Foundations of Computational Mathematics
https://doi.org/10.1007/s10208-025-09726-6

Improved global performance guarantees of second-order methods in convex minimization

Pavel Dvurechensky[1] · **Yurii Nesterov**[2,3,4]

Received: 16 October 2024 / Revised: 30 May 2025 / Accepted: 14 July 2025

Abstract

In this paper, we attempt to compare two distinct branches of research on second-order optimization methods. The first one studies self-concordant functions and barriers, the main assumption being that the third derivative of the objective is bounded by the second derivative. The second branch studies cubic regularized Newton methods (CRNMs) with the main assumption that the second derivative is Lipschitz continuous. We develop a new theoretical analysis for a path-following scheme (PFS) for general self-concordant functions, as opposed to the classical path-following scheme developed for self-concordant barriers. We show that the complexity bound for this scheme is better than that of the Damped Newton Method (DNM) and show that our method has global superlinear convergence. We propose also a new predictor-corrector path-following scheme (PCPFS) that leads to further improvement of constant factors in the complexity guarantees for minimizing general self-concordant functions. We also apply path-following schemes to different classes of constrained optimization problems and obtain the resulting complexity bounds. Finally, we analyze an important subclass of general self-concordant functions, namely a class of strongly convex functions with Lipschitz continuous second derivative, and show that for this subclass CRNMs give even better complexity bounds.

Communicated by Damek Davis.

✉ Pavel Dvurechensky
pavel.dvurechensky@wias-berlin.de

Yurii Nesterov
Yurii.Nesterov@uclouvain.be

1 Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, Germany

2 Corvinus Centre for Operations Research, Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Budapest Fövám tér 8, 1093, Hungary

3 School of Data Sciences, Chinese University of Hong Kong (Shenzhen), part time, Dao Yuan Building, 2001 Longxiang Boulevard, Longgang District, Shenzhen, China

4 Professor emeritus at CORE/INMA, Catholic University of Louvain (UCL), 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium

Published online: 13 August 2025

Keywords Self-concordant function · Damped Newton Method · Cubic Regularized Newton Method · Path-following method

Mathematics Subject Classification 90C25 · 90C30 · 68Q25

1 Introduction

Motivation. Local performance guarantees for the second-order methods are known since 1948 [21]. In that paper, the author proved a local quadratic convergence of the Newton method under some natural assumptions (non-degeneracy of the Hessian at solution and local Lipschitz continuity of the Hessian). However, in some sense, the quadratic convergence is too fast: each step of such methods doubles the number of right digits in the approximate solution. Therefore, questions on the acceleration of these schemes were never raised in the literature [7]. Moreover, for many years the only global complexity results for second-order methods were obtained in the framework of the theory of self-concordant functions and barriers [34, 35].

The situation was changed after the paper [36], where the first global complexity bounds were obtained for the cubic regularized Newton method (CRNM). Namely, it was shown that for convex functions with globally Lipschitz continuous Hessian the CRNM converges in terms of the function value as $O(\frac{1}{k^2})$, where k is the iteration counter. Very soon it was shown that this method can be accelerated up to the rate $O(\frac{1}{k^3})$ using the technique of estimating sequences [31]. Under the same assumptions, based on a special line-search procedure, the authors of [30] proposed a second-order method with convergence rate $O\left(\frac{1}{k^{7/2}}\right)$.

Thus, at this moment there exist two main, nearly independent, frameworks for global complexity analysis of the second-order methods. One is based on the affine-invariant theory of self-concordant functions. And the second one assumes bounded third derivatives of the objective in a fixed Euclidean norm. In the recent years, many new results were obtained within the second framework, including development of accelerated schemes. The main goal of this paper is to revisit the first framework and try to improve existing complexity bounds. The secondary goal of this paper is to show that these two classes of problems do intersect and we can compare the efficiency of the corresponding methods. Towards the first goal, we derive new complexity bounds for a path-following scheme (PFS) as applied to the unconstrained minimization of a self-concordant *function*. This result is new since the known complexity bounds for path-following methods are related to self-concordant *barriers* (see, for example, Section 4.2 in [34]). Moreover, this result shows that our PFS exhibits *global superlinear convergence*, and the resulting complexity bound is better than that of the classical Damped Newton Method (DNM). We next propose a predictor-corrector path-following scheme (PCPFS) that leads to further improvement of constant factors in the complexity guarantees for minimizing self-concordant functions. Additionally, we show that our results lead to improved complexity for constrained problems where one can additionally use the self-concordant barrier property. We further compare our bounds with the complexity results for different versions of the CRNM on the class of strongly convex functions with Lipschitz continuous Hessian. It appears that such functions are self-concordant. We conclude that the latter methods are much more efficient when applied to this subclass of self-concordant functions.

Related work. Local convergence rate analysis of Newton method was extended in [24] to the case of composite optimization with the objective given as a sum of a twice-differentiable function and a simple convex function. Inexact proximal Newton methods for self-concordant functions with composite term were studied in [25], where global and local convergence were studied. For the Machine Learning applications, the authors of [39] propose an incremental Newton method with local superlinear convergence and global linear convergence. The authors of [44] study distributed minimization of self-concordant empirical losses via DNM combined with distributed preconditioned conjugate gradient method. A new version of Newton method with adaptive stepsizes was proposed in [38] for non-linear systems of equations. Generalization of DNM for generalized self-concordant functions is developed together with local convergence analysis in [41]. The authors of [20] analyze global and local convergence of a DNM for minimizing a subclass of self-concordant functions called semi-strongly self-concordant functions. Beyond classical second-order Newton-type methods, local superlinear convergence was obtained for path-following interior-point methods [32]. The papers [8, 27] study inexact path-following method for optimization with special separable structure. In [42], the authors propose a new homotopy proximal variable-metric framework for composite convex minimization and show that under appropriate assumptions such as strong convexity-type and smoothness, or self-concordance, their schemes can achieve global linear convergence. A general Adaptive Regularization algorithm using Cubics was proposed in [5], where global and local convergence were proved under relaxed assumptions. Global complexity analysis was extended for the case of Hölder-continuous Hessians in [19] and even higher order derivatives in [6, 18, 40]. Linear convergence of the CRNM under Hölder continuity of the Hessians and uniform convexity was proved in [11]. The author of [9] obtains global complexity bounds for minimizing quasi-self-concordant functions by gradient regularization of Newton method. Frank-Wolfe method for minimization of self-concordant functions was recently proposed in [15] and a refined analysis for minimizing self-concordant barriers was made in [45]. The latter allowed also to construct a path-following Frank-Wolfe method in [13]. A combination of Frank-Wolfe and Newton method for minimization of self-concordant functions was proposed in [26], and close approach, but for functions with Lipschitz-continuous Hessian, was studied in [2]. In [16, 17] the authors combine the ideas of cubic regularization and self-concordant barriers to propose algorithms for constrained non-convex optimization with global complexity guarantees.

Previous works mainly analyze either the class of self-concordant functions or the class of functions with Lipschitz-continuous Hessians. In this work, we attempt to compare these two classes in terms of the global complexity of solving the corresponding minimization problem. At the same time, previous works analyzed the path-following methods for self-concordant barriers rather than for self-concordant functions. Here we do not assume the barrier property and analyze path-following schemes for general self-concordant functions. In particular, to the best of our knowledge, this is the first time when global superlinear convergence is obtained for this class of methods and problems.

Contributions. Our contributions can be summarized as follows.

1. We give a new theoretical analysis of a path-following scheme (PFS) for *general* self-concordant *functions*. This is in contrast to [34], where the classical analysis is made under additional assumption that the objective is self-concordant *barrier*. Moreover, the obtained global complexity bound of the PFS is better than that of the Damped Newton Method (DNM). In particular, we demonstrate global superlinear convergence of the PFS.
2. We consider a feasibility problem and compare complexity bounds for this problem by the DNM, PFS and dual PFS. Importantly, the proposed PFS gives the best complexity, improving upon known complexities.
3. We propose a predictor-corrector path-following scheme (PCPFS) for *general* self-concordant *functions*. This scheme has even better global complexity bound in terms of the constant factors than the PFS for minimizing general self-concordant functions and also possesses superlinear convergence.
4. We apply the PCPFS to two types of constrained optimization problems and under additional self-concordant barrier property show that the resulting primal and dual PCPFS have better constant factors in the complexity compared to the existing bounds in [34].
5. We propose variants of our path-following schemes with adaptive stepsizes that may take longer steps adapting to local behavior of the central path.
6. We make an observation that the above two classes of functions intersect. Namely, strongly convex functions with Lipschitz continuous Hessian belong to both classes. Thus, all the discussed methods can be applied to this class. Our conclusion is that on this class CRNMs possess better complexity that the DNM and the PFS.

We would also like to mention that besides the two classes of functions considered in this paper, important results on global performance of second-order methods are obtained for minimization of quasi-self-concordant functions [1], generalized self-concordant functions [41], functions with stable Hessians [4, 22]. An interesting open question for future research is to extend our comparison between two classes to include also these classes.

Contents. In Section 2 we recall the properties of self-concordant functions extending them to the case of non-standard self-concordant functions and give the complexity of the DNM. In Section 3, we provide description and new analysis of the PFS for general self-concordant functions. In Section 4 we discuss implications of our analysis and algorithms when applied to a feasibility problem. Section 5 is devoted to the construction and analysis of our PCPFS for minimizing general self-concordant functions. In Section 6, we apply the PCPFS to two types of constrained optimization problems that involve self-concordant barriers and obtain the resulting complexity. Section 7 contains complexity analysis of strongly convex functions with Lipschitz-continuous Hessians by CRNMs and their comparison to the complexity by the DNM and PFS.

Notation. Let $\mathbb{E}$ be a finite-dimensional real vector space and $\mathbb{E}^*$ be its dual. For a function f, we denote its domain as $\mathrm{dom} f = \{x \in \mathbb{E} : f(x) < +\infty\}$, and the closure of the domain as Dom f. Given a function f with Hessian $\nabla^2 f(x)$ that is non-degenerate at any $x \in \mathbb{E}$, we denote

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

$$\text{and} \;\; \lambda_f(x) = \|\nabla f(x)\|_x^*, \quad x \in \mathbb{E}. \tag{1}$$

We use $D^3 f(x)[h] = \lim_{\alpha \to 0} \frac{1}{\alpha}(\nabla^2 f(x+\alpha h) - \nabla^2 f(x))$ to denote the third-order derivative at a point $x \in \mathbb{E}$ in a direction $h \in \mathbb{E}$. For $h, h_1, h_2, h_3 \in \mathbb{E}$ we use a shortcut notation

$$D^3 f(x)[h]^3 = \langle D^3 f(x)[h]h, h\rangle, \quad D^3 f(x)[h_1, h_2, h_3] = \langle D^3 f(x)[h_1]h_2, h_3\rangle.$$

We define also, for $\tau \geq 0$, $\omega(\tau) = \tau - \ln(1+\tau)$, $\omega_*(\tau) = -\tau - \ln(1-\tau)$.

2 Minimizing self-concordant functions: Damped Newton Method

We consider the following minimization problem:

$$f^* = \min_{x \in \mathbb{E}} f(x), \tag{2}$$

where f is a general self-concordant function whose definition is given below. We assume that the Hessian of this function at any point is positive definite and that the solution x^* of problem (2) exists.

2.1 Preliminaries on general self-concordant functions

In this technical subsection we give the definition and properties of general self-concordant functions which we use throughout the paper. Let us start from a variant of the definition of self-concordant functions.

Definition 1 Let function f from $\mathbb{C}^3$, i.e., three times continuously differentiable, be convex on $\mathbb{E}$. It is called a general self-concordant function if there exists a constant $M_f \geq 0$ such that for any point $x \in \mathbb{E}$ and direction $h \in \mathbb{E}$:

$$|D^3 f(x)[h]^3| \leq 2M_f \langle \nabla^2 f(x)h, h\rangle^{3/2}. \tag{3}$$

If $M_f = 1$, then the function is called *standard self-concordant*. □

An equivalent characterization of general self-concordant functions is given by the following result [34, Lemma 4.1.2].

Lemma 1 *A function f from $\mathbb{C}^3$ is general self-concordant iff for any point $x \in \mathbb{E}$ and any $h_1, h_2, h_3 \in E$ we have*

$$|D^3 f(x)[h_1, h_2, h_3]| \leq 2M_f \|h_1\|_x \|h_2\|_x \|h_3\|_x. \tag{4}$$

It is clear that for any self-concordant function f, function

$$\tilde{f}(x) = M_f^2 f(x), \quad x \in \mathbb{E}, \tag{5}$$

is standard self-concordant. Standard self-concordant functions are more convenient for defining self-concordant barriers (see [35]). However, for the main algorithms in this paper we work directly with definition (3) and do not impose an additional assumption that f is a self-concordant barrier.

Taking into account normalization (5), we can rewrite all known properties of standard self-concordant functions for the general ones. Let us present the most important of them (see Section 4.1 in [34]). For all $y \in \mathbb{E}$ with $\|y - x\|_x < \frac{1}{M_f}$ we have

$$(1 - M_f\|y - x\|_x)^2 \nabla^2 f(x) \preceq \nabla^2 f(y) \preceq \frac{1}{(1-M_f\|y-x\|_x)^2} \nabla^2 f(x), \tag{6}$$

$$\|y - x\|_y \le \frac{\|y-x\|_x}{1-M_f\|y-x\|_x}, \tag{7}$$

$$f(y) \le f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{M_f^2}\omega_*(M_f\|y - x\|_x), \tag{8}$$

and for all $y \in \mathbb{E}$ we have

$$f(y) \ge f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{M_f^2}\omega(M_f\|y - x\|_x). \tag{9}$$

Inequality (9) leads to the following bound for $\lambda_f(x) < \frac{1}{M_f}$

$$f(x) - \min_{y\in\mathbb{E}} f(y) \le \frac{1}{M_f^2}\omega_*(M_f\lambda_f(x)). \tag{10}$$

Moreover, for $\lambda_f(x) < \frac{1}{M_f}$ and $x^* = \arg\min_{y\in\mathbb{E}} f(y)$,

$$\|x - x^*\|_x \le \frac{1}{M_f}\omega_*'(M_f\lambda_f(x)) = \frac{M_f\lambda_f(x)}{1-M_f\lambda_f(x)}. \tag{11}$$

Similarly to (6), if $\delta \equiv \|\nabla f(x) - \nabla f(y)\|_x^* < \frac{1}{M_f}$, then

$$(1 - M_f\delta)^2 \nabla^2 f(x) \preceq \nabla^2 f(y) \preceq \frac{1}{(1-M_f\delta)^2} \nabla^2 f(x). \tag{12}$$

This relation follows from the observation [35] that the Fenchel conjugate f_* for f defined as

$$f_*(s) = \sup_{x\in\mathbb{E}}[\langle s, x\rangle - f(x)]$$

is also self-concordant on its domain with the same constant M_f. Note also that, for $x(s) = \arg\max_{x\in\mathrm{dom}\, f}[\langle s, x\rangle - f(x)]$,

$$\nabla f_*(s) = x(s), \tag{13}$$

$$\nabla^2 f_*(s) = [\nabla^2 f(x(s))]^{-1}. \tag{14}$$

The following result was not proved in [34] even for the case of $M_f = 1$.

Lemma 2 *Let* $x, y \in \operatorname{dom} f$ *be such that* $\|y - x\|_x < \frac{1}{M_f}$. *Then*

$$\|\nabla f(y) - \nabla f(x)\|_x^* \le \frac{\|y - x\|_x}{1 - M_f \|y - x\|_x}.$$

Proof The proof follows from inequality (4.1.7) in [34], Cauchy-Schwarz and the duality relation (13). □

We finish this subsection with several examples of functions that are self-concordant, but not self-concordant barriers. For shortness, we denote this class of functions as $\mathcal{SCF}\backslash\mathcal{B}$. The simplest examples of such functions are linear and convex quadratic functions, see, e.g., [34, Example 4.2.1]. Further examples can be obtained as a sum of a self-concordant barrier and a self-concordant function. Here the examples are $\langle c, x\rangle + F(x)$, where $F(x)$ is a self-concordant barrier, and augmented self-concordant barriers [33] which have the form $\frac{1}{2}\langle Qx, x\rangle + F(x)$, where $F(x)$ is a self-concordant barrier for a cone K, and Q is a positive-semidefinite matrix. The entropy-barrier function which for $x \in \mathbb{R}$ s.t. $x > 0$ is defined as $x \ln x - \ln x$ is another example of a function from $\mathcal{SCF}\backslash\mathcal{B}$. Clearly, one can obtain a multivariate version of this function. As we will see in detail below in Section 7, another prominent example of functions in $\mathcal{SCF}\backslash\mathcal{B}$ is the class of strongly convex functions with Lipschitz Hessian. A particular example is the regularized logistic regression problem [41] which is one of the cornerstone optimization problems in machine learning.

Another example of functions from $\mathcal{SCF}\backslash\mathcal{B}$ may be constructed using Fenchel's duality. Consider a feasibility problem

$$\text{Find } x \text{ s.t. } x \in Q \text{ and } Ax = b, \tag{15}$$

where $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$. To solve this problem, we can introduce a self-concordant barrier $F(x)$ for the set Q and minimize it over the affine manifold:

$$\min_x F(x) \text{ s.t. } Ax = b. \tag{16}$$

Introducing a Lagrange multiplier y for the constraints $Ax = b$, we obtain

$$\min_x \{F(x): \; Ax = b\} = \max_y \{\langle b, y\rangle - F_*(A^T y)\}$$

and the dual problem for (16) is

$$\min_y \{\Phi(y) - \langle b, y\rangle\}, \tag{17}$$

where $\Phi(y) \overset{\text{def}}{=} F_*(A^T y)$, F_* being the Fenchel conjugate for F. In this case the dual is an unconstrained problem of minimization of $\Phi(y)$ that is a standard self-concordant function, but not a self-concordant barrier. We consider the above primal-dual pair of problems in more detail in Section 4.

2.2 Newton methods for general self-concordant functions.

We now move to standard algorithms used for minimizing general self-concordant functions. When applied to this class of functions, Standard Newton Method has local quadratic convergence.

Theorem 1 *[Theorem 4.1.14 in [34]] Define the Standard Newton Step*

$$x_+ = x - [\nabla^2 f(x)]^{-1} \nabla f(x). \tag{18}$$

Then,

$$\lambda_f(x_+) \le \tfrac{1}{M_f} \left(\tfrac{M_f \lambda_f(x)}{1 - M_f \lambda_f(x)} \right)^2. \tag{19}$$

Minimizing the right-hand side of inequality (8) in y, we come to the following result.

Theorem 2 *Define the Damped Newton Step*

$$x_+ = x - \tfrac{[\nabla^2 f(x)]^{-1} \nabla f(x)}{1 + M_f \lambda_f(x)}. \tag{20}$$

Then,

$$f(x_+) \le f(x) - \tfrac{1}{M_f^2} \omega(M_f \lambda_f(x)). \tag{21}$$

Moreover,

$$\lambda_f(x_+) \le 2 M_f \lambda_f^2(x). \tag{22}$$

Proof Inequality (21) follows from Theorem 4.1.2 in [34].

We now prove inequality (22), which was not proved in [34] even for $M_f = 1$. Denote $\lambda = \lambda_f(x)$, $h = x_+ - x$, $r = \|h\|_x = \frac{\lambda}{1 + M_f \lambda}$. Hence, $\lambda = \frac{r}{1 - M_f r}$. Note that, since $r < \frac{1}{M_f}$,

$$\lambda_+^2 \equiv \langle \nabla f(x_+), [\nabla^2 f(x_+)]^{-1} \nabla f(x_+) \rangle \overset{(6)}{\le} \tfrac{1}{(1 - M_f r)^2} \langle \nabla f(x_+), [\nabla^2 f(x)]^{-1} \nabla f(x_+) \rangle.$$

Without changing the notation, we can associate with the Hessians symmetric positive-definite matrices. Then, denoting $G = [\nabla^2 f(x)]^{1/2} \succ 0$, we have

$$\begin{aligned} \nabla f(x_+) &= \nabla f(x) + \int_0^1 \nabla^2 f(x + \tau h) h d\tau \\ &\overset{(20)}{=} -(1 + M_f \lambda) \nabla^2 f(x) h + \int_0^1 \nabla^2 f(x + \tau h) h d\tau \\ &= G \left[-(1 + M_f \lambda) I + G^{-1} \left(\int_0^1 \nabla^2 f(x + \tau h) d\tau \right) G^{-1} \right] G h. \end{aligned} \tag{23}$$

By Corollary 4.1.4 in [34], we have that

$$(1 - M_f r + \frac{1}{3} M_f^2 r^2) \nabla^2 f(x) \preceq \int_0^1 \nabla^2 f(x + \tau h) d\tau \preceq \frac{1}{1 - M_f r} \nabla^2 f(x).$$

Thus, denoting $H = [\cdot]$ in (23), we can see that $H \preceq 0$ and $H \succeq [-(1 + M_f \lambda) + (1 - M_f r)]I = [-(1 + M_f \lambda) + \frac{1}{1+M_f \lambda}]I \succeq -2\lambda M_f I$. So, we conclude that

$$\lambda_+^2 \leq \frac{1}{(1-M_f r)^2} (\|GHGh\|_x^*)^2 = \frac{1}{(1-M_f r)^2} \langle GH^2Gh, h\rangle \leq \frac{(2M_f\lambda)^2}{(1-M_f r)^2} \langle G^2 h, h\rangle$$

$$= \frac{(2M_f\lambda)^2}{(1-M_f r)^2} \langle \nabla^2 f(x)h, h\rangle = \frac{(2M_f\lambda)^2 r^2}{(1-M_f r)^2} = 4M_f^2 \lambda^4.$$

□

We can now analyze the efficiency of the Damped Newton Method (DNM)

$$x_{k+1} = x_k - \frac{[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)}{1+M_f \lambda_f(x_k)}, \quad k \geq 0 \tag{24}$$

in terms of iteration complexity as applied to the minimization problem (2). In view of inequality (22), method (24) starts converging quadratically when it enters the region

$$\mathbb{Q} = \left\{ x \in \mathbb{E} : \ \lambda_f(x) \leq \frac{1}{2M_f} \right\}. \tag{25}$$

This convergence is very fast and, in view of inequality (10), any reasonable accuracy in function value can be reached in a small number of iterations. Therefore, the main computational time is spent when $\lambda_f(x_k) \geq \frac{1}{2M_f}$. Denote by N the last iteration such that $\lambda_f(x_k) \geq \frac{1}{2M_f}$, $k = 0, \ldots, N$. Then, in view of inequality (21), we have

$$N \leq \frac{\Delta(x_0)}{\omega\left(\frac{1}{2}\right)}, \quad \Delta(x_0) \overset{\text{def}}{=} M_f^2 (f(x_0) - f^*). \tag{26}$$

This is the main result of this section. Namely, when applied to the class of general self-concordant functions, DNM has complexity $O(\Delta(x_0))$. In the following sections, we will show that this complexity may be improved by proposing a different algorithm.

Let us show that $\Delta(x_0)$ is a natural complexity measure of our problem class. In order to see this, let us attribute to our objects some physical units. Denote the units for measuring the function value by μ_f, and the units for measuring the argument by μ_x. Then, the units for measuring the gradient are $\mu_g = \mu_f / \mu_x$. The Hessian is measured in $\mu_h = \mu_f / \mu_x^2$, and the third derivative is measured in $\mu_t = \mu_f / \mu_x^3$. Thus, in view of definition (3), the units for measuring the constant M_f are $\mu_s = \mu_t \mu_x^3 / (\mu_h \mu_x^2)^{3/2} = \mu_f^{-1/2}$. Note that the number of iterations is an integer number with no physical dimension (scalar). Therefore, for using the constant M_f in the bounds for the number of iterations, it must be multiplied by something having physical dimension $\mu_f^{1/2}$. The simplest way to do this is to define the characteristic $\Delta(x_0)$ as in (26). In the sequel, we will use $\Delta(x_0)$ as the main characteristic of complexity of problem (2). Importantly, we can use the characteristics of our problem as arguments of nonlinear univariate functions only by transforming them to a scalar form. For example, the values $M_f \lambda_f(x)$ and $M_f \|h\|_x$ have no physical dimension and can be substituted to the functions ω, ω_*.

In addition to the global iteration complexity (26), we can show that the DNM accelerates to superlinear convergence when $\lambda_f(x_k) \le \frac{1}{M_f}$.

Theorem 3 *For the DNM* (24) *when* $\lambda_f(x_k) \le \frac{1}{M_f}$ *it holds that*

$$M_f^2(f(x_{k+1}) - f^*) \le M_f^2(f(x_k) - f^*)\left(1 - \frac{\omega(\omega_*^{-1}(M_f^2(f(x_k) - f^*)))}{M_f^2(f(x_k) - f^*)}\right). \quad (27)$$

In other words, the method has superlinear convergence.

Proof Denote for $k \ge 0$ $\Delta_k = f(x_k) - f^*$ and $\lambda_k = \lambda_f(x_k)$. Then, by (10), we have that $\omega_*^{-1}(M_f^2 \Delta_k) \le M_f \lambda_k$. Combining this with (21), we obtain

$$M_f^2 \Delta_{k+1} \le M_f^2 \Delta_k - \omega(\omega_*^{-1}(M_f^2 \Delta_k)),$$

which is (27). Denoting $\omega_*^{-1}(M_f^2 \Delta_k) = \xi$, we see that $\frac{\omega(\omega_*^{-1}(M_f^2 \Delta_k))}{M_f^2 \Delta_k} = \frac{\omega(\xi)}{\omega_*(\xi)}$. Recalling the definition of $\omega(\xi)$, $\omega_*(\xi)$ and using their Taylor expansion, we readily see that $\frac{\omega(\xi)}{\omega_*(\xi)} < 1$ and, hence, (27) leads to convergence. Further, since ω_*^{-1} monotonically increases, we have that ξ decreases as Δ_k decreases to 0. Considering the function $\frac{\omega(\xi)}{\omega_*(\xi)}$, its derivative, and Taylor expansions for $\omega(\xi)$, $\omega_*(\xi)$, it is easy to show that this function increases when ξ decreases and its limit when $\xi \to 0$ is equal to 1. Combining all the observations, we see that indeed the DNM has superlinear convergence when $\lambda_f(x_k) \le \frac{1}{M_f}$, i.e., local superlinear convergence. In the next section we propose a path-following scheme that has *global superlinear convergence*. □

3 Minimizing self-concordant functions: path-following scheme

In this section, we estimate the complexity of solving problem (2) by a path-following scheme (PFS). Note that the full justification of path-following methods is done so far in [34] only for self-concordant *barriers*. On the contrary, our analysis relies only on the less restrictive assumption that f is a self-concordant *function*. Moreover, we make the analysis for *general* self-concordant functions with $M_f > 0$ rather than $M_f = 1$. As for the DNM above, our main goal is to estimate the complexity to enter the region of quadratic convergence $\mathbb{Q}$ defined in (25). Let us start from some $x_0 \in \mathbb{E}$. Define the central path $x(t)$, $0 \le t \le 1$, by the following equation:

$$\nabla f(x(t)) = t \nabla f(x_0). \quad (28)$$

Clearly, $x(1) = x_0$ and $x(0) = x^*$. Note that this is a trajectory of minimizers of the following parametric family of general self-concordant functions:

$$x(t) = \arg\min_{x\in\mathbb{E}} \left\{ f_t(x) \stackrel{\text{def}}{=} f(x) - t\langle \nabla f(x_0), x\rangle \right\}, \quad 0 \le t \le 1. \tag{29}$$

Let us introduce two constants

$$\beta = 0.026, \quad \gamma \ = \ 0.1125 < \tfrac{\sqrt{\beta}}{1+\sqrt{\beta}} - \beta. \tag{30}$$

We say that a point x satisfies an *approximate centering condition* if

$$\lambda_{f_t}(x) \equiv \|\nabla f(x) - t\nabla f(x_0)\|_x^* \ \le \ \tfrac{\beta}{M_f}. \tag{31}$$

Consider the path-following iterate:

$$(t_+, x_+) \ = \ \mathcal{P}(t,x) \ \stackrel{\text{def}}{=} \ \begin{cases} t_+ = \max\left\{t - \frac{\gamma}{M_f\|\nabla f(x_0)\|_x^*}, 0\right\}, \\ x_+ = x - [\nabla^2 f(x)]^{-1}(\nabla f(x) - t_+\nabla f(x_0)). \end{cases} \tag{32}$$

The following statement is a counterpart of Theorem 4.2.8 in [34].

Theorem 4 *If the pair (x,t) satisfies (31), and β, γ are chosen such that*

$$|\gamma| \le \frac{\sqrt{\beta}}{1+\sqrt{\beta}} - \beta, \tag{33}$$

then the pair (x_+, t_+) satisfies (31) too.

Proof Let us denote $\lambda_0 = \|\nabla f(x) - t\nabla f(x_0)\|_x^*$, $\lambda_1 = \|\nabla f(x) - t_+\nabla f(x_0)\|_x^*$ and $\lambda_+ = \|\nabla f(x_+) - t_+\nabla f(x_0)\|_{x_+}^*$. Clearly, $\lambda_0 \le \frac{\beta}{M_f}$. If $t_+ > 0$, we have

$$\lambda_1 = \left\| \nabla f(x) - t\nabla f(x_0) + \frac{\gamma}{M_f\|\nabla f(x_0)\|_x^*}\nabla f(x_0) \right\|_x^* \le \frac{\beta + |\gamma|}{M_f}.$$

If $t_+ = 0$, we have that $t \le \frac{\gamma}{M_f\|\nabla f(x_0)\|_x^*}$. Hence,

$$\lambda_1 = \|\nabla f(x)\|_x^* = \|\nabla f(x) - t\nabla f(x_0) + t\nabla f(x_0)\|_x^* \le \frac{\beta + |\gamma|}{M_f}.$$

Since x_+ is obtained from x as the Standard Newton Step for the function $f_{t_+}(x)$, by (19), $\lambda_+ \le M_f \left(\frac{\lambda_1}{1-M_f\lambda_1}\right)^2$. The statement of the theorem follows from the fact that inequality $M_f\left(\frac{\lambda_1}{1-M_f\lambda_1}\right)^2 \le \frac{\beta}{M_f}$ is equivalent to inequality $\lambda_1 \le \frac{1}{M_f}\frac{\sqrt{\beta}}{1+\sqrt{\beta}}$. □

Remark 1 If at some iterate we obtain that $t_+ = 0$, then t is not updated in the later iterates and the algorithm automatically switches to the standard full-step Newton method: $x_+ = x - [\nabla^2 f(x)]^{-1} \nabla f(x)$. According to (19), for this method it holds that

$$\lambda_f(x_+) \leq \frac{M_f \lambda_f^2(x)}{(1 - M_f \lambda_f(x))^2}. \tag{34}$$

From Theorem 4, we know that, if $t_+ = 0$, then the approximate centering condition (31) holds at (x_+, t_+), which means that, since $t_+ = 0$,

$$\lambda_f(x_+) = \|\nabla f(x_+)\|_{x_+}^* \leq \frac{\beta}{M_f}. \tag{35}$$

Combining (34), (35), and our choice of the value β in (30), we see that the point x_+ belongs to the region of quadratic convergence of the standard Newton method. Thus, when $t_+ = 0$, the PFS automatically switches to the quadratically-convergent Newton method.

Let us prove the first main result of this section that gives convergence rate for the penalty parameter t_k in the PFS as applied to problem (2) when the region of quadratic convergence $\mathbb{Q}$ defined in (25) is not yet reached. The main novelty of this result is that we rely only on the assumption that f is a self-concordant function rather than a self-concordant barrier as in [34].

Theorem 5 *Consider the path-following scheme (PFS):*

$$t_0 = 1, \ x_0 \in \mathbb{E}, \quad (t_{k+1}, x_{k+1}) = \mathcal{P}(t_k, x_k), \quad k \geq 0, \tag{36}$$

where $\mathcal{P}$ is defined in (32). Assume that $\lambda_f(x_k) \geq \frac{1}{2M_f}$ for all $k = 0, \ldots, N$. Then,

$$t_N \leq \exp\left\{-\frac{\gamma(\gamma - 2\beta)N^2}{2M_f^2(f(x_0) - f^*)}\right\}. \tag{37}$$

Proof Denote $c = -\nabla f(x_0)$. Then,

$$t_{k+1} \overset{(32)}{=} t_k - \frac{\gamma}{M_f \|c\|_{x_k}^*} = t_k \left(1 - \frac{\gamma}{M_f t_k \|c\|_{x_k}^*}\right) \leq t_k \exp\left\{-\frac{\gamma}{M_f t_k \|c\|_{x_k}^*}\right\}.$$

Thus, $t_N \leq \exp\left\{-\frac{\gamma}{M_f} S_N\right\}$, where $S_N = \sum\limits_{k=0}^{N} \frac{1}{t_k \|c\|_{x_k}^*}$.

Let us estimate the value S_N from below. Note that

$$x_k - x_{k+1} \overset{(32)}{=} [\nabla^2 f(x_k)]^{-1} \left(t_k c + \nabla f(x_k) - \frac{\gamma c}{M_f \|c\|_{x_k}^*}\right). \tag{38}$$

Therefore,

$$r_k \overset{\text{def}}{=} \|x_k - x_{k+1}\|_{x_k} \overset{(31)}{\leq} \frac{\beta + \gamma}{M_f}. \tag{39}$$

On the other hand, $\frac{\beta^2}{M_f^2} \overset{(31)}{\geq} \lambda_f^2(x_k) + 2t_k \langle \nabla f(x_k), [\nabla^2 f(x_k)]^{-1} c\rangle + t_k^2 (\|c\|_{x_k}^*)^2$. Hence,

$$-\langle \nabla f(x_k), [\nabla^2 f(x_k)]^{-1} c\rangle \geq \frac{1}{2t_k}\left[\lambda_f^2(x_k) + t_k^2(\|c\|_{x_k}^*)^2 - \frac{\beta^2}{M_f^2}\right]. \tag{40}$$

Therefore, denoting $\lambda_k = \|\nabla f(x_k) - t_k \nabla f(x_0)\|_{x_k}^*$,

$$\begin{aligned}
f(x_k) - f(x_{k+1}) &\overset{(8)}{\geq} \langle \nabla f(x_k), x_k - x_{k+1}\rangle - \tfrac{1}{M_f^2}\omega_*(M_f r_k) \\
&\overset{(38)}{=} \langle \nabla f(x_k), [\nabla^2 f(x_k)]^{-1}\left(t_k c + \nabla f(x_k) - \tfrac{\gamma c}{M_f \|c\|_{x_k}^*}\right)\rangle - \tfrac{1}{M_f^2}\omega_*(M_f r_k) \\
&= \lambda_k^2 - t_k \langle c, [\nabla^2 f(x_k)]^{-1}(t_k c + \nabla f(x_k))\rangle + \langle \nabla f(x_k), [\nabla^2 f(x_k)]^{-1}\left(\tfrac{-\gamma c}{M_f\|c\|_{x_k}^*}\right)\rangle \\
&\quad - \tfrac{1}{M_f^2}\omega_*(M_f r_k) \\
&\geq \lambda_k^2 - t_k \|c\|_{x_k}^* \lambda_k - \tfrac{\gamma}{M_f\|c\|_{x_k}^*}\langle \nabla f(x_k), [\nabla^2 f(x_k)]^{-1} c\rangle - \tfrac{1}{M_f^2}\omega_*(M_f r_k) \\
&\overset{(40)}{\geq} \lambda_k^2 - t_k\|c\|_{x_k}^*\lambda_k + \tfrac{\gamma}{2M_f t_k \|c\|_{x_k}^*}\left[\lambda_f^2(x_k) + t_k^2(\|c\|_{x_k}^*)^2 - \tfrac{\beta^2}{M_f^2}\right] - \tfrac{1}{M_f^2}\omega_*(M_f r_k) \\
&\overset{(39)}{\geq} \tfrac{\gamma - 2M_f\lambda_k}{2M_f} t_k \|c\|_{x_k}^* + \rho_k \overset{(31)}{\geq} \tfrac{\gamma - 2\beta}{2M_f} t_k\|c\|_{x_k}^* + \rho_k,
\end{aligned} \tag{41}$$

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

Our next goal is to show that $\rho_k \geq 0$. Note that $t_k\|c\|_{x_k}^* \overset{(31)}{\leq} \lambda_f(x_k) + \frac{\beta}{M_f}$. Since $\lambda_f(x_k) \geq \frac{1}{2M_f}$, we have

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

Using the values (30), by direct computation we can see that the right-hand side of this inequality is positive.

Thus, we have proved that $f(x_k) - f(x_{k+1}) \geq \frac{\gamma - 2\beta}{2M_f} t_k \|c\|_{x_k}^*$. Therefore,

$$\begin{aligned}
S_N &\geq \sum_{k=0}^{N} \frac{\gamma - 2\beta}{2M_f(f(x_k) - f(x_{k+1}))} \\
&\geq \frac{\gamma - 2\beta}{2M_f} \min_{\tau \in \mathbb{R}_+^{N+1}} \left\{ \sum_{i=1}^{N+1} \frac{1}{\tau^{(i)}} : \sum_{i=1}^{N+1} \tau^{(i)} = f(x_0) - f(x_{N+1}) \right\} \\
&= \frac{(\gamma - 2\beta)(N+1)^2}{2M_f(f(x_0) - f(x_{N+1}))} \geq \frac{(\gamma-2\beta)(N+1)^2}{2M_f(f(x_0) - f^*)}.
\end{aligned} \tag{42}$$

□

Remark 2 The bound (37) can be slightly improved to give a superlinear convergence. Indeed, on the one hand, we have

$$t_{k+1} \overset{(32)}{=} t_k - \frac{\gamma}{M_f \|c\|^*_{x_k}} = t_k \left(1 - \frac{\gamma}{M_f t_k \|c\|^*_{x_k}}\right).$$

On the other hand, as we showed in the proof of Theorem 5, $f(x_k) - f(x_{k+1}) \geq \frac{\gamma-2\beta}{2M_f} t_k \|c\|^*_{x_k}$. Whence, if $f(x_k) - f(x_{k+1}) \leq \frac{\gamma(\gamma-2\beta)}{2M_f^2}$, we have $\frac{\gamma(\gamma-2\beta)}{2M_f^2} \geq \frac{\gamma-2\beta}{2M_f} t_k \|c\|^*_{x_k}$, and $t_k \leq \frac{\gamma}{M_f \|c\|^*_{x_k}}$. Thus, recalling that $c = -\nabla f(x_0)$, from (32), we have that $t_{k+1} = 0$, and, according to Remark 1, the PFS automatically switches to the quadratically-convergent full-step Newton method. Thus, since we are interested in the complexity of reaching the region of quadratic convergence, we assume that for $k = 0, ..., N$, $f(x_k) - f(x_{k+1}) > \frac{\gamma(\gamma-2\beta)}{2M_f^2}$. Then, for $k = 0, ..., N$ we have

$$t_{k+1} = t_k \left(1 - \frac{\gamma}{M_f t_k \|c\|^*_{x_k}}\right) \leq t_k \left(1 - \frac{\gamma(\gamma-2\beta)}{2M_f^2 (f(x_k) - f(x_{k+1}))}\right).$$

Thus, denoting $\tau = \frac{\gamma(\gamma-2\beta)}{2M_f^2}$, $\Delta_k = f(x_k) - f(x_{k+1})$, we obtain

$$\ln t_N \leq \sum_{k=0}^{N} \ln\left(1 - \frac{\tau}{\Delta_k}\right) \leq \max_{\Delta_k \geq 0: \sum_{k=0}^{N} \Delta_k = f(x_0) - f(x_{N+1})} \sum_{k=0}^{N} \ln\left(1 - \frac{\tau}{\Delta_k}\right).$$

The latter is a concave maximization problem. By symmetry, the minimum is achieved when $\Delta_k = \frac{f(x_0)-f(x_{N+1})}{N+1} \leq \frac{f(x_0)-f(x^*)}{N+1}$. So, we obtain

$$t_N \leq \left(1 - \frac{\gamma(\gamma-2\beta)(N+1)}{2M_f^2 (f(x_0) - f(x^*))}\right)^{N+1}, \tag{43}$$

which holds while $N + 1 < \frac{f(x_0)-f(x^*)}{\tau}$. Thus, we obtain that the PFS has *global superlinear convergence*. We note also that the closer $f(x_0)$ to $f(x^*)$, the faster the algorithm converges. This is reasonable since if $f(x_0)$ is close to $f(x^*)$, the point x_0 is close to the region of local quadratic convergence.

Let us now estimate the number of iterations, which is sufficient for method (36) to enter the region of quadratic convergence $\mathbb{Q}$ defined in (25). This is the second main result of this section. Denote

$$D = \max_{x,y \in \mathrm{dom} f} \{\|x - y\|_{x_0} : \; f(x) \leq f(x_0), \; f(y) \leq f(x_0)\}.$$

Theorem 6 *Let sequence $\{x_k\}_{k\geq 0}$ be generated by the method (36). Then, if*

$$N \geq \left[\frac{2\Delta(x_0)}{\gamma(\gamma-2\beta)} \ln \frac{M_f D \omega^{-1}(\Delta(x_0))}{\omega\left(\frac{(1-\beta)(1-2\beta)}{2}\right)} \right]^{1/2} \tag{44}$$

we have $x_N \in \mathbb{Q}$.

Proof Indeed,

$$f(x(t_k)) - f^* \leq \langle \nabla f(x(t_k)), x(t_k) - x^* \rangle \overset{(28)}{=} t_k \langle \nabla f(x_0), x(t_k) - x^* \rangle \leq t_k \lambda_f(x_0) D,$$

where we used that $f(x_{k+1}) \leq f(x_k)$, $k \geq 0$, see (41). Since $\omega(M_f \lambda_f(x_0)) \overset{(21)}{\leq} M_f^2(f(x_0) - f^*)$, we have

$$\frac{1}{M_f^2}\omega(M_f \lambda_f(x(t_k))) \overset{(21)}{\leq} f(x(t_k)) - f^* \leq \frac{t_k}{M_f}\omega^{-1}(\Delta(x_0))D.$$

Note that $\|\nabla f(x_k) - \nabla f(x(t_k))\|_{x_k}^* \overset{(28)}{=} \|\nabla f(x_k) - t_k \nabla f(x_0)\|_{x_k}^* \overset{(31)}{\leq} \frac{\beta}{M_f} < \frac{1}{M_f}$. Therefore,

$$\begin{aligned}\lambda_f(x_k) &\overset{(31)}{\leq} t_k \|\nabla f(x_0)\|_{x_k}^* + \frac{\beta}{M_f} \overset{(28)}{=} \langle \nabla f(x(t_k)), [\nabla^2 f(x_k)]^{-1} \nabla f(x(t_k)) \rangle^{\frac{1}{2}} + \frac{\beta}{M_f} \\ &\overset{(12)}{\leq} \frac{1}{1-\beta}\lambda_f(x(t_k)) + \frac{\beta}{M_f}.\end{aligned}$$

Thus, inclusion $x_k \in \mathbb{Q}$, is ensured by inequality $\lambda_f(x(t_k)) \leq \frac{(1-\beta)(1-2\beta)}{2M_f}$. Consequently, we need

$$\frac{t_k}{M_f}\omega^{-1}(\Delta(x_0))D \leq \frac{1}{M_f^2}\omega\left(\frac{(1-\beta)(1-2\beta)}{2}\right) \tag{45}$$

It remains to use inequality (37). □

As we can see from the estimate (44) for the global complexity, up to a logarithmic factor, the number of iterations of the PFS is proportional to $\Delta^{1/2}(x_0)$, where $\Delta(x_0)$ is defined in (26). This is much better than the guarantee (26) for the DNM (24). However, as we will see in Section 7, for some special subclasses of self-concordant functions the performance estimate (44) can be significantly improved.

Remark 3 Note that, in the complexity bound (44), the constant $\left[\frac{2}{\gamma(\gamma-2\beta)}\right]^{1/2} \leq 17.1$. The choice of the parameters β and γ is governed by the following aspects. First, from Theorem 4, these parameters should satisfy (33). Second, ρ_k in the proof of Theorem 5 should be non-negative. Third, the complexity in (44) is proportional to $(\gamma(\gamma-2\beta))^{-1/2}$, which is desired to be as small as it is possible. This motivates the following maximization problem for optimal choice of β, γ.

$$\max \gamma(\gamma - 2\beta) \ \text{ s.t.} \tag{46}$$

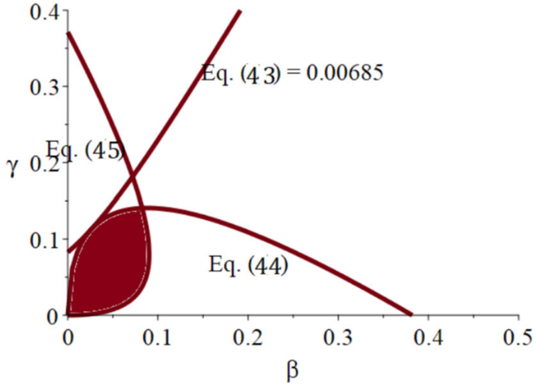

Fig. 1 Optimal choice of β and γ. The feasible set given by (47) and (48) is filled with red color

$$\frac{\sqrt{\beta}}{1+\sqrt{\beta}} - \beta - \gamma \geq 0 \tag{47}$$

$$\frac{\gamma(1-2\beta)}{4} - \omega_*(\beta+\gamma) \geq 0. \tag{48}$$

Figure 1 illustrates this optimization problem and the optimal objective value.

The above results prescribe specific values for the accuracy of following the central path β and the stepsize γ. Nevertheless, if we use a larger stepsize γ and after the Newton step the approximate centering condition holds, we can continue to follow the path. This leads to a PFS with adaptive choice of the stepsize γ outlined below.

Adaptive Path-Following Scheme
– Set an initial point x_0, initial value of the penalty parameter $t_0 = 1$, initial stepsize value $\gamma_{-1} \geq 0.1125$. – k-th iteration. Find the minimum value $i_k \geq 0$ s.t. the path-following step (32) with the stepsize $\gamma = 2^{1-i_k}\gamma_{k-1}$ outputs a point satisfying approximate centering condition (31). – Set $\gamma_k = 2^{1-i_k}\gamma_{k-1}$, $x_{k+1} = x_+$, $t_{k+1} = t_+$.

By Theorem 4, there exists $\hat{\gamma}$ s.t. the Newton step after the update of the parameter t with the stepsize $\hat{\gamma}$ outputs a point satisfying the approximate centering condition. Hence, the search for i_k is finite and $\gamma_k = 2^{1-i_k}\gamma_{k-1} \geq \frac{\hat{\gamma}}{2}$. Hence, the total number of Newton steps can be estimated as follows

$$\sum_{j=0}^{k} i_k = \sum_{j=0}^{k}\left(1 + \log_2 \frac{\gamma_{j-1}}{\gamma_j}\right) = k + 1 + \log_2 \frac{\gamma_{-1}}{\gamma_k} \leq k + \log_2 \frac{4\gamma_{-1}}{\hat{\gamma}}.$$

As we see, the price for the adaptivity is reasonable, taking into account that the practical performance of the adaptive algorithm is expected to be better since the penalty parameter t potentially decreases faster.

Interestingly, increasing the stepsize γ looks equivalent to decreasing the constant M_f. We underline that we assume the constant M_f to be fixed and known. If the method would be adaptive also to the constant M_f, then the local estimate M_f should be increased also in the inequality of the approximate centering condition. Also, note that for a fixed and known constant M_f the stepsize for the DNM is optimal since it is obtained by the minimization of the upper bound for the self-concordant function. Thus, adaptive stepsize for this method does not make much sense for a fixed and known constant M_f.

Remark 4 After the first version of this paper appeared as the preprint [14], we were made aware of a concurrent work [42]. In particular, the latter paper considers path-following methods for functions involving general self-concordant functions that are not barriers. At the same time there are several important differences between our work and theirs. They consider a different to our problem (2) composite optimization problem of the form

$$\min_{x \in \mathbb{E}} f(x) + g(x),$$

where f is a self-concordant function and g is a simple closed convex function. Their algorithms and results are very specific to that problem and do not allow to obtain our algorithms and results as a particular case with $g = 0$. First, their central path is defined by the inclusion

$$0 \in t\nabla f(x(t)) - (1-t)g'(x_0) + \partial g(x(t)),$$

where $t \in [0, 1]$, $\partial g(x)$ denotes the subdifferential of g, $g'(x_0) \in \partial g(x_0)$. If $g = 0$, then $g'(x_0) = 0$ and the above equation, for any $t \in [0, 1]$, is equivalent to the optimality condition for the problem (2) and thus, following the path approximately does not make sense since for each $t \in [0, 1]$ approximating $x(t)$ is equivalent to solving the original problem. In the same way, their algorithm is different to ours and when $g = 0$, their main algorithmic step (12) is just the Standard Newton Step (18) that may diverge and has only local convergence. Thus, for their results it is essential that $g \neq 0$ and our results do not follow from theirs as a particular case. Finally, they prove a linear convergence for their scheme, whereas we prove global superlinear convergence with explicit rate. We can say that their results and ours are complementary to each other. Finally, in the next sections we propose also a new predictor-corrector scheme, which they do not consider.

z

4 Path-following scheme: implications for a feasibility problem

In this section, we first recall some properties of an important subclass of self-concordant functions, ν-self-concordant barriers. These properties are then used to compare complexity of different methods applied to a feasibility problem which is solved by the minimization of a barrier on an affine subspace. Importantly, the proposed above PFS gives the best complexity, improving upon known complexities. Moreover, the obtained complexity corresponds to the *global superlinear convergence* of the proposed scheme.

4.1 Preliminaries on self-concordant barriers

In this subsection, we recall the definition of self-concordant barriers and state their properties necessary to obtain the results of this section.

Definition 2 Let F be a standard self-concordant function. We call it a ν-self-concordant barrier for the set $\mathrm{Dom} F$ iff

$$\sup_{y\in\mathbb{E}}[2\langle\nabla F(x), y\rangle - \langle\nabla^2 F(x)y, y\rangle] \leq \nu, \tag{49}$$

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

Note that, if $\nabla^2 F(x)$ is non-degenerate, inequality (49) is equivalent to

$$\langle[\nabla^2 F(x)]^{-1}\nabla F(x), \nabla F(x)\rangle \leq \nu. \tag{50}$$

The above inequality together with the duality relations (13) and (14) imply

$$\langle u, \nabla^2 F_*(u)u\rangle \leq \nu, \quad u \in \mathrm{dom}\, F_*. \tag{51}$$

Theorem 7 *[Theorem 4.2.4 in [34]] 1. Let $F(x)$ be a ν-self-concordant barrier. Then, for any $x \in \mathrm{dom}\, F$ and $y \in \mathrm{Dom} F$, we have*

$$\langle\nabla F(x), y - x\rangle \leq \nu. \tag{52}$$

2. A standard self-concordant function F is a ν-self-concordant barrier iff

$$F(y) \geq F(x) - \nu \ln\left(1 - \frac{1}{\nu}\langle\nabla F(x), y - x\rangle\right). \tag{53}$$

4.2 Feasibility problem

Let us consider the following feasibility problem

$$\text{Find } x \text{ s.t. } x \in Q \text{ and } Ax = b, \tag{54}$$

where $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$. We assume that $Q \subset \mathbb{R}^n$ is a closed and convex set and $0 \in \mathrm{int}\, Q$. To solve this problem, we introduce a ν-self-concordant *barrier*

$F(x)$ for the set Q and minimize it over the affine manifold:

$$\min_x F(x) \quad \text{s.t.} \quad Ax = b. \tag{55}$$

Without loss of generality, we assume that 0 is the analytic center, i.e. $\nabla F(0) = 0$, and that $F(0) = 0$. Denote by x^* a solution of (55). Then, clearly, x^* solves the feasibility problem. We also assume that the *feasibility depth* ε satisfies

$$\varepsilon \stackrel{\text{def}}{=} \max\{\delta \geq 0 \ : \ (1-\delta)Q \cap \{x : Ax = b\} \neq \emptyset\} > 0. \tag{56}$$

Lemma 3 *Under the above assumptions, we have that*

$$\frac{1}{\varepsilon} \geq 1 + \frac{1}{\nu}\langle \nabla F(x^*), x^* \rangle, \tag{57}$$

$$F(x^*) - F(0) \leq \nu \ln \frac{1}{\varepsilon}. \tag{58}$$

Proof By the definition of the feasibility depth ε, there exists $\hat{x} \in (1-\varepsilon)Q$ such that $A\hat{x} = b$. Since x^* is a solution to (55), there exists a Lagrange multiplier y^* such that $\nabla F(x^*) = A^T y^*$. Thus, for any x satisfying $Ax = b$, we have

$$\langle \nabla F(x^*), x \rangle = \langle A^T y^*, x \rangle = \langle y^*, Ax \rangle = \langle y^*, b \rangle. \tag{59}$$

Whence, $\langle \nabla F(x^*), \hat{x} \rangle = \langle \nabla F(x^*), x^* \rangle$. Thus, since $\hat{x} \in (1-\varepsilon)Q$, we have

$$\begin{aligned}
\langle \nabla F(x^*), x^* \rangle = \langle \nabla F(x^*), \hat{x} \rangle &\leq \max_{u \in (1-\varepsilon)Q} \langle \nabla F(x^*), u \rangle \\
&= (1-\varepsilon) \max_{u \in Q} \{\langle \nabla F(x^*), u - x^* \rangle + \langle \nabla F(x^*), x^* \rangle\} \\
&\stackrel{(52)}{\leq} (1-\varepsilon)(\nu + \langle \nabla F(x^*), x^* \rangle).
\end{aligned}$$

Thus, we have

$$\varepsilon \langle \nabla F(x^*), x^* \rangle \leq (1-\varepsilon)\nu \Rightarrow \frac{1}{\nu}\langle \nabla F(x^*), x^* \rangle \leq \frac{1}{\varepsilon} - 1,$$

which is (57). Applying (53) with $y = 0$ and $x = x^*$, we obtain, by (57),

$$F(0) \geq F(x^*) - \nu \ln\left(1 + \frac{1}{\nu}\langle \nabla F(x^*), x^* \rangle\right) \geq F(x^*) - \nu \ln \frac{1}{\varepsilon},$$

which is (58). □

Let us now consider the dual problem to (55). Introducing a Lagrange multiplier y for the constraints $Ax = b$, we obtain

$$\begin{aligned} &\min_x \{F(x) : \quad Ax = b\} = \min_x \max_y \{F(x) + \langle y, b - Ax\rangle\} \\ &= \quad \max_y \langle b, y\rangle + \min_x\{F(x) - \langle A^T y, x\rangle\}\} = \max_y\{\langle b, y\rangle - F_*(A^T y)\} \end{aligned}$$

and the dual problem for (55) is

$$\min_y \{\Phi(y) - \langle b, y\rangle\}, \tag{60}$$

where $\Phi(y) \stackrel{\text{def}}{=} F_*(A^T y)$, F_* being the Fenchel conjugate for F. We also have $\Phi(y^*) - \langle b, y^*\rangle = -F(x^*)$, where y^* is the solution of the dual problem (60). Note that $\Phi(y)$ is a standard self-concordant function and is not a self-concordant barrier. Yet, it has a useful property (51) which allows to minimize it with the complexity standard for self-concordant barriers.

Assume now that we solve the dual problem (60) starting from $y = 0$. Denote also $\widetilde{\Phi}(y) = \Phi(y) - \langle b, y\rangle$. Then, since $\widetilde{\Phi}(0) = \min_x F(x) = F(0) = 0$,

$$\widetilde{\Phi}(0) - \widetilde{\Phi}^* = -\widetilde{\Phi}^* = F(x^*) = F(x^*) - F(0) \stackrel{(58)}{\leq} \nu \ln \frac{1}{\varepsilon}. \tag{61}$$

We can apply the DNM (24) and the PFS (36) to solve the dual problem (60).

Let us also consider dual path-following scheme [35]. To do so, consider the central path of a parametric family of problems

$$y_\sigma = \arg\min_y\{\Phi(y) : \sigma = \langle b, y\rangle\}, \quad \sigma \geq 0. \tag{62}$$

Since $\nabla\Phi(0) = A(\arg\min_x F(x)) = 0$, we start with $y_0 = 0$ and $\sigma_0 = 0$, and the goal is to follow the path by increasing σ until $\sigma = \sigma^* = \langle b, y^*\rangle$. Using (59) and (57), we obtain

$$\sigma^* = \langle b, y^*\rangle = \langle \nabla F(x^*), x^*\rangle \leq \nu\left(\frac{1}{\varepsilon} - 1\right) \leq \frac{\nu}{\varepsilon}. \tag{63}$$

The Hessian of Φ induces at y local norm $\|\cdot\|_y$ and its conjugate $\|\cdot\|_y^*$. For the sake of simplicity, let us assume that we can follow the path (62) exactly. Namely, we use the following procedure

$$(\sigma_+, y_+) \; = \; \mathcal{DP}(\sigma, y) \; \stackrel{\text{def}}{=} \; \begin{cases} \sigma_+ = \sigma + \gamma\|b\|_y^*, \\ y_+ = \arg\min_z\{\Phi(z) : \sigma_+ = \langle b, z\rangle\}. \end{cases} \tag{64}$$

and iterate until $\sigma \geq \sigma^*$. We have

$$\langle b, y_+\rangle = \sigma_+ = \sigma + \gamma\|b\|_y^* = \langle b, y\rangle + \gamma\|b\|_y^* = \langle b, y\rangle\left(1 + \frac{\gamma\|b\|_y^*}{\langle b, y\rangle}\right)$$
$$\geq \langle b, y\rangle\left(1 + \frac{\gamma}{\|y\|_y}\right) \geq \langle b, y\rangle\left(1 + \frac{\gamma}{\sqrt{\nu}}\right) = \sigma\left(1 + \frac{\gamma}{\sqrt{\nu}}\right),$$

where we used that Φ is Fenchel conjugate for ν-self-concordant barrier F and, by (51), $\|y\|_y = \langle\nabla^2\Phi(y)y, y\rangle^{1/2} \leq \sqrt{\nu}$ for all y. Hence, for the iterates $(\sigma_{k+1}, y_{k+1}) = \mathcal{DP}(\sigma_k, y_k)$ it holds that $\sigma_k \geq \sigma_1\left(1 + \frac{\gamma}{\sqrt{\nu}}\right)^k = \gamma\|b\|_{y_0}^*\left(1 + \frac{\gamma}{\sqrt{\nu}}\right)^k$ and, by (63), the complexity to get $\sigma_k \geq \sigma^*$ is bounded as $O\left(\frac{\sqrt{\nu}}{\gamma}\ln\frac{\nu}{\gamma\varepsilon\|b\|_{y_0}^*}\right)$.

To sum up, we can apply three different strategies to solve problem (55) by solving its dual (60).

1. Damped Newton Method (24). In this case, we have that the complexity is given by (26), i.e. is $O\left(\widetilde{\Phi}(0) - \widetilde{\Phi}^*\right) = O\left(\nu\ln\frac{1}{\varepsilon}\right)$.
2. Path-following scheme (36). In this case, we have that the complexity is given by (44), i.e. is $O\left(\sqrt{\widetilde{\Phi}(0) - \widetilde{\Phi}^*}\right) = O\left(\sqrt{\nu\ln\frac{1}{\varepsilon}}\right)$.
3. Dual path-following scheme (64) has the complexity $O\left(\sqrt{\nu}\ln\frac{\nu}{\varepsilon}\right)$.

Interestingly, the complexity $\sqrt{\nu\ln\frac{1}{\varepsilon}}$ corresponds to the superlinear convergence of the order $\exp\left(-\left(\frac{k}{\sqrt{\nu}}\right)^2\right)$, which is faster than the linear convergence $\exp\left(-\frac{k}{\sqrt{\nu}}\right)$ corresponding to the complexity $\sqrt{\nu}\ln\frac{1}{\varepsilon}$ that is typical for path-following methods. What is even more surprising is that we use the short-step PFS which was for a long time believed to be inferior to long-step methods having complexity $\sqrt{\nu}\ln\frac{1}{\varepsilon}$. Moreover to get the complexity $\sqrt{\nu\ln\frac{1}{\varepsilon}}$ we use the barrier property only to estimate the initial objective residual. We do not use the barrier property in the analysis of the method, only the self-concordance property is used.

Let us show that, actually, the feasibility problem (54) is very general. In particular, one can reduce a linear program in standard form to this feasibility problem. The primal-dual pair of linear programs corresponding to a linear program in standard form is

$$\min_x\{\langle c, x\rangle : Ax = b, x \geq 0\} = \max_{s,y}\{\langle b, y\rangle : s + A^T y = c, s \geq 0\}, \tag{65}$$

where $x, s, c \in \mathbb{R}^n$, $y, b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$. If the matrix A has full row rank m, without loss of generality, we can assume that the matrix A has the form (I_m, B), where I_m is the identity matrix in $\mathbb{R}^{m\times m}$. Then, the vector c can be divided in two blocks (c_1, c_2) s.t. $c_1 \in \mathbb{R}^m$, and, similarly, for the vector $s = (s_1, s_2)$. Moreover, from the equation $s + A^T y = c$, we have $s_1 + y = c_1$, $s_2 + B^T y = c_2$. Hence, we have that the problem $\max_{s,y}\{\langle b, y\rangle : s + A^T y = c, s \geq 0\}$ is equivalent to $\max_s\{\langle b, c_1 - s_1\rangle : s_2 - B^T s_1 = c_2 - B^T c_1, s \geq 0\}$. By the strong duality, we have that solving this primal-dual pair of problems is equivalent to solving the system of equations and inequalities

$$\langle c, x\rangle - \langle b, c_1 - s_1\rangle = 0, \; Ax = b, \; s_2 - B^T s_1 = c_2 - B^T c_1, \; x \geq 0, \; s \geq 0.$$

Introducing a new variable $\tau \geq 0$, we have that the above system is equivalent to the following homogenous system of linear inequalities and equations

$$\langle c, x\rangle + \langle b, s_1\rangle - \tau\langle b, c_1\rangle = 0, \; Ax = \tau b, \; s_2 - B^T s_1 = \tau(c_2 - B^T c_1), \; x, s, \tau \geq 0.$$

Thus, we obtain that solving the primal-dual pair of linear programs is equivalent to solving the system $z \in \mathbb{R}^q, \;\; Qz = 0, \;\; z \geq 0$, where $z = (x, s, \tau)$ and the matrix Q is composed of all the data in the previous system of equations (sometimes this representation is called *homogeneous self-dual embedding*, see [29, 37, 43] and references therein). This system has a trivial solution $z = 0$. To find a non-trivial solution, we add a constraint that z belongs to the standard simplex $S_q(1) = \{z \in \mathbb{R}^q : z \geq 0, \langle e_q, z\rangle = 1\}$, where e_q is the vector of all ones. This gives us a feasibility problem of finding z s.t. $z \in S_q(1)$ and $Qz = 0$. The only issue of this reformulation is that the standard simplex has empty interior. This can be easily resolved by transition to a $(q-1)$-dimensional problem. Indeed, by the simplex constraint, $z_q = 1 - \sum_{i=1}^{q-1} z_i \geq 0$. Thus, if we introduce the variable $\bar{z}$ which consists of the first $q-1$ components of z, matrix $\bar{Q}$ which consists of the first $q-1$ columns of Q, vector $\bar{q}$ which is the last column of Q, we have that the considered feasibility problem is equivalent to $\bar{z} \geq 0, \; \langle e_{q-1}, \bar{z}\rangle \leq 1, \; (\bar{Q} - \bar{q}e_{q-1}^T)\bar{z} = -\bar{q}$.

As we see, linear programming problem in the standard form (65) can be reformulated as the feasibility problem (54). Thus, the obtained above complexity bounds for the feasibility problem apply also to linear programming problems.

5 Minimizing self-concordant functions: predictor-corrector path-following scheme

In this section, we return back to problem (2) and propose an alternative algorithm, which we call predictor-corrector path-following scheme (PCPFS). First, we prove an important technical lemma that the central path is sufficiently smooth. This allows us to use longer predictor steps in approximate tangent direction to the path, which is then followed by a corrector step to guarantee approximate centering condition. Notably, due to longer predictor step, we obtain better constant in the complexity bounds compared to the constant for the PFS (36).

5.1 Main lemma

Given a fixed vector c, we consider the trajectory of minimizers of the following parametric family of general self-concordant functions:

$$x(t) = \arg\min_{x \in \mathbb{E}} \left\{ f_t(x) \stackrel{\mathrm{def}}{=} f(x) + t\langle c, x\rangle \right\}, \quad t \in \mathbb{R}. \tag{66}$$

Recall that $\lambda_{f_t}(x) \equiv \|\nabla f(x) + tc\|_x^*$, $t \in \mathbb{R}$.

The following lemma is the main result of this subsection and it will be used to justify our predictor-corrector path-following scheme (PCPFS).

Lemma 4 *Let f be an M_f-self-concordant function. Define $h = [\nabla^2 f(x)]^{-1}c$ and $r = \|h\|_x = \|c\|_x^*$. Then, for all τ such that $|\tau| M_f \|h\|_x < 1$,*

$$\lambda_{f_{t-\tau}}(x + \tau h) \leq \lambda_{f_t}(x)\left(1 + \frac{\tau M_f r}{1 - \tau M_f r}\right) + \frac{1}{M_f}\left(\frac{\tau M_f r}{1 - \tau M_f r}\right)^2. \tag{67}$$

Proof Define $g_\tau = \nabla f(x + \tau h) + (t - \tau)c$ and $H_\tau = \nabla^2 f(x + \tau h)$. Then, for

$$\begin{aligned}
\xi(\tau) &\stackrel{\mathrm{def}}{=} (\lambda_{f_{t-\tau}}(x + \tau h))^2 = \langle g_\tau, H_\tau^{-1} g_\tau\rangle \\
&= \left\langle \nabla f(x + \tau h) + (t - \tau)c, [\nabla^2 f(x + \tau h)]^{-1}(\nabla f(x + \tau h) + (t - \tau)c)\right\rangle,
\end{aligned}$$

we have

$$\begin{aligned}
\xi'(\tau) &= -\left\langle g_\tau, H_\tau^{-1} D^3 f(x + \tau h)[h] H_\tau^{-1} g_\tau\right\rangle + 2\left\langle H_\tau^{-1}(H_\tau h - c), g_\tau\right\rangle \\
&\stackrel{(4)}{\leq} 2M_f \|h\|_{x+\tau h} \|H_\tau^{-1} g_\tau\|_{x+\tau h}^2 + 2\left\langle (H_0^{-1} - H_\tau^{-1})c, g_\tau\right\rangle \\
&= 2M_f \|h\|_{x+\tau h} \xi(\tau) + 2\left\langle (H_\tau^{\frac{1}{2}} H_0^{-1} H_\tau^{\frac{1}{2}} - I) H_\tau^{-\frac{1}{2}} c, H_\tau^{-\frac{1}{2}} g_\tau\right\rangle.
\end{aligned} \tag{68}$$

Let us estimate $\|h\|_{x+\tau h}$ and the second term in the r.h.s. Since $|\tau| M_f \|h\|_x < 1$,

$$\|h\|_{x+\tau h} = \frac{1}{\tau}\|(x + \tau h) - x\|_{x+\tau h} \stackrel{(7)}{\leq} \frac{1}{\tau}\frac{\|(x + \tau h) - x\|_x}{1 - M_f\|(x + \tau h) - x\|_x} = \frac{r}{1 - \tau M_f r}. \tag{69}$$

Further, since $|\tau| M_f \|h\|_x < 1$, from (6), we have

$$(1 - \tau M_f r)^2 H_0^{-1} \preceq H_\tau^{-1} \preceq \tfrac{1}{(1-\tau M_f r)^2} H_0^{-1}$$

$$(1 - \tau M_f r)^2 H_\tau^{\frac{1}{2}} H_0^{-1} H_\tau^{\frac{1}{2}} \preceq I \preceq \tfrac{1}{(1-\tau M_f r)^2} H_\tau^{\frac{1}{2}} H_0^{-1} H_\tau^{\frac{1}{2}}$$

$$(1 - \tau M_f r)^2 I \preceq H_\tau^{\frac{1}{2}} H_0^{-1} H_\tau^{\frac{1}{2}} \preceq \tfrac{1}{(1-\tau M_f r)^2} I$$

$$\left((1 - \tau M_f r)^2 - 1\right) I \preceq H_\tau^{\frac{1}{2}} H_0^{-1} H_\tau^{\frac{1}{2}} - I \preceq \left(\tfrac{1}{(1-\tau M_f r)^2} - 1\right) I.$$

Hence,

$$\left\langle (H_\tau^{\frac{1}{2}} H_0^{-1} H_\tau^{\frac{1}{2}} - I) H_\tau^{-\frac{1}{2}} c, H_\tau^{-\frac{1}{2}} g_\tau \right\rangle \leq \|(H_\tau^{\frac{1}{2}} H_0^{-1} H_\tau^{\frac{1}{2}} - I) H_\tau^{-\frac{1}{2}} c\| \|H_\tau^{-\frac{1}{2}} g_\tau\|$$
$$\leq \left(\tfrac{1}{(1-\tau M_f r)^2} - 1\right) \|H_\tau^{-\frac{1}{2}} c\| \sqrt{\xi(\tau)} \leq \left(\tfrac{1}{(1-\tau M_f r)^2} - 1\right) \tfrac{r}{1-\tau M_f r} \sqrt{\xi(\tau)}, \tag{70}$$

where we used that

$$\|H_\tau^{-\frac{1}{2}} c\|^2 = \left\langle H_\tau^{-\frac{1}{2}} c, H_\tau^{-\frac{1}{2}} c \right\rangle \overset{(6)}{\leq} \tfrac{1}{(1-\tau M_f r)^2} \left\langle c, H_0^{-1} c \right\rangle = \tfrac{r^2}{(1-\tau M_f r)^2}.$$

Combininig (68), (69) and (70), we obtain

$$\xi'(\tau) \leq \frac{2M_f r}{1-\tau M_f r} \xi(\tau) + \left(\frac{1}{(1-\tau M_f r)^2} - 1\right) \frac{2r}{1-\tau M_f r} \sqrt{\xi(\tau)}.$$

Using this inequality and denoting $\phi(\tau) = (1-\tau M_f r)^2 \xi(\tau)$, we obtain

$$\begin{aligned} \phi'(\tau) &= -2M_f r(1-\tau M_f r)\xi(\tau) + (1-\tau M_f r)^2 \xi'(\tau) \\ &= (1-\tau M_f r)^2 \left(\xi'(\tau) - \tfrac{2M_f r}{1-\tau M_f r}\xi(\tau)\right) \\ &\leq (1-\tau M_f r)^2 \left(\tfrac{1}{(1-\tau M_f r)^2} - 1\right) \tfrac{2r}{1-\tau M_f r} \sqrt{\xi(\tau)} \\ &= 2r \left(\tfrac{1}{(1-\tau M_f r)^2} - 1\right) \sqrt{\phi(\tau)}. \end{aligned}$$

Hence,

$$d\sqrt{\phi(\tau)} \leq \tfrac{1}{M_f} \left(\tfrac{1}{(1-\tau M_f r)^2} - 1\right) d(\tau M_f r).$$

Integrating both sides from 0 to τ, we obtain

$$\sqrt{\phi(\tau)} - \sqrt{\phi(0)} \leq \tfrac{1}{M_f} \left(\tfrac{1}{1-\tau M_f r} - \tau M_f r - 1\right) = \tfrac{\tau^2 M_f r^2}{1-\tau M_f r}$$

$$\text{and} \qquad (1-\tau M_f r)\sqrt{\xi(\tau)} - \sqrt{\xi(0)} \leq \tfrac{\tau^2 M_f r^2}{1-\tau M_f r},$$

which finishes the proof since $\sqrt{\xi(\tau)} = \lambda_{f_{t-\tau}}(x + \tau h)$. □

Let us make a remark on the meaning of the above Lemma. As we know from Section 3, the quantity $\lambda_{f_t}(x)$ shows how good we follow the central path and this quantity should be kept sufficiently small during the iterations. The above lemma shows how this quantity evolves when we update the penalty parameter t. Unlike the PFS (36), we will make simultaneously the update $t_+ = t - \tau$ and $y_+ = x + \tau h$ as the predictor step. The above lemma allows us to estimate the value $\lambda_{f_{t-\tau}}(x + \tau h)$ and choose τ such that the corrector Newton step for $f_{t-\tau}$ again guarantees the approximate centering condition, but for the updated function $f_{t-\tau}$. This construction allows us to decrease t faster leading to faster convergence.

5.2 Predictor-corrector path-following scheme

In this subsection, we describe our predictor-corrector path-following scheme (PCPFS) for solving problem (2). We again underline that we focus on the case when f is a general self-concordant function without the barrier property. In the next section we will also show that the PCPFS has implications for problems with the barrier property of f. As for the DNM and PFS above, our main goal in this section is to estimate the complexity to enter the region of quadratic convergence $\mathbb{Q}$ defined in (25).

For convenience, we recall from Section 3 the main objects important for our derivations. Namely, we start from some $x_0 \in \mathbb{E}$ and define the central path $x(t)$, $0 \le t \le 1$, by the following equation:

$$\nabla f(x(t)) = t\nabla f(x_0), \tag{71}$$

or, equivalently,

$$x(t) = \arg\min_{x\in\mathbb{E}} \left\{ f_t(x) \stackrel{\text{def}}{=} f(x) - t\langle \nabla f(x_0), x\rangle \right\}, \quad 0 \le t \le 1. \tag{72}$$

In this section, we change the definition of the main parameters for following the central path (cf. (30) where the stepsize γ is smaller)

$$\beta = 0.0015, \quad \gamma \; = \; 0.158. \tag{73}$$

As before, we say that point x satisfies an *approximate centering condition* if

$$\lambda_{f_t}(x) \stackrel{\text{def}}{=} \|\nabla f(x) - t\nabla f(x_0)\|_x^* \; \le \; \tfrac{\beta}{M_f}. \tag{74}$$

Instead of the PFS (36) we propose the predictor-corrector path-following iterate:

$$(t_+, x_+) \; = \; \mathcal{PC}(t,x) \; \stackrel{\text{def}}{=} \; \begin{cases} t_+ = \max\left\{ t - \tfrac{\gamma}{M_f\|\nabla f(x_0)\|_x^*}, 0\right\}, \\ y = x - \tfrac{\gamma}{M_f\|\nabla f(x_0)\|_x^*}[\nabla^2 f(x)]^{-1}\nabla f(x_0), \\ x_+ = y - [\nabla^2 f(y)]^{-1}(\nabla f(y) - t_+\nabla f(x_0)). \end{cases} \tag{75}$$

We refer to the y-step as the predictor and x_+-step as the corrector.

Theorem 8 *If the pair (x,t) satisfies (74), then the pair (x_+, t_+) satisfies (74) too.*

Proof Applying Lemma 4 with $c = -\nabla f(x_0)$, since, in this case, $r = \|h\|_x = \|\nabla f(x_0)\|_x^*$, we obtain

$$\lambda_{f_{t_+}}(y) \le \lambda_{f_t}(x)\left(1 + \frac{\gamma}{1-\gamma}\right) + \frac{1}{M_f}\left(\frac{\gamma}{1-\gamma}\right)^2 \le \frac{\beta}{M_f}\cdot\frac{1}{1-\gamma} + \frac{1}{M_f}\left(\frac{\gamma}{1-\gamma}\right)^2. \tag{76}$$

As x_+ is defined as the Standard Newton Step for f_{t_+} from y, by (19), we have

$$\lambda_{f_{t_+}}(x_+) \le \frac{1}{M_f}\left(\frac{M_f\lambda_{f_{t_+}}(y)}{1-M_f\lambda_{f_{t_+}}(y)}\right)^2 = \frac{1}{M_f}\left(\omega_*'(M_f\lambda_{f_{t_+}}(y))\right)^2.$$

By the choice of β and γ, we have $\omega_*'(M_f\lambda_{f_{t_+}}(y)) \le \omega_*'\left(\frac{\beta}{1-\gamma}+\left(\frac{\gamma}{1-\gamma}\right)^2\right) \le \sqrt{\beta}$, which finishes the proof. □

Let us prove the first main result of this subsection that gives convergence rate for the penalty parameter t_k in the PCPFS as applied to problem (2) when the region of quadratic convergence $\mathbb{Q}$ defined in (25) is not yet reached.

Theorem 9 *Consider the predictor-corrector path-following scheme (PCPFS):*

$$t_0 = 1,\ x_0 \in \mathbb{E}, \quad (t_{k+1}, x_{k+1}) = \mathcal{PC}(t_k, x_k), \quad k \ge 0, \tag{77}$$

where $\mathcal{PC}$ is defined in (75). Assume that $\lambda_f(x_k) \ge \frac{1}{2M_f}$ for all $k = 0, \ldots, N$. Then

$$t_N \le \exp\left\{-\frac{\kappa\gamma N^2}{M_f^2(f(x_0)-f^*)}\right\}, \quad \kappa = \frac{\gamma}{2} - \frac{\beta}{(1-\gamma)^2} - \frac{\gamma^2}{(1-\gamma)^3}. \tag{78}$$

Proof Denote $c = -\nabla f(x_0)$. In the same way as in the proof of Theorem 5, we obtain that $t_N \le \exp\left\{-\frac{\gamma}{M_f} S_N\right\}$, where $S_N = \sum\limits_{k=0}^{N} \frac{1}{t_k\|c\|_{x_k}^*}$.

Let us estimate the value S_N from below. Note that

$$\frac{\beta^2}{M_f^2} \overset{(74)}{\ge} \lambda_f^2(x_k) + 2t_k\langle \nabla f(x_k), [\nabla^2 f(x_k)]^{-1}c\rangle + t_k^2(\|c\|_{x_k}^*)^2.$$

Hence,

$$-\langle \nabla f(x_k), [\nabla^2 f(x_k)]^{-1}c\rangle \ge \frac{1}{2t_k}\left[\lambda_f^2(x_k) + t_k^2(\|c\|_{x_k}^*)^2 - \frac{\beta^2}{M_f^2}\right]. \tag{79}$$

Thus, we obtain

$$\begin{aligned} f(x_k) - f(y_k) &\overset{(8)}{\ge} \langle \nabla f(x_k), x_k - y_k\rangle - \frac{1}{M_f^2}\omega_*(M_f\|x_k - y_k\|_{x_k}) \\ &\overset{(75)}{=} -\frac{\gamma}{M_f\|c\|_{x_k}^*}\langle \nabla f(x_k), [\nabla^2 f(x_k)]^{-1}c\rangle - \frac{1}{M_f^2}\omega_*(\gamma) \\ &\overset{(79)}{\ge} \frac{\gamma}{2M_f t_k\|c\|_{x_k}^*}\left[\lambda_f^2(x_k) + t_k^2(\|c\|_{x_k}^*)^2 - \frac{\beta^2}{M_f^2}\right] - \frac{1}{M_f^2}\omega_*(\gamma) \\ &= \frac{\gamma}{2M_f}t_k\|c\|_{x_k}^* + \frac{\gamma}{2M_f t_k\|c\|_{x_k}^*}\left[\lambda_f^2(x_k) - \frac{\beta^2}{M_f^2}\right] - \frac{1}{M_f^2}\omega_*(\gamma). \end{aligned} \tag{80}$$

Note that

$$y_k - x_{k+1} \overset{(75)}{=} [\nabla^2 f(y_k)]^{-1} (\nabla f(y_k) + t_{k+1}c)\,. \tag{81}$$

Hence,

$$r_k \overset{\text{def}}{=} \|y_k - x_{k+1}\|_{y_k} \;= \lambda_{f_{t_{k+1}}}(y_k) \overset{(76)}{\leq} \;\; \tfrac{1}{M_f}\left(\tfrac{\beta}{1-\gamma} + \left(\tfrac{\gamma}{1-\gamma}\right)^2\right). \tag{82}$$

Therefore,

$$\begin{aligned} f(y_k) - f(x_{k+1}) &\overset{(8)}{\geq} \langle \nabla f(y_k), y_k - x_{k+1}\rangle - \tfrac{1}{M_f^2}\omega_*(M_f r_k) \\ &\overset{(81)}{=} \langle \nabla f(y_k), [\nabla^2 f(y_k)]^{-1} (\nabla f(y_k) + t_{k+1}c)\rangle - \tfrac{1}{M_f^2}\omega_*(M_f r_k) \\ &= \lambda^2_{f_{t_{k+1}}}(y_k) - \left(t_k - \tfrac{\gamma}{M_f\|c\|^*_{x_k}}\right)\langle c, [\nabla^2 f(y_k)]^{-1} (\nabla f(y_k) + t_{k+1}c)\rangle - \tfrac{\omega_*(M_f r_k)}{M_f^2} \\ &\geq \left(\lambda_{f_{t_{k+1}}}(y_k)\right)^2 - \left(t_k - \tfrac{\gamma}{M_f\|c\|^*_{x_k}}\right)\|c\|^*_{y_k}\lambda_{f_{t_{k+1}}}(y_k) - \tfrac{1}{M_f^2}\omega_*(M_f r_k) \\ &\geq -t_k\|c\|^*_{x_k}\tfrac{\lambda_{f_{t_{k+1}}}(y_k)}{1-\gamma} - \tfrac{1}{M_f^2}\omega_*(M_f r_k), \end{aligned} \tag{83}$$

where we used that $\|y_k - x_k\|_{x_k} \leq \gamma$ and, hence, $\|c\|^*_{y_k} \leq \frac{\|c\|^*_{x_k}}{1-\gamma}$ by (6). Combining (80) and (83), we obtain

$$\begin{aligned} f(x_k) - f(x_{k+1}) \;\;\geq\;\; & t_k\|c\|^*_{x_k}\left(\tfrac{\gamma}{2M_f} - \tfrac{\lambda_{f_{t_{k+1}}}(y_k)}{1-\gamma}\right) + \tfrac{\gamma}{2M_f t_k\|c\|^*_{x_k}}\left[\lambda^2_f(x_k) - \tfrac{\beta^2}{M_f^2}\right] \\ & -\tfrac{1}{M_f^2}\omega_*(\gamma) - \tfrac{1}{M_f^2}\omega_*(M_f r_k) \\ \overset{(82)}{\geq}\;\; & \tfrac{\kappa}{M_f} t_k\|c\|^*_{x_k} + \rho_k, \end{aligned}$$

where $\kappa = \frac{\gamma}{2} - \frac{\beta}{(1-\gamma)^2} - \frac{\gamma^2}{(1-\gamma)^3}$,

$$\rho_k = \frac{\gamma}{2M_f t_k\|c\|^*_{x_k}}\left[\lambda^2_f(x_k) - \frac{\beta^2}{M_f^2}\right] - \frac{1}{M_f^2}\omega_*(\gamma) - \frac{1}{M_f^2}\omega_*\left(\frac{\beta}{1-\gamma} - \frac{\gamma^2}{(1-\gamma)^2}\right).$$

Our next goal is to show that $\rho_k \geq 0$. Note that $t_k\|c\|^*_{x_k} \overset{(74)}{\leq} \lambda_f(x_k) + \frac{\beta}{M_f}$. Since $\lambda_f(x_k) \geq \frac{1}{2M_f}$, we have

$$\begin{aligned} \rho_k &\geq \tfrac{\gamma}{2M_f}\left[\lambda_f(x_k) - \tfrac{\beta}{M_f}\right] - \tfrac{1}{M_f^2}\omega_*(\gamma) - \tfrac{1}{M_f^2}\omega_*\left(\tfrac{\beta}{1-\gamma} - \tfrac{\gamma^2}{(1-\gamma)^2}\right) \\ &\geq \tfrac{\gamma(1-2\beta)}{4M_f^2} - \tfrac{1}{M_f^2}\omega_*(\gamma) - \tfrac{1}{M_f^2}\omega_*\left(\tfrac{\beta}{1-\gamma} - \tfrac{\gamma^2}{(1-\gamma)^2}\right). \end{aligned}$$

Using the values (73), by direct computation we can see that the right-hand side of this inequality is positive.

Thus, we have proved that $f(x_k) - f(x_{k+1}) \geq \frac{\kappa}{M_f} t_k \|c\|^*_{x_k}$. Using the values (73), we can see that $\kappa > 0$. Using the same steps as in the derivation of (42), we obtain $S_N \geq \frac{\kappa(N+1)^2}{M_f(f(x_0)-f(x_{N+1}))} \geq \frac{\kappa(N+1)^2}{M_f(f(x_0)-f^*)}$. □

Note that applying the same arguments as in Remark 2, we see that the sequence t_k for the PCPFS also has superlinear convergence. Let us estimate now the number of iterations, which is sufficient for PCPFS (77) to enter the region of quadratic convergence $\mathbb{Q}$ defined in (25). Recall that we denote $D = \max\limits_{x,y \in \mathrm{dom}\, f} \{\|x-y\|_{x_0} : \; f(x) \leq f(x_0), \;\; f(y) \leq f(x_0)\}$.

Theorem 10 *Let sequence $\{x_k\}_{k \geq 0}$ be generated by the method (77). Then,*

$$N \geq \left[\frac{\Delta(x_0)}{\gamma\kappa} \ln \left(\frac{M_f D \omega^{-1}(\Delta(x_0))}{\omega\left(\frac{(1-\beta)(1-2\beta)}{2}\right)} \right) \right]^{1/2} \tag{84}$$

guarantees that $x_N \in \mathbb{Q}$.

Proof The result follows from (78) and (45) that ensures that $x_k \in \mathbb{Q}$. □

As we can see from the estimate (84), up to a logarithmic factor, the number of iterations of the PCPFS is proportional to $\Delta^{1/2}(x_0)$. This is much better than the guarantee (26) for the DNM (24). Second, the constant $\left(\frac{1}{\gamma\kappa}\right)^{1/2} \leq 13.5$ in the complexity of the PCPFS is better than that of the standard PFS (36), where it is about 17.

The above results prescribe specific values for the accuracy of following the central path β and stepsize γ. Similarly to adaptive PFS in Section 3, we can propose also an adaptive PCPFS.

6 Predictor-corrector path-following scheme: implications for minimization problems

In this section, we apply the predictor-corrector path-following scheme (PCPFS) for minimization problems using an important subclass of self-concordant functions, ν-self-concordant barriers. The existing theory [34] shows that path-following schemes have faster convergence when applied to self-concordant barriers. We show that our PCPFS as well possesses improved convergence properties if we use an additional barrier property. We first apply this approach to the classical minimization of a linear function over a set. Then, we consider applications to linearly constrained problems.

6.1 Minimization problem: primal approach

In this subsection, we consider the following minimization problem

$$\min \langle c, x \rangle \quad \text{s.t.} \quad x \in Q, \tag{85}$$

where $Q = \mathrm{Dom}\, F$ is a bounded closed convex set with non-empty interior and F is a ν-self-concordant barrier meaning also that $M_F = 1$. Similar problem was considered in [34], but we propose a different algorithm and obtain a better constant in the complexity bound. The reason is that we can make longer steps using the PCPFS.

To solve problem (85), we start from $x_0 = x^*_F$, the analytic center of Q. For the purposes of this section we slightly redefine the main objects used in the path-following constructions. We define the central path $x(t)$, $t \geq 0$, as

$$\nabla F(x(t)) = -tc. \tag{86}$$

Clearly, $x(0) = x_0$. Note that

$$x(t) = \arg\min_{x \in \mathbb{E}} \left\{ f_t(x) \overset{\mathrm{def}}{=} F(x) + t\langle c, x\rangle \right\}, \quad t \geq 0. \tag{87}$$

We again redefine the main parameters for following the central path (cf. (30) and (73) where the stepsize γ is smaller)

$$\beta = 0.06, \quad \gamma \;=\; 0.254. \tag{88}$$

We say that point x satisfies an *approximate centering condition* if

$$\lambda_{f_t}(x) \overset{\mathrm{def}}{=} \|\nabla F(x) + tc\|^*_x \;\leq\; \beta. \tag{89}$$

Consider the following iterate which is a counterpart of (75):

$$(t_+, x_+) \;=\; \mathcal{PC}(t, x) \;\overset{\mathrm{def}}{=}\; \begin{cases} t_+ = t + \frac{\gamma}{\|c\|^*_x}, \\ y = x - \frac{\gamma}{\|c\|^*_x}[\nabla^2 F(x)]^{-1}c, \\ x_+ = y - [\nabla^2 F(y)]^{-1}(\nabla F(y) + t_+ c). \end{cases} \tag{90}$$

Similarly to Theorem 8, we can prove that if the pair (x, t) satisfies (89), then the pair (x_+, t_+) satisfies (89) too. Note that the chosen value γ is nearly 2 times larger than the classical choice $\gamma = \frac{5}{36}$ in [34]. This allows to make larger steps and the scheme converges faster.

Lemma 5 *Let x, t be generated by the iterates (90). Then* $t_+ \geq t\left(1 + \frac{\gamma}{\beta + \sqrt{\nu}}\right)$.

Proof From (89) and (50), we obtain

$$t\|c\|^*_x = \|\nabla f_t(x) - \nabla F(x)\|^*_x \leq \|\nabla f_t(x)\|^*_x + \|\nabla F(x)\|^*_x \leq \beta + \sqrt{\nu}$$

Further, $t_+ = t\left(1 + \frac{\gamma}{t\|c\|^*_x}\right) \geq t\left(1 + \frac{\gamma}{\beta + \sqrt{\nu}}\right)$. □

Lemma 6 *Let x^* be an optimal solution in (85). Then, for all $t > 0$,*

$$0 \le \langle c, x(t)\rangle - \langle c, x^*\rangle \le \frac{\nu}{t}. \tag{91}$$

If x satisfies approximate centering condition (89), then

$$|\langle c, x\rangle - \langle c, x^*\rangle| \le \frac{1}{t}\left(\nu + \frac{(\beta+\sqrt{\nu})\beta}{1-\beta}\right). \tag{92}$$

Proof Since x^* is an optimal solution in (85), by (86) and (52)

$$0 \le \langle c, x(t) - x^*\rangle = \frac{1}{t}\langle \nabla F(x(t)), x^* - x(t)\rangle \le \frac{\nu}{t}.$$

Further, let x satisfy (89). Denote $\lambda = \lambda_{f_t}(x)$. Then

$$\begin{aligned} t|\langle c, x - x(t)\rangle| &= |\langle \nabla f_t(x) - \nabla F(x), x - x(t)\rangle| \\ &\le (\lambda + \sqrt{\nu})\|x - x(t)\|_x \le (\lambda+\sqrt{\nu})\tfrac{\lambda}{1-\lambda} \le \tfrac{(\beta+\sqrt{\nu})\beta}{1-\beta}, \end{aligned}$$

where we used (50), (11), and (89). □

Note that inequality (92) gives us a reliable accuracy certificate based on the value of t.

The following is the main result of this subsection and gives complexity of the scheme (90) for solving problem (85).

Theorem 11 *Consider the predictor-corrector path-following scheme (PCPFS):*

$$t_0 = 0,\ x_0 \in \operatorname{dom} F \text{ s.t. } \|\nabla F(x_0)\|^*_{x_0} \le \beta, \quad (t_{k+1}, x_{k+1}) = \mathcal{PC}(t_k, x_k), \quad k \ge 0, \tag{93}$$

where β is defined in (88), $\mathcal{PC}$ is defined in (90). Then, for any $\varepsilon > 0$ and

$$N \ge N_\varepsilon = O\left(\frac{\sqrt{\nu}}{\gamma}\ln\left(\frac{\nu\|c\|^*_{x^*_F}}{\varepsilon}\right)\right),$$

we have $x_N \in Q$ and $|\langle c, x_N\rangle - \langle c, x^*\rangle| \le \varepsilon$.

Proof By construction we have that $x_k \in Q$ for $k \ge 0$. Note that, by (11), $r_0 \overset{\text{def}}{=} \|x_0 - x^*_F\|_{x_0} \le \frac{\beta}{1-\beta}$. Hence, by (6),

$$\frac{\gamma}{t_1} = \|c\|^*_{x_0} \le \frac{1}{1-r_0}\|c\|^*_{x^*_F} \le \frac{1-\beta}{1-2\beta}\|c\|^*_{x^*_F}.$$

Hence, for all $k \ge 0$, from Lemma 5, $t_k \ge \frac{\gamma(1-2\beta)}{(1-\beta)\|c\|^*_{x^*_F}}\left(1 + \frac{\gamma}{\beta+\sqrt{\nu}}\right)^{k-1}$. As noted above, we have that (89) holds during the iterations. The result follows by the combination of (92) and the above lover bound for t_k. □

As we can see, the main part of the obtained complexity bound is

$$3.94\sqrt{\nu}\ln\frac{\nu\|c\|^*_{x^*_F}}{\varepsilon},$$

which has a better constant 3.94 than the constant 7.2 for the one-step classical scheme [34]. Similarly to the previous sections, the stepsize γ may be chosen adaptively for better practical performance.

6.2 Minimization problem: dual approach

In this subsection, we consider the following maximization problem

$$\max -\langle c, x\rangle \quad \text{s.t.} \quad Bx = 0, \quad x \in Q, \tag{94}$$

where $B \in \mathbb{R}^{m\times n}$ and $0 \in \operatorname{int} Q$. The main difference with the previous subsection is that we now have linear constraints. Let us introduce

$$A = \begin{pmatrix} -c^T \\ B \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 0_m \end{pmatrix}.$$

Then, problem (94) is equivalent to

$$\max_{x,\alpha} \alpha \quad \text{s.t.} \quad Ax = \alpha b, \quad x \in Q. \tag{95}$$

Introducing penalty parameter σ and a self-concordant barrier F for the set Q s.t. $\nabla F(0) = 0$, we consider the following parametric family of problems

$$\max_{x,\alpha} \{\alpha\sigma - F(x) : \quad Ax = \alpha b\}, \quad \sigma \geq 0. \tag{96}$$

Introducing a Lagrange multiplier u for constraints $Ax = \alpha b$, we obtain

$$\begin{aligned} &\max\nolimits_{x,\alpha} \{\alpha\sigma - F(x) : \quad Ax = \alpha b\} = \max\nolimits_{x,\alpha} \min\nolimits_u \{\alpha\sigma - F(x) + \langle u, Ax - \alpha b\rangle\} \\ &= \min\nolimits_u \{\max\nolimits_\alpha \{\alpha(\sigma - \langle u, b\rangle) + \max\nolimits_x \{-F(x) + \langle A^T u, x\rangle\}\} \\ &= \min\nolimits_u \{F_*(A^T u) : \sigma = \langle b, u\rangle\} \end{aligned}$$

and the dual problem for (96) is

$$\min_u \{\Phi(u) : \sigma = \langle b, u\rangle\}, \quad \sigma \geq 0, \tag{97}$$

where $\Phi(u) \stackrel{\text{def}}{=} F_*(A^T u)$, F_* being the Fenchel conjugate for F. Note that $\Phi(u)$ is a standard self-concordant function with $M_\Phi = 1$, but not a self-concordant barrier.

Our goal is to follow the central path

$$u_\sigma \stackrel{\text{def}}{=} \arg\min_u \{\Phi(u) : \sigma = \langle b, u\rangle\}$$

of the dual problem (97) for σ from 0 to $+\infty$. Optimality conditions in (97) give the following characterization of the dual central path

$$\begin{aligned} \nabla\Phi(u_\sigma) = A\nabla F_*(A^T u_\sigma) = \alpha_\sigma b, \\ \langle b, u_\sigma\rangle = \sigma, \end{aligned} \tag{98}$$

where u_σ is an optimal solution for (97) and α_σ is an optimal Lagrange multiplier. Let us define $x_\sigma \stackrel{\text{def}}{=} \nabla F_*(A^T u_\sigma)$. Since F_* is the Fenchel conjugate for a barrier F for Q, $x_\sigma \in Q$ and $(\alpha_\sigma, x_\sigma)$ is a feasible point for (95), see (13).

We use the same β, γ as in the previous subsection (cf. (88)):

$$\beta = 0.06, \quad \gamma = 0.254. \tag{99}$$

For the dual problem (97), we define $\Phi_t(u) \stackrel{\text{def}}{=} \Phi(u) - tb$ and say that a point u satisfies an *approximate centering condition* if, for some $t \in \mathbb{R}$,

$$\lambda_{\Phi_t}(u) \stackrel{\text{def}}{=} \|\nabla\Phi(u) - tb\|_u^* \le \beta, \tag{100}$$

where the local norm $\|\cdot\|_u$ is induced by the Hessian of $\Phi(u)$. We also define $t(u) \stackrel{\text{def}}{=} \arg\min_t \lambda_{\Phi_t}(u)$ and $\lambda(u) \stackrel{\text{def}}{=} \lambda_{\Phi_{t(u)}}(u)$. Note that, for any v

$$\begin{aligned} t(v) &= \arg\min_t \lambda_{\Phi_t}(v) = \arg\min_t (\|\nabla\Phi(v) - tb\|_v^*)^2 \\ &= \arg\min_t \{\lambda_\Phi(v)^2 - 2t\langle b, [\nabla^2\Phi(v)]^{-1}\nabla\Phi(v)\rangle + t^2(\|b\|_v^*)^2\} \\ &= \frac{\langle b, [\nabla^2\Phi(v)]^{-1}\nabla\Phi(v)\rangle}{(\|b\|_v^*)^2}. \end{aligned} \tag{101}$$

We consider the dual PCPFS that approximately follows the central path of the dual problem (97) in order to solve the primal problem (94)

$$(\sigma_+, u_+) = \mathcal{PC}(\sigma, u) \stackrel{\text{def}}{=} \begin{cases} v = u + \frac{\gamma}{\|b\|_u^*}[\nabla^2\Phi(u)]^{-1}b, \\ \sigma_+ = \langle b, v\rangle, \\ u_+ = v + \arg\min_{h:\langle b,h\rangle=0}\left\{\langle\nabla\Phi(v), h\rangle + \frac{1}{2}\langle\nabla^2\Phi(v)h, h\rangle\right\}. \end{cases} \tag{102}$$

Note that by the optimality condition for the minimization problem defining u_+, there exist a Lagrange multiplier κ s.t.

$$\nabla\Phi(v) + \nabla^2\Phi(v)h - \kappa b = 0 \quad \Leftrightarrow \quad h = -[\nabla^2\Phi(v)]^{-1}(\nabla\Phi(v) - \kappa b).$$

Substituting this into the equality constraint $\langle b, h\rangle = 0$, we obtain

$$\kappa = \frac{\langle b, [\nabla^2 \Phi(v)]^{-1} \nabla \Phi(v)\rangle}{(\|b\|_v^*)^2} \overset{(101)}{=} t(v). \tag{103}$$

Moreover,

$$u_+ = v - [\nabla^2 \Phi(v)]^{-1} (\nabla \Phi(v) - t(v) b). \tag{104}$$

Lemma 7 *If $\lambda(u) \le \beta$, then $\lambda(u_+) \le \beta$ too.*

Proof Note that Lemma 4 holds for all t. Thus, applying Lemma 4 for $t = t(u)$ with $c = b$, $h = [\nabla^2 \Phi(u)]^{-1} b$, $\tau = \frac{\gamma}{\|b\|_u^*}$, and Φ with $M_\Phi = 1$ instead of f, since, in this case, $r = \|h\|_u = \|b\|_u^*$, we obtain, by (102) that $v = u + \tau h$ and

$$\lambda(v) = \lambda_{\Phi_{t(v)}}(v) = \min_t \lambda_{\Phi_t}(v) \le \lambda_{\Phi_{t(u)-\tau}}(v) \le \frac{\lambda(u)}{1-\gamma} + \left(\frac{\gamma}{1-\gamma}\right)^2 \le \frac{\beta}{1-\gamma} + \left(\frac{\gamma}{1-\gamma}\right)^2.$$

Since, by (104), u_+ is obtained by the Standard Newton Step for $\Phi_{t(v)}$ from v, by (19), we have

$$\lambda(u_+) \le \lambda_{\Phi_{t(v)}}(u_+) \le \left(\frac{\lambda_{\Phi_{t(v)}}(v)}{1 - \lambda_{\Phi_{t(v)}}(v)}\right)^2 = \left(\omega_*'(\lambda_{\Phi_{t(v)}}(v))\right)^2.$$

By the choice of β and γ in (99) and the estimate for $\lambda_{\Phi_{t(v)}}(v)$ above, we have $\omega_*'(\lambda_{\Phi_{t(v)}}(v)) \le \omega_*'\left(\frac{\beta}{1-\gamma} + \left(\frac{\gamma}{1-\gamma}\right)^2\right) \le \sqrt{\beta}$, which finishes the proof. □

Lemma 8 *Let u, σ be generated by the iterates (102). Then $\sigma_+ \ge \sigma\left(1 + \frac{\gamma}{\sqrt{\nu}}\right)$.*

Proof We have

$$\begin{aligned}\sigma_+ = \langle b, u_+\rangle = \langle b, v\rangle = \langle b, u\rangle + \gamma \|b\|_u^* = \langle b, u\rangle \left(1 + \tfrac{\gamma \|b\|_u^*}{\langle b, u\rangle}\right) \\ \ge \langle b, u\rangle \left(1 + \tfrac{\gamma}{\|u\|_u}\right) \ge \langle b, u\rangle \left(1 + \tfrac{\gamma}{\sqrt{\nu}}\right) = \sigma\left(1 + \tfrac{\gamma}{\sqrt{\nu}}\right),\end{aligned}$$

where we used that Φ is Fenchel conjugate for ν-self-concordant barrier F and, by (51), $\|u\|_u = \langle \nabla^2 \Phi(u) u, u\rangle^{1/2} \le \sqrt{\nu}$ for all u. □

Lemma 9 *Let α^* be an optimal function value in (95). Then, for all $\sigma \ge 0$,*

$$0 \le \alpha^* - \alpha_\sigma \le \frac{\nu}{\sigma}. \tag{105}$$

If u satisfies $\lambda(u) \le \beta$ and $\langle b, u\rangle = \sigma$, then

$$|\alpha^* - t(u)| \le \frac{1}{\sigma}\left(\nu + \frac{2\beta(1-\beta)}{1-2\beta}\sqrt{\nu}\right). \tag{106}$$

Proof Since (α^*, x^*) is an optimal solution for (95), $A(x^* - x_\sigma) = (\alpha^* - \alpha_\sigma)b$ and

$$0 \le \alpha^* - \alpha_\sigma = \frac{\langle A(x^* - x_\sigma), u_\sigma\rangle}{\langle b, u_\sigma\rangle} = \frac{\langle x^* - x_\sigma, A^T u_\sigma\rangle}{\langle b, u_\sigma\rangle} = \frac{\langle x^* - x_\sigma, \nabla F(x_\sigma)\rangle}{\langle b, u_\sigma\rangle}$$
$$\overset{(52)}{\le} \frac{\nu}{\langle b, u_\sigma\rangle} = \frac{\nu}{\sigma},$$

where we used that $x_\sigma = \nabla F_*(A^T u_\sigma)$ and, hence, $A^T u_\sigma = \nabla F(x_\sigma)$.

Since u_σ is a minimizer of self-concordant function $\Phi(v) - t(u)\langle b, v\rangle$ on the affine set $\langle b, v\rangle = \sigma$, by (11), we have

$$\|u_\sigma - u\|_u \le \omega_*'\left(\lambda_{\Phi(\cdot) - t(u)\langle b,\cdot\rangle}(u)\right) = \omega_*'\left(\lambda(u)\right) \le \frac{\beta}{1-\beta}.$$

Then, using Lemma 2, we obtain

$$\begin{aligned}
|\alpha_\sigma - t(u)|\sigma &= |\alpha_\sigma - t(u)|\langle b, u\rangle = |\langle \nabla\Phi(u_\sigma) - \nabla\Phi(u) + \nabla\Phi(u) - t(u)b, u\rangle| \\
&\le |\langle \nabla\Phi(u_\sigma) - \nabla\Phi(u), u\rangle| + |\langle \nabla\Phi(u) - t(u)b, u\rangle| \\
&\le \|\nabla\Phi(u_\sigma) - \nabla\Phi(u)\|_u^*\|u\|_u + \|\nabla\Phi(u) - t(u)b\|_u^*\|u\|_u \\
&\le \tfrac{\|u_\sigma - u\|_u}{1-\|u_\sigma - u\|_u}\|u\|_u + \|\nabla\Phi(u) - t(u)b\|_u^*\|u\|_u \le \left(\tfrac{\beta}{1-2\beta} + \beta\right)\sqrt{\nu} = \tfrac{2\beta(1-\beta)}{1-2\beta}\sqrt{\nu},
\end{aligned}$$

where we used that Φ is Fenchel conjugate for ν-self-concordant barrier F and, by (51), $\|u\|_u = \langle \nabla^2\Phi(u)u, u\rangle^{1/2} \le \sqrt{\nu}$ for all u. □

Note that inequality (106) gives us a reliable accuracy certificate based on the value of σ.

The following is the main result of this subsection and gives complexity of the scheme (102) for solving problem (94) in its equivalent form (95).

Theorem 12 *Consider the dual predictor-corrector path-following scheme:*

$$\sigma_0 = 0, u_0 = 0, \quad (\sigma_{k+1}, u_{k+1}) = \mathcal{PC}(\sigma_k, u_k), \quad k \ge 0, \tag{107}$$

where $\mathcal{PC}$ is defined in (102). Let also, for $k \ge 0$, $x_k = \nabla F_*(A^T u_k)$ and $\hat{x}_k$ be the solution to the following problem

$$\min\left\{\frac{1}{2}\|x - x_k\|_{x_k}^2 \quad s.t. \quad Ax = t(u_k)b\right\}.$$

Then, $\hat{x}_k \in Q$, $A\hat{x}_k = t(u_k)b$ for $k \ge 0$. Moreover, for any $\varepsilon > 0$ and

$$N \ge N_\varepsilon = O\left(\frac{\sqrt{\nu}}{\gamma}\ln\left(\frac{\nu}{\varepsilon\gamma\|b\|_0^*}\left(1 + \frac{2\beta(1-\beta)}{(1-2\beta)\sqrt{\nu}}\right)\right)\right), \tag{108}$$

we have $|\alpha^* - t(u_N)| \le \varepsilon$.

Proof By construction, we have $A\hat{x}_k = t(u_k)b$ and our next step is to show that $\hat{x}_k \in Q$. We do that by showing that $\|\hat{x}_k - x_k\|_{x_k} \le \beta < 1$, which since $x_k \in Q$ implies that $\hat{x}_k$ is in the Dikin ellipsoid around x_k and belongs to Q.

By the optimality conditions for the problem defining $\hat{x}_k$, we have that there exist a dual variable y such that

$$\begin{aligned} \nabla^2 F(x_k)(\hat{x}_k - x_k) + A^T y &= 0, \\ A\hat{x}_k &= t(u_k)b. \end{aligned}$$

From the first equality, we obtain $\hat{x}_k - x_k = -[\nabla^2 F(x_k)]^{-1} A^T y$. Then, from the second equality, we obtain $y = \left(A[\nabla^2 F(x_k)]^{-1} A^T\right)^{-1} (Ax_k - t(u_k)b)$, which finally gives

$$\hat{x}_k - x_k = -[\nabla^2 F(x_k)]^{-1} A^T \left(A[\nabla^2 F(x_k)]^{-1} A^T\right)^{-1} (Ax_k - t(u_k)b). \tag{109}$$

Recall that $Ax_k = A\nabla F_*(A^T u_k) = \nabla \Phi(u_k)$. Further, since $\Phi(u) = F_*(A^T u)$, by (14), we have that $\nabla^2 \Phi(u_k) = A[\nabla^2 F(x_k)]^{-1} A^T$. Thus, from (109), we have that $\hat{x}_k - x_k = -[\nabla^2 F(x_k)]^{-1} A^T [\nabla^2 \Phi(u_k)]^{-1} (\nabla \Phi(u_k) - t(u_k)b)$. Hence,

$$\begin{aligned} \|\hat{x}_k - x_k\|_{x_k}^2 &= \langle [\nabla^2 F(x_k)][\nabla^2 F(x_k)]^{-1} A^T [\nabla^2 \Phi(u_k)]^{-1} (\nabla \Phi(u_k) - t(u_k)b), \\ &\quad [\nabla^2 F(x_k)]^{-1} A^T [\nabla^2 \Phi(u_k)]^{-1} (\nabla \Phi(u_k) - t(u_k)b)\rangle \\ &= \langle A[\nabla^2 F(x_k)]^{-1} A^T [\nabla^2 \Phi(u_k)]^{-1} (\nabla \Phi(u_k) - t(u_k)b), \\ &\quad [\nabla^2 \Phi(u_k)]^{-1} (\nabla \Phi(u_k) - t(u_k)b)\rangle \\ &= \langle \nabla \Phi(u_k) - t(u_k)b, [\nabla^2 \Phi(u_k)]^{-1} (\nabla \Phi(u_k) - t(u_k)b)\rangle \\ &= (\|\nabla \Phi(u_k) - t(u_k)b\|_{u_k}^*)^2 \le \beta^2, \end{aligned}$$

where in the last inequality we used the approximate centering condition $\lambda(u_k) \le \beta$. Thus, we have that $\hat{x}_k$ belongs to the Dikin ellipsoid around x_k and, hence, belongs to Q.

Finally, we show the complexity result. Since $\nabla F(0) = 0$, we have that $\nabla \Phi(0) = 0$ and $\lambda(u_0) = \lambda(0) = 0 \le \beta$. Hence, by Lemma 7, we have that $\lambda(u_k) \le \beta$ for all $k \ge 0$. By (102) we have that $\sigma_k = \langle b, u_k\rangle$. Thus, by Lemma 9, we have that (106) holds with $\sigma = \sigma_k$, $u = u_k$ for all $k \ge 0$. Further, $\sigma_1 = \gamma \|b\|_0^*$ and, by Lemma 8, $\sigma_k \ge \gamma \|b\|_0^* \left(1 + \frac{\gamma}{\sqrt{\nu}}\right)^{k-1}$. Combining these observations with (108), we obtain that $|\alpha^* - t(u_N)| \le \varepsilon$. □

As we can see, the main part of the obtained complexity bound is

$$3.94\sqrt{\nu} \ln \frac{\nu}{\varepsilon \|b\|_0^*},$$

which has a better constant 3.94 than the constant 7.2 for the one-step classical scheme [34]. Similarly to the previous sections, the stepsize γ may be chosen adaptively for better practical performance. Note also that it is not necessary to calculate the point $\hat{x}_k$ in each iteration.

7 Minimizing strongly convex functions with Lipschitz Hessian

Let $B = B^* \succ 0$ map $\mathbb{E}$ to $\mathbb{E}^*$. Define the Euclidean metric $\|x\|^2 = \langle Bx, x\rangle^{1/2}$, $x \in \mathbb{E}$. In this section, we return back to the unconstrained problem (2), where f is a strongly convex function with parameter $\sigma_f > 0$:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \tfrac{1}{2}\sigma_f \|y - x\|^2, \quad x, y \in \mathbb{E}. \tag{110}$$

We also assume that $f \in \mathbb{C}^3(\mathbb{E})$ and its Hessian is Lipschitz continuous:

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq H_f \|x - y\|, \quad x, y \in \mathbb{E}. \tag{111}$$

Lemma 10 *Under assumptions above, function f is self-concordant with*

$$M_f = \frac{H_f}{2\sigma_f^{3/2}}. \tag{112}$$

Proof Indeed, for any point $x \in \mathbb{E}$ and direction $h \in \mathbb{E}$ we have

$$D^3 f(x)[h]^3 \overset{(111)}{\leq} H_f \left[\|h\|^2\right]^{3/2} \overset{(110)}{\leq} H_f \left[\tfrac{1}{\sigma_f}\langle \nabla^2 f(x)h, h\rangle\right]^{3/2}.$$

It remains to use definition (3). □

Thus, problem (2) can be solved by methods (24) and (36). The corresponding complexity bounds can be given in terms of the complexity measure $\Delta(x_0) = \frac{H_f^2}{\sigma_f^3}(f(x_0) - f^*)$. As we saw, the first and the second methods need $O(\Delta(x_0))$ and $\widetilde{O}(\Delta^{1/2}(x_0))$ iterations respectively. Let us show that for our particular subclass of self-concordant functions these bounds can be significantly improved.

We consider methods based on the *cubic regularization* of the Newton method (CRNM). Let us define quadratic approximation of f at point $x \in \mathbb{E}$:

$$Q(x, y) = f(x) + \langle \nabla f(x), y - x\rangle + \tfrac{1}{2}\langle \nabla^2 f(x)(y - x), y - x\rangle.$$

Then, from (111), $|f(y) - Q(x, y)| \leq \frac{H_f}{6}\|y - x\|^3, \quad y \in \mathbb{E}$, which justifies the cubic regularized Newton step given a parameter $M > 0$:

$$x_+ = T_M(x) \overset{\text{def}}{=} \arg\min_{y \in \mathbb{E}}\{Q(x, y) + \tfrac{1}{6}M\|y - x\|^3\}. \tag{113}$$

The cubic regularized Newton method (CRNM) iterating these steps converges [36] for functions satisfying (111) as $O(\frac{1}{k^2})$, where k is the iteration counter.

Define the region of quadratic convergence of the CRNM in terms of the function value (see (6.4) in [31]):

$$\mathbb{Q}_f = \left\{ x \in \mathbb{E} : \ f(x) - f^* \le \frac{\sigma_f^3}{2H_f^2} \overset{(112)}{=} \frac{1}{8M_f^2} \right\}.$$

Let us check how many iterations we need for entering $\mathbb{Q}_f$ by different schemes based on the step (113). Let our method have the following convergence rate

$$\begin{aligned} f(x_k) - f^* &\le \frac{cH_f \|x_0 - x^*\|^3}{k^p} \overset{(110)}{\le} \frac{cH_f}{k^p} \left(\frac{2}{\sigma_f} (f(x_0) - f^*) \right)^{3/2} \\ &\overset{(112)}{=} \frac{2^{5/2} c M_f}{k^p} (f(x_0) - f^*)^{3/2}, \end{aligned} \tag{114}$$

where c is an absolute constant and $p > 0$. Thus, we need

$$O\left(\left[M_f^3 (f(x_0) - f^*)^{3/2} \right]^{1/p} \right) = O\left(\Delta^{\frac{3}{2p}}(x_0) \right)$$

iterations for entering the region of superlinear convergence $\mathbb{Q}_f$. For the CRNM we have $p = 2$ (see [36]). Thus, it ensures complexity $O(\Delta^{3/4}(x_0))$. For the accelerated CRNM [31] we have $p = 3$. Thus, it needs $O(\Delta^{1/2}(x_0))$ iterations (which is slightly better than (44) by PFS). For the optimal up lo logarithmic factors [30] and optimal up to constant factors second-order methods [3, 23] we have $p = 3.5$ and these methods need $O(\Delta^{3/7}(x_0))$ iterations. However, note that there exists a powerful tool for accelerating these schemes, the *restarting strategy*.

Let us define k_p as the first integer, for which the right-hand side of inequality (114) is smaller than $\frac{1}{2}(f(x_0) - f^*)$:

$$\frac{2^{5/2} c M_f}{k_p^p} (f(x_0) - f^*)^{3/2} \le \frac{1}{2}(f(x_0) - f^*).$$

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

Multi-stage Acceleration Scheme	
At the first stage, we perform $t_1 = \lceil k_p \rceil$ iterations of our method starting from the point $y_0 = x_0$ and get the point y_1, which is the starting point for the next stage. In general, kth stage starts from the point y_{k-1} and its length is $t_k = \left\lceil \frac{k_p}{2^{(k-1)/(2p)}} \right\rceil$. Method stops when $y_k \in \mathbb{Q}_f$.	(115)

Theorem 13 *The total number of stages T in the optimizations strategy (115) satisfies inequality $T \leq 4 + \log_2 \Delta(x_0)$. The total number N of the lower-level iterations in this scheme does not exceed $4 + \log_2 \Delta(x_0) + \frac{2^{1/(2p)}}{2^{1/(2p)}-1} k_p$.*

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,

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

Hence, T satisfies inequality $\left(\frac{1}{2}\right)^{T-1} (f(x_0) - f^*) \geq \frac{1}{8 M_f^2}$. Finally,

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

□

Applying Theorem 13 to different CRNMs, we get the following complexity bounds.

- **Cubic Regularized Newton Method [36].** For this method $p = 2$. Therefore, the complexity bound of this scheme, as applied in the framework of multi-stage method (115) is of the order $O(\Delta^{1/4}(x_0))$. In fact, this method does not need a restarting strategy. Thus, Theorem 13 provides the CRNM with a better way of estimating its rate of convergence.
- **Accelerated Cubic Regularized Newton Method [31].** For this method $p = 3$. Hence, the complexity bound of the corresponding multi-stage scheme (115) becomes $O(\Delta^{1/6}(x_0))$.
- **Optimal second-order method [3, 23, 30].** For this method $p = 3.5$. Therefore, the corresponding complexity bound is $O(\Delta^{1/7}(x_0))$.

Remark 5 Note that the knowledge of f^* is not required for the restarting procedure. Indeed, if we know a lower bound $\tilde{f}$ such that $f^* \geq \tilde{f}$, we obtain that after $\tilde{k}_p = O\left(\left[M_f (f(x_0) - \tilde{f})^{1/2}\right]^{1/p}\right)$ steps of the inner method the right-hand side of inequality (114) is smaller than $\frac{1}{2}(f(x_0) - f^*)$ since $\tilde{k}_p \geq k_p$. Then, the overall number of lower-level iterations is of the order of $\tilde{k}_p = O\left(\tilde{\Delta}(x_0)\right)^{\frac{1}{2p}}$, where $\tilde{\Delta}(x_0) = M_f^2 (f(x_0) - \tilde{f})$.

Remark 6 To simplify this section we focused on CRNMs trying to compare complexities for two standard classes of functions for second-order methods. Similar complexities on the class of functions satisfying inequalities (110) and (111) can be achieved by using quadratic regularization instead of cubic [3, 10, 12, 28].

As we can see, the methods considered in this section have much better complexity bounds for problem (2) than the methods based on the framework of self-concordant functions. A possible explanation of this phenomena is that these methods use a more precise model of the objective function, which is based on two independent inequalities (110) and (111) instead of single inequality (3). Nevertheless, the methods in this section rely on Lipschitz property of the Hessian and, thus, are not applicable for a wide class of self-concordant functions, namely, self-concordant barriers, a typical example being log-barrier $-\ln x$. On the contrary, the DNM and the PFS can be still applied for the minimization of self-concordant barriers.

Acknowledgements We would like to thank two anonymous referees for the excellent remarks that helped us improve the quality of the paper.

Funding Open Access funding enabled and organized by Projekt DEAL.

Funding and conflicts of interests This research was supported by the National Research, Development and Innovation Office (NKFIH) under grant number 2024-1.2.3-HU-RIZONT-2024-00030 and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689). The authors declare that they have no conflict of interest.

References

1. Francis Bach. Self-concordant analysis for logistic regression. *Electron. J. Statist.*, 4:384–414, 2010.
2. Alejandro Carderera and Sebastian Pokutta. Second-order conditional gradient sliding. arXiv:2002.08907, 2020.
3. Yair Carmon, Danielle Hausler, Arun Jambulapati, Yujia Jin, and Aaron Sidford. Optimal and adaptive Monteiro-Svaiter acceleration. *Advances in Neural Information Processing Systems*, 35:20338–20350, 2022.
4. Yair Carmon, Arun Jambulapati, Qijia Jiang, Yujia Jin, Yin Tat Lee, Aaron Sidford, and Kevin Tian. Acceleration with a ball optimization oracle. *Advances in Neural Information Processing Systems*, 33:19052–19063, 2020.
5. Coralia Cartis, Nicholas IM Gould, and Philippe L Toint. Adaptive cubic regularisation methods for unconstrained optimization. part i: motivation, convergence and numerical results. *Mathematical Programming*, 127(2):245–295, 2011.
6. Coralia. Cartis, Nick I. Gould, and Philippe L. Toint. Universal regularization methods: Varying the power, the smoothness and the accuracy. *SIAM Journal on Optimization*, 29(1):595–615, 2019.
7. Andrew Conn, Nicholas Gould, and Philippe Toint. *Trust Region Methods*. Society for Industrial and Applied Mathematics, 2000.
8. Quoc Tran Dinh, Ion Necoara, Carlo Savorgnan, and Moritz Diehl. An inexact perturbed path-following method for lagrangian decomposition in large-scale separable convex optimization. *SIAM Journal on Optimization*, 23(1):95–125, 2013.
9. Nikita Doikov. Minimizing quasi-self-concordant functions by gradient regularization of newton method. arXiv:2308.14742, 2023.

10. Nikita Doikov, Konstantin Mishchenko, and Yurii Nesterov. Super-universal regularized Newton method. *SIAM Journal on Optimization*, 34(1):27–56, 2024.
11. Nikita Doikov and Yurii Nesterov. Minimizing uniformly convex functions by cubic regularization of newton method. *Journal of Optimization Theory and Applications*, 189(1):317–339, Apr 2021.
12. Nikita Doikov and Yurii Nesterov. Gradient regularization of newton method with bregman distances. *Mathematical Programming*, 204(1):1–25, Mar 2024.
13. Pavel Dvurechensky, Gabriele Iommazzo, Shimrit Shtern, and Mathias Staudigl. A conditional gradient homotopy method with applications to semidefinite programming. *IMA Journal of Numerical Analysis*, 2025. (accepted), arXiv:2207.03101.
14. Pavel Dvurechensky and Yurii Nesterov. Global performance guarantees of second-order methods for unconstrained convex minimization. CORE Discussion Paper 2018/32, CORE UCL, 2018.
15. Pavel Dvurechensky, Petr Ostroukhov, Kamil Safin, Shimrit Shtern, and Mathias Staudigl. Self-concordant analysis of Frank-Wolfe algorithms. In Hal Daume III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 2814–2824, Virtual, 13–18 Jul 2020. PMLR. arXiv:2002.04320.
16. Pavel Dvurechensky and Mathias Staudigl. Barrier algorithms for constrained non-convex optimization. In *Proceedings of the 41st International Conference on Machine Learning*, Proceedings of Machine Learning Research. PMLR, 2024. arXiv:2404.18724.
17. Pavel Dvurechensky and Mathias Staudigl. Hessian barrier algorithms for non-convex conic optimization. *Mathematical Programming*, 209(1):171–229, Jan 2025.
18. G. N. Grapiglia and Yu. Nesterov. Tensor methods for minimizing convex functions with hölder continuous higher-order derivatives. *SIAM Journal on Optimization*, 30(4):2750–2779, 2020.
19. Geovani Nunes Grapiglia and Yu Nesterov. Regularized newton methods for minimizing functions with Hölder continuous hessians. *SIAM Journal on Optimization*, 27(1):478–506, 2017.
20. Slavomír Hanzely, Dmitry Kamzolov, Dmitry Pasechnyuk, Alexander Gasnikov, Peter Richtarik, and Martin Takac. A damped newton method achieves global $o(1/k^2)$ and local quadratic convergence rate. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 25320–25334. Curran Associates, Inc., 2022.
21. L.V. Kantorovich. On newton's method for functional equations. *Dokl. Akad. Nauk SSSR*, 59(7):1237–1240, 1948.
22. Sai Praneeth Karimireddy, Sebastian U Stich, and Martin Jaggi. Global linear convergence of newton's method without strong-convexity or lipschitz gradients. arXiv:1806.00413, 2018.
23. Dmitry Kovalev and Alexander Gasnikov. The first optimal acceleration of high-order methods in smooth convex optimization. *Advances in Neural Information Processing Systems*, 35:35339–35351, 2022.
24. Jason D. Lee, Yuekai. Sun, and Michael A. Saunders. Proximal newton-type methods for minimizing composite functions. *SIAM Journal on Optimization*, 24(3):1420–1443, 2014.
25. Jinchao Li, Martin S. Andersen, and Lieven Vandenberghe. Inexact proximal newton methods for self-concordant functions. *Mathematical Methods of Operations Research*, 85(1):19–41, Feb 2017.
26. Deyi Liu, Volkan Cevher, and Quoc Tran-Dinh. A newton frank–wolfe method for constrained self-concordant minimization. *Journal of Global Optimization*, 83(2):273–299, Jun 2022.
27. Deyi Liu and Quoc Tran-Dinh. An inexact interior-point lagrangian decomposition algorithm with inexact oracles. *Journal of Optimization Theory and Applications*, 185(3):903–926, Jun 2020.
28. Konstantin Mishchenko. Regularized Newton method with global $\mathcal{O}(1/k^2)$ convergence. *SIAM Journal on Optimization*, 33(3):1440–1462, 2023.
29. Shinji Mizuno and Michael J Todd. On two homogeneous self-dual approaches to linear programming and its extensions. *Mathematical programming*, 89:517–534, 2001.
30. R. Monteiro and B. Svaiter. An accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods. *SIAM Journal on Optimization*, 23(2):1092–1125, 2013.
31. Yu. Nesterov. Accelerating the cubic regularization of newton's method on convex problems. *Mathematical Programming*, 112(1):159–181, Mar 2008.
32. Yu. Nesterov and L. Tunçel. Local superlinear convergence of polynomial-time interior-point methods for hyperbolicity cone optimization problems. *SIAM Journal on Optimization*, 26(1):139–170, 2016.
33. Yu. Nesterov and J.-Ph. Vial. Augmented self-concordant barriers and nonlinear optimization problems with finite complexity. *Mathematical Programming*, 99(1):149–174, Jan 2004.

34. Yurii Nesterov. *Introductory Lectures on Convex Optimization: a basic course*. Kluwer Academic Publishers, Massachusetts, 2004.
35. Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.
36. Yurii Nesterov and Boris Polyak. Cubic regularization of newton method and its global performance. *Mathematical Programming*, 108(1):177–205, 2006.
37. Brendan O'Donoghue, Eric Chu, Neal Parikh, and Stephen Boyd. Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications*, 169(3):1042–1068, Jun 2016.
38. Boris Polyak and Andrey Tremba. New versions of newton method: step-size choice, convergence domain and under-determined equations. *Optimization Methods and Software*, 0(0):1–32, 2019.
39. Anton Rodomanov and Dmitry Kropotov. A superlinearly-convergent proximal newton-type method for the optimization of finite sums. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2597–2605, New York, New York, USA, 20–22 Jun 2016. PMLR.
40. Chaobing Song, Yong Jiang, and Yi Ma. Unified acceleration of high-order algorithms under general hölder continuity. *SIAM Journal on Optimization*, 31(3):1797–1826, 2021.
41. Tianxiao Sun and Quoc Tran-Dinh. Generalized self-concordant functions: a recipe for newton-type methods. *Mathematical Programming*, 178(1-2):145–213, 2019.
42. Quoc Tran-Dinh, Liang Ling, and Kim-Chuan Toh. A new homotopy proximal variable-metric framework for composite convex minimization. arXiv:1812.05243, 2018.
43. Yinyu Ye, Michael J. Todd, and Shinji Mizuno. An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. *Mathematics of Operations Research*, 19(1):53–67, 1994.
44. Yuchen Zhang and Lin Xiao. *Communication-Efficient Distributed Optimization of Self-concordant Empirical Loss*, pages 289–341. Springer International Publishing, Cham, 2018.
45. Renbo Zhao and Robert M. Freund. Analysis of the frank–wolfe method for convex composite optimization involving a logarithmically-homogeneous barrier. *Mathematical Programming*, May 2022.