

# Self-concordant inclusions: a unified framework for path-following generalized Newton-type algorithms

Quoc Tran-Dinh<sup>1</sup>  · Tianxiao Sun<sup>1</sup> · Shu Lu<sup>1</sup>

Received: 18 October 2016 / Accepted: 21 March 2018 / Published online: 30 March 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature and Mathematical Optimization Society 2018

**Abstract** We study a class of monotone inclusions called “self-concordant inclusion” which covers three fundamental convex optimization formulations as special cases. We develop a new generalized Newton-type framework to solve this inclusion. Our framework subsumes three schemes: full-step, damped-step, and path-following methods as specific instances, while allows one to use inexact computation to form generalized Newton directions. We prove the local quadratic convergence of both full-step and damped-step algorithms. Then, we propose a new two-phase inexact path-following scheme for solving this monotone inclusion which possesses an  $\mathcal{O}(\sqrt{\nu} \log(1/\varepsilon))$ -worst-case iteration-complexity to achieve an  $\varepsilon$ -solution, where  $\nu$  is the barrier parameter and  $\varepsilon$  is a desired accuracy. As byproducts, we customize our scheme to solve three convex problems: the convex–concave saddle-point problem, the nonsmooth constrained convex program, and the nonsmooth convex program with linear constraints. We also provide three numerical examples to illustrate our theory and compare with existing methods.

**Keywords** Self-concordant inclusion · Generalized Newton-type methods · Path-following schemes · Monotone inclusion · Constrained convex programming · Saddle-point problems

---

✉ Quoc Tran-Dinh  
quoctd@email.unc.edu

Tianxiao Sun  
tianxias@email.unc.edu

Shu Lu  
shulu@email.unc.edu

<sup>1</sup> Department of Statistics and Operations Research, The University of North Carolina at Chapel Hill (UNC), 318 Hanes Hall, Chapel Hill, NC 27599, USA

**Mathematics Subject Classification** 90C25 · 90C06 · 90-08

## 1 Introduction

### 1.1 Problem statement

This paper is devoted to studying the following monotone inclusion which covers three important convex optimization templates [2, 16, 45]:

$$\text{Find } \mathbf{z}^* \in \mathbb{R}^p \text{ such that: } 0 \in \mathcal{A}_{\mathcal{Z}}(\mathbf{z}^*) := \mathcal{A}(\mathbf{z}^*) + \mathcal{N}_{\mathcal{Z}}(\mathbf{z}^*), \quad (1)$$

where  $\mathcal{Z}$  is a nonempty, closed, and convex set in  $\mathbb{R}^p$ ;  $\mathcal{A} : \mathbb{R}^p \rightrightarrows 2^{\mathbb{R}^p}$  is a multivalued and maximally monotone operator (*cf.* Definition 1);  $\mathcal{N}_{\mathcal{Z}}(\mathbf{z})$  is the normal cone of  $\mathcal{Z}$  at  $\mathbf{z}$  given by  $\{\mathbf{w} \in \mathbb{R}^p \mid \langle \mathbf{w}, \mathbf{z} - \hat{\mathbf{z}} \rangle \geq 0, \forall \hat{\mathbf{z}} \in \mathcal{Z}\}$  if  $\mathbf{z} \in \mathcal{Z}$ , and  $\emptyset$  otherwise; and “ $=$ ” stands for “is defined as”. Throughout this paper, we assume that  $\mathcal{Z}$  is endowed with a “ $v$  - self-concordant barrier”  $F$  (*cf.* Definition 3). We denote by  $\mathcal{Z}^* := \{\mathbf{z}^* \mid 0 \in \mathcal{A}(\mathbf{z}^*) + \mathcal{N}_{\mathcal{Z}}(\mathbf{z}^*)\}$  the solution set of (1).

Without the self-concordance of  $\mathcal{Z}$ , (1) is a classical monotone inclusion [2, 45], and can be reformulated into a multivalued variational inequality problem [16]. In particular, (1) covers the optimality (or KKT) conditions of unconstrained and constrained convex programs, and convex–concave saddle-point problems as described in Sect. 1.2. Therefore, (1) can be used as a unified tool to study and develop numerical methods for these problems [2, 16]. Methods for solving (1) and its special instances are well-developed under different structure assumptions imposed on  $\mathcal{A}$  and  $\mathcal{Z}$  [2, 16]. See Sect. 6 for a more thorough discussion.

We instead focus on a class of (1), where  $\mathcal{Z}$  is equipped with a “*self-concordant*” barrier (*cf.* Definition 3). The self-concordance notion was introduced by Nesterov and Nemirovskii [31, 36] in the 1990s to develop a unified theory and polynomial time algorithms in interior-point methods for structural convex programming, but has not been well exploited in other classes of optimization methods in both the convex and nonconvex cases.

Our approach in this paper can briefly be described as follows. Let  $\mathcal{Z}$  be equipped with a  $v$ -self-concordant barrier  $F$ . Since  $\mathcal{N}_{\mathcal{Z}}(\mathbf{z}) = \{\mathbf{0}^p\}$  for any  $\mathbf{z} \in \text{int}(\mathcal{Z})$ , the interior of  $\mathcal{Z}$ , we can define the following barrier problem associated with (1):

$$\text{Find } \mathbf{z}_t^* \in \text{int}(\mathcal{Z}) \text{ such that: } 0 \in \mathcal{A}_t(\mathbf{z}_t^*) := t \nabla F(\mathbf{z}_t^*) + \mathcal{A}(\mathbf{z}_t^*), \quad (2)$$

where  $t > 0$  is a penalty parameter. For any  $t > 0$ ,  $\mathcal{A}_t$  remains a maximally monotone operator. Hence, (2) is a parametric monotone inclusion depending on the parameter  $t$ . As we will show in Lemma 1, the solution  $\mathbf{z}_t^*$  of (2) exists and is unique for any  $t > 0$  under mild conditions. By perturbation theory [13, 42], one can show that  $\mathbf{z}_t^*$  is continuous w.r.t.  $t > 0$ . The set  $\{\mathbf{z}_t^* \mid t > 0\}$  containing solutions of (2) for each  $t$  generates a trajectory called the central path of (1). Each point  $\mathbf{z}_t^*$  on this path is called a central point. Our objective is to design efficiently numerical methods for solving (1) from the linearization of (2).

## 1.2 Three fundamental convex optimization templates

We present three basic problems in convex optimization covered by (1) to motivate our work.

### 1.2.1 Constrained convex programs

Consider a general constrained convex optimization problem as studied in [51, 52]:

$$g^* := \min_{\mathbf{x}} \{g(\mathbf{x}) \mid \mathbf{x} \in \mathcal{X}\}, \quad (3)$$

where  $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is proper, closed, and convex, and  $\mathcal{X}$  is a nonempty, closed and convex set in  $\mathbb{R}^n$  endowed with a  $v$ -self-concordant barrier  $f$  (*cf.* Definition 3). Let  $\partial g$  be the subdifferential of  $g$  (*cf.* Sect. 2). The following optimality condition is necessary and sufficient for  $\mathbf{x}^* \in \mathbb{R}^n$  to be an optimal solution of (3) under a given constraint qualification:

$$0 \in \partial g(\mathbf{x}^*) + \mathcal{N}_{\mathcal{X}}(\mathbf{x}^*).$$

By letting  $\mathbf{z} := \mathbf{x}$ ,  $\mathcal{A} := \partial g$  and  $\mathcal{Z} := \mathcal{X}$ , this inclusion exactly has the same form as (1). The barrier problem associated with (3) becomes

$$\mathcal{B}^*(t) := \min_{\mathbf{x} \in \mathbb{R}^n} \{\mathcal{B}(\mathbf{x}; t) := g(\mathbf{x}) + tf(\mathbf{x}) \mid \mathbf{x} \in \text{int}(\mathcal{X})\},$$

where  $t > 0$  is a penalty parameter. The optimality condition of this barrier problem is  $0 \in t\nabla f(\mathbf{x}_t^*) + \partial g(\mathbf{x}_t^*)$  which is exactly (2) with  $F := f$ .

### 1.2.2 Constrained convex programs with linear constraints

We are interested in the following constrained convex optimization problem:

$$\mathcal{G}^* := \max_{\mathbf{x} \in \mathbb{R}^n, \mathbf{s} \in \mathbb{R}^m} \{\mathcal{G}(\mathbf{x}, \mathbf{s}) := \langle \mathbf{c}, \mathbf{x} \rangle - g(\mathbf{s}) \mid L\mathbf{x} - W\mathbf{s} = \mathbf{b}, \mathbf{x} \in \mathcal{K}\}, \quad (4)$$

where  $\mathbf{c} \in \mathbb{R}^n$ ,  $\mathbf{b} \in \mathbb{R}^p$ ,  $L : \mathbb{R}^n \rightarrow \mathbb{R}^p$  and  $W : \mathbb{R}^{n_s} \rightarrow \mathbb{R}^p$  are linear operators,  $g : \mathbb{R}^{n_s} \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper, closed, convex and possibly nonsmooth function, and  $\mathcal{K}$  is a proper, nonempty, closed, pointed, and convex cone endowed with a  $v$ -self-concordant logarithmically homogeneous barrier  $f$  (*cf.* Definition 3). In addition, we assume that  $n \leq p$ .

The corresponding dual problem of (4) can be written as follows:

$$\mathcal{H}^* := \min_{\mathbf{y}} \{\mathcal{H}(\mathbf{y}) := g^*(W^*\mathbf{y}) + \langle \mathbf{b}, \mathbf{y} \rangle \mid L^*\mathbf{y} - \mathbf{c} \in \mathcal{K}^*\}, \quad (5)$$

where  $\mathcal{K}^* := \{\mathbf{u} \in \mathbb{R}^n \mid \langle \mathbf{x}, \mathbf{u} \rangle \geq 0, \forall \mathbf{x} \in \mathcal{K}\}$  is the dual cone of  $\mathcal{K}$ ,  $L^*$  and  $W^*$  are the adjoint operators of  $L$  and  $W$ , respectively, and  $g^*(\mathbf{u}) := \sup_{\mathbf{s}} \{\langle \mathbf{u}, \mathbf{s} \rangle - g(\mathbf{s})\}$  is the

conjugate of  $g$ . Let  $\mathcal{Y} := \{\mathbf{y} \in \mathbb{R}^p \mid L^*\mathbf{y} - \mathbf{c} \in \mathcal{K}^*\}$ . Then, the optimality condition of (5) becomes

$$0 \in \partial g^*(W^*\mathbf{y}^*) + \mathbf{b} + \mathcal{N}_{\mathcal{Y}}(\mathbf{y}^*),$$

which fits the form of (1). The barrier problem associated with the dual problem (5) is

$$\min_{\mathbf{y} \in \mathbb{R}^p} \{g^*(W^*\mathbf{y}) + \langle \mathbf{b}, \mathbf{y} \rangle + t f^*(\mathbf{c} - L^*\mathbf{y})\}, \quad (6)$$

where  $f^*$  is the Fenchel conjugate of  $f$ . If we define  $\psi(\cdot) = g^*(W^*(\cdot)) + \langle \mathbf{b}, \cdot \rangle$  and  $\varphi(\cdot) := f^*(\mathbf{c} - L^*(\cdot))$  the barrier of  $\mathcal{Y}$ , then the optimality condition of (6) becomes

$$0 \in -tL(\nabla f^*(\mathbf{c} - L^*\mathbf{y}_t^*)) + \partial\psi(\mathbf{y}_t^*), \quad (7)$$

which falls into the form (2) with  $\mathbf{z} := \mathbf{y}$ ,  $F(\cdot) := \varphi(\cdot) = f^*(\mathbf{c} - L^*(\cdot))$ , and  $\mathcal{A}(\cdot) := \partial\psi(\cdot)$ .

### 1.2.3 Convex–concave saddle-point problems

Consider the following convex–concave saddle-point problem that covers many applications including signal/image processing and duality theory [9, 11]:

$$\Phi^* := \min_{\mathbf{y} \in \mathcal{Y}} \{\Phi(\mathbf{y}) := \psi(\mathbf{y}) + \max_{\mathbf{x} \in \mathcal{X}} \{\langle \mathbf{y}, L\mathbf{x} \rangle - g(\mathbf{x})\}\}, \quad (8)$$

where  $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper, closed, and convex function;  $\psi : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  is also a proper, closed, and convex function;  $\mathcal{X}$  and  $\mathcal{Y}$  are two nonempty, closed and convex sets in  $\mathbb{R}^n$  and  $\mathbb{R}^m$ , respectively; and  $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a given linear operator. The optimality condition of (8) for a saddle point  $(\mathbf{x}^*, \mathbf{y}^*)$  is

$$\begin{cases} 0 \in \partial g(\mathbf{x}^*) - L^*\mathbf{y}^* + \mathcal{N}_{\mathcal{X}}(\mathbf{x}^*), \\ 0 \in \partial\psi(\mathbf{y}^*) + L\mathbf{x}^* + \mathcal{N}_{\mathcal{Y}}(\mathbf{y}^*). \end{cases} \quad (9)$$

where  $\mathcal{N}_{\mathcal{X}}$  and  $\mathcal{N}_{\mathcal{Y}}$  are the normal cones of  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively. If we define  $\mathbf{z} := (\mathbf{x}, \mathbf{y})$ ,  $\mathcal{Z} := \mathcal{X} \times \mathcal{Y}$ ,

$$\mathcal{A}(\mathbf{z}) := \begin{pmatrix} \partial g(\mathbf{x}) - L^*\mathbf{y} \\ \partial\psi(\mathbf{y}) + L\mathbf{x} \end{pmatrix}, \text{ and } \mathcal{N}_{\mathcal{Z}}(\mathbf{z}) := \mathcal{N}_{\mathcal{X}}(\mathbf{x}) \times \mathcal{N}_{\mathcal{Y}}(\mathbf{y}), \quad (10)$$

then (9) can be cast into the form (1).

Let  $\mathcal{X}$  and  $\mathcal{Y}$  be endowed with self-concordant barriers  $f$  and  $\varphi$ , respectively. Then, we can write down the barrier problem of (8) as

$$\mathcal{B}^*(t) := \min_{\mathbf{y} \in \text{int}(\mathcal{Y})} \{\mathcal{B}(\mathbf{y}; t) := \psi(\mathbf{y}) + t\varphi(\mathbf{y}) + \max_{\mathbf{x} \in \text{int}(\mathcal{X})} \{\langle \mathbf{y}, L\mathbf{x} \rangle - g(\mathbf{x}) - tf(\mathbf{x})\}\},$$

where  $t > 0$  is a penalty parameter. Hence, its optimality condition becomes

$$\begin{cases} 0 \in t\nabla f(\mathbf{x}_t^*) - L^*\mathbf{y}_t^* + \partial g(\mathbf{x}_t^*) \\ 0 \in t\nabla\varphi(\mathbf{y}_t^*) + L\mathbf{x}_t^* + \partial\psi(\mathbf{y}_t^*). \end{cases} \quad (11)$$

If we define  $F(\mathbf{z}) := f(\mathbf{x}) + \varphi(\mathbf{y})$ , then (11) can be written into the form (2).

### 1.3 Our contribution

We unify the proximal-point and the path-following interior-point schemes to design a joint treatment between these methods for solving the monotone inclusion (1). Our approach is fundamentally different from existing methods, where we use the means of self-concordant barriers of the feasible set  $\mathcal{Z}$  in (1) to develop generalized Newton-type algorithms.

We propose a unified framework that covers three fundamental convex problems as previously described. We develop three different generalized Newton-type methods for solving (1). Our framework covers the previous work in [51, 52] for the convex problem (3) as special cases. Our approach relies on specific structure of  $\mathcal{Z}$  in (1) where we can treat (1) via the linearization of its barrier formulation (2). By introducing a new scaled resolvent mapping and generalized proximal Newton decrement, we develop a generalized Newton framework for solving (1). Then, we combine it and a homotopy strategy for the penalty parameter  $t$  to obtain a path-following scheme for solving (1). Our approach relates to classical proximal-point and interior-point methods in the literature as discussed in Sect. 6.

*Contribution* To this end, we can summarize the contribution of this paper as follows:

- (a) (*Theory*) We study a class of monotone inclusions, which we call “self-concordant inclusions”, that provides a unified framework using self-concordant barriers to investigate three fundamental classes of convex optimization problems. We prove the existence and uniqueness of the central path of (2) under mild assumptions.
- (b) (*Algorithms*) We propose a generalized Newton-type framework for solving (1). This framework covers three methods: full-step generalized Newton, damped-step generalized Newton, and full-step path-following generalized Newton schemes. Our methods allow one to use inexact computation to form generalized Newton-search directions, and adaptively update the contraction factor for the penalty parameter  $t$  associated with  $F$ .
- (c) (*Convergence theory*) We prove the local quadratic convergence of the first two inexact generalized Newton-methods, and estimate the worst-case iteration-complexity of the third inexact path-following scheme to achieve an  $\varepsilon$ -solution, where  $\varepsilon$  is a desired accuracy. Surprisingly, this worst-case complexity is  $\mathcal{O}(\sqrt{v} \log(1/\varepsilon))$  which is the same as in standard path-following methods for smooth convex programming [31, 35].
- (d) (*Special instances*) We customize our path-following framework to solve three convex problems: (3), (4) and (8), and investigate the overall worst-case iteration-

complexity for each method. In addition, we provide an explicit scheme to recover primal solutions from the dual ones in the linear constrained case (4) with a rigorous convergence guarantee.

Let us emphasize the following points of our contribution. First, using a barrier function for the constraint set  $\mathcal{Z}$  in (1) allows us to handle a wide class of problems where projections onto  $\mathcal{Z}$  are no longer efficient, e.g., when  $\mathcal{Z}$  is a general polyhedron, or a hyperbolic cone. Second, these are second-order methods which often achieve high accuracy solutions and have a fast local convergence rate. This is an advantage when the evaluation of barrier function values and its derivatives is expensive. In addition, they are known to be robust to inexact computation and noise. However, as a compensation, the complexity-per-iteration is often higher than first-order methods. Fortunately, inexact computation allows us to apply iterative methods for computing generalized Newton search directions. Third, when applied to (3), (4), and (8), the efficiency of our algorithms depends on the cost of the scaled proximal operator of  $g$  and  $\psi$  which is a key component in first-order, primal-dual, and splitting methods. Finally, our framework is sufficiently general and can be customized to specific classes of structural convex problems such as conic and geometric programming.

## 1.4 Outline of the paper

The rest of this paper is organized as follows. In Sect. 2, we recall some preliminary results including monotone operators and self-concordance notions [35] used in this paper. Section 3 presents a unified generalized Newton-type framework that covers three different methods and analyzes their local convergence properties as well as their worst-case iteration-complexity. Section 4 customizes our path-following framework to solve the convex–concave minimax problem (8), the primal constrained convex problem (3), and the linear constrained convex problem (4). Section 5 deals with specific applications and illustrates numerically the performance of our algorithms. For clarity of exposition, technical proofs of the results in the main text are deferred to the appendix.

## 2 Preliminaries: monotonicity, convexity, and self-concordance

We recall some preliminary results from classical convex analysis including monotonicity, convexity, and self-concordance which will be used in the sequel.

### 2.1 Basic definitions

Let  $\langle \mathbf{u}, \mathbf{v} \rangle$  or  $\mathbf{u}^\top \mathbf{v}$  denote the inner product, and  $\|\mathbf{u}\|_2$  denote the Euclidean norm for any  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^p$ . For a proper, closed, and convex function  $F : \mathbb{R}^p \rightarrow \mathbb{R} \cup \{+\infty\}$ ,  $\text{dom}(F) := \{\mathbf{z} \in \mathbb{R}^p \mid F(\mathbf{z}) < +\infty\}$  denotes its domain, and  $\text{Dom}(F) := \text{cl}(\text{dom}(F))$  denotes the closure of  $\text{dom}(F)$ ,  $\partial F(\mathbf{z}) := \{\mathbf{w} \in \mathbb{R}^p \mid F(\mathbf{u}) \geq F(\mathbf{z}) + \langle \mathbf{w}, \mathbf{u} - \mathbf{z} \rangle, \forall \mathbf{u} \in \text{dom}(F)\}$  denotes its subdifferential at  $\mathbf{z}$  [44]. We also use  $\mathcal{C}^3(\mathcal{Z})$  to denote the class of three-time continuously differentiable functions from  $\mathcal{Z} \subseteq \mathbb{R}^p$ .

to  $\mathbb{R}$ . Given a multivalued operator  $\mathcal{A} : \mathbb{R}^p \rightrightarrows 2^{\mathbb{R}^p}$ ,  $\text{dom}(\mathcal{A}) := \{\mathbf{z} \in \mathbb{R}^p \mid \mathcal{A}(\mathbf{z}) \neq \emptyset\}$  denotes the domain of  $\mathcal{A}$ , and  $\text{gr}(\mathcal{A}) := \{(\mathbf{z}, \mathbf{w}) \in \mathbb{R}^p \times \mathbb{R}^p \mid \mathbf{w} \in \mathcal{A}(\mathbf{z})\}$  denotes the graph of  $\mathcal{A}$ .  $\mathcal{S}_+^p$  stands for the symmetric positive semidefinite cone of dimension  $p$ , and  $\mathcal{S}_{++}^p$  is its interior, i.e.,  $\mathcal{S}_{++}^p = \text{int}(\mathcal{S}_+^p)$ . For any  $\mathbf{Q} \in \mathcal{S}_{++}^p$ , we denote  $\|\mathbf{z}\|_{\mathbf{Q}} := \langle \mathbf{Q}\mathbf{z}, \mathbf{z} \rangle^{1/2}$  the weighted norm of  $\mathbf{z}$ , and  $\|\mathbf{z}\|_{\mathbf{Q}}^* := \langle \mathbf{Q}^{-1}\mathbf{z}, \mathbf{z} \rangle^{1/2}$  is its dual norm.

For the three-time continuously differentiable and convex function  $F : \mathbb{R}^p \rightarrow \mathbb{R}$  defined in (2) such that  $\nabla^2 F(\mathbf{z}) \succ 0$  at some  $\mathbf{z} \in \text{dom}(F)$  (i.e.,  $\nabla^2 F(\mathbf{z})$  is symmetric positive definite), we define a local norm, and its dual norm, respectively as

$$\|\mathbf{u}\|_{\mathbf{z}} := \langle \nabla^2 F(\mathbf{z})\mathbf{u}, \mathbf{u} \rangle^{1/2}, \quad \text{and} \quad \|\mathbf{v}\|_{\mathbf{z}}^* := \langle \nabla^2 F(\mathbf{z})^{-1}\mathbf{v}, \mathbf{v} \rangle^{1/2}, \quad (12)$$

for given  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^p$ . Clearly, with this definition, the well-known Cauchy–Schwarz inequality  $\langle \mathbf{u}, \mathbf{v} \rangle \leq \|\mathbf{u}\|_{\mathbf{z}} \|\mathbf{v}\|_{\mathbf{z}}^*$  holds.

## 2.2 Maximally monotone, resolvent, and proximal operators

**Definition 1** Given a multivalued operator  $\mathcal{A} : \mathbb{R}^p \rightrightarrows 2^{\mathbb{R}^p}$ , we say that  $\mathcal{A}$  is monotone if for any  $\mathbf{z}, \hat{\mathbf{z}} \in \text{dom}(\mathcal{A})$ ,  $\langle \mathbf{w} - \hat{\mathbf{w}}, \mathbf{z} - \hat{\mathbf{z}} \rangle \geq 0$  for  $\mathbf{w} \in \mathcal{A}(\mathbf{z})$  and  $\hat{\mathbf{w}} \in \mathcal{A}(\hat{\mathbf{z}})$ ; and  $\mathcal{A}$  is maximal if its graph is not properly contained in the graph of any other monotone operator.

Given a maximally monotone operator  $\mathcal{A} : \mathbb{R}^p \rightrightarrows 2^{\mathbb{R}^p}$ , and  $\mathbf{Q} \in \mathcal{S}_{++}^p$ , we define

$$J_{\mathbf{Q}^{-1}\mathcal{A}}(\mathbf{z}) = (\mathbb{I} + \mathbf{Q}^{-1}\mathcal{A})^{-1}(\mathbf{z}) := \{\mathbf{w} \in \mathbb{R}^p \mid 0 \in \mathbf{Q}(\mathbf{w} - \mathbf{z}) + \mathcal{A}(\mathbf{w})\}, \quad (13)$$

the scaled resolvent operator of  $\mathcal{A}$  [2, 45]. It is well-known that  $\text{dom}(J_{\mathbf{Q}^{-1}\mathcal{A}}) = \mathbb{R}^p$  and  $J_{\mathbf{Q}^{-1}\mathcal{A}}$  is well-defined and single-valued. If  $\mathbf{Q} = \mathbb{I}$ , the identity operator, then  $J_{\mathbb{I}^{-1}\mathcal{A}} \equiv J_{\mathcal{A}}$  is the standard resolvent of  $\mathcal{A}$ . When  $\mathcal{A} = \partial g$ , the subdifferential of a proper, closed and convex function  $g$ ,  $J_{\mathbf{Q}^{-1}\mathcal{A}}$  becomes a scaled proximal operator of  $g$ , which is defined as follows:

$$\text{prox}_{\mathbf{Q}^{-1}g}(\mathbf{x}) := \underset{\mathbf{u}}{\operatorname{argmin}} \left\{ g(\mathbf{u}) + (1/2)\|\mathbf{u} - \mathbf{x}\|_{\mathbf{Q}}^2 \mid \mathbf{u} \in \text{dom}(g) \right\}. \quad (14)$$

Methods for evaluating  $\text{prox}_{\mathbf{Q}^{-1}g}$  have been discussed in the literature, see, e.g., [4, 18]. If  $\mathbf{Q} = \mathbb{I}$ , then  $\text{prox}_{\mathbf{Q}^{-1}g} = \text{prox}_g$ , the standard proximal operator of  $g$ . Examples of such functions can be found, e.g., in [2, 10, 39].

## 2.3 Self-concordant functions and self-concordant barriers

We also use the self-concordance concept introduced by Nesterov and Nemirovskii [31, 35].

**Definition 2** A univariate convex function  $\varphi \in \mathcal{C}^3(\text{dom}(\varphi))$  is called *standard self-concordant* if  $|\varphi'''(\tau)| \leq 2\varphi''(\tau)^{3/2}$  for all  $\tau \in \text{dom}(\varphi)$ , where  $\text{dom}(\varphi)$  is an open

set in  $\mathbb{R}$ . A function  $F : \text{dom}(F) \subseteq \mathbb{R}^p \rightarrow \mathbb{R}$  is standard self-concordant if for any  $\mathbf{z} \in \text{dom}(F)$  and  $\mathbf{v} \in \mathbb{R}^p$ , the univariate function  $\varphi$  defined by  $\tau \mapsto \varphi(\tau) := F(\mathbf{z} + \tau\mathbf{v})$  is standard self-concordant.

**Definition 3** A standard self-concordant function  $F : \mathcal{Z} \subset \mathbb{R}^p \rightarrow \mathbb{R}$  is a  $v$ -self-concordant barrier for a convex set  $\mathcal{Z}$  with parameter  $v > 0$  if  $\text{dom}(F) = \text{int}(\mathcal{Z})$  and

$$\sup_{\mathbf{u} \in \mathbb{R}^p} \left\{ 2\langle \nabla F(\mathbf{z}), \mathbf{u} \rangle - \|\mathbf{u}\|_{\mathbf{z}}^2 \right\} \leq v, \quad \forall \mathbf{z} \in \text{dom}(F).$$

In addition,  $F(\mathbf{z})$  tends to  $+\infty$  as  $\mathbf{z}$  approaches the boundary of  $\mathcal{Z}$ . A function  $F$  is called a  $v$ -self-concordant logarithmically homogeneous barrier function of  $\mathcal{Z}$  if  $F(\tau\mathbf{z}) = F(\mathbf{z}) - v \log(\tau)$  for all  $\mathbf{z} \in \text{int}(\mathcal{Z})$  and  $\tau > 0$ .

Several simple sets are equipped with a self-concordant logarithmically homogeneous barrier. For instance,  $F_{\mathbb{R}_+^p}(\mathbf{z}) := -\sum_{i=1}^p \log(z_i)$  is a  $p$ -self-concordant barrier of  $\mathbb{R}_+^p$ ,  $F_{\mathcal{S}_+^n}(\mathbf{Z}) := -\log \det(\mathbf{Z})$  is an  $n$ -self-concordant barrier of  $\mathcal{S}_+^n$ , and  $F(\mathbf{z}, t) = -\log(t^2 - \|\mathbf{z}\|_2^2)$  is a 2-self-concordant barrier of the Lorentz cone  $\mathcal{L}_{p+1} := \{(\mathbf{z}, t) \in \mathbb{R}^p \times \mathbb{R}_+ \mid \|\mathbf{z}\|_2 \leq t\}$ .

When  $\mathcal{Z}$  is bounded and  $F$  is a  $v$ -self-concordant barrier for  $\mathcal{Z}$ , the analytical center  $\bar{\mathbf{z}}_f^*$  of  $f$  exists and is unique. It is defined by

$$\bar{\mathbf{z}}_f^* := \underset{\mathbf{z} \in \text{int}(\mathcal{Z})}{\operatorname{argmin}} F(\mathbf{z}), \quad (\text{and its optimality condition is } \nabla F(\bar{\mathbf{z}}_f^*) = 0). \quad (15)$$

Let us define  $\kappa := v + 2\sqrt{v}$  for a general self-concordant barrier, and  $\kappa := 1$  for a self-concordant logarithmically homogeneous barrier. Then, we have  $\|\mathbf{v}\|_{\mathbf{z}}^* \leq \kappa \|\mathbf{v}\|_{\bar{\mathbf{z}}_f^*}^*$  for any  $\mathbf{z} \in \text{int}(\mathcal{Z})$  and  $\mathbf{v} \in \mathbb{R}^p$ .

Let  $\mathcal{K}$  be a proper, closed, and pointed convex cone. If  $\mathcal{K}$  is endowed with a  $v$ -self-concordant logarithmically homogeneous barrier function  $F$ , then its Fenchel conjugate (also called the Legendre transformation [35])

$$F^*(\mathbf{w}) := \sup_{\mathbf{z}} \{ \langle \mathbf{w}, \mathbf{z} \rangle - F(\mathbf{z}) \mid \mathbf{z} \in \mathcal{K} \}$$

is also a  $v$ -self-concordant logarithmically homogeneous barrier of the anti-dual cone  $-\mathcal{K}^*$  of  $\mathcal{K}$ . For instance, if  $\mathcal{K} = \mathcal{S}_+^n$ , then  $\mathcal{K}^* = \mathcal{S}_+^n = \mathcal{K}$  (self-dual cone). A barrier function of  $\mathcal{S}_+^n$  is  $F(\mathbf{z}) := -\log \det(\mathbf{z})$ . Hence,  $F^*(\mathbf{w}) = -n - \log \det(-\mathbf{w})$  is a barrier function of  $-\mathcal{K}^*$ .

### 3 Generalized Newton-type methods for self-concordant inclusions

We propose a novel generalized Newton-type scheme for solving (1). Then, we develop three inexact generalized Newton-type schemes: full-step, damped-step, and path-following algorithms based on the linearization of (2). We provide a unified analysis for convergence.

### 3.1 Fundamental assumptions and fixed-point characterization

Throughout this paper, we rely on the following fundamental, but standard assumption.

- Assumption A.1** (a) The feasible set  $\mathcal{Z}$  is nonempty, closed, and convex. Moreover,  $\mathcal{Z}$  is equipped with a  $v$ -self-concordant barrier  $F$ .  
 (b) The operator  $\mathcal{A}$  is maximally monotone,  $\text{int}(\mathcal{Z}) \cap \text{dom}(\mathcal{A}) \neq \emptyset$ , and  $\text{dom}(\mathcal{A})$  is either an open set or a closed set.  
 (c) The solution set  $\mathcal{Z}^*$  of (1) is nonempty.

Note that since  $\text{dom}(\nabla F) = \text{int}(\mathcal{Z})$ , Assumption A.1 is sufficient for  $\mathcal{A}_t$  defined by (2) to be maximally monotone [2, Corollary 25.5]. This assumption can be relaxed to different conditions as discussed in [2, Section 25.1], which we omit here.

Our aim is to compute an approximate solution of (1) up to a given accuracy as follows:

**Definition 4** Given  $\varepsilon \geq 0$ , we say that  $\tilde{\mathbf{z}}_\varepsilon^* \in \text{int}(\mathcal{Z})$  is an  $\varepsilon$ -solution to (1) if

$$\text{dist}_{\tilde{\mathbf{z}}_\varepsilon^*}(\mathbf{0}, \mathcal{A}(\tilde{\mathbf{z}}_\varepsilon^*)) := \min_{\mathbf{e}} \left\{ \|\mathbf{e}\|_{\tilde{\mathbf{z}}_\varepsilon^*}^* \mid \mathbf{e} \in \mathcal{A}(\tilde{\mathbf{z}}_\varepsilon^*) \right\} \leq \varepsilon.$$

Here,  $\text{dist}_{\mathbf{z}}(\mathbf{w}, \Omega)$  defines a weighted distance from  $\mathbf{w} \in \mathbb{R}^p$  to a nonempty, closed and convex set  $\Omega$  in  $\mathbb{R}^p$ , and  $\mathbf{0}$  is the zero vector. Since  $\tilde{\mathbf{z}}_\varepsilon^* \in \text{int}(\mathcal{Z})$ , we have  $\mathcal{N}_{\mathcal{Z}}(\tilde{\mathbf{z}}_\varepsilon^*) = \{\mathbf{0}\}$ . Hence,  $\mathcal{A}_{\mathcal{Z}}(\tilde{\mathbf{z}}_\varepsilon^*) \equiv \mathcal{A}(\tilde{\mathbf{z}}_\varepsilon^*)$ .

If  $\varepsilon = 0$ , then Definition 4 says that  $0 \in \mathcal{A}(\tilde{\mathbf{z}}_\varepsilon^*)$ . Hence,  $\tilde{\mathbf{z}}_\varepsilon^*$  is an exact solution of (1) in the interior of  $\mathcal{Z}$ . If all solutions  $\mathbf{z}^*$  of (1) are on the boundary of  $\mathcal{Z}$ , then Definition 4 only works if  $\varepsilon > 0$ .

We can modify Definition 4 as  $\text{dist}_{\mathbf{z}}(\mathbf{0}, \mathcal{A}_{\mathcal{Z}}(\tilde{\mathbf{z}}_\varepsilon^*)) \leq \varepsilon$ , where  $\mathbf{z} \in \text{int}(\mathcal{Z})$  is fixed a priori. Then, all the results in the next sections remain preserved but require a slight justification. In the sequel, we develop different numerical methods to generate a sequence  $\{\mathbf{z}^k\}$  from the interior of  $\mathcal{Z}$ .

*The scaled resolvent operator of  $\mathcal{A}$ :* Let us fix  $\hat{\mathbf{z}} \in \text{int}(\mathcal{Z})$  and  $t > 0$ . Then, we have  $\nabla^2 F(\hat{\mathbf{z}}) \in \mathcal{S}_{++}^p$ . For simplicity of presentation, using (13) we denote by

$$\mathcal{P}_{\hat{\mathbf{z}}}(\cdot; t) := J_{(t\nabla^2 F(\hat{\mathbf{z}}))^{-1}} \mathcal{A}(\cdot) = \left( \mathbb{I} + t^{-1} \nabla^2 F(\hat{\mathbf{z}})^{-1} \mathcal{A} \right)^{-1}(\cdot), \quad (16)$$

the scaled resolvent of  $\mathcal{A}$ . Using  $\mathcal{P}_{\hat{\mathbf{z}}}(\cdot; t)$ , we can formulate the monotone inclusion (2) as a fixed-point equation

$$\mathbf{z}_t^* = \mathcal{P}_{\hat{\mathbf{z}}} \left( \mathbf{z}_t^* - \nabla^2 F(\hat{\mathbf{z}})^{-1} \nabla F(\mathbf{z}_t^*); t \right). \quad (17)$$

Clearly, if we define  $R_{\hat{\mathbf{z}}}(\cdot) := \mathcal{P}_{\hat{\mathbf{z}}}(\cdot - \nabla^2 F(\hat{\mathbf{z}})^{-1} \nabla F(\cdot); t)$ , then  $\mathbf{z}_t^*$  is a fixed-point of  $R_{\hat{\mathbf{z}}}(\cdot)$ .

*The existence of the central path* We prove in “Appendix 7.1” the following existence result for (2). Let us recall that the horizon cone of a convex set  $C$  consists of vectors  $\omega$  such that  $\mathbf{z} + \tau\omega \in \text{cl}(C)$  for any  $\mathbf{z} \in C$  and any  $\tau > 0$ , where  $\text{cl}(C)$  stands for the closure of  $C$ .

**Lemma 1** Suppose that for any nonzero  $\omega$  in the horizon cone of  $\text{int}(\mathcal{Z}) \cap \text{dom}(\mathcal{A})$ , there exists some  $\hat{\mathbf{z}} \in \text{int}(\mathcal{Z}) \cap \text{dom}(\mathcal{A})$  with  $\hat{\mathbf{a}} \in \mathcal{A}(\hat{\mathbf{z}})$  such that  $\langle \hat{\mathbf{a}}, \omega \rangle > 0$ . Then, for each  $t > 0$ , problem (2) has a unique solution. Moreover, we have  $\text{dist}_{\mathbf{z}_t^*}(\mathbf{0}, \mathcal{A}(\mathbf{z}_t^*)) \leq t\sqrt{v}$ , so  $\mathbf{z}_t^*$  is an  $\varepsilon$ -solution to (1) in the sense of Definition 4 if  $t \leq \frac{\varepsilon}{\sqrt{v}}$ .

The assumption in Lemma 1 is quite general. There are two special cases in which this assumption holds. First, if  $\text{int}(\mathcal{Z}) \cap \text{dom}(\mathcal{A})$  is bounded, then the only element in the horizon cone of  $\text{int}(\mathcal{Z}) \cap \text{dom}(\mathcal{A})$  is  $\mathbf{0}$ , and the assumption trivially holds. Second, if the solution set  $\mathcal{Z}^*$  of (1) is nonempty and bounded, and the set-valued map  $\mathcal{A}$  is continuous at points in  $\text{bdry}(\mathcal{Z} \cap \text{dom}(\mathcal{A}))$  relative to  $\mathcal{Z} \cap \text{dom}(\mathcal{A})$ , then this assumption also holds as can be shown using [45, Theorem 12.51]. Here,  $\text{bdry}(\mathcal{Z})$  stands for the boundary of  $\mathcal{Z}$ , and we refer to [45, Definition 5.4] for the definition of the continuity of a set-valued map.

*Generalized gradient mapping* Fix  $\hat{\mathbf{z}} \in \text{int}(\mathcal{Z})$  with  $\nabla^2 F(\hat{\mathbf{z}}) \succ 0$ . We consider the following linear monotone inclusion in  $\mathbf{s}$ :

$$0 \in t\nabla F(\mathbf{z}) + t\nabla^2 F(\hat{\mathbf{z}})(\mathbf{s} - \mathbf{z}) + \mathcal{A}(\mathbf{s}). \quad (18)$$

If we take  $\mathbf{z} = \hat{\mathbf{z}}$ , then it becomes a linearization (with respect to  $\nabla F$ ) of (2) at a given point  $\mathbf{z}$ . It is obvious that (18) is strongly and maximally monotone so that its solution exists and is unique. We denote this solution by  $\mathbf{s}_{\hat{\mathbf{z}}}(\mathbf{z}; t)$ , and, by using  $\mathcal{P}_{\hat{\mathbf{z}}}(\cdot; t)$ , it can be written as

$$\mathbf{s}_{\hat{\mathbf{z}}}(\mathbf{z}; t) := \mathcal{P}_{\hat{\mathbf{z}}} \left( \mathbf{z} - \nabla^2 F(\hat{\mathbf{z}})^{-1} \nabla F(\mathbf{z}); t \right). \quad (19)$$

Next, we define the following mapping

$$G_{\hat{\mathbf{z}}}(\mathbf{z}; t) := \nabla^2 F(\hat{\mathbf{z}})(\mathbf{z} - \mathbf{s}_{\hat{\mathbf{z}}}(\mathbf{z}; t)) \equiv \nabla^2 F(\hat{\mathbf{z}}) \left( \mathbf{z} - \mathcal{P}_{\hat{\mathbf{z}}} \left( \mathbf{z} - \nabla^2 F(\hat{\mathbf{z}})^{-1} \nabla F(\mathbf{z}); t \right) \right), \quad (20)$$

When  $\mathcal{A} = 0$ ,  $G_{\hat{\mathbf{z}}}(\mathbf{z}; t) = \nabla F(\mathbf{z})$ , which is exactly the gradient of  $F$ . Then, we adopt the name in [31] to call  $G_{\hat{\mathbf{z}}}(\cdot; t)$  a generalized gradient mapping.

Given  $G_{\mathbf{z}}(\mathbf{z}; t)$  as in (20) with  $\hat{\mathbf{z}} = \mathbf{z}$ , we define the following generalized Newton decrement  $\lambda_t(\mathbf{z})$  to analyze the convergence of generalized Newton-type methods below:

$$\lambda_t(\mathbf{z}) := \|G_{\mathbf{z}}(\mathbf{z}; t)\|_{\mathbf{z}}^* = \left\| \mathbf{z} - \mathcal{P}_{\mathbf{z}} \left( \mathbf{z} - \nabla^2 F(\mathbf{z})^{-1} \nabla F(\mathbf{z}); t \right) \right\|_{\mathbf{z}}. \quad (21)$$

If  $\mathcal{A}(\mathbf{z}) = \mathbf{c}$ , a constant operator, then  $\lambda_t(\mathbf{z}) = \|t^{-1}\mathbf{c} + \nabla F(\mathbf{z})\|_{\mathbf{z}}^*$ , which is exactly the Newton decrement defined in [31, Formula 4.2.16].

To conclude, we summarize the result of this subsection in the following lemma. This result is a direct consequence of the definition of  $G_{\mathbf{z}}(\cdot; t)$  and  $\lambda_t(\cdot)$ . We omit the proof.

**Lemma 2** *The solution  $s_{\bar{\mathbf{z}}}(\cdot; t)$  of (18) exists and is unique for any  $\mathbf{z} \in \text{dom}(F)$ . Consequently,  $G_{\bar{\mathbf{z}}}(\cdot; t)$  given by (20) is well-defined on  $\text{dom}(F)$ .*

*Let  $\mathbf{z}_t^* \in \text{int}(\mathcal{Z})$  be a given point and  $\lambda_t(\cdot)$  be defined by (21). Then,  $\lambda_t(\mathbf{z}_t^*) = 0$  if and only if  $\mathbf{z}_t^*$  is a solution to (2).*

In the sequel, we only work with the solution  $s_{\bar{\mathbf{z}}}(\mathbf{z}; t)$  of (18) which exists and is unique. However, we assume throughout this paper that the assumptions of Lemma 1 hold so that the solution  $\mathbf{z}_t^*$  of (2) exists and is unique for each  $t > 0$ . We do not use  $\mathbf{z}_t^*$  of (2) at any step of our algorithms. Since  $\mathbf{z}_t^*$  is on the central path of (1) at  $t > 0$ ,  $\mathbf{z}_t^* \in \mathcal{Z}$ . If  $t > 0$  is sufficiently small, e.g.,  $t := \varepsilon/\sqrt{v}$ , then we can say that  $\mathbf{z}_t^*$  is also an  $\varepsilon$ -solution of (1) as stated in Lemma 1 in the sense of Definition 4.

### 3.2 Inexact generalized Newton-type schemes

The main step of the generalized Newton method is presented as follows: For a fixed value  $t > 0$ , and a given iterate  $\mathbf{z} \in \text{int}(\mathcal{Z})$ , we approximate  $F$  by its Taylor's expansion and define

$$\widehat{\mathcal{A}}_t(\mathbf{w}; \mathbf{z}) := t \left[ \nabla F(\mathbf{z}) + \nabla^2 F(\mathbf{z})(\mathbf{w} - \mathbf{z}) \right] + \mathcal{A}(\mathbf{w}). \quad (22)$$

Since  $\nabla^2 F(\mathbf{z}) \succ 0$ , we can compute the unique solution of the linearized inclusion:

$$\mathbf{s}_{\mathbf{z}}(\mathbf{z}; t) := \{ \mathbf{w} \in \text{int}(\mathcal{Z}) \mid 0 \in \widehat{\mathcal{A}}_t(\mathbf{w}; \mathbf{z}) \} \equiv (\widehat{\mathcal{A}}_t(\cdot; \mathbf{z}))^{-1}(\mathbf{0}). \quad (23)$$

Computing  $\mathbf{s}_{\mathbf{z}}(\mathbf{z}; t)$  exactly is often impractical, so we allow one to approximate it as follows.

**Definition 5** Given an accuracy  $\delta \in [0, 1)$ , we say that  $\mathbf{z}_+$  is a  $\delta$ -approximation to the true solution  $\bar{\mathbf{z}}_+ := \mathbf{s}_{\mathbf{z}}(\mathbf{z}; t)$  defined in (23) (and is denoted by  $\mathbf{z}_+ \approx \bar{\mathbf{z}}_+$ ) if

$$\text{dist}_{\mathbf{z}}(\mathbf{0}, \widehat{\mathcal{A}}_t(\mathbf{z}_+; \mathbf{z})) = \min_{\mathbf{e}} \{ \|\mathbf{e}\|_{\mathbf{z}}^* \mid \mathbf{e} \in \widehat{\mathcal{A}}_t(\mathbf{z}_+; \mathbf{z}) \} \leq t\delta. \quad (24)$$

First, we show that, under (24), we have  $\|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}} \leq \delta$ . Next, since we are working with the linearization (23) of (2), the following lemma, whose proof is in “Appendix 7.2”, shows that an approximate solution of (23) is also an approximate solution of problem (1).

**Lemma 3** *Let  $\mathbf{z}_+$  be a  $\delta$ -approximate solution to  $\bar{\mathbf{z}}_+$  of (23) in the sense of Definition 5. Then, we have  $\|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}} \leq \delta$ . Furthermore, if  $\lambda_t(\mathbf{z}) + \delta < 1$ , then*

$$\text{dist}_{\mathbf{z}_+}(\mathbf{0}, \mathcal{A}_{\mathcal{Z}}(\mathbf{z}_+)) \leq (1 - \lambda_t(\mathbf{z}) - \delta)^{-1} (\sqrt{v} + \lambda_t(\mathbf{z}) + 2\delta) t. \quad (25)$$

*If we choose  $t > 0$  such that  $t \leq (1 - \lambda_t(\mathbf{z}) - \delta) (\sqrt{v} + \lambda_t(\mathbf{z}) + 2\delta)^{-1} \varepsilon$  for a given  $\varepsilon > 0$ , then  $\mathbf{z}_+ \in \text{int}(\mathcal{Z})$ , and  $\mathbf{z}_+$  is an  $\varepsilon$ -solution to (1) in the sense of Definition 4.*

We now investigate the convergence of the inexact full-step, damped-step, and path-following generalized Newton methods.

### 3.2.1 A key estimate

The following theorem provides a key estimate to analyze the convergence of the generalized Newton-type scheme above, whose proof can be found in “Appendix 7.3”.

**Theorem 1** *For a given  $\mathbf{z} \in \text{int}(\mathcal{Z})$ , let  $\mathbf{z}_+$  be the point generated by the inexact generalized Newton scheme (in the sense of Definition 5):*

$$\mathbf{z}_+ \approx \bar{\mathbf{z}}_+ := \mathcal{P}_{\mathbf{z}} \left( \mathbf{z} - \nabla^2 F(\mathbf{z})^{-1} \nabla F(\mathbf{z}); t_+ \right). \quad (26)$$

*Then, if  $\lambda_{t_+}(\mathbf{z}) + \delta(\mathbf{z}) < 1$ , where  $\lambda_{t_+}(\mathbf{z})$  is defined by (21) and  $\delta(\mathbf{z}) := \|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}}$ , then  $\mathbf{z}_+ \in \text{int}(\mathcal{Z})$ , and the following estimate holds:*

$$\lambda_{t_+}(\mathbf{z}_+) \leq \left( \frac{\lambda_{t_+}(\mathbf{z}) + \delta(\mathbf{z})}{1 - \lambda_{t_+}(\mathbf{z}) - \delta(\mathbf{z})} \right)^2 + \frac{\delta(\mathbf{z})}{(1 - \lambda_{t_+}(\mathbf{z}) - \delta(\mathbf{z}))^3}. \quad (27)$$

*Moreover, the right-hand side of (27) is monotonically increasing w.r.t.  $\lambda_{t_+}(\mathbf{z})$  and  $\delta(\mathbf{z})$ .*

Clearly, if  $\mathbf{z}_+ = \bar{\mathbf{z}}_+$  (i.e., the subproblem (26) is solved exactly), then (27) reduces to

$$\lambda_{t_+}(\mathbf{z}_+) \leq \left( \frac{\lambda_{t_+}(\mathbf{z})}{1 - \lambda_{t_+}(\mathbf{z})} \right)^2, \quad (28)$$

which is in the form of [31, Theorem 4.1.14], but for the exact variant of (26).

### 3.2.2 Neighborhood of the central path and quadratic convergence region

Given the generalized Newton decrement  $\lambda_t(\cdot)$  defined by (21), we consider the following set

$$\Omega_t(\beta) := \{ \mathbf{z} \in \text{int}(\mathcal{Z}) \mid \lambda_t(\mathbf{z}) \leq \beta \}, \quad (29)$$

where  $\beta \in (0, 1)$ . We call  $\Omega_t(\beta)$  a neighborhood of the central path of (2) with the radius  $\beta$ .

If we can choose  $\beta \in (0, 1)$  such that:

- (i) the sequence  $\{\mathbf{z}^k\}$  generated by a generalized Newton scheme starting from  $\mathbf{z}^0 \in \Omega_t(\beta)$  belongs to  $\Omega_t(\beta)$ , and
- (ii) the corresponding sequence of the generalized Newton decrements  $\{\lambda_t(\mathbf{z}^k)\}$  converges quadratically to zero,

then we call  $\Omega_t(\beta)$  a quadratic convergence region of this method, and denote it by  $\mathcal{Q}_t(\beta)$ .

Next, we propose two inexact generalized Newton schemes: full-step and damped step, to generate a sequence  $\{\mathbf{z}^k\}$  starting from  $\mathbf{z}^0 \in \mathcal{Q}_t(\beta)$  for some predefined  $\beta \in (0, 1)$ , and show that  $\{\lambda_t(\mathbf{z}^k)\}$  converges quadratically to zero. In these schemes,

the penalty parameter  $t$  is fixed at a sufficiently small value *a priori*, which may cause some difficulty for computing  $\mathbf{z}^{k+1}$  from  $\mathbf{z}^k$  due to the ill-condition of  $\nabla^2 F(\mathbf{z}^k)$ . To avoid this situation, we then suggest using a path-following scheme to gradually decrease  $t$  starting from a larger value  $t = t_0 > 0$ .

### 3.2.3 Inexact full-step generalized Newton method (FGN): local convergence

We investigate the convergence of the FGN and maximize the radius of its quadratic convergence region  $\mathcal{Q}_t(\beta)$ . The following theorem shows a quadratic convergence of the inexact generalized Newton scheme, whose proof is deferred to “Appendix 7.4”.

**Theorem 2** *Given a fixed parameter  $t > 0$ , let  $\{\mathbf{z}^k\}$  be a sequence generated by the following inexact full-step generalized Newton scheme (FGN):*

$$\mathbf{z}^{k+1} \approx \bar{\mathbf{z}}^{k+1} := \mathcal{P}_{\mathbf{z}^k} \left( \mathbf{z}^k - \nabla^2 F(\mathbf{z}^k)^{-1} \nabla F(\mathbf{z}^k); t \right). \quad (\text{FGN})$$

where the approximation  $\approx$  is in the sense of Definition 5. Then, we have three statements:

- (a) Let  $0 < \beta < \frac{1}{2}(3 - \sqrt{5})$  be a given radius, and  $\Omega_t(\beta)$  be defined by (29). If we choose  $\mathbf{z}^0 \in \Omega_t(\beta)$  and the tolerance  $\delta_k$  in Definition 5 such that

$$\|\mathbf{z}^{k+1} - \bar{\mathbf{z}}^{k+1}\|_{\mathbf{z}^k} \leq \delta_k \leq \bar{\delta}_k(\beta) := \frac{\beta(1 - 3\beta + \beta^2)(1 - \beta)^4}{2\beta^3 - 5\beta^2 + 3\beta + 1},$$

then  $\{\mathbf{z}^k\}$  generated by FGN belongs to  $\Omega_t(\beta)$ .

- (b) If we choose  $\delta_k \leq \frac{\lambda_t(\mathbf{z}^k)^2}{1 - \lambda_t(\mathbf{z}^k)}$ , then, for  $k \geq 0$  and  $\lambda_t(\mathbf{z}^0) < 1$ , we have

$$\lambda_t(\mathbf{z}^{k+1}) \leq \left( \frac{2 - 4\lambda_t(\mathbf{z}^k) + \lambda_t(\mathbf{z}^k)}{(1 - 2\lambda_t(\mathbf{z}^k))^3} \right) \lambda_t(\mathbf{z}^k)^2 < 1. \quad (30)$$

For any  $\beta \in (0, 0.18858]$ , if we choose  $\mathbf{z}^0 \in \mathcal{Q}_t(\beta)$ , where  $\mathcal{Q}_t(\beta)$  is the quadratic convergence region of FGN, then  $\{\mathbf{z}^k\} \subset \mathcal{Q}_t(\beta)$ , and  $\{\lambda_t(\mathbf{z}^k)\}$  quadratically converges to zero.

- (c) Let  $c := \frac{2-4\beta+\beta^2}{(1-2\beta)^3} \in (0, 1)$ , and  $\varepsilon > 0$  be a given tolerance for  $\beta \in (0, 0.18858]$ .

If we choose  $t := (1 - \epsilon)(\sqrt{v} + \epsilon + 2\epsilon^2/(1 - \epsilon))^{-1}\varepsilon$  for a sufficiently small  $\epsilon \in (0, \beta)$ , and update  $\delta_k := \frac{2\beta_k^2}{1 - \beta_k}$  with  $\beta_k := c^{2^{k-1}}\beta^{2^k}$ , then after at most  $k := \mathcal{O}(\ln(\ln(1/\varepsilon)))$  iterations,  $\mathbf{z}^k$  is an  $\varepsilon$ -solution of (1) in the sense of Definition 4.

By a numerical experiment, we can show that  $\bar{\delta}_k$  defined in Theorem 2(a) is maximized at  $\beta_* = 0.0997 \in (0, 0.18858]$  with  $\bar{\delta}_k^* = 0.0372$ . Therefore, if we choose these values, we can maximize the tolerance  $\delta_k$ . Note that  $\bar{\delta}_k$  is decreasing when  $\beta$  is increasing in  $(0.0997, \frac{1}{2}(3 - \sqrt{5}))$  and vice versa. Hence, we can trade-off between the radius  $\beta$  of  $\Omega_t(\beta)$  and the tolerance  $\delta_k$  of the subproblem in (FGN).

### 3.2.4 Inexact damped-step GN method (**DGN**): local convergence

We now consider a damped-step generalized Newton scheme. The following theorem summarizes the result whose proof is moved to “Appendix 7.5”.

**Theorem 3** *Given a fixed parameter  $t > 0$ , let  $\{\mathbf{z}^k\}$  be the sequence generated by the following inexact damped-step generalized Newton scheme (**DGN**):*

$$\begin{cases} \tilde{\mathbf{z}}^{k+1} \approx \bar{\mathbf{z}}^{k+1} := \mathcal{P}_{\mathbf{z}^k}(\mathbf{z}^k - \nabla^2 F(\mathbf{z}^k)^{-1} \nabla F(\mathbf{z}^k); t), \\ \alpha_k := \frac{1}{(1+\tilde{\lambda}_t(\mathbf{z}^k))} \text{ with } \tilde{\lambda}_t(\mathbf{z}^k) := \|\tilde{\mathbf{z}}^{k+1} - \mathbf{z}^k\|_{\mathbf{z}^k}, \\ \mathbf{z}^{k+1} := (1 - \alpha_k)\mathbf{z}^k + \alpha_k \tilde{\mathbf{z}}^{k+1}. \end{cases} \quad (\text{DGN})$$

Then, we have the following three statements:

- (a) If we choose  $\delta_k$  such that  $\delta_k \leq \frac{\tilde{\lambda}_t(\mathbf{z}^k)^2}{1+\tilde{\lambda}_t(\mathbf{z}^k)}$ , then

$$\tilde{\lambda}_t(\mathbf{z}^{k+1}) \leq \left( \frac{2\tilde{\lambda}_t(\mathbf{z}^k)^2 + 4\tilde{\lambda}_t(\mathbf{z}^k) + 3}{1 - \tilde{\lambda}_t(\mathbf{z}^k)^2 (2\tilde{\lambda}_t(\mathbf{z}^k)^2 + 4\tilde{\lambda}_t(\mathbf{z}^k) + 3)} \right) \tilde{\lambda}_t(\mathbf{z}^k)^2. \quad (31)$$

For any  $\beta \in (0, 0.21027]$ , the sequence  $\{\mathbf{z}^k\}$  generated by **DGN** starting from any  $\mathbf{z}^0 \in \Omega_t(\beta)$  belongs to  $\Omega_t(\beta)$ , i.e.,  $\{\mathbf{z}^k\} \subset \Omega_t(\beta)$ .

- (b) If we choose  $\beta \in (0, 0.21027]$ , then the sequence  $\{\tilde{\lambda}_t(\mathbf{z}^k)\}$  generated by **DGN** starting from any  $\mathbf{z}^0 \in \mathcal{Q}_t(\beta)$  also quadratically converges to zero.
- (c) For  $\beta \in (0, 0.21027]$ , let  $\bar{c} := \frac{2\beta^2+4\beta+3}{1-\beta^2(2\beta^2+4\beta+3)} \in (0, 1)$ , and  $\varepsilon > 0$  be a given tolerance. If we choose  $t := (1 - 2\epsilon^2)(\sqrt{\nu}(1 + \epsilon) + \epsilon + 3\epsilon^2)^{-1}\varepsilon$  for a sufficiently small  $\epsilon \in (0, \beta)$ , and update  $\delta_k := (1 + \beta_k)^{-1}\beta_k^2$  with  $\beta_k := \bar{c}^{2^k-1}\beta^{2^k}$ , then after at most  $k := \mathcal{O}(\ln(\ln(1/\varepsilon)))$  iterations,  $\mathbf{z}^k$  is an  $\varepsilon$ -solution of (1) in the sense of Definition 4.

Note that the quadratic convergence stated in Theorem 3 is given through  $\{\tilde{\lambda}_t(\mathbf{z}^k)\}$ , which is computable as opposed to  $\{\lambda_t(\mathbf{z}^k)\}$  in Theorem 2. Due to the fact that  $\lambda_t(\mathbf{z}^k) \leq \tilde{\lambda}_t(\mathbf{z}^k) + \delta(\mathbf{z}^k) \leq \tilde{\lambda}_t(\mathbf{z}^k) + \frac{\tilde{\lambda}_t(\mathbf{z}^k)^2}{1+\tilde{\lambda}_t(\mathbf{z}^k)} \rightarrow 0^+$  as  $k \rightarrow \infty$ , we conclude that  $\{\lambda_t(\mathbf{z}^k)\}$  also converges to zero at a quadratic rate in the **DGN** scheme.

### 3.2.5 Inexact path-following GN method (**PFGN**): the worst-case iteration-complexity

We consider the following inexact path-following generalized Newton scheme (**PFGN**) for solving (1) directly by simultaneously updating both  $\mathbf{z}$  and  $t$  at each iteration:

$$\begin{cases} t_{k+1} := (1 - \sigma_\beta)t_k \\ \mathbf{z}^{k+1} \approx \bar{\mathbf{z}}^{k+1} := \mathcal{P}_{\mathbf{z}^k}(\mathbf{z}^k - \nabla^2 F(\mathbf{z}^k)^{-1} \nabla F(\mathbf{z}^k); t_{k+1}) \end{cases}, \quad (\text{PFGN})$$

where  $\sigma_\beta \in (0, 1)$  is a given factor. As before, the approximation  $\mathbf{z}^{k+1} \approx \bar{\mathbf{z}}^{k+1}$  is in the sense of Definition 5 with a tolerance  $\delta_k \geq 0$ .

We emphasize that our PFGN scheme updates  $t$  by decreasing it at each iteration, while the standard path-following scheme in [31, 4.2.23] increases the penalty parameter at each iteration. When  $\mathcal{A}(\mathbf{z}) = \mathbf{c}$  is constant, we can define  $s := \frac{1}{t}$  to obtain the scheme [31, 4.2.23], and it allows us to start from  $s = 0$ . This is not the case in our scheme when  $\mathcal{A}(\mathbf{z}) \neq \mathbf{c}$ .

Given  $\beta \in (0, \frac{1}{2}(3 - \sqrt{5}))$ , we first find  $\sigma_\beta \in (0, 1)$  such that if  $\mathbf{z}^k \in \Omega_{t_k}(\beta)$ , then the new point  $\mathbf{z}^{k+1}$  at a new parameter  $t_{k+1}$  still satisfies  $\mathbf{z}^{k+1} \in \Omega_{t_{k+1}}(\beta)$ . The following lemma proves this key property, whose proof is deferred to “Appendix 7.6”.

**Lemma 4** *Let  $\{(\mathbf{z}^k, t_k)\}$  be the sequence generated by the inexact path-following generalized Newton scheme (PFGN). Then, for  $\mathbf{z}^k$  with  $\lambda_{t_k}(\mathbf{z}^k) < 1$ , we have*

$$\begin{aligned}\lambda_{t_{k+1}}(\mathbf{z}^k) &\leq \lambda_{t_k}(\mathbf{z}^k) + \left( \frac{\sigma_\beta}{1 - \sigma_\beta} \right) \left[ \|\nabla F(\mathbf{z}^k)\|_{\mathbf{z}^k}^* + \lambda_{t_k}(\mathbf{z}^k) \right] \\ &\leq \lambda_{t_k}(\mathbf{z}^k) + \left( \frac{\sigma_\beta}{1 - \sigma_\beta} \right) \left[ \sqrt{v} + \lambda_{t_k}(\mathbf{z}^k) \right].\end{aligned}\quad (32)$$

Let us fix  $c \in (0, 1]$ . Then, for any  $0 < \beta < 0.5(1 + 2c^2 - \sqrt{1 + 4c^2})$ , if the factor  $\sigma_\beta$  and the tolerance  $\delta_k$  are respectively chosen such that

$$\begin{aligned}0 < \sigma_\beta \leq \bar{\sigma}_\beta &:= \frac{c\sqrt{\beta} - \beta(1 + c\sqrt{\beta})}{(1 + c\sqrt{\beta})\sqrt{v} + c\sqrt{\beta}}, \quad \text{and} \\ 0 \leq \delta_k \leq \bar{\delta}_t(\beta) &:= \frac{(1 - c^2)\beta}{(1 + c\sqrt{\beta})^3 [3c\sqrt{\beta} + c^2\beta + (1 + c\sqrt{\beta})^3]},\end{aligned}\quad (33)$$

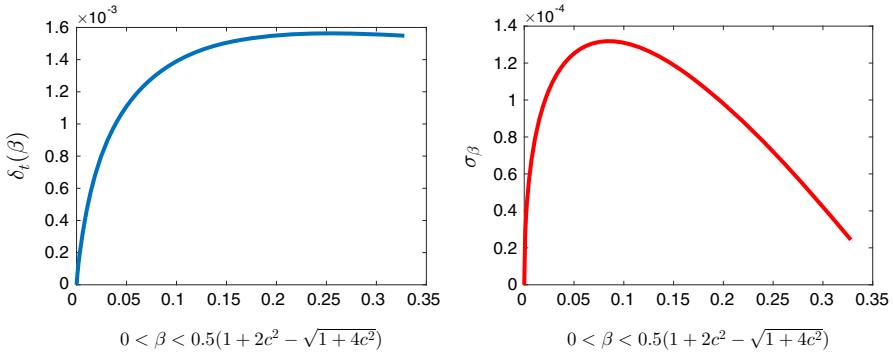
then  $\lambda_{t_k}(\mathbf{z}^k) \leq \beta$  implies  $\lambda_{t_{k+1}}(\mathbf{z}^{k+1}) \leq \beta$ . In addition,  $\lambda_{t_{k+1}}(\mathbf{z}^k) \leq \frac{c\sqrt{\beta}}{1+c\sqrt{\beta}}$ .

As an example, if we choose  $c := 0.95$ , then the possible interval for  $\beta$  is  $(0, 0.32895)$ . Now, if we choose  $\beta := \frac{1}{9c^2} \in (0, 0.32895)$  (i.e.,  $\beta \approx 0.12311$ ), then  $\bar{\sigma}_\beta = \frac{5}{36\sqrt{v} + 9}$ , which is the same as in the standard path-following method in [31]. In this case, the tolerance  $\delta_k$  for the subproblem at the second line of PFGN must be chosen such that  $0 \leq \delta_k \leq 7.45933 \times 10^{-4}$ . Figure 1 plots the values of  $\bar{\delta}_t(\beta)$  and  $\bar{\sigma}_\beta$  in (33) as a function of  $\beta$ , respectively for given  $c = 0.95$  and  $v = 1000$ . This figure shows that  $\bar{\delta}_t$  is an increasing function of  $\beta$ , while  $\bar{\sigma}_\beta$  has the maximum point at  $\beta = 0.0870$ . Hence, a good choice of  $\beta$  is  $\beta = 0.0870$ .

The following theorem investigates the worst-case iteration-complexity of PFGN using the update rule (33) for  $\sigma_\beta$ . The proof of this theorem can be found in “Appendix 7.7”.

**Theorem 4** *Let  $\{(\mathbf{z}^k, t_k)\}$  be generated by PFGN under the following configuration:*

- (i)  $c \in (0, 1]$  is given, and  $\beta$  is chosen such that  $0 < \beta < 0.5(1 + 2c^2 - \sqrt{1 + 4c^2})$ ; and



**Fig. 1** The graph of the two functions  $\bar{\delta}_t$  and  $\bar{\sigma}_\beta$  with respect to  $\beta$

(ii) the initial points  $\mathbf{z}^0$  and  $t_0 > 0$  are chosen such that  $\mathbf{z}^0 \in \text{int}(\mathcal{Z})$  and  $\lambda_{t_0}(\mathbf{z}^0) \leq \beta$ .

Then, the following conclusions hold:

- (a)  $\lambda_{t_k}(\mathbf{z}^k) \leq \beta$  for all  $k \geq 0$  (i.e.,  $\mathbf{z}^k \in \Omega_{t_k}(\beta)$  for  $k \geq 0$ ).
- (b) The number of iterations  $k$  to achieve an  $\varepsilon$ -solution  $\mathbf{z}^k$  of (1) in the sense of Definition 4 does not exceed

$$k_{\max} := \left\lfloor \left( \frac{(1 + c\sqrt{\beta})\sqrt{v} + c\sqrt{\beta}}{c\sqrt{\beta} - \beta(1 + c\sqrt{\beta})} \right) \ln \left( \frac{M_0 t_0}{\varepsilon} \right) \right\rfloor + 1,$$

where  $M_0 := \left( 1 - \frac{c\sqrt{\beta}}{1+c\sqrt{\beta}} - \bar{\delta}_t(\beta) \right)^{-1} \left( \sqrt{v} + \frac{c\sqrt{\beta}}{1+c\sqrt{\beta}} + 2\bar{\delta}_t(\beta) \right) = \mathcal{O}(\sqrt{v})$ .

- (c) Consequently, the worst-case iteration-complexity of PFGN is  $\mathcal{O}\left(\sqrt{v} \ln\left(\frac{\sqrt{v} t_0}{\varepsilon}\right)\right)$ .

Theorem 4 requires a starting point  $\mathbf{z}^0 \in \Omega_{t_0}(\beta)$  at a given penalty parameter  $t_0 > 0$ . In order to find  $\mathbf{z}^0$ , we perform an initial phase (called Phase 1) as described below.

### 3.2.6 Finding an initial point with the path-following iterations using auxiliary problem

When  $\mathcal{A}(\cdot) = \partial H(\cdot)$  the subgradient of a proper, closed, and convex function  $H$ , we can find  $\mathbf{z}^0 \in \Omega_{t_0}(\beta)$  for PFGN by applying the [inexact] damped-step generalized Newton scheme (FGN) to solve (2) for a fixed penalty parameter  $t = t_0$ . This scheme has a sublinear convergence rate [51]. However, it is still unclear how to generalize this method to (1).

We instead propose a new auxiliary problem for (2), and then apply PFGN to solve this auxiliary problem in order to obtain  $\mathbf{z}^0$ . Then, we estimate the maximum number of the path-following iterations needed in this phase.

Let us fix some  $\hat{\mathbf{z}}^0 \in \text{int}(\mathcal{Z})$ . We first compute a vector  $\hat{\xi}^0 \in \mathcal{A}(\hat{\mathbf{z}}^0)$  and evaluate  $\nabla F(\hat{\mathbf{z}}^0)$ . Then, we define  $\hat{\zeta}^0 := t_0 \nabla F(\hat{\mathbf{z}}^0) + \hat{\xi}^0$ , and consider the following auxiliary problem of (2):

$$\text{Find } \hat{\mathbf{z}}_\tau^* \in \text{int}(\mathcal{Z}) \text{ such that: } 0 \in t_0 \nabla F(\hat{\mathbf{z}}_\tau^*) - \tau \hat{\xi}^0 + \mathcal{A}(\hat{\mathbf{z}}_\tau^*), \quad (34)$$

where  $\tau \in [0, 1]$  is a new homotopy parameter. Clearly, (34) has a similar form as (2).

- (a) When  $\tau = 1$ , we have  $0 \in t_0 \nabla F(\hat{\mathbf{z}}^0) - \hat{\xi}^0 + \mathcal{A}(\hat{\mathbf{z}}^0)$  due to the choice of  $\hat{\xi}^0$ . Hence,  $\hat{\mathbf{z}}^0$  is an exact solution of (34) at  $\tau = 1$ .
- (b) When  $\tau = 0$ , we have  $0 \in t_0 \nabla F(\hat{\mathbf{z}}_\tau^*) + \mathcal{A}(\hat{\mathbf{z}}_\tau^*)$ . Hence, any solution  $\hat{\mathbf{z}}_\tau^*$  of (34) is also a solution of (2) at  $t = t_0$ .

Now, we can apply PFGN to solve (34) starting from  $\tau_0 = 1$  but using a different update rule for  $\tau$ . More precisely, this scheme can be written as follows:

$$\begin{cases} \tau_{j+1} := \tau_j - \Delta_j, \\ \hat{\mathbf{z}}^{j+1} \approx \bar{\mathbf{z}}^{j+1} := \mathcal{P}_{\hat{\mathbf{z}}^j} \left( \hat{\mathbf{z}}^j - \nabla^2 F(\hat{\mathbf{z}}^j)^{-1} \left( \nabla F(\hat{\mathbf{z}}^j) - \tau_{j+1} t_0^{-1} \hat{\xi}^0 \right); t_0 \right), \end{cases} \quad (35)$$

where  $\Delta_j > 0$  is a given decrement. Here, the approximation  $\hat{\mathbf{z}}^{j+1} \approx \bar{\mathbf{z}}^{j+1}$  is in the sense of Definition 5 with a given tolerance  $\hat{\delta}_j \geq 0$ . We also use the index  $j$  to distinguish with the index  $k$  in PFGN, and using the notation “hat” for the iterate vectors.

Similar to (21), we define the following generalized Newton decrement for (34):

$$\hat{\lambda}_\tau(\hat{\mathbf{z}}) := \left\| \hat{\mathbf{z}} - \mathcal{P}_{\hat{\mathbf{z}}} \left( \hat{\mathbf{z}} - \nabla^2 F(\hat{\mathbf{z}})^{-1} \left( \nabla F(\hat{\mathbf{z}}) - \tau t_0^{-1} \hat{\xi}^0 \right); t_0 \right) \right\|_{\hat{\mathbf{z}}}. \quad (36)$$

The theorem below provides the number of iterations  $j$  needed to find an initial point  $\mathbf{z}^0 \in \Omega_{t_0}(\beta)$  for PFGN using (35), whose proof can be found in “Appendix 7.8”.

**Theorem 5** Let  $c \in (0, 1]$  and  $\beta$  be chosen as in Theorem 4, and  $\eta$  be chosen such that  $0 < \eta < \beta$ . Let  $\{(\hat{\mathbf{z}}^j, \tau_j)\}$  be generated by (35). If  $\Delta_j$  and  $\hat{\delta}_j$  are chosen such that

$$\begin{aligned} 0 \leq \Delta_j &\leq \frac{\bar{\mu}_\eta}{\|\hat{\xi}_0\|_{\hat{\mathbf{z}}^j}^*} \quad \text{with } \bar{\mu}_\eta := \frac{t_0}{\|\hat{\xi}_0\|_{\hat{\mathbf{z}}^j}^*} \left( \frac{c\sqrt{\eta}}{1+c\sqrt{\eta}} - \eta \right), \quad \text{and} \\ 0 \leq \hat{\delta}_j &\leq \bar{\delta}_\tau(\eta) := \frac{(1-c^2)\eta}{(1+c\sqrt{\eta})^3 [3c\sqrt{\eta} + c^2\eta + (1+c\sqrt{\eta})^3]}, \end{aligned} \quad (37)$$

then  $\hat{\lambda}_{\tau_j}(\hat{\mathbf{z}}^j)$  defined in (36) satisfies  $\hat{\lambda}_{\tau_j}(\hat{\mathbf{z}}^j) \leq \eta$  for all  $j \geq 0$ .

Let  $\mathbf{z}^0 := \hat{\mathbf{z}}^{j_{\max}}$  be obtained after  $j_{\max}$  iterations. Then,  $\hat{\lambda}_{\tau_0}(\hat{\mathbf{z}}^0) = 0$  and we have

$$\lambda_{t_0}(\mathbf{z}^0) \leq \hat{\lambda}_{\tau_j}(\hat{\mathbf{z}}^j) + t_0^{-1} \tau_j \|\hat{\xi}^0\|_{\hat{\mathbf{z}}^j}^* \leq \eta + \frac{\kappa \|\hat{\xi}^0\|_{\bar{\mathbf{z}}_F^*}^*}{t_0} - j \left( \frac{c\sqrt{\eta}}{1+c\sqrt{\eta}} - \eta \right), \quad \forall j \geq j_{\max}, \quad (38)$$

where  $\lambda_t(\mathbf{z})$  is defined by (21), and  $\bar{\mathbf{z}}_F^*$  and  $\kappa$  are defined by (15) and below (15), respectively.

The number of iterations  $j$  to achieve  $\mathbf{z}^0 := \hat{\mathbf{z}}^j$  such that  $\lambda_{t_0}(\mathbf{z}^0) \leq \beta$  does not exceed

$$j_{\max} := \left\lfloor \frac{\kappa(1 + c\sqrt{\eta})\|\hat{\zeta}^0\|_{\hat{\mathbf{z}}_F^*}^*}{t_0(c\sqrt{\eta} - \eta(1 + c\sqrt{\eta}))} - \frac{(\beta - \eta)(1 + c\sqrt{\eta})}{c\sqrt{\eta} - \eta(1 + c\sqrt{\eta})} \right\rfloor + 1.$$

The worst-case iteration-complexity of (35) to obtain  $\mathbf{z}^0$  such that  $\lambda_{t_0}(\mathbf{z}^0) \leq \beta$  is  $\mathcal{O}\left(\frac{\kappa\|\hat{\zeta}^0\|_{\hat{\mathbf{z}}_F^*}^*}{t_0}\right)$ .

Theorem 5 suggests us to choose  $t_0 := \kappa$ . In this case, the maximum number of iterations in Phase 1 does not exceed  $\frac{(1+c\sqrt{\eta})\|\hat{\zeta}^0\|_{\hat{\mathbf{z}}_F^*}^*}{c\sqrt{\eta}-\eta(1+c\sqrt{\eta})}$ , which is a constant.

*Remark 1* From (38), we can compute  $\|\hat{\zeta}^0\|_{\hat{\mathbf{z}}_F^*}^*$  directly in order to terminate (35) by checking  $\tau_j\|\hat{\zeta}^0\|_{\hat{\mathbf{z}}_F^*}^* \leq t_0(\beta - \eta)$ . Hence, (35) does not require to evaluate the analytical center  $\hat{\mathbf{z}}_F^*$  of  $F$ . If  $F$  is a self-concordant logarithmically homogeneous barrier, then we can simply choose  $t_0 := 1$ . Otherwise, we can choose  $t_0 := v + 2\sqrt{v}$ .

### 3.2.7 Two-phase inexact path-following generalized Newton algorithm

Putting two schemes (35) and PFGN together, we obtain a two-phase inexact path-following generalized Newton algorithm for solving (1) as described in Algorithm 1.

The main computational cost of Algorithm 1 is the solution of the two linear monotone inclusions in (35) and PFGN, respectively. When  $\mathcal{A} = \partial H$  the subdifferential of a convex function  $H$ , various methods including fast gradient, primal-dual methods, and splitting techniques can be used to solve these problems [3, 4, 7, 15, 18, 34].

The overall worst-case iteration-complexity of Algorithm 1 is given in the following theorem which is a direct consequence of Lemma 3, Theorems 4 and 5.

**Theorem 6** Let us choose  $t_0 := \kappa$  as defined below (15). Then, the overall worst-case iteration-complexity of Algorithm 1 to achieve an  $\varepsilon$ -solution  $\mathbf{z}^k$  of (1) as in Definition 4 is

$$\mathcal{O}\left(\frac{\kappa\|\hat{\zeta}^0\|_{\hat{\mathbf{z}}_F^*}^*}{t_0} + \sqrt{v}\ln\left(\frac{M_0 t_0}{\varepsilon}\right)\right) \quad \left(\text{or simpler } \mathcal{O}\left(\sqrt{v}\ln\left(\frac{\kappa\sqrt{v}}{\varepsilon}\right)\right)\right),$$

where  $t_0 > 0$  is an initial penalty parameter and  $M_0 = \mathcal{O}(\sqrt{v})$  is defined in Theorem 4.

*Proof* The total number of iterations requires in Phase 1 and Phase 2 of Algorithm 1 is

$$K_{\max} \geq \frac{\kappa(1 + c\sqrt{\eta})\|\hat{\zeta}^0\|_{\hat{\mathbf{z}}_F^*}^*}{t_0(c\sqrt{\eta} - \eta(1 + c\sqrt{\eta}))} + C_2 \ln\left(\frac{M_0 t_0}{\varepsilon}\right) - C_1.$$

**Algorithm 1** (*Two-phase inexact path-following generalized Newton algorithm*)

- 1: **Initialization:**
- 2: Choose an arbitrary initial point  $\hat{\mathbf{z}}^0 \in \text{int}(\mathcal{Z})$  and a desired accuracy  $\varepsilon > 0$ .
- 3: Fix  $t_0 > 0$  and  $c \in (0, 1]$  (e.g.,  $t_0 := \kappa$ , and  $c := 0.95$ ).
- 4: Compute  $\hat{\xi}^0 \in \mathcal{A}(\hat{\mathbf{z}}^0)$  and evaluate  $\nabla F(\hat{\mathbf{z}}^0)$ . Set  $\hat{\zeta}^0 := t_0 \nabla F(\hat{\mathbf{z}}^0) + \hat{\xi}^0$  and  $\tau_0 := 1$ .
- 5: Fix  $\beta$  as in Theorem 4 (e.g.,  $\beta := \frac{1}{9c^2}$ ) and choose  $\eta < \beta$  (e.g.,  $\eta := 0.5\beta$ ).
- 6: Compute  $\bar{\delta}_\tau$ ,  $\bar{\mu}_\eta$ ,  $\bar{\delta}_t$ , and  $\bar{\sigma}_\beta$  by (37) and (33), respectively, and  $M_0$  from Theorem 4.

**7: Phase 1: Computing an initial point by path-following iterations**

- 8: **For**  $j = 0, \dots, j_{\max}$ , **perform:**
- 9: If  $\tau_j \|\hat{\zeta}^0\|_{\hat{\mathbf{z}}^j}^* \leq t_0(\beta - \eta)$ , then set  $\mathbf{z}^0 := \hat{\mathbf{z}}^j$  and TERMINATE.
- 10: Update  $(\hat{\mathbf{z}}^{j+1}, \tau_{j+1})$  by (35) with  $\Delta_j := \frac{\bar{\mu}_\eta}{\|\hat{\zeta}^0\|_{\hat{\mathbf{z}}^j}^*}$  up to an accuracy  $\hat{\delta}_j \leq \bar{\delta}_\tau$ .
- 11: **End for**

**12: Phase 2: Inexact path-following generalized Newton iterations**

- 13: **For**  $k = 0, \dots, k_{\max}$ , **perform:**
- 14: If  $M_0 t_k \leq \varepsilon$ , then return  $\mathbf{z}^k$  as an  $\varepsilon$ -solution of (1), and TERMINATE.
- 15: Update  $(\mathbf{z}^{k+1}, t_{k+1})$  by (PFGN) up to an accuracy  $\delta_k \leq \bar{\delta}_t$ .
- 16: **End for**

where  $C_1 := \frac{(\beta-\eta)(1+c\sqrt{\eta})}{c\sqrt{\eta}-\eta(1+c\sqrt{\eta})}$ , and  $C_2 := \left( \frac{(1+c\sqrt{\beta})\sqrt{\nu}+c\sqrt{\beta}}{c\sqrt{\beta}-\beta(1+c\sqrt{\beta})} \right)$ . Note that  $C_1$  is a constant, while  $C_2 = \mathcal{O}(\sqrt{\nu})$ . Hence,  $K_{\max} \geq C_3 \frac{\kappa \|\hat{\zeta}^0\|_{\hat{\mathbf{z}}_F^*}^*}{t_0} + \mathcal{O}(\sqrt{\nu}) \ln \left( \frac{M_0 t_0}{\varepsilon} \right) - C_1$ , where  $C_3 := \frac{1+c\sqrt{\eta}}{c\sqrt{\eta}-\eta(1+c\sqrt{\eta})}$ . We can write this as  $K_{\max} \geq \mathcal{O} \left( \frac{\kappa \|\hat{\zeta}^0\|_{\hat{\mathbf{z}}_F^*}^*}{t_0} + \sqrt{\nu} \ln \left( \frac{M_0 t_0}{\varepsilon} \right) \right)$ .

We finally note that  $M_0 = \mathcal{O}(\sqrt{\nu})$  and  $t_0 = \kappa$ , and the first term is a constant and independent of  $\varepsilon$ , which is dominated by the second term. Hence, we obtain the second estimate of Theorem 6.  $\square$

The complexity bound in Theorem 6 also depends on the choice of  $\beta$ ,  $\eta$  and  $t_0$ . Adjusting these parameters allows us to trade-off between Phase 1 and Phase 2 in Algorithm 1. Clearly, if  $t_0$  is large, the number of iterations required in Phase 1 is small, but the number of iterations in Phase 2 is large, and vice versa.

*Remark 2* Note that we can recover the convergence guarantee of the exact generalized Newton-type schemes as consequences of Theorems 2, 3 and 6, respectively. For instance, in the exact variant of PFGN, if we can choose  $\beta \in (0, 0.5(3 - \sqrt{5}))$ , then the upper bound  $\bar{\sigma}_\beta$  in (33) reduces to  $\bar{\sigma}_\beta := \frac{\sqrt{\beta}-\beta(1+\sqrt{\beta})}{(1+\sqrt{\beta})\sqrt{\nu}+\sqrt{\beta}}$ . Hence, we can show that the worst-case iteration-complexity estimate of this exact scheme coincides with the standard path-following scheme for smooth structural convex programming given in [31, Theorem 4.2.9].

## 4 Inexact path-following proximal Newton algorithms

We now specify our framework, Algorithm 1, to solve three problems: (3), (4), and (8).

### 4.1 Inexact primal-dual path-following algorithm for saddle-point problems

We recall the convex-concave saddle-point problem (8). Our primal-dual path-following proximal Newton method relies on the following assumption.

- Assumption A.2** (a) The feasible set  $\mathcal{X}$  (respectively,  $\mathcal{Y}$ ) is a nonempty, closed, and convex cone with nonempty interior, and is endowed with a  $v_f$ -self-concordant barrier  $f$  (respectively, a  $v_\varphi$ -self-concordant barrier  $\varphi$ ) such that  $\text{Dom}(f) = \mathcal{X}$  (respectively,  $\text{Dom}(\varphi) = \mathcal{Y}$ ).  
(b) Both  $g$  and  $\psi$  in (8) are proper, closed, and convex such that  $\text{int}(\mathcal{X}) \cap \text{dom}(g) \neq \emptyset$  and  $\text{int}(\mathcal{Y}) \cap \text{dom}(\psi) \neq \emptyset$ .  
(c) The solution set  $\mathcal{Z}^*$  of (8) is nonempty.

For any  $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ ,  $\hat{\mathbf{z}} = (\hat{\mathbf{x}}, \hat{\mathbf{y}})$ ,  $(\xi_g, \xi_\psi) \in \partial g(\mathbf{x}) \times \partial \psi(\mathbf{y})$ , and  $(\hat{\xi}_g, \hat{\xi}_\psi) \in \partial g(\hat{\mathbf{x}}) \times \partial \psi(\hat{\mathbf{y}})$ :

$$\begin{pmatrix} \xi_g - L^* \mathbf{y} - \hat{\xi}_g + L^* \hat{\mathbf{y}} \\ \xi_\psi + L \mathbf{x} - \hat{\xi}_\psi - L \hat{\mathbf{x}} \end{pmatrix}^\top \begin{pmatrix} \mathbf{x} - \hat{\mathbf{x}} \\ \mathbf{y} - \hat{\mathbf{y}} \end{pmatrix} \geq 0.$$

This shows that  $\mathcal{A}$  defined by (10) is maximally monotone. In addition,  $F$  is a self-concordant barrier of  $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$  with the barrier parameter  $v := v_f + v_\varphi$ .

#### 4.1.1 Inexact primal-dual path-following proximal Newton method

We specify PFGN to solve (8). Let  $\mathbf{z}^k := (\mathbf{x}^k, \mathbf{y}^k) \in \text{int}(\mathcal{Z})$  be a given point at  $t_k > 0$ . We update  $\mathbf{z}^{k+1} := (\mathbf{x}^{k+1}, \mathbf{y}^{k+1})$  and  $t_{k+1}$  using PFGN, which reduces to the following form:

$$\begin{cases} 0 \in t_{k+1} [\nabla f(\mathbf{x}^k) + \nabla^2 f(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k)] - L^* \mathbf{y} + \partial g(\mathbf{x}), \\ 0 \in t_{k+1} [\nabla \varphi(\mathbf{y}^k) + \nabla^2 \varphi(\mathbf{y}^k)(\mathbf{y} - \mathbf{y}^k)] + L \mathbf{x} + \partial \psi(\mathbf{y}). \end{cases} \quad (39)$$

Here, we approximately solve (39) as done in PFGN. Hence, PFGN can be rewritten as

$$\begin{cases} t_{k+1} := (1 - \sigma_\beta) t_k, \\ \mathbf{z}^{k+1} \approx \bar{\mathbf{z}}^{k+1} := \underset{\mathbf{y}}{\text{argmin}} \underset{\mathbf{x}}{\max} \left\{ t_{k+1} Q_\varphi(\mathbf{y}; \mathbf{y}^k) + \psi(\mathbf{y}) + \langle L \mathbf{x}, \mathbf{y} \rangle - t_{k+1} Q_f(\mathbf{x}; \mathbf{x}^k) - g(\mathbf{x}) \right\}, \end{cases} \quad (40)$$

where  $Q_f(\cdot; \mathbf{x}^k)$  and  $Q_\varphi(\cdot; \mathbf{y}^k)$  are quadratic surrogates of  $f$  and  $\varphi$ , respectively, i.e.:

$$\begin{cases} Q_f(\mathbf{x}; \mathbf{x}^k) := \langle \nabla f(\mathbf{x}^k), \mathbf{x} - \mathbf{x}^k \rangle + \frac{1}{2} \langle \nabla^2 f(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k), \mathbf{x} - \mathbf{x}^k \rangle, \\ Q_\varphi(\mathbf{y}; \mathbf{y}^k) := \langle \nabla \varphi(\mathbf{y}^k), \mathbf{y} - \mathbf{y}^k \rangle + \frac{1}{2} \langle \nabla^2 \varphi(\mathbf{y}^k)(\mathbf{y} - \mathbf{y}^k), \mathbf{y} - \mathbf{y}^k \rangle. \end{cases} \quad (41)$$

The second line of (40) is again a linear convex–concave saddle-point problem with strongly convex objectives. Methods for solving this problem can be found, e.g., in [2, 9, 15].

#### 4.1.2 Finding an initial point

We specify (35) for finding an initial point as follows.

1. Provide a value  $t_0 > 0$  (e.g.,  $t_0 := \kappa$ ), and an initial point  $\hat{\mathbf{z}}^0 := (\hat{\mathbf{x}}^0, \hat{\mathbf{y}}^0) \in \text{int}(\mathcal{Z})$ .
2. Compute a subgradient  $\hat{\xi}_g^0 \in \partial g(\hat{\mathbf{x}}^0)$  and  $\hat{\xi}_\psi^0 \in \partial \psi(\hat{\mathbf{y}}^0)$ , and evaluate  $\nabla f(\hat{\mathbf{x}}^0)$  and  $\nabla \varphi(\hat{\mathbf{y}}^0)$ .
3. Define  $\hat{\zeta}_g^0 := t_0 \nabla f(\hat{\mathbf{x}}^0) - L^* \hat{\mathbf{y}}^0 + \hat{\xi}_g^0$  and  $\hat{\zeta}_\psi^0 := t_0 \nabla \varphi(\hat{\mathbf{y}}^0) + L \hat{\mathbf{x}}^0 + \hat{\xi}_\psi^0$ .
4. Perform Phase 1 of Algorithm 1 applied to solve (8) as follows:

$$\begin{cases} \tau_{j+1} := \tau_j - \bar{\Delta}_j, \\ \hat{\mathbf{z}}^{j+1} \approx \bar{\mathbf{z}}^{j+1} := \arg \min_{\mathbf{y}} \max_{\mathbf{x}} \left\{ t_0 Q_\varphi(\mathbf{y}; \hat{\mathbf{y}}^j) - \tau_{j+1} \langle \hat{\zeta}_\psi^0, \mathbf{y} \rangle + \psi(\mathbf{y}) + \langle L \mathbf{x}, \mathbf{y} \rangle \right. \\ \quad \left. - t_0 Q_f(\mathbf{x}; \hat{\mathbf{x}}^j) + \tau_{j+1} \langle \hat{\zeta}_g^0, \mathbf{x} \rangle - g(\mathbf{x}) \right\}. \end{cases} \quad (42)$$

Here,  $\tau \in (0, 1]$  is referred to as a new homotopy parameter starting from  $\tau_0 := 1$ .

Now, we substitute this scheme into Phase 1, and (40) into Step 15 of Algorithm 1, respectively, to obtain a new variant to solve (8), which we call Algorithm 1(a).

The worst-case iteration-complexity of Algorithm 1(a) to achieve an  $\varepsilon$ -primal-dual solution  $\mathbf{z}^k := (\mathbf{x}^k, \mathbf{y}^k)$  in the sense of Definition 4 for the optimality condition (9) instead of (1) is guaranteed by Theorem 6. We omit the detailed proof in this paper.

## 4.2 Inexact path-following primal proximal Newton algorithm

We present an inexact primal path-following proximal Newton method obtained from Algorithm 1 to solve (3). This algorithm has several new features compared to [51, 52].

First, associated with the barrier function  $f$  of  $\mathcal{X}$  in (3), we define a local norm  $\|\mathbf{u}\|_{\mathbf{x}} := \langle \nabla^2 f(\mathbf{x}) \mathbf{u}, \mathbf{u} \rangle^{1/2}$  and its corresponding dual norm  $\|\mathbf{v}\|_{\mathbf{x}}^* := \langle \nabla^2 f(\mathbf{x})^{-1} \mathbf{v}, \mathbf{v} \rangle^{1/2}$  for a given  $\mathbf{x} \in \text{dom}(f)$ . Next, let  $Q_f$  be the quadratic surrogate of  $f$  around  $\mathbf{x}^k$  as defined in (41). Then, the main step of Algorithm 1 applied to (3) performs the following inexact path-following proximal Newton scheme:

$$\begin{cases} t_{k+1} := (1 - \sigma_\beta) t_k, \\ \mathbf{x}^{k+1} \approx \bar{\mathbf{x}}^{k+1} := \operatorname{argmin}_{\mathbf{x}} \left\{ h_k(\mathbf{x}) := Q_f(\mathbf{x}; \mathbf{x}^k) + t_{k+1}^{-1} g(\mathbf{x}) \right\}, \end{cases} \quad (43)$$

Here, the approximation  $\approx$  is in the sense of Definition 5, and implies  $\|\mathbf{x}^{k+1} - \bar{\mathbf{x}}^{k+1}\|_{\mathbf{x}^k} \leq \delta_k$  for a given tolerance  $\delta_k \geq 0$ . As shown in [51], this condition is satisfied if

$$h_k(\mathbf{x}^{k+1}) - h_k(\bar{\mathbf{x}}^{k+1}) \leq 0.5 \delta_k^2,$$

where  $h_k(\cdot)$  is the objective function of (43). This condition is different from Definition 5, where we can check it by evaluating the objective values.

We redefine the following proximal Newton decrement using (14) as follows:

$$\lambda_t(\mathbf{x}) := \left\| \mathbf{x} - \text{prox}_{(t\nabla^2 f(\mathbf{x}))^{-1} g} \left( \mathbf{x} - \nabla^2 f(\mathbf{x})^{-1} \nabla f(\mathbf{x}) \right) \right\|_{\mathbf{x}}. \quad (44)$$

Although the scheme (43) has been studied in [51, 52], the following features are new.

1. *Phase 1: Finding an initial point  $\mathbf{x}^0$*  We solve the following auxiliary problem by applying (35) to obtain an initial point  $\mathbf{x}^0 \in \text{int}(\mathcal{X})$  such that  $\lambda_{t_0}(\mathbf{x}^0) \leq \beta$ :

$$\min_{\mathbf{x}} \left\{ f(\mathbf{x}) - \tau \langle \nabla f(\hat{\mathbf{x}}^0) + t_0^{-1} \hat{\xi}^0, \mathbf{x} \rangle + t_0^{-1} g(\mathbf{x}) \right\}, \quad (45)$$

where  $\hat{\mathbf{x}}^0 \in \text{int}(\mathcal{X})$  is an arbitrary initial point, and  $\hat{\xi}^0 \in \partial g(\hat{\mathbf{x}}^0)$ . The inexact proximal path-following scheme for solving (45) rendering from (35) becomes

$$\begin{cases} \tau_{j+1} := \tau_j - \bar{\Delta}_j, \\ \hat{\mathbf{x}}^{j+1} \approx \tilde{\hat{\mathbf{x}}}^{j+1} := \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ Q_f(\mathbf{x}; \hat{\mathbf{x}}^j) - \tau_{j+1} \langle \nabla f(\hat{\mathbf{x}}^0) + t_0^{-1} \hat{\xi}^0, \mathbf{x} \rangle + t_0^{-1} g(\mathbf{x}) \right\}. \end{cases} \quad (46)$$

2. *New neighborhood of the central path* We choose  $\beta \in (0, 0.329)$ , which is approximately twice larger than  $\beta \in (0, 0.15)$  as given in [51].
3. *Adaptive rule for  $t$*  We can update  $t_k$  in (43) adaptively using the value  $\|\nabla f(\mathbf{x}^k)\|_{\mathbf{x}^k}^*$  as

$$t_{k+1} := (1 - \sigma_k) t_k, \quad \text{where } \sigma_k := \frac{c\sqrt{\beta} - \beta(1 + c\sqrt{\beta})}{(1 + c\sqrt{\beta})\|\nabla f(\mathbf{x}^k)\|_{\mathbf{x}^k}^* + c\sqrt{\beta}} \geq \bar{\sigma}_\beta.$$

4. *Implementable stopping condition* We can terminate Phase 1 using  $\tau_j \|\nabla f(\hat{\mathbf{x}}^0) + t_0^{-1} \hat{\xi}^0\|_{\hat{\mathbf{x}}^j}^* \leq (\beta - \eta)$ , which is implementable without incurring significantly computational cost.

Let us denote this algorithmic variant by Algorithm 1(b). Instead of terminating this algorithmic variant with  $t_k \leq \frac{\varepsilon}{M_0}$ , we use  $\Delta(\beta, \nu)t_k \leq \varepsilon$  to terminate Algorithm 1(b), where  $\Delta(\beta, \nu)$  is a function defined as in [51, Lemma 5.1.]. The following corollary provides the worst-case iteration-complexity of Algorithm 1(b) as a direct consequence of Theorem 6.

**Corollary 1** *Let us choose  $t_0 := \kappa$  defined below the formula (15). Then, the worst-case iteration-complexity of Algorithm 1(b) to achieve an  $\varepsilon$ -solution  $\mathbf{x}^k$  of (3) such that  $\mathbf{x}^k \in \mathcal{X}$  and  $g(\mathbf{x}^k) - g^* \leq \varepsilon$  is*

$$\mathcal{O} \left( \frac{\kappa \|\nabla f(\hat{\mathbf{x}}^0) + t_0^{-1} \hat{\xi}^0\|_{\hat{\mathbf{x}}^*}^*}{t_0} + \sqrt{\nu} \ln \left( \frac{\Delta(\beta, \nu)t_0}{\varepsilon} \right) \right) \quad \left( \text{or simpler } \mathcal{O} \left( \sqrt{\nu} \ln \left( \frac{\kappa \nu}{\varepsilon} \right) \right) \right).$$

Note that the worst-case iteration-complexity bound in Corollary 1 is the overall iteration-complexity. It is similar to the one given in [52] but the method is different.

### 4.3 Inexact dual path-following proximal Newton algorithm

We develop an inexact dual path-following scheme to solve (4), which works in the dual space. For simplicity of presentation, we assume that  $W = \mathbb{I}$ . Otherwise, we can use  $\hat{g}(\cdot) := g(W(\cdot))$ . We first write the barrier formulation of the dual problem (5) as follows:

$$\min_{\mathbf{y} \in \mathbb{R}^p} \{t f^*(\mathbf{c} - L^* \mathbf{y}) + g^*(\mathbf{y}) + \langle \mathbf{b}, \mathbf{y} \rangle\},$$

where  $t > 0$  is a penalty parameter. This problem can shortly read as

$$\phi_t^* := \min_{\mathbf{y} \in \mathbb{R}^p} \{\phi_t(\mathbf{y}) := \varphi(\mathbf{y}) + t^{-1}\psi(\mathbf{y})\}, \quad (47)$$

where  $\varphi$  and  $\psi$  are two convex functions defined by

$$\varphi(\mathbf{y}) := f^*(\mathbf{c} - L^* \mathbf{y}), \quad \text{and} \quad \psi(\mathbf{y}) := g^*(\mathbf{y}) + \langle \mathbf{b}, \mathbf{y} \rangle. \quad (48)$$

In order to characterize the relation between the primal problem (4) and its dual form (5), we formally impose the following assumption.

**Assumption A.3** The objective function  $g$  in (4) is proper, closed, and convex. The linear operator  $L : \mathbb{R}^n \rightarrow \mathbb{R}^p$  is full-row rank with  $n \leq p$ . The following Slater condition holds:

$$(\text{int } (\mathcal{K}) \times \text{ri}(\text{dom}(g))) \cap \{(\mathbf{x}, \mathbf{s}) \mid L\mathbf{x} - \mathbf{s} = \mathbf{b}\} \neq \emptyset.$$

In addition,  $\mathcal{K}$  is a nonempty, closed, and pointed convex cone such that  $\text{int } (\mathcal{K}) \neq \emptyset$ , and  $\mathcal{K}$  is endowed with a  $v$ -self-concordant logarithmically homogeneous barrier  $f$  with  $\text{Dom}(f) = \mathcal{K}$ . The solution set  $\mathcal{S}^*$  of (4) is nonempty.

The following lemma shows that  $\varphi(\cdot) := f^*(\mathbf{c} - L^*(\cdot))$  remains a  $v$ -self-concordant barrier associated with the dual feasible set, while the scaled proximal operator of  $\psi$  can be computed from the one of  $g$ . The proof of this lemma is classical and is omitted, see [2, 35].

**Lemma 5** *Under Assumption A.3,  $\varphi(\cdot)$  defined by (48) is a  $v$ -self-concordant barrier of the dual feasible set  $\mathcal{D}_Y := \{\mathbf{y} \in \mathbb{R}^p \mid L^* \mathbf{y} - \mathbf{c} \in \mathcal{K}^*\}$ . The proximal operator of  $\psi$  defined in (47) is computed as  $\text{prox}_{\mathbf{Q}\psi}(\mathbf{y}) = \mathbf{y} - \mathbf{Q}\mathbf{b} - \mathbf{Q}\text{prox}_{\mathbf{Q}^{-1}g}(\mathbf{Q}^{-1}\mathbf{y} - \mathbf{b})$  for any  $\mathbf{Q} \in \mathcal{S}_{++}^p$ .*

Together with the primal local norm  $\|\cdot\|_x$  given in Sect. 4.2, we also define a local norm with respect to  $\varphi(\cdot)$  as  $\|\mathbf{u}\|_Y := \langle \nabla^2 \varphi(\mathbf{y}) \mathbf{u}, \mathbf{u} \rangle^{1/2}$  and its dual norm  $\|\mathbf{v}\|_Y^* := \langle \nabla^2 \varphi(\mathbf{y})^{-1} \mathbf{v}, \mathbf{v} \rangle^{1/2}$  for a given  $\mathbf{y} \in \text{dom}(\varphi)$ . Under Assumption A.3, any primal-dual

solution  $(\mathbf{x}^*, \mathbf{s}^*) \in \mathcal{S}^*$  and  $\mathbf{y}^* \in \mathbb{R}^p$  of (4) is also the KKT point of (4) and vice versa, i.e.:

$$0 \in L^* \mathbf{y}^* - \mathbf{c} + \mathcal{N}_{\mathcal{K}}(\mathbf{x}^*), \quad \mathbf{y}^* \in \partial g(\mathbf{s}^*), \quad \text{and } L \mathbf{x}^* - \mathbf{s}^* = \mathbf{b}. \quad (49)$$

In practice, we cannot solve (4) and (5) (or equivalently, (49)) exactly to obtain an optimal solution  $(\mathbf{x}^*, \mathbf{s}^*) \in \mathcal{S}^*$  and  $\mathbf{y}^* \in \mathbb{R}^p$  as indicated by the KKT condition (49). We can only find an  $\varepsilon$ -approximate solution  $(\mathbf{x}_\varepsilon^*, \mathbf{s}_\varepsilon^*)$  and  $\mathbf{y}_\varepsilon^*$  as defined in Definition 6.

**Definition 6** Given a tolerance  $\varepsilon > 0$ , we say that  $(\mathbf{x}_\varepsilon^*, \mathbf{s}_\varepsilon^*)$  is an  $\varepsilon$ -solution for (4) associated with an  $\varepsilon$ -dual solution  $\mathbf{y}_\varepsilon^* \in \mathbb{R}^p$  of (5) if  $\mathbf{x}_\varepsilon^* \in \text{int}(\mathcal{K})$  and

$$\begin{cases} \mathbf{y}_\varepsilon^* \in \partial g(\mathbf{s}_\varepsilon^*), \\ \|L^* \mathbf{y}_\varepsilon^* - \mathbf{c}\|_{\mathbf{x}_\varepsilon^*}^* \leq \varepsilon, \\ \|L \mathbf{x}_\varepsilon^* - \mathbf{s}_\varepsilon^* - \mathbf{b}\|_{\mathbf{y}_\varepsilon^*}^* \leq \varepsilon. \end{cases}$$

Note that our path-following method always generates  $\mathbf{x}_\varepsilon^* \in \text{int}(\mathcal{K})$ , which implies  $\mathbf{x}_\varepsilon^* \in \mathcal{K}$ .

Next, we specify Algorithm 1 to solve the dual problem (5) and provide a recovery strategy to obtain an  $\varepsilon$ -solution of (4).

#### 4.3.1 The inexact path-following proximal Newton scheme for the dual problem

By Lemma 5, the function  $\varphi$  defined by (48) is also self-concordant, and its gradient and Hessian-vector product are given explicitly as

$$\nabla \varphi(\mathbf{y}) = -L \nabla f^*(\mathbf{c} - L^* \mathbf{y}) \quad \text{and} \quad \nabla^2 \varphi(\mathbf{y}) \mathbf{d} = L \nabla^2 f^*(\mathbf{c} - L^* \mathbf{y}) L^* \mathbf{d}. \quad (50)$$

Let us denote by  $Q_\varphi$  the quadratic surrogate of  $\varphi$  at  $\mathbf{y}^k$  defined by (41). Under Assumption A.3,  $\nabla^2 \varphi$  is positive definite and hence,  $Q_\varphi(\cdot; \mathbf{y}^k)$  is strongly convex. The main step of our inexact dual path-following proximal Newton method can be presented as follows:

$$\begin{cases} t_{k+1} := (1 - \sigma_\beta) t_k, \\ \mathbf{y}^{k+1} \approx \bar{\mathbf{y}}^{k+1} := \underset{\mathbf{y} \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ \phi(\mathbf{y}; \mathbf{y}^k) := Q_\varphi(\mathbf{y}; \mathbf{y}^k) + t_{k+1}^{-1} \psi(\mathbf{y}) \right\}, \end{cases} \quad (51)$$

where the approximation  $\approx$  is defined as in Definition 5 with a given tolerance  $\delta_k$ , and  $\sigma_\beta \in (0, 1)$  is a given factor.

The second line of (51) is a composite convex quadratic minimization problem, which has the same form as (43). To analyze (51), we define

$$\lambda_t(\mathbf{y}) := \|\mathbf{y} - \mathcal{P}_{\mathbf{y}}(\mathbf{y} - \nabla^2 \varphi(\mathbf{y})^{-1} \nabla \varphi(\mathbf{y}); t)\|_{\mathbf{y}}, \quad (52)$$

where  $\mathcal{P}_{\mathbf{y}}(\cdot; t) = \text{prox}_{t^{-1} \nabla^2 \varphi(\mathbf{y})^{-1} \psi}(\cdot)$  is defined by (16).

### 4.3.2 Finding a starting point via an auxiliary problem

Let us fix  $t_0 > 0$  (e.g.,  $t_0 := \kappa$ ), and choose  $\beta \in (0, 1)$  such that  $\Omega_{t_0}(\beta)$  defined in (29) is a central path neighborhood of (52). The aim is to find a starting point  $\mathbf{y}^0 \in \Omega_{t_0}(\beta)$ . We again apply (51) to solve an auxiliary problem of (47) for finding  $\mathbf{y}^0 \in \Omega_{t_0}(\beta)$ .

Given an arbitrary  $\hat{\mathbf{y}}^0 \in \text{int}(\mathcal{D}_{\mathcal{Y}})$ , let  $\hat{\xi}_0 \in \partial\psi(\hat{\mathbf{y}}^0)$  be an arbitrary subgradient of  $\psi$  at  $\hat{\mathbf{y}}^0$ , and  $\hat{\zeta}^0 := \nabla\varphi(\hat{\mathbf{y}}^0) + t_0^{-1}\hat{\xi}_0$ . We consider the following auxiliary convex problem:

$$\min_{\mathbf{y} \in \mathbb{R}^p} \left\{ \hat{\phi}_{\tau}(\mathbf{y}) := \varphi(\mathbf{y}) - \tau \langle \hat{\zeta}^0, \mathbf{y} \rangle + t_0^{-1}\psi(\mathbf{y}) \right\}, \quad (53)$$

where  $\tau \in [0, 1]$  is a given continuation parameter.

As seen before, when  $\tau = 0$ , (53) becomes (47) at  $t := t_0$ , while with  $\tau = 1$  we have  $\nabla\varphi(\hat{\mathbf{y}}^0) - \hat{\xi}_0 = \nabla\varphi(\hat{\mathbf{y}}^0) - \nabla\varphi(\hat{\mathbf{y}}^0) - t_0^{-1}\hat{\xi}_0 = -t_0^{-1}\hat{\xi}_0 \in -t_0^{-1}\partial\psi(\hat{\mathbf{y}}^0)$ . Hence,  $0 \in \nabla\varphi(\hat{\mathbf{y}}^0) - \hat{\xi}_0 + t_0^{-1}\partial\psi(\hat{\mathbf{y}}^0)$ , which implies that  $\hat{\mathbf{y}}^0$  is a solution of (53) at  $\tau = 1$ .

We customize (51) to solve (53) by tracking the path  $\{\tau_j\}$  starting from  $\tau_0 := 1$  such that  $\{\tau_j\}$  converges to zero. We use the index  $j$  instead of  $k$  to distinguish with Phase 2.

Given  $\hat{\mathbf{y}}^j \in \text{int}(\mathcal{D}_{\mathcal{Y}})$  and  $\tau_j > 0$ , similar to (51), we update

$$\begin{cases} \tau_{j+1} := \tau_j - \Delta_j \\ \hat{\mathbf{y}}^{j+1} \approx \bar{\mathbf{y}}^{j+1} := \underset{\mathbf{y} \in \mathbb{R}^p}{\text{argmin}} \left\{ \hat{\phi}(\mathbf{y}; \hat{\mathbf{y}}^j) := Q_{\varphi}(\mathbf{y}; \hat{\mathbf{y}}^j) - \tau_{j+1} \langle \hat{\zeta}^0, \mathbf{y} \rangle + t_0^{-1}\psi(\mathbf{y}) \right\}, \end{cases} \quad (54)$$

where  $\Delta_j > 0$  is given, and  $\approx$  is in the sense of Definition 5 with a given tolerance  $\delta_j$ .

Since (54) has the same form as (51) applied to (53), we define

$$\hat{\lambda}_j := \|\hat{\mathbf{y}}_j - \text{prox}_{t_0^{-1}\nabla^2\varphi(\hat{\mathbf{y}}^j)^{-1}\psi} \left( \hat{\mathbf{y}}_j - \nabla^2\varphi(\hat{\mathbf{y}}_j)^{-1}\nabla\hat{\varphi}(\hat{\mathbf{y}}_j) \right)\|_{\hat{\mathbf{y}}_j}, \quad (55)$$

as the dual proximal Newton decrement for (54).

### 4.3.3 Primal solution recovery and the worst-case complexity

Our next step is to recover an approximate primal solution  $(\mathbf{x}_\varepsilon^*, \mathbf{s}_\varepsilon^*)$  of the primal problem (3) from the dual one  $\mathbf{y}_\varepsilon^*$  of (5). The following theorem provides such a strategy whose proof can be found in “Appendix 7.9”. The notation  $\pi_{\partial g^*(\mathbf{y}^k)}(L\mathbf{x}^k - \mathbf{b})$  stands for the projection of  $L\mathbf{x}^k - \mathbf{b}$  onto  $\partial g^*(\mathbf{y}^k)$  which is a nonempty, closed, and convex set.

**Theorem 7** *Let  $\{(\mathbf{y}^k, t_k)\}$  be the sequence generated by (51) and (54) to approximate a solution of the dual problem (5). Then,  $(\mathbf{x}^k, \mathbf{s}^k)$  computed by*

$$\mathbf{x}^k := \nabla f^*(t_k^{-1}(\mathbf{c} - L^*\mathbf{y}^k)) \in \text{int}(\mathcal{K}) \quad \text{and} \quad \mathbf{s}^k = \pi_{\partial g^*(\mathbf{y}^k)}(L(\mathbf{x}^k) - \mathbf{b}), \quad (56)$$

together with  $\mathbf{y}^k$  satisfy the following estimate

$$\begin{cases} \mathbf{y}^k \in \partial g(\mathbf{s}^k), \\ \|L^*\mathbf{y}^k - \mathbf{c}\|_{\mathbf{x}^k}^* \leq \sqrt{v}t_k, \\ \|L\mathbf{x}^k - \mathbf{s}^k - \mathbf{b}\|_{\mathbf{y}^k}^* \leq \theta(c, \beta)t_k, \end{cases} \quad (57)$$

where

$$\begin{aligned} \theta(c, \beta) := & \frac{(1-c^2)\beta}{(1+c\sqrt{\beta})^2[3c\sqrt{\beta}+c^2\beta+(1+c\sqrt{\beta})^3]-(1-c^2)\beta} \\ & + \left( \frac{(1-c^2)\beta+c\sqrt{\beta}(1+c\sqrt{\beta})^2[3c\sqrt{\beta}+c^2\beta+(1+c\sqrt{\beta})^3]}{(1+c\sqrt{\beta})^2[3c\sqrt{\beta}+c^2\beta+(1+c\sqrt{\beta})^3]-(1-c^2)\beta} \right)^2 \leq 1, \end{aligned} \quad (58)$$

is a constant for fixed  $c$  and  $\beta$  chosen as in Lemma 4.

Consequently, if  $\max \{\sqrt{v}, \theta(c, \beta)\} t_k = \sqrt{v}t_k \leq \varepsilon$ , then  $(\mathbf{x}^k, \mathbf{s}^k)$  is an  $\varepsilon$ -solution to (3) in the sense of Definition 6 associated with the  $\varepsilon$ -dual solution  $\mathbf{y}^k$  of (5).

#### 4.3.4 Two-phase inexact dual path-following proximal Newton algorithm

Now, we specify Algorithm 1 to solve (4) using (54) and (51) as in Algorithm 2.

---

#### Algorithm 2 (Two-phase inexact dual path-following proximal Newton algorithm)

---

- 1: **Initialization:**
  - 2: Choose  $\hat{\mathbf{y}}^0 \in \mathbb{R}^p$  such that  $\mathcal{L}^*\hat{\mathbf{y}}^0 - \mathbf{c} \in \text{int}(\mathcal{K}^*)$ . Fix  $t_0 := \kappa$ , and an accuracy  $\varepsilon > 0$ .
  - 3: Compute a vector  $\hat{\xi}_0 \in \partial\psi(\hat{\mathbf{y}}^0)$  and evaluate  $\nabla\varphi(\hat{\mathbf{y}}^0)$ .
  - 4: Set  $\hat{\xi}^0 := \nabla\varphi(\hat{\mathbf{y}}^0) + t_0^{-1}\hat{\xi}_0$  and  $\tau_0 := 1$ .
  - 5: Choose  $\beta, \eta$ , then compute  $\bar{\delta}_\tau, \bar{\mu}_\eta, \bar{\delta}_t, \bar{\sigma}_\beta$  as in Algorithm 1. Compute  $\theta(c, \beta)$  by (58).
- 

#### 6: Phase 1: Computing an initial point

---

- 7: **For**  $j = 0, \dots, j_{\max}$ , **perform**:
  - 8: If  $\tau_j \|\hat{\xi}^0\|_{\hat{\mathbf{y}}^j}^* \leq (\beta - \eta)$ , then TERMINATE.
  - 9: Perform (54) with  $\Delta_j := \frac{\bar{\mu}_\eta}{\|\hat{\xi}^0\|_{\hat{\mathbf{y}}^j}^*}$  up to an accuracy  $\hat{\delta}_j \leq \bar{\delta}_\tau$ .
  - 10: **End for**
- 

#### 11: Phase 2: Inexact dual path-following proximal iterations

---

- 12: **For**  $k = 0, \dots, k_{\max}$ , **perform**:
  - 13: If  $\sqrt{v}t_k \leq \varepsilon$ , then TERMINATE.
  - 14: Perform (51) up to an accuracy  $\delta_k \leq \bar{\delta}_t$ .
  - 15: **End for**
  - 16: **Primal recovery:** Recover  $(\mathbf{x}^k, \mathbf{s}^k)$  from  $\mathbf{y}^k$  as in (56). Then, return  $(\mathbf{x}^k, \mathbf{s}^k, \mathbf{y}^k)$ .
-

The main per-iteration complexity of Algorithm 2 lies at Steps 9 and 14, where we need to solve two composite and strongly convex quadratic programs in (54) and (51), respectively. The primal solution recovery at Step 16 does not significantly increase the computational cost of Algorithm 2. The worst-case iteration-complexity of Algorithm 2 remains the same as in Theorem 6 with  $M_0 := \sqrt{v}$ , and we do not restate it here.

## 5 Preliminary numerical experiments

We present three numerical examples to illustrate three algorithmic variants described in the previous sections, respectively. We compare our methods with three state-of-the-art interior-point solvers: SDPT3 [50], SeDuMi [48], and Mosek (a commercial software package). We also compare our methods with the SDPNAL+ solver (version 0.5) in [62], which implemented a majorized semi-smooth Newton-CG augmented Lagrangian method. For the second example, we implement an alternating direction method of multipliers (ADMM) [7] to compare with our method. Our numerical experiments are carried out in a Matlab R2014b environment, running on a MacBook Pro Laptop (Retina, 2.7GHz Intel Core i5, 16GM Memory).

### 5.1 Example 1: minimizing the maximum eigenvalue with constraints

We illustrate Algorithm 1(a) via the well-known maximum eigenvalue problem [33]:

$$\lambda_{\max}^* := \min_{\mathbf{y} \in \mathcal{Y}} \{\lambda_{\max}(C + L\mathbf{y})\}, \quad (59)$$

where  $\lambda_{\max}(U)$  is the maximum eigenvalue of a symmetric matrix  $U \in \mathcal{S}^n$ ,  $C \in \mathcal{S}^n$  is a given matrix,  $L$  is a linear operator from  $\mathbb{R}^p \rightarrow \mathcal{S}^n$ , and  $\mathcal{Y}$  is a nonempty, closed, and convex set in  $\mathbb{R}^p$  endowed with a self-concordant barrier  $\varphi$ .

As a consequence of J. von Neumann's trace inequality, we can show that  $\lambda_{\max}(U) = \max_{\mathbf{x}} \{\mathbf{x}^\top U \mathbf{x} \mid \|\mathbf{x}\|_2 = 1\} = \max \{\text{trace}(UX) \mid \text{trace}(X) = 1, X \in \mathcal{S}_+^n\}$ . Hence, if we define  $\mathcal{X} := \mathcal{S}_+^n$  and  $g(X) := \delta_{\{X \mid \text{trace}(X)=1\}}(X) - \text{trace}(CX)$ , and using  $\langle L\mathbf{y}, X \rangle = \text{trace}((L\mathbf{y})X)$ , then we can rewrite (59) as (8), which is of the form:

$$\tilde{\lambda}_{\max}^* := \min_{\mathbf{y} \in \mathcal{Y}} \left\{ \max_{X \in \mathcal{S}_+^n} \{\langle L\mathbf{y}, X \rangle - g(X)\} \right\}. \quad (60)$$

The corresponding barrier for  $\mathcal{S}_+^n$  is  $f(X) := -\log \det(X)$ .

Now, we can apply Algorithm 1(a) to solve (60). The main computation of this algorithm is the solution of (40) and (42), which can be written explicitly as follows for (60):

$$\min_{\mathbf{y} \in \mathbb{R}^p} \left\{ \max_{X \in \mathcal{S}^n} \left\{ \langle L\mathbf{y}, X \rangle - t Q_f(X; X^k) - g(X) \right\} + t Q_\varphi(\mathbf{y}; \mathbf{y}^k) \right\}, \quad (61)$$

where  $Q_f(X; X^k) := \langle \nabla f(X^k), X - X^k \rangle + \frac{1}{2} \langle \nabla^2 f(X^k)(X - X^k), X - X^k \rangle$  and  $Q_\varphi(\mathbf{y}; \mathbf{y}^k) := \langle \nabla \varphi(\mathbf{y}^k), \mathbf{y} - \mathbf{y}^k \rangle + \frac{1}{2} \langle \nabla^2 \varphi(\mathbf{y}^k)(\mathbf{y} - \mathbf{y}^k), \mathbf{y} - \mathbf{y}^k \rangle$ . We can solve (61) in a closed form as follows:

$$\begin{cases} X_k^* := \text{mat} \left( \left[ \frac{\text{trace}(\text{mat}(H_k^{-1} h_k)) + 1}{\text{trace}(\text{mat}(H_k^{-1} \text{vec}(\mathbb{I})))} \right] H_k^{-1} \text{vec}(\mathbb{I}) - H_k^{-1} h_k \right), \\ \mathbf{y}_k^* := \mathbf{y}^k - \nabla^2 \varphi(\mathbf{y}^k)^{-1} (\nabla \varphi(\mathbf{y}^k) + t^{-1} L^* X_k^*), \end{cases} \quad (62)$$

where  $H_k := \nabla^2 f(X^k) + t^{-2} L \nabla^2 \varphi(\mathbf{y}^k)^{-1} L^* \succ 0$  and  $h_k := [\nabla f(X^k) - \nabla^2 f(X^k) \text{vec}(X^k)] - t^{-1} L (\mathbf{y}^k - \nabla^2 \varphi(\mathbf{y}^k)^{-1} \nabla \varphi(\mathbf{y}^k)) - t^{-1} \text{vec}(C)$ .

We consider a simple case, where  $\mathcal{Y} := \{\mathbf{y} \in \mathbb{R}^p \mid \|\mathbf{y}\|_\infty \leq 1\}$ . Then, the barrier function of  $\mathcal{Y}$  is simply  $\varphi(\mathbf{y}) := -\sum_{i=1}^p \log(1 - \mathbf{y}_i^2)$ . In this case, we can compute both  $\nabla \varphi(\cdot)$  and  $\nabla^2 \varphi(\cdot)^{-1}$  in a closed form. The barrier parameter for  $F := f + \varphi$  is  $v := 2p + n$ .

We test 5 solvers: Algorithm 1(a), SDPT3, SeDuMi, Mosek, and SDPNAL+ on 10 medium-size problems, where  $n$  varies from 5 to 50, and  $p = 10n^2$  (varies from 250 to 25,000). For these three IP solvers and SDPNAL+, we reformulate (60) into the following SDP problem:

$$\min_{s, X, \mathbf{y}} \{s \mid s\mathbb{I} - X - L\mathbf{y} = C, X \succeq 0, -1 \leq \mathbf{y}_j \leq 1, j = 1, \dots, p\}.$$

The linear operator  $L$  and matrix  $C$  are generated randomly using the standard Gaussian distribution `randn` in Matlab, which are completely dense. In Phase 1 of Algorithm 1(a), instead of performing a path-following scheme on the auxiliary problem, we simply perform a damped step variant on the original problem. We set the initial penalty parameter  $t_0 := 0.1$ . We terminate our algorithm if  $t_k \leq 10^{-6}$  and  $\tilde{\lambda}_k \leq 10^{-8}$ . When  $t_k \leq 10^{-6}$ , if  $\tilde{\lambda}_k$  does not reach the  $10^{-8}$  accuracy, we fix  $t_k = 10^{-6}$  and perform at most 15 additional iterations to decrease  $\tilde{\lambda}_k$ . We terminate SDPT3, SeDiMi, and Mosek with the same accuracy  $\sqrt{\varepsilon} = 1.49 \times 10^{-8}$ , where  $\varepsilon$  is Matlab's machine precision. We terminate SDPNAL+ using its default setting, but set the maximum number of iterations at 1000.

The result and performance of these solvers are reported in Table 1, where  $n \times p$  is the size of  $L$ , `iter` is the number of iterations in Phase 1 and Phase 2, and  $\lambda^*$  is the reported objective value of each solver. The most intensive computation of Algorithm 1(a) is  $L \text{diag}(\nabla^2 \varphi(\mathbf{y}^k)^{-1} L^\top)$ , which costs from 40% to 80% the overall computational time.

In this test, Mosek is the fastest when the size is increasing, while SDPNAL+ is the slowest. SDPT3 is slow but is slightly better than SDPNAL+ in this test. Our algorithm produces nearly optimal objective value while requires reasonable computational time compared to the other solvers. SDPNAL+ gives a slightly lower objective value in some problems, but also violates the bound constraint  $\|\mathbf{y}\|_\infty \leq 1$ . We emphasize that our algorithm is naively implemented in Matlab without optimizing the code or using `mex` files as other solvers. Mosek is a well-known commercial software implemented in C++ using several advanced heuristic strategies, and SDPNAL+ has been developed through several releases using both Matlab and C codes.

**Table 1** Summary of the result and performance of 5 solvers for solving problem (59)

Problem	$n$	$p$	Algorithm 1(a)			SDPT3			SeDuMi			Mosek			SDPNAL+		
			Iter	Time (s)	$\lambda^*$ <sub>ours</sub>	Time (s)	$\lambda^*$ <sub>sdp3</sub>	Time (s)	$\lambda^*$ <sub>sedumi</sub>	Time (s)	$\lambda^*$ <sub>mosek</sub>	Time (s)	$\lambda^*$ <sub>sdpnal+</sub>	Time (s)			
5	250	19/40	0.37	-80.34	2.49	-80.34	1.32	-80.34	3.12	-80.34	3.12	-80.34	27.53	-80.34			
10	1000	3/2/60	0.30	-255.92	3.66	-255.92	1.56	-255.92	2.06	-255.92	2.06	-255.92	12.25	-255.92			
15	2250	41/72	1.27	-453.51	13.35	-453.52	4.95	-453.52	2.53	-453.52	2.53	-453.52	36.99	-453.52			
20	4000	49/81	6.88	-684.18	55.58	-684.18	21.67	-684.18	4.14	-684.18	4.14	-684.18	38.38	-684.19			
25	6250	57/81	23.06	-952.12	309.29	-952.12	135.91	-952.12	7.91	-952.12	7.91	-952.12	315.46	-952.13			
30	9000	65/86	155.37	-1265.71	518.82	-1265.71	209.27	-1265.71	14.53	-1265.71	14.53	-1265.71	1202.07	-1257.17			
35	12,250	71/104	181.21	-1582.48	1262.64	-1582.49	494.84	-1582.49	30.40	-1582.49	30.40	-1582.49	2912.92	-1582.21			
35	12,250	71/104	181.21	-1582.48	1262.64	-1582.49	494.84	-1582.49	30.40	-1582.49	30.40	-1582.49	2912.92	-1582.21			
40	16,000	78/110	400.89	-1931.65	2795.90	-1931.66	1064.91	-1931.66	68.78	-1931.66	68.78	-1931.66	2487.30	-1925.06			
45	20,250	84/117	831.43	-2322.09	4777.96	-2322.12	2052.40	-2322.12	77.91	-2322.11	77.91	-2322.11	1840.10	-2322.91			
50	25,000	89/125	1367.36	-2694.27	9474.14	-2694.29	4184.44	-2694.29	130.61	-2694.29	130.61	-2694.29	13948.03	-2696.67			

## 5.2 Example 2: sparse and low-rank matrix approximation

The problem of approximating a given  $n \times n$ -symmetric matrix  $M$  as the sum of a low-rank positive semidefinite matrix  $X$  with bounded magnitudes, and a sparse matrix  $M - X$  can be formulated into the following convex optimization problem (see [51]):

$$\begin{cases} \min_X \rho \|\text{vec}(X - M)\|_1 + (1 - \rho)\text{trace}(X) \\ \text{s.t. } X \succeq 0, \quad L_{ij} \leq X_{ij} \leq U_{ij}, \quad 1 \leq i < j \leq n. \end{cases} \quad (63)$$

Here,  $\rho \in (0, 1)$  is a regularization parameter, and  $L$  and  $U$  are the lower and upper bounds.

Let us define  $\mathcal{X} := \mathcal{S}_{++}^n$  and  $g(X) := \rho \|\text{vec}(X - M)\|_1 + (1 - \rho)\text{trace}(X) + \delta_{[L, U]}(X)$ , where  $\delta_{[L, U]}$  is the indicator of  $[L, U] := \{X \in \mathcal{S}^n \mid L_{ij} \leq X_{ij} \leq U_{ij}, 1 \leq i < j \leq n\}$ . Then, we can formulate (63) into the form (3) with  $f(X) := -\log \det(X)$ .

We implement Algorithm 1(b) for solving (3) and compare it with Mosek, SDPNAL+, and ADMM. The initial parameter is set to  $t_0 := 0.1$ . We use a restarting accelerated proximal-gradient algorithm proposed in [49] to solve the subproblems (43) and (46) with at most 150 iterations. We apply the same strategy as in Sect. 5.1 to terminate this algorithm. For Mosek and SDPNAL+, we use their default configuration to solve an equivalent mixed SDP reformulation of (63) by transforming the  $\ell_1$ -norm term into second-order cone constraints.

Since problems of the form (63) have been successfully solved by ADMM [57], we also implement an ADMM variant [7] to solve (63). We terminate ADMM using a criterion in [7, page 19] with a tolerance  $\varepsilon = 0.5 \times 10^{-5}$ . With this  $\varepsilon$ , ADMM nearly reaches the same order of accuracy as in other methods. We also set the maximum number of iterations at 5000. We tune ADMM to find a reasonable penalty parameter for all problems, which is  $\beta = 10.0$ .

We test four algorithms on 12 problems with the size reported in Table 2.

We limit our test to  $n = 240$  since Mosek can only solve up to this size in our computer. The data is generated as follows. We generate a symmetric matrix  $M_0$  using standard Gaussian distribution with the rank of  $\lfloor 0.25n \rfloor$  and the sparsity of 25%. Then, we add a sparse Gaussian noise  $E$  with the sparsity of 10% and the variance of  $10^{-4}$  as  $M := M_0 + E$  to obtain  $M$ . We generate the lower bound  $L_{ij} := 0.9 \min \{M_{ij} \mid 1 \leq i < j \leq n\}$  and the upper bound  $U_{ij} := 1.1 \max \{M_{ij} \mid 1 \leq i < j \leq n\}$ . We choose  $\rho = 0.2$  for all problem instances.

The performance and result of three algorithms are reported in Table 2.

Here, `iter` is the number of iterations for Phase 1 and Phase 2 of Algorithm 1(b); `time` is the computational time in second;  $g(X^k)$  (respectively,  $g^{\star}_{\text{sdpn}}$ ,  $g^{\star}_{\text{mosek}}$ , and  $g^{\star}_{\text{admm}}$ ) is the objective value of Algorithm 1(b) (respectively, SDPNAL+, Mosek, and ADMM); and `spr/rank` is the sparsity level of  $X^k - M$  (e.g.,  $\text{spr} := \text{nnz}(X^k - M)/n^2$ ), and the rank of  $X^k$  (rounding up to  $10^{-6}$  accuracy);

As we can see from Table 2, four solvers give similar results in terms of the objective value. The computational time of Algorithm 1(b), SDPNAL+, and ADMM is in the same order, while Mosek is much slower than other three solvers when  $p$  is increasing.

**Table 2** Summary of the result and performance of 4 solvers on 12 problem instances

n	Algorithm 1(b)				SDPNAL+ [62]				Mosek				ADMM [7, 57]			
	Iter	Time	$g(X^k)$	spr/rank	Iter	Time	$g^*$ <sub>sdpnal+</sub>	spr/rank	Time	$g^*$ <sub>mosek</sub>	spr/rank	Time	$g^*$ <sub>admm</sub>	spr/rank		
20	11/74	0.39	11.37	0.39/13	294	0.91	11.37	0.73/18	0.13	11.37	0.46/4	0.88	11.37	0.39/3		
40	16/101	0.93	57.06	0.36/32	500	4.54	57.06	0.82/39	0.65	57.06	0.60/19	1.67	57.06	0.80/15		
60	23/124	2.16	162.75	0.27/53	350	1.93	162.74	0.96/59	3.07	162.75	0.89/44	3.87	162.75	0.83/25		
80	25/147	2.36	245.39	0.21/73	210	1.62	245.38	0.88/79	8.99	245.39	0.85/53	4.61	245.39	0.86/36		
100	29/166	3.24	427.29	0.18/92	216	1.71	427.27	0.90/100	23.32	427.29	0.79/66	5.48	427.29	0.89/46		
120	33/184	3.71	662.17	0.14/113	211	2.66	662.14	0.89/119	57.58	662.17	0.89/87	5.99	662.17	0.90/55		
140	35/200	4.45	893.47	0.12/133	191	2.74	893.44	0.90/139	141.89	893.47	0.82/96	11.24	893.49	0.90/65		
160	38/216	5.98	1185.71	0.10/154	193	3.82	1185.67	0.90/159	266.47	1185.71	0.80/127	9.75	1185.73	0.90/75		
180	41/244	8.01	1493.43	0.10/173	191	3.89	1493.38	0.90/179	542.29	1493.43	0.89/169	12.73	1493.45	0.91/85		
200	43/272	10.97	1741.32	0.08/196	194	4.30	1741.26	0.90/199	1049.39	1741.33	0.92/194	15.54	1741.35	0.91/99		
220	46/297	11.52	2082.55	0.06/214	191	7.13	2082.47	0.90/218	1374.15	2082.55	0.86/218	16.62	2082.58	0.91/111		
240	49/329	15.31	2577.35	0.04/236	194	6.21	2577.25	0.90/237	2644.77	2577.36	0.91/219	17.83	2577.38	0.91/117		

More precisely, Algorithm 1(b) is slightly slower than SDPNAL+, but slightly faster than ADMM in terms of time. It is also much faster than Mosek.

Both Algorithm 1(b) and Mosek perfectly satisfy the positive definiteness constraint with  $\lambda_{\min}(X^k) > 0$ , while SDPNAL+ still slightly violates this constraint with  $\lambda_{\min}(X^k) \leq -\mathcal{O}(10^{-4})$ . ADMM also slightly violates the positivity constraint  $X \succeq 0$  in 6 problem instances out of 12. Algorithm 1 gives the best results in terms of the sparsity of  $X^k - M$ , while achieves similar rank as Mosek and SDPNAL+ in the majority of the test. ADMM gives a better rank than other methods but its solution leads to a dense residual  $X^k - M$ .

As often observed in first-order methods, ADMM easily achieves a low-accuracy solution (i.e.,  $10^{-4}$  accuracy) in less than five hundred iterations (i.e., from 282 to 476 iterations). However, in order to reach a high accuracy solution as Algorithm 1(b), or IP solvers, it requires many iterations (i.e., from 700 to 3210 iterations depending on problem instances).

### 5.3 Example 3: cluster recovery

Finally, we test Algorithm 2 for solving the following well-studied clustering recovery problem via an SDP relaxation as studied in [22]:

$$\begin{cases} \max_{X \in \mathcal{S}_+^n} \text{trace}(A^\top X) \\ \text{s.t. } X_{ii} \leq 1, \quad X_{ij} \geq 0, \quad i, j = 1, \dots, n, \\ \quad \text{trace}(X) = s_1, \quad \text{trace}(E_n X) = s_2, \end{cases} \quad (64)$$

where  $A$  is the adjacency matrix of a given graph,  $E_n$  is the all-one matrix in  $\mathbb{R}^{n \times n}$ ,  $s_1 = \sum_{i=1}^r K_i$ , and  $s_2 := \sum_{i=1}^r K_i^2$  with  $K_1, K_2, \dots, K_r$  being the size of  $r$  clusters.

Let us define  $\mathcal{K} := \mathcal{S}_+^n$ ,  $LX := [\text{trace}(X), \text{trace}(E_n X), X_{ii}, X_{ij}] : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n(n+1)+2}$ ,  $g(\mathbf{s}) := \delta_{\{0\}^2}(\mathbf{s}_{1:2}) + \delta_{-\mathbb{R}_+^n}(\mathbf{s}_{3:n+2}) + \delta_{\mathbb{R}_+^{n^2}}(\mathbf{s}_{n+3:n(n+1)+2})$ , and  $\mathbf{b} := (s_1, s_2, \mathbf{e}_n, \mathbf{0}^{n^2})^\top \in \mathbb{R}^{n(n+1)+2}$ , where  $\mathbf{e}_n$  is the all-one vector,  $\delta_X$  is the indicator function of  $X$ , and  $\mathbf{s}_{k_1:k_2}$  is the subvector of  $\mathbf{s}$  concatenating from the  $k_1$ -th entry to the  $k_2$ -entry. Using these notations, we can reformulate (64) into the constrained convex problem (4).

Since  $\mathcal{K} := \mathcal{S}_+^n$  is a self-dual cone, i.e.,  $\mathcal{K}^* = \mathcal{S}_+^n$ , and the corresponding self-concordant logarithmically homogeneous barrier of  $-\mathcal{K}^*$  is  $f^*(S) := -\log \det(-S) - n$ .

We implement Algorithm 2 to solve (64). We use a restarting proximal-gradient algorithm [3] to solve the corresponding subproblems in (51) and (54). Since we can compute the dual objective values, we use a damped-step Newton scheme in Phase 1 to compute an initial point  $\mathbf{y}^0$ . Since we use the first-order method for solving the subproblems, we set the precision of those solvers above to be low such that the relative error is guaranteed to be less than  $\mathcal{O}(10^{-4})$  when it is terminated. We choose the number of clusters such that the average number of points of each cluster is between 10 and 20. The initial value of  $t_0$  is set to  $t_0 := 0.5$  if  $n \leq 120$ , and to  $t_0 := 1.0$ , otherwise. All optimal values of our algorithm have the relative error smaller than

$10^{-4}$  when the algorithm is terminated, which matches the precision of the above solvers.

The results and performance of these solvers are reported in Table 3 for small-scale problems. In this table, we summarize the result and performance of 12 problems of the size from 60 to 250, where  $K$  is the number of clusters;  $\text{iter}$  is the number of iterations in Phase 1 and Phase 2 of Algorithm 2; and  $\text{spu}$  is the speed up ratio (i.e.,  $\text{spu} = \frac{\text{time}_{\text{Algorithm 2}}}{\text{time}_{\text{Algorithm 2}}} - 1$ ) in terms of time compared to other solvers.

Table 3 shows that Algorithm 2 can achieve the same order of accuracy as other three solvers while it highly outperforms SDPT3 and Mosek in terms of computational time. We can speed up Algorithm 2 up to 18 times compared to Mosek, and 62 times over SDPT3. This is due to the low cost computation of the projections when we work directly on the dual of the original problem. The other IP solvers require to convert it to an appropriate SDP format which substantially increases the problem size as seen from Sect. 5.2. Since problem (64) fits SDPNAL+ very well, we also compare it with our method. Clearly, SDPNAL+ takes advantages of splitting techniques in ADMM, semi-smooth Newton-CG methods, and heuristic procedures to perform well in this test. It has a similar performance as Algorithm 2 on small problems, while it becomes faster when the problem size increases.

## 6 Discussion

We have studied a class of self-concordant inclusions of the form (1), and designed an inexact generalized Newton-type framework for solving it. Problem (1) is sufficiently general to cope with three fundamental convex optimization formulations discussed in Sect. 1.2. Moreover, since this problem can be reformulated into a multivalued variational inequality problem [45], theory and methods from this area can be used to deal with (1), see [2, 16, 45].

Most existing methods for solving (1) exploit specific structures of  $\mathcal{A}$  and  $\mathcal{Z}$ . When  $\mathcal{A}$  is single-valued, (1) is a standard single-valued variational inequality, and it becomes a complementarity problem if additionally  $\mathcal{Z}$  is a box. The most commonly used methods to solve complementarity problems are based on generalized Newton methods developed for nonsmooth equations, including the path following methods [41] and semi-smooth Newton-type methods [12, 26, 40, 58, 62]. The basic idea is to reformulate the complementarity problem as an equation defined by nonsmooth functions, and at each iteration, one can approximately solve the equation obtained by some first-order approximation or a generalized Jacobian matrix of the nonsmooth functions. Another important class of methods to solve (1) are based on projection and splitting [54, 55, 60]. These methods can be considered as special cases of the forward-backward splitting scheme when the second operator is simply the normal cone of the convex set  $\mathcal{Z}$ . When  $\mathcal{A}$  is maximally monotone, its resolvent is well-defined and single-valued. Splitting methods using proximal-point and projected schemes such as Douglas–Rachford’s methods can be applied to solve (1). Other approaches such as augmented Lagrangian [62], extragradient, mirror descent, hybrid-gradient, gap functions, smoothing techniques, and interior-point proximal methods are also widely studied in the literature.

**Table 3** Summary of performance of 4 solvers on 12 problem instances

Problem	n	K	$\rho$ (%)	Algorithm 2		SDPT3		Mosek		SDPNAL+		
				Iter	Time (s)	$\mathcal{G}_k$	Time (s)	$\mathcal{G}_{\text{sdpf3}}^*$	spu	Time (s)	$\mathcal{G}_{\text{mosek}}^*$	spu
60	5	18.6	52/215	3.63	660.07	6.28	660.00	1.7	4.41	660.00	1.2	1.58
70	7	13.0	42/232	3.37	630.02	9.27	630.00	2.8	6.62	630.00	2.0	3.09
80	8	11.3	45/247	4.30	720.02	14.46	720.00	3.4	8.39	720.00	2.0	3.37
90	9	10.1	48/262	6.61	810.03	27.80	810.00	4.2	14.40	810.00	2.2	3.47
100	10	9.1	50/276	7.54	900.03	44.24	900.00	5.9	23.45	900.00	3.1	3.21
110	10	9.2	57/289	8.91	1100.05	57.26	1100.00	6.4	33.01	1100.00	3.7	2.34
120	10	9.2	78/302	11.13	1320.13	88.61	1320.00	8.0	50.80	1320.00	4.6	3.58
140	10	9.4	57/359	20.18	1820.13	203.74	1820.00	10.1	110.82	1820.00	5.5	4.22
160	10	9.4	76/383	26.57	2400.23	340.58	2400.00	12.8	214.13	2400.00	8.1	5.02
180	15	6.2	63/406	36.73	1980.16	808.40	1980.00	22.0	423.28	1980.00	11.5	4.52
200	20	4.5	64/428	66.34	1800.14	1945.14	1798.69	29.3	787.78	1800.00	11.9	4.31
250	25	3.6	72/478	184.18	2250.21	11,378.33	2250.00	61.8	3415.16	2250.00	18.5	5.48

for different classes of (1), see, e.g., [1, 6, 15, 16, 19, 20, 25, 28, 30, 32, 47, 59] and the references quoted therein.

From a theoretical point of view, the setting (1) can be used as a unified tool to handle a wide range of convex problems. Three specific instances (3), (4) and (8) of (1) are well-studied and have a great impact in different fields including operations research, statistics, machine learning, signal and image processing, and controls [2, 5, 8, 31]. Methods for solving these instances include sequential quadratic programming [37], interior-points [35], augmented Lagrangian-type methods [56, 62] (e.g., implemented in SDPAD, SDPNAL/SDPNAL+), first-order/second-order primal-dual and splitting methods [2, 7, 9, 11, 14, 46, 53, 54], Frank-Wolfe-type algorithms [17, 23], and stochastic gradient descents [24, 29], just to name a few.

Perhaps, the interior-point method [5, 31, 36, 61] is one of the most well-developed methods for solving standard conic programs covered by (3) and (4). Interior-point methods together with disciplined programming approach [21] allow us to solve a large class of convex optimization problems arising in different fields. These techniques have been systematically implemented in several off-the-shelf software packages such as CVX [21], YALMIP [27], CPLEX, and Gurobi for both commercial and academic use. While interior-point methods provide a powerful framework to solve a large class of constrained convex problems with high accuracy and numerically robust performance, their high per-iteration complexity prevents them from solving large-scale applications in modern convex optimization.

Although the interior-point method and the proximal-type method have been separately well-developed for several classes of convex problems, their joint treatment was first proposed in [51, 52] to the best of our knowledge. In these papers, the authors proposed a novel path-following proximal Newton framework for the instance (3) of (1). They characterized an  $\mathcal{O}(\sqrt{v} \log(1/\varepsilon))$ -worst-case iteration-complexity as in standard path-following methods [31] to achieve an  $\varepsilon$ -solution of (3), where  $v$  is the barrier parameter of a barrier function of the feasible set  $\mathcal{X}$ , and  $\varepsilon$  is a desired accuracy. However, [51, 52] obtained a smaller neighborhood for the central path compared to standard path-following methods [31]. In addition, these algorithms used the points on the central path to measure the proximal Newton decrement which leads to a unimplementable stopping criterion. In contrast, this paper focuses on developing a unified theory using self-concordant inclusion (1) as a generic framework. The main component of our methods is the generalized Newton method studied, e.g., in [6, 38, 42, 43], but we have extended it to self-concordant settings. Moreover, we use a generalized gradient mapping to measure a neighborhood of the central path as well as a quadratic convergence region of the generalized Newton iterations, and this neighborhood has the same size as in standard path-following methods [31]. When this framework is specified to solve (3) and (4), such a generalized gradient mapping allows us to obtain an implementable stopping criterion, an adaptive rule for the penalty parameter, and the overall polynomial time worst-case iteration-complexity bounds.

**Acknowledgements** This work was supported in part by the NSF Grant, USA, Award Number: 1619884.

## 7 Appendix: the proofs of technical results

This appendix provides the full proofs of all lemmas and theorems in the main text.

### 7.1 The proof of Lemma 1: the existence and uniqueness of the solution of (2).

Under Assumption A.1, the operator  $t \nabla F(\cdot) + \mathcal{A}(\cdot)$  is maximally monotone for any  $t > 0$ . We use [45, Theorem 12.51] to prove the solution existence of (2).

To this end, let  $\omega \neq \mathbf{0}$  be chosen from the horizon cone of  $\text{int}(\mathcal{Z}) \cap \text{dom}(\mathcal{A})$ . We need to find  $\mathbf{z} \in \text{int}(\mathcal{Z}) \cap \text{dom}(\mathcal{A})$  with  $\mathbf{v} \in t \nabla F(\mathbf{z}) + \mathcal{A}(\mathbf{z})$  such that  $\langle \mathbf{v}, \omega \rangle > 0$ . By assumption, there exists  $\hat{\mathbf{z}} \in \text{int}(\mathcal{Z}) \cap \text{dom}(\mathcal{A})$  with  $\hat{\mathbf{a}} \in \mathcal{A}(\hat{\mathbf{z}})$  such that  $\langle \hat{\mathbf{a}}, \omega \rangle > 0$ .

First, we show that  $\mathbf{z}_\tau = \hat{\mathbf{z}} + \tau \omega$  belongs to  $\text{int}(\mathcal{Z}) \cap \text{dom}(\mathcal{A})$  for any  $\tau > 0$ . To see this, note that the assumption  $\text{int}(\mathcal{Z}) \cap \text{dom}(\mathcal{A}) \neq \emptyset$  implies that  $\text{int}(\mathcal{Z}) \cap \text{ri dom}(\mathcal{A}) \neq \emptyset$ , which implies that the closure of  $\text{int}(\mathcal{Z}) \cap \text{dom}(\mathcal{A})$  is exactly  $\mathcal{Z} \cap \text{cl}(\text{dom}(\mathcal{A}))$ . Choose  $\tau' > \tau$ ; by definition of the horizon cone,  $\mathbf{z}_{\tau'}$  belongs to the closure of  $\text{int}(\mathcal{Z}) \cap \text{dom}(\mathcal{A})$ , so  $\mathbf{z}_{\tau'} \in \mathcal{Z}$  and  $\mathbf{z}_{\tau'} \in \text{cl}(\text{dom}(\mathcal{A}))$ . Since  $\mathbf{z}_\tau$  is a convex combination of  $\hat{\mathbf{z}}$  and  $\mathbf{z}_{\tau'}$ , it belongs to  $\text{int}(\mathcal{Z}) \cap \text{dom}(\mathcal{A})$ , where we use the assumption that  $\text{dom}(\mathcal{A})$  is either closed or open.

Next, for any  $\mathbf{a}_\tau \in \mathcal{A}(\mathbf{z}_\tau)$ , we have

$$\langle \mathbf{a}_\tau, \omega \rangle = \langle \mathbf{a}_\tau - \hat{\mathbf{a}}, \omega \rangle + \langle \hat{\mathbf{a}}, \omega \rangle = \langle \mathbf{a}_\tau - \hat{\mathbf{a}}, \tau^{-1}(\mathbf{z}_\tau - \hat{\mathbf{z}}) \rangle + \langle \hat{\mathbf{a}}, \omega \rangle \geq \langle \hat{\mathbf{a}}, \omega \rangle > 0.$$

On the other hand,  $\langle t \nabla F(\mathbf{z}_\tau), \omega \rangle = \langle t \nabla F(\mathbf{z}_\tau), \tau^{-1}(\mathbf{z}_\tau - \hat{\mathbf{z}}) \rangle \geq -\tau^{-1}t\nu$  by [31, Theorem 4.2.4]. Combining the above two inequalities, we can see that

$$\langle t \nabla F(\mathbf{z}_\tau) + \mathbf{a}_\tau, \omega \rangle \geq -\tau^{-1}t\nu + \langle \hat{\mathbf{a}}, \omega \rangle > 0$$

as long as  $\tau^{-1}t\nu < \langle \hat{\mathbf{a}}, \omega \rangle$ . We have thereby verified the condition in [45, Theorem 12.51], which needs to guarantee (2) for having a nonempty (and bounded) solution set. Since  $\nabla F$  is strictly monotone, the solution of (2) is unique.

Note that  $\mathbf{z}_t^*$  is the solution of (2) and  $\mathbf{z}_t^* \in \text{int}(\mathcal{Z})$ , we have  $-t \nabla F(\mathbf{z}_t^*) \in \mathcal{A}(\mathbf{z}_t^*) = \mathcal{A}_{\mathcal{Z}}(\mathbf{z}_t^*)$ . Hence,  $\text{dist}_{\mathbf{z}_t^*}(\mathbf{0}, \mathcal{A}_{\mathcal{Z}}(\mathbf{z}_t^*)) \leq t \|\nabla F(\mathbf{z}_t^*)\|_{\mathbf{z}_t^*}^* \leq t\sqrt{\nu}$  due to the property of  $F$  [31]. Using Definition 4, we have the last conclusion.  $\square$

### 7.2 The proof of Lemma 3: approximate solution

First, since  $\bar{\mathbf{z}}_+$  is a zero point of  $\widehat{\mathcal{A}}_t(\cdot; z)$ , i.e.,  $0 \in \widehat{\mathcal{A}}_t(\bar{\mathbf{z}}_+, z)$ , we have  $-t \nabla F(\mathbf{z}) - t \nabla^2 F(\mathbf{z})(\bar{\mathbf{z}}_+ - \mathbf{z}) \in \mathcal{A}(\bar{\mathbf{z}}_+)$ . Second, since  $\mathbf{z}_+$  is a  $\delta$ -solution to (23), there exists  $\mathbf{e}$  such that  $\mathbf{e} \in t \nabla F(\mathbf{z}) + t \nabla^2 F(\mathbf{z})(\mathbf{z}_+ - \mathbf{z}) + \mathcal{A}(\mathbf{z}_+)$  with  $\|\mathbf{e}\|_{\mathbf{z}}^* \leq t\delta$  by Definition 5. Combining these expressions, and using the monotonicity of  $\mathcal{A}$  in Definition 1, we can show that  $\langle t[\nabla F(\mathbf{z}) + \nabla^2 F(\mathbf{z})(\mathbf{z}_+ - \mathbf{z}) - \nabla F(\mathbf{z}) - \nabla^2 F(\mathbf{z})(\bar{\mathbf{z}}_+ - \mathbf{z})] - \mathbf{e}, \bar{\mathbf{z}}_+ - \mathbf{z}_+ \rangle \geq 0$ . This inequality leads to

$$t \|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}}^2 \leq \langle \mathbf{e}, \mathbf{z}_+ - \bar{\mathbf{z}}_+ \rangle \leq \|\mathbf{e}\|_{\mathbf{z}}^* \|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}}, \quad (65)$$

which implies  $\|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}} \leq t^{-1}\|\mathbf{e}\|_{\mathbf{z}}^*$ . Hence,  $\|\mathbf{e}\|_{\mathbf{z}}^* \leq t\delta$  implies  $\|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}} \leq \delta$ .

Next, since  $\mathbf{z}_+$  is a  $\delta$ -approximate solution to (23) at  $t$  in the sense of Definition 5 up to the accuracy  $\delta$ , there exists  $\mathbf{e} \in \mathbb{R}^P$  such that

$$\mathbf{e} \in t \left[ \nabla F(\mathbf{z}) + \nabla^2 F(\mathbf{z})(\mathbf{z}_+ - \mathbf{z}) \right] + \mathcal{A}(\mathbf{z}_+) \text{ with } \|\mathbf{e}\|_{\mathbf{z}}^* \leq t\delta.$$

In addition, we have  $\mathbf{z}_+ \in \text{int}(\mathcal{Z})$  due to Theorem 1. Hence, we have  $\mathcal{A}_{\mathcal{Z}}(\mathbf{z}_+) = \mathcal{A}(\mathbf{z}_+)$ . Using this relation and the above inclusion, we can show that

$$\begin{aligned} \text{dist}_{\mathbf{z}}(\mathbf{0}, \mathcal{A}_{\mathcal{Z}}(\mathbf{z}_+)) &\leq \|\mathbf{e} - t [\nabla F(\mathbf{z}) + \nabla^2 F(\mathbf{z})(\mathbf{z}_+ - \mathbf{z})]\|_{\mathbf{z}}^* \\ &\leq \|\mathbf{e}\|_{\mathbf{z}}^* + t \|\nabla F(\mathbf{z})\|_{\mathbf{z}}^* + t \|\nabla^2 F(\mathbf{z})(\mathbf{z}_+ - \mathbf{z})\|_{\mathbf{z}}^* \\ &\leq t [\delta + \sqrt{v} + \|\nabla^2 F(\mathbf{z})(\bar{\mathbf{z}}_+ - \mathbf{z})\|_{\mathbf{z}}^* + \|\nabla^2 F(\mathbf{z})(\bar{\mathbf{z}}_+ - \mathbf{z}_+)\|_{\mathbf{z}}^*] \\ &\leq t [\delta + \sqrt{v} + \lambda_t(\mathbf{z}) + \|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}}] \\ &\leq t (\sqrt{v} + \lambda_t(\mathbf{z}) + 2\delta). \end{aligned} \tag{66}$$

Here, we have used  $\|\nabla F(\mathbf{z})\|_{\mathbf{z}}^* \leq \sqrt{v}$ , and  $\|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}} \leq \delta$  by the first part of this lemma. Note that if  $\lambda_t(\mathbf{z}) + \delta < 1$ , then  $\text{dist}_{\mathbf{z}_+}(\mathbf{0}, \mathcal{A}_{\mathcal{Z}}(\mathbf{z}_+)) \leq (1 - \lambda_t(\mathbf{z}) - \delta)^{-1} \text{dist}_{\mathbf{z}}(\mathbf{0}, \mathcal{A}_{\mathcal{Z}}(\mathbf{z}_+))$ . Combining this inequality and the last estimate, we obtain (25). Finally, if we choose  $t \leq (1 - \lambda_t(\mathbf{z}) - \delta)(\sqrt{v} + \lambda_t(\mathbf{z}) + 2\delta)^{-1}\varepsilon$ , then  $\text{dist}_{\mathbf{z}_+}(\mathbf{0}, \mathcal{A}_{\mathcal{Z}}(\mathbf{z}_+)) \leq \varepsilon$ . Hence,  $\mathbf{z}_+$  is an  $\varepsilon$ -solution to (1) in the sense of Definition 4.  $\square$

### 7.3 The proof of Theorem 1: a key estimate of generalized Newton-type schemes

First, similar to [2], we can easily show the the following non-expansive property holds

$$\|\mathcal{P}_{\hat{\mathbf{z}}}(\mathbf{u}; t) - \mathcal{P}_{\hat{\mathbf{z}}}(\mathbf{v}; t)\|_{\hat{\mathbf{z}}} \leq \|\mathbf{u} - \mathbf{v}\|_{\hat{\mathbf{z}}}, \quad \forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^P. \tag{67}$$

Note that  $\|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}} \leq \|\bar{\mathbf{z}}_+ - \mathbf{z}\|_{\mathbf{z}} + \|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}} = \lambda_t(\mathbf{z}) + \delta(\mathbf{z}) < 1$  by our assumption. This shows that  $\mathbf{z}_+ \in \text{int}(\mathcal{Z})$  due to [31, Theorem 4.1.5 (1)].

Next, we consider the generalized gradient mappings  $G_{\mathbf{z}}(\mathbf{z}; t_+)$  and  $G_{\mathbf{z}_+}(\mathbf{z}_+; t_+)$  at  $\mathbf{z}$  and  $\mathbf{z}_+$ , respectively defined by (20) as follows:

$$\begin{aligned} G_{\mathbf{z}}(\mathbf{z}; t_+) &:= \nabla^2 F(\mathbf{z}) (\mathbf{z} - \mathcal{P}_{\mathbf{z}}(\mathbf{z} - \nabla^2 F(\mathbf{z})^{-1} \nabla F(\mathbf{z}); t_+)), \\ G_{\mathbf{z}_+}(\mathbf{z}_+; t_+) &:= \nabla^2 F(\mathbf{z}_+) (\mathbf{z}_+ - \mathcal{P}_{\mathbf{z}_+}(\mathbf{z}_+ - \nabla^2 F(\mathbf{z}_+)^{-1} \nabla F(\mathbf{z}_+); t_+)). \end{aligned} \tag{68}$$

Let  $r_{\mathbf{z}}(\bar{\mathbf{z}}_+) := \nabla F(\mathbf{z}) + \nabla^2 F(\mathbf{z})(\bar{\mathbf{z}}_+ - \mathbf{z})$ . Then, by using  $\bar{\mathbf{z}}_+ := \mathcal{P}_{\mathbf{z}}(\mathbf{z} - \nabla^2 F(\mathbf{z})^{-1} \nabla F(\mathbf{z}); t_+)$  from (26), we can show that

$$-r_{\mathbf{z}}(\bar{\mathbf{z}}_+) := -[\nabla F(\mathbf{z}) + \nabla^2 F(\mathbf{z})(\bar{\mathbf{z}}_+ - \mathbf{z})] \in t_+^{-1} \mathcal{A}(\bar{\mathbf{z}}_+). \tag{69}$$

Clearly, we can rewrite (69) as  $\bar{\mathbf{z}}_+ - \nabla^2 F(\mathbf{z}_+)^{-1} r_{\mathbf{z}}(\bar{\mathbf{z}}_+) \in \bar{\mathbf{z}}_+ + t_+^{-1} \nabla^2 F(\mathbf{z}_+)^{-1} \mathcal{A}(\bar{\mathbf{z}}_+)$ . Then, using the definition (16) of  $\mathcal{P}_{\mathbf{z}_+}(\cdot) := (\mathbb{I} + t_+^{-1} \nabla^2 F(\mathbf{z}_+)^{-1} \mathcal{A})^{-1}(\cdot)$ , we can derive

$$\mathbf{z}_+ = \mathcal{P}_{\mathbf{z}_+} \left( \bar{\mathbf{z}}_+ - \nabla^2 F(\mathbf{z}_+)^{-1} r_{\mathbf{z}}(\bar{\mathbf{z}}_+); t_+ \right) + (\mathbf{z}_+ - \bar{\mathbf{z}}_+). \quad (70)$$

Now, we can estimate  $\lambda_{t_+}(\mathbf{z}_+)$  defined by (21) using (68), (70), (67), and (69) as follows:

$$\begin{aligned} \lambda_{t_+}(\mathbf{z}_+) &:= \|G_{\mathbf{z}_+}(\mathbf{z}_+; t_+)\|_{\mathbf{z}_+}^* \stackrel{(68)}{=} \|\mathbf{z}_+ - \mathcal{P}_{\mathbf{z}_+} \left( \mathbf{z}_+ - \nabla^2 F(\mathbf{z}_+)^{-1} \nabla F(\mathbf{z}_+); t_+ \right)\|_{\mathbf{z}_+} \\ &\stackrel{(70)}{=} \left\| \mathcal{P}_{\mathbf{z}_+} \left( \bar{\mathbf{z}}_+ - \nabla^2 F(\mathbf{z}_+)^{-1} r_{\mathbf{z}}(\bar{\mathbf{z}}_+); t_+ \right) \right. \\ &\quad \left. - \mathcal{P}_{\mathbf{z}_+} \left( \mathbf{z}_+ - \nabla^2 F(\mathbf{z}_+)^{-1} \nabla F(\mathbf{z}_+); t_+ \right) + (\mathbf{z}_+ - \bar{\mathbf{z}}_+) \right\|_{\mathbf{z}_+} \\ &\leq \left\| \mathcal{P}_{\mathbf{z}_+} \left( \bar{\mathbf{z}}_+ - \nabla^2 F(\mathbf{z}_+)^{-1} r_{\mathbf{z}}(\bar{\mathbf{z}}_+); t_+ \right) \right. \\ &\quad \left. - \mathcal{P}_{\mathbf{z}_+} \left( \mathbf{z}_+ - \nabla^2 F(\mathbf{z}_+)^{-1} \nabla F(\mathbf{z}_+); t_+ \right) \right\|_{\mathbf{z}_+} \\ &\quad + \|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}_+} \\ &\stackrel{(67)}{\leq} \left\| \nabla^2 F(\mathbf{z}_+)^{-1} [\nabla F(\mathbf{z}_+) - r_{\mathbf{z}}(\bar{\mathbf{z}}_+)] + (\bar{\mathbf{z}}_+ - \mathbf{z}_+) \right\|_{\mathbf{z}_+} + \|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}_+} \\ &\stackrel{(69)}{=} \left\| \nabla^2 F(\mathbf{z}_+)^{-1} [\nabla F(\mathbf{z}_+) - \nabla F(\mathbf{z}) - \nabla^2 F(\mathbf{z})(\mathbf{z}_+ - \mathbf{z}) + (\nabla^2 F(\mathbf{z}_+) \right. \\ &\quad \left. - \nabla^2 F(\mathbf{z}))(\bar{\mathbf{z}}_+ - \mathbf{z}_+)] \right\|_{\mathbf{z}_+} + \|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}_+} \\ &\leq \|\nabla F(\mathbf{z}_+) - \nabla F(\mathbf{z}) - \nabla^2 F(\mathbf{z})(\mathbf{z}_+ - \mathbf{z})\|_{\mathbf{z}_+}^* \\ &\quad + \|(\nabla^2 F(\mathbf{z}_+) - \nabla^2 F(\mathbf{z}))(\bar{\mathbf{z}}_+ - \mathbf{z}_+)\|_{\mathbf{z}_+}^* + \|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}_+} \\ &\leq \frac{1}{1 - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}}} \left[ \|\nabla F(\mathbf{z}_+) - \nabla F(\mathbf{z}) - \nabla^2 F(\mathbf{z})(\mathbf{z}_+ - \mathbf{z})\|_{\mathbf{z}}^* \right. \\ &\quad \left. + \|(\nabla^2 F(\mathbf{z}_+) - \nabla^2 F(\mathbf{z}))(\bar{\mathbf{z}}_+ - \mathbf{z}_+)\|_{\mathbf{z}}^* \right] + \frac{\|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}}}{1 - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}}}. \end{aligned} \quad (71)$$

Here, in the last equality of (71), we have used the fact that  $\|\mathbf{w}\|_{\mathbf{z}_+}^2 = \langle \nabla^2 F(\mathbf{z}_+) \mathbf{w}, \mathbf{w} \rangle \leq (1 - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}})^{-2} \langle \nabla^2 F(\mathbf{z}) \mathbf{w}, \mathbf{w} \rangle = (1 - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}})^{-2} \|\mathbf{w}\|_{\mathbf{z}}^2$  for any  $\mathbf{w}$  and  $\mathbf{z}, \mathbf{z}_+$  such that  $\|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}} < 1$ , and the analogous fact for the dual norms. Both facts can be derived from [31, Theorem 4.1.6]. The condition  $\|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}} < 1$  is guaranteed since  $\|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}} \leq \|\mathbf{z} - \bar{\mathbf{z}}_+\|_{\mathbf{z}} + \|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}} = \lambda_{t_+}(\mathbf{z}) + \delta(\mathbf{z}) < 1$  by our assumption.

Similar to the proof of [31, Theorem 4.1.14], we can show that

$$\left\| \nabla F(\mathbf{z}_+) - \nabla F(\mathbf{z}) - \nabla^2 F(\mathbf{z})(\mathbf{z}_+ - \mathbf{z}) \right\|_{\mathbf{z}}^* \leq \frac{\|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}}^2}{1 - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}}}. \quad (72)$$

Next, we need to estimate  $B := \|(\nabla^2 F(\mathbf{z}_+) - \nabla^2 F(\mathbf{z}))(\bar{\mathbf{z}}_+ - \mathbf{z}_+)\|_{\mathbf{z}}^*$ . We define

$$\Sigma := \nabla^2 F(\mathbf{z})^{-1/2} \left( \nabla^2 F(\mathbf{z}_+) - \nabla^2 F(\mathbf{z}) \right) \nabla^2 F(\mathbf{z})^{-1/2}.$$

By [31, Theorem 4.1.6], we can show that

$$\begin{aligned}\|\Sigma\| &\leq \max \left\{ 1 - (1 - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}})^2, (1 - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}})^{-2} - 1 \right\} \\ &= \frac{2\|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}} - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}}^2}{(1 - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}})^2}.\end{aligned}$$

Using this inequality we can estimate  $B$  as

$$\begin{aligned}B^2 &= (\bar{\mathbf{z}}_+ - \mathbf{z}_+)^T \nabla^2 F(\mathbf{z})^{1/2} \Sigma^2 \nabla^2 F(\mathbf{z})^{1/2} (\bar{\mathbf{z}}_+ - \mathbf{z}_+) \leq \|\Sigma\|^2 \|\bar{\mathbf{z}}_+ - \mathbf{z}_+\|_{\mathbf{z}}^2 \\ &\leq \left( \frac{2\|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}} - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}}^2}{(1 - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}})^2} \right)^2 \|\bar{\mathbf{z}}_+ - \mathbf{z}_+\|_{\mathbf{z}}^2,\end{aligned}$$

which implies

$$B \leq \left( \frac{2\|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}} - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}}^2}{(1 - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}})^2} \right) \|\bar{\mathbf{z}}_+ - \mathbf{z}_+\|_{\mathbf{z}}. \quad (73)$$

Substituting (72) and (73) into (71) we get

$$\begin{aligned}\lambda_{t+}(\mathbf{z}_+) &\leq \frac{\|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}}^2}{(1 - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}})^2} \\ &\quad + \frac{[2\|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}} - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}}^2] \|\bar{\mathbf{z}}_+ - \mathbf{z}_+\|_{\mathbf{z}}}{(1 - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}})^3} + \frac{\|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}}}{1 - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}}} \\ &= \frac{\|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}}^2}{(1 - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}})^2} + \frac{\|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}}}{(1 - \|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}})^3}.\end{aligned} \quad (74)$$

Finally, we note that  $\lambda_{t+}(\mathbf{z}) := \|G_{\mathbf{z}}(\mathbf{z}; t_+)\|_{\mathbf{z}}^* = \|\mathbf{z} - \mathcal{P}_{\mathbf{z}}(\mathbf{z} - \nabla^2 F(\mathbf{z})^{-1} \nabla F(\mathbf{z}); t_+)\|_{\mathbf{z}} = \|\mathbf{z} - \bar{\mathbf{z}}_+\|_{\mathbf{z}}$  due to (26). Using the triangle inequality we have  $\|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}} \leq \|\mathbf{z} - \bar{\mathbf{z}}_+\|_{\mathbf{z}} + \|\mathbf{z}_+ - \bar{\mathbf{z}}_+\|_{\mathbf{z}} = \lambda_{t+}(\mathbf{z}) + \delta(\mathbf{z}) < 1$ . Since the right-hand side of (74) is monotonically increasing with respect to  $\|\mathbf{z}_+ - \mathbf{z}\|_{\mathbf{z}}$ , using the last inequality into (74), we obtain (27).  $\square$

## 7.4 The proof of Theorem 2: local quadratic convergence of FGN

We first prove (a). Given a fixed parameter  $t > 0$  sufficiently small, our objective is to find  $\beta \in (0, 1)$  such that if  $\lambda_t(\mathbf{z}^k) \leq \beta$ , then  $\lambda_t(\mathbf{z}^{k+1}) \leq \beta$ . Indeed, using the key estimate (27) with  $t$  instead of  $t_+$ , we can see that to guarantee  $\lambda_t(\mathbf{z}^{k+1}) \leq \beta$ , we require

$$\left( \frac{\lambda_t(\mathbf{z}^k) + \delta(\mathbf{z}^k)}{1 - \lambda_t(\mathbf{z}^k) - \delta(\mathbf{z}^k)} \right)^2 + \frac{\delta(\mathbf{z}^k)}{(1 - \lambda_t(\mathbf{z}^k) - \delta(\mathbf{z}^k))^3} \leq \beta.$$

Since the left-hand side of this inequality is monotonically increasing when  $\lambda_t(\mathbf{z}^k)$  and  $\delta(\mathbf{z}^k)$  are increasing, we can overestimate it by

$$\left( \frac{\beta + \delta}{1 - \beta - \delta} \right)^2 + \frac{\delta}{(1 - \beta - \delta)^3} \leq \beta.$$

Using the identity  $\frac{\beta + \delta}{1 - \beta - \delta} = \frac{\beta}{1 - \beta} + \frac{\delta}{(1 - \beta)(1 - \beta - \delta)}$ , we can write the last inequality as

$$\left[ \frac{2\beta}{(1 - \beta)^2(1 - \beta - \delta)} + \frac{\delta}{(1 - \beta)^2(1 - \beta - \delta)^2} + \frac{1}{(1 - \beta - \delta)^3} \right] \delta \leq \beta - \left( \frac{\beta}{1 - \beta} \right)^2. \quad (75)$$

Clearly, the left-hand side of (75) is positive if  $0 < \delta < 1 - \beta$ . Hence, we need to choose  $\beta \in (0, 0.5(3 - \sqrt{5}))$  such that the right-hand side of (75) is also positive. Now, we choose  $\delta \geq 0$  such that  $\delta \leq \beta(1 - \beta) < 1 - \beta$ . Then, (75) can be one more time overestimated by

$$\left( \frac{2\beta^3 - 5\beta^2 + 3\beta + 1}{(1 - \beta)^4} \right) \delta \leq \beta(1 - 3\beta + \beta^2),$$

which implies

$$0 \leq \delta \leq \frac{\beta(1 - 3\beta + \beta^2)(1 - \beta)^4}{2\beta^3 - 5\beta^2 + 3\beta + 1} < \beta(1 - \beta), \quad \forall \beta \in (0, 0.5(3 - \sqrt{5})).$$

This inequality suggests that we can choose  $\delta := \frac{\beta(1 - 3\beta + \beta^2)(1 - \beta)^4}{2\beta^3 - 5\beta^2 + 3\beta + 1} > 0$ . In this case, we also have  $\delta(\mathbf{z}) + \lambda_t(\mathbf{z}) \leq \delta + \beta < 1$ , which guarantees the condition of Theorem 1. Hence, we can conclude that  $\lambda_t(\mathbf{z}^k) \leq \beta$  implies  $\lambda_t(\mathbf{z}^{k+1}) \leq \beta$ . In other words,  $\{\mathbf{z}^k\}$  belongs to  $\mathcal{Q}_t(\beta)$ .

(b) Next, to guarantee a quadratic convergence, we can choose  $\delta_k$  such that  $\delta(\mathbf{z}^k) \leq \delta_k \leq \bar{\delta}_k := \frac{\lambda_t(\mathbf{z}^k)^2}{1 - \lambda_t(\mathbf{z}^k)}$ . Substituting the upper bound  $\bar{\delta}_k$  of  $\delta(\mathbf{z}^k)$  into (27) we obtain

$$\lambda_t(\mathbf{z}^{k+1}) \leq \left( \frac{2 - 4\lambda_t(\mathbf{z}^k) + \lambda_t(\mathbf{z}^k)}{(1 - 2\lambda_t(\mathbf{z}^k))^3} \right) \lambda_t(\mathbf{z}^k)^2.$$

Let us consider the function  $s(r) := \frac{(2 - 4r + r^2)r^2}{(1 - 2r)^3}$  on  $[0, 1]$ . We can easily check that  $s(r) < 1$  for all  $r \in [0, 1]$ . Hence,  $\lambda_t(\mathbf{z}^{k+1}) < 1$  as long as  $\lambda_t(\mathbf{z}^k) < 1$ . This proves the estimate (30).

Now, let us choose some  $\beta \in (0, 1)$  such that  $\lambda_t(\mathbf{z}^k) \leq \beta$ . Then (30) leads to

$$\lambda_t(\mathbf{z}^{k+1}) \leq \left( \frac{2 - 4\beta + \beta^2}{(1 - 2\beta)^3} \right) \lambda_t(\mathbf{z}^k)^2 = c\lambda_t(\mathbf{z})^2,$$

where  $c := \frac{2-4\beta+\beta^2}{(1-2\beta)^3} > 0$ . We need to choose  $\beta \in (0, 1)$  such that  $c\lambda_t(\mathbf{z}^k) < 1$ . Since  $\lambda_t(\mathbf{z}^k) \leq \beta$ , we choose  $\beta$  such that  $c\beta < 1$ , which is equivalent to  $9\beta^3 - 16\beta^2 + 8\beta - 1 < 0$ . If  $\beta \in (0, 0.18858]$ , then  $9\beta^3 - 16\beta^2 + 8\beta - 1 < 0$ . Therefore, the radius of the quadratic convergence region of  $\{\lambda_t(\mathbf{z}^k)\}$  is  $r := 0.18858$ .

(c) Finally, for any  $\beta \in (0, 0.18858]$ , we can write  $c\lambda_t(\mathbf{z}^{k+1}) \leq (c\lambda_t(\mathbf{z}^k))^2$ . By induction,  $c\lambda_t(\mathbf{z}^k) \leq (c\lambda_t(\mathbf{z}^0))^{2^k} \leq c^{2^k}\beta^{2^k} < 1$ . We obtain  $\lambda_t(\mathbf{z}^k) \leq c^{2^{k-1}}\beta^{2^k}$ . Let us choose  $\delta_k := \frac{\lambda_t(\mathbf{z}^k)^2}{1-\lambda_t(\mathbf{z}^k)}$ . For  $\epsilon \in (0, \beta)$ , assume that  $c^{2^{k-1}}\beta^{2^k} \leq \epsilon$ . From Lemma 3, we can choose  $t := (1-\epsilon)(\sqrt{v} + \epsilon + 2\epsilon^2/(1-\epsilon))^{-1}\epsilon$ . Then,  $\mathbf{z}^k$  is an  $\epsilon$ -solution of (1). It remains to use the fact that  $c^{2^{k-1}}\beta^{2^k} \leq \epsilon$  to upper bound the number of iterations  $k := \mathcal{O}(\ln(\ln(1/\epsilon)))$ .  $\square$

## 7.5 The proof of Theorem 3: local quadratic convergence of DGN

(a) Given a fixed parameter  $t > 0$  sufficiently small, it follows from DGN and (70) that

$$\begin{aligned}\bar{\mathbf{z}}^{k+2} &= \mathcal{P}_{\mathbf{z}^{k+1}}(\mathbf{z}^{k+1} - \nabla^2 F(\mathbf{z}^{k+1})^{-1}\nabla F(\mathbf{z}^{k+1}); t), \\ \mathbf{z}^{k+1} &= \mathcal{P}_{\mathbf{z}^{k+1}}(\bar{\mathbf{z}}^{k+1} - \nabla^2 F(\mathbf{z}^{k+1})^{-1}r_{\mathbf{z}^k}(\bar{\mathbf{z}}^{k+1}); t) + (\mathbf{z}^{k+1} - \bar{\mathbf{z}}^{k+1}).\end{aligned}$$

Hence, using these notations and the same proof as (74) with  $t$  instead of  $t_+$ , and assuming  $\|\mathbf{z}^{k+1} - \mathbf{z}^k\|_{\mathbf{z}^k} < 1$ , we can derive

$$\|\bar{\mathbf{z}}^{k+2} - \mathbf{z}^{k+1}\|_{\mathbf{z}^{k+1}} \leq \left( \frac{\|\mathbf{z}^{k+1} - \mathbf{z}^k\|_{\mathbf{z}^k}}{1 - \|\mathbf{z}^{k+1} - \mathbf{z}^k\|_{\mathbf{z}^k}} \right)^2 + \frac{\|\mathbf{z}^{k+1} - \bar{\mathbf{z}}^{k+1}\|_{\mathbf{z}^k}}{(1 - \|\mathbf{z}^{k+1} - \mathbf{z}^k\|_{\mathbf{z}^k})^3}. \quad (76)$$

Now, let us define  $\tilde{\lambda}_t(\mathbf{z}^k) := \|\tilde{\mathbf{z}}^{k+1} - \mathbf{z}^k\|_{\mathbf{z}^k}$  and  $\alpha_k := (1 + \tilde{\lambda}_t(\mathbf{z}^k))^{-1}$  as in DGN. From the update  $\mathbf{z}^{k+1} := (1 - \alpha_k)\mathbf{z}^k + \alpha_k\tilde{\mathbf{z}}^{k+1}$  of DGN, we have

$$\begin{aligned}\|\mathbf{z}^{k+1} - \mathbf{z}^k\|_{\mathbf{z}^k} &= \alpha_k \|\tilde{\mathbf{z}}^{k+1} - \mathbf{z}^k\|_{\mathbf{z}^k} = \alpha_k \tilde{\lambda}_t(\mathbf{z}^k), \quad \text{and} \\ \|\mathbf{z}^{k+1} - \bar{\mathbf{z}}^{k+1}\|_{\mathbf{z}^k} &\leq \|\mathbf{z}^{k+1} - \tilde{\mathbf{z}}^{k+1}\|_{\mathbf{z}^k} + \|\tilde{\mathbf{z}}^{k+1} - \bar{\mathbf{z}}^{k+1}\|_{\mathbf{z}^k} \\ &= (1 - \alpha_k) \|\tilde{\mathbf{z}}^{k+1} - \mathbf{z}^k\|_{\mathbf{z}^k} + \delta(\mathbf{z}^k) \\ &= (1 - \alpha_k) \tilde{\lambda}_t(\mathbf{z}^k) + \delta(\mathbf{z}^k).\end{aligned}$$

Substituting these expressions into (76) we get

$$\|\bar{\mathbf{z}}^{k+2} - \mathbf{z}^{k+1}\|_{\mathbf{z}^{k+1}} \leq \left( \frac{\alpha_k \tilde{\lambda}_t(\mathbf{z}^k)}{1 - \alpha_k \tilde{\lambda}_t(\mathbf{z}^k)} \right)^2 + \frac{\delta(\mathbf{z}^k) + (1 - \alpha_k) \tilde{\lambda}_t(\mathbf{z}^k)}{\left(1 - \alpha_k \tilde{\lambda}_t(\mathbf{z}^k)\right)^3}.$$

Substituting  $\alpha_k := (1 + \tilde{\lambda}_t(\mathbf{z}^k))^{-1}$  into the last inequality and simplifying the result, we get

$$\|\bar{\mathbf{z}}^{k+2} - \mathbf{z}^{k+1}\|_{\mathbf{z}^{k+1}} \leq \left( 2 + 2\tilde{\lambda}_t(\mathbf{z}^k) + \tilde{\lambda}_t(\mathbf{z}^k)^2 \right) \tilde{\lambda}_t(\mathbf{z}^k)^2 + \left( 1 + \tilde{\lambda}_t(\mathbf{z}^k) \right)^3 \delta(\mathbf{z}^k).$$

Next, by the triangle inequality, it follows from (68) and the definition of  $\lambda_t(\mathbf{z})$  and  $\tilde{\lambda}_t(\mathbf{z})$  that  $\tilde{\lambda}_t(\mathbf{z}^{k+1}) = \|\tilde{\mathbf{z}}^{k+2} - \mathbf{z}^{k+1}\|_{\mathbf{z}^{k+1}} \leq \|\bar{\mathbf{z}}^{k+2} - \mathbf{z}^{k+1}\|_{\mathbf{z}^{k+1}} + \|\tilde{\mathbf{z}}^{k+2} - \bar{\mathbf{z}}^{k+2}\|_{\mathbf{z}^{k+1}} = \|\bar{\mathbf{z}}^{k+2} - \mathbf{z}^{k+1}\|_{\mathbf{z}^{k+1}} + \delta(\mathbf{z}^{k+1})$ . Combining this estimate and the above inequality we get

$$\tilde{\lambda}_t(\mathbf{z}^{k+1}) \leq \left(2 + 2\tilde{\lambda}_t(\mathbf{z}^k) + \tilde{\lambda}_t(\mathbf{z}^k)^2\right) \tilde{\lambda}_t(\mathbf{z}^k)^2 + \left(1 + \tilde{\lambda}_t(\mathbf{z}^k)\right)^3 \delta(\mathbf{z}^k) + \delta(\mathbf{z}^{k+1}).$$

If we choose  $\delta(\mathbf{z}^k) \leq \delta_k \leq \frac{\tilde{\lambda}_t(\mathbf{z}^k)^2}{1+\tilde{\lambda}_t(\mathbf{z}^k)}$ , then, by induction,  $\delta(\mathbf{z}^{k+1}) \leq \delta_{k+1} \leq \frac{\tilde{\lambda}_t(\mathbf{z}^{k+1})^2}{1+\tilde{\lambda}_t(\mathbf{z}^{k+1})}$ . Substituting these bounds into the last inequality and simplifying the result, we obtain

$$\tilde{\lambda}_t(\mathbf{z}^{k+1}) \leq \left( \frac{2\tilde{\lambda}_t(\mathbf{z}^k)^2 + 4\tilde{\lambda}_t(\mathbf{z}^k) + 3}{1 - \tilde{\lambda}_t(\mathbf{z}^k)^2 (2\tilde{\lambda}_t(\mathbf{z}^k)^2 + 4\tilde{\lambda}_t(\mathbf{z}^k) + 3)} \right) \tilde{\lambda}_t(\mathbf{z}^k)^2,$$

which is indeed (31).

From (31), after a few elementary calculations, we can see that  $\tilde{\lambda}_t(\mathbf{z}^{k+1}) \leq \tilde{\lambda}_t(\mathbf{z}^k)$  if  $\tilde{\lambda}_t(\mathbf{z}^k)(1 + \tilde{\lambda}_t(\mathbf{z}^k))(2\tilde{\lambda}_t(\mathbf{z}^k)^2 + 4\tilde{\lambda}_t(\mathbf{z}^k) + 3) \leq 1$ . Note that the function  $s(\tau) := \tau(1 + \tau)(2\tau^2 + 4\tau + 3)$  is increasing on  $[0, 0.5(3 - \sqrt{5})]$ . By numerically computing  $\tilde{\lambda}_t(\mathbf{z}^k)$  we can observe that if  $\tilde{\lambda}_t(\mathbf{z}^k) \in [0, 0.21027]$ , then  $\tilde{\lambda}_t(\mathbf{z}^{k+1}) \leq \tilde{\lambda}_t(\mathbf{z}^k)$ . Hence, if  $\tilde{\lambda}_t(\mathbf{z}^k) \leq \beta$  then  $\tilde{\lambda}_t(\mathbf{z}^{k+1}) \leq \beta$ . In other words, we can say that  $\{\mathbf{z}^k\} \subset \Omega_t(\beta)$ .

We now prove (b). Indeed, if we take any  $\beta \in (0, 0.21027]$ , we can show from (31) that

$$\tilde{\lambda}_t(\mathbf{z}^{k+1}) \leq \left( \frac{2\beta^2 + 4\beta + 3}{1 - \beta^2 (2\beta^2 + 4\beta + 3)} \right) \tilde{\lambda}_t(\mathbf{z}^k)^2,$$

where  $\bar{c} := \left( \frac{2\beta^2 + 4\beta + 3}{1 - \beta^2 (2\beta^2 + 4\beta + 3)} \right) \in (0, +\infty)$ . To guarantee  $\bar{c}\beta < 1$ , we need to choose  $\beta > 0$  such that  $2\beta^4 + 6\beta^3 + 7\beta^2 + 3\beta - 1 < 0$ . This condition leads to  $\beta \in (0, 0.21027]$ . Hence, for any  $0 < \beta \leq 0.21027$ , if  $\mathbf{z}^0 \in \mathcal{Q}_t(\beta)$ , then  $\tilde{\lambda}_t(\mathbf{z}^{k+1}) \leq \bar{c}\tilde{\lambda}_t(\mathbf{z}^k)^2 < 1$  and, therefore,  $\{\tilde{\lambda}_t(\mathbf{z}^k)\}$  quadratically converges to zero.

(c) To prove the last conclusion in (c), from (66), we can show that

$$\begin{aligned} \text{dist}_{\mathbf{z}^k}(\mathbf{0}, \mathcal{A}_{\mathcal{Z}}(\mathbf{z}^{k+1})) &\leq t\delta_k + t \left\| \nabla F(\mathbf{z}^k) \right\|_{\mathbf{z}^k}^* + t \left\| \mathbf{z}^{k+1} - \mathbf{z}^k \right\|_{\mathbf{z}^k} \\ &\leq t(\delta_k + \sqrt{v} + \alpha_k \tilde{\lambda}_t(\mathbf{z}^k)). \end{aligned}$$

Since  $\tilde{\lambda}_t(\mathbf{z}^k) \leq \bar{c}^{2^k-1} \lambda_t(\mathbf{z}^0)^{2^k} \leq \bar{c}^{2^k-1} \beta^{2^k}$ ,  $\delta_k \leq \frac{\tilde{\lambda}_t(\mathbf{z}^k)^2}{1+\tilde{\lambda}_t(\mathbf{z}^k)}$ , and  $\alpha_k = \frac{\tilde{\lambda}_t(\mathbf{z}^k)}{1+\tilde{\lambda}_t(\mathbf{z}^k)}$ , we obtain the last conclusion as a consequence of Lemma 3 with the same proof as in Theorem 2.  $\square$

## 7.6 The proof of Lemma 4: the update rule for the penalty parameter

Let us define  $\bar{\mathbf{u}}^k := \mathcal{P}_{\mathbf{z}^k}(\mathbf{z}^k - \nabla^2 F(\mathbf{z}^k)^{-1} \nabla F(\mathbf{z}^k); t_k)$ . Then,  $\lambda_{t_k}(\mathbf{z}^k)$  defined by (21) becomes  $\lambda_{t_k}(\mathbf{z}^k) := \|G_{\mathbf{z}^k}(\mathbf{z}^k; t_k)\|_{\mathbf{z}^k}^* = \|\mathbf{z}^k - \mathcal{P}_{\mathbf{z}^k}(\mathbf{z}^k - \nabla^2 F(\mathbf{z}^k)^{-1} \nabla F(\mathbf{z}^k); t_k)\|_{\mathbf{z}^k} = \|\mathbf{z}^k - \bar{\mathbf{u}}^k\|_{\mathbf{z}^k}$ . Note that  $\bar{\mathbf{u}}^k = \mathcal{P}_{\mathbf{z}^k}(\mathbf{z}^k - \nabla^2 F(\mathbf{z}^k)^{-1} \nabla F(\mathbf{z}^k); t_k)$  leads to

$$-t_k \left( \nabla F(\mathbf{z}^k) + \nabla^2 F(\mathbf{z}^k)(\bar{\mathbf{u}}^k - \mathbf{z}^k) \right) \in \mathcal{A}(\bar{\mathbf{u}}^k).$$

Combining this inclusion and (69) and using the monotonicity of  $\mathcal{A}$ , we can derive

$$\langle t_{k+1} \left[ \nabla F(\mathbf{z}^k) + \nabla^2 F(\mathbf{z}^k)(\bar{\mathbf{z}}^{k+1} - \mathbf{z}^k) \right] - t_k \left[ \nabla F(\mathbf{z}^k) + \nabla^2 F(\mathbf{z}^k)(\bar{\mathbf{u}}^k - \mathbf{z}^k) \right], \bar{\mathbf{z}}^{k+1} - \bar{\mathbf{u}}^k \rangle \leq 0.$$

By rearranging this expression using  $t_{k+1} := (1 - \sigma_\beta)t_k$  from PFGN, we finally obtain

$$\begin{aligned} \|\bar{\mathbf{z}}^{k+1} - \bar{\mathbf{u}}^k\|_{\mathbf{z}^k}^2 &\leq \frac{\sigma_\beta}{1 - \sigma_\beta} \langle \nabla F(\mathbf{z}^k) + \nabla^2 F(\mathbf{z}^k)(\bar{\mathbf{u}}^k - \mathbf{z}^k), \bar{\mathbf{z}}^{k+1} - \bar{\mathbf{u}}^k \rangle \\ &\leq \frac{\sigma_\beta}{1 - \sigma_\beta} \|\nabla F(\mathbf{z}^k) + \nabla^2 F(\mathbf{z}^k)(\bar{\mathbf{u}}^k - \mathbf{z}^k)\|_{\mathbf{z}^k}^* \|\bar{\mathbf{z}}^{k+1} - \bar{\mathbf{u}}^k\|_{\mathbf{z}^k}, \end{aligned}$$

where the last inequality follows from the elementary Cauchy–Schwarz inequality. This inequality eventually leads to

$$\begin{aligned} \|\bar{\mathbf{z}}^{k+1} - \bar{\mathbf{u}}^k\|_{\mathbf{z}^k} &\leq \frac{\sigma_\beta}{1 - \sigma_\beta} \|\nabla F(\mathbf{z}^k) + \nabla^2 F(\mathbf{z}^k)(\bar{\mathbf{u}}^k - \mathbf{z}^k)\|_{\mathbf{z}^k}^* \\ &\leq \frac{\sigma_\beta}{1 - \sigma_\beta} \left[ \|\nabla F(\mathbf{z}^k)\|_{\mathbf{z}^k}^* + \|\nabla^2 F(\mathbf{z}^k)(\bar{\mathbf{u}}^k - \mathbf{z}^k)\|_{\mathbf{z}^k}^* \right] \\ &\leq \frac{\sigma_\beta}{1 - \sigma_\beta} \left[ \|\nabla F(\mathbf{z}^k)\|_{\mathbf{z}^k}^* + \|\bar{\mathbf{u}}^k - \mathbf{z}^k\|_{\mathbf{z}^k} \right]. \end{aligned}$$

Now, by the triangle inequality, we have  $\|\bar{\mathbf{z}}^{k+1} - \mathbf{z}^k\|_{\mathbf{z}^k} \leq \|\bar{\mathbf{z}}^{k+1} - \bar{\mathbf{u}}^k\|_{\mathbf{z}^k} + \|\bar{\mathbf{u}}^k - \mathbf{z}^k\|_{\mathbf{z}^k}$ . This inequality is equivalent to  $\lambda_{t_{k+1}}(\mathbf{z}^k) \leq \|\bar{\mathbf{z}}^{k+1} - \bar{\mathbf{u}}^k\|_{\mathbf{z}^k} + \lambda_{t_k}(\mathbf{z}^k)$  due to the definitions  $\lambda_{t_{k+1}}(\mathbf{z}^k) = \|\bar{\mathbf{z}}^{k+1} - \mathbf{z}^k\|_{\mathbf{z}^k}$  and  $\lambda_{t_k}(\mathbf{z}^k) = \|\bar{\mathbf{u}}^k - \mathbf{z}^k\|_{\mathbf{z}^k}$ . Using the last estimate in the above inequality we get

$$\lambda_{t_{k+1}}(\mathbf{z}^k) \leq \lambda_{t_k}(\mathbf{z}^k) + \frac{\sigma_\beta}{1 - \sigma_\beta} \left[ \|\nabla F(\mathbf{z}^k)\|_{\mathbf{z}^k}^* + \lambda_{t_k}(\mathbf{z}^k) \right],$$

which is (32). The second inequality of (32) follows from the fact that  $\|\nabla F(\mathbf{z}^k)\|_{\mathbf{z}^k}^* \leq \sqrt{\nu}$ .

Let us denote by  $\gamma_k := \left( \frac{\sigma_\beta}{1 - \sigma_\beta} \right) (\sqrt{\nu} + \lambda_{t_k}(\mathbf{z}^k))$ . For a given  $\beta \in (0, 1)$ , we now assume that  $\lambda_{t_k}(\mathbf{z}^k) \leq \beta$ . Then, by using (32) in (27), and the monotonic increase of its right-hand side with respect to  $\lambda_{t_{k+1}}(\mathbf{z}^k)$ , we can derive

$$\begin{aligned} \lambda_{t_{k+1}}(\mathbf{z}^{k+1}) &\leq \left( \frac{\lambda_{t_k}(\mathbf{z}^k) + |\gamma_k| + \delta_k}{1 - \lambda_{t_k}(\mathbf{z}^k) - |\gamma_k| - \delta_k} \right)^2 + \frac{\delta_k}{(1 - \lambda_{t_k}(\mathbf{z}^k) - |\gamma_k| - \delta_k)^3} \\ &\leq \left( \frac{\beta + |\gamma_k| + \delta_k}{1 - \beta - |\gamma_k| - \delta_k} \right)^2 + \frac{\delta_k}{(1 - \beta - |\gamma_k| - \delta_k)^3}, \end{aligned}$$

as long as  $\beta + |\gamma_k| + \delta_k < 1$ . Let us denote  $\theta_k := \beta + |\gamma_k|$ . By using the identity  $\frac{\beta + |\gamma_k| + \delta_k}{1 - \beta - |\gamma_k| - \delta_k} = \frac{\beta + |\gamma_k|}{1 - \beta - |\gamma_k|} + \frac{\delta_k}{(1 - \theta_k)(1 - \theta_k - \delta_k)}$ , we can rewrite the last inequality as

$$\lambda_{t_{k+1}}(\mathbf{z}^{k+1}) \leq \left( \frac{\theta_k}{1 - \theta_k} \right)^2 + \left[ \frac{2\theta_k}{(1 - \theta_k)^2(1 - \theta_k - \delta_k)} + \frac{\delta_k}{(1 - \theta_k)^2(1 - \theta_k - \delta_k)^2} + \frac{1}{(1 - \theta_k - \delta_k)^3} \right] \delta_k.$$

If we choose  $\delta_k$  such that  $0 \leq \delta_k \leq \theta_k(1 - \theta_k) < 1 - \theta_k$ , then the above inequality implies

$$\begin{aligned} \lambda_{t_{k+1}}(\mathbf{z}^{k+1}) &\leq \left( \frac{\theta_k}{1 - \theta_k} \right)^2 \\ &\quad + \left[ \frac{2\theta_k(1 - \theta_k)^2 + \theta_k(1 - \theta_k) + 1}{(1 - \theta_k)^6} \right] \delta_k := \left( \frac{\theta_k}{1 - \theta_k} \right)^2 + M_k \delta_k. \end{aligned} \quad (77)$$

Take any  $c \in (0, 1)$ , e.g.,  $c := 0.95$ , and choose  $\delta_k$  such that  $0 \leq \delta_k \leq \frac{(1-c^2)}{c^2 M_k} \left( \frac{\theta_k}{1 - \theta_k} \right)^2$ .

Hence, in order to guarantee  $\lambda_{t_{k+1}}(\mathbf{z}^{k+1}) \leq \beta$ , by using (77), we can impose the condition  $\left( \frac{\theta_k}{1 - \theta_k} \right)^2 + M_k \delta_k \leq \frac{1}{c^2} \left( \frac{\theta_k}{1 - \theta_k} \right)^2 \leq \beta$ , which is equivalent to  $\frac{\theta_k}{1 - \theta_k} \leq c\sqrt{\beta}$ .

This condition leads to  $\theta_k \geq \frac{c\sqrt{\beta}}{1+c\sqrt{\beta}}$ , and therefore,  $|\gamma_k| \leq \frac{c\sqrt{\beta}}{1+c\sqrt{\beta}} - \beta$ . Since  $|\gamma_k| > 0$ , we need to choose  $\beta$  such that  $0 < \beta < 0.5(1 + 2c^2 - \sqrt{1 + 4c^2})$ .

Next, by the choice of  $\delta_k$ , we require  $0 \leq \delta_k \leq \min \left\{ \frac{(1-c^2)}{c^2 M_k} \left( \frac{\theta_k}{1 - \theta_k} \right)^2, \theta_k(1 - \theta_k) \right\}$ .

Using the fact that  $M_k = \frac{2\theta_k(1 - \theta_k)^2 + \theta_k(1 - \theta_k) + 1}{(1 - \theta_k)^6}$  from (77) and  $0 \leq \theta_k \leq \frac{c\sqrt{\beta}}{1+c\sqrt{\beta}}$ , we can show that the condition on  $\delta_k$  holds if we choose

$$\delta_k \leq \bar{\delta} := \frac{(1 - c^2)\beta}{(1 + c\sqrt{\beta})^3 [3c\sqrt{\beta} + c^2\beta + (1 + c\sqrt{\beta})^3]}.$$

On the other hand, we have  $|\gamma_k| = \left| \left( \frac{\sigma_\beta}{1 - \sigma_\beta} \right) (\sqrt{\nu} + \lambda_{t_k}(\mathbf{z}^k)) \right| \leq \left( \frac{\sigma_\beta}{1 - \sigma_\beta} \right) (\sqrt{\nu} + \beta)$ .

In order to guarantee that  $|\gamma_k| \leq \frac{c\sqrt{\beta}}{1+c\sqrt{\beta}} - \beta$ , we use the above estimate to impose a condition  $\left( \frac{\sigma_\beta}{1 - \sigma_\beta} \right) \leq \frac{1}{\sqrt{\nu} + \beta} \left( \frac{c\sqrt{\beta}}{1+c\sqrt{\beta}} - \beta \right)$ , which leads to

$$\sigma_\beta \leq \bar{\sigma}_\beta := \frac{c\sqrt{\beta} - \beta(1 + c\sqrt{\beta})}{(1 + c\sqrt{\beta})\sqrt{\nu} + c\sqrt{\beta}}.$$

This estimate is exactly the right-hand side of (33). Finally, using (32) and the definition of  $\gamma_k$ , we can easily show that  $\lambda_{t_{k+1}}(\mathbf{z}^k) \leq \lambda_{t_k}(\mathbf{z}^k) + |\gamma_k| \leq \beta + |\gamma_k| \equiv \theta_k \leq \frac{c\sqrt{\beta}}{1+c\sqrt{\beta}}$ .  $\square$

## 7.7 The proof of Theorem 4: the worst-case iteration-complexity of PFGN

By Lemma 3 and  $\lambda_{t_{k+1}}(\mathbf{z}^k) \leq \frac{c\sqrt{\beta}}{1+c\sqrt{\beta}}$ , we can see that  $\mathbf{z}^k$  is an  $\varepsilon$ -solution of (1) if  $t_k := M_0^{-1}\varepsilon$ , where  $M_0 := \left(1 - \frac{c\sqrt{\beta}}{1+c\sqrt{\beta}}\right)^{-1} \left(\sqrt{v} + \frac{c\sqrt{\beta}}{1+c\sqrt{\beta}} + 2\bar{\delta}_t(\beta)\right) = \mathcal{O}(\sqrt{v})$ .

On the other hand, by induction, it follows from the update rule  $t_{k+1} = (1 - \sigma_\beta)t_k$  of PFGN that  $t_k = (1 - \sigma_\beta)^k t_0$ . Hence,  $\mathbf{z}^k$  is an  $\varepsilon$ -solution of (1) if we have  $t_k = (1 - \sigma_\beta)^k t_0 \leq \frac{\varepsilon}{M_0}$ . This condition leads to  $k \ln(1 - \sigma_\beta) \geq \ln\left(\frac{\varepsilon}{M_0 t_0}\right)$ , which implies  $k \leq \frac{\ln(\varepsilon/(M_0 t_0))}{\ln(1 - \sigma_\beta)}$ . Using an elementary inequality  $\ln(1 - \sigma_\beta) \leq -\sigma_\beta$ , we can upper bound  $k$  as

$$k \geq \frac{1}{\bar{\sigma}_\beta} \ln\left(\frac{M_0 t_0}{\varepsilon}\right) = \frac{(1 + c\sqrt{\beta})\sqrt{v} + c\sqrt{\beta}}{c\sqrt{\beta} - \beta(1 + c\sqrt{\beta})} \ln\left(\frac{M_0 t_0}{\varepsilon}\right).$$

Consequently, the worst-case iteration-complexity of PFGN is  $\mathcal{O}\left(\sqrt{v} \ln\left(\frac{\sqrt{v} t_0}{\varepsilon}\right)\right)$ .  $\square$

## 7.8 The proof of Theorem 5: finding an initial point for PFGN

From (35), if we define  $\nabla \hat{F}(\hat{\mathbf{z}}^j) := \nabla F(\hat{\mathbf{z}}^j) - t_0^{-1} \tau_{k+1} \xi_0$ , then we still have  $\nabla^2 \hat{F}(\hat{\mathbf{z}}^j) = \nabla^2 F(\hat{\mathbf{z}}^j)$ . Hence, the estimate (27) still holds for  $\hat{\lambda}_\tau(\hat{\mathbf{z}}^j)$ .

Next, if we define  $\bar{\mathbf{v}}^j := \mathcal{P}_{\hat{\mathbf{z}}^j} \left( \hat{\mathbf{z}}^j - \nabla^2 F(\hat{\mathbf{z}}^j)^{-1} \left( \nabla F(\hat{\mathbf{z}}^j) - \tau_j t_0^{-1} \hat{\xi}^0 \right); t_0 \right)$ , then, by the definition of  $\mathcal{P}_{\hat{\mathbf{z}}^j}$ , we have

$$-t_0 \left[ \nabla^2 F(\hat{\mathbf{z}}^j)(\bar{\mathbf{v}}^j - \hat{\mathbf{z}}^j) + \nabla F(\hat{\mathbf{z}}^j) - \tau_j t_0^{-1} \hat{\xi}^0 \right] \in \mathcal{A}(\bar{\mathbf{v}}^j). \quad (78)$$

Similarly, since  $\bar{\hat{\mathbf{z}}}^{j+1} := \mathcal{P}_{\hat{\mathbf{z}}^j} \left( \hat{\mathbf{z}}^j - \nabla^2 F(\hat{\mathbf{z}}^j)^{-1} \left( \nabla F(\hat{\mathbf{z}}^j) - \tau_{j+1} t_0^{-1} \hat{\xi}^0 \right); t_0 \right)$ , we have

$$-t_0 \left[ \nabla^2 F(\hat{\mathbf{z}}^j)(\bar{\hat{\mathbf{z}}}^{j+1} - \hat{\mathbf{z}}^j) + \nabla F(\hat{\mathbf{z}}^j) - \tau_{j+1} t_0^{-1} \hat{\xi}_0 \right] \in \mathcal{A}(\bar{\hat{\mathbf{z}}}^{j+1}). \quad (79)$$

Using (78), (79), and the monotonicity of  $\mathcal{A}$ , we have

$$t_0 \langle \nabla^2 F(\hat{\mathbf{z}}^j)(\bar{\hat{\mathbf{z}}}^{j+1} - \bar{\mathbf{v}}^j), \bar{\hat{\mathbf{z}}}^{j+1} - \bar{\mathbf{v}}^j \rangle \leq (\tau_j - \tau_{j+1}) \langle \hat{\xi}_0, \bar{\mathbf{v}}^j - \bar{\hat{\mathbf{z}}}^{j+1} \rangle.$$

Using  $\tau_{j+1} := \tau_j - \Delta_j$  and the Cauchy–Schwarz inequality, the last inequality leads to

$$t_0 \left\| \bar{\hat{\mathbf{z}}}^{j+1} - \bar{\mathbf{v}}^j \right\|_{\hat{\mathbf{z}}^j} \leq \Delta_j \|\hat{\xi}_0\|_{\hat{\mathbf{z}}^j}^*. \quad (80)$$

Now, similar to the proof of Lemma 4, using (80), we can derive

$$\hat{\lambda}_{\tau_{j+1}}(\hat{\mathbf{z}}^j) \leq \hat{\lambda}_{\tau_j}(\hat{\mathbf{z}}^j) + \frac{\Delta_j}{t_0} \|\hat{\xi}_0\|_{\hat{\mathbf{z}}^j}^*. \quad (81)$$

By the same argument as the proof of (33), we can show that with  $\hat{\gamma}_k := \frac{\Delta_j}{t_0} \|\hat{\xi}_0\|_{\hat{\mathbf{z}}^j}^*$ , we have  $|\hat{\gamma}_k| \leq \frac{c\sqrt{\eta}}{1+c\sqrt{\eta}} - \eta$ . This shows that  $\Delta_j \leq \frac{t_0}{\|\hat{\xi}_0\|_{\hat{\mathbf{z}}^j}^*} \left( \frac{c\sqrt{\eta}}{1+c\sqrt{\eta}} - \eta \right)$ , which is the first estimate of (37). The second estimate of (37) can be derived as in Lemma 4 using  $\eta$  instead of  $\beta$ .

We prove (38). From (21) and (36), using the triangle inequality, we can upper bound

$$\begin{aligned} \lambda_{t_0}(\mathbf{z}^0) &:= \|\mathbf{z}^0 - \mathcal{P}_{\mathbf{z}^0}(\mathbf{z}^0 - \nabla^2 F(\mathbf{z}^0)^{-1} \nabla F(\mathbf{z}^0); t_0)\|_{\mathbf{z}^0} \\ &\stackrel{\mathbf{z}^0 = \hat{\mathbf{z}}^j}{=} \|\hat{\mathbf{z}}^j - \mathcal{P}_{\hat{\mathbf{z}}^j}(\hat{\mathbf{z}}^j - \nabla^2 F(\hat{\mathbf{z}}^j)^{-1} \nabla F(\hat{\mathbf{z}}^j); t_0)\|_{\hat{\mathbf{z}}^j} \\ &\leq \|\hat{\mathbf{z}}^j - \mathcal{P}_{\hat{\mathbf{z}}^j}(\hat{\mathbf{z}}^j - \nabla^2 F(\hat{\mathbf{z}}^j)^{-1} (\nabla F(\hat{\mathbf{z}}^j) - \tau_j t_0^{-1} \hat{\xi}^0); t_0)\|_{\hat{\mathbf{z}}^j} \\ &\quad + \|\mathcal{P}_{\hat{\mathbf{z}}^j}(\hat{\mathbf{z}}^j - \nabla^2 F(\hat{\mathbf{z}}^j)^{-1} (\nabla F(\hat{\mathbf{z}}^j) - \tau_j t_0^{-1} \hat{\xi}^0); t_0)\|_{\hat{\mathbf{z}}^j} \\ &\stackrel{(36), (67)}{\leq} \hat{\lambda}_{\tau_j}(\hat{\mathbf{z}}^j) + \|t_0^{-1} \tau_j \nabla^2 F(\hat{\mathbf{z}}^j)^{-1} \hat{\xi}^j\|_{\hat{\mathbf{z}}^j} \\ &= \hat{\lambda}_{\tau_j}(\hat{\mathbf{z}}^j) + \tau_j t_0^{-1} \|\hat{\xi}^0\|_{\hat{\mathbf{z}}^j}^*, \end{aligned}$$

which proves the first inequality of (38).

By [31, Corollary 4.2.1], we have  $\|\hat{\xi}^0\|_{\hat{\mathbf{z}}^j}^* \leq \kappa \|\hat{\xi}^0\|_{\bar{\mathbf{z}}_F^*}^*$ , where  $\bar{\mathbf{z}}_F^*$  and  $\kappa$  are given by (15) and below (15), respectively. Hence,  $\bar{\Delta}_\eta := \frac{\mu_\eta}{\kappa \|\hat{\xi}^0\|_{\bar{\mathbf{z}}_F^*}^*} \leq \bar{\Delta}_j$ . The second estimate of (38) follows from  $\tau_j := \tau - \sum_{l=0}^{j-1} \Delta_l \leq 1 - j \bar{\Delta}_\eta$  due to the update rule (35) with  $\Delta_j := \bar{\Delta}_j \geq \bar{\Delta}_\eta$ . In order to guarantee  $\lambda_{t_0}(\mathbf{z}^0) \leq \beta$ , it follows from (38) and the update rule of  $\tau_j$  that

$$j \geq \frac{1}{\bar{\Delta}_\eta} \left( 1 - \frac{(\beta - \eta)t_0}{\kappa \|\hat{\xi}^0\|_{\bar{\mathbf{z}}_F^*}^*} \right).$$

Finally, substituting  $\bar{\Delta}_\eta = \frac{t_0}{\kappa \|\hat{\xi}_0\|_{\bar{\mathbf{z}}_F^*}^*} \left( \frac{c\sqrt{\eta}}{1+c\sqrt{\eta}} - \eta \right)$  into this estimate and after simplifying the result, we obtain the remaining conclusion of Theorem 5.  $\square$

## 7.9 The proof of Theorem 7: primal recovery for (4) in Algorithm 2

By the definition of  $\varphi$ , we have  $\varphi(\mathbf{y}) := f^*(\mathbf{c} - L^*\mathbf{y}) = f^*(t^{-1}(\mathbf{c} - L^*\mathbf{y})) - v \ln(t)$  due to the self-concordant logarithmic homogeneity of  $f$ . Using the property of the Legendre transformation  $f^*$  of  $f$ , we can express this function as

$$\varphi(\mathbf{y}) = t^{-1} \max_{\mathbf{x} \in \text{int}(\mathcal{K})} \{ \langle \mathbf{c} - L^*\mathbf{y}, \mathbf{x} \rangle - t f(\mathbf{x}) \} - v \ln(t).$$

We show that the point  $\mathbf{x}^k$  given by (56) solves the above maximization problem. We can write down the optimality condition of the above maximization problem as

$$\mathbf{c} - L^* \mathbf{y}^{k+1} - t_{k+1} \nabla f(\mathbf{x}^{k+1}) = 0,$$

which leads to  $\nabla f(\mathbf{x}^{k+1}) = t_{k+1}^{-1}(\mathbf{c} - L^* \mathbf{y}^{k+1})$ . On the other hand, by the well-known property of  $f$  [31], we have  $\mathbf{x}^{k+1} = \nabla f^*(\nabla f(\mathbf{x}^{k+1})) = \nabla f^*\left(t_{k+1}^{-1}(\mathbf{c} - L^* \mathbf{y}^{k+1})\right) \in \text{int}(\mathcal{K})$ .

Now, we prove (57). Note that  $\mathbf{c} - L^* \mathbf{y}^{k+1} - t_{k+1} \nabla f(\mathbf{x}^{k+1}) = 0$  and  $\|\nabla f(\mathbf{x})\|_{\mathbf{x}}^* \leq \sqrt{\nu}$ , which leads to

$$\|L^* \mathbf{y}^{k+1} - \mathbf{c}\|_{\mathbf{x}^{k+1}}^* = t_{k+1} \|\nabla f(\mathbf{x}^{k+1})\|_{\mathbf{x}^{k+1}}^* \leq t_{k+1} \sqrt{\nu}.$$

Since  $t_{k+1} \leq \varepsilon$ , this estimate leads to the first inequality of (57).

From (24), there exists  $\mathbf{e}^k \in \mathbb{R}^p$  such that  $\mathbf{e}^k \in \nabla \varphi(\mathbf{y}^k) + \nabla^2 \varphi(\mathbf{y}^k)(\mathbf{y}^{k+1} - \mathbf{y}^k) + t_{k+1}^{-1} \partial \psi(\mathbf{y}^{k+1})$  and  $\|\mathbf{e}^k\|_{\mathbf{y}^k}^* \leq \delta_k$ . This condition leads to

$$\mathbf{e}^k + \nabla \varphi(\mathbf{y}^{k+1}) - \nabla \varphi(\mathbf{y}^k) - \nabla^2 \varphi(\mathbf{y}^k)(\mathbf{y}^{k+1} - \mathbf{y}^k) \in \nabla \varphi(\mathbf{y}^{k+1}) + t_{k+1}^{-1} \partial \psi(\mathbf{y}^{k+1}).$$

Therefore, we have

$$\begin{aligned} \text{dist}_{\mathbf{y}^{k+1}}(0, \nabla \varphi(\mathbf{y}^{k+1}) + t_{k+1}^{-1} \partial \psi(\mathbf{y}^{k+1})) &\leq \|\mathbf{e}^k + \nabla \varphi(\mathbf{y}^{k+1}) \\ &\quad - \nabla \varphi(\mathbf{y}^k) - \nabla^2 \varphi(\mathbf{y}^k)(\mathbf{y}^{k+1} - \mathbf{y}^k)\|_{\mathbf{y}^{k+1}}^* \\ &\leq \|\mathbf{e}^k\|_{\mathbf{y}^{k+1}}^* + \|\nabla \varphi(\mathbf{y}^{k+1}) - \nabla \varphi(\mathbf{y}^k) \\ &\quad - \nabla^2 \varphi(\mathbf{y}^k)(\mathbf{y}^{k+1} - \mathbf{y}^k)\|_{\mathbf{y}^{k+1}}^*. \end{aligned} \tag{82}$$

To estimate the right-hand side of this inequality, we define  $M_k := \|\nabla \varphi(\mathbf{y}^{k+1}) - \nabla \varphi(\mathbf{y}^k) - \nabla^2 \varphi(\mathbf{y}^k)(\mathbf{y}^{k+1} - \mathbf{y}^k)\|_{\mathbf{y}^{k+1}}^*$ . With the same proof as [31, Theorem 4.1.14], we can show that

$$M_k \leq \left(1 - \|\mathbf{y}^{k+1} - \mathbf{y}^k\|_{\mathbf{y}^k}\right)^{-2} \|\mathbf{y}^{k+1} - \mathbf{y}^k\|_{\mathbf{y}^k}^2 \leq \frac{(\delta(\mathbf{y}^k) + \lambda_{t_{k+1}}(\mathbf{y}^k))^2}{(1 - \lambda_{t_{k+1}}(\mathbf{y}^k) - \delta(\mathbf{y}^k))^2}. \tag{83}$$

Here, we use  $\|\mathbf{y}^{k+1} - \mathbf{y}^k\|_{\mathbf{y}^k} \leq \|\mathbf{y}^{k+1} - \bar{\mathbf{y}}^{k+1}\|_{\mathbf{y}^k} + \|\bar{\mathbf{y}}^{k+1} - \mathbf{y}^k\|_{\mathbf{y}^k} = \delta(\mathbf{y}^k) + \lambda_{t_{k+1}}(\mathbf{y}^k)$  by the definitions of  $\lambda_{t_{k+1}}(\mathbf{y})$  in (21) and of  $\delta(\mathbf{y})$  above (27). Substituting (83) into (82) we get

$$\text{dist}_{\mathbf{y}^{k+1}}(0, \nabla \varphi(\mathbf{y}^{k+1}) + t_{k+1}^{-1} \partial \psi(\mathbf{y}^{k+1})) \leq \|\mathbf{e}^k\|_{\mathbf{y}^{k+1}}^* + \frac{(\delta(\mathbf{y}^k) + \lambda_{t_{k+1}}(\mathbf{y}^k))^2}{(1 - \lambda_{t_{k+1}}(\mathbf{y}^k) - \delta(\mathbf{y}^k))^2}. \tag{84}$$

Next, it remains to estimate  $\|\mathbf{e}^k\|_{\mathbf{y}^{k+1}}^*$ . Indeed, we have

$$\begin{aligned}\|\mathbf{e}^k\|_{\mathbf{y}^{k+1}}^* &\leq \left(1 - \|\mathbf{y}^{k+1} - \mathbf{y}_t^k\|_{\mathbf{y}^k}\right)^{-1} \|\mathbf{e}^k\|_{\mathbf{y}^k} \leq \left(1 - \lambda_{t_{k+1}}(\mathbf{y}^k) - \delta(\mathbf{y}^k)\right)^{-1} \|\mathbf{e}^k\|_{\mathbf{y}^k} \\ &\leq \frac{\delta_k}{1 - \lambda_{t_{k+1}}(\mathbf{y}^k) - \delta_k}.\end{aligned}$$

Using this estimate into (84) and  $\lambda_{t_{k+1}}(\mathbf{y}^k) \leq c\sqrt{\beta}(1 + c\sqrt{\beta})^{-1}$  from Lemma 4, we obtain

$$\begin{aligned}\text{dist}_{\mathbf{y}^{k+1}}\left(0, \nabla\varphi(\mathbf{y}^{k+1}) + t_{k+1}^{-1}\partial\psi(\mathbf{y}^{k+1})\right) &\leq \frac{\delta_k(1 + c\sqrt{\beta})}{(1 - \delta_k(1 + c\sqrt{\beta}))} \\ &+ \frac{(\delta_k(1 + c\sqrt{\beta}) + c\sqrt{\beta})^2}{(1 - \delta_k(1 + c\sqrt{\beta}))^2}.\end{aligned}$$

Substituting an upper bound  $\delta_t := \frac{(1-c^2)\beta}{(1+c\sqrt{\beta})^3[3c\sqrt{\beta}+c^2\beta+(1+c\sqrt{\beta})^3]}$  of  $\delta_k$  from Lemma 4 into the last estimate and simplifying the result, we get

$$\text{dist}_{\mathbf{y}^{k+1}}\left(0, \nabla\varphi(\mathbf{y}^{k+1}) + t_{k+1}^{-1}\partial\psi(\mathbf{y}^{k+1})\right) \leq \theta(c, \beta), \quad (85)$$

where  $\theta(c, \beta)$  is defined as

$$\begin{aligned}\theta(c, \beta) &:= \frac{(1 - c^2)\beta}{(1 + c\sqrt{\beta})^2[3c\sqrt{\beta} + c^2\beta + (1 + c\sqrt{\beta})^3] - (1 - c^2)\beta} \\ &+ \left( \frac{(1 - c^2)\beta + c\sqrt{\beta}(1 + c\sqrt{\beta})^2[3c\sqrt{\beta} + c^2\beta + (1 + c\sqrt{\beta})^3]}{(1 + c\sqrt{\beta})^2[3c\sqrt{\beta} + c^2\beta + (1 + c\sqrt{\beta})^3] - (1 - c^2)\beta} \right)^2.\end{aligned} \quad (86)$$

Using the fact that  $c \in (0, 1)$  and  $0 \leq \beta < 0.5(1 + 2c^2 - \sqrt{1 + 4c^2})$ , we have  $\theta(c, \beta) \leq 1$ . Since  $\nabla\varphi(\cdot) = -L\nabla f^*(\mathbf{c} - L^*(\cdot)) = -t_{k+1}^{-1}L\nabla f^*(t_{k+1}^{-1}(\mathbf{c} - L^*(\cdot)))$  due to (48), using (56) we can show that  $\nabla\varphi(\mathbf{y}^{k+1}) = t_{k+1}^{-1}L\mathbf{x}^{k+1}$ . Plugging this expression into (85) and noting that  $\partial\psi(\cdot) = \partial g^*(\cdot) + \mathbf{b}$ , we obtain

$$\begin{aligned}\text{dist}_{\mathbf{y}^{k+1}}\left(L\mathbf{x}^{k+1} - \mathbf{b}, \partial g^*(\mathbf{y}^{k+1})\right) &= \text{dist}_{\mathbf{y}^{k+1}}\left(0, \mathbf{b} - L\mathbf{x}^{k+1} + \partial g^*(\mathbf{y}^{k+1})\right) \\ &\leq t_{k+1}\theta(c, \beta).\end{aligned}$$

Let  $\mathbf{s}^{k+1} = \pi_{\partial g^*(\mathbf{y}^{k+1})}(L\mathbf{x}^{k+1} - \mathbf{b})$  be the projection of  $L\mathbf{x}^{k+1} - \mathbf{b}$  onto  $\partial g^*(\mathbf{y}^{k+1})$ . Then,  $\mathbf{s}^{k+1} \in \partial g^*(\mathbf{y}^{k+1})$ , and hence,  $\mathbf{y}^{k+1} \in \partial g(\mathbf{s}^{k+1})$ , which shows the second term of (57). Using this relation in the last inequality and the definition of  $\mathbf{s}^{k+1}$ , we obtain  $\|L\mathbf{x}^{k+1} - \mathbf{b} - \mathbf{s}^{k+1}\|_{\mathbf{y}^{k+1}}^* \leq t_{k+1}\theta(c, \beta)$ , which is the third term of (57). Finally, since  $\theta(c, \beta) \leq 1$ , we have  $\max\{\sqrt{v}, \theta(c, \beta)\} = \sqrt{v}$ . Using (57), we can conclude that  $(\mathbf{x}^k, \mathbf{s}^k)$  is an  $\varepsilon$ -solution of (3) if  $\sqrt{v}t_k \leq \varepsilon$ .  $\square$

## References

1. Auslender, A., Teboulle, M., Ben-Tiba, S.: A logarithmic-quadratic proximal method for variational inequalities. *Comput. Optim. Appl.* **12**(1–3), 31–40 (1999)
2. Bauschke, H.H., Combettes, P.: Convex Analysis and Monotone Operators Theory in Hilbert Spaces, 2nd edn. Springer, Berlin (2017)
3. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**(1), 183–202 (2009)
4. Becker, S., Fadili, M.J.: A quasi-Newton proximal splitting method. In: Proceedings of Neural Information Processing Systems Foundation (NIPS) (2012)
5. Ben-Tal, A., Nemirovski, A.: Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications. Volume 3 of MPS/SIAM Series on Optimization. SIAM, Philadelphia (2001)
6. Bonnans, J.F.: Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Appl. Math. Optim.* **29**, 161–186 (1994)
7. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–122 (2011)
8. Boyd, S., Vandenberghe, L.: Convex Optimization. University Press, Cambridge (2004)
9. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**(1), 120–145 (2011)
10. Combettes, P., Pesquet, J.-C.: Signal recovery by proximal forward-backward splitting. In: Fixed-Point Algorithms for Inverse Problems in Science and Engineering, pp. 185–212. Springer, Berlin (2011)
11. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward–backward splitting. *Multiscale Model. Simul.* **4**, 1168–1200 (2005)
12. De Luca, T., Facchinei, F., Kanzow, C.: A semismooth equation approach to the solution of nonlinear complementarity problems. *Math. Program.* **75**(3), 407–439 (1996)
13. Dontchev, A.L., Rockafellar, R.T.: Implicit Functions and Solution Mappings: A View from Variational Analysis. Springer, Berlin (2014)
14. Eckstein, J., Bertsekas, D.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.* **55**, 293–318 (1992)
15. Esser, J.E.: Primal-dual algorithm for convex models and applications to image restoration, registration and nonlocal inpainting. Ph.D. Thesis, University of California, Los Angeles (2010)
16. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, vol. 1–2. Springer, Berlin (2003)
17. Frank, M., Wolfe, P.: An algorithm for quadratic programming. *Nav. Res. Logist. Q.* **3**, 95–110 (1956)
18. Friedlander, M., Goh, G.: Efficient evaluation of scaled proximal operators. *Electron. Trans. Numer. Anal.* **46**, 1–22 (2017)
19. Fukushima, M.: Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. *Math. Program.* **53**, 99–110 (1992)
20. Goldstein, T., Li, M., Yuan, X., Esser, E., Baraniuk, R.: Adaptive primal-dual hybrid gradient methods for saddle-point problems, pp. 1–15 (2015). [arXiv:1305.0546v2](https://arxiv.org/abs/1305.0546v2)
21. Grant, M., Boyd, S., Ye, Y.: Disciplined convex programming. In: Liberti, L., Maculan, N. (eds.) Global Optimization: From Theory to Implementation, Nonconvex Optimization and Its Applications, pp. 155–210. Springer, Berlin (2006)
22. Hajek, B., Wu, Y., Xu, J.: Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Trans. Inf. Theory* **62**, 2788–2797 (2016)
23. Jaggi, M.: Revisiting Frank–Wolfe: projection-free sparse convex optimization. In: JMLR W&CP, vol. 28, no. 1, pp. 427–435 (2013)
24. Johnson, R., Zhang, T.: Accelerating stochastic gradient descent using predictive variance reduction. In: Advances in Neural Information Processing Systems (NIPS), pp. 315–323 (2013)
25. Korpelevic, G.M.: An extragradient method for finding saddle-points and for other problems. *Èkon. Mat. Metody* **12**(4), 747–756 (1976)
26. Kummer, B.: Newton's method for non-differentiable functions. *Adv. Math. Optim.* **45**, 114–125 (1988)
27. Löfberg, J.: YALMIP : a toolbox for modeling and optimization in MATLAB. In: Proceedings of the CACSD Conference, Taipei, Taiwan (2004)
28. Monteiro, R.D.C., Svaiter, B.F.: Iteration-complexity of a Newton proximal extragradient method for monotone variational inequalities and inclusion problems. *SIAM J. Optim.* **22**(3), 914–935 (2012)

29. Nemirovski, A., Juditsky, A., Lan, G., Shapiro, A.: Robust stochastic approximation approach to stochastic programming. *SIAM J. Optim.* **19**(4), 1574–1609 (2009)
30. Nemirovskii, A.: Prox-method with rate of convergence  $\mathcal{O}(1/t)$  for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. Optim.* **15**(1), 229–251 (2004)
31. Nesterov, Y.: *Introductory Lectures on Convex Optimization: A Basic Course*. Volume 87 of Applied Optimization. Kluwer Academic Publishers, Dordrecht (2004)
32. Nesterov, Y.: Dual extrapolation and its applications to solving variational inequalities and related problems. *Math. Program.* **109**(2–3), 319–344 (2007)
33. Nesterov, Y.: Smoothing technique and its applications in semidefinite optimization. *Math. Program.* **110**(2), 245–259 (2007)
34. Nesterov, Y.: Gradient methods for minimizing composite objective function. *Math. Program.* **140**(1), 125–161 (2013)
35. Nesterov, Y., Nemirovski, A.: *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial Mathematics, Philadelphia (1994)
36. Nesterov, Y., Todd, M.J.: Self-scaled barriers and interior-point methods for convex programming. *Math. Oper. Res.* **22**(1), 1–42 (1997)
37. Nocedal, J., Wright, S.J.: *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering, 2nd edn. Springer, New York (2006)
38. Pang, J.-S.: A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems. *Math. Program.* **51**(1), 101–131 (1991)
39. Parikh, N., Boyd, S.: Proximal algorithms. *Found. Trends Optim.* **1**(3), 123–231 (2013)
40. Qi, L., Sun, J.: A nonsmooth version of Newton’s method. *Math. Program.* **58**, 353–367 (1993)
41. Ralph, D.: Global convergence of damped Newton’s method for nonsmooth equations via the path search. *Math. Oper. Res.* **19**(2), 352–389 (1994)
42. Robinson, S.M.: Strongly regular generalized equations. *Math. Oper. Res.* **5**(1), 43–62 (1980)
43. Robinson, S.M.: Newton’s method for a class of nonsmooth functions. *Set Valued Var. Anal.* **2**, 291–305 (1994)
44. Rockafellar, R.T.: *Convex Analysis*. Volume 28 of Princeton Mathematics Series. Princeton University Press, Princeton (1970)
45. Rockafellar, R.T., Wets, R. J.-B.: *Variational Analysis*. Springer, Berlin (1997)
46. Shefi, R., Teboulle, M.: Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. *SIAM J. Optim.* **24**(1), 269–297 (2014)
47. Solodov, M.V., Svaiter, B.F.: A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set Valued Var. Anal.* **7**(4), 323–345 (1999)
48. Sturm, F.: Using SeDuMi 1.02: A Matlab toolbox for optimization over symmetric cones. *Optim. Methods Softw.* **11–12**, 625–653 (1999)
49. Su, W., Boyd, S., Candes, E.: A differential equation for modeling Nesterov’s accelerated gradient method: theory and insights. In: *Advances in Neural Information Processing Systems (NIPS)*, pp. 2510–2518 (2014)
50. Toh, K.-C., Todd, M.J., Tütüncü, R.H.: On the implementation and usage of SDPT3—a Matlab software package for semidefinite-quadratic-linear programming. Technical Report 4, NUS Singapore (2010)
51. Tran-Dinh, Q., Kyriolidis, A., Cevher, V.: An inexact proximal path-following algorithm for constrained convex minimization. *SIAM J. Optim.* **24**(4), 1718–1745 (2014)
52. Tran-Dinh, Q., Kyriolidis, A., Cevher, V.: A single phase proximal path-following framework. *Math. Oper. Res.* (2018) (**accepted**)
53. Tran-Dinh, Q., Necula, I., Savorgnan, C., Diehl, M.: An inexact perturbed path-following method for Lagrangian decomposition in large-scale separable convex optimization. *SIAM J. Optim.* **23**(1), 95–125 (2013)
54. Tseng, P.: Applications of splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM J. Control Optim.* **29**, 119–138 (1991)
55. Tseng, P.: Alternating projection-proximal methods for convex programming and variational inequalities. *SIAM J. Optim.* **7**(4), 951–965 (1997)
56. Wen, Z., Goldfarb, D., Yin, W.: Alternating direction augmented Lagrangian methods for semidefinite programming. *Math. Program. Comput.* **2**, 203–230 (2010)

57. Wen, Z., Yin, W., Zhang, Y.: Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. *Math. Program. Comput.* **4**(4), 333–361 (2012)
58. Womersley, R.S., Sun, D., Qi, H.: A feasible semismooth asymptotically Newton method for mixed complementarity problems. *Math. Program.* **94**(1), 167–187 (2002)
59. Wright, S.J.: Applying new optimization algorithms to model predictive control. In: Kantor J.C., Garcia C.E., Carnahan B. (eds) Fifth International Conference on Chemical Process Control—CPCV, pp. 147–155. American Institute of Chemical Engineers (1996)
60. Xiu, N., Zhang, J.: Some recent advances in projection-type methods for variational inequalities. *J. Comput. Appl. Math.* **152**(1), 559–585 (2003)
61. Yamashita, H., Yabe, H., Harada, K.: A primal-dual interior point method for nonlinear semidefinite programming. *Math. Program.* **135**, 89–121 (2012)
62. Yang, L., Sun, D., Toh, K.-C.: SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints. *Math. Program. Comput.* **7**(3), 331–366 (2015)