\omega}\langle \nabla^2 f(x)h, h \rangle + \frac{1}{\omega^2}(\langle \nabla f(x), h \rangle - \delta)^2.$$

Define $\xi = \left[\hat{\omega}\langle \nabla^2 f(x)h, h \rangle\right]^{1/2}$ and $\lambda = \frac{1}{\omega}(\langle \nabla f(x), h \rangle - \delta)$. Note that

$$\phi'(0) \overset{(5.3.6)}{\leq} \sqrt{\nu}\langle \nabla^2 f(x)h, h \rangle^{1/2} + \lambda = \xi\sqrt{\frac{\nu}{\hat{\omega}}} + \lambda.$$

It remains to note that the maximum of the right-hand side of this inequality subject to the constraint $\xi^2 + \lambda^2 = 1$ is equal to $\left[\frac{\nu}{\hat{\omega}} + 1\right]^{1/2} \leq \sqrt{\nu + 1}$. Thus, in view of definition (5.3.6), the parameter of the barrier F can be chosen as $\nu + 1$.

Let us estimate now the third derivative of the function ϕ at zero, assuming that its second derivative is less or equal to one. Note that

$$\phi'''(0) = D^3 f(x)[h, h, h] + \frac{2}{\omega^3}(\langle \nabla f(x), h \rangle - \delta)^3$$

$$+ \frac{3}{\omega^2}(\langle \nabla f(x), h \rangle - \delta)\langle \nabla^2 f(x)h, h \rangle + \frac{1}{\omega}D^3 f(x)[h, h, h]$$

$$\overset{(5.1.4)}{\leq} 2\hat{\omega}\langle \nabla^2 f(x)h, h \rangle^{3/2} + \frac{2}{\omega^3}(\langle \nabla f(x), h \rangle - \delta)^3$$

$$+ \frac{3}{\omega^2}(\langle \nabla f(x), h \rangle - \delta)\langle \nabla^2 f(x)h, h \rangle$$

$$= 2\sqrt{\frac{\omega}{1+\omega}}\xi^3 + 2\lambda^3 + \frac{3}{1+\omega}\xi^2\lambda = 2\gamma\xi^3 + 2\lambda^3 + 3(1 - \gamma^2)\xi^2\lambda,$$

where $\gamma^2 = \frac{\omega}{1+\omega}$. We need to maximize the right-hand side of the above inequality subject to constraints $\xi^2 + \lambda^2 \leq 1$ and $\gamma \in [0, 1]$:

$$\varkappa^* = \max_{\gamma, \lambda, \xi}\{2\gamma\xi^3 + 2\lambda^3 + 3(1 - \gamma^2)\xi^2\lambda : \xi^2 + \lambda^2 \leq 1, \ 0 \leq \gamma \leq 1\}.$$

Let us maximize this objective in γ. From the first-order optimality condition for γ,

$$2\xi^3 - 6\gamma\xi^2\lambda = 0,$$

we have $\gamma_* = \min\left\{1, \frac{\xi}{3\lambda}\right\}$. Assume that $\xi \geq 3\lambda$. Then $\gamma_* = 1$ and we need to maximize $2\xi^3 + 2\lambda^3$ with constraints $\xi^2 + \lambda^2 = 1$ and $\xi \geq 3\lambda$. Introducing new

variables $\hat{\xi} = \xi^2$ and $\hat{\lambda} = \lambda^2$, we come to the problem

$$\max_{\hat{\xi}, \hat{\lambda} \geq 0} \{2\hat{\xi}^{3/2} + 2\hat{\lambda}^{3/2} : \hat{\xi} + \hat{\lambda} \leq 1, \ \hat{\xi} \geq 9\hat{\lambda}\}.$$

Its objective is convex. Hence, by inspecting the extreme points of its feasible set we find the optimal solution $\hat{\xi}_* = 1$, $\hat{\lambda}_* = 0$. Thus, the maximal value of this problem is two.

Assume now that $\xi \leq 3\lambda$. Then $\gamma_* = \frac{\xi}{3\lambda}$ and we get the following objective:

$$2\frac{\xi}{3\lambda}\xi^3 + 2\lambda^3 + 3\left(1 - \frac{\xi^2}{9\lambda^2}\right)\xi^2\lambda = \frac{\xi^4}{3\lambda} + 2\lambda^3 + 3\xi^2\lambda.$$

Note that the maximum of this expression is attained at the boundary of the unit circle: $\xi^2 + \lambda^2 = 1$. Thus, we need to show that

$$\frac{(1-\lambda^2)^2}{3\lambda} + 2\lambda^3 + 3(1 - \lambda^2)\lambda \leq 2,$$

with constraint $3\lambda \geq \sqrt{1 - \lambda^2}$. In other words, we need to prove that

$$p(\lambda) \stackrel{\text{def}}{=} (1 - \lambda^2)^2 + 3\lambda(3\lambda - \lambda^3) - 6\lambda \ \leq \ 0, \quad \frac{1}{\sqrt{10}} \leq \lambda \leq 1.$$

Note that $p(\lambda) = (1 - \lambda)^2(3 - 2(1 + \lambda)^2) \leq 0$ for all $\lambda \geq \sqrt{\frac{3}{2}} - 1 = \frac{1}{2+\sqrt{6}}$, and this constant is smaller than our lower bound for λ: $\frac{1}{\sqrt{10}} > \frac{1}{2+\sqrt{6}}$.

Thus, $\varkappa^* \leq 2$, which means that F is a standard self-concordant function. $\square$

Corollary 5.3.1 *If f is a standard self-concordant function, then F is also a standard self-concordant function with Dom $F = \mathcal{E}_f$.*

Finally, let us prove the Implicit Barrier Theorem. Let Φ be a ν-self-concordant barrier for dom $\Phi \subset \mathbb{E}$. We partition the space as follows: $\mathbb{E} = \mathbb{E}_1 \times \mathbb{E}_2$. Define

$$F(x) = \min_y \{\Phi(x, y) : \ (x, y) \in \text{dom } \Phi\}. \tag{5.3.9}$$

We assume that for any $x \in \text{dom } F \subset \mathbb{E}_1$ the solution $y(x)$ of this optimization problem exists and is unique. Then, as we have seen in the proof of Theorem 5.1.11,

$$\nabla_y \Phi(x, y(x)) = 0, \quad \nabla_x \Phi(x, y(x)) \ = \ \nabla F(x).$$

Theorem 5.3.6 *The function F defined by (5.3.9) is a ν-self-concordant barrier.*

Proof In view of Theorem 5.1.11 the function F is standard self-concordant. Let us fix $x \in \text{dom } F$. Then for any direction $z = (h, \delta) \in \mathbb{E}_1 \times \mathbb{E}_2$ we have

$$\langle \nabla F(x), h \rangle_{\mathbb{E}_1}^2 \quad = \quad \langle \nabla_x \Phi(x, y(x)), h \rangle_{\mathbb{E}_1}^2 \quad = \quad \langle \nabla \Phi(x, y(x)), z \rangle_{\mathbb{E}}^2$$

$$\overset{(5.3.6)}{\leq} \quad \nu \langle \nabla^2 \Phi(x, y(x)) z, z \rangle_{\mathbb{E}}.$$

As was shown in the proof of Theorem 5.1.11,

$$\min_{\delta \in \mathbb{E}_2} \langle \nabla^2 \Phi(x, y(x)) z, z \rangle_{\mathbb{E}} = \langle \nabla^2 F(x) h, h \rangle_{\mathbb{E}_1}.$$

Thus, F satisfies definition (5.3.6) of a ν-self-concordant barrier. □

5.3.3 Main Inequalities

Let us show that the local characteristics of a self-concordant barrier (gradient and Hessian) provide us with *global* information about the structure of its domain.

Theorem 5.3.7 *1. Let F be a ν-self-concordant barrier. For any x and y from dom F, we have*

$$\langle \nabla F(x), y - x \rangle \; < \; \nu. \tag{5.3.10}$$

Moreover, if $\langle \nabla F(x), y - x \rangle \geq 0$, then

$$\langle \nabla F(y) - \nabla F(x), y - x \rangle \geq \frac{\langle \nabla F(x), y - x \rangle^2}{\nu - \langle \nabla F(x), y - x \rangle}. \tag{5.3.11}$$

2. A standard self-concordant function F is a ν-self-concordant barrier if and only if

$$F(y) \geq F(x) - \nu \ln \left(1 - \tfrac{1}{\nu} \langle \nabla F(x), y - x \rangle \right) \quad \forall x, y \in \text{dom } F. \tag{5.3.12}$$

Proof 1. Let us fix two points $x, y \in \text{dom } F$. Consider the univariate function

$$\phi(t) = \langle \nabla F(x + t(y - x)), y - x \rangle, \quad t \in [0, 1].$$

If $\phi(0) \leq 0$, then (5.3.10) is trivial. If $\phi(0) = 0$, then (5.3.11) is valid in view of convexity of f. Suppose that $\phi(0) > 0$. In view of inequality (5.3.6), we have

$$\phi'(t) = \langle \nabla^2 F(x + t(y - x))(y - x), y - x \rangle$$

$$\geq \tfrac{1}{\nu} \langle \nabla F(x + t(y - x)), y - x \rangle^2 \; = \; \tfrac{1}{\nu} \phi^2(t).$$

Therefore, $\phi(t)$ increases and is positive for $t \in [0, 1]$. Moreover, for any $t \in [0, 1]$ we have

$$-\tfrac{1}{\phi(t)} + \tfrac{1}{\phi(0)} = \int\limits_0^t \tfrac{\phi'(\tau)}{\phi^2(\tau)} \, d\tau \overset{(5.3.6)}{\geq} \tfrac{1}{\nu} t.$$

This implies that $\langle \nabla F(x), y - x \rangle = \phi(0) < \tfrac{\nu}{t}$ for all $t \in [0, 1]$. Thus, (5.3.10) is proved. At the same time,

$$\phi(t) - \phi(0) \geq \tfrac{\nu \phi(0)}{\nu - t \phi(0)} - \phi(0) = \tfrac{t \phi(0)^2}{\nu - t \phi(0)}, \quad t \in [0, 1].$$

Choosing $t = 1$, we get inequality (5.3.11).

2. Let $\psi(x) = e^{-\frac{1}{\nu} F(x)}$. In view of Lemma 5.3.1, this function is concave. It remains to note that inequality (5.3.12) is equivalent to the condition

$$\psi(y) \leq \psi(x) + \langle \nabla \psi(x), y - x \rangle$$

up to a logarithmic transformation of both sides.   □

Corollary 5.3.2 *Let F be a ν-self-concordant barrier and $h \in \mathbb{E}$ be a recession direction of dom F: $x + \tau h \in dom\, F$ for any $x \in dom\, F$ and $\tau \geq 0$. Then,*

$$\langle \nabla^2 F(x)h, h \rangle^{1/2} \leq \langle -\nabla F(x), h \rangle. \tag{5.3.13}$$

Proof In view of inequality (5.3.10), $\langle \nabla F(x), h \rangle \leq 0$. If dom F does not contain the line $\{x + \tau h, \ \tau \in \mathbb{R}\}$, then inequality (5.3.13) follows from (5.1.27). If it contains the line, then $\langle \nabla F(x), h \rangle = 0$ for all $x \in dom\, F$. This means that F is constant along this line and both sides of inequality (5.3.13) vanish.   □

Corollary 5.3.3 *Let $x, y \in dom\, F$. Then for any $\alpha \in [0, 1)$ we have*

$$F(x + \alpha(y - x)) \leq F(x) - \nu \ln(1 - \alpha). \tag{5.3.14}$$

Proof Let $y(t) = x + t(y - x)$ and $\phi(t) = F(y(t))$. Then

$$\phi'(t) = \langle \nabla F(y(t)), y - x \rangle = \tfrac{1}{1-t} \langle \nabla F(y(t)), y - y(\alpha) \rangle \overset{(5.3.10)}{\leq} \tfrac{\nu}{1-t}.$$

Integrating this inequality in $t \in [0, \alpha)$, we get inequality (5.3.14).   □

Theorem 5.3.8 *Let F be a ν-self-concordant barrier. Then for any $x \in dom\, F$ and $y \in Dom\, F$ such that*

$$\langle \nabla F(x), y - x \rangle \geq 0, \tag{5.3.15}$$

we have

$$\| \, y - x \, \|_x \; \leq \; \nu + 2\sqrt{\nu}. \tag{5.3.16}$$

Proof Let $r = \| \, y - x \, \|_x$ and suppose $r > \sqrt{\nu}$ (otherwise (5.3.16) is trivial). Consider the point $y_\alpha = x + \alpha(y - x)$ with $\alpha = \frac{\sqrt{\nu}}{r} < 1$. In view of our assumption (5.3.15) and inequality (5.1.13) we have

$$\omega \equiv \langle \nabla F(y_\alpha), y - x \rangle \geq \langle \nabla F(y_\alpha) - \nabla F(x), y - x \rangle$$

$$= \tfrac{1}{\alpha} \langle \nabla F(y_\alpha) - \nabla F(x), y_\alpha - x \rangle$$

$$\geq \tfrac{1}{\alpha} \cdot \frac{\|y_\alpha - x\|_x^2}{1 + \|y_\alpha - x\|_x} = \frac{\alpha \|y - x\|_x^2}{1 + \alpha \|y - x\|_x} = \frac{r\sqrt{\nu}}{1 + \sqrt{\nu}}.$$

On the other hand, in view of (5.3.10), we obtain

$$(1 - \alpha)\omega \; = \; \langle \nabla F(y_\alpha), y - y_\alpha \rangle \; \leq \; \nu.$$

Thus,

$$\left(1 - \frac{\sqrt{\nu}}{r} \right) \frac{r\sqrt{\nu}}{1 + \sqrt{\nu}} \leq \nu,$$

and this is exactly (5.3.16). □

We conclude this section by studying the properties of one special point of a convex set.

Definition 5.3.3 Let F be a ν-self-concordant barrier for the set Dom F. The point

$$x_F^* \; = \; \arg \min_{x \in \mathrm{dom}\, F} \; F(x)$$

is called the *analytic center* of the convex set Dom F, generated by the barrier F.

Theorem 5.3.9 *Assume that the analytic center of a ν-self-concordant barrier F exists. Then for any $x \in$ Dom F we have*

$$\| \, x - x_F^* \, \|_{x_F^*} \; \leq \; \nu + 2\sqrt{\nu}.$$

On the other hand, for any $x \in \mathbb{R}^n$ such that $\| \, x - x_F^* \, \|_{x_F^*} \leq 1$, we have $x \in$ Dom F.

Proof The first statement follows from Theorem 5.3.8 since $\nabla F(x_F^*) = 0$. The second statement follows from Theorem 5.1.5. □

Thus, the *asphericity* of the set Dom F with respect to x_F^*, computed in the metric $\| \cdot \|_{x_F^*}$, does not exceed $\nu + 2\sqrt{\nu}$. It is well known that for any convex set in $\mathbb{R}^n$ there exists a metric in which the asphericity of this set is less than or equal to n

(John's Theorem). However, we managed to estimate the asphericity in terms of the *parameter* of the self-concordant barrier. This value does not depend directly on the dimension of the space of variables.

Recall also that if Dom F contains no straight lines the existence of x_F^* implies the boundedness of Dom F (since then $\nabla^2 F(x_F^*)$ is nondegenerate, see Theorem 5.1.6).

Corollary 5.3.4 *Let Dom F be bounded. Then for any $x \in dom\, F$ and $v \in \mathbb{R}^n$ we have*

$$\| v \|_x^* \leq (\nu + 2\sqrt{\nu}) \| v \|_{x_F^*}^* .$$

In other words, for any $x \in dom\, F$ we have

$$\nabla^2 F(x) \succeq \tfrac{1}{(\nu+2\sqrt{\nu})^2}\nabla^2 F(x_F^*). \tag{5.3.17}$$

Proof By Lemma 3.1.20, we get the following representation:

$$\| v \|_x^* \equiv \langle v, [\nabla^2 F(x)]^{-1}v \rangle^{1/2} \;=\; \max\{\langle v, u \rangle \mid \langle \nabla^2 F(x)u, u \rangle \leq 1\}.$$

On the other hand, in view of Theorems 5.1.5 and 5.3.9, we have

$$B \equiv \{y \in \mathbb{R}^n \mid \| y - x \|_x \leq 1\} \subseteq \text{Dom } F$$

$$\subseteq \{y \in \mathbb{R}^n \mid \| y - x_F^* \|_{x_F^*} \leq \nu + 2\sqrt{\nu}\} \equiv B_*.$$

Therefore, using again Theorem 5.3.9, we get the following relations:

$$\| v \|_x^* = \max\{\langle v, y - x \rangle \mid y \in B\} \leq \max\{\langle v, y - x \rangle \mid y \in B_*\}$$

$$= \langle v, x_F^* - x \rangle + (\nu + 2\sqrt{\nu}) \| v \|_{x_F^*}^* .$$

Note that $\| v \|_x^* = \| -v \|_x^*$. Therefore, we can always ensure $\langle v, x_F^* - x \rangle \leq 0$.   □

5.3.4   The Path-Following Scheme

Now we are ready to describe a *barrier model* of the minimization problem. This is a *standard* minimization problem

$$\min\{\langle c, x \rangle \mid x \in Q\} \tag{5.3.18}$$

where Q is a *bounded* closed convex set with nonempty interior, which is a closure of the domain of some ν-self-concordant barrier F.

We are going to solve (5.3.18) by tracing the *central path*:

$$x^*(t) \ = \ \arg \min_{x \in \text{dom} F} f(t; x), \tag{5.3.19}$$

where $f(t; x) = t\langle c, x \rangle + F(x)$ and $t \geq 0$. In view of the first-order optimality condition (1.2.4), any point of the central path satisfies the equation

$$tc + \nabla F(x^*(t)) = 0. \tag{5.3.20}$$

Since the set Q is bounded and F is a closed convex function, the *analytic center* of this set x_F^* exists and it is uniquely defined (see Item 4 of Theorems 3.1.4 and 5.1.6). Moreover, it is a starting point for the central path:

$$x^*(0) \ = \ x_F^*. \tag{5.3.21}$$

In order to follow the central path, we are going to update the points satisfying an *approximate centering condition*:

$$\lambda_{f(t;\cdot)}(x) \ \equiv \ \| \, f'(t; x) \, \|_x^* = \| \, tc + \nabla F(x) \, \|_x^* \ \leq \ \beta, \tag{5.3.22}$$

where the *centering parameter* β is small enough.

Let us show that this is a reasonable goal.

Theorem 5.3.10 *For any $t > 0$, we have*

$$\langle c, x^*(t) \rangle - c^* \leq \tfrac{\nu}{t}, \tag{5.3.23}$$

where c^* is the optimal value of problem (5.3.18). If a point x satisfies the approximate centering condition (5.3.22), then

$$\langle c, x \rangle - c^* \leq \tfrac{1}{t} \left(\nu + \tfrac{(\beta + \sqrt{\nu})\beta}{1 - \beta} \right). \tag{5.3.24}$$

Proof Let x^* be a solution to (5.3.18). In view of (5.3.20) and (5.3.10), we have

$$\langle c, x^*(t) - x^* \rangle = \tfrac{1}{t} \langle \nabla F(x^*(t)), x^* - x^*(t) \rangle \ \leq \ \tfrac{\nu}{t}.$$

Further, let x satisfy (5.3.22). Let $\lambda \ = \ \lambda_{f(t;\cdot)}(x)$. Then, in view of (5.3.5), Theorem 5.2.1, and (5.3.22), we have

$$t\langle c, x - x^*(t) \rangle = \langle f'(t; x) - \nabla F(x), x - x^*(t) \rangle \ \leq \ (\lambda + \sqrt{\nu}) \, \| \, x - x^*(t) \, \|_x$$

$$\leq (\lambda + \sqrt{\nu}) \frac{\lambda}{1 - \lambda} \ \leq \ \frac{(\beta + \sqrt{\nu})\beta}{1 - \beta}. \qquad\qquad \square$$

Let us analyze now one step of a path-following scheme. It differs from the updating rule (5.2.14) only by the origin of the objective vector.

Assume that $x \in \text{dom}\, F$. Consider the following iterate:

$$
\begin{aligned}
t_+ &= t + \frac{\gamma}{\|c\|_x^*}, \\[2mm]
x_+ &= x - \frac{1}{1+\xi}[\nabla^2 F(x)]^{-1}(t_+ c + \nabla F(x)), \\[2mm]
&\text{where } \xi = \frac{\lambda^2}{1+\lambda} \text{ and } \lambda = \|t_+ c + \nabla F(x)\|_x^*.
\end{aligned}
\tag{5.3.25}
$$

From Lemma 5.2.2 we know that if $\beta = \beta(\tau) = \tau^2(1 + \tau + \frac{\tau}{1+\tau+\tau^2})$ with $\tau \in [0, \frac{1}{2}]$ and x satisfies approximate centering condition (5.3.22), then for γ, such that

$$
|\gamma| \leq \tau - \tau^2(1 + \tau + \tfrac{\tau}{1+\tau+\tau^2}),
\tag{5.3.26}
$$

we have again $\|t_+ c + \nabla F(x_+)\|_{x_+}^* \leq \beta$.

Let us prove now that the increase of t in the scheme (5.3.25) is sufficiently large.

Lemma 5.3.2 *Let x satisfy (5.3.22). Then*

$$
\|c\|_x^* \leq \tfrac{1}{t}(\beta + \sqrt{\nu}).
\tag{5.3.27}
$$

Proof Indeed, in view of (5.3.22) and (5.3.5), we have

$$
t\,\|c\|_x^* = \|f'(t;x) - \nabla F(x)\|_x^* \leq \|f'(t;x)\|_x^* + \|\nabla F(x)\|_x^*
$$

$$
\leq \beta + \sqrt{\nu}. \qquad \qquad \qquad \qquad \square
$$

Let us now fix some reasonable values of parameters in method (5.3.25). In the remaining part of this chapter we always assume that

$$
\begin{aligned}
\tau &= 0.29, \quad \beta = \beta(\tau) \approx 0.126, \\[2mm]
\gamma &= \tau - \beta(\tau) \approx 0.164 \quad \Rightarrow \quad \gamma^{-1} < 6.11.
\end{aligned}
\tag{5.3.28}
$$

We have proved that it is possible to follow the central path, using the rule (5.3.25). Note that we can either increase or decrease the current value of t. The lower

estimate for the rate of *increasing t* is

$$t_+ \geq \left(1 + \frac{\gamma}{\beta+\sqrt{\nu}}\right) \cdot t,$$

and the upper estimate for the rate of *decreasing t* is

$$t_+ \leq \left(1 - \frac{\gamma}{\beta+\sqrt{\nu}}\right) \cdot t.$$

Thus, the general scheme for solving the problem (5.3.18) is as follows.

Main path-following scheme

---

0. Set $t_0 = 0$. Choose an accuracy $\epsilon > 0$ and $x_0 \in \operatorname{dom} F$ such that

$$\| \nabla F(x_0) \|_{x_0}^* \leq \beta.$$

1. kth iteration ($k \geq 0$). Set

$$t_{k+1} = t_k + \frac{\gamma}{\|c\|_{x_k}^*},$$

$$x_{k+1} = x_k - \frac{1}{1+\xi_k}[\nabla^2 F(x_k)]^{-1}(t_{k+1}c + \nabla F(x_k)),$$

where $\xi_k = \frac{\lambda_k^2}{1+\lambda_k}$, and $\lambda_k = \|t_{k+1}c + \nabla F(x_k)\|_{x_k}^*$.

2. Stop the process if $t_k \geq \frac{1}{\epsilon}\left(\nu + \frac{(\beta+\sqrt{\nu})\beta}{1-\beta}\right)$.

$$(5.3.29)$$

Let us derive a complexity bound for the above scheme.

Theorem 5.3.11 *Method (5.3.29) terminates after N steps at most, where*

$$N \leq O\left(\sqrt{\nu} \ln \frac{\nu\|c\|_{x_F}^*}{\epsilon}\right).$$

Moreover, at the moment of termination we have $\langle c, x_N \rangle - c^* \leq \epsilon$.

Proof Note that $r_0 \equiv \| x_0 - x_F^* \|_{x_0} \le \frac{\beta}{1-\beta}$ (see Theorem 5.2.1). Therefore, in view of Theorem 5.1.7 we have

$$\frac{\gamma}{t_1} = \| c \|_{x_0}^* \le \frac{1}{1-r_0} \| c \|_{x_F^*}^* \le \frac{1-\beta}{1-2\beta} \| c \|_{x_F^*}^* .$$

Thus, $t_k \ge \frac{\gamma(1-2\beta)}{(1-\beta)\|c\|_{x_F^*}^*} \left(1 + \frac{\gamma}{\beta+\sqrt{\nu}} \right)^{k-1}$ for all $k \ge 1$. $\quad\square$

Let us discuss now the above complexity bound. The main term there is

$$6.11 \sqrt{\nu} \ \ln \frac{\nu\|c\|_{x_F^*}^*}{\epsilon}.$$

Note that the value $\nu \| c \|_{x_F^*}^*$ estimates from above the variation of the linear function $\langle c, x \rangle$ over the set Dom F (see Theorem 5.3.9). Thus, the ratio $\frac{\epsilon}{\nu\|c\|_{x_F^*}^*}$ can be seen as the *relative accuracy* of the solution.

The process (5.3.29) has one drawback. Sometimes it is difficult to satisfy its starting condition

$$\| \nabla F(x_0) \|_{x_0}^* \le \beta.$$

In this case, we need an additional process for *computing* an appropriate starting point. We analyze the corresponding strategies in the next section.

5.3.5 Finding the Analytic Center

Thus, our current goal is to find an approximation to the *analytic center* of the set Dom F. Let us look at the following minimization problem:

$$\min\{F(x) \mid x \in \text{dom } F\}, \tag{5.3.30}$$

where F is a ν-self-concordant barrier. In view of the needs of the previous section, we accept an approximate solution $\bar{x} \in \text{dom } F$ to this problem, which satisfies the inequality

$$\| \nabla F(\bar{x}) \|_{\bar{x}}^* \le \beta,$$

for certain $\beta \in (0, 1)$.

As we have already discussed in Sect. 5.2, we can apply two different minimization strategies. The first one is a straightforward implementation of the Intermediate Newton's Method and the second one is based on a path-following approach.

Consider the first scheme.

Intermediate Newton's Method for finding the analytic center

---

   0. Choose $y_0 \in \operatorname{dom} F$.

   1. *k*th iteration ($k \geq 0$). Set

$$y_{k+1} = y_k - \frac{[\nabla^2 F(y_k)]^{-1} \nabla F(y_k)}{1+\xi_k},$$

(5.3.31)

   where $\xi_k = \frac{\lambda_k^2}{1+\lambda_k}$ and $\lambda_k = \| \nabla F(y_k) \|_{y_k}^*$.

   2. Stop the process if $\| \nabla F(y_k) \|_{y_k}^* \leq \beta$.

As we have seen already, this method needs $O(F(y_0) - F(x_F^*))$ iterations to enter to the region of quadratic convergence.

To implement the path-following approach, we need to choose some $y_0 \in \operatorname{dom} F$ and define the *auxiliary central path*:

$$y^*(t) = \arg \min_{y \in \operatorname{dom} F} [-t \langle \nabla F(y_0), y \rangle + F(y)],$$

where $t \geq 0$. Since this trajectory satisfies the equation

$$\nabla F(y^*(t)) = t \nabla F(y_0),$$

(5.3.32)

it connects two points, the starting point y_0 and the analytic center x_F^*:

$$y^*(1) = y_0, \quad y^*(0) = x_F^*.$$

As was shown in Lemma 5.2.2, we can follow this trajectory by the process (5.3.25) with *decreasing* t.

Let us estimate the rate of convergence of the auxiliary central path $y^*(t)$ to the analytic center in terms of the *barrier parameter*.

Lemma 5.3.3 *For any $t \geq 0$, we have*

$$\| \nabla F(y^*(t)) \|_{y^*(t)}^* \leq (\nu + 2\sqrt{\nu}) \| \nabla F(y_0) \|_{x_F^*}^* \cdot t.$$

Proof This estimate follows from (5.3.32) and Corollary 5.3.4. □

Let us look now at the corresponding algorithmic scheme.

Auxiliary Path-Following Scheme

---

0. Choose $y_0 \in \operatorname{dom} F$. Set $t_0 = 1$.
1. k**th iteration** ($k \geq 0$). Set

$$t_{k+1} = t_k - \frac{\gamma}{\|\nabla F(y_0)\|_{y_k}^*},$$

$$y_{k+1} = y_k - \frac{1}{1+\xi_k}[\nabla^2 F(y_k)]^{-1}(-t_{k+1}\nabla F(y_0) + \nabla F(y_k)),$$

where $\xi_k = \frac{\lambda_k^2}{1+\lambda_k}$ and $\lambda_k = \|t_{k+1}\nabla F(y_0) - \nabla F(y_k)\|_{y_k}^*$.

2. Stop the process if $\| \nabla F(y_k) \|_{y_k}^* \leq \tau$. Set $\xi_k = \frac{\lambda_F(y_k)^2}{1+\lambda_F(y_k)}$
and $\bar{x} = y_k - \frac{1}{1+\xi_k}[\nabla^2 F(y_k)]^{-1}\nabla F(y_k)$.

$$(5.3.33)$$

Note that the above scheme follows the auxiliary central path $y^*(t)$ as $t_k \to 0$. It updates the points $\{y_k\}$ satisfying the approximate centering condition

$$\| -t_k\nabla F(y_0) + \nabla F(y_k) \|_{y_k}^* \leq \beta.$$

The termination criterion of this process,

$$\lambda_k = \| \nabla F(y_k) \|_{y_k}^* \leq \tau,$$

guarantees that $\| \nabla F(\bar{x}) \|_{\bar{x}}^* \leq \beta(\tau)$ (see Theorem 5.2.2). Let us derive a complexity bound for this process.

Theorem 5.3.12 *The process (5.3.33) terminates no later than after*

$$\frac{1}{\gamma}(\beta + \sqrt{\nu}) \ln \left[\frac{1}{\gamma}(\nu + 2\sqrt{\nu}) \| \nabla F(y_0) \|_{x_F^*}^* \right]$$

iterations.

Proof Recall that our parameters are fixed by (5.3.28). Note that $t_0 = 1$. Therefore, in view of Lemmas 5.2.2 and 5.3.2, we have

$$t_{k+1} \leq \left(1 - \frac{\gamma}{\beta+\sqrt{\nu}}\right) t_k \leq \exp\left(-\frac{\gamma(k+1)}{\beta+\sqrt{\nu}}\right) t_0.$$

Further, in view of Lemma 5.3.3, we obtain

$$\| \nabla F(y_k) \|_{y_k}^* = \| (-t_k \nabla F(x_0) + \nabla F(y_k)) + t_k \nabla F(y_0) \|_{y_k}^*$$

$$\leq \beta + t_k \| \nabla F(y_0) \|_{y_k}^* \leq \beta + t_k(\nu + 2\sqrt{\nu}) \| \nabla F(y_0) \|_{x_F^*}^* .$$

Thus, the process is terminated at most when the following inequality holds:

$$t_k(\nu + 2\sqrt{\nu}) \| \nabla F(y_0) \|_{x_F^*}^* \leq \tau - \beta(\tau) = \gamma. \qquad \square$$

The principal term in the complexity bound of the auxiliary path-following scheme is

$$6.11\sqrt{\nu}[\ln \nu + \ln \| \nabla F(y_0) \|_{x_F^*}^*]$$

and for the auxiliary Intermediate Newton's method it is $O(F(y_0) - F(x_F^*))$. These estimates cannot be compared directly. However, as we have proved in Sect. 5.2.2 by another reasoning the path-following approach is much more efficient. Note also that its complexity estimate naturally fits the complexity of the main path-following process. Indeed, if we apply (5.3.29) with (5.3.33), we get the following complexity bound for the whole process:

$$6.11\sqrt{\nu}\left[2 \ln \nu + \ln \| \nabla F(y_0) \|_{x_F^*}^* + \ln \| c \|_{x_F^*}^* + \ln \frac{1}{\epsilon}\right].$$

To conclude this section, note that for some problems it is difficult even to point out a starting point $y_0 \in \text{dom } F$. In such cases, we should apply one more auxiliary minimization process, which is similar to the process (5.3.33). We discuss this situation in the next section.

5.3.6 Problems with Functional Constraints

Let us consider the following minimization problem:

$$\min_{x \in Q}\{f_0(x) : f_j(x) \leq 0, \ j = 1 \ldots m\}, \tag{5.3.34}$$

where Q is a simple bounded closed convex set with nonempty interior and all functions f_j, $j = 0 \ldots m$, are convex. We assume that the problem satisfies the Slater condition: There exists an $\bar{x} \in \text{int } Q$ such that $f_j(\bar{x}) < 0$ for all $j = 1 \ldots m$.

Let us assume that we know an upper bound $\bar{\xi}$ such that $f_0(x) < \bar{\xi}$ for all $x \in Q$. Then, introducing two additional variables ξ and $\varkappa$, we can rewrite this problem in the standard form:

$$\min_{\substack{\xi \leq \bar{\xi}, \, \varkappa \leq 0, \\ x \in Q}} \{\xi : \ f_0(x) \leq \xi, \ f_j(x) \leq \varkappa, \ j = 1 \ldots m\}. \tag{5.3.35}$$

Note that we can apply interior-point methods to this problem only if we are able to construct a self-concordant barrier for the feasible set. In the current situation, this means that we should be able to construct the following barriers:

- A self-concordant barrier $F_Q(x)$ for the set Q.
- A self-concordant barrier $F_0(x, \xi)$ for the epigraph of the objective function $f_0(x)$.
- Self-concordant barriers $F_j(x, \varkappa)$ for the epigraphs of functional constraints $f_j(x)$.

Let us assume that we can do that. Then the resulting self-concordant barrier for the feasible set of problem (5.3.35) is as follows:

$$\hat{F}(x, \xi, \varkappa) = F_Q(x) + F_0(x, \xi) + \sum_{j=1}^{m} F_j(x, \varkappa) - \ln(\bar{\xi} - \xi) - \ln(-\varkappa).$$

The parameter of this barrier is

$$\hat{\nu} = \nu_Q + \nu_0 + \sum_{j=1}^{m} \nu_j + 2, \tag{5.3.36}$$

where $\nu_{(\cdot)}$ are the parameters of the corresponding barriers.

Note that it could still be difficult to find a starting point from $\text{dom} \, \hat{F}$. This domain is an intersection of the set Q with epigraphs of the objective function and constraints, and with two additional linear constraints $\xi \leq \bar{\xi}$ and $\varkappa \leq 0$. If we have a point $x_0 \in \text{int} \, Q$, then we can choose ξ_0 and $\varkappa_0$ large enough to guarantee

$$f_0(x_0) \ < \ \xi_0 \ < \ \bar{\xi}, \quad f_j(x_0) \ < \ \varkappa_0, \ j = 1 \ldots m.$$

Then, only constraint $\varkappa \leq 0$ will be violated.

In order to simplify our analysis, let us change the notation. From now on, we consider the problem

$$\min_{z \in S} \{\langle c, z \rangle : \ \langle d, z \rangle \leq 0\}, \tag{5.3.37}$$

where $z = (x, \xi, \varkappa)$, $\langle c, z \rangle \equiv \xi$, $\langle d, z \rangle \equiv \varkappa$ and S is the feasible set of problem (5.3.35) without the constraint $\varkappa \leq 0$. Note that we know a self-concordant barrier $F(z)$ for the set S, and we can easily find a point $z_0 \in \text{int} \, S$. Moreover, in

view of our assumptions, the set

$$S(\alpha) \ = \ \{z \in S \mid \langle d, z \rangle \leq \alpha\}$$

is bounded and, for α large enough, it has nonempty interior.

The process of solving problem (5.3.37) consists of three stages.

1. Choose a starting point $z_0 \in \text{int } S$ and some initial gap $\Delta > 0$. Set $\alpha = \langle d, z_0 \rangle + \Delta$. If $\alpha \leq 0$, then we can use the two-stage process described in Sect. 5.3.5. Otherwise, we do the following. First, we find an approximate analytic center of the set $S(\alpha)$, generated by the barrier

$$\tilde{F}(z) = F(z) - \ln(\alpha - \langle d, z \rangle).$$

Namely, we find a point $\tilde{z}$ satisfying the condition

$$\lambda_{\tilde{F}}(\tilde{z}) \ \equiv \ \langle \nabla F(\tilde{z}) + \tfrac{d}{\alpha - \langle d, \tilde{z} \rangle}, [\nabla^2 \tilde{F}(\tilde{z})]^{-1} \left(\nabla F(\tilde{z}) + \tfrac{d}{\alpha - \langle d, \tilde{z} \rangle} \right) \rangle^{1/2} \leq \beta.$$

In order to generate such a point, we can use the auxiliary schemes discussed in Sect. 5.3.5.

2. The next stage consists in following the central path $z(t)$ defined by the equation

$$td + \nabla \tilde{F}(z(t)) \ = \ 0, \quad t \geq 0.$$

Note that the previous stage provides us with a reasonable approximation to the analytic center $z(0)$. Therefore, we can follow this path, using the process (5.3.25). This trajectory leads us to the solution of the minimization problem

$$\min\{\langle d, z \rangle \mid z \in S(\alpha)\}.$$

In view of the Slater condition for problem (5.3.37), the optimal value of this problem is strictly negative.

The goal of this stage consists in finding an approximation to the analytic center of the set

$$\bar{S} = \{z \in S(\alpha) \mid \langle d, z \rangle \leq 0\}$$

generated by the barrier $\bar{F}(z) = \tilde{F}(z) - \ln(-\langle d, z \rangle)$. This point, z_*, satisfies the equation

$$\nabla \tilde{F}(z_*) - \tfrac{d}{\langle d, z_* \rangle} = 0.$$

Therefore, z_* is a point of the central path $z(t)$. The corresponding value of the penalty parameter t_* is

$$t_* = -\frac{1}{\langle d, z_* \rangle} > 0.$$

This stage terminates with a point $\bar{z}$ satisfying the condition

$$\lambda_{\tilde{F}}(\bar{z}) \;\equiv\; \langle \nabla \tilde{F}(\bar{z}) - \frac{d}{\langle d, \bar{z} \rangle}, [\nabla^2 \tilde{F}(\bar{z})]^{-1} \left(\nabla \tilde{F}(\bar{z}) - \frac{d}{\langle d, \bar{z} \rangle} \right) \rangle^{1/2} \leq \beta.$$

3. Note that $\nabla^2 \bar{F}(z) \succeq \nabla^2 \tilde{F}(z)$. Therefore, the point $\bar{z}$, computed at the previous stage, satisfies the inequality

$$\lambda_{\bar{F}}(\bar{z}) \;\equiv\; \langle \nabla \tilde{F}(\bar{z}) - \frac{d}{\langle d, \bar{z} \rangle}, [\nabla^2 \bar{F}(\bar{z})]^{-1} \left(\nabla \tilde{F}(\bar{z}) - \frac{d}{\langle d, \bar{z} \rangle} \right) \rangle^{1/2} \leq \beta.$$

This means that we have a good approximation of the analytic center of the set $\bar{S}$, and we can apply the main path-following scheme (5.3.29) to solve the problem

$$\min\{\langle c, z \rangle : \; z \in \bar{S}\}.$$

Clearly, this problem is equivalent to (5.3.37).

We omit the detailed complexity analysis of the above three-stage scheme. It can be done similarly to the analysis of Sect. 5.3.5. The main term in the complexity of this scheme is proportional to the product of $\sqrt{\hat{\nu}}$ (see (5.3.36)) and the sum of the logarithm of the desired accuracy ϵ with logarithms of some structural characteristics of the problem (size of the region, depth of Slater condition, etc.).

Thus, we have shown that the interior point methods can be applied to all problems, for which we can point out some self-concordant barriers for the basic feasible set Q and for the epigraphs of functional constraints. Our main goal now is to describe the classes of convex problems for which such barriers can be constructed in a computable form. Note that we have an exact characteristic of the quality of self-concordant barrier. This is the value of its parameter. The smaller it is, the more efficient will be the corresponding path-following scheme. In the next section, we discuss our possibilities in applying the developed theory to particular convex problems.

5.4   Applications to Problems with Explicit Structure

(Bounds on parameters of self-concordant barriers; Linear and quadratic optimization; Semidefinite optimization; Extremal ellipsoids; Constructing self-concordant barriers for particular sets; Separable problems; Geometric optimization; Approximation in ℓ_p-norms; Choice of optimization scheme.)

5.4.1 Lower Bounds for the Parameter of a Self-concordant Barrier

In the previous section, we discussed a path-following scheme for solving the following problem:

$$\min_{x \in Q} \langle c, x \rangle, \tag{5.4.1}$$

where Q is a closed convex set with nonempty interior, for which we know a ν-self-concordant barrier $F(\cdot)$. Using such a barrier, we can solve (5.4.1) in $O\left(\sqrt{\nu} \cdot \ln \frac{\nu}{\epsilon}\right)$ iterations of a path-following scheme. Recall that the most difficult part of each iteration is the solution of a system of linear equations.

In this section, we study the limits of applicability of this approach. We discuss the lower and upper bounds for the parameters of self-concordant barriers. We also discuss some classes of convex problems for which the model (5.4.1) can be created in a computable form.

Let us start from the lower bounds on the barrier parameters.

Lemma 5.4.1 *Let f be a ν-self-concordant barrier for the interval $(\alpha, \beta) \subset \mathbb{R}$, $\alpha < \beta < \infty$, where we admit the value $\alpha = -\infty$. Then*

$$\nu \geq \varkappa \overset{def}{=} \sup_{t \in (\alpha, \beta)} \frac{(f'(t))^2}{f''(t)} \geq 1.$$

Proof Note that $\nu \geq \varkappa$ by definition. Let us assume that $\varkappa < 1$. Since f is a convex barrier function for (α, β), there exists a value $\bar{\alpha} \in (\alpha, \beta)$ such that $f'(t) > 0$ for all $t \in [\bar{\alpha}, \beta)$.

Consider the function $\phi(t) = \frac{(f'(t))^2}{f''(t)}$, $t \in [\bar{\alpha}, \beta)$. Then, since $f'(t) > 0$, $f(\cdot)$ is standard self-concordant, and $\phi(t) \leq \varkappa < 1$, we have

$$\phi'(t) = 2 f'(t) - \left(\frac{f'(t)}{f''(t)} \right)^2 f'''(t)$$

$$= f'(t) \left(2 - \frac{f'(t)}{\sqrt{f''(t)}} \cdot \frac{f'''(t)}{[f''(t)]^{3/2}} \right) \geq 2(1 - \sqrt{\varkappa}) f'(t).$$

Hence, for all $t \in [\bar{\alpha}, \beta)$ we obtain $\phi(t) \geq \phi(\bar{\alpha}) + 2(1 - \sqrt{\varkappa})(f(t) - f(\bar{\alpha}))$. This is a contradiction since f is a barrier function and ϕ is bounded from above. $\square$

Corollary 5.4.1 *Let F be a ν-self-concordant barrier for $Q \subset \mathbb{E}$. Then $\nu \geq 1$.*

Proof Indeed, let $x \in \text{int } Q$. Since $Q \subset \mathbb{E}$, there exists a nonzero direction $u \in \mathbb{E}$ such that the line $\{y = x + tu, \ t \in \mathbb{R}\}$ intersects the boundary of the set Q. Therefore, considering function $f(t) = F(x + tu)$, and using Lemma 5.4.1, we get the result. $\square$

Let us prove a simple lower bound for parameters of self-concordant barriers for unbounded sets.

Let Q be a closed convex set with nonempty interior. Consider $\bar{x} \in \operatorname{int} Q$. Assume that there exists a nontrivial set of *recession* directions $\{p_1, \ldots, p_k\}$ of the set Q:

$$\bar{x} + \alpha p_i \in Q \quad \forall \alpha \geq 0, \quad i = 1, \ldots, k.$$

Theorem 5.4.1 *Let the positive coefficients $\{\beta_i\}_{i=1}^k$ satisfy the condition*

$$\bar{x} - \beta_i \ p_i \notin \operatorname{int} Q, \quad i = 1, \ldots, k.$$

If for some positive $\alpha_1, \ldots, \alpha_k$ we have $\bar{y} = \bar{x} - \sum\limits_{i=1}^k \alpha_i p_i \in Q$, then the parameter v of any self-concordant barrier for the set Q satisfies the inequality:

$$v \geq \sum_{i=1}^k \frac{\alpha_i}{\beta_i}.$$

Proof Let F be a v-self-concordant barrier for the set Q. Since p_i is a recession direction, by Theorem 5.1.14 we have

$$\langle \nabla F(\bar{x}), -p_i \rangle \geq \langle \nabla^2 F(\bar{x}) p_i, p_i \rangle^{1/2} \equiv \| p_i \|_{\bar{x}} .$$

Note that $\bar{x} - \beta_i \ p_i \notin Q$. Therefore, in view of Theorem 5.1.5, the norm of direction p_i is large enough: $\beta_i \| p_i \|_{\bar{x}} \geq 1$. Hence, in view of Theorem 5.3.7, we obtain

$$v \geq \langle \nabla F(\bar{x}), \bar{y} - \bar{x} \rangle = \langle \nabla F(\bar{x}), -\sum_{i=1}^k \alpha_i p_i \rangle$$

$$\geq \sum_{i=1}^k \alpha_i \| p_i \|_{\bar{x}} \geq \sum_{i=1}^k \frac{\alpha_i}{\beta_i}. \qquad \qquad \square$$

5.4.2 Upper Bound: Universal Barrier and Polar Set

Let us present now an existence theorem for self-concordant barriers. Consider a closed convex set Q, $\operatorname{int} Q \neq \emptyset$, and assume that Q contains no straight lines. Define a *polar set* of Q with respect to some point $\bar{x} \in \operatorname{int} Q$ as follows:

$$P(\bar{x}) = \{s \in \mathbb{R}^n \mid \langle s, x - \bar{x} \rangle \leq 1, \ \forall x \in Q\}.$$

It can be proved that for any $x \in \text{int } Q$ the set $P(x)$ is a bounded closed convex set with nonempty interior. It always contains the origin.

Define $V(x) = \text{vol}_n P(x)$.

Theorem 5.4.2 *There exist absolute constants c_1 and c_2, such that the function*

$$U(x) = c_1 \cdot \ln V(x)$$

is a $(c_2 \cdot n)$-self-concordant barrier for Q. □

We drop the proof of this statement since it is very technical. □

The function $U(\cdot)$ is called the *Universal Barrier* for the set Q. Note that the analytical complexity of problem (5.4.1), equipped with a universal barrier, is $O\left(\sqrt{n} \cdot \ln \frac{n}{\epsilon}\right)$ calls of oracle. Recall that such efficiency estimate is *impossible* for the methods based on a local Black-Box oracle (see Theorem 3.2.8).

The statement of Theorem 5.4.2 is mainly of theoretical interest. Indeed, in general, the value $U(x)$ cannot easily be computed. However, Theorem 5.4.2 demonstrates that self-concordant barriers, in principle, can be found for *any* convex set. Thus, the applicability of this approach is restricted only by our ability to construct a *computable* self-concordant barrier, hopefully with a small value of the parameter. The process of creating the *barrier model* of the initial problem can hardly be described in a formal way. For each particular problem, there could be many different barrier models, and we should choose the best one, taking into account the value of the parameter of the self-concordant barrier, the complexity of the computation of its gradient and Hessian, and the complexity of the solution of the corresponding Newton system.

In the remaining part of this section we will see how this can be done for some *standard* problem classes of Convex Optimization.

5.4.3 Linear and Quadratic Optimization

Let us start from a problem of Linear Optimization:

$$\min_{x \in \mathbb{R}^n_+} \{\langle c, x \rangle : Ax = b\}, \tag{5.4.2}$$

where A is an $(m \times n)$-matrix, $m < n$. The basic feasible set in this problem is represented by the *positive orthant*, the set of all vectors with nonnegative coefficients in $\mathbb{R}^n$. It can be equipped with the following self-concordant barrier:

$$F(x) = -\sum_{i=1}^n \ln x^{(i)}, \quad \nu = n, \tag{5.4.3}$$

(see Example 5.3.1 and Theorem 5.3.2). This barrier is called the *standard logarithmic barrier* for $\mathbb{R}^n_+$.

In order to solve problem (5.4.2), we have to use a restriction of the barrier F onto the affine subspace $\{x : Ax = b\}$. Since this restriction is an n-self-concordant barrier (see Theorem 5.3.3), the complexity bound for problem (5.4.2) is $O\left(\sqrt{n} \cdot \ln \frac{n}{\epsilon}\right)$ iterations of a path-following scheme.

Let us prove that the standard logarithmic barrier is optimal for $\mathbb{R}^n_+$.

Lemma 5.4.2 *The parameter ν of any self-concordant barrier for $\mathbb{R}^n_+$ satisfies inequality $\nu \geq n$.*

Proof Let us choose

$$\bar{x} = \bar{e}_n \equiv (1, \ldots, 1)^T \in \operatorname{int} \mathbb{R}^n_+,$$

$$p_i = e_i, \quad i = 1 \ldots n,$$

where e_i is the ith coordinate vector of $\mathbb{R}^n$. In this case the conditions of Theorem 5.4.1 are satisfied with $\alpha_i = \beta_i = 1, i = 1 \ldots n$. Therefore,

$$\nu \geq \sum_{i=1}^n \frac{\alpha_i}{\beta_i} = n. \qquad \square$$

Note that the above lower bound is valid only for the whole set $\mathbb{R}^n_+$. The lower bound for the intersection $\{x \in \mathbb{R}^n_+ \mid Ax = b\}$ can be smaller.

Self-concordant barriers for cones usually have one important property, which is called *logarithmic homogeneity* (e.g. (5.4.3)).

Definition 5.4.1 A function $F \in C^2(\mathbb{E})$ with $\operatorname{Dom} F = K$, where K is a closed convex cone, is called logarithmically homogeneous if there exists a constant $\nu \geq 1$ such that

$$F(\tau x) = F(x) - \nu \ln \tau, \quad \forall x \in \operatorname{int} K, \ \tau > 0. \tag{5.4.4}$$

This simple property has surprisingly many interesting consequences, one of which makes the computation of the barrier parameter completely trivial.

Lemma 5.4.3 *Let F be a logarithmically homogeneous self-concordant barrier for a convex cone K which contains no straight lines. Then for any $x \in \operatorname{int} K$ and $\tau > 0$ we have*

$$\nabla F(\tau x) = \tfrac{1}{\tau} \nabla F(x), \quad \nabla^2 F(\tau x) = \tfrac{1}{\tau^2} \nabla^2 F(x), \tag{5.4.5}$$

$$\langle \nabla F(x), x \rangle = -\nu, \quad \nabla^2 F(x) x = -\nabla F(x), \tag{5.4.6}$$

$$\langle \nabla^2 F(x) x, x \rangle = \nu, \quad \langle \nabla F(x), [\nabla^2 F(x)]^{-1} \nabla F(x) \rangle = \nu. \tag{5.4.7}$$

Proof Differentiating identity (5.4.4) in x, we get the first identity in (5.4.5). Differentiating the latter identity in x again, we get the second relation in (5.4.5).

Differentiating identity (5.4.4) in τ and taking $\tau = 1$, we get the first identity in (5.4.6). Differentiating it in x, we obtain the second identity in this line.

Finally, substituting the last expression in (5.4.6) into the first one, we get the first identity in (5.4.7). Since K contains no straight lines $\nabla^2 F(x)$ is non-degenerate. Therefore, $x = -[\nabla^2 F(x)]^{-1}\nabla F(x)$, and we get the last expression in (5.4.7).   □

Thus, for logarithmically homogeneous barriers, the degree of homogeneity is always equal to the barrier parameter (see the second identity in (5.4.7)).

Let us look now at the quadratically constrained quadratic optimization problem:

$$\min_{x \in \mathbb{R}^n} \{\, q_0(x) = \alpha_0 + \langle a_0, x \rangle + \tfrac{1}{2}\langle A_0 x, x \rangle,$$

(5.4.8)

$$q_i(x) = \alpha_i + \langle a_i, x \rangle + \tfrac{1}{2}\langle A_i x, x \rangle \leq \beta_i, \ i = 1 \ldots m\},$$

where A_i are some positive semidefinite $(n \times n)$-matrices. Let us rewrite this problem in the standard form:

$$\min_{x \in \mathbb{R}^n, \tau \in \mathbb{R}} \{\tau : \ q_0(x) \leq \tau, \ q_i(x) \leq \beta_i, \ i = 1 \ldots m\}.$$

(5.4.9)

The feasible set of this problem can be equipped with the following self-concordant barrier:

$$F(x, \tau) = -\ln(\tau - q_0(x)) - \sum_{i=1}^{m} \ln(\beta_i - q_i(x)), \quad \nu = m + 1,$$

(see Example 5.3.1, and Theorem 5.3.2). Thus, the complexity bound for problem (5.4.8) is $O\left(\sqrt{m + 1} \cdot \ln \frac{m}{\epsilon}\right)$ iterations of a path-following scheme. Note that this estimate *does not depend* on n.

In some applications, the functional components of the problem include a nonsmooth quadratic term of the form $\| Ax - b \|$, where the norm is standard Euclidean. Let us show that we can treat such terms using an interior-point technique.

Lemma 5.4.4 *The function*

$$F(x, t) = -\ln(t^2 - \| x \|^2)$$

is a 2-self-concordant barrier for the convex cone[5]

$$K_2 = \{(x, t) \in \mathbb{R}^{n+1} \mid t \geq \| x \|\}.$$

---

[5]Depending on the field, this set has different names: Lorentz cone, ice-cream cone, second-order cone.

Proof Let us fix a point $z = (x, t) \in$ int K_2 and a nonzero direction $u = (h, \tau) \in \mathbb{R}^{n+1}$. Let $\xi(\alpha) = (t + \alpha\tau)^2 - \| x + \alpha h \|^2$. We need to compare the derivatives of the function

$$\phi(\alpha) = F(z + \alpha u) = -\ln \xi(\alpha)$$

at $\alpha = 0$. Let $\phi^{(\cdot)} = \phi^{(\cdot)}(0), \xi^{(\cdot)} = \xi^{(\cdot)}(0)$. Then

$$\xi' = 2(t\tau - \langle x, h\rangle), \quad \xi'' = 2(\tau^2 - \| h \|^2), \quad \xi''' = 0,$$

$$\phi' = -\frac{\xi'}{\xi}, \quad \phi'' = \left(\frac{\xi'}{\xi}\right)^2 - \frac{\xi''}{\xi}, \quad \phi''' = 3\frac{\xi'\xi''}{\xi^2} - 2\left(\frac{\xi'}{\xi}\right)^3.$$

Note that inequality $2\phi'' \geq (\phi')^2$ is equivalent to $(\xi')^2 \geq 2\xi\xi''$. Thus, we need to prove that for any (h, τ) we have

$$(t\tau - \langle x, h\rangle)^2 \geq (t^2 - \| x \|^2)(\tau^2 - \| h \|^2).$$

After opening the brackets and cancellation, we come to the inequality

$$\tau^2 \|x\|^2 + t^2 \|h\|^2 + \langle x, h\rangle^2 - 2\tau t\langle x, h\rangle \geq \|x\|^2\|h\|^2.$$

Minimizing the left-hand side in τ, we get inequality

$$t^2 \|h\|^2 + \langle x, h\rangle^2 - t^2 \frac{\langle x, h\rangle^2}{\|x\|^2} \geq \|x\|^2\|h\|^2,$$

$$\Updownarrow$$

$$\|h\|^2(t^2 - \|x\|^2) \geq \langle x, h\rangle^2 \left(\frac{t^2}{\|x\|^2} - 1\right),$$

which is valid since $t \geq \|x\|$.

Finally, since $0 \leq \frac{\xi\xi''}{(\xi')^2} \leq \frac{1}{2}$ and $[1 - \xi]^{3/2} \geq 1 - \frac{3}{2}\xi$, we get the following:

$$\frac{|\phi'''|}{(\phi'')^{3/2}} = 2\frac{|\xi'|\cdot|(\xi')^2 - \frac{3}{2}\xi\xi''|}{[(\xi')^2 - \xi\xi'']^{3/2}} \leq 2. \qquad \square$$

Let us prove that the barrier described in the above statement is optimal for the second-order cone.

Lemma 5.4.5 *The parameter v of any self-concordant barrier for the set K_2 satisfies the inequality $v \geq 2$.*

Proof Let us choose $\bar{z} = (0, 1) \in$ int K_2 and some $h \in \mathbb{R}^n$, $\| h \| = 1$. Define

$$p_1 = (h, 1), \quad p_2 = (-h, 1), \quad \alpha_1 = \alpha_2 = \frac{1}{2}, \quad \beta_1 = \beta_2 = \frac{1}{2}.$$

Note that for all $\gamma \geq 0$ we have $\bar{z} + \gamma p_i = (\pm \gamma h, 1 + \gamma) \in K_2$ and

$$\bar{z} - \beta_i p_i = (\pm \tfrac{1}{2} h, \tfrac{1}{2}) \notin \operatorname{int} K_2,$$

$$\bar{z} - \alpha_1 p_1 - \alpha_2 p_2 = (-\tfrac{1}{2} h + \tfrac{1}{2} h, 1 - \tfrac{1}{2} - \tfrac{1}{2}) = 0 \in K_2.$$

Therefore, the conditions of Theorem 5.4.1 are satisfied and

$$\nu \geq \tfrac{\alpha_1}{\beta_1} + \tfrac{\alpha_2}{\beta_2} = 2. \qquad \square$$

5.4.4 Semidefinite Optimization

In Semidefinite Optimization, the decision variables are matrices. Let

$$X = \{X^{(i,j)}\}_{i,j=1}^n$$

be a symmetric $n \times n$-matrix (notation: $X \in \mathbb{S}^n$). The real vector space $\mathbb{S}^n$ can be provided with the following inner product: for any $X, Y \in \mathbb{S}^n$ define

$$\langle X, Y \rangle_F = \sum_{i=1}^n \sum_{j=1}^n X^{(i,j)} Y^{(i,j)}, \quad \| X \|_F = \langle X, X \rangle_F^{1/2}.$$

Sometimes the value $\| X \|_F$ is called the *Frobenius norm* of the matrix X. For symmetric matrices X and Y, we have the following identity:

$$\langle X, Y \cdot Y \rangle_F = \sum_{i=1}^n \sum_{j=1}^n X^{(i,j)} \sum_{k=1}^n Y^{(i,k)} Y^{(j,k)} = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n X^{(i,j)} Y^{(i,k)} Y^{(j,k)}$$

$$= \sum_{k=1}^n \sum_{j=1}^n Y^{(k,j)} \sum_{i=1}^n X^{(j,i)} Y^{(i,k)} = \sum_{k=1}^n \sum_{j=1}^n Y^{(k,j)} (XY)^{(j,k)}$$

$$= \sum_{k=1}^n (YXY)^{(k,k)} = \operatorname{Trace}(YXY) = \langle YXY, I_n \rangle_F.$$

$$(5.4.10)$$

In Semidefinite Optimization, a nontrivial part of the constraints is formed by the *cone of positive semidefinite $n \times n$-matrices* $\mathbb{S}^N_+ \subset \mathbb{S}^n$. Recall that $X \in \mathbb{S}^n_+$ if and only if $\langle Xu, u \rangle \geq 0$ for any $u \in \mathbb{R}^n$. If $\langle Xu, u \rangle > 0$ for all nonzero u, we call X *positive definite*. Such matrices form the interior of cone $\mathbb{S}^n_+$. Note that $\mathbb{S}^n_+$ is a closed convex set.

The general formulation of the Semidefinite Optimization problem is as follows:

$$\min_{X \in \mathbb{S}^n_+} \{\langle C, X \rangle_F : \langle A_i, X \rangle_F = b_i, \ i = 1 \ldots m\}, \qquad (5.4.11)$$

where C and all A_i belong to $\mathbb{S}^n$. In order to apply a path-following scheme to this problem, we need a self-concordant barrier for the cone $\mathbb{S}^n_+$.

Let the matrix X belong to int $\mathbb{S}^n_+$. Define $F(X) = -\ln \det X$. Clearly

$$F(X) = -\sum_{i=1}^{n} \ln \lambda_i(X),$$

where $\{\lambda_i(X)\}_{i=1}^n$ is the set of eigenvalues of matrix X.

Lemma 5.4.6 *Function F is convex and $\nabla F(X) = -X^{-1}$. Moreover, for any direction $\Delta \in \mathbb{S}^n$, we have*

$$\langle \nabla^2 F(X)\Delta, \Delta \rangle_F = \| X^{-1/2} \Delta X^{-1/2} \|_F^2 = \langle X^{-1}\Delta X^{-1}, \Delta \rangle_F$$

$$= Trace\left([X^{-1/2}\Delta X^{-1/2}]^2\right),$$

$$D^3 F(x)[\Delta, \Delta, \Delta] = -2\langle I_n, [X^{-1/2}\Delta X^{-1/2}]^3 \rangle_F$$

$$= -2Trace\left([X^{-1/2}\Delta X^{-1/2}]^3\right).$$

Proof Let us fix some $\Delta \in \mathbb{S}^n$ and $X \in$ int $\mathbb{S}^n_+$ such that $X + \Delta \in \mathbb{S}^n_+$. Then

$$F(X + \Delta) - F(X) = -\ln \det(X + \Delta) - \ln \det X$$

$$= -\ln \det(I_n + X^{-1/2}\Delta X^{-1/2})$$

$$\geq -\ln \left[\tfrac{1}{n}Trace\,(I_n + X^{-1/2}\Delta X^{-1/2})\right]^n$$

$$= -n \ln \left[1 + \tfrac{1}{n}\langle I_n, X^{-1/2}\Delta X^{-1/2}\rangle_F\right]$$

$$\geq -\langle I_n, X^{-1/2}\Delta X^{-1/2}\rangle_F = -\langle X^{-1}, \Delta \rangle_F.$$

Thus, $-X^{-1} \in \partial F(X)$. Therefore, F is convex (Lemma 3.1.6) and $\nabla F(x) = -X^{-1}$ (Lemma 3.1.7).

Further, consider the function $\phi(\alpha) \equiv \langle \nabla F(X + \alpha \Delta), \Delta \rangle_F$, $\alpha \in [0, 1]$. Then

$$\phi(\alpha) - \phi(0) = \langle X^{-1} - (X + \alpha \Delta)^{-1}, \Delta \rangle_F$$

$$= \langle (X + \alpha \Delta)^{-1}[(X + \alpha \Delta) - X]X^{-1}, \Delta \rangle_F$$

$$= \alpha \langle (X + \alpha \Delta)^{-1} \Delta X^{-1}, \Delta \rangle_F.$$

Thus, $\phi'(0) = \langle \nabla^2 F(X)\Delta, \Delta \rangle_F = \langle X^{-1}\Delta X^{-1}, \Delta \rangle_F$.

The last expression can be proved in a similar way by differentiating the function $\psi(\alpha) = \langle (X + \alpha \Delta)^{-1} \Delta (X + \alpha \Delta)^{-1}, \Delta \rangle_F$. $\square$

Theorem 5.4.3 *The function F is an n-self-concordant barrier for $\mathbb{S}_+^n$.*

Proof Let us fix $X \in \text{int}\,\mathbb{S}_+^n$ and $\Delta \in \mathbb{S}^n$. Define $Q = X^{-1/2}\Delta X^{-1/2}$ and $\lambda_i = \lambda_i(Q)$, $i = 1 \ldots n$. Then, in view of Lemma 5.4.6, we have

$$\langle \nabla F(X), \Delta \rangle_F = \sum_{i=1}^n \lambda_i,$$

$$\langle \nabla^2 F(X)\Delta, \Delta \rangle_F = \sum_{i=1}^n \lambda_i^2,$$

$$D^3 F(X)[\Delta, \Delta, \Delta] = -2 \sum_{i=1}^n \lambda_i^3.$$

Using the two standard inequalities

$$\left(\sum_{i=1}^n \lambda_i \right)^2 \leq n \sum_{i=1}^n \lambda_i^2, \quad \left| \sum_{i=1}^n \lambda_i^3 \right| \leq \left(\sum_{i=1}^n \lambda_i^2 \right)^{3/2},$$

we obtain

$$\langle \nabla F(X), \Delta \rangle_F^2 \leq n \langle \nabla^2 F(X)\Delta, \Delta \rangle_F,$$

$$|D^3 F(X)[\Delta, \Delta, \Delta]| \leq 2 \langle \nabla^2 F(X)\Delta, \Delta \rangle_F^{3/2}. \quad \square$$

Let us prove that $F(X) = -\ln \det X$ is the optimal barrier for $\mathbb{S}_+^n$.

Lemma 5.4.7 *The parameter ν of any self-concordant barrier for the cone $\mathbb{S}_+^n$ satisfies the inequality $\nu \geq n$.*

Proof Let us choose $\bar{X} = I_n \in \text{int}\,\mathbb{S}_+^n$ and directions $P_i = e_i e_i^T$, $i = 1 \ldots n$, where e_i is the ith coordinate vector of $\mathbb{R}^n$. Note that for any $\gamma \geq 0$ we have

$I_n + \gamma P_i \in \text{int}\, \mathbb{S}_+^n$. Moreover,

$$I_n - e_i e_i^T \notin \text{int}\, \mathbb{S}_+^n, \quad I_n - \sum_{i=1}^{n} e_i e_i^T = 0 \in \mathbb{S}_+^n.$$

Therefore conditions of Theorem 5.4.1 are satisfied with $\alpha_i = \beta_i = 1,\ i = 1 \ldots n$, and we obtain $\nu \geq \sum_{i=1}^{n} \frac{\alpha_i}{\beta_i} = n.$ $\square$

As in Linear Optimization problem (5.4.2), in problem (5.4.11) we need to use the restriction of F onto the affine subspace

$$\mathscr{L} = \{X :\ \langle A_i, X \rangle_F = b_i,\ i = 1 \ldots m\}.$$

This restriction is an n-self-concordant barrier in view of Theorem 5.3.3. Thus, the complexity bound of the problem (5.4.11) is $O\left(\sqrt{n} \cdot \ln \frac{n}{\epsilon}\right)$ iterations of a path-following scheme. Note that this estimate is very encouraging since the dimension of the problem (5.4.11) is $\frac{1}{2} n(n+1)$.

Let us estimate the arithmetical cost of each iteration of a path-following scheme (5.3.29) as applied to the problem (5.4.11). Note that we work with a restriction of the barrier F to the set $\mathscr{L}$. In view of Lemma 5.4.6, each Newton step consists in solving the following problem:

$$\min_{\Delta} \{ \langle U, \Delta \rangle_F + \tfrac{1}{2} \langle X^{-1} \Delta X^{-1}, \Delta \rangle_F :\ \langle A_i, \Delta \rangle_F = 0,\ i = 1 \ldots m \},$$

where $X \succ 0$ belongs to $\mathscr{L}$ and U is a combination of the cost matrix C and the gradient $\nabla F(X)$. In accordance with the statement (3.1.59), the solution of this problem can be found from the following system of linear equations:

$$U + X^{-1} \Delta X^{-1} = \sum_{j=1}^{m} \lambda^{(j)} A_j,$$

$$\langle A_i, \Delta \rangle_F = 0, \quad i = 1 \ldots m.$$

(5.4.12)

From the first equation in (5.4.12) we get

$$\Delta = X \left[-U + \sum_{j=1}^{m} \lambda^{(j)} A_j \right] X.$$

(5.4.13)

Substituting this expression into the second equation in (5.4.12), we get the linear system

$$\sum_{j=1}^{m} \lambda^{(j)} \langle A_i, X A_j X \rangle_F = \langle A_i, X U X \rangle_F, \quad i = 1 \ldots m,$$

(5.4.14)

which can be written in matrix form as $S\lambda = d$ with

$$S^{(i,j)} = \langle A_i, XA_jX \rangle_F, \quad d^{(j)} = \langle U, XA_jX \rangle_F, \quad i, j = 1 \dots n.$$

Thus, a straightforward strategy of solving system (5.4.12) consists in the following steps.

- Compute the matrices XA_jX, $j = 1 \dots m$. Cost: $O(mn^3)$ operations.
- Compute the elements of S and d. Cost: $O(m^2n^2)$ operations.
- Compute $\lambda = S^{-1}d$. Cost: $O(m^3)$ operations.
- Compute Δ by (5.4.13). Cost: $O(mn^2)$ operations.

Taking into account that $m \leq \frac{n(n+1)}{2}$ we conclude that the complexity of one Newton step does not exceed

$$\boxed{O(n^2(m+n)m) \text{ arithmetic operations.}} \qquad (5.4.15)$$

However, if the matrices A_j possess a certain structure, then this estimate can be significantly improved. For example, if all A_j are of rank 1:

$$A_j = a_ja_j^T, \quad a_j \in \mathbb{R}^n, \quad j = 1 \dots m,$$

then the computation of the Newton step can be done in

$$\boxed{O((m+n)^3) \text{ arithmetic operations.}} \qquad (5.4.16)$$

We leave the justification of this claim as an exercise for the reader.

To conclude this section, note that in many important applications we can use the barrier $-\ln\det(\cdot)$ to treat some functions of eigenvalues. Consider, for example, a matrix $\mathscr{A}(x) \in \mathbb{S}^n$ which depends linearly on x. Then the convex region

$$\{(x,t) \mid \max_{1 \leq i \leq n} \lambda_i(\mathscr{A}(x)) \leq t\}$$

can be described by a self-concordant barrier

$$F(x,t) = -\ln\det(tI_n - \mathscr{A}(x)).$$

The value of the parameter of this barrier is equal to n.

5.4.5   *Extremal Ellipsoids*

In some applications, we are interested in approximating different sets by ellipsoids. Let us consider the most important examples.

5.4.5.1   Circumscribed Ellipsoid

> Given a set of points $a_1, \ldots, a_m \in \mathbb{R}^n$, find an ellipsoid W with the minimal volume which contains all points $\{a_i\}$.

Let us pose this problem in a formal way. First of all, note that any bounded ellipsoid $W \subset \mathbb{R}^n$ can be represented as

$$W = \{x \in \mathbb{R}^n \mid x = H^{-1}(v + u), \ \| u \| \leq 1\},$$

where $H \in \text{int}\, \mathbb{S}^n_+$, $v \in \mathbb{R}^n$, and the norm is standard Euclidean. Then the inclusion $a \in W$ is equivalent to the inequality $\| Ha - v \| \leq 1$. Note also that

$$\text{vol}_n W = \text{vol}_n B_2(0, 1) \cdot \det H^{-1} \;=\; \frac{\text{vol}_n B_2(0,1)}{\det H}.$$

Thus, our problem is as follows:

$$\min_{\substack{H \in \mathbb{S}^n_+, \\ v \in \mathbb{R}^n, \tau \in \mathbb{R}}} \ \{\tau : \ -\ln \det H \leq \tau, \ \| Ha_i - v \| \leq 1, \ i = 1 \ldots m\}. \tag{5.4.17}$$

In order to solve this problem by an interior-point scheme, we need to find a self-concordant barrier for the feasible set. In view of Theorems 5.4.3 and 5.3.5, we know self-concordant barriers for all components. Indeed, we can use the following barrier:

$$F(H, v, \tau) = -\ln \det H - \ln(\tau + \ln \det H) - \sum_{i=1}^m \ln(1 - \| Ha_i - v \|^2),$$

$$\nu = m + n + 1.$$

The corresponding complexity bound is $O\left(\sqrt{m + n + 1} \cdot \ln \frac{m+n}{\epsilon}\right)$ iterations of a path-following scheme.

5.4.5.2 Inscribed Ellipsoid with Fixed Center

Let Q be a convex polytope defined by a set of linear inequalities:

$$Q = \{x \in \mathbb{R}^n \mid \langle a_i, x \rangle \leq b_i, \ i = 1 \ldots m\},$$

and let $v \in \operatorname{int} Q$. Find an ellipsoid $W \subset Q$ with the biggest volume which is centered at v.

Let us fix some $H \in \operatorname{int} \mathbb{S}^n_+$. We can represent the ellipsoid W as

$$W = \{x \in \mathbb{R}^n \mid \langle H^{-1}(x - v), x - v \rangle \leq 1\}.$$

We need the following simple result.

Lemma 5.4.8 *Let $\langle a, v \rangle < b$. The inequality $\langle a, x \rangle \leq b$ is valid for all $x \in W$ if and only if*

$$\langle Ha, a \rangle \leq (b - \langle a, v \rangle)^2.$$

Proof In view of Lemma 3.1.20, we have

$$\max_u \{\langle a, u \rangle \mid \langle H^{-1}u, u \rangle \leq 1\} = \langle Ha, a \rangle^{1/2}.$$

Therefore, we need to ensure

$$\max_{x \in W} \langle a, x \rangle = \max_{x \in W} [\langle a, x - v \rangle + \langle a, v \rangle]$$

$$= \langle a, v \rangle + \max_x \{\langle a, u \rangle \mid \langle H^{-1}u, u \rangle \leq 1\}$$

$$= \langle a, v \rangle + \langle Ha, a \rangle^{1/2} \leq b.$$

This proves our statement since $\langle a, v \rangle < b$. $\square$

Note that $\operatorname{vol}_n W = \operatorname{vol}_n B_2(0, 1)[\det H]^{1/2}$. Hence, our problem is as follows:

$$\min_{H \in \mathbb{S}^n_+, \tau \in \mathbb{R}} \{\tau : \ -\ln \det H \leq \tau, \ \langle Ha_i, a_i \rangle \leq (b_i - \langle a_i, v \rangle)^2, \ i = 1 \ldots m\}.$$

$$(5.4.18)$$

In view of Theorems 5.4.3 and 5.3.5, we can use the following self-concordant barrier:

$$F(H, \tau) = -\ln \det H - \ln(\tau + \ln \det H) - \sum_{i=1}^{m} \ln[(b_i - \langle a_i, v \rangle)^2 - \langle H a_i, a_i \rangle],$$

with barrier parameter $\nu = m + n + 1$. The complexity bound of the corresponding path-following scheme is

$$O\left(\sqrt{m + n + 1} \cdot \ln \tfrac{m+n}{\epsilon}\right)$$

iterations.

5.4.5.3 Inscribed Ellipsoid with Free Center

Let Q be a convex polytope defined by a set of linear inequalities:

$$Q = \{x \in \mathbb{R}^n \mid \langle a_i, x \rangle \le b_i, \ i = 1 \ldots m\},$$

and let int $Q \neq \emptyset$. Find an ellipsoid W with the biggest volume which is contained in Q.

Let $G \in \text{int } \mathbb{S}_+^n$ and $v \in \text{int } Q$. We can represent W as follows:

$$W = \{x \in \mathbb{R}^n \mid \| G^{-1}(x - v) \| \le 1\}$$

$$\equiv \{x \in \mathbb{R}^n \mid \langle G^{-2}(x - v), x - v \rangle \le 1\}.$$

In view of Lemma 5.4.8, inequality $\langle a, x \rangle \le b$ is valid for any $x \in W$ if and only if

$$\| Ga \|^2 \equiv \langle G^2 a, a \rangle \le (b - \langle a, v \rangle)^2.$$

This gives us a convex feasible set for parameters (G, v):

$$\| Ga \| \le b - \langle a, v \rangle.$$

Note that $\text{vol}_n W = \text{vol}_n B_2(0, 1) \det G$. Therefore, our problem can be written as follows:

$$\min_{\substack{G \in \mathbb{S}_+^n, \\ v \in \mathbb{R}^n, \tau \in \mathbb{R}}} \{\tau : -\ln \det G \le \tau, \ \| Ga_i \| \le b_i - \langle a_i, v \rangle, \ i = 1 \ldots m\}. \tag{5.4.19}$$

In view of Theorems 5.4.3, 5.3.5 and Lemma 5.4.4, we can use the following self-concordant barrier:

$$F(G, v, \tau) = -\ln \det G - \ln(\tau + \ln \det G) - \sum_{i=1}^{m} \ln[(b_i - \langle a_i, v \rangle)^2 - \| G a_i \|^2]$$

with barrier parameter $v = 2m + n + 1$. The corresponding efficiency estimate is $O\left(\sqrt{2m + n + 1} \cdot \ln \frac{m+n}{\epsilon}\right)$ iterations of a path-following scheme.

5.4.6 Constructing Self-concordant Barriers for Convex Sets

In this section we develop a general framework for constructing self-concordant barriers for convex cones. First of all, let us define the objects we are working with. They are related to three different real vector spaces, $\mathbb{E}_1$, $\mathbb{E}_2$, and $\mathbb{E}_3$.

Consider a function $\xi(\cdot) : \mathbb{E}_1 \to \mathbb{E}_2$ defined on a closed convex set $Q_1 \subset \mathbb{E}_1$. Assume that ξ is three times continuously differentiable and *concave* with respect to a closed convex cone $K \subset E_2$:

$$-D^2\xi(x)[h, h] \in K \quad \forall x \in \text{int } Q_1, \ h \in \mathbb{E}_1. \tag{5.4.20}$$

It is convenient to write this inclusion as $D^2\xi(x)[h, h] \preceq_K 0$.

Definition 5.4.2 Let $F(\cdot)$ be a v-self-concordant barrier for Q_1 and $\beta \geq 1$. We say that a function ξ is β-compatible with F if for all $x \in \text{int } Q_1$ and $h \in \mathbb{E}_1$ we have

$$D^3\xi(x)[h, h, h] \preceq_K -3\beta \cdot D^2\xi(x)[h, h] \cdot \langle \nabla^2 F(x)h, h \rangle^{1/2}. \tag{5.4.21}$$

Alternating the sign of direction h in (5.4.21), we get the following equivalent condition:

$$-D^3\xi(x)[h, h, h] \preceq_K -3\beta \cdot D^2\xi(x)[h, h] \cdot \langle \nabla^2 F(x)h, h \rangle^{1/2}. \tag{5.4.22}$$

Note that the set of β-compatible functions is a convex cone: if functions ξ_1 and ξ_2 are β-compatible with barrier F, then the sum $\alpha_1\xi_1 + \alpha_2\xi_2$, with arbitrary $\alpha_1, \alpha_2 > 0$, is also β-compatible with F.

Let us construct a self-concordant barrier for a *composition* of the set

$$\mathscr{S}_1 = \{(x, y) \in Q_1 \times \mathbb{E}_2 : \xi(x) \succeq_K y\}$$

and a convex set $Q_2 \subset \mathbb{E}_2 \times \mathbb{E}_3$. That is

$$\mathscr{Q} = \{(x, z) \in Q_1 \times \mathbb{E}_3 : \exists y, \ \xi(x) \succeq_K y, \ (y, z) \in Q_2\}.$$

The necessity of such a structure is clear from the following example.

Example 5.4.1 Let us fix some $\alpha \in (0, 1)$. Consider the following *power cone*

$$K_\alpha = \left\{ (x^{(1)}, x^{(2)}, z) \in \mathbb{R}^2_+ \times \mathbb{R} : (x^{(1)})^\alpha \cdot (x^{(2)})^{1-\alpha} \geq |z| \right\}.$$

For our representation, we need the following objects:

$$\mathbb{E}_1 = \mathbb{R}^2, \quad Q_1 = \mathbb{R}^2_+, \quad F(x) = -\ln x^{(1)} - \ln x^{(2)}, \quad \nu = 2,$$

$$\mathbb{E}_2 = \mathbb{R}, \quad \xi(x) = (x^{(1)})^\alpha \cdot (x^{(2)})^{1-\alpha}, \quad K = \mathbb{R}_+ \subset \mathbb{E}_2,$$

$$\mathbb{E}_3 = \mathbb{R}, \quad Q_2 = \{(y, z) \in \mathbb{E}_2 \times \mathbb{E}_3 : y \geq |z|\}. \qquad \square$$

In our construction, we also need a μ-self-concordant barrier $\Phi(y, z)$ for the set Q_2. We assume that all directions from the cone $K_0 \stackrel{\text{def}}{=} K \times \{0\} \subset \mathbb{E}_2 \times \mathbb{E}_3$ are recession directions of the set Q_2. Consequently, for any $s \in K$ and $(y, z) \in \text{int } Q_2$ we have

$$\langle \nabla_y \Phi(y, z), s \rangle = \langle \nabla \Phi(y, z), (s, 0) \rangle \stackrel{(5.3.13)}{\leq} 0. \qquad (5.4.23)$$

Consider the barrier

$$\Psi(x, z) = \Phi(\xi(x), z) + \beta^3 F(x).$$

Let us fix a point $(x, z) \in \text{int } \mathscr{Q}$ and choose an arbitrary direction $d = (h, v) \in \mathbb{E}_1 \times \mathbb{E}_3$. Define

$$\xi' = D\xi(x)[h], \quad \xi'' = D^2\xi(x)[h, h], \quad \xi''' = D^3\xi(x)[h, h, h], \quad l = (\xi', v).$$

Let $\psi(x, z) = \Phi(\xi(x), z)$. Consider the following directional derivatives:

$$\Delta_1 \stackrel{\text{def}}{=} D\psi(x, z)[d] = \langle \nabla_y \Phi(\xi(x), z), \xi' \rangle + \langle \nabla_z \Phi(\xi(x), z), v \rangle = \langle \nabla \Phi(\xi(x), z), l \rangle.$$

Note that $l \equiv l(x)$. Therefore $l' \stackrel{\text{def}}{=} Dl(x)[d] = (\xi'', 0) \stackrel{(5.4.20)}{\in} -K_0$. Thus, we can continue:

$$\Delta_2 \stackrel{\text{def}}{=} D^2\psi(x, z)[d, d] = \langle \nabla^2\Phi(\xi(x), z)l, l \rangle + \langle \nabla\Phi(\xi(x), z), l' \rangle$$

$$(5.4.24)$$

$$= \langle \nabla^2\Phi(\xi(x), z)l, l \rangle + \langle \nabla_y \Phi(\xi(x), z), \xi'' \rangle \stackrel{\text{def}}{=} \sigma_1 + \sigma_2.$$

Since $-l'$ is a recession direction of Q_2, by (5.3.13) we have $\sigma_2 \geq 0$. Finally,

$$\Delta_3 \stackrel{\text{def}}{=} D^3 \psi(x, z)[d, d, d]$$

$$= D^3 \Phi(\xi(x), z)[l, l, l] + 3\langle \nabla^2 \Phi(\xi(x), z)l, l' \rangle + \langle \nabla_y \Phi(\xi(x), z), \xi''' \rangle. \tag{5.4.25}$$

Again, since $-l'$ is a recession direction of Q_2,

$$\langle \nabla^2 \Phi(\xi(x), z)l, l' \rangle \quad \leq \quad \langle \nabla^2 \Phi(\xi(x), z)l, l \rangle^{1/2} \cdot \langle \nabla^2 \Phi(\xi(x), z)l', l' \rangle^{1/2}$$

$$\stackrel{(5.3.13)}{\leq} \langle \nabla^2 \Phi(\xi(x), z)l, l \rangle^{1/2} \cdot \langle -\nabla \Phi(\xi(x), z), -l' \rangle = \sigma_1^{1/2} \sigma_2.$$

Further, let $\sigma_3 = \langle \nabla^2 F(x)h, h \rangle$. Since ξ is β-compatible with F (see (5.4.22)), we have

$$\langle -\nabla_y \Phi(\xi(x), z), -\xi''' \rangle \stackrel{(5.4.23)}{\leq} 3\beta \langle -\nabla_y \Phi(\xi(x), z), -\xi'' \rangle \cdot \sigma_3^{1/2} = 3\beta \cdot \sigma_2 \cdot \sigma_3^{1/2}.$$

Thus, substituting these inequalities into (5.4.25) and using (5.1.4), we obtain

$$\Delta_3 \leq 2\sigma_1^{3/2} + 3\sigma_1^{1/2}\sigma_2 + 3\beta \cdot \sigma_2 \cdot \sigma_3^{1/2}.$$

Consider now D_k, $k = 1 \ldots 3$, the directional derivatives of the function Ψ. Note that

$$D_2 = \Delta_2 + \beta^3 \sigma_3 = \sigma_1 + \sigma_2 + \beta^3 \sigma_3 \geq \sigma_1 + \sigma_2 + \beta^2 \sigma_3. \tag{5.4.26}$$

Therefore,

$$D_3 = \Delta_3 + \beta^3 D^3 F(x)[h, h, h] \stackrel{(5.1.4)}{\leq} \Delta_3 + 2\beta^3 \sigma_3^{3/2}$$

$$\leq 2\sigma_1^{3/2} + 3\sigma_1^{1/2}\sigma_2 + 3\beta \cdot \sigma_2 \cdot \sigma_3^{1/2} + 2\beta^3 \sigma_3^{3/2}$$

$$= (\sigma_1^{1/2} + \beta\sigma_3^{1/2})(2\sigma_1 - 2\beta\sigma_1^{1/2}\sigma_3^{1/2} + 2\beta^2\sigma_3 + 3\sigma_2)$$

$$\stackrel{(5.4.26)}{\leq} (\sigma_1^{1/2} + \beta\sigma_3^{1/2})(3D_2 - (\sigma_1^{1/2} + \beta\sigma_3^{1/2})^2) \leq 2D_2^{3/2}.$$

Thus, we come to the following statement.

Theorem 5.4.4 *Let the function $\xi(\cdot) : E_1 \to E_2$ satisfy the following conditions.*

- *It is concave with respect to a convex cone $K \subset E_2$.*
- *It is β-compatible with self-concordant barrier $F(\cdot)$ for a set $Q \subseteq \text{dom}\,\xi$.*

Assume in addition that $\Phi(\cdot, \cdot)$ is a μ-self-concordant barrier for a closed convex set $Q_2 \subset E_2 \times E_3$, and the cone $K \times \{0\} \subset E_2 \times E_3$ contains only the recession directions of the set Q_2. Then the function

$$\Psi(x, z) = \Phi(\xi(x), z) + \beta^3 F(x) \tag{5.4.27}$$

is a self-concordant barrier for the set $\{(x, z) \in Q \times \mathbb{E}_3 : \exists y, \ \xi(x) \succeq_K y, \ (y, z) \in Q_2\}$ with barrier parameter $\hat{v} = \mu + \beta^3 v$.

Proof We need to justify only the value of the barrier parameter $\hat{v}$. Indeed,

$$D_1 = \langle \nabla \Phi(\xi(x), z), l \rangle + \beta^3 \langle \nabla F(x), h \rangle \ \leq \ \sqrt{v} \cdot \sigma_1^{1/2} + \beta^3 \sqrt{\mu} \cdot \sigma_3^{1/2}$$

$$\leq \ \max_{\sigma_1, \sigma_3 \geq 0} \{ \sqrt{v} \cdot \sigma_1^{1/2} + \beta^3 \cdot \sqrt{\mu}\sigma_3^{1/2} : \sigma_1 + \beta^3 \sigma_3 \stackrel{(5.4.26)}{\leq} D_2 \}$$

$$= \ \sqrt{\hat{v}} \cdot D_2^{1/2}.$$

It remains to use definition (5.3.6). $\square$

Note that in construction (5.4.27) the function ξ must be compatible only with the barrier F. The function Φ can be an arbitrary self-concordant barrier for the set Q_2.

5.4.7 Examples of Self-concordant Barriers

Despite its complicated formulation, Theorem 5.4.4 is very convenient for constructing a good self-concordant barrier for convex cones. Let us confirm this claim with several examples.

1. The power cone and epigraph of the p-norm. Let us fix some $\alpha \in (0, 1)$. To the description of the representation of the power cone

$$K_\alpha = \left\{ (x^{(1)}, x^{(2)}, z) \in R_+^2 \times R : \ (x^{(1)})^\alpha \cdot (x^{(2)})^{1-\alpha} \geq |z| \right\},$$

given in Example 5.4.1, we need to add only a definition of the barrier function for the set Q_2. In view of Lemma 5.4.4, we can take

$$\Phi(y, z) = -\ln(y^2 - z^2),$$

with barrier parameter $\mu = 2$. Thus, all conditions of Theorem 5.4.4 are clearly satisfied except β-compatibility.

Let us prove that the function $\xi(x) = (x^{(1)})^\alpha \cdot (x^{(2)})^{1-\alpha}$ is β-comptible with barrier $F(x) = -\ln x^{(1)} - \ln x^{(2)}$. Let us choose a direction $h \in \mathbb{R}^2$ and $x \in \text{int } \mathbb{R}_+^2$.

Define

$$\delta_1 = \frac{h^{(1)}}{x^{(1)}}, \quad \delta_2 = \frac{h^{(2)}}{x^{(2)}}, \quad \sigma = \delta_1^2 + \delta_2^2.$$

Let us compute the directional derivatives:

$$D\xi(x)[h] = \left[\frac{\alpha h^{(1)}}{x^{(1)}} + \frac{(1-\alpha)h^{(2)}}{x^{(2)}} \right] \cdot \xi(x) = [\alpha \delta_1 + (1-\alpha)\delta_2] \cdot \xi(x),$$

$$D^2\xi(x)[h, h] = -[\alpha \delta_1^2 + (1-\alpha)\delta_2^2] \cdot \xi(x) + [\alpha \delta_1 + (1-\alpha)\delta_2] \cdot D\xi(x)[h]$$

$$= -\alpha(1-\alpha)(\delta_1 - \delta_2)^2 \cdot \xi(x),$$

$$D^3\xi(x)[h, h, h] = 2\alpha(1-\alpha)(\delta_1 - \delta_2) \cdot (\delta_1^2 - \delta_2^2) \cdot \xi(x)$$

$$-\alpha(1-\alpha)(\delta_1 - \delta_2)^2 \cdot D\xi(x)[h]$$

$$= \xi(x) \cdot \alpha(1-\alpha)(\delta_1 - \delta_2)^2 \cdot [2\delta_1 + 2\delta_2 - \alpha \delta_1 - (1-\alpha)\delta_2]$$

$$= -D^2\xi(x)[h, h] \cdot [(2-\alpha)\delta_1 + (1+\alpha)\delta_2].$$

Since $(2-\alpha)\delta_1 + (1+\alpha)\delta_2 \le [(2-\alpha)^2 + (1+\alpha)^2]^{1/2}\sigma^{1/2} < 3\sigma^{1/2}$, we conclude that ξ is 1-compatible with F. Therefore, in view of Theorem 5.4.4, function

$$\Psi_P(x, z) = -\ln\left((x^{(1)})^{2\alpha} \cdot (x^{(2)})^{2(1-\alpha)} - z^2 \right) - \ln x^{(1)} - \ln x^{(2)} \qquad (5.4.28)$$

is a 4-self-concordant barrier for cone K_α.

A similar structure can be used to construct a self-concordant barrier for the cone

$$K_\alpha^+ = \left\{ (x^{(1)}, x^{(2)}, z) \in \mathbb{R}_+^2 \times \mathbb{R} : (x^{(1)})^\alpha \cdot (x^{(2)})^{1-\alpha} \ge z \right\}.$$

In this case, we can choose $\Phi(y, z) = \ln(y - z)$ with parameter $\mu = 1$. Thus, by Theorem 5.4.4, we get the following 3-self-concordant barrier:

$$\Psi_P^+(x, z) = -\ln\left((x^{(1)})^\alpha \cdot (x^{(2)})^{(1-\alpha)} - z \right) - \ln x^{(1)} - \ln x^{(2)}. \qquad (5.4.29)$$

Let us show that this barrier has the best possible value of parameter.

Lemma 5.4.9 *Any ν-self-concordant barrier for the cone K_α^+ has $\nu \ge 3$.*

Proof Note that the cone K_α^+ has three recession directions:

$$p_1 = (1, 0, 0)^T, \quad p_2 = (0, 1, 0)^T, \quad p_3 = (0, 0, -1)^T.$$

Let us choose a parameter $\tau > 0$ and define $\bar{x} = (1, 1, -\tau)^T$. Note that

$$\bar{x} - p_1 \notin \text{int } K_\alpha^+, \quad \bar{x} - p_2 \notin \text{int } K_\alpha^+, \quad \bar{x} - (1+\tau)p_3 \in \partial K_\alpha^+.$$

On the other hand, $\bar{x} - p_1 - p_2 - \tau p_3 = 0 \in K_\alpha^+$. Thus, to apply Theorem 5.4.1, we can choose

$$\alpha_1 = \alpha_2 = 1, \quad \alpha_3 = \tau, \quad \beta_1 = \beta_2 = 1, \quad \beta_3 = 1 + \tau.$$

Hence, $\nu \geq \sum\limits_{i=1}^{3} \frac{\alpha_i}{\beta_i} = 2 + \frac{\tau}{1+\tau}$. It remains to compute the limit as $\tau \to +\infty$. $\square$

Note that the barrier $\Psi_P(x, z)$ can be used to construct $4n$-self-concordant barrier for the epigraph of an ℓ_p-norm in $\mathbb{R}^n$:

$$\mathcal{K}_p = \left\{ (\tau, z) \in \mathbb{R} \times \mathbb{R}^n : \tau \geq \|z\|_{(p)} \right\}, \quad 1 \leq p \leq \infty,$$

where $\|z\|_{(p)} = \left[\sum\limits_{i=1}^{n} |z^{(i)}|^p \right]^{1/p}$. Let us assume that $\alpha \overset{\text{def}}{=} \frac{1}{p} \in (0, 1)$. Then, it is easy to prove that the point (τ, z) belongs to $\mathcal{K}_p$ if and only if there exists an $x \in \mathbb{R}^n_+$ satisfying the conditions

$$(x^{(i)})^\alpha \cdot \tau^{1-\alpha} \geq |z^{(i)}|, \quad i = 1, \ldots, n,$$

$$\sum\limits_{i=1}^{n} x^{(i)} = \tau. \tag{5.4.30}$$

Thus, a self-concordant barrier for the cone $\mathcal{K}_p$ can be implemented by restricting the $(4n)$-self-concordant barrier

$$\Psi_\alpha(\tau, x, z) = -\sum\limits_{i=1}^{n} \left[\ln \left((x^{(i)})^{2\alpha} \cdot \tau^{2(1-\alpha)} - (z^{(i)})^2 \right) + \ln x^{(i)} + \ln \tau \right] \tag{5.4.31}$$

onto the hyperplane $\sum\limits_{i=1}^{n} x^{(i)} = \tau$.

2. The conic hull of the epigraph of the entropy function. We need to describe the conic hull of the following set:

$$\left\{ (x^{(1)}, z) : z \geq x^{(1)} \ln x^{(1)}, \ x^{(1)} > 0 \right\}.$$

Introducing a projective variable $x^{(2)} > 0$, we obtain the cone

$$\mathcal{Q} = \left\{ (x^{(1)}, x^{(2)}, z) : z \geq x^{(1)} \cdot [\ln x^{(1)} - \ln x^{(2)}], \ x^{(1)}, x^{(2)} > 0 \right\}. \tag{5.4.32}$$

Let us represent it in the format of Theorem 5.4.4:

$$\mathbb{E}_1 = \mathbb{R}^2, \quad Q_1 = R_+^2, \quad F(x) = -\ln x^{(1)} - \ln x^{(2)}, \quad \nu = 2,$$

$$\mathbb{E}_2 = \mathbb{R}, \quad \xi(x) = -x^{(1)} \cdot [\ln x^{(1)} - \ln x^{(2)}], \quad K = \mathbb{R}_+,$$

$$\mathbb{E}_3 = \mathbb{R}, \quad Q_2 = \{(y, z) : y + z \geq 0\}, \quad \Phi(y, z) = -\ln(y + z), \quad \mu = 1.$$

Let us show that ξ is 1-compatible with F. We use the notation of the previous example.

$$D\xi(x)[h] = \delta_1 \cdot \xi(x) - x^{(1)} \cdot [\delta_1 - \delta_2].$$

$$D^2\xi(x)[h, h] = -\delta_1^2 \cdot \xi(x) + \delta_1 \cdot D\xi(x)[h] - h^{(1)} \cdot [\delta_1 - \delta_2] + x^{(1)} \cdot [\delta_1^2 - \delta_2^2]$$

$$= x^{(1)} \cdot [-2\delta_1(\delta_1 - \delta_2) + \delta_1^2 - \delta_2^2] = -x^{(1)} \cdot (\delta_1 - \delta_2)^2.$$

$$D^3\xi(x)[h, h, h] = -h^{(1)} \cdot (\delta_1 - \delta_2)^2 + 2x^{(1)} \cdot (\delta_1 - \delta_2) \cdot (\delta_1^2 - \delta_2^2)$$

$$= x^{(1)}(\delta_1 - \delta_2)^2 \cdot [-\delta_1 + 2(\delta_1 + \delta_2)]$$

$$= -D^2\xi(x)[h, h] \cdot [\delta_1 + 2\delta_2].$$

Since $\delta_1 + 2\delta_2 \leq \sqrt{5} \cdot \sigma^{1/2} < 3\sigma^{1/2}$, we conclude that ξ is 1-compatible with F. Therefore, in view of Theorem 5.4.4 the function

$$\Psi_E(x, z) = -\ln\left(z - x^{(1)} \cdot \ln \frac{x^{(1)}}{x^{(2)}}\right) - \ln x^{(1)} - \ln x^{(2)} \tag{5.4.33}$$

is a 3-self-concordant barrier for the cone $\mathcal{Q}$. It is interesting that the same barrier can also describe the epigraph of logarithmic and exponent functions. Indeed,

$$\mathcal{Q} \bigcap \{x : x^{(1)} = 1\} = \{(x^{(2)}, z) : z \geq -\ln x^{(2)}\} = \{(x^{(2)}, z) : x^{(2)} \geq e^{-z}\}.$$

Let us show that we can use the 3-self-concordant barrier

$$\psi_E(x, y, \tau) = -\ln\left(\tau \ln \frac{y}{\tau} - x\right) - \ln y - \ln \tau,$$

$$(x, y, \tau) \in \text{int } \mathscr{E} \overset{\text{def}}{=} \{y \geq \tau e^{x/\tau}, \ \tau > 0\} \subset \mathbb{R}^3, \tag{5.4.34}$$

in more complicated situations. Consider the conic hull of the epigraph of the following function:

$$f_n(x) \stackrel{\text{def}}{=} \ln\left(\sum_{i=1}^n e^{x^{(i)}}\right), \quad x \in \mathbb{R}^n,$$

(5.4.35)

$$Q \stackrel{\text{def}}{=} \left\{(x, t, \tau) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} : t \geq \tau f_n\left(\tfrac{x}{\tau}\right), \ \tau > 0\right\}.$$

Clearly $(x, t, \tau) \in Q$ if and only if

$$f_n\left(\tfrac{1}{\tau}(x - t \cdot \bar{e}_n)\right) \leq 1,$$

where $\bar{e}_n \in \mathbb{R}^n$ is the vector of all ones. Therefore, we can model Q as a projection of the following cone:

$$\hat{Q} = \left\{(x, y, t, \tau) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} : y^{(i)} \geq \tau e^{(x^{(i)} - t)/\tau}, \ i = 1, \dots, n,\right.$$

$$\left. \sum_{i=1}^n y^{(i)} = \tau\right\}.$$

This cone admits a $3n$-self-concordant barrier, obtained as a restriction of the function

$$\Psi_L(x, y, t, \tau) = -\sum_{i=1}^n \left[\ln\left(t + \tau \ln y^{(i)} - x^{(i)} - \tau \ln \tau\right) + \ln y^{(i)} + \ln \tau\right],$$

(5.4.36)

onto the hyperplane $\sum_{i=1}^n y^{(i)} = \tau$.

3. The geometric mean. Let $x \in \mathbb{R}^n_+$ and $a \in \Delta_n \stackrel{\text{def}}{=} \left\{y \in \mathbb{R}^n_+ : \sum_{i=1}^n y^{(i)} = 1\right\}$. Without loss of generality, we can consider a with positive components. Define

$$\xi(x) = x^a \stackrel{\text{def}}{=} \prod_{i=1}^n (x^{(i)})^{a^{(i)}}.$$

Let us write down the directional derivatives of this function along some $h \in \mathbb{R}^n$. Define

$$\delta_x^{(i)}(h) = \tfrac{h^{(i)}}{x^{(i)}}, \ i = 1, \dots, n,$$

$$\delta_x(h) = \left(\delta_x^{(1)}(h), \dots, \delta_x^{(n)}(h)\right)^T,$$

$$F(x) = -\sum_{i=1}^n \ln x^{(i)}.$$

Clearly, $\|h\|_x \overset{\text{def}}{=} \langle F''(x)h, h\rangle^{1/2} = \|\delta_x(h)\|$, where the norm is standard Euclidean. Note that

$$D(\ln \xi(x))[h] = \tfrac{1}{\xi(x)} D\xi(x)[h] = \langle a, \delta_x(h)\rangle.$$

Thus, $D\xi(x)[h] = \xi(x) \cdot \langle a, \delta_x(h)\rangle$. Denoting by $[x]^k \in \mathbb{R}^n$ a component-wise power of a vector $x \in \mathbb{R}^n$, we obtain:

$$D^2\xi(x)[h, h] = \xi(x) \cdot \langle a, \delta_x(h)\rangle^2 - \xi(x) \cdot \langle a, [\delta_x(h)]^2\rangle$$

$$= -\xi(x) \cdot \langle a, [\delta_x(h) - \langle a, \delta_x(h)\rangle \cdot \bar{e}_n]^2\rangle \overset{\text{def}}{=} -\xi(x) \cdot S_2.$$

Further, defining $\xi = \xi(x)$ and $\delta = \delta_x(h)$, we obtain:

$$D^3\xi(x)[h, h, h] = \xi\langle a, \delta\rangle^3 + 2\xi\langle a, \delta\rangle\langle a, -[\delta]^2\rangle - \xi\langle a, \delta\rangle\langle a, [\delta]^2\rangle - \xi\langle a, -2[\delta]^3\rangle$$

$$= \xi\left(\langle a, \delta\rangle^3 - 3\langle a, \delta\rangle\langle a, [\delta]^2\rangle + 2\langle a, [\delta]^3\rangle\right).$$

Define

$$S_3 = \langle a, [\delta - \langle a, \delta\rangle\bar{e}_n]^3\rangle = \langle a, [\delta]^3 - 3\langle a, \delta\rangle[\delta]^2 + 3\langle a, \delta\rangle^2\delta - \langle a, \delta\rangle^3\bar{e}_n\rangle$$

$$= \langle a, [\delta]^3\rangle - 3\langle a, \delta\rangle\langle a, [\delta]^2\rangle + 2\langle a, \delta\rangle^3.$$

Then, in this new notation we have

$$D^3\xi(x)[h, h, h] = \xi\Big(\langle a, \delta\rangle^3 - 3\langle a, \delta\rangle\langle a, [\delta]^2\rangle$$

$$+ 2\left[S_3 + 3\langle a, \delta\rangle\langle a, [\delta]^2\rangle - 2\langle a, \delta\rangle^3\right]\Big)$$

$$= \xi\left(2S_3 + 3\langle a, \delta\rangle\langle a, [\delta]^2\rangle - 3\langle a, \delta\rangle^3\right) = \xi(2S_3 + 3\langle a, \delta\rangle S_2).$$

Therefore,

$$D^3\xi(x)[h, h, h] \le \xi S_2 \left(3\langle a, \delta\rangle + 2\max_{1\le i\le n} [\delta^{(i)} - \langle a, \delta\rangle]\right)$$

$$\le \xi S_2 \left(\langle a, \delta\rangle + 2\max_{1\le i\le n} |\delta^{(i)}|\right)$$

$$\le -3D^2\xi(x)[h, h] \cdot \langle F''(x)\delta, \delta\rangle^{1/2}.$$

Thus, we have proved that ξ is 1-compatible with F. This means that the function

$$\Psi(x,t) = -\ln(\xi(x) - t) + F(x), \quad x > 0 \in R^n, \tag{5.4.37}$$

is an $(n+1)$-self-concordant barrier for the hypograph of the function ξ. Moreover, since the set of β-compatible functions is a convex cone, any sum

$$\xi(x) = \sum_{k=1}^{m} \alpha_k x^{a_k}, \tag{5.4.38}$$

with $\alpha_k > 0$, and $a_k \in \Delta_n$, $k = 1, \ldots, m$, is 1-compatible with F. Hence, for such functions formula (5.4.37) is also applicable and the parameter of this barrier remains equal to $n+1$.

Note that the functions in the form (5.4.38) sometimes arise in optimization problems related to polynomials. Indeed, assume we need to solve the problem

$$\max_{y} \left\{ p(y) = \sum_{k=1}^{m} \alpha_k y^{b_k} : \ y \geq 0, \ \|y\|_{(d)} \leq 1 \right\},$$

where all b_k belong to $d \cdot \Delta_n$ and $\|y\|_{(d)} = \left[\sum_{i=1}^{n} (y^{(i)})^d \right]^{1/d}$. Then for new variables $y^{(i)} = \left[x^{(i)} \right]^{1/d}$, $i = 1, \ldots, n$, our problem becomes convex with a concave objective $\xi(\cdot)$ given by (5.4.38).

4. The hypograph of the exponent of the self-concordant barrier. Let $F(\cdot)$ be ν-self-concordant barrier for the set Dom F. Let us fix $p \geq \nu$ and consider the function $\xi_p(x) = \exp\left\{ -\frac{1}{p} F(x) \right\}$. As we have proved in Lemma 5.3.1, this function is concave on dom F. Consider the following set:

$$\mathscr{H}_p = \left\{ (x,t) \in \text{dom } F \times \mathbb{R} : \ \xi_p(x) \geq t \right\}.$$

Let us construct a self-concordant barrier for this set.

In our framework, $Q_1 = \text{Dom } F$, $Q_2 = \{(y,t) \in \mathbb{R}^2 : \ y \geq t\}$, $K = \mathbb{R}_+$, and $\Phi(y,t) = -\ln(y - t)$ with $\mu = 1$. Let us prove that $\xi_p(x)$ is concave with respect to K, and it is β-compatible with F.

Let us fix $x \in \text{dom } F$ and an arbitrary direction $h \in \mathbb{E}$. Then

$$\xi' \stackrel{\text{def}}{=} D\xi_p(x)[h] = -\frac{1}{p} \langle \nabla F(x), h \rangle \xi_p(x),$$

$$\xi'' \stackrel{\text{def}}{=} D^2 F(x)[h,h] = \frac{1}{p^2} \langle \nabla F(x), h \rangle^2 \xi_p(x) - \frac{1}{p} \langle \nabla^2 F(x)h, h \rangle \xi_p(x),$$

$$\xi''' \stackrel{\text{def}}{=} D^3 F(x)[h,h,h] = -\frac{1}{p^3} \langle \nabla F(x), h \rangle^3 \xi_p(x)$$

$$+\frac{3}{p^2} \langle \nabla F(x), h \rangle \cdot \langle \nabla^2 F(x)h, h \rangle \xi_p(x) - \frac{1}{p} D^3 F(x)[h,h,h] \xi_p(x).$$

As we have already seen, in view of (5.3.6), we have $\xi'' \le 0$. This means that it is concave with respect to K.

Let $\xi = \xi_p(x)$, $D_1 = \langle \nabla F(x), h \rangle$, $D_2 = \langle \nabla^2 F(x)h, h \rangle^{1/2}$, and $\tau = \frac{\xi}{p} D_2^2$. Then

$$\xi'' = \frac{\xi}{p^2} D_1^2 - \tau \le 0,$$

$$\xi''' \overset{(5.1.4)}{\le} \frac{2\xi}{p} D_2^3 + \frac{3\xi}{p^2} D_1 D_2^2 - \frac{\xi}{p^3} D_1^3 = 2\tau D_2 + \frac{1}{p} D_1 \left(3\tau - \frac{\xi}{p^2} D_1^2 \right)$$

$$= 2\tau D_2 + \frac{1}{p} D_1 \left(2\tau - \xi'' \right) \overset{(5.3.6)}{\le} 2\tau D_2 + \frac{\sqrt{\nu}}{p} D_2 \left(2\tau - \xi'' \right).$$

Note that $\xi'' + \tau = \frac{\xi}{p^2} D_1^2 \overset{(5.3.6)}{\le} \frac{\xi \nu}{p^2} D_2^2 = \frac{\nu}{p} \tau$. Thus, $\tau \le \frac{p}{p-\nu}(-\xi'')$, and therefore

$$\xi''' \le D_2 \left(2(1 + \frac{\sqrt{\nu}}{p})\tau + \frac{\sqrt{\nu}}{p}(-\xi'') \right) \le D_2 \left(\frac{2}{\sqrt{p}-\sqrt{\nu}} + \frac{\sqrt{\nu}}{p} \right) (-\xi'').$$

This means that for $p \ge (1 + \sqrt{\nu})^2$ the function $\xi_p(x)$ is 1-compatible with F and by Theorem 5.4.4 we get a $(\nu + 1)$-self-concordant barrier

$$\Psi_H(x, t) = - \ln \left(\exp \left\{ -\frac{1}{p} F(x) \right\} - t \right) + F(x) \tag{5.4.39}$$

for the set $\mathcal{H}_p$.

5. The matrix epigraph of the inverse matrix. Consider the following set

$$\mathcal{I}_n = \{(X, Y) \in \mathbb{S}_+^n \times \mathbb{S}_+^n : X^{-1} \preceq Y\}.$$

In order to construct a barrier for this set, consider the mapping $\xi(X) = -X^{-1}$. It is defined on the set of positive definite matrices, for which we know a ν-self-concordant barrier $F(X) = -\ln \det X$ with the barrier parameter $\nu = n$ (see Theorem 5.4.3). Let us show that ξ is 1-compatible with F.

Indeed, let us fix an arbitrary direction $H \in \mathbb{S}^n$. By the same reasoning as in Lemma 5.4.6, we can prove that

$$D\xi(X)[H] = X^{-1} H X^{-1},$$

$$D^2\xi(X)[H, H] = -2X^{-1} H X^{-1} H X^{-1} \in -S_+^n,$$

$$D^3\xi(X)[H, H, H] = 6X^{-1} H X^{-1} H X^{-1} H X^{-1}.$$

Let $A = X^{-1/2} H X^{-1/2}$ and $\rho = \max\limits_{1 \le i \le n} |\lambda_i(A)|$. Then, in view of Lemma 5.4.6,

$$\langle \nabla^2 F(X) H, H \rangle = \|A\|_F^2 = \sum_{i=1}^n \lambda_i^2(A) \ge \rho^2.$$

On the other hand,

$$D^3 \xi(X)[H, H, H] = 6 X^{-1/2} A^3 X^{-1/2} \preceq 6\rho X^{-1/2} A^2 X^{-1/2}$$

$$\preceq 6 \langle \nabla^2 F(X) H, H \rangle^{1/2} X^{-1/2} A^2 X^{-1/2}$$

$$= 3 \langle \nabla^2 F(X) H, H \rangle^{1/2} D^2 F(X)[H, H].$$

Thus, condition (5.4.21) is satisfied with $\beta = 1$. Hence, by Theorem 5.4.4 the function

$$F(X, Y) = -\ln \det(Y - X^{-1}) - \ln \det X \qquad (5.4.40)$$

is a ν-self-concordant barrier for $\mathscr{I}_n$ with $\nu = 2n$.

Lemma 5.4.10 *Any self-concordant barrier for the set $\mathscr{I}_n$ has parameter $\nu \ge 2n$.*

Proof Let us choose $\gamma > 1$ and consider matrices $\bar{X} = \bar{Y} = \gamma I_n$. Clearly the point $(\bar{X}, \bar{Y})$ belongs to int $\mathscr{I}_n$. Note that for positive definite matrices, relation $Y \succeq X^{-1}$ holds if and only if $X \succeq Y^{-1}$. Therefore, all directions

$$p_i = (e_i e_i^T, 0), \quad q_i = (0, e_i e_i^T), \quad i = 1, \dots, n,$$

are recession directions of the set $\mathscr{I}_n$. It is easy to check that for $\beta = \gamma - \frac{1}{\gamma}$ we get

$$(\bar{X}, \bar{Y}) - \beta p_i \in \partial \mathscr{I}_n, \quad (\bar{X}, \bar{Y}) - \beta q_i \in \partial \mathscr{I}_n, \quad i = 1, \dots, n.$$

On the other hand, for $\alpha = \gamma - 1$, we have $\bar{Y} - \alpha \sum\limits_{i=1}^n e_i e_i^T = I_n = (\bar{X} - \alpha \sum\limits_{i=1}^n e_i e_i^T)^{-1}$. Therefore, in the conditions of Theorem 5.4.1 we can get all $\alpha_i = \alpha$ and all $\beta_i = \beta$. Thus, we obtain $\nu \ge 2n \frac{\alpha}{\beta} = \frac{2n\gamma}{1+\gamma}$. Since γ can be arbitrarily big, we come to the bound $\nu \ge 2n$. $\square$

5.4.8 Separable Optimization

In problems of Separable Optimization all nonlinear terms in functional components are represented by univariate functions. A general formulation of such a problem is

as follows:

$$\min_{x \in \mathbb{R}^n} \left\{ q_0(x) = \sum_{j=1}^{m_0} \alpha_{0,j} f_{0,j}(\langle a_{0,j}, x \rangle + b_{0,j}), \right.$$

$$\left. q_i(x) = \sum_{j=1}^{m_i} \alpha_{i,j} f_{i,j}(\langle a_{i,j}, x \rangle + b_{i,j}) \le \beta_i, \ i = 1 \dots m \right\}, \tag{5.4.41}$$

where $\alpha_{i,j}$ are some positive coefficients, $a_{i,j} \in \mathbb{R}^n$ and $f_{i,j}(\cdot)$ are convex functions of one variable. Let us rewrite this problem in the standard form:

$$\min_{x \in \mathbb{R}^n, \tau \in \mathbb{R}^{m+1}, t \in \mathbb{R}^M} \left\{ \tau_0 : \sum_{j=1}^{m_i} \alpha_{i,j} t_{i,j} \le \tau_i, \ i = 0 \dots m, \ \tau_i \le \beta_i, \ i = 1 \dots m, \right.$$

$$\left. f_{i,j}(\langle a_{i,j}, x \rangle + b_{i,j}) \le t_{i,j}, \ j = 1 \dots m_i, \ i = 0 \dots m, \right\}, \tag{5.4.42}$$

where $M = \sum_{i=0}^{m} m_i$. Thus, in order to construct a self-concordant barrier for the feasible set of this problem, we need barriers for epigraphs of univariate convex functions $f_{i,j}$. Let us point out such barriers for several important examples.

5.4.8.1   Logarithm and Exponent

By fixing the first coordinate in the barrier (5.4.33), we obtain the barrier function $F_1(x, t) = -\ln x - \ln(\ln x + t)$, which is a 3-self-concordant barrier for the set

$$Q_1 = \{(x, t) \in \mathbb{R}^2 \mid x > 0, \ t \ge -\ln x\}.$$

Similarly, we obtain the function $F_2(x, t) = -\ln t - \ln(\ln t - x)$ as a 3-self-concordant barrier for the set

$$Q_2 = \{(x, t) \in \mathbb{R}^2 \mid t \ge e^x\}.$$

5.4.8.2   Entropy Function

By fixing the second coordinate in the barrier (5.4.33), we obtain the barrier function $F_3(x, t) = -\ln x - \ln(t - x \ln x)$, which is a 3-self-concordant barrier for the set

$$Q_3 = \{(x, t) \in \mathbb{R}^2 \mid x \ge 0, \ t \ge x \ln x\}.$$

5.4.8.3   Increasing Power Functions

Let $p \geq 1$ and define $\alpha = \frac{1}{p}$. By fixing the second variable in barrier (5.4.28), $x^{(2)} = 1$, we get function $F_4(x, t) = -\ln t - \ln(t^{2/p} - x^2)$, which is a 4-self-concordant barrier for the set

$$Q_4 = \{(x, t) \in \mathbb{R}^2 \mid t \geq |x|^p\}, \quad p \geq 1.$$

If $p < 1$, then a similar operation with the barrier (5.4.29) gives us the function $F_5(x, t) = -\ln t - \ln(t^p - x)$, which is a 3-self-concordant barrier for the set

$$Q_5 = \{(x, t) \in \mathbb{R}^2 \mid t \geq 0, \ t^p \geq x\}, \quad 0 < p \leq 1.$$

5.4.8.4   Decreasing Power Functions

Let $p > 0$. Define $\alpha = \frac{p}{p+1}$. Then by fixing $z = 1$ in the barrier (5.4.29), we get the function $F_6(x, t) = -\ln x - \ln t - \ln(x^\alpha t^{1-\alpha} - 1)$, which is a 3-self-concordant barrier for the set

$$Q_6 = \left\{(x, t) \in \mathbb{R}^2 \mid x > 0, \ t \geq \frac{1}{x^p}\right\}.$$

Let us conclude our discussion with two examples.

5.4.8.5   Geometric Optimization

The initial formulation of such problems is as follows:

$$\min_{x \in \mathbb{R}_{++}^n} \left\{ q_0(x) = \sum_{j=1}^{m_0} \alpha_{0,j} \prod_{j=1}^{n} (x^{(j)})^{\sigma_{0,j}^{(j)}}, \right.$$

$$q_i(x) = \sum_{j=1}^{m_i} \alpha_{i,j} \prod_{j=1}^{n} (x^{(j)})^{\sigma_{i,j}^{(j)}} \leq 1, \ i = 1 \ldots m \right\}, \quad (5.4.43)$$

where $\mathbb{R}_{++}^N$ is the interior of the positive orthant, and $\alpha_{i,j}$ are some positive coefficients. Note that the problem (5.4.43) is not convex.

Let us introduce vectors $a_{i,j} = (\sigma_{i,j}^{(1)}, \ldots, \sigma_{i,j}^{(n)}) \in \mathbb{R}^n$, and change variables:

$$x^{(i)} = e^{y^{(i)}}, \quad i = 1, \ldots, n.$$

Then problem (5.4.43) can be written in a *convex* form.

$$\min_{y \in \mathbb{R}^n} \left\{ \sum_{j=1}^{m_0} \alpha_{0,j} \exp(\langle a_{0,j}, y \rangle) : \ \sum_{j=1}^{m_i} \alpha_{i,j} \exp(\langle a_{i,j}, y \rangle) \le 1, \ i = 1 \ldots m \right\}.$$
(5.4.44)

Let $M = \sum_{i=0}^{m} m_i$. The complexity of solving (5.4.44) by a path-following scheme is $O\left(M^{1/2} \cdot \ln \frac{M}{\epsilon}\right)$ iterations.

5.4.8.6 Approximation in an ℓ_p-Norm

The simplest problem of this type is as follows:

$$\min_{x \in \mathbb{R}^n} \left\{ \sum_{i=1}^{m} | \langle a_i, x \rangle - b^{(i)} |^p : \ \alpha \le x \le \beta \right\},$$
(5.4.45)

where $p \ge 1$ and $\alpha, \beta \in \mathbb{R}^n$. Clearly, we can rewrite this problem in an equivalent standard form:

$$\min_{x \in \mathbb{R}^n, \tau \in \mathbb{R}^{m+1}} \left\{ \tau^{(0)} : | \langle a_i, x \rangle - b^{(i)} |^p \le \tau^{(i)}, \ i = 1 \ldots m, \right.$$
$$\left. \sum_{i=1}^{m} \tau^{(i)} \le \tau^{(0)}, \ \alpha \le x \le \beta \right\}.$$
(5.4.46)

The complexity bound of this problem is $O\left(\sqrt{m+n} \cdot \ln \frac{m+n}{\epsilon}\right)$ iterations of a path-following scheme.

We have discussed the performance of Interior-Point Methods for several *pure* optimization problems. However, it is important that we can apply these methods to *mixed* problems. For example, in problems (5.4.11) or (5.4.45) we can also treat the quadratic constraints. To do this, we need to construct a corresponding self-concordant barrier. Such barriers are known for all important functional components arising in practical applications.

5.4.9 Choice of Minimization Scheme

We have seen that the majority of convex optimization problems can be solved by Interior-Point Methods. However, the same problems can also be solved by methods of Nonsmooth Optimization. In general, we cannot say which approach is better, since the answer depends on the individual structure of a particular problem. However, the complexity estimates for optimization schemes are often helpful in making the choice. Let us consider a simple example.

Assume we are going to solve a problem of finding the best approximation in an ℓ_p-norm:

$$\min_{x \in \mathbb{R}^n} \left\{ \sum_{i=1}^{m} |\langle a_i, x \rangle - b^{(i)}|^p : \; \alpha \le x \le \beta \right\}, \tag{5.4.47}$$

where $p \ge 1$. We have two available numerical methods:

- The Ellipsoid Method (Sect. 3.2.8).
- The Interior-Point Path-Following Scheme.

Which of them should we use? The answer can be derived from the complexity analysis of the corresponding schemes.

Firstly, let us estimate the performance of the Ellipsoid Method as applied to problem (5.4.47).

Complexity of the Ellipsoid Method

Number of iterations: $\qquad O\left(n^2 \ln \frac{1}{\epsilon}\right),$

Complexity of the oracle: $\quad O(mn)$ operations,

Complexity of the iteration: $O(n^2)$ operations.

Total complexity: $\qquad\qquad O\left(n^3(m+n) \ln \frac{1}{\epsilon}\right)$ operations.

The analysis of the Path-Following Method is more involved. First of all, we should form a barrier model of the problem:

$$\min_{x \in \mathbb{R}^n, \tau \in \mathbb{R}^m, \xi \in \mathbb{R}} \left\{ \xi : |\langle a_i, x \rangle - b^{(i)}|^p \le \tau^{(i)}, \; i = 1 \ldots m, \right.$$

$$\left. \sum_{i=1}^{m} \tau^{(i)} \le \xi, \; \alpha \le x \le \beta \right\},$$

$$\tag{5.4.48}$$

$$F(x, \tau, \xi)) = \sum_{i=1}^{m} f(\langle a_i, x \rangle - b^{(i)}, \tau^{(i)}) - \ln(\xi - \sum_{i=1}^{m} \tau^{(i)})$$

$$- \sum_{i=1}^{n} [\ln(x^{(i)} - \alpha^{(i)}) + \ln(\beta^{(i)} - x^{(i)})],$$

where $f(y, t) = -\ln t - \ln(t^{2/p} - y^2)$.

We have seen that the parameter of barrier $F(x, \tau, \xi)$ is $\nu = 4m + n + 1$. Therefore, the Path-Following Scheme needs $O\left(\sqrt{4m + n + 1} \cdot \ln \frac{m+n}{\epsilon}\right)$ iterations at most.

At each iteration of this method, we need to compute the gradient and the Hessian of barrier $F(x, \tau, \xi)$. Define

$$g_1(y, t) = \nabla_y f(y, t), \quad g_2(y, t) = f'_t(y, t).$$

Then

$$\nabla_x F(x, \tau, \xi) = \sum_{i=1}^{m} g_1(\langle a_i, x \rangle - b^{(i)}, \tau^{(i)}) a_i - \sum_{i=1}^{n} \left[\frac{1}{x^{(i)} - \alpha^{(i)}} - \frac{1}{\beta^{(i)} - x^{(i)}} \right] e_i,$$

$$F'_{\tau^{(i)}}(x, \tau, \xi) = g_2(\langle a_i, x \rangle - b^{(i)}, \tau^{(i)}) + \left[\xi - \sum_{j=1}^{m} \tau^{(j)} \right]^{-1},$$

$$F'_{\xi}(x, \tau, \xi) = - \left[\xi - \sum_{i=1}^{m} \tau^{(i)} \right]^{-1}.$$

Further, defining

$$h_{11}(y, t) = \nabla^2_{yy} F(y, t), \quad h_{12}(y, t) = \nabla^2_{yt} F(y, t), \quad h_{22}(y, t) = F''_{tt}(y, t),$$

we obtain

$$\nabla^2_{xx} F(x, \tau, \xi) = \sum_{i=1}^{m} h_{11}(\langle a_i, x \rangle - b^{(i)}, \tau^{(i)}) a_i a_i^T$$

$$+ \mathrm{diag}\left[\frac{1}{(x^{(i)} - \alpha^{(i)})^2} + \frac{1}{(\beta^{(i)} - x^{(i)})^2} \right],$$

$$\nabla^2_{\tau^{(i)} x} F(x, \tau, \xi) = h_{12}(\langle a_i, x \rangle - b^{(i)}, \tau^{(i)}) a_i,$$

$$F''_{\tau^{(i)}, \tau^{(i)}}(x, \tau, \xi) = h_{22}(\langle a_i, x \rangle - b^{(i)}, \tau^{(i)}) + \left(\xi - \sum_{i=1}^{m} \tau^{(i)} \right)^{-2},$$

$$F''_{\tau^{(i)}, \tau^{(j)}}(x, \tau, \xi) = \left(\xi - \sum_{i=1}^{m} \tau^{(i)} \right)^{-2}, \quad i \neq j,$$

$$\nabla^2_{x, \xi} F(x, \tau, \xi) = 0, \quad F''_{\tau^{(i)}, \xi}(x, \tau, \xi) = -\left(\xi - \sum_{i=1}^{m} \tau^{(i)} \right)^{-2},$$

$$F''_{\xi, \xi}(x, \tau, \xi) = \left(\xi - \sum_{i=1}^{m} \tau^{(i)} \right)^{-2}.$$

Thus, the complexity of the second-order oracle in the Path-Following Scheme is $O(mn^2)$ arithmetic operations.

Let us estimate now the complexity of each iteration. The main source of computations at each iteration is the solution of the Newton system. Let

$$\varkappa = \left(\xi - \sum_{i=1}^{m} \tau^{(i)}\right)^{-2}, \quad s_i = \langle a_i, x \rangle - b^{(i)}, \ i = 1 \ldots n,$$

and

$$\Lambda_0 = \text{diag}\left[\frac{1}{(x^{(i)} - \alpha^{(i)})^2} + \frac{1}{(\beta^{(i)} - x^{(i)})^2}\right]_{i=1}^{n} \ \Lambda_1 = \text{diag}\left(h_{11}(s_i, \tau^{(i)})\right)_{i=1}^{m},$$

$$\Lambda_2 = \text{diag}\left(h_{12}(s_i, \tau^{(i)})\right)_{i=1}^{m}, \qquad\qquad D = \text{diag}\left(h_{22}(s_i, \tau^{(i)})\right)_{i=1}^{m}.$$

Then, using the notation $A = (a_1, \ldots, a_m)$, $\bar{e}_m = (1, \ldots, 1) \in \mathbb{R}^m$, the Newton system can be written in the following form:

$$[A(\Lambda_0 + \Lambda_1)A^T]\Delta x + A\Lambda_2\Delta\tau = \nabla_x F(x, \tau, \xi),$$

$$\Lambda_2 A^T \Delta x + [D + \varkappa I_m]\Delta\tau + \varkappa\bar{e}_m\Delta\xi = F'_\tau(x, \tau, \xi), \qquad\qquad (5.4.49)$$

$$\varkappa\langle\bar{e}_m, \Delta\tau\rangle + \varkappa\Delta\xi = F'_\xi(x, \tau, \xi) + t,$$

where t is the penalty parameter. From the second equation in (5.4.49), we obtain

$$\Delta\tau = [D + \varkappa I_m]^{-1}(F'_\tau(x, \tau, \xi) - \Lambda_2 A^T \Delta x - \varkappa\bar{e}_m\Delta\xi).$$

Substituting $\Delta\tau$ into the first equation in (5.4.49), we have

$$\Delta x = [A(\Lambda_0 + \Lambda_1 - \Lambda_2^2[D + \varkappa I_m]^{-1})A^T]^{-1}\{\nabla_x F(x, \tau, \xi)$$

$$- A\Lambda_2[D + \varkappa I_m]^{-1}(F'_\tau(x, \tau, \xi) - \varkappa\bar{e}_m\Delta\xi)\}.$$

Using these relations, we can find $\Delta\xi$ from the last equation in (5.4.49).

Thus, the Newton system (5.4.49) can be solved in $O(n^3 + mn^2)$ operations. This implies that the total complexity of the Path-Following Scheme can be estimated as

$$O\left(n^2(m + n)^{3/2} \cdot \ln \frac{m+n}{\epsilon}\right)$$

arithmetic operations. Comparing this estimate with the bound for the Ellipsoid Method, we conclude that the Interior-Point Method is more efficient if m is not too big, namely, if $m \leq O(n^2)$.

Of course, this analysis is valid only if the methods behave in accordance with their worst-case complexity bounds. For the Ellipsoid Method this is indeed true. However, Interior-Point Path-Following Schemes can be accelerated by long-step strategies. The explanation of these abilities requires the introduction of a primal-dual setting of the optimization problems, posed in a conic form. Because of the volume constraints, we have decided not to touch on this deep theory in the present book.