



Set-Limited Functions and Polynomial-Time Interior-Point Methods

Yurii Nesterov¹ 

Received: 24 March 2022 / Accepted: 13 January 2023
© The Author(s) 2023

Abstract

In this paper, we revisit some elements of the theory of self-concordant functions. We replace the notion of self-concordant barrier by a new notion of set-limited function, which forms a wider class. We show that the proper set-limited functions ensure polynomial time complexity of the corresponding path-following method (PFM). Our new PFM follows a deviated path, which connects an arbitrary feasible point with the solution of the problem. We present some applications of our approach to the problems of unconstrained optimization, for which it ensures a global linear rate of convergence even in for nonsmooth objective function.

Keywords Convex optimization · Interior-point methods · Self-concordant functions · Polynomial-time methods

Mathematics Subject Classification 90C25

1 Introduction

Motivation In the end of the last century, Polynomial-Time interior-point methods (IPM) were the most popular tools for solving convex optimization problems. Starting from the famous papers by Karmarkar [7], Renegar [8], Gonzaga [5], and many others, these methods completely changed our abilities in solving Linear Optimization Problems. The further extension onto nonlinear optimization problems was done by the theory of self-concordant functions [15]. Due to this theory, it became possi-

Communicated by Goran Lesaja.

✉ Yurii Nesterov
Yurii.Nesterov@uclouvain.be

¹ Center of Operations Research and Econometrics (CORE), Catholic University of Louvain (UCLouvain), Louvain-la-Neuve, Belgium

ble to develop linearly convergent methods for Linear Matrix Inequalities and other important classes of structured convex optimization problems (see [2, 12]).

However, the machinery underlying the general IPM is not simple. The most efficient methods require rewriting the initial problem into the primal-dual conic form (see [10, 17]). Moreover, the process of solution of these problem has to be divided on two or even three stages, which makes the practical implementation of the new methods a nontrivial task.

The algorithmic complexity of IPM and their high iteration cost were the main factors in switching the research priorities onto much simpler gradient-type methods. Development of the smoothing technique [13], supported by the demands of Big Data, opened for these optimization schemes new possibilities of acceleration, which overpass the limits of the classical Complexity Theory [11]. In parallel, we have seen intensive development of the new second-order methods [6, 9, 14, 16], which now benefit from the global complexity analysis even for nonconvex problems [1, 4].

However, an unavoidable drawback of the first- and second-order Black Box Schemes consists in their sublinear global rate of convergence on general classes of convex problems. Since in the most practical applications the internal structure of convex problem is quite visible, these methods compete with linearly convergent IPM, which use this structure for creating a powerful descriptor of the problem, the *self-concordant barrier* of the feasible set.

In this paper, we revisit some basic elements of the theory of self-concordant functions, having in mind simplification of the existing IPM. This theory is based on two concepts, the notions of self-concordant function and notion of self-concordant barrier. The first one is necessary for a proper description of the behavior of Newton Method, and the second one is responsible for the global polynomial-time complexity of IPM. Both concepts are local in the sense that they assume some relations between directional derivatives of convex function, computed at the same point.

In our presentation, we replace the notion of self-concordant barrier by a new concept of *set-limited function*, which requires boundedness of the variation of the gradient with respect to the current point. This condition is clearly global and it is much easier for verification. Thus, we significantly increase the class of good barriers. On the other hand, using a new line of arguments, we show that the polynomial-time complexity of the corresponding schemes is preserved.

Our second development is the Greedy Path-Following Method. In the standard framework of IPM, it is necessary to follow the *central path*, which starts in a close neighborhood of the *analytic center* of feasible set. Thus, usually a preliminary stage is needed for approaching this center (e.g. Section 4.2 in [12]). In our new method, we start moving towards the optimum immediately from the starting point, by following a *deviated path*. We present some simple characteristics of the starting point ensuring the polynomial-time complexity of this procedure.

As important application examples, we consider problems of unconstrained minimization, where we know barriers for the epigraph of the objective function. We show how to choose the starting point in the epigraphs, which ensure the global linear rate of convergence of the corresponding methods. Note that our scheme updates some objects in the epigraph of the objective function, similarly to the methods based on the

overestimating technique (e.g. [3, 16]). However, in contrast to them, our new method benefits from the global linear rate of convergence even for nonsmooth functions.

Contents In Sect. 2, we introduce the notion of a proper set-limited function, which replaces in our presentation the notion of self-concordant barrier. The class of such functions is much wider. However, we prove that the main properties of self-concordant barrier are preserved.

In Sect. 3, we describe the Greedy Path-Following Method, which follows the *deviated path*, starting from an arbitrary feasible point. We highlight the conditions ensuring polynomial-time complexity of this scheme. As a side result, we show that in the case of sharp minimum, the deviated path asymptotically approached the standard central path.

In the last Sect. 4, we discuss several applications of our results to problem of unconstrained optimization. In all cases, we show how to choose the starting point in the epigraph of the objective function, in order to ensure the linear rate of convergence of the scheme. Note that this rate is achieved even if the objective function is nonsmooth.

Notation and Generalities In what follows, we denote by \mathbb{E} a finite-dimensional real vector space, and by \mathbb{E}^* its dual space composed by linear functions on \mathbb{E} . For such a function $s \in \mathbb{E}^*$, we denote by $\langle s, x \rangle$ its value at $x \in \mathbb{E}$.

We measure distances in \mathbb{E} by arbitrary norm $\|\cdot\|_{\mathbb{E}}$, denoting by \mathcal{B} the corresponding unit ball. Then the dual norm is defined in the standard way:

$$\|g\|_{\mathbb{E}}^* = \max_{x \in \mathcal{B}} \langle g, x \rangle, \quad g \in \mathbb{E}^*.$$

For two sets Q_1 and Q_2 in \mathbb{E} , we say that $Q_1 \subset Q_2$ if there exists some $\epsilon > 0$ such that $Q_1 + \epsilon \mathcal{B} \subseteq Q_2$. Sometimes we measure distances in \mathbb{E} by Euclidean norm $\|\cdot\|$. It is defined by a self-adjoint positive-definite operator linear $B : \mathbb{E} \rightarrow \mathbb{E}^*$ in the following way:

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

For a smooth function $f : \mathbb{E} \rightarrow \mathbb{R}$, denote by $\nabla f(x)$ its gradient, and by $\nabla^2 f(x)$ its Hessian evaluated at point $x \in \text{dom } f$. Note that

$$\nabla f(x) \in \mathbb{E}^*, \quad \nabla^2 f(x)h \in \mathbb{E}^*, \quad h \in \mathbb{E}.$$

Another possibility for measuring distances is given by the local norms defined by a self-concordant function (see Sect. 4.2 in [12]). Recall that function $f(\cdot)$ is called self-concordant if it is a closed convex function with open domain, which satisfies the following condition:

$$D^3 f(x)[h]^3 \leq 2 \langle \nabla^2 f(x)h, h \rangle^{3/2}, \quad x \in \text{dom } f, \quad h \in \mathbb{E},$$

where notation in the left-hand side corresponds to the third directional derivative of function $f(\cdot)$ along direction h . The Hessian of self-concordant function provides us with the local Euclidean norms:

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

Among many useful properties of self-concordant functions, we employ the following inequality (e.g. Theorem 4.1.7 in [12]):

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{\|y - x\|_x^2}{1 + \|y - x\|_x}, \quad x, y \in \text{dom } f. \quad (1)$$

In what follows, we often use different statements from Chapter 4 of [12]. We put the corresponding references in bold. Thus, reference **T.1.2** means Theorem 4.1.2 from [12].

2 Set-Limited Functions

Let Q be a closed convex set in \mathbb{E} with nonempty interior. Denote by $F(\cdot)$ a self-concordant function with $\text{dom } F = \text{int } Q$. In this paper, we consider the following *standard* minimization problem:

$$c_* = \min_{x \in Q} \langle c, x \rangle, \quad (2)$$

where c is a linear functional from \mathbb{E}^* .

In order to justify efficiency bounds for corresponding optimization methods, we need to introduce additional assumptions on $F(\cdot)$. In the standard theory of Polynomial-Time Interior-Point Methods [15], we assume that $F(\cdot)$ is a ν -self-concordant barrier:

$$\langle \nabla F(x), h \rangle^2 \leq \nu \langle \nabla^2 F(x)h, h \rangle, \quad x \in \text{dom } F, \quad h \in E, \quad (3)$$

where the *barrier parameter* $\nu \geq 1$ is responsible for the complexity of problem (2). In this way, it is possible to justify polynomial-time solvability for many important classes of optimization problems (see [15]). However, the barrier property (3) is very fragile. Even an addition to $F(\cdot)$ a linear function destroys the values of the barrier parameter. The main goal of this paper is to replace (3) by a more robust definition, which still gives us a possibility to prove polynomial-time complexity of interior-point methods.

Note that the condition (3) is local. It looks as an upper bound on the size of the gradient in the local norm defined by the Hessian of the barrier. And maybe this is the main reason of its fragility. In what follows, we replace it by a global condition, related to the size of the gradient with respect to the *feasible set*.

Definition 2.1 We call convex function $F(\cdot)$ \varkappa -*set-limited* with respect convex set $Q \subseteq \text{dom } F$ if there exists a constant $\varkappa \geq 0$ such that for any $x, y \in Q$ we have

$$\langle \nabla F(x), y - x \rangle \leq \varkappa. \quad (4)$$

This definition has a geometric interpretation. Define the polar set at point $x \in \text{int } Q$ as follows:

$$P_Q(x) = \left\{ g \in \mathbb{E}^* : \langle g, y - x \rangle \leq 1, \ y \in Q \right\}.$$

Then the condition (4) is equivalent to the following inclusion:

$$\frac{1}{\varkappa} \nabla F(x) \in P_Q(x), \quad x \in \text{int } Q. \tag{5}$$

Note that inequality (4) with $\varkappa = \nu$ is one of the main *properties* of self-concordant barrier (see **T.2.4(1)**). By lifting it up to the status of definition, we significantly increase the class of good barrier functions.

Let us mention simple properties of set-limited functions, which do not need proofs.

- If $F(\cdot)$ is \varkappa -set-limited, then for any $\lambda > 0$ function $\lambda F(\cdot)$ is $(\lambda \varkappa)$ -set-limited.
- If functions $F_i(\cdot)$ are \varkappa_i -set-limited with respect to the sets $Q_i, i = 1, 2$, then function $F(x) = F_1(x) + F_2(x)$ is $(\varkappa_1 + \varkappa_2)$ -set-limited with respect to $Q = Q_1 \cap Q_2$.
- If $F(\cdot)$ has bounded variation on the set Q , then it is set-limited with respect to Q with parameter $\varkappa = \text{Var}_Q(F) \stackrel{\text{def}}{=} \sup_{x,y \in Q} [F(y) - F(x)]$.

The next property is also simple. However, we put it separately for future references.

Lemma 2.1 *Let function $F(\cdot)$ be set-limited and p be a recession direction of its domain. Then*

$$\langle \nabla F(z), p \rangle \leq 0, \quad z \in \text{dom } F. \tag{6}$$

Proof Indeed, since the inequality $\langle \nabla F(z), (z + \tau p) - z \rangle \stackrel{(4)}{\leq} \varkappa$ is valid for arbitrarily large $\tau > 0$, we get (6). □

The following result sometimes is useful.

Lemma 2.2 *Let $f(\cdot)$ be differentiable and concave on the set $\Omega = \{x \in \mathbb{E} : f(x) \geq 0\}$ and $\text{int } \Omega \neq \emptyset$. Then function $F(x) = -\ln f(x)$ is set-limited on Ω with $\varkappa = 1$.*

Proof Indeed, let $f(x) > 0$ and $f(y) \geq 0$. Then, since $f(\cdot)$ is concave, we have

$$\langle \nabla F(x), y - x \rangle = \frac{1}{f(x)} \langle \nabla f(x), x - y \rangle \leq \frac{1}{f(x)} (f(x) - f(y)) \leq 1.$$

□

For the needs of interior-point methods, we specify an additional property of set-limited functions.

Definition 2.2 If set-limited function is self-concordant with respect to its domain, we call it the *proper* set-limited function.

Recall that self-concordant functions are closed and convex and they have open domains. Let us prove that proper set-limited functions belong to the family of self-concordant barriers.

Lemma 2.3 *Any proper \varkappa -self-limited function is also a \varkappa^2 -self-concordant barrier.*

Proof For $x \in \text{dom } F$, let us choose an arbitrary direction $h \in \mathbb{E}$. If necessary, we can multiply it by -1 , ensuring anyway the inequality $\langle \nabla F(x), h \rangle \geq 0$. In view of **T.4.1.5(1)**, the point $y = x + h/\|h\|_x$ belongs to $\text{Cl}(\text{dom } F)$. Since $F(\cdot)$ is \varkappa -set-limited, we have

$$0 \leq \langle \nabla F(x), h \rangle = \langle \nabla F(x), y - x \rangle \|h\|_x \stackrel{(4)}{\leq} \varkappa \|h\|_x.$$

Thus, the definition (3) is valid with $v = \varkappa^2$. □

One of the consequences of this statement is that for proper set-limited functions we always have $\varkappa \geq 1$ (see **L.3.1**).

Let us show that the proper set-limited functions inherit one of the most important properties of self-concordant barriers. It is important that for this property we can use parameter \varkappa , not \varkappa^2 .

Theorem 2.1 *Let $F(\cdot)$ be a proper set-limited function with parameter \varkappa . Then for all $x, y \in \text{dom } F$ with $\langle \nabla F(x), y - x \rangle \geq 0$ we have*

$$\|y - x\|_x \leq \varkappa + 2\sqrt{\varkappa}. \tag{7}$$

Proof Denote $r = \|x - y\|_x$, and let $r > \sqrt{\varkappa}$ (otherwise, (7) is trivial). Choosing $\alpha = \frac{1}{r}\sqrt{\varkappa} < 1$ for $y_\alpha = x + \alpha(y - x)$, we get

$$\begin{aligned} \omega &\stackrel{\text{def}}{=} \langle \nabla F(y_\alpha), y - x \rangle \geq \langle \nabla F(y_\alpha) - \nabla F(x), y - x \rangle \\ &= \frac{1}{\alpha} \langle \nabla F(y_\alpha) - \nabla F(x), y_\alpha - x \rangle \stackrel{(1)}{\geq} \frac{1}{\alpha} \cdot \frac{\|y_\alpha - x\|_x^2}{1 + \|y_\alpha - x\|_x} \\ &= \frac{\alpha \|y - x\|_x^2}{1 + \alpha \|y - x\|_x} = \frac{r\sqrt{\varkappa}}{1 + \sqrt{\varkappa}}. \end{aligned}$$

At the same time, $(1 - \alpha)\omega = \langle \nabla F(y_\alpha), y - y_\alpha \rangle \stackrel{(4)}{\leq} \varkappa$. Thus, $(1 - \frac{1}{r}\sqrt{\varkappa}) \frac{r\sqrt{\varkappa}}{1 + \sqrt{\varkappa}} \leq \varkappa$, and this is exactly inequality (7). □

The following result provides us with many examples of proper set-limited functions.

Lemma 2.4 *Let $F_1(\cdot)$ be v -self-concordant barrier for set $Q \stackrel{\text{def}}{=} \text{dom } F_1$ and $F_2(\cdot)$ be a self-concordant function with $Q \subset \text{dom } F_2$. Then function $F(x) = F_1(x) + F_2(x)$ is a proper \varkappa -set-limited with respect to Q with $\varkappa = v + \text{Var } Q(F_2)$.*

Proof Indeed, for any x and y from Q , we have

$$\begin{aligned} \langle \nabla F(x), y - x \rangle &= \langle \nabla F_1(x), y - x \rangle + \langle \nabla F_2(x), y - x \rangle \stackrel{(4)}{\leq} \nu + F_2(y) - F_2(x) \\ &\leq \nu + \text{Var}_Q(F_2). \end{aligned} \quad \square$$

It is convenient to use Lemma 2.4 when $\text{dom } F_2 = \mathbb{E}$. Therefore, let us recall the most important examples of such functions.

- $F_2(x) = \frac{1}{2} \langle Qx, x \rangle$, where matrix Q is positive semidefinite.
- $F_2(x) = \phi^*(x) \stackrel{\text{def}}{=} \sup_{g \in \text{dom } \phi} [\langle g, x \rangle - \phi(g)]$, where $\phi(\cdot)$ is a self-concordant function defined on a bounded open convex set $\text{dom } \phi \subset \mathbb{E}^*$ (see Theorem 2.4.1 in [15]).

3 Greedy Path-Following Method

Let $F(\cdot)$ be a proper set-limited function with respect to the set Q with parameter \varkappa . For solving optimization problem (2), we propose the following simple scheme.

Greedy Path-Following Method

Initialization. Choose $x_0 \in \text{int } Q$ and $\beta \in (0, \frac{1}{3}]$.

Set $\gamma = \frac{\sqrt{\beta}}{1 + \sqrt{\beta}} - \beta$ and $g_0 = -\nabla F(x_0)$. (8)

k th iteration ($k \geq 0$): Update $g_{k+1} = g_k + \frac{\gamma c}{\|c\|_{x_k}^*}$.

Set $x_{k+1} = x_k - [\nabla^2 F(x_k)]^{-1}(\nabla F(x_k) + g_{k+1})$.

We call this method greedy since it immediately attacks the problem (2), without preliminary finding the analytic center of the set Q , as it is advised by the standard theory (e.g. Section 4.2 in [12]). Note that the bounds on β ensure $\gamma > 0$.

Defining $t_0 = 0$ and $t_{k+1} = t_k + \frac{\gamma}{\|c\|_{x_k}^*}$ for $k \geq 0$, we can see that $g_k = g_0 + t_k c$. Let us prove that $t_k \rightarrow \infty$ and method (8) follows approximately the sequence of minimizers of the auxiliary problems

$$\min_{x \in Q} \left\{ f_k(x) \stackrel{\text{def}}{=} \langle g_k, x \rangle + F(x) \right\}, \quad k \geq 0.$$

Lemma 3.1 Let $\beta \in (0, \frac{1}{3}]$. Then, for any $k \geq 0$, we have

$$\|\nabla F(x_k) + g_k\|_{x_k}^* \leq \beta. \tag{9}$$

Proof For $k = 0$, the left-hand side of inequality (9) is zero, so it is satisfied.

Note that the main step at each iteration of method (8) is the Newton Step from the point x_k for the potential $f_{k+1}(\cdot)$. Assume that (9) is valid for certain $k \geq 0$. Then

$$\|\nabla f_{k+1}(x_k)\|_{x_k}^* = \left\| \nabla F(x_k) + g_k + \frac{\gamma c}{\|c\|_{x_k}^*} \right\|_{x_k}^* \stackrel{(9)}{\leq} \beta + \gamma.$$

Therefore, in view of **T.1.14**, we have

$$\|\nabla f_{k+1}(x_{k+1})\|_{x_{k+1}}^* \leq \left(\frac{\beta + \gamma}{1 - \beta - \gamma} \right)^2 = \beta.$$

Hence, inequality (9) is proved for all $k \geq 0$. □

Thus, we have proved that method (8) follows approximately a *deviated path* $x_d(\cdot)$, defined by the equation

$$\nabla F(x_d(t)) + tc - \nabla F(x_0) = 0, \quad t \geq 0. \tag{10}$$

It starts at an arbitrary point $x_d(0) = x_0 \in \text{int } Q = \text{dom } F$, and we can expect that it approaches an optimal solution of the problem (2) as $t \rightarrow \infty$. In the standard theory of Polynomial-Time Interior-Point Methods, this property is ensured by two additional assumptions. Firstly, we assume that $F(\cdot)$ is also a ν -self-concordant barrier. Another assumption is that we follow the *central path*, which starts from the *analytic center* x_F of the set Q , defined by condition $\nabla F(x_F) = 0$. In our analysis, we drop the last assumption and study the behavior of $x_d(\cdot)$ assuming that $F(\cdot)$ is a proper set-limited function.

First of all, we need to prove that the big values of t provide us with a good approximation of the optimal solution.

Theorem 3.1 *Let $F(\cdot)$ be a proper \varkappa -set-limited function. Then, for all $t > 0$ we have*

$$\langle c, x_d(t) \rangle - c_* \leq \frac{\kappa_0}{t}, \quad \kappa_0 \stackrel{\text{def}}{=} 2\varkappa + \langle \nabla F(x_0), x_0 - x^* \rangle. \tag{11}$$

Moreover, if x_k satisfies the approximate centering condition (9), then

$$\langle c, x_k \rangle - c_* \leq \frac{(1 - \beta)\kappa_0}{(1 - 2\beta)t_k}. \tag{12}$$

Proof Indeed,

$$\begin{aligned} t \langle c, x_d(t) - x^* \rangle &\stackrel{(10)}{=} \langle -\nabla F(x_d(t)) + \nabla F(x_0), x_d(t) - x^* \rangle \\ &\stackrel{(4)}{\leq} \varkappa + \langle \nabla F(x_0), x_d(t) - x^* \rangle \stackrel{(4)}{\leq} 2\varkappa + \langle \nabla F(x_0), x_0 - x^* \rangle. \end{aligned}$$

Note that by **T.1.5(1)**, $\|c\|_{x_k}^* \leq \langle c, x_k \rangle - c_*$. Therefore, in view of **T.1.13**, we have

$$\langle c, x_k - x_d(t_k) \rangle \leq \|c\|_{x_k}^* \|x_k - x_d(t_k)\|_{x_k} \stackrel{(9)}{\leq} \frac{\beta}{1 - \beta} \|c\|_{x_k}^* \leq \frac{\beta}{1 - \beta} (\langle c, x_k \rangle - c_*).$$

Thus, $\langle c, x_k \rangle - c_* \leq \frac{1 - \beta}{1 - 2\beta} (\langle c, x_d(t_k) \rangle - c_*)$, and (12) follows from (11). □

From the theory of self-concordant barriers, it is known that the path-following strategy along the central path has linear rate of convergence, which depends only on the value of barrier parameter ν (e.g. [12]). Let us show that for the deviated paths this role is now taken by the constant κ_0 .

Theorem 3.2 *Let $\beta \in (0, \frac{2 - \sqrt{3}}{2})$. Then, for any $k \geq 1$, we have*

$$t_k \geq \frac{t_1}{2} \cdot 2^{k/\hat{k}_\beta}, \tag{13}$$

where $\hat{k}_\beta = 1 + \sqrt{\frac{\kappa_\beta}{\gamma(\gamma - \beta)}}$ with $\kappa_\beta = \frac{1 - \beta}{1 - 2\beta} \kappa_0$.

Proof Denote by \hat{k} the smallest integer such that $\hat{k} \geq \sqrt{\frac{\kappa_\beta}{\gamma(\gamma - \beta)}}$. Since the values t_k are monotonically increasing, for all $k, 1 \leq k \leq \hat{k}$, we have

$$t_k \geq t_1 \geq t_1 \cdot 2^{(k - \hat{k})/\hat{k}}. \tag{14}$$

Let us assume now that inequality (14) is valid for all $k, 1 \leq k \leq n$, where $n \geq \hat{k}$. Note that the upper bound for β ensures $\gamma > \beta$. At the same time,

$$\begin{aligned} \langle c, x_{k+1} \rangle &= \langle c, x_k \rangle - \left\langle c, [\nabla^2 F(x_k)]^{-1} \left(\nabla F(x_k) + g_k + \frac{\gamma c}{\|c\|_{x_k}^*} \right) \right\rangle \\ &\stackrel{(9)}{\leq} \langle c, x_k \rangle + \beta \|c\|_{x_k}^* - \gamma \|c\|_{x_k}^*. \end{aligned}$$

Thus, $0 < (\gamma - \beta) \|c\|_{x_k}^* \leq \langle c, x_k \rangle - \langle c, x_{k+1} \rangle$, and we conclude that

$$\begin{aligned} t_{n+1} - t_{n+1 - \hat{k}} &= \sum_{i=n+1 - \hat{k}}^n \frac{\gamma}{\|c\|_{x_i}^*} \geq \sum_{i=n+1 - \hat{k}}^n \frac{\gamma(\gamma - \beta)}{\langle c, x_i \rangle - \langle c, x_{i+1} \rangle} \\ &\geq \frac{\gamma(\gamma - \beta)}{\langle c, x_{n+1 - \hat{k}} \rangle - c_*} \cdot \hat{k}^2 \end{aligned}$$

since $\sum_{i=n+1 - \hat{k}}^n [\langle c, x_i \rangle - \langle c, x_{i+1} \rangle] = \langle c, x_{n+1 - \hat{k}} \rangle - \langle c, x_{n+1} \rangle \leq \langle c, x_{n+1 - \hat{k}} \rangle - c_*$. Note that

$$\langle c, x_{n+1 - \hat{k}} \rangle - c_* \stackrel{(12)}{\leq} \frac{\kappa_\beta}{t_{n+1 - \hat{k}}}.$$

Therefore, $t_{n+1} \geq t_{n+1-\hat{k}} \left(1 + \frac{\gamma(\gamma-\beta)}{\kappa\beta} \hat{k}^2\right) \geq 2t_{n+1-\hat{k}}$. Hence, in view of assumption (14),

$$t_{n+1} \geq 2t_1 \cdot 2^{(n+1-2\hat{k})/\hat{k}} = t_1 \cdot 2^{(n+1-\hat{k})/\hat{k}}.$$

Thus, inequality (14) is proved for all $k \geq 1$. It remains to note that $\hat{k} \leq \sqrt{\frac{\kappa\beta}{\gamma(\gamma-\beta)}} + 1$. □

Thus, if we start from a point x_0 with $\kappa_0 \leq O(\varkappa)$, then the efficiency of Greedy Path-Following Method (8) remains on the level of a standard path-following scheme, equipped with a \varkappa -self-concordant barrier. Note that we keep the standard possibility to start from a neighborhood of the analytic center x_F since

$$\begin{aligned} \langle \nabla F(x_0), x_0 - x^* \rangle &\leq \|\nabla F(x_0)\|_{x_F}^* (\|x_0 - x_F\|_{x_F} + \|x^* - x_F\|_{x_F}) \\ &\stackrel{(7)}{\leq} (\|x_0 - x_F\|_{x_F} + \varkappa + 2\sqrt{\varkappa}) \|\nabla F(x_0)\|_{x_F}^*. \end{aligned}$$

Thus, if $\|\nabla F(x_0)\|_{x_F}^*$ is smaller than an absolute constant, we have $\kappa_0 \leq O(\varkappa)$. However, sometimes we have another possibilities for reaching this relation (see Sect. 4).

In the opposite case, if κ_0 is very big, then we cannot guarantee good complexity bounds. However, in one particular situation, any deviated path nevertheless approaches asymptotically the central path.

Definition 3.1 We say that problem (2) has sharp minimum if there exists a constant $\rho > 0$ such that for all $x \in Q$ we have

$$\langle c, x \rangle - c_* \geq \rho \|x - x^*\|. \tag{15}$$

Our analysis is based on the following result, which was initially proved in [18] for the dual formulation of a conic optimization problem.

Lemma 3.2 *Let function $F(\cdot)$ be self-concordant with $\text{dom } F = \text{int } Q$ and the minimum of problem (2) be sharp. Then for any $x \in \text{int } Q$ we have*

$$\nabla^2 F(x) \geq \frac{\rho^2}{4[\langle c, x \rangle - c_*]^2} B. \tag{16}$$

Proof Let $x \in \text{int } Q$. By **T.1.5(1)**, the ellipsoid $W = \{y \in \mathbb{E} : \|y - x\|_x \leq 1\}$ belongs to Q . Denote $\tau = \langle c, x \rangle - c_*$. Then $\langle c, x \rangle - \|c\|_x^* \geq c_*$. Thus, $\|c\|_x^* \leq \tau$ and we conclude that for any $y \in W$ we have $\langle c, y \rangle - c_* \leq 2\tau$. This means that $\|y - x^*\| \leq \frac{(15) 2\tau}{\rho}$. Consequently, for any $g \in \mathbb{E}^*$, we have

$$\max_{y \in W} \langle g, y \rangle = \langle g, x \rangle + \|g\|_x^* \leq \langle g, x^* \rangle + \frac{2\tau}{\rho} \|g\|_x^*.$$

Without loss of generality, we assume that $\langle g, x^* \rangle \leq \langle g, x \rangle$ (otherwise, multiply g by -1). Thus, we have proved the inequality $\|g\|_x^* \leq \frac{2\tau}{\rho} \|g\|^*$. Since g is arbitrary, we get (16). \square

In order to get point x_k in a neighborhood of the central path, we need to ensure that the norm $\|\nabla F(x_k) + t_k c\|_{x_k}^*$ is smaller than an absolute constant. Note that

$$\begin{aligned} \|\nabla F(x_k) + t_k c\|_{x_k}^* &\leq \|\nabla F(x_k) + g_k\|_{x_k}^* + \|\nabla F(x_0)\|_{x_k}^* \\ &\stackrel{(16)}{\leq} \beta + \frac{2}{\rho} [\langle c, x_k \rangle - c_*] \|\nabla F(x_0)\|^*. \end{aligned}$$

Hence, by inequality (12), we obtain an upper bound for this moment.

4 Second-Order Methods for Structured Unconstrained Minimization

Let us consider the following problem of unconstrained convex optimization:

$$\min_{x \in \mathbb{E}} f(x), \tag{17}$$

where function $f(\cdot)$ has an explicit structure, which allows us to point out a proper set-limited function F with respect to its epigraph $\mathcal{E}_f = \{z = (\tau, x) \in \mathbb{R} \times \mathbb{E} : \tau \geq f(x)\}$. This gives us a possibility to solve problem (17) by a linearly convergent method (8) as applied to the initial optimization problem, rewritten in the standard form:

$$\min_z \left\{ \langle c, z \rangle \equiv \tau : z = (\tau, x) \in \mathcal{E}_f \right\}. \tag{18}$$

Note that a special form of the feasible set in (18) gives us additional possibilities in keeping the value $\langle \nabla F(z_0), z_0 - z^* \rangle$ small enough (see (11) and (13) for its role in the complexity bounds). Indeed, given by arbitrary starting point $x_0 \in \mathbb{E}$, we are free to choose any starting value $\tau_0 > f(x_0)$. If τ_0 is big enough, then we can expect the value of $\langle \nabla F(z_0), z_0 - z^* \rangle$ to be small, or even negative. Note that the condition

$$\langle \nabla F(z_0), z_0 - z^* \rangle \leq 0 \tag{19}$$

allows us to estimate the distance to the solution from the starting point by Theorem 2.1. At the same time, it is much weaker than any of the "centering" conditions for the starting point. Let us look at the following example. \square

Example 4.1 Let the objective function in problem (17) is $f(x) = |x|$, $x \in \mathbb{R} \equiv \mathbb{E}$. We can rewrite this problem in the standard form as follows:

$$\min_{z=(\tau,x) \in \mathbb{R}^2} \left\{ \tau : \tau \geq |x| \right\}. \tag{20}$$

Thus, a natural self-concordant barrier for epigraph of the objective is $F(z) = -\ln(\tau^2 - x^2)$ with barrier parameter $\nu = 2$. The central path $z(t)$ of problem (2) is the trajectory of minimizers of the following function:

$$t\tau - \ln(\tau^2 - x^2), \quad t > 0.$$

Thus, $z(t) = (\frac{2}{t}, 0)$. A narrow neighborhood \mathcal{N}_β of the central path, which is appropriate for the standard path-following scheme, can be found from the following condition:

$$z \in \mathcal{N}_\beta \Leftrightarrow \min_{t>0} \|z - z(t)\|_z \leq \beta,$$

where $\beta \in (0, \frac{1}{3})$ (see Sect. 4.2 in [12]). Simple but tedious computations tell us that

$$\min_{t>0} \|z - z(t)\|_z^2 = \frac{2x^2}{x^2 + \tau^2}.$$

Hence, the neighborhood \mathcal{N}_β has the following representation:

$$\mathcal{N}_\beta = \left\{ (\tau, x) : \tau \geq \sqrt{\frac{2}{\beta^2} - 1} \cdot |x| \right\}.$$

Thus, if β is small, this is a tiny cone around the horizontal axis in \mathbb{R}^2 . The standard theory of Interior-Point Methods requires to start the path-following process exactly from this region. Let us look what happens with condition (19). Since $z^* = (0, 0)$, for any $z \in \text{dom } F$ we have:

$$\langle \nabla F(z), z - z^* \rangle = \langle \nabla F(z), z \rangle = -2.$$

Thus, *any point* in $\text{dom } F$ is appropriate for starting the corresponding deviated path, which approaches the optimal solution of problem (20) with linear rate. □

In the remaining part of this section, we show how to ensure condition (19) for different problem settings.

4.1 Lipschitz-Continuous Functions

Let us assume that the objective function of problem (17) satisfies the following condition:

$$f(x) - f^* \leq L_f \|x - x^*\|_{\mathbb{E}}, \quad x \in \mathbb{E}, \tag{21}$$

where $\|\cdot\|_{\mathbb{E}}$ is an arbitrary norm on \mathbb{E} . This means that all elements of the cone

$$\mathcal{R} = \left\{ z = (\tau, x) \in \mathbb{R} \times \mathbb{E} : \tau \geq L_f \|x\|_{\mathbb{E}} \right\}$$

are recession directions of the epigraph \mathcal{E}_f .

Let us choose arbitrary $x_0 \in \mathbb{E}$ with a known bound to the optimum: $\|x_0 - x^*\|_{\mathbb{E}} \leq R$. Then we can take

$$z_0 = (f(x_0) + L_f R, x_0). \tag{22}$$

In this case, for direction $p_0 = z_0 - z^* = (f(x_0) - f^* + L_f R, x_0 - x^*)$, we have

$$L_f \|x_0 - x^*\|_{\mathbb{E}} \leq L_f R \leq f(x_0) - f^* + L_f R.$$

Thus, $p_0 \in \mathcal{R}$, and by Lemma 2.1 we conclude that

$$\varkappa_0 \stackrel{(11)}{=} 2\varkappa + \langle \nabla F(z_0), z_0 - z^* \rangle \stackrel{(6)}{\leq} 2\varkappa.$$

Hence, the choice (22) of the starting point ensures the following complexity of method (8) as applied to problem (17):

$$O\left(\sqrt{\varkappa} \ln \frac{\|c\|_{z_0}^*}{\epsilon}\right) \tag{23}$$

iterations of the path-following scheme.

From the view point of iteration complexity, method (8) is a second-order scheme. However, it can be applied to the objective functions in (17) satisfying no specific assumptions on their smoothness except boundedness of the first-order derivatives (21). An alternative approach could consist in applying to $f(\cdot)$ a variant of smoothing technique [13] and minimizing the result by a second-order method. However, to the best of our knowledge, all existing strategies of this kind have sublinear rate of convergence.

4.2 Max-Type Functions

Let $f(x) = \max_{1 \leq i \leq m} f_i(x)$, where the functions $f_i(\cdot)$ are closed and convex, with epigraphs

$$\mathcal{E}_i = \{z = (\tau, x) : \tau \geq f_i(x)\}$$

admitting v_i -self-concordant barriers in the form $F_i(z) = -\ln(\tau - f_i(x))$, $i = 1, \dots, m$. For an important example of quadratic $f_i(\cdot)$, we have all $v_i = 1$. Note that quadratic functions do not satisfy assumption (21) of Sect. 4.1.

For the standard barrier $F(z) = -\sum_{i=1}^m \ln(\tau - f_i(x))$, we have $\varkappa = \sum_{i=1}^m \nu_i$. At the same time,

$$\begin{aligned} \langle \nabla F(z_0), z_0 - z^* \rangle &= \sum_{i=1}^m \frac{1}{\tau_0 - f_i(x_0)} [\tau^* - \tau_0 + \langle \nabla f_i(x_0), x_0 - x^* \rangle] \\ &\leq \sum_{i=1}^m \frac{1}{\tau_0 - f_i(x_0)} [f(x_0) - \tau_0 + \langle \nabla f_i(x_0), x_0 - x^* \rangle] \\ &\leq \sum_{i=1}^m \frac{f(x_0) - f_i(x_0)}{\tau_0 - f(x_0)} + \sum_{i=1}^m \frac{\|\nabla f_i(x_0)\|_{\mathbb{E}}^*}{\tau_0 - f(x_0)} R - m, \end{aligned}$$

where $R \geq \|x_0 - x^*\|_{\mathbb{E}}$. Thus, in order to have $\langle \nabla F(z_0), z_0 - z^* \rangle \leq 0$, it is enough to choose

$$\tau_0 = f(x_0) + \frac{1}{m} \sum_{i=1}^m \left\{ f(x_0) - f_i(x_0) + R \|\nabla f_i(x_0)\|_{\mathbb{E}}^* \right\}. \tag{24}$$

In this case, the corresponding implementation of method (8) admits the polynomial-time complexity bound (23).

4.3 Average Function

Let $f(x) = \frac{1}{m} \sum_{i=1}^m f_i(\langle a_i, x \rangle - b_i)$, where the univariate functions $f_i(\cdot)$ are convex on \mathbb{R} . We assume also that they are Lipschitz-continuous:

$$|f_i(s_1) - f_i(s_2)| \leq L_i |s_1 - s_2|, \quad s_1, s_2 \in \mathbb{R}, \quad i = 1, \dots, m. \tag{25}$$

Let the epigraphs $\mathcal{E}_i = \{z = (\tau, s) \in \mathbb{R}^2 : \tau \geq f_i(s)\}$ admit ν_i -self-concordant barriers $F_i(\cdot, \cdot)$ with parameters $\nu_i, i = 1, \dots, m$. Then we can use the following proper set-limited function

$$F(\tau, x) = \sum_{i=1}^m F_i(\tau^{(i)}, \langle a_i, x \rangle - b_i), \quad \tau = (\tau^{(1)}, \dots, \tau^{(m)}) \in \mathbb{R}^m, \quad x \in \mathbb{E},$$

with parameter $\varkappa = \sum_{i=1}^m \nu_i$. Our objective function now is $\frac{1}{m} \sum_{i=1}^m \tau^{(i)}$.

Let $z = (\tau, x)$. Our goal is to find the starting point z_0 such that $\langle \nabla F(z_0), z_0 - z^* \rangle \leq 0$. In view of Lemma 2.1, for that we need to ensure the inclusion

$$z_0 - z^* \in \mathcal{R} = \left\{ (\tau, x) : \tau^{(i)} \geq L_i |\langle a_i, x \rangle|, \quad i = 1, \dots, m \right\}.$$

Thus, we can handle each coordinate independently. Note that for $R \geq \|x_0 - x^*\|_{\mathbb{E}}$ we have

$$\begin{aligned}
 & L_i |\langle a_i, x_0 - x^* \rangle| \\
 & \leq L_i \|a_i\|_{\mathbb{E}}^* R \stackrel{(25)}{\leq} 2L_i \|a_i\|_{\mathbb{E}}^* \cdot R + f_i(\langle a_i, x_0 \rangle - b_i) - f_i(\langle a_i, x^* \rangle - b_i).
 \end{aligned}$$

Therefore, we can take

$$\tau_0^{(i)} = f_i(\langle a_i, x_0 \rangle - b_i) + 2L_i \|a_i\|_{\mathbb{E}}^* \cdot R, \quad i = 1, \dots, m. \tag{26}$$

In this case, the corresponding second-order scheme (8) has the complexity bound (23).

As an example, consider the following problem, arising in Machine Learning:

$$\min_{x \in \mathbb{E}} \left\{ \frac{1}{m} \sum_{i=1}^m \phi_0(\langle a_i, x \rangle - b_i) \right\}, \quad \phi_0(s) = \max\{0, s\}, \quad L_{\phi_0} = 1. \tag{27}$$

This is a convex nonsmooth optimization problem, which admits only slowly convergent optimization schemes. Therefore, for accelerating the methods, very often the function $\phi_0(\cdot)$ is replaced by its smooth approximation $\phi_\mu(s) = \mu \ln(1 + e^{s/\mu})$ with $\mu > 0$, augmenting sometimes the objective in (27) by a strongly convex regularization term.

In our approach, we get linearly convergent scheme directly for the problem (27). For that, we need to endow the epigraph $\mathcal{E}_{\phi_0} = \{(\tau, s) \in \mathbb{R}^2 : \tau \geq \phi_0(s)\}$ with the standard 2-self-concordant barrier

$$F(\tau, s) = -\ln(\tau - s) - \ln \tau, \quad \tau > 0, \quad s \in \mathbb{R},$$

and use method (8) for the epigraph of the objective function.

Acknowledgements This paper has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (Grant agreement No 788368). It was also supported by Multidisciplinary Institute in Artificial intelligence MIAI@Grenoble Alpes (ANR-19-P3IA-0003). The author is thankful to two anonymous refereed for their useful suggestions.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Birgin, E.G., Gardenghi, J.L., Martinez, J.M., Santos, S.A., Toint, Ph.L.: Worst-case evaluation complexity for unconstrained nonlinear optimization using high-order regularized models. *Math. Program.* **163**, 359–368 (2017)
2. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, Cambridge (2004)

3. Cartis, C., Gould, N., Toint, Ph.: Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. *Math. Program.* **127**(2), 245–295 (2011)
4. Cartis, C., Gould, N., Toint, Ph.: Complexity bounds for second-order optimality in unconstrained optimization. *J. Complex.* **28**(1), 93–108 (2012)
5. Gonzaga, C.: Path-following methods for linear programming. *SIAM Rev.* **34**(2), 167–224 (1992)
6. Grapiglia, G., Nesterov, Yu.: Accelerated regularized Newton method for minimizing composite convex functions. *SIOPT* **29**(1), 77–99 (2019)
7. Karmarkar, N.: A new polynomial time algorithm for linear programming. *Combinatorica* **4**(4), 373–395 (1984)
8. Renegar, J.: A polynomial-time algorithm, based on Newton’s method, for linear programming. *Math. Program.* **40**, 59–93 (1988)
9. Monteiro, R.D.C., Svaiter, B.F.: An accelerated hybrid proximal extragradient method for convex optimization and its implications to the second-order methods. *SIOPT* **23**(2), 1092–1125 (2013)
10. Nemirovski, A.: Advances in convex optimization: conic programming. *Int. Congr. Math.* **1**, 413–444 (2007)
11. Nemirovsky, A., Yudin, D.: *Problem Complexity and Method Efficiency in Optimization*. J Wiley @ Sons, New York (1983)
12. Nesterov, Yu.: *Introductory Lectures on Convex Optimization*. Kluwer, Boston (2004)
13. Nesterov, Yu.: Smooth minimization of non-smooth functions. *Math. Program. (A)* **103**(1), 127–152 (2005)
14. Nesterov, Yu.: Accelerating the cubic regularization of Newton’s method on convex problems. *Math. Program.* **112**(1), 159–181 (2008)
15. Nesterov, Yu., Nemirovskii, A.: *Interior Point Polynomial Methods in Convex Programming: Theory and Applications*. SIAM, Philadelphia (1994)
16. Nesterov, Yu., Polyak, B.: Cubic regularization of Newton’s method and its global performance. *Math. Program.* **108**(1), 177–205 (2006)
17. Nesterov, Yu., Nemirovskii, A.: Conic duality and its applications in convex programming. *Optim. Methods Softw.* **1**, 95–115 (1992)
18. Nesterov, Yu., Tunçel, L.: Local superlinear convergence of polynomial-time interior-point methods for hyperbolicity cone optimization problems. *SIOPT* **26**(1), 139–170 (2016)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.