



Hessian barrier algorithms for non-convex conic optimization

Pavel Dvurechensky¹ · Mathias Staudigl²

Received: 17 October 2022 / Accepted: 16 January 2024
© The Author(s) 2024

Abstract

A key problem in mathematical imaging, signal processing and computational statistics is the minimization of non-convex objective functions that may be non-differentiable at the relative boundary of the feasible set. This paper proposes a new family of first- and second-order interior-point methods for non-convex optimization problems with linear and conic constraints, combining logarithmically homogeneous barriers with quadratic and cubic regularization respectively. Our approach is based on a potential-reduction mechanism and, under the Lipschitz continuity of the corresponding derivative with respect to the local barrier-induced norm, attains a suitably defined class of approximate first- or second-order KKT points with worst-case iteration complexity $O(\varepsilon^{-2})$ (first-order) and $O(\varepsilon^{-3/2})$ (second-order), respectively. Based on these findings, we develop new path-following schemes attaining the same complexity, modulo adjusting constants. These complexity bounds are known to be optimal in the unconstrained case, and our work shows that they are upper bounds in the case with complicated constraints as well. To the best of our knowledge, this work is the first which achieves these worst-case complexity bounds under such weak conditions for general conic constrained non-convex optimization problems.

Keywords Non-convex optimization · Interior-point methods · Self-concordant barrier · Cubic regularization of Newton method · Conic constraints

Mathematics Subject Classification 90C26 · 90C30 · 90C60 · 68Q25

✉ Pavel Dvurechensky
pavel.dvurechensky@wias-berlin.de
Mathias Staudigl
mathias.staudigl@uni-mannheim.de

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

² Department of Business Informatics and Mathematics, University of Mannheim, 68131 Mannheim, Germany

1 Introduction

Non-convex optimization is an active area of research in optimization with one of the goals being to establish complexity guarantees for finding approximate stationary points, see the review [40] and the references therein. Two large groups of algorithms that allow one to achieve this goal are first-order methods [2, 20, 23, 32, 51] and second-order methods [18, 26, 27, 29–32, 37, 38, 42, 79]. Higher-order algorithms are also analyzed, e.g., in [17, 28, 33, 74]. One of the challenges in the development and analysis of algorithms for non-convex optimization is dealing with constraints. The class of optimization problems we consider in this paper is described as follows.

Let \mathbb{E} be a finite dimensional vector space with inner product $\langle \cdot, \cdot \rangle$ and Euclidean norm $\|\cdot\|$. We are concerned with solving constrained conic optimization problems of the form

$$\min_x f(x) \quad \text{s.t.: } \mathbf{A}x = b, \quad x \in \bar{K}. \quad (\text{Opt})$$

The main working assumption underlying our developments is as follows:

- Assumption 1**
1. $\bar{K} \subset \mathbb{E}$ is a pointed (i.e., $\bar{K} \cap (-\bar{K}) = \{0\}$) closed convex cone with nonempty interior K ;
 2. $\mathbf{A} : \mathbb{E} \rightarrow \mathbb{R}^m$ is a linear operator assigning each element $x \in \mathbb{E}$ to a vector in \mathbb{R}^m and having full rank, i.e., $\text{im}(\mathbf{A}) = \mathbb{R}^m$, $b \in \mathbb{R}^m$;
 3. The feasible set $\bar{X} = \bar{K} \cap L$, where $L = \{x \in \mathbb{E} | \mathbf{A}x = b\}$, is compact and has nonempty relative interior denoted by $X = K \cap L$;
 4. $f : \mathbb{E} \rightarrow \mathbb{R}$ is possibly non-convex, continuous on \bar{X} and continuously differentiable on X ;
 5. Problem (Opt) admits a global solution. We let $f_{\min}(X) = \min\{f(x) | x \in \bar{X}\}$.

Note that f is not assumed to be globally differentiable, but only over the relative interior of the feasible set. Problem (Opt) contains many important classes of optimization problems as special cases. We summarize the three most important ones below.

Example 1 (NLP with non-negativity constraints) For $\mathbb{E} = \mathbb{R}^n$ and $\bar{K} \equiv \bar{K}_{\text{NN}} = \mathbb{R}_+^n$ we recover non-linear programming problems with linear equality constraints and non-negativity constraints: $\bar{X} = \{x \in \mathbb{R}^n | \mathbf{A}x = b, \text{ and } x_i \geq 0 \text{ for all } i = 1, \dots, n\}$. \diamond

Example 2 (Optimization over the second-order cone) Consider $\mathbb{E} = \mathbb{R}^n$ and $\bar{K} \equiv \bar{K}_{\text{SOC}} = \{x = (x_0, \underline{x}) \in \mathbb{R} \times \mathbb{R}^{n-1} | x_0 \geq \|\underline{x}\|_2\}$, the second-order cone (SOC). In this case problem (Opt) becomes a non-linear second-order conic optimization problem. Such problems have a huge number of applications, including energy systems [75], network localization [84], among many others [3]. \diamond

Example 3 (Semi-definite programming) If $\mathbb{E} = \mathcal{S}^n$ is the space of real symmetric $n \times n$ matrices and $\bar{K} \equiv \bar{K}_{\text{SDP}} = \mathcal{S}_+^n$ is the cone of positive semi-definite matrices, we obtain a non-linear semi-definite programming problem. Endow this space with the

standard inner product $\langle a, b \rangle = \text{tr}(ab)$. In this case, the linear operator \mathbf{A} assigns a matrix $x \in \mathbb{S}^n$ to a vector $\mathbf{A}x = [\langle a_1, x \rangle, \dots, \langle a_m, x \rangle]^\top$. Such mathematical programs have received enormous attention due to the large number of applications in control theory, combinatorial optimization, and engineering [13, 41, 70]. \diamond

Example 4 (Exponential cone programming) Consider the exponential cone defined as

$$K_{\text{exp}} = \{x \in \mathbb{R}^3 \mid x_1 \geq x_2 e^{x_3/x_2}, x_2 > 0\}$$

with the closure $\bar{K}_{\text{exp}} = \text{cl}(K_{\text{exp}}) = K_{\text{exp}} \cup \{(x_1, 0, x_3) \mid x_1 \geq 0, x_3 \leq 0\}$. K_{exp} is an important convex cone that is implemented in standard numerical solution packages like YALMIP, MOSEK, and CVX, as it can be used to represent a lot of interesting convex sets arising in optimization; see [34] and the PhD thesis [35]. \diamond

1.1 Motivating applications

1.1.1 Inverse problems with non-convex regularization

An important instance of (Opt) is the composite optimization problem

$$\min_x \left\{ f(x) = \ell(x) + \lambda \sum_{i=1}^n \varphi(x_i^p) \right\} \quad \text{s.t.: } x \in \bar{K}_{\text{NN}}, \tag{1}$$

where $\ell : \mathbb{R}^n \rightarrow \mathbb{R}$ is a smooth data fidelity function, $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is a convex function, $p \in (0, 1)$, and $\lambda > 0$ is a regularization parameter. A common use of this problem formulation is the regularized empirical risk-minimization problem in high-dimensional statistics, or the variational regularization technique in inverse problems. Non-negativity constraints can be motivated by some prior knowledge about the observed object which needs to be reconstructed. For example, the true signal may represent an image with positive intensities of the pixels, or one can consider Poisson Inverse Problem as in [59]. Common specifications for the regularizing function are $\varphi(s) = s$, or $\varphi(s) = s^{2/p}$. In the first case, we obtain $\sum_{i=1}^n \varphi(x_i^p) = \sum_{i=1}^n x_i^p = \|x\|_p^p$ on K_{NN} , whereas in the second case, we get $\sum_{i=1}^n \varphi(x_i^p) = \sum_{i=1}^n x_i^2 = \|x\|^2$. Note that the first case yields the objective f which is non-convex and non-differentiable at the boundary of the feasible set. It has been reported in imaging sciences that the use of such non-convex and non-differentiable regularizer has advantages in the restoration of piecewise constant images. Bian and Chen [15] contains a nice survey of studies supporting this observation. Moreover, in variable selection, the L_p penalty function with $p \in (0, 1)$ owns the oracle property [44] in statistics, while L_1 (called the LASSO) does not; problem (1) with $p \in (0, 1)$ can be used for variable selection at the group and individual variable levels simultaneously, while the very same problem with $p = 1$ can only work for individual variable selection [66]. See [36, 50] for a complexity-theoretic analysis of this problem.

1.1.2 Low rank matrix recovery

On the space of symmetric matrices $\mathbb{E} = \mathbb{S}^n$, together with the feasible set $\bar{X} = \{x \in \mathbb{E} | Ax = b, x \in \bar{K}_{\text{SDP}}\}$ we can consider the composite model $f(x) = \ell(x) + r(x)$, with smooth loss function $\ell : \mathbb{E} \rightarrow \mathbb{R}$, and with regularizer $r : \mathbb{E} \rightarrow \mathbb{R} \cup \{+\infty\}$. In applications the regularizer is given in the form of a matrix function $r(x) = \sum_i \sigma_i(x)^p$ on $x \in K_{\text{SDP}}$, where $p \in (0, 1)$ and $\sigma_i(x)$ is the i -th singular value of the matrix x . The resulting optimization problem is a matrix version of the non-convex regularized problem (1). See [67, 86] for a wealth of optimization problems following this description.

1.2 Challenges and contribution

One of the challenges in approaching problem (Opt) algorithmically is to deal with the feasible set $L \cap K$. A projection-based approach faces the computational bottleneck to project onto the intersection of a cone with an affine subspace, which makes the majority of the existing first-order [2, 20, 23, 32, 51, 57, 77] and second-order [18, 26, 27, 31, 32, 37, 38, 42, 79] methods practically less attractive, as they either are designed for unconstrained problems or use proximal steps in the updates. When primal feasibility is not a major concern, augmented Lagrangian algorithms [5, 7, 19, 54] are an alternative, though they do not always come with complexity guarantees. These observations motivate us to focus on primal barrier-penalty methods that allow us to decompose the feasible set and treat \bar{K} and L separately. Barrier methods are classical and powerful for convex optimization in the form of interior-point methods [78]. In the non-convex optimization setting results are in a sense fragmentary, with many different algorithms existing for different particular instantiations of (Opt). In particular, the main focus of barrier methods for non-convex optimization has been on particular cases, such as non-negativity constraints [16, 22, 58, 82, 85, 87] and quadratic programming [47, 73, 87]. In this paper we develop a flexible and unifying algorithmic framework that is able to accommodate first- and second-order interior-point algorithms for (Opt) with potentially non-convex and non-smooth at the relative boundary objective functions, and general, possibly non-symmetric, conic constraints. To the best of our knowledge, our framework is the first one providing complexity results for first- and second-order algorithms to reach points satisfying, respectively, suitably defined approximate first- and second-order necessary optimality conditions, under such weak assumptions and for such a general setting.

1.2.1 Our approach

At the core of our approach is the assumption that the cone \bar{K} admits a *logarithmically homogeneous self-concordant barrier* (LHSCB) $h(x)$ ([78], cf. Definition 1), for which we can retrieve information about the function value $h(x)$, the gradient $\nabla h(x)$ and the Hessian $H(x) = \nabla^2 h(x)$ with relative ease. This is not a very restrictive assumption, since all standard conic restrictions in optimization (i.e., \bar{K}_{NN} , \bar{K}_{SOC} , \bar{K}_{SDP} and \bar{K}_{exp}) have this property. Using this barrier function, our algorithms are designed to reduce

the *potential function*

$$F_\mu(x) = f(x) + \mu h(x), \quad (2)$$

where $\mu > 0$ is a (typically) small penalty parameter. By definition, the domain of the potential function F_μ is the interior of the cone \bar{K} . Therefore, any algorithm designed to reduce the potential will automatically respect the conic constraints, and the satisfaction of the linear constraints L can be ensured by choosing search directions from the nullspace of the linear operator A . Our target is to find points satisfying suitably defined approximate necessary first- and second-order optimality conditions for problem (Opt) expressed in terms of ε -KKT and $(\varepsilon_1, \varepsilon_2)$ -2KKT points respectively (cf. Sect. 3 for a precise definition).¹

1.2.2 Finding points satisfying approximate necessary first-order conditions

To produce an ε -KKT point for the model problem (Opt), we construct a novel gradient-based method, which we call the *first-order adaptive Hessian barrier algorithm* (FAHBA, Algorithm 1). The main computational steps involved in FAHBA are the identification of a search direction and a step-size policy, guaranteeing feasibility and sufficient decrease in the potential function value. The algorithm starts from an approximate analytic center of the feasible set. To find a search direction, we employ a linear model for F_μ , regularized by the squared local norm induced by the Hessian of the barrier function h , which is then minimized over the tangent space of the affine subspace L . The step-size is adaptively chosen to ensure feasibility and sufficient decrease in the objective function value f . For a judiciously chosen value of μ , we prove that this gradient-based method enjoys the upper iteration complexity bound $O(\varepsilon^{-2})$ for reaching an ε -KKT point when a “descent Lemma” holds relative to the local norm induced by the Hessian of h (cf. Assumption 3 and Theorem 1 in Sect. 4). We then move on in proving that FAHBA can be embedded within a path-following scheme that iteratively reduces the value of μ . This renders our first-order interior-point method parameter-free and any-time convergent, with basically the same iteration complexity of $O(\varepsilon^{-2})$.

1.2.3 Finding points satisfying approximate necessary second-order conditions

To obtain $(\varepsilon_1, \varepsilon_2)$ -2KKT points, we construct a *second-order adaptive Hessian barrier algorithm* (SAHBA, Algorithm 3). As for FAHBA, the search direction subroutine minimizes a local model of the potential function F_μ over the tangent space of the affine subspace L . The minimized model is composed of the linear model for F_μ , augmented by second-order information on the objective function f and regularized by the cube of the local norm induced by the Hessian of h . The regularization parameter is chosen adaptively to allow for potentially larger steps in the areas of small curvature. For a judiciously chosen value of μ , we establish (cf. Theorem 3) the worst-case upper bound

¹ Following the existing literature [58, 61, 82], we use weak second-order necessary optimality conditions. We refer to Sect. 3.3 for the detailed discussion.

$O(\max\{\varepsilon_1^{-3/2}, \varepsilon_2^{-3/2}\})$ on the number of iterations for reaching an $(\varepsilon_1, \varepsilon_2)$ -2KKT point, under a weaker version of the assumption that the Hessian of f is Lipschitz relative to the local norm induced by the Hessian of h (cf. Assumption 4 in Sect. 5 for a precise definition). We then propose a path-following version of **SAHBA** that iteratively reduces the value of μ making the algorithm parameter-free and any-time convergent, with $O(\max\{\varepsilon_1^{-3/2}, \varepsilon_2^{-3/2}\})$ complexity.

1.3 Related work

To the best of our knowledge, **FAHBA** and **SAHBA** are the first interior-point algorithms that achieve such complexity bounds for the general non-convex problem template (**Opt**). Closest to our approach, but within the trust-region framework, are the works [58, 65, 82]. All these papers focus on the special case of non-negativity constraints with $\bar{K} = \bar{K}_{\text{NN}}$ (Example 1) and fix μ before the start of the algorithm based on the desired accuracy ε , which may require some hyperparameter tuning in practice and may not work if the desired accuracy is not yet known. Interestingly, for the special case $\bar{K} = \bar{K}_{\text{NN}}$, our algorithms provide stronger results under weaker assumptions, compared to the first- and second-order methods in [58], the second-order method in [65] specified to our setting of linear constraints, and the first-order implementation of the second-order method in [82]. We make this claim precise in Sects. 4.4 and 5.3.

1.3.1 First-order methods

In the unconstrained setting, when the gradient is Lipschitz continuous, the standard gradient descent achieves the lower iteration complexity bound $O(\varepsilon^{-2})$ to find a first-order ε -stationary point \hat{x} such that $\|\nabla f(\hat{x})\| \leq \varepsilon$ [24, 25, 76]. The original inspiration for the construction of our methods comes from the paper [22] on Hessian barrier algorithms, which in turn was strongly influenced by continuous-time techniques [4, 21]. We extend the first-order method of [22] to general conic constraints, and develop a unified complexity analysis, which goes far beyond the quadratic optimization case studied in detail in that reference.

1.3.2 Second-order methods

In unconstrained optimization with Lipschitz continuous Hessian, cubic-regularized Newton methods [55, 79] and second-order trust region algorithms [27, 37, 38] achieve the lower iteration complexity bound $O(\max\{\varepsilon_1^{-3/2}, \varepsilon_2^{-3/2}\})$ [24, 25] to find a second-order $(\varepsilon_1, \varepsilon_2)$ -stationary point, i.e., a point \hat{x} satisfying $\|\nabla f(\hat{x})\| \leq \varepsilon_1$ and $\lambda_{\min}(\nabla^2 f(\hat{x})) \geq -\sqrt{\varepsilon_2}$, where $\lambda_{\min}(\cdot)$ denotes the minimal eigenvalue of a matrix.² The existing literature on non-convex problems with non-linear constraints either considers only equality constraints [39], only inequality constraints [65], or both, but

² A number of works, e.g. [27, 82], consider an $(\varepsilon_1, \varepsilon_2)$ -stationary point defined as \hat{x} such that $\|\nabla f(\hat{x})\| \leq \varepsilon_1$ and $\lambda_{\min}(\nabla^2 f(\hat{x})) \geq -\varepsilon_2$ and the corresponding complexity $O(\max\{\varepsilon_1^{-3/2}, \varepsilon_2^{-3}\})$. Our definition and complexity bound are the same up to redefinition of ε_2 .

require projection [32]. Moreover, they do not consider general conic constraints as we do in this paper.

1.3.3 Approximate optimality conditions

Bian et al. [16] consider box-constrained minimization of the same objective as in (1) and propose a notion of ε -scaled KKT points. Their definition is tailored to the geometry of the optimization problem, mimicking the complementarity slackness condition of the classical KKT theorem for the non-negative orthant. In particular, their first-order condition consists of feasibility of x along with a scaled gradient condition. Haeser et al. and O’Neill–Wright [58, 82] convincingly argue that, without additional assumptions on the objective function, points that satisfy the scaled gradient condition may not approach KKT points as ε decreases. Thus, [58, 82], provide alternative notions of approximate first- and second-order KKT conditions for the domain K_{NN} . Inspired by [58], we define the corresponding notions for general cones. Our first-order conditions turn out to be stronger than those of [58, 82] and the second-order condition is equivalent to theirs in the particular case of non-negativity constraints (cf. Sects. 3.3, 4.4, and 5.3). The proof that our algorithms are guaranteed to find such approximate KKT points requires some fine analysis exploiting the structural properties of logarithmically homogeneous barriers associated to the cone K , and are not simple extensions of arguments used for K_{NN} .

We remark that after the first release of our preprint on arXiv on October 29, 2021 (<https://arxiv.org/abs/2111.00100v1>), the paper [61] appeared in July 2022, of which we became aware during the revision of our paper after its submission. They also consider the model problem (Opt) and also build their algorithmic developments on a barrier construction. They propose a Newton-CG based method for finding $(\varepsilon, \varepsilon)$ -second-order approximate KKT points with similar to ours iteration complexity guarantee $O(\varepsilon^{-3/2})$, yet the dependence of their complexity bound on the barrier parameter is not clear. Their algorithm also runs for a fixed parameter μ , and thus is not parameter-free, unlike our parameter-free versions. Also, unlike them, we propose a first-order algorithm.

1.4 Notation

In what follows, \mathbb{E} denotes a finite-dimensional real vector space, and \mathbb{E}^* the dual space, which is formed by all linear functions on \mathbb{E} . The value of $s \in \mathbb{E}^*$ at $x \in \mathbb{E}$ is denoted by $\langle s, x \rangle$. In the particular case where $\mathbb{E} = \mathbb{R}^n$, we have $\mathbb{E} = \mathbb{E}^*$. The gradient vector of a differentiable function $f : \mathbb{E} \rightarrow \mathbb{R}$ is denoted as $\nabla f(x) \in \mathbb{E}^*$. For an operator $\mathbf{H} : \mathbb{E} \rightarrow \mathbb{E}^*$, denote by \mathbf{H}^* is adjoint operator, defined by the identity

$$(\forall u, v \in \mathbb{E}) : \quad \langle \mathbf{H}u, v \rangle = \langle u, \mathbf{H}^*v \rangle.$$

Thus, $\mathbf{H}^* : \mathbb{E} \rightarrow \mathbb{E}^*$. It is called self-adjoint if $\mathbf{H} = \mathbf{H}^*$. We use $\lambda_{\max}(\mathbf{H})$ ($\lambda_{\min}(\mathbf{H})$), to denote the maximum (minimum) eigenvalue of such operators. Important examples of

such self-adjoint operators are Hessians of twice differentiable functions $f : \mathbb{E} \rightarrow \mathbb{R}$:

$$\langle \nabla^2 f(x)u, v \rangle = \langle u, \nabla^2 f(x)v \rangle \quad (\forall u, v \in \mathbb{E})$$

Operator $\mathbf{H} : \mathbb{E} \rightarrow \mathbb{E}^*$ is positive semi-definite if $\langle \mathbf{H}u, u \rangle \geq 0$ for all $u \in \mathbb{E}$. If the inequality is always strict for non-zero u , then \mathbf{H} is called positive definite. These attributes are denoted as $\mathbf{H} \succeq 0$ and $\mathbf{H} \succ 0$, respectively. By fixing a positive definite self-adjoint operator $\mathbf{H} : \mathbb{E} \rightarrow \mathbb{E}^*$, we can define the following Euclidean norms

$$\|u\| = \langle \mathbf{H}u, u \rangle^{1/2}, \quad \|s\|^* = \langle s, \mathbf{H}^{-1}s \rangle^{1/2} \quad u \in \mathbb{E}, s \in \mathbb{E}^*.$$

In some cases we use notation $\|u\|_{\mathbf{H}}$ to explicitly indicate the operator used to define the norm. If $\mathbb{E} = \mathbb{R}^n$, then \mathbf{H} is usually taken as the identity matrix $\mathbf{H} = \mathbf{I}$. The L_∞ -norm of $x \in \mathbb{R}^n$ is denoted as $\|x\|_\infty = \max_{i=1, \dots, n} |x_i|$. The directional derivative of function $f : \mathbb{E} \rightarrow \mathbb{R}$ is defined in the usual way:

$$Df(x)[v] = \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} [f(x + \varepsilon v) - f(x)].$$

More generally, for $v_1, \dots, v_p \in \mathbb{E}$, we define $D^p f(x)[v_1, \dots, v_p]$ the p -th directional derivative at x along directions $v_i \in \mathbb{E}$. In that way we define $\nabla f(x) \in \mathbb{E}^*$ by $Df(x)[u] = \langle \nabla f(x), u \rangle$ and the Hessian $\nabla^2 f(x) : \mathbb{E} \rightarrow \mathbb{E}^*$ by $\langle \nabla^2 f(x)u, v \rangle = D^2 f(x)[u, v]$. We denote by $L_0 \triangleq \{v \in \mathbb{E} | \mathbf{A}v = 0\} \triangleq \ker(\mathbf{A})$ the tangent space associated with the affine subspace $L \subset \mathbb{E}$.

2 Preliminaries

2.1 Cones and their self-concordant barriers

Let $\bar{K} \subset \mathbb{E}$ be a *regular cone*: \bar{K} is closed convex with nonempty interior $K = \text{int}(\bar{K})$, and pointed (i.e. $\bar{K} \cap (-\bar{K}) = \{0\}$). Any such cone admits a self-concordant logarithmically homogeneous barrier $h(x)$ with finite parameter value θ [78].

Definition 1 A function $h : \bar{K} \rightarrow (-\infty, \infty]$ with $\text{dom } h = K$ is called a *θ -logarithmically homogeneous self-concordant barrier* (θ -LHSCB) for the cone \bar{K} if:

(a) h is a θ -self-concordant barrier for \bar{K} , i.e., for all $x \in K$ and $u \in \mathbb{E}$

$$\begin{aligned} |D^3 h(x)[u, u, u]| &\leq 2D^2 h(x)[u, u]^{3/2}, \text{ and} \\ \sup_{u \in \mathbb{E}} |2Dh(x)[u] - D^2 h(x)[u, u]| &\leq \theta. \end{aligned}$$

(b) h is logarithmically homogeneous:

$$h(tx) = h(x) - \theta \ln(t) \quad \forall x \in K, t > 0.$$

We denote the set of θ -logarithmically homogeneous barriers by $\mathcal{H}_\theta(K)$.

The constant θ is called the parameter of the barrier function. Just like in standard interior point methods, it affects the iteration complexity of our methods. Given $h \in \mathcal{H}_\theta(K)$, from [76, Thm 5.1.3] we know that for any $\bar{x} \in \text{bd}(\bar{K})$, any sequence $(x_k)_{k \geq 0}$ with $x_k \in K$ and $\lim_{k \rightarrow \infty} x_k = \bar{x}$ satisfies $\lim_{k \rightarrow \infty} h(x_k) = +\infty$. For a pointed cone \bar{K} , we have $\theta \geq 1$ and the Hessian $H(x) \triangleq \nabla^2 h(x) : \mathbb{E} \rightarrow \mathbb{E}^*$ is a positive definite linear operator defined by $\langle H(x)u, v \rangle \triangleq D^2 h(x)[u, v]$ for all $u, v \in \mathbb{E}$, see [76, Thm. 5.1.6]. The Hessian gives rise to a (primal) local norm

$$(\forall x \in K)(\forall u \in \mathbb{E}) : \|u\|_x \triangleq \langle H(x)u, u \rangle^{1/2}.$$

We also define a dual local norm on \mathbb{E}^* as

$$(\forall x \in K)(\forall s \in \mathbb{E}^*) : \|s\|_x^* \triangleq \langle [H(x)]^{-1}s, s \rangle^{1/2}.$$

The *Dikin ellipsoid* with center $x \in K$ and radius $r > 0$ is defined as the open set $\mathcal{W}(x; r) \triangleq \{u \in \mathbb{E} \mid \|u - x\|_x < r\}$. The usage of the local norm adapts the unit ball to the local geometry of the set K . Indeed, the following classical results are key to the development of our methods, since they allow us to ensure feasibility of the iterates and sufficient decrease of the potential function in each iteration of our algorithms.

Lemma 1 (Theorem 5.1.5 [76]) *For all $x \in K$ we have $\mathcal{W}(x; 1) \subseteq K$.*

Proposition 1 (Theorem 5.1.9 [76]) *Let $h \in \mathcal{H}_\theta(K)$, $x \in \text{dom } h$, and a fixed direction $d \in \mathbb{E}$. For all $t \in [0, \frac{1}{\|d\|_x})$, with the convention that $\frac{1}{\|d\|_x} = +\infty$ if $\|d\|_x = 0$, we have:*

$$h(x + td) \leq h(x) + t \langle \nabla h(x), d \rangle + t^2 \|d\|_x^2 \omega(t \|d\|_x),$$

where $\omega(t) = \frac{-t - \ln(1-t)}{t^2}$.

We will also use the following inequality for the function $\omega(t)$ [76, Lemma 5.1.5]:

$$\omega(t) \leq \frac{1}{2(1-t)}, \quad t \in [0, 1). \tag{3}$$

Appendix A contains some more technical properties of LHSCBs which are used in the proofs. We close this section with important examples of conic domains to which our methods can be directly applied.

Example 5 (Non-negativity constraints) For $\mathbb{E} = \mathbb{R}^n$ and $\bar{K} = \bar{K}_{\text{NN}}$, we define the log-barrier $h(x) = -\sum_{i=1}^n \ln(x_i)$ for all $x \in K_{\text{NN}} = \mathbb{R}_{++}^n$. It is readily seen that $h \in \mathcal{H}_n(K)$. ◇

Example 6 (SOC constraints) Let $\mathbb{E} = \mathbb{R}^n$ and $\bar{K} = \bar{K}_{\text{SOC}}$, defined in Example 2. For $x = (x_0, \underline{x}) \in \bar{K}_{\text{SOC}}$, we define the barrier $h(x) = -\ln(x_0^2 - \underline{x}^\top \underline{x})$. It is well known that $h \in \mathcal{H}_2(K_{\text{SOC}})$ [78]. ◇

Example 7 (SDP constraints) Let $\mathbb{E} = \mathbb{S}^n$ and $\bar{K} = \bar{K}_{\text{SDP}}$, defined in Example 3. Consider the barrier $h(x) = -\ln \det(x)$. It is well known that $h \in \mathcal{H}_n(K_{\text{SDP}})$. \diamond

Example 8 (The exponential cone) Consider the exponential cone K_{exp} with closure \bar{K}_{exp} introduced in Example 4. This set admits a 3-LHSB

$$h(x_1, x_2, x_3) = -\ln(x_2 \ln(x_1/x_2) - x_3) - \ln(x_1) - \ln(x_2) \in \mathcal{H}_3(K_{\text{exp}}).$$

We remark that this cone is not self-dual (cf. Definition 2), but

$$G\bar{K}_{\text{exp}} = \bar{K}_{\text{exp}}^* = \text{cl} \left(\{y \in \mathbb{R}^3 \mid y_1 \geq -y_3 e^{y_2/y_3-1}, y_1 > 0, y_3 < 0\} \right),$$

under the linear transformation $G = \begin{bmatrix} 1/e & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix}$. There are many convex sets that can be represented using the exponential cone. We list some examples below, but refer to the PhD thesis [35] for further details.

- Exponential: $\{(t, u) \mid t \geq e^u\} \iff (t, 1, u) \in \bar{K}_{\text{exp}}$;
- Logarithm: $\{(t, u) \mid t \leq \ln(u)\} \iff (u, 1, t) \in \bar{K}_{\text{exp}}$;
- Entropy: $t \leq -u \ln(u) \iff t \leq u \ln(1/u) \iff (1, u, t) \in \bar{K}_{\text{exp}}$;
- Relative Entropy: $t \geq u \log(u/w) \iff (w, u, t) \in \bar{K}_{\text{exp}}$;
- Softplus function: $t \geq \ln(1 + e^u) \iff a + b \leq 1, (a, 1, u - t) \in \bar{K}_{\text{exp}}, (b, 1, -t) \in \bar{K}_{\text{exp}}$.

\diamond

2.2 Exploiting the structure of symmetric cones

Nesterov and Todd [80] introduced *self-scaled barriers*, which later were realized as LHSCBs for *symmetric cones*. Such barriers are nowadays key to define primal–dual interior point methods with potentially larger step-sizes for convex problems. Our methods can also exploit the structure of self-scaled barriers, leading to potentially larger step-sizes and eventual faster convergence in practice. For a given closed convex nonempty cone \bar{K} , its *dual cone* is the closed convex and nonempty cone $\bar{K}^* \triangleq \{s \in \mathbb{E}^* \mid \langle s, x \rangle \geq 0 \ \forall x \in \bar{K}\}$. If $h \in \mathcal{H}_\theta(K)$, then the *dual barrier* is defined as $h_*(s) \triangleq \sup_{x \in K} \{\langle -s, x \rangle - h(x)\}$ for $s \in \bar{K}^*$. Note that $h_* \in \mathcal{H}_\theta(K^*)$ [78, Theorem 2.4.4].

Definition 2 An open convex cone is *self-dual* if $K^* = K$. K is *homogeneous* if for all $x, y \in K$ there exists a linear bijection $G : \mathbb{E} \rightarrow \mathbb{E}$ such that $Gx = y$ and $GK = K$. An open convex cone K is called *symmetric* if it is self-dual and homogeneous.

The class of symmetric cones can be characterized within the language of *Euclidean Jordan algebras* [45, 46, 63, 64]. For optimization, the three symmetric cones of most relevance are $\bar{K}_{\text{NN}}, \bar{K}_{\text{SOC}},$ and \bar{K}_{SDP} .

Definition 3 ([80]) $h \in \mathcal{H}_\theta(K)$ is a θ -*self-scaled barrier* (θ -SSB) if for all $x, w \in K$ we have $H(w)x \in K$ and $h_*(H(w)x) = h(x) - 2h(w) - \theta$. Let $\mathcal{B}_\theta(K)$ denote the class of θ -SSBs.

Clearly, $\mathcal{B}_\theta(K) \subset \mathcal{H}_\theta(K)$. Hauser and Güler [60] showed that every symmetric cone admits a θ -SSB for some $\theta \geq 1$, while a characterization of the barrier parameter θ has been obtained in [56]. The main advantage of working with SSBs instead of LHSCBs is that we can make potentially longer steps in the interior of the cone K towards the direction of its boundary, and have larger decrease of the potential. Let $x \in K$ and $d \in \mathbb{E}$. Denote

$$\sigma_x(d) \triangleq (\sup\{t : x - td \in K\})^{-1}.$$

Since $\mathcal{W}(x; 1) \subseteq K$ for all $x \in K$, we have that $\sigma_x(d) \leq \|d\|_x$ and $\sigma_x(-d) \leq \|d\|_x$ for all $d \in \mathbb{E}$. Therefore $[0, \frac{1}{\|d\|_x}] \subseteq [0, \frac{1}{\sigma_x(d)}]$. Hence, if the scalar quantity $\sigma_x(d)$ can be computed efficiently, it would allow us to make a larger step without violating feasibility.

Example 9 For $\bar{K} = \bar{K}_{NN}$, to guarantee $x - td \in K_{NN}$, we need $x_i - td_i > 0$ for all $i \in \{1, \dots, n\}$. Hence, if $d_i \leq 0$, this is satisfied for all $t \geq 0$. If $d_i > 0$, we obtain the restriction $t \leq \frac{x_i}{d_i}$. Therefore, $\sigma_x(-d) = \max\{\frac{d_i}{x_i} : d_i > 0\}$. \diamond

Example 10 For $\bar{K} = \bar{K}_{SDP}$, we see that $x - td > 0$ if and only if $\text{Id} > tx^{-1/2}dx^{-1/2}$, where Id is the identity matrix and $x^{1/2}$ denotes the square root of a matrix $x \in K_{SDP}$. Hence, if $\lambda_{\max}(x^{-1/2}dx^{-1/2}) > 0$, then $t < \frac{1}{\lambda_{\max}(x^{-1/2}dx^{-1/2})}$. Thus, $\sigma_x(d) = \max\{\lambda_{\max}(x^{-1/2}dx^{-1/2}), 0\}$. \diamond

The analogous result to Proposition 1 for barriers $h \in \mathcal{B}_\theta(K)$ reads as follows:

Proposition 2 (Theorem 4.2 [80]) *Let $h \in \mathcal{B}_\theta(K)$ and $x \in K$. Let $d \in \mathbb{E}$ be such that $\sigma_x(-d) > 0$. Then, for all $t \in [0, \frac{1}{\sigma_x(-d)})$, we have:*

$$h(x + td) \leq h(x) + t \langle \nabla h(x), d \rangle + t^2 \|d\|_x^2 \omega(t\sigma_x(-d)).$$

2.3 Unified notation

Our algorithms work on any conic domain on which we can efficiently evaluate a θ -LHSCB.

Assumption 2 \bar{K} is a regular cone admitting an efficient barrier setup $h \in \mathcal{H}_\theta(K)$. By this we mean that at a given query point $x \in K$, we can construct an oracle that returns to us information about the values $h(x)$, $\nabla h(x)$ and $H(x) = \nabla^2 h(x)$, with low computational efforts.

Note that when $h \in \mathcal{B}_\nu(K)$, we have the flexibility to treat h either as $h \in \mathcal{H}_\nu(K)$ or as $h \in \mathcal{B}_\nu(K)$. Given the potential advantages when working on symmetric cones, it is useful to develop a unified notation handling the case $h \in \mathcal{H}_\theta(K)$ and $h \in \mathcal{B}_\theta(K)$ at the same time. We therefore introduce the notation

$$(\forall(x, d) \in X \times \mathbb{E}) : \zeta(x, d) \triangleq \begin{cases} \|d\|_x & \text{if } h \in \mathcal{H}_\theta(K) \setminus \mathcal{B}_\theta(K), \\ \sigma_x(-d) & \text{if } h \in \mathcal{B}_\theta(K). \end{cases} \tag{4}$$

Note that

$$(\forall(x, d) \in X \times \mathbb{E}) : \zeta(x, d) \leq \|d\|_x, \tag{5}$$

$$(\forall(x, d) \in X \times \mathbb{E}) \left(\forall t \in \left[0, \frac{1}{\zeta(x, d)} \right) \right) : x + td \in K. \tag{6}$$

Propositions 1 and 2 together with Eq. (4), give us the one-and-for-all *Bregman bound*

$$h(x + td) - h(x) - \langle \nabla h(x), td \rangle \leq t^2 \|d\|_x^2 \omega(t\zeta(x, d)), \tag{7}$$

valid for all $(x, d) \in X \times \mathbb{E}$ and $t \in \left[0, \frac{1}{\zeta(x, d)} \right)$.

3 Approximate optimality conditions

Consider the non-convex optimization problem (Opt). If x^* is a local solution of the optimization problem at which the objective function f is continuously differentiable, then there exists $y^* \in \mathbb{R}^m$ such that

$$\nabla f(x^*) - \mathbf{A}^* y^* \in -\text{NC}_{\bar{K}}(x^*). \tag{8}$$

This is equivalent to the standard local optimality condition [81]

$$\langle \nabla f(x^*), x - x^* \rangle \geq 0 \quad \forall x \in \bar{X} = \bar{K} \cap L. \tag{9}$$

Remark 1 Since $-\text{NC}_{\bar{K}}(x^*) \subseteq \bar{K}^*$, the inclusion (8) implies $s^* \triangleq \nabla f(x^*) - \mathbf{A}^T y^* \in \bar{K}^*$. ◇

3.1 First-order approximate KKT conditions

The next definition specifies our notion of an approximate first-order KKT point for problem (Opt).

Definition 4 Given $\varepsilon \geq 0$, a point $\bar{x} \in \mathbb{E}$ is an ε -KKT point for problem (Opt) if there exists $\bar{y} \in \mathbb{R}^m$ such that

$$\mathbf{A}\bar{x} = b, \bar{x} \in K, \tag{10}$$

$$\bar{s} = \nabla f(\bar{x}) - \mathbf{A}^* \bar{y} \in \bar{K}^*, \tag{11}$$

$$\langle \bar{s}, \bar{x} \rangle \leq \varepsilon. \tag{12}$$

Note that by conditions (10) and (11) $\bar{x} \in K$ and $\bar{s} \in \bar{K}^*$, i.e., are feasible and $\langle \bar{s}, \bar{x} \rangle \geq 0$. Thus, condition (12) is reasonable since it recovers the standard complementary slackness when $\varepsilon \rightarrow 0$. Although our definition explicitly requires *strict*

primal feasibility, similar approximate KKT conditions have been introduced for candidates \bar{x} which may be infeasible. This gains relevance in primal–dual settings and sequential optimality conditions for Lagrangian-based methods. In such setups a projection should be used [7, 9].

Definition 4 can be motivated as an approximate version of the exact first-order KKT condition stated as Theorem 5 in Appendix B. It is also easy to show that the above definition readily implies the standard approximate first-order stationarity condition $\langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq -\varepsilon$, for all $x \in \bar{X}$ (cf. (9)). Indeed, let $x \in \bar{X}$ be arbitrary and \bar{x} satisfy Definition 4. Then,

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle = \langle \bar{s}, x - \bar{x} \rangle + \langle \bar{y}, \mathbf{A}(x - \bar{x}) \rangle = \langle \bar{s}, x \rangle - \langle \bar{s}, \bar{x} \rangle \geq -\varepsilon,$$

where we used that $\mathbf{A}x = \mathbf{A}\bar{x} = b$, $\langle \bar{s}, x \rangle \geq 0$ since $x \in \bar{K}$ and $\bar{s} \in \bar{K}^*$, and $\langle \bar{s}, \bar{x} \rangle \leq \varepsilon$.

The next result uses a perturbation argument based on an interior penalty approach, inspired by [58], and shows that our definition of an ε -KKT point is attainable and can be read as an *approximate KKT condition*.

Proposition 3 *Let $x^* \in \bar{X}$ be a local solution of problem (Opt) where f is continuously differentiable on X . Then, there exist sequences $(x_k)_{k \geq 1} \subset X$, $(y_k)_{k \geq 1} \subset \mathbb{R}^m$ and $s_k = \nabla f(x_k) - \mathbf{A}^*y_k$, $k \geq 1$ satisfying the following:*

- (i) $x_k \rightarrow x^*$.
- (ii) For all $\varepsilon > 0$ we have $|\langle s_k, x_k \rangle| \leq \varepsilon$ for all k sufficiently large.
- (iii) All accumulation points of the sequence $(s_k)_{k \geq 1}$ are contained in $-\text{NC}_{\bar{K}}(x^*) \subseteq \bar{K}^*$.

Proof (i) Consider the following perturbed version of problem (Opt), for which x^* is the unique global solution if we take $\delta > 0$ sufficiently small,

$$\min_x f(x) + \frac{1}{2}\|x - x^*\|^2 \quad \text{s.t.: } \mathbf{A}x = b, x \in \bar{K}, \|x - x^*\|^2 \leq \delta. \quad (13)$$

Further, consider the penalty function

$$\varphi_k(x) \triangleq f(x) + \mu_k h(x) + \frac{1}{2}\|x - x^*\|^2$$

for a sequence $\mu_k \rightarrow 0^+$ as $k \rightarrow \infty$ and the optimization problem

$$\min_x \varphi_k(x) \quad \text{s.t.: } \mathbf{A}x = b, x \in K, \|x - x^*\|^2 \leq \delta.$$

It is well known [48] that a global solution x_k exists for this problem for all $k \geq 1$, and that cluster points of the sequence $(x_k)_{k \geq 1}$ are global solutions of (13). Since $\|x_k - x^*\|^2 \leq \delta$, the sequence x_k is bounded and, thus, $x_k \rightarrow x^*$ as $k \rightarrow \infty$.

- (ii) For k large enough, x_k is a local solution of

$$\min_x \varphi_k(x) \quad \text{s.t.: } \mathbf{A}x = b.$$

The optimality conditions for this problem read as

$$\nabla f(x_k) + (x_k - x^*) + \mu_k \nabla h(x_k) - \mathbf{A}^* y_k = 0,$$

where $y_k \in \mathbb{R}^m$ is a Lagrange multiplier. Setting $s_k \triangleq \nabla f(x_k) - \mathbf{A}^* y_k$, we can continue with

$$|\langle s_k, x_k \rangle| = |\langle x^* - x_k, x_k \rangle - \mu_k \langle \nabla h(x_k), x_k \rangle| \stackrel{(81)}{=} |\langle x^* - x_k, x_k \rangle + \mu_k \theta|.$$

As $x_k \rightarrow x^*$ and $\mu_k \rightarrow 0$, we conclude that $\lim_{k \rightarrow \infty} |\langle s_k, x_k \rangle| = 0$.

(iii) Let $x \in \bar{K}$ be arbitrary. Using part (ii) and the Cauchy–Schwarz inequality, we see that

$$\begin{aligned} \langle s_k, x \rangle &= \langle s_k, x - x_k \rangle + \langle s_k, x_k \rangle \\ &= \langle x^* - x_k, x - x_k \rangle + \langle s_k, x_k \rangle - \mu_k \langle \nabla h(x_k), x - x_k \rangle \\ &\stackrel{(83)}{\geq} -\|x^* - x_k\| \cdot \|x - x_k\| + \langle s_k, x_k \rangle - \mu_k \theta \rightarrow 0. \end{aligned}$$

Hence, $\liminf_{k \rightarrow \infty} \langle s_k, x \rangle \geq 0$ for all $x \in \bar{K}$, which proves that accumulation points of $(s_k)_{k \geq 1}$ are contained in the dual cone \bar{K}^* .

Let $x_k \rightarrow x^*$. Assume first that $x^* \in X = K \cap L$. Then the sequence $(s_k)_{k \geq 1}$ constructed as in part (ii) satisfies $s_k = -\mu_k \nabla h(x_k) + (x^* - x_k)$ for all $k \geq 1$. Consequently, $s_k \rightarrow 0$ as $k \rightarrow \infty$. Now assume that $x^* \in \text{bd}(X)$. Choosing $\mu_k \triangleq \frac{1}{\|\nabla h(x_k)\|} \rightarrow 0$ as $k \rightarrow \infty$, gives $s_k = x^* - x_k - \frac{1}{\|\nabla h(x_k)\|} \nabla h(x_k)$. Lemma 4 in Appendix A shows that the sequence $\left(\frac{\nabla h(x_k)}{\|\nabla h(x_k)\|}\right)_{k \geq 1}$ is bounded with all its accumulation points contained in $\text{NC}_{\bar{K}}(x^*)$. This readily implies that accumulation points of the sequence $(s_k)_{k \geq 1}$ must be contained in $-\text{NC}_{\bar{K}}(x^*)$. □

3.2 Second-order approximate KKT conditions

Our definition of an approximate second-order KKT point is motivated by the exact definition in Theorem 6, stated in Appendix B. The natural inexact version of this definition reads as follows.

Definition 5 Given $\varepsilon_1, \varepsilon_2 \geq 0$, a point $\bar{x} \in \mathbb{E}$ is an $(\varepsilon_1, \varepsilon_2)$ -2KKT point for problem (Opt) if there exists $\bar{y} \in \mathbb{R}^m$ such that for $\bar{s} = \nabla f(\bar{x}) - \mathbf{A}^* \bar{y}$ conditions (10) and (11) hold, as well as

$$\langle \bar{s}, \bar{x} \rangle \leq \varepsilon_1, \tag{14}$$

$$\nabla^2 f(\bar{x}) + \sqrt{\varepsilon_2} H(\bar{x}) \succeq 0 \text{ on } L_0. \tag{15}$$

Given $x \in X$, define the set of *feasible directions* as $\mathcal{F}_x \triangleq \{v \in \mathbb{E} \mid x+v \in X\}$. Lemma 1 implies that

$$\mathcal{T}_x \triangleq \{v \in \mathbb{E} \mid \mathbf{A}v = 0, \|v\|_x < 1\} \subseteq \mathcal{F}_x. \tag{16}$$

Upon defining $d = [H(x)]^{1/2}v$ for $v \in \mathcal{T}_x$, we obtain a direction d satisfying $\mathbf{A}[H(x)]^{-1/2}d = 0$ and $\|d\| = \|v\|_x$. Hence, for $x \in K$, we can equivalently characterize the set \mathcal{T}_x as $\mathcal{T}_x = \{[H(x)]^{-1/2}d \mid \mathbf{A}[H(x)]^{-1/2}d = 0, \|d\| < 1\}$. In terms of feasible directions contained in the set \mathcal{T}_x , we observe that condition (15) can be rewritten as follows:

$$\begin{aligned} &\nabla^2 f(\bar{x}) + \sqrt{\varepsilon_2}H(\bar{x}) \geq 0 \text{ on } L_0 \\ &\iff \langle \nabla^2 f(x)[H(x)]^{-1/2}u, [H(x)]^{-1/2}u \rangle \geq -\sqrt{\varepsilon_2} \\ &\quad \|u\|^2 \quad \forall u \in \ker(\mathbf{A}[H(x)]^{-1/2}) \\ &\iff \langle \nabla^2 f(x)[H(x)]^{-1/2}u, [H(x)]^{-1/2}u \rangle \geq -\sqrt{\varepsilon_2} \\ &\quad \forall u \in \ker(\mathbf{A}[H(x)]^{-1/2}) \cap \{u \in \mathbb{E} \mid \|u\| < 1\} \\ &\iff \langle \nabla^2 f(x)v, v \rangle \geq -\sqrt{\varepsilon_2} \quad \forall v \in \mathcal{T}_x. \end{aligned}$$

The last line connects our approximate KKT condition with the exact condition stated in Theorem 6 in Appendix B. The next Proposition gives a justification of Definition 5. We again use a perturbation argument based on an interior penalty approach, inspired by [58].

Proposition 4 *Let $x^* \in \bar{X}$ be a local solution of problem (Opt), where f is twice continuously differentiable on X . Then, there exist sequences $(x_k)_{k \geq 1} \subset X$, $(y_k)_{k \geq 1} \subset \mathbb{R}^m$ and $s_k = \nabla f(x_k) - \mathbf{A}^*y_k$, $k \geq 1$, satisfying the following:*

- (i) $x_k \rightarrow x^*$.
- (ii) For all $\varepsilon_1 > 0$ we have $|\langle s_k, x_k \rangle| \leq \varepsilon_1$ for all k sufficiently large.
- (iii) All accumulation points of the sequence $(s_k)_{k \geq 1}$ are contained in $-\text{NC}_{\bar{K}}(x^*) \subseteq \bar{K}^*$.
- (iv) For all sequences $v_k \in \mathcal{T}_{x_k}$ we have

$$\liminf_{k \rightarrow \infty} \langle \nabla^2 f(x_k)v_k, v_k \rangle \geq 0.$$

Proof Consider the penalty function

$$\varphi_k(x) \triangleq f(x) + \mu_k h(x) + \frac{1}{4} \|x - x^*\|^4.$$

It is easy to see that the proof of Proposition 3 still applies, *mutatis mutandis*, showing that items (i)–(iii) of the current Proposition hold.

(iv) The point x_k satisfies the necessary second-order condition

$$\left\langle \left(\nabla^2 f(x_k) + \mu_k \nabla^2 h(x_k) + \|x_k - x^*\|^2 \text{Id}_{\mathbb{E}} + 2(x_k - x^*) \otimes (x_k - x^*) \right) v, v \right\rangle \geq 0$$

for all $v \in L_0$. Here $\text{Id}_{\mathbb{E}}$ is the identity operator on \mathbb{E} . This reads as

$$\langle \nabla^2 f(x_k)v, v \rangle \geq -\mu_k \|v\|_{x_k}^2 - \rho_k \|v\|^2,$$

where ρ_k is the largest eigenvalue of $\|x_k - x^*\|^2 \text{Id}_{\mathbb{E}} + 2(x_k - x^*) \otimes (x_k - x^*)$. We know that $x_k \rightarrow x^*$. Let $(v_k)_{k \geq 1}$ be an arbitrary sequence with $v_k \in \mathcal{T}_{x_k}$ for all $k \geq 1$. Then, there exists a sequence $(u_k)_{k \geq 1}$ satisfying $v_k = [H(x_k)]^{-1/2}u_k$ and $\|u_k\| < 1$. Therefore, $\|v_k\|_{x_k}^2 = \|u_k\|^2 < 1$, and $\|v_k\| = \|u_k\|_{x_k}^* \leq \|x_k\| e_*(u_k)$, where we have used Lemma 5 and the definition $e_*(v) \triangleq \sup_{x \in K: \|x\|=1} \|v\|_x^*$. We thus obtain the bound

$$\langle \nabla^2 f(x_k)v_k, v_k \rangle \geq -\mu_k \|u_k\|^2 - \rho_k \|x_k\|^2 e_*(u_k)^2.$$

Since the sequence $(u_k)_{k \geq 1}$ is bounded, $\mu_k \rightarrow 0^+$ and $\rho_k \rightarrow 0^+$ as $k \rightarrow \infty$, the claim follows. □

3.3 Discussion

3.3.1 Comparison with approximate KKT conditions for interior-point methods

To compare our approximate KKT conditions with the ones previously formulated in the literature we consider the particular case $\bar{K} = \bar{K}_{\text{NN}} \subset \mathbb{E} = \mathbb{R}^n$ as in [58, 82]. The exact optimality condition for problem (Opt), assuming that f is continuously differentiable at x^* , reads in this case as the complementarity system for the primal-dual triple $(x^*, y^*, s^*) \in \bar{K}_{\text{NN}} \times \mathbb{R}^m \times \bar{K}_{\text{NN}}$ given by

$$\begin{aligned} \mathbf{A}x^* &= b, \quad s^* = \nabla f(x^*) - \mathbf{A}^*y^* \in \bar{K}_{\text{NN}}, \\ (\forall i = 1, \dots, n) : x_i^*s_i^* &= 0. \end{aligned}$$

These conditions directly motivate the approximate optimality conditions used in the interior-point method of [58], which are $\mathbf{A}^*\bar{x} = b, \bar{x} > 0$, and

$$\bar{s} = \nabla f(\bar{x}) - \mathbf{A}^*\bar{y} \geq -\varepsilon \mathbf{1}_n, \tag{17}$$

$$\max_{1 \leq i \leq n} |\bar{x}_i[\nabla f(\bar{x}) - \mathbf{A}^*\bar{y}]_i| \leq \varepsilon, \tag{18}$$

$$d^\top \left(\text{diag}[\bar{x}] \nabla^2 f(\bar{x}) \text{diag}[\bar{x}] + \sqrt{\varepsilon} \mathbf{I} \right) d \geq 0 \quad \forall d \in \ker(\mathbf{A}[H(\bar{x})]^{-1/2}), \tag{19}$$

where $\mathbf{1}_n \in \mathbb{R}^n$ is the vector of all ones and $\mathbf{I} = \text{Id}_{\mathbb{R}^n}$ is the $n \times n$ identity matrix. Note that, in this particular case, our first-order conditions (10)–(12) are similar but a bit stronger since they imply (17) and (18). Note also that the change of variable $v = [H(\bar{x})]^{-1/2}d = \text{diag}[\bar{x}]d$ shows the equivalence between our second-order condition (15) and the condition (19). These observations provide additional motivation for our definitions of approximate KKT points since similar in spirit definitions of approximate KKT points were previously used in the literature, i.e., in [58].

As pointed out in [82], conditions (18) and (19) are commonly used approximate optimality conditions for (Opt) [15, 36]. However, these two conditions alone are insufficient to guarantee that a sequence of points that satisfies these conditions as $\varepsilon \rightarrow 0$ converges to a KKT point for f [58]. For this reason the condition (17) is added in [58]. These conditions can be overly stringent for coordinates i in which \bar{x}_i is positive and numerically large (which is possible only if the norm of the vectors is not too much restricted in concrete instances, due to our compactness assumption). In this case, the complementarity condition (18) requires the corresponding dual variable \bar{s}_i to be very small. Similarly, (19) requires that the Hessian in the subspace spanned by these coordinates can have only minimal negative curvature. Such requirements contrast sharply with the case of unconstrained minimization. In the limiting scenario in which all of the coordinates of \bar{x} are far from the boundary, these approximate first-order conditions are significantly harder to satisfy than in the (equivalent) unconstrained formulation. To remedy this potential problem, O’Neill and Wright [82] proposed scaling in Eqs. (18) and (19) only when $\bar{x}_i \in (0, 1]$. This operation aims for an interpolation between the bound-constrained case (when \bar{x}_i is small) and the unconstrained case (when \bar{x}_i is large), while also controlling the norm of the matrix used in their optimality conditions. Since [82] only assume non-negativity constraints without linear equality constraints, and no further upper bounds on the decision variables, this clipping of variables makes a lot of sense. However, since we assume compactness, the coordinates of the approximate solutions can become large, but only up to a pre-defined and known upper bound. Hence, the hardness issue of identifying an approximate KKT point is less pronounced in our work. Moreover, we prove that our algorithms produce approximate KKT points with standard scaling and with a similar iteration complexity as in the setting of [82]. Finally, it is easy to show that the conditions with interpolating scaling of [82] follow in our setting for $\bar{K} = \bar{K}_{NN}$ as well since $0 \leq \min\{x_i, 1\} \leq x_i$ in this case.

3.3.2 Complementarity conditions for symmetric cones

If \bar{K} is a symmetric cone, the approximate complementarity conditions (12) and (14) are equivalent to approximate complementarity conditions formulated in terms of the multiplication \circ under which K becomes an Euclidean Jordan algebra. [72, Prop.2.1] shows that $x \circ y = 0$ if and only if $\langle x, y \rangle = 0$, where $\langle \cdot, \cdot \rangle$ is the inner product of the space \mathbb{E} . Moreover, if \bar{K} is a primitive symmetric cone, then by [45, Prop.III.4.1], there exists a constant $a > 0$ such that $a \operatorname{tr}(x \circ y) = \langle x, y \rangle$ for all $x, y \in K$. In view of this relation, our approximate complementarity conditions can be written in the form of condition $\bar{s} \circ \bar{x} \leq \varepsilon$. Hence, our approximate KKT conditions reduce to the ones reported in [6]. In particular, for $\bar{K} = \bar{K}_{NN}$ we recover the standard approximate complementary slackness condition $\bar{s}_i^k \bar{x}_i^k \leq \varepsilon$ for all i , as in this case the Jordan product \circ gives rise to the Hadamard product. See [5] for more details.

3.3.3 On the relation to scaled critical points

In absence of differentiability at the boundary, a popular formulation of necessary optimality conditions involves the definition of scaled critical points. Indeed, at a local minimizer x^* , the scaled first-order optimality condition $x_i^* [\nabla f(x^*)]_i = 0, 1 \leq i \leq n$

holds, where the product is taken to be 0 when the derivative does not exist. Based on this characterization, one may call a point $x \in K_{NN}$ with $|x_i[\nabla f(x)]_i| \leq \varepsilon$ for all $i = 1, \dots, n$ an ε -scaled first-order point. Algorithms designed to produce ε -scaled first-order points, with some small $\varepsilon > 0$, have been introduced in [15, 16]. As reported in [58], there are several problems associated with this weak definition of a critical point. First, when derivatives are available on \bar{K}_{NN} , the standard definition of a critical point would entail the inequality $\langle \nabla f(x), x' - x \rangle \geq 0$ for all $x' \in \bar{K}_{NN}$. Hence, $[\nabla f(x)]_i = 0$ for $x_i > 0$ and $[\nabla f(x)]_i \geq 0$ for $x_i = 0$. It follows that $\nabla f(x) \in \bar{K}_{NN}$, a condition that is absent in the definition of a scaled critical point. Second, scaled critical points come with no measure of strength, as they hold trivially when $x = 0$, regardless of the objective function. Third, there is a general gap between local minimizers and limits of ε -scaled first-order points, when $\varepsilon \rightarrow 0^+$ (see [58]). Similar remarks apply to the scaled second-order condition, considered in [15]. Our definition of approximate KKT points overcome these issues. In fact, our definitions of approximate first- and second-order KKT points are continuous in ε , and therefore in the limit our approximate KKT conditions coincide with first- and *weak* second-order necessary conditions for a local minimizer. This is achieved without assuming global differentiability of the objective function or performing an additional smoothing of the problem data as in [14, 15].

3.3.4 Second-order conditions in the literature on interior-point methods compared to standard second-order conditions

As discussed above, our necessary second-order optimality conditions stated in Proposition 4, and their approximate counterpart given in Definition 5, are in alignment with the existing body of work that studies non-convex conic optimization [58, 61, 82]. In particular, condition (15) coincides with those used in [58, 82] for the particular case of non-negativity constraints, and in [61] for general conic constraints. However, the necessary second-order optimality conditions in Proposition 4, [58, Theorem 1], [82, Theorem 3.1], [61, Theorem 4] are weaker than the standard ones that we refer to as *strong* conditions. This can be easily illustrated with the following example.³

Example 11 (Strong vs. weak necessary second-order optimality conditions) Consider $\mathbb{E} = \mathbb{R}$ and $\bar{K} = \bar{K}_{NN} = \mathbb{R}_+$. Also, assume that there are no linear equality constraints. Consider the function $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ defined by

$$f(x) = \begin{cases} -\frac{1}{2}x^2 & \text{if } x \in [0, 1], \\ \frac{1}{2}(x - 2)^2 - 1 & \text{if } x \in (1, 3], \\ x - \frac{7}{2} & \text{if } x \in (3, \infty). \end{cases}$$

This function is continuously differentiable on \mathbb{R}_{++} , bounded from below, and has Lipschitz gradient. The minimization problem $\min\{f(x) | x \in \mathbb{R}_+\}$ has two first-order KKT points $x \in \{0, 2\}$, both supported by Lagrange multiplier $s = 0$. A strong second-order necessary optimality condition involves the *strong* critical cone [81,

³ We thank an anonymous referee who proposed to us a similar example.

Section 12.5], which in the current example reads as

$$C^s(x) = \{d \in \mathbb{R} \mid d \geq 0 \text{ if } x = 0\}.$$

A strong second-order necessary condition is then

$$\langle d, \nabla^2 f(\bar{x})d \rangle \geq 0 \quad \forall d \in C^s(\bar{x}).$$

Since $C^s(0) = [0, \infty)$, but $\nabla^2 f(0) = -1$, it follows that $\bar{x} = 0$ does not satisfy the strong second-order necessary condition. However, $C^s(2) = \mathbb{R}$, and $\nabla^2 f(2) = 1$. In fact $(x^*, s^*) = (2, 0)$ is the global solution to the problem. When the strong critical cone is replaced by the weak critical cone [52, 53]

$$C^w(x) = \{d \in \mathbb{R} \mid d = 0 \text{ if } x = 0\},$$

the weak second-order necessary condition reads as

$$\langle d, \nabla^2 f(\bar{x})d \rangle \geq 0 \quad \forall d \in C^w(\bar{x}). \tag{20}$$

Clearly $\bar{x} = 0$ satisfies this weak second-order condition.

At the same time, $\bar{x} = 0$ is a second-order stationary point in the sense of Proposition 4 since the necessary conditions (i)–(iv) of that Proposition hold. Indeed, let us consider the sequence $x_k = \frac{1}{k}$, $k \geq 1$ and the corresponding sequence $s_k = \nabla f(x_k) = -x_k = -\frac{1}{k}$. Then, clearly $x_k \rightarrow 0 = \bar{x}$ and condition (i) holds. Condition (ii) also holds since $|\langle s_k, x_k \rangle| = \frac{1}{k^2} \leq \varepsilon_1$ for sufficiently large k . Since $s_k \rightarrow 0$ as $k \rightarrow \infty$, we have that condition (iii) holds as well. We finally show that (iv) holds. Let us consider arbitrary $v_k \in \mathcal{J}_{x_k}$. Then, by the Definition (16) of \mathcal{J}_{x_k} we have that $1 > \|v_k\|_{x_k} = \sqrt{\langle H(x_k)v_k, v_k \rangle} = |v_k/x_k| = |kv_k|$, which implies $|v_k| \leq \frac{1}{k}$ and $\langle \nabla^2 f(x_k)v_k, v_k \rangle = -v_k^2 \rightarrow 0$ as $k \rightarrow \infty$, which finishes the proof of (iv). Thus, we proved that $\bar{x} = 0$ is a second-order stationary point in the sense of Proposition 4. Importantly, one can show that in this example the point $\bar{x} = 0$ satisfies also second-order necessary optimality conditions in [58, Theorem 1], [82, Theorem 3.1], [61, Theorem 4]. Overall, in this example we see that $\bar{x} = 0$ is not a strong second-order stationary point, but is a weak second-order stationary point since (20) as well as conditions in Proposition 4, [58, Theorem 1], [82, Theorem 3.1], [61, Theorem 4] hold. \diamond

Our second-order necessary conditions involve the set \mathcal{J}_x . Theorem 3.2 in [73] demonstrates that the cone generated by \mathcal{J}_x coincides with the weak critical cone for conic programming, which is used in weak second-order necessary conditions [8, 52]. At the same time, weak second-order necessary conditions are appropriate notions for barrier algorithms. Indeed, [53] gives an explicit example illustrating that accumulation points of trajectories generated by barrier algorithms will converge to stationary points satisfying the weak second-order necessary condition, but not the strong version. Later, Andreani and Secchin [10] made a small modification in Gould and Toint's

counterexample to come to the same conclusion for augmented Lagrangian-type algorithms. Hence, the most we can expect from our method is that it generates points that approximately satisfy the weak second-order necessary optimality conditions.

4 A first-order Hessian barrier algorithm

In this section we introduce a first-order potential reduction method for solving (Opt) that uses a barrier $h \in \mathcal{H}_\theta(\mathbb{K})$ and potential function (2). We assume that we are able to compute an approximate analytic center at low computational cost. Specifically, our algorithm relies on the availability of a θ -analytic center, i.e. a point $x^0 \in X$ such that

$$h(x) \geq h(x^0) - \theta \quad \forall x \in X. \tag{21}$$

To obtain such a point x^0 , one can apply interior point methods to the convex programming problem $\min_{x \in X} h(x)$. Moreover, since $\theta \geq 1$ we do not need to solve it with high precision, making the application of computationally cheap first-order method, such as [43], an appealing choice for this preprocessing step.

4.1 Local properties

Our complexity analysis relies on the ability to control the behavior of the objective function along the set of feasible directions and with respect to the local norm.

Assumption 3 (Local smoothness) $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{+\infty\}$ is continuously differentiable on X and there exists a constant $M > 0$ such that for all $x \in X$ and $v \in \mathcal{T}_x$, where \mathcal{T}_x is defined in (16), we have

$$f(x + v) - f(x) - \langle \nabla f(x), v \rangle \leq \frac{M}{2} \|v\|_x^2. \tag{22}$$

Remark 2 If the set \bar{X} is bounded, we have $\lambda_{\min}(H(x)) \geq \sigma$ for some $\sigma > 0$. In this case, assuming f has an M -Lipschitz continuous gradient, the classical descent lemma [76] implies Assumption 3. Indeed,

$$f(x + v) - f(x) - \langle \nabla f(x), v \rangle \leq \frac{M}{2} \|v\|^2 \leq \frac{M}{2\sigma} \|v\|_x^2.$$

◇

Remark 3 We emphasize that the local Lipschitz smoothness condition (22) does not require global differentiability. Consider the composite non-smooth and non-convex model (1) on \bar{K}_{NN} , with $\varphi(s) = s$ for $s \geq 0$. This means $\sum_{i=1}^n \varphi(x_i^p) = \|x\|_p^p$ for $p \in (0, 1)$ and $x \in \bar{K}_{\text{NN}}$. As a concrete example for the smooth part of the problem let us consider the L_2 -loss $\ell(x) = \frac{1}{2} \|\mathbf{N}x - \mathbf{p}\|^2$. This gives rise to the $L_2 - L_p$ minimization problem, an important optimization formulation arising in

phase retrieval, mathematical statistics, signal processing and image recovery [36, 49, 50, 71]. Setting $M = \lambda_{\max}(\mathbf{N}^*\mathbf{N})$, the descent lemma yields

$$\ell(x^+) \leq \ell(x) + \langle \nabla \ell(x), x^+ - x \rangle + \frac{M}{2} \|x^+ - x\|^2,$$

for $x, x^+ \in K_{\text{NN}}$. Since $t \mapsto t^p$ is concave for $t > 0$ and $p \in (0, 1)$, we have, for $x, x^+ \in K_{\text{NN}}$,

$$(x_i^+)^p \leq x_i^p + px_i^{p-1}(x_i^+ - x_i) \quad i = 1, \dots, n.$$

Adding these inequalities, we immediately arrive at condition (22) in terms of the Euclidean norm. Over a bounded feasible set X , this implies Assumption 3 (cf. Remark 2). At the same time, f is not differentiable for $x \in \text{bd}(K_{\text{NN}}) = \{x \in \mathbb{R}_+^n \mid x_i = 0 \text{ for some } i\}$. ◇

We emphasize that in Assumption 3, the Lipschitz modulus M is hardly known exactly in practice, and it is also not an easy task to obtain universal upper bounds that can be efficiently used in the algorithm. Therefore, adaptive techniques should be used to estimate it and are likely to improve the practical performance of the method.

Considering $x \in X, v \in \mathcal{T}_x$ and combining Eq. (22) with Eq. (7) (with $d = v$ and $t = 1 < \frac{1}{\|v\|_x} \stackrel{(5)}{\leq} \frac{1}{\zeta(x,v)}$) reveals a suitable quadratic model, to be used in the design of our first-order algorithm.

Lemma 2 (Quadratic overestimation) *For all $x \in X, v \in \mathcal{T}_x$ and $L \geq M$, we have*

$$F_\mu(x + v) \leq F_\mu(x) + \langle \nabla F_\mu(x), v \rangle + \frac{L}{2} \|v\|_x^2 + \mu \|v\|_x^2 \omega(\zeta(x, v)). \quad (23)$$

4.2 Algorithm description and its complexity

4.2.1 Defining the step direction

Let $x \in X$ be given. Our first-order method employs a quadratic model $Q_\mu^{(1)}(x, v)$ to compute a search direction $v_\mu(x)$, given by

$$v_\mu(x) \triangleq \operatorname{argmin}_{v \in \mathbb{E}: \mathbf{A}v=0} \left\{ Q_\mu^{(1)}(x, v) \triangleq F_\mu(x) + \langle \nabla F_\mu(x), v \rangle + \frac{1}{2} \|v\|_x^2 \right\}. \quad (24)$$

For the above problem, we have the following system of optimality conditions involving the dual variable $y_\mu(x) \in \mathbb{R}^m$:

$$\nabla F_\mu(x) + H(x)v_\mu(x) - \mathbf{A}^*y_\mu(x) = 0, \quad (25)$$

$$\mathbf{A}v_\mu(x) = 0. \quad (26)$$

Since $H(x) \succ 0$ for $x \in X$, any standard solution method [81] can be applied for the above linear system. Moreover, this system can be solved explicitly. Indeed, since $H(x) \succ 0$ for $x \in X$, and \mathbf{A} has full row rank, the linear operator $\mathbf{A}[H(x)]^{-1}\mathbf{A}^*$ is invertible. Hence, $v_\mu(x)$ is given explicitly as

$$v_\mu(x) = -([H(x)]^{-1}\mathbf{A}^*(\mathbf{A}[H(x)]^{-1}\mathbf{A}^*)^{-1}\mathbf{A}[H(x)]^{-1} - [H(x)]^{-1})\nabla F_\mu(x) \triangleq -\mathbf{S}_x\nabla F_\mu(x).$$

To give some intuition behind this expression, observe that we can give an alternative representation of \mathbf{S}_x as $\mathbf{S}_x v = [H(x)]^{-1/2}\Pi_x[H(x)]^{-1/2}v$, where

$$\Pi_x v \triangleq v - [H(x)]^{-1/2}\mathbf{A}^*(\mathbf{A}[H(x)]^{-1}\mathbf{A}^*)^{-1}\mathbf{A}[H(x)]^{-1/2}v.$$

This shows that \mathbf{S}_x is just the $\|\cdot\|_x$ -orthogonal projection operator onto $\ker(\mathbf{A}[H(x)]^{-1/2})$.

4.2.2 Defining the step-size

To determine an acceptable step-size, consider a point $x \in X$. The search direction $v_\mu(x)$ gives rise to a family of parameterized arcs $x^+(t) \triangleq x + tv_\mu(x)$, where $t \geq 0$. Our aim is to choose this step-size to ensure feasibility of iterates and decrease of the potential. By (6) and (26), we know that $x^+(t) \in X$ for all $t \in I_{x,\mu} \triangleq [0, \frac{1}{\zeta(x, v_\mu(x))})$. Multiplying (25) by $v_\mu(x)$ and using (26), we obtain $\langle \nabla F_\mu(x), v_\mu(x) \rangle = -\|v_\mu(x)\|_x^2$. Choosing $t \in I_{x,\mu}$, we bound

$$t^2\|v_\mu(x)\|_x^2\omega(t\zeta(x, v_\mu(x))) \stackrel{(3)}{\leq} \frac{t^2\|v_\mu(x)\|_x^2}{2(1 - t\zeta(x, v_\mu(x)))}.$$

Therefore, if $t\zeta(x, v_\mu(x)) \leq 1/2$, we readily see from (23) that

$$\begin{aligned} F_\mu(x^+(t)) - F_\mu(x) &\leq -t\|v_\mu(x)\|_x^2 + \frac{t^2M}{2}\|v_\mu(x)\|_x^2 + \mu t^2\|v_\mu(x)\|_x^2 \\ &= -t\|v_\mu(x)\|_x^2 \left(1 - \frac{M + 2\mu}{2}t\right) \triangleq -\eta_x(t). \end{aligned} \tag{27}$$

The function $t \mapsto \eta_x(t)$ is strictly concave with the unique maximum at $\frac{1}{M+2\mu}$, and two real roots at $t \in \left\{0, \frac{2}{M+2\mu}\right\}$. Thus, maximizing the per-iteration decrease $\eta_x(t)$ under the restriction $0 \leq t \leq \frac{1}{2\zeta(x, v_\mu(x))}$, we choose the step-size

$$\tau_{\mu,M}(x) \triangleq \min \left\{ \frac{1}{M + 2\mu}, \frac{1}{2\zeta(x, v_\mu(x))} \right\}.$$

Algorithm 1: First-order Adaptive Hessian Barrier Algorithm - **FAHBA**($\mu, \varepsilon, L_0, x^0$)

Data: $h \in \mathcal{H}_\theta(K), \mu > 0, \varepsilon > 0, L_0 > 0, x^0 \in X$.

Result: $(x^k, y^k, s^k, L_k) \in X \times \mathbb{R}^m \times K^* \times \mathbb{R}_+$, where $s^k = \nabla f(x^k) - A^*y^k$, and L_k is the last estimate of the Lipschitz constant in (22).

Set $k = 0$;

repeat

Set $i_k = 0$. Find $v^k \triangleq v_\mu(x^k)$ and the corresponding dual variable $y^k \triangleq y_\mu(x^k)$ as the solution to

$$\min_{v \in \mathbb{E}: Av=0} \{F_\mu(x^k) + \langle \nabla F_\mu(x^k), v \rangle + \frac{1}{2} \|v\|_{x^k}^2\}. \tag{28}$$

repeat

$$\alpha_k \triangleq \min \left\{ \frac{1}{2^{i_k} L_k + 2\mu}, \frac{1}{2\zeta(x^k, v^k)} \right\}, \text{ where } \zeta(\cdot, \cdot) \text{ as in (4)} \tag{29}$$

Set $z^k = x^k + \alpha_k v^k, i_k = i_k + 1$;

until

$$f(z^k) \leq f(x^k) + \langle \nabla f(x^k), z^k - x^k \rangle + 2^{i_k-1} L_k \|z^k - x^k\|_{x^k}^2. \tag{30}$$

;

Set $L_{k+1} = 2^{i_k-1} L_k, x^{k+1} = z^k, k = k + 1$;

until $\|v^k\|_{x^k} < \frac{\varepsilon}{\theta}$;

4.2.3 Backtracking on the Lipschitz modulus

The above step-size rule, however, requires knowledge of the parameter M . To boost numerical performance, we employ a backtracking scheme in the spirit of [79] to estimate the constant M at each iteration. This procedure generates a sequence of positive numbers $(L_k)_{k \geq 0}$ for which the local Lipschitz smoothness condition (22) holds. More specifically, suppose that x^k is the current position of the algorithm with the corresponding initial local Lipschitz estimate L_k and $v^k = v_\mu(x^k)$ is the corresponding search direction. To determine the next iterate x^{k+1} , we iteratively try step-sizes α_k of the form $\tau_{\mu, 2^{i_k} L_k}(x^k)$ for $i_k \geq 0$ until the local smoothness condition (22) holds with $x = x^k, v = \alpha_k v^k$ and local Lipschitz estimate $M = 2^{i_k} L_k$, see (30). This process must terminate in finitely many steps, since when $2^{i_k} L_k \geq M$, inequality (22) with M changed to $2^{i_k} L_k$, i.e., (30), follows from Assumption 3.

4.2.4 First-order algorithm and its complexity result

Combining the search direction finding problem (24) with the just outlined backtracking strategy, yields a First-order Adaptive Hessian Barrier Algorithm (**FAHBA**, Algorithm 1).

Our main result on the iteration complexity of Algorithm 1 is the following Theorem, whose proof is given in Sect. 4.3.

Theorem 1 *Let Assumptions 1–3 hold. Fix the error tolerance $\varepsilon > 0$, the regularization parameter $\mu = \frac{\varepsilon}{\theta}$, and some initial guess $L_0 > 0$ for the Lipschitz constant in (22). Let $(x^k)_{k \geq 0}$ be the trajectory generated by FAHBA($\mu, \varepsilon, L_0, x^0$), where x^0 is a θ -analytic center satisfying (21). Then the algorithm stops in no more than*

$$\mathbb{K}_I(\varepsilon, x^0) = \left\lceil 4(f(x^0) - f_{\min}(X) + \varepsilon) \frac{\theta^2(\max\{M, L_0\} + \varepsilon/\theta)}{\varepsilon^2} \right\rceil \tag{31}$$

outer iterations, and the number of inner iterations is no more than $2(\mathbb{K}_I(\varepsilon, x^0) + 1) + \max\{\log_2(M/L_0), 0\}$. Moreover, the last iterate obtained from FAHBA($\mu, \varepsilon, L_0, x^0$) constitutes a 2ε -KKT point for problem (Opt) in the sense of Definition 4.

Remark 4 The line-search process of finding the appropriate i_k is simple since only recalculating z^k is needed, and repeatedly solving problem (28) is not required. Furthermore, the sequence of constants L_k is allowed to decrease along subsequent iterations, which is achieved by the division by the constant factor 2 in the final updating step of each iteration. This potentially leads to longer steps and faster decrease of the potential. \diamond

Remark 5 Since $\theta \geq 1$, $f(x^0) - f_{\min}(X)$ is expected to be larger than ε , and the constant M is potentially large, we see that the main term in the complexity bound (31) is $O\left(\frac{M\theta^2(f(x^0) - f_{\min}(X))}{\varepsilon^2}\right) = O\left(\frac{\theta^2}{\varepsilon^2}\right)$, i.e., has the same dependence on ε as the standard complexity bounds [24, 25, 69] of first-order methods for non-convex problems under the standard Lipschitz-gradient assumption, which on bounded sets is subsumed by our Assumption 3. Further, if the function f is linear, Assumption 3 holds with $M = 0$ and we can take $L_0 = 0$. In this case, the complexity bound (31) improves to $O\left(\frac{\theta(f(x^0) - f_{\min}(X))}{\varepsilon}\right)$. \diamond

Remark 6 Just like classical interior-point methods, the iteration complexity of FAHBA depends on the barrier parameter $\theta \geq 1$. For conic domains, the characterization of this barrier parameter has thus been an active research line. [56] demonstrated that for symmetric cones, the barrier parameter is equivalent to algebraic properties of the cone and identified it with the rank of the cone (see [45] for a definition of the rank of a symmetric cone). This deep analysis gives an exact characterization of the optimal barrier parameter for the most important conic domains in optimization. For K_{NN} and K_{SDP} , it is known that $\theta = n$ is optimal, whereas for K_{SOC} the optimal barrier parameter is $\theta = 2$ (and therefore independent of the dimension n). \diamond

4.2.5 Connection with interior point flows on polytopes

Consider $\bar{K} = \bar{K}_{NN}$, and $X = K_{NN} \cap L$. We are given a function $f : \bar{X} \rightarrow \mathbb{R}$ which is the restriction of a smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. The canonical barrier for this setting is $h(x) = -\sum_{i=1}^n \ln(x_i)$, so that $H(x) = \text{diag}[x_1^{-2}, \dots, x_n^{-2}] \triangleq X^{-2}$ for $x \in X$. Applying our first-order method on this domain gives the search direction $v_\mu(x) = -S_x \nabla F_\mu(x) = -X(I - XA^\top(A X^2 A^\top)^{-1}AX)X \nabla F_\mu(x)$. This explicit formula yields

various interesting connections between our approach and classical methods. First, for $f(x) = c^\top x$ and $\mu = 0$, we obtain from this formula the search direction employed in *affine scaling* methods for linear programming [1, 11, 12]. Second, [22] partly motivated their algorithm as a discretization of the Hessian-Riemannian gradient flows introduced in [4, 21]. Heuristically, we can think of **FAHBA** as an explicit Euler discretization (with non-monotone adaptive step-size policies) of the gradient-like flow $\dot{x}(t) = -\mathbf{S}_{x(t)} \nabla F_\mu(x(t))$, which resembles very much the class of dynamical systems introduced in [21]. This gives an immediate connection to a large class of interior point flows on polytopes, heavily studied in control theory [62].

4.3 Proof of Theorem 1

Our proof proceeds in several steps. First, we show that procedure **FAHBA**($\mu, \varepsilon, L_0, x^0$) produces points in X , and, thus, is indeed an interior-point method. Next, we show that the line-search process of finding appropriate L_k in each iteration is finite, and estimate the total number of trials in this process. Then we enter the core of our analysis where we prove that if the stopping criterion does not hold at iteration k , i.e., $\|v^k\|_{x^k} \geq \frac{\varepsilon}{\theta}$, then the objective f is decreased by a quantity $O(\varepsilon^2)$, and, since the objective is globally lower bounded, we conclude that the method stops in at most $O(\varepsilon^{-2})$ iterations. Finally, we show that when the stopping criterion holds, the method has generated an ε -KKT point.

4.3.1 Interior-point property of the iterates

By construction $x^0 \in X$. Proceeding inductively, let $x^k \in X$ be the k -th iterate of the algorithm, delivering the search direction $v^k \triangleq v_\mu(x^k)$. By Eq. (29), the step-size α_k satisfies $\alpha_k \leq \frac{1}{2\zeta(x^k, v^k)}$, and, hence, $\alpha_k \zeta(x^k, v^k) \leq 1/2$ for all $k \geq 0$. Thus, by (6) $x^{k+1} = x^k + \alpha_k v^k \in K$. Since, by (28), $\mathbf{A}v^k = 0$, we have that $x^{k+1} \in L$. Thus, $x^{k+1} \in K \cap L = X$. By induction, we conclude that $(x^k)_{k \geq 0} \subset X$.

4.3.2 Bounding the number of backtracking steps

Let us fix iteration k . Since the sequence $2^{i_k} L_k$ is increasing as i_k is increasing, and Assumption 3 holds, we know that when $2^{i_k} L_k \geq \max\{M, L_k\}$, the line-search process for sure stops since inequality (30) holds. Hence, $2^{i_k} L_k \leq 2 \max\{M, L_k\}$ must be the case, and, consequently, $L_{k+1} = 2^{i_k-1} L_k \leq \max\{M, L_k\}$, which, by induction, gives $L_{k+1} \leq \bar{M} \triangleq \max\{M, L_0\}$. At the same time, $\log_2\left(\frac{L_{k+1}}{L_k}\right) = i_k - 1, \forall k \geq 0$. Let $N(k)$ denote the number of inner line-search iterations up to the k -th iteration of **FAHBA**($\mu, \varepsilon, L_0, x^0$). Then, using that $L_{k+1} \leq \bar{M} = \max\{M, L_0\}$,

$$N(k) = \sum_{j=0}^k (i_j + 1) = \sum_{j=0}^k (\log_2(L_{j+1}/L_j) + 2) \leq 2(k + 1) + \max\{\log_2(M/L_0), 0\}.$$

This shows that on average the inner loop ends after two trials.

4.3.3 Per-iteration analysis and a bound for the number of iterations

Let us fix iteration counter k . Since $L_{k+1} = 2^{i_k-1}L_k$, the step-size (29) reads as $\alpha_k = \min \left\{ \frac{1}{2L_{k+1}+2\mu}, \frac{1}{2\zeta(x^k, v^k)} \right\}$. Hence, $\alpha_k \zeta(x^k, v^k) \leq 1/2$, and (27) with the specification $t = \alpha_k = \tau_{\mu, 2L_{k+1}}(x^k)$, $M = 2L_{k+1}$, $x = x^k$, $v_\mu(x^k) \triangleq v^k$ gives:

$$F_\mu(x^{k+1}) - F_\mu(x^k) \leq -\alpha_k \|v^k\|_{x^k}^2 (1 - (L_{k+1} + \mu)\alpha_k) \leq -\frac{\alpha_k \|v^k\|_{x^k}^2}{2}, \tag{32}$$

where we used that $\alpha_k \leq \frac{1}{2(L_{k+1} + \mu)}$ in the last inequality. Substituting into (32) the two possible values of the step-size α_k in (29) gives

$$F_\mu(x^{k+1}) - F_\mu(x^k) \leq \begin{cases} -\frac{\|v^k\|_{x^k}^2}{4(L_{k+1} + \mu)} & \text{if } \alpha_k = \frac{1}{2(L_{k+1} + \mu)} \\ -\frac{\|v^k\|_{x^k}^2}{4\zeta(x^k, v^k)} \stackrel{(5)}{\leq} -\frac{\|v^k\|_{x^k}}{4} & \text{if } \alpha_k = \frac{1}{2\zeta(x^k, v^k)}. \end{cases} \tag{33}$$

Recalling $L_{k+1} \leq \bar{M}$ (see Sect. 4.3.2), we obtain that

$$F_\mu(x^{k+1}) - F_\mu(x^k) \leq -\frac{\|v^k\|_{x^k}}{4} \min \left\{ 1, \frac{\|v^k\|_{x^k}}{\bar{M} + \mu} \right\} \triangleq -\delta_k.$$

Rearranging and summing these inequalities for k from 0 to $K - 1$ gives

$$\begin{aligned} K \min_{k=0, \dots, K-1} \delta_k &\leq \sum_{k=0}^{K-1} \delta_k \leq F_\mu(x^0) - F_\mu(x^K) \\ &\stackrel{(2)}{=} f(x^0) - f(x^K) + \mu(h(x^0) - h(x^K)) \leq f(x^0) - f_{\min}(\mathbf{X}) + \varepsilon, \end{aligned} \tag{34}$$

where we used that, by the assumptions of Theorem 1, x^0 is a θ -analytic center defined in (21) and $\mu = \varepsilon/\theta$, implying that $h(x^0) - h(x^K) \leq \theta = \varepsilon/\mu$. Thus, up to passing to a subsequence, $\delta_k \rightarrow 0$, and consequently $\|v^k\|_{x^k} \rightarrow 0$ as $k \rightarrow \infty$. This shows that the stopping criterion in Algorithm 1 is achievable.

Assume now that the stopping criterion $\|v^k\|_{x^k} < \frac{\varepsilon}{\theta}$ does not hold for K iterations of FAHBA. Then, for all $k = 0, \dots, K - 1$, it holds that $\delta_k \geq \min \left\{ \frac{\varepsilon}{4\theta}, \frac{\varepsilon^2}{4\theta^2(\bar{M} + \mu)} \right\}$. Together with the parameter coupling $\mu = \frac{\varepsilon}{\theta}$, it follows from (34) that

$$K \frac{\varepsilon^2}{4\theta^2(\bar{M} + \varepsilon/\theta)} = K \min \left\{ \frac{\varepsilon}{4\theta}, \frac{\varepsilon^2}{4\theta^2(\bar{M} + \varepsilon/\theta)} \right\} \leq f(x^0) - f_{\min}(\mathbf{X}) + \varepsilon.$$

Hence, recalling that $\bar{M} = \max\{M, L_0\}$, we obtain

$$K \leq 4(f(x^0) - f_{\min}(\mathbf{X}) + \varepsilon) \cdot \frac{\theta^2(\max\{M, L_0\} + \varepsilon/\theta)}{\varepsilon^2},$$

i.e., the algorithm stops for sure after no more than this number of iterations. This, combined with the bound for the number of inner steps in Sect. 4.3.2, proves the first statement of Theorem 1.

4.3.4 Generating ε -KKT point

To finish the proof of Theorem 1, we now show that when Algorithm 1 stops for the first time, it returns a 2ε -KKT point of (Opt) according to Definition 4.

Let the stopping criterion hold at iteration k . By the optimality condition (25) and the definition of the potential (2), we have

$$\begin{aligned} \nabla f(x^k) - \mathbf{A}^*y^k + \mu \nabla h(x^k) = -H(x^k)v^k &\iff [H(x^k)]^{-1} \\ &\left(\nabla f(x^k) - \mathbf{A}^*y^k + \mu \nabla h(x^k)\right) = -v^k. \end{aligned} \tag{35}$$

Denoting $g^k \triangleq -\mu \nabla h(x^k)$, multiplying both equations, and using the stopping criterion $\|v^k\|_{x^k} < \frac{\varepsilon}{\theta}$, we conclude

$$\|\nabla f(x^k) - \mathbf{A}^*y^k - g^k\|_{x^k}^* = \|v^k\|_{x^k} < \frac{\varepsilon}{\theta}.$$

Whence, setting $s^k \triangleq \nabla f(x^k) - \mathbf{A}^*y^k \in \mathbb{E}^*$, the definition of the dual norm yields

$$\begin{aligned} \frac{\varepsilon}{\theta} > \|v^k\|_{x^k} &= \|s^k - g^k\|_{x^k}^* \\ &= \|s^k - g^k\|_{[H(x^k)]^{-1}} \stackrel{(79)}{=} \|s^k - g^k\|_{\nabla^2 h_*(-\nabla h(x^k))} \\ &= \|s^k - g^k\|_{\nabla^2 h_*(\frac{1}{\mu}g^k)} = \mu \|s^k - g^k\|_{\nabla^2 h_*(g^k)}, \end{aligned} \tag{36}$$

where in the last equality we used the fact $h_* \in \mathcal{H}_\theta(K^*)$, so that Eq. (80) delivers the identity $\nabla^2 h_*(\frac{1}{\mu}g^k) = \mu^2 \nabla^2 h_*(g^k)$. Since we set $\mu = \frac{\varepsilon}{\theta}$, it follows

$$\|s^k - g^k\|_{\nabla^2 h_*(g^k)} = \frac{\|v^k\|_{x^k}}{\mu} < \frac{\varepsilon}{\mu\theta} = 1. \tag{37}$$

Thus, since, by (78), $g^k = -\mu \nabla h(x^k) \in \text{int}(K^*)$, applying Lemma 1, we can now conclude that $\nabla f(x^k) - \mathbf{A}^*y^k = s^k \in \text{int}(K^*)$ and therefore (11) holds. By construction, $x^k \in K$ and $\mathbf{A}x^k = b$. Thus, (10) holds as well. Finally, since $(x^k, s^k) \in K \times K^*$, we see

$$\begin{aligned} 0 \leq \langle s^k, x^k \rangle &= \langle s^k - g^k, x^k \rangle + \langle g^k, x^k \rangle \\ &\leq \|s^k - g^k\|_{x^k}^* \cdot \|x^k\|_{x^k} - \mu \langle \nabla h(x^k), x^k \rangle \\ &\stackrel{(36), (82), (81)}{=} \|v^k\|_{x^k} \sqrt{\theta} + \mu\theta \\ &< \sqrt{\theta} \frac{\varepsilon}{\theta} + \varepsilon \leq 2\varepsilon, \end{aligned} \tag{38}$$

where the last inequality uses $\theta \geq 1$. Hence, the complementarity condition (12) holds as well. This finishes the proof of Theorem 1.

4.4 Discussion

4.4.1 Special case of non-negativity constraints

For $\bar{K} = \bar{K}_{\text{NN}} = \mathbb{R}_+^n$, i.e., a particular case covered by our general problem template (Opt), [58] consider a first-order potential reduction method employing the standard log-barrier $h(x) = -\sum_{i=1}^n \ln(x_i)$ and using a trust-region subproblem for obtaining the search direction. For $x \in K_{\text{NN}}$, we have $\nabla h(x) = [-x_1^{-1}, \dots, -x_n^{-1}]^\top$, $H(x) = \text{diag}[x_1^{-2}, \dots, x_n^{-2}] \triangleq \mathbf{X}^{-2}$, which makes the problem in a sense simpler since we have a natural coupling between the variable x and the Hessian $H(x)$ expressed as $[H(x^k)^{-\frac{1}{2}}s^k]_i = x_i^k s_i^k$ and simplifying the derivation of the complementarity condition. More precisely, combining (35), the information $[H(x^k)]^{-1/2} \nabla h(x^k) = -\mathbf{1}_n$, $\theta = n$, and the stopping criterion of Algorithm 1 at iteration k , saying that $\|v^k\|_{x^k} < \frac{\varepsilon}{\theta}$, we see

$$\begin{aligned} \|[H(x^k)]^{-\frac{1}{2}}(\nabla f(x^k) - \mathbf{A}^*y^k) - \mu\mathbf{1}_n\|_\infty &\leq \|[H(x^k)]^{-\frac{1}{2}}(\nabla f(x^k) - \mathbf{A}^*y^k) - \mu\mathbf{1}_n\| \\ &= \|[H(x^k)]^{\frac{1}{2}}v^k\| < \frac{\varepsilon}{n}. \end{aligned}$$

Therefore, since $\mu = \varepsilon/n$ and $s^k = \nabla f(x^k) - \mathbf{A}^*y^k$, we have $H(x^k)^{-\frac{1}{2}}s^k = \mathbf{X}^k s^k > 0$ and $x^k \in \mathbb{R}_{++}^n$ implies $s^k \in \mathbb{R}_{++}^n$. Additionally, from the triangle inequality, we have

$$\|\mathbf{X}^k s^k\|_\infty \leq \|H(x^k)^{-1/2}s^k - \mu\mathbf{1}_n\|_\infty + \mu \leq \frac{2\varepsilon}{n}. \tag{39}$$

By Remark 5, these results are achieved after $O\left(\frac{Mn^2(f(x^0) - f_{\min}(X))}{\varepsilon^2}\right)$ iterations of FAHBA. This bound is by the factor $\frac{1}{n}$ sharper than the complementarity measure employed in [58] who have just ε in the r.h.s. of relation (39), although they work under an assumption similar to our Assumption 3 and the additional assumption that the level sets of f are bounded. Conversely, in order to attain an approximate KKT point with the same strength as in [58], the above calculations suggest that we can weaken our tolerance from ε to $\varepsilon \cdot n$, which results in an overall iteration complexity of $O\left(\frac{M(f(x^0) - f_{\min}(X))}{\varepsilon^2}\right)$, and a complementarity measure $\|\mathbf{X}^k s^k\|_\infty \leq 2\varepsilon$. Thus, in the particular case of non-negativity constraints our general algorithm is able to obtain results similar to [58], but under weaker assumptions. At the same time, our algorithm ensures a stronger measure of complementarity. Indeed, our algorithm guarantees that $x^k \in K_{\text{NN}}$, $s^k = \nabla f(x^k) - \mathbf{A}^*y^k \in K_{\text{NN}}$, i.e., $x^k, s^k > 0$, and approximate complementary $0 \leq \sum_{i=1}^n |x_i^k s_i^k| = \sum_{i=1}^n x_i^k s_i^k \leq 2\varepsilon$ after $O\left(\frac{Mn^2(f(x^0) - f_{\min}(X))}{\varepsilon^2}\right)$ iterations, which is stronger than $\max_{1 \leq i \leq n} |x_i^k s_i^k| \leq 2\varepsilon$ guaranteed by [58]. Indeed, $\max_{1 \leq i \leq n} |x_i^k s_i^k| \leq \sum_{i=1}^n |x_i^k s_i^k| \leq n \max_{1 \leq i \leq n} |x_i^k s_i^k|$, and both equalities are achieved

able. Moreover, to match our stronger guarantee, one has to change $\varepsilon \rightarrow \varepsilon/n$ in the complexity bound of [58], which leads to the same $O\left(\frac{Mn^2(f(x^0)-f_{\min}(X))}{\varepsilon^2}\right)$ complexity bound. Besides these important insights, our algorithm is designed for general cones, rather than only for \tilde{K}_{NN} . For more general cones we can not use the identity $[H(x)]^{-\frac{1}{2}} = \mathbf{X}$, which was very helpful for the derivation of (39) and for the analysis in [58]. Thus, for general, potentially non-symmetric, cones we have to exploit suitable properties of the barrier class $\mathcal{H}_\theta(K)$ and develop a new analysis technique. Finally, our method does not rely on the trust-region techniques as in [58] that may slow down the convergence in practice since the radius of the trust region is no greater than $O(\varepsilon)$ leading to short steps.

4.4.2 Exploiting the structure of symmetric cones

In (33) we can clearly observe the benefit of the use of θ -SSB in our algorithm, whenever K is a symmetric cone. Indeed, when $\alpha_k = \frac{1}{2\zeta(x^k, v^k)}$, the per-iteration decrease of the potential is $\frac{\|v^k\|_{x^k}^2}{4\zeta(x^k, v^k)} \geq \frac{\varepsilon\|v^k\|_{x^k}}{4\theta\zeta(x^k, v^k)}$ which may be large if $\zeta(x^k, v^k) = \sigma_{x^k}(-v^k) \ll \|v^k\|_{x^k}$.

4.4.3 The role of the penalty parameter

Next, we discuss more explicitly, how the algorithm and complexity bounds depend on the parameter μ . The first observation is that from (37), to guarantee that $s^k \in K^*$, we need the stopping criterion to be $\|v^k\|_{x^k} < \mu$, which by (38) leads to the error $2\mu\theta$ in the complementarity conditions. From the analysis following equation (34), we have that

$$K \frac{\mu^2}{4(\bar{M} + \mu)} = K \min \left\{ \frac{\mu}{4}, \frac{\mu^2}{4(\bar{M} + \mu)} \right\} \leq f(x^0) - f_{\min}(X) + \mu\theta.$$

Whence, recalling that $\bar{M} = \max\{M, L_0\}$,

$$K \leq 4(f(x^0) - f_{\min}(X) + \mu\theta) \cdot \frac{\max\{M, L_0\} + \mu}{\mu^2}.$$

Thus, we see that after $O(\mu^{-2})$ iterations the algorithm finds a $(2\mu\theta)$ -KKT point, and if $\mu \rightarrow 0$, we have convergence to a KKT point, but the complexity bound tends to infinity and becomes non-informative. At the same time, as it is seen from (28), when $\mu \rightarrow 0$, the algorithm itself can be interpreted as a preconditioned gradient method with the local preconditioner given by the Hessian $H(x)$.

We also see from the above explicit expressions in terms of μ that the design of the algorithm requires careful balance between the desired accuracy of the approximate KKT point expressed mainly by the complementarity condition, stopping criterion, and complexity. Moreover, the step-size should be also taken carefully to ensure the feasibility of the iterates, and the standard for first-order methods step-size $1/M$ may not work.

4.5 Anytime convergence via restarting FAHBA

The analysis of Algorithm 1 is based on the a-priori fixed tolerance $\varepsilon > 0$ and the parameter coupling $\mu = \varepsilon/\theta$. This coupling allows us to embed Algorithm 1 within a restarting scheme featuring a decreasing sequence $\{\mu_i\}_{i \geq 0}$, combined with restarts of FAHBA for each particular μ_i . This restarting strategy frees Algorithm 1 from hard-coded parameters and connects it well to traditional barrier methods. Moreover, the complexity of such path-following method is the same as for FAHBA, up to a constant factor.

To describe this double-loop algorithm, we fix $\varepsilon_0 > 0$ and select the starting point x_0^0 as a θ -analytic center of X with respect to $h \in \mathcal{H}_\theta(K)$. We let $i \geq 0$ denote the counter for the restarting epochs at the start of which the value μ_i is decreased. In epoch i , we generate a sequence $\{x_i^k\}_{k=0}^{K_i}$ by calling FAHBA($\mu_i, \varepsilon_i, L_0^{(i)}, x_i^0$) until the stopping condition is reached. This will take at most $\mathbb{K}_I(\varepsilon_i, x_i^0)$ iterations, specified in Eq. (31). We store the last iterate $\hat{x}_i = x_i^{K_i}$ and the last estimate of the Lipschitz constant $\hat{M}_i = L_{K_i}^{(i)}$ obtained from procedure FAHBA($\mu_i, \varepsilon_i, L_0^{(i)}, x_i^0$) and then restart the algorithm using the “warm start” $x_{i+1}^0 = \hat{x}_i, L_0^{(i+1)} = \hat{M}_i/2, \varepsilon_{i+1} = \varepsilon_i/2, \mu_{i+1} = \varepsilon_{i+1}/\theta$. If $\varepsilon \in (0, \varepsilon_0)$ is the target accuracy of the final solution, it suffices to perform $\lceil \log_2(\varepsilon_0/\varepsilon) \rceil + 1$ restarts since, by construction, $\varepsilon_i = \varepsilon_0 \cdot 2^{-i}$.

Algorithm 2: Restarting FAHBA

Data: $h \in \mathcal{H}_\theta(K), \varepsilon_0 > 0, x_0^0 \in X$ satisfying (21), $L_0^{(0)} > 0$.
Result: Point \hat{x}_i , dual variables $\hat{y}_i, \hat{s}_i = \nabla f(\hat{x}_i) - \mathbf{A}^* \hat{y}_i$.
for $i = 0, 1, \dots$ **do**
 Set $\varepsilon_i = 2^{-i} \varepsilon_0, \mu_i = \frac{\varepsilon_i}{\theta}$;
 Obtain $(\hat{x}_i, \hat{y}_i, \hat{s}_i, \hat{M}_i)$ from FAHBA($\mu_i, \varepsilon_i, L_0^{(i)}, x_i^0$);
 Set $x_{i+1}^0 = \hat{x}_i$ and $L_0^{(i+1)} = \hat{M}_i/2$.
end

Theorem 2 *Let Assumptions 1–3 hold. Then, for any $\varepsilon \in (0, \varepsilon_0)$, Algorithm 2 finds a 2ε -KKT point for problem (Opt) in the sense of Definition 4 after no more than $I(\varepsilon) = \lceil \log_2(\varepsilon_0/\varepsilon) \rceil + 1$ restarts and at most*

$$\left\lceil \frac{16}{3\varepsilon^2} (f(x^0) - f_{\min}(X) + \varepsilon_0)\theta^2 (\max\{M, L_0^{(0)}\} + \varepsilon_0/\theta) \right\rceil$$

iterations of FAHBA.

Proof Let us consider a restart $i \geq 0$ and repeat the proof of Theorem 1 with the change $\varepsilon \rightarrow \varepsilon_i, \mu \rightarrow \mu_i = \varepsilon_i/\theta, L_0 \rightarrow L_0^{(i)} = \hat{M}_{i-1}/2, \bar{M} = \max\{M, L_0\} \rightarrow \bar{M}_i = \max\{M, L_0^{(i)}\}, x^0 \rightarrow x_i^0 = \hat{x}_{i-1}$. Let K_i be the last iteration of FAHBA($\mu_i, \varepsilon_i, L_0^{(i)}, x_i^0$) meaning that $\|v^{K_i}\|_{x^{K_i}} < \frac{\varepsilon_i}{\theta}, \|v^{K_i-1}\|_{x^{K_i-1}} \geq \frac{\varepsilon_i}{\theta}$,

and the stopping criterion does not hold for K_i iterations $k = 0, \dots, K_i - 1$. From the analysis following Eq. (34), we have that

$$K_i \frac{\varepsilon_i^2}{4\theta^2(\bar{M}_i + \varepsilon_i/\theta)} \leq K_i \min_{k=0, \dots, K_i-1} \delta_k^i \leq \sum_{k=0}^{K_i-1} \delta_k^i \leq F_{\mu_i}(x_i^0) - F_{\mu_i}(x_i^{K_i}). \quad (40)$$

Further, using the fact that μ_i is a decreasing sequence and (21), it is easy to deduce

$$\begin{aligned} F_{\mu_{i+1}}(x_{i+1}^0) &= F_{\mu_{i+1}}(x_i^{K_i}) \stackrel{(2)}{=} f(x_i^{K_i}) + \mu_{i+1}h(x_i^{K_i}) \stackrel{(2)}{=} F_{\mu_i}(x_i^{K_i}) \\ &\quad + (\mu_{i+1} - \mu_i)h(x_i^{K_i}) \\ &\stackrel{(21)}{\leq} F_{\mu_i}(x_i^{K_i}) + (\mu_{i+1} - \mu_i)(h(x_0^0) - \theta) \\ &\stackrel{(40)}{\leq} F_{\mu_i}(x_i^0) - K_i \frac{\varepsilon_i^2}{4\theta^2(\bar{M}_i + \varepsilon_i/\theta)} + (\mu_{i+1} - \mu_i)(h(x_0^0) - \theta). \end{aligned} \quad (41)$$

Letting $I \equiv I(\varepsilon) = \lceil \log_2(\frac{\varepsilon_0}{\varepsilon}) \rceil + 1$, by Theorem 1 applied to the restart $I - 1$, we see that **FAHBA**($\mu_{I-1}, \varepsilon_{I-1}, L_0^{(I-1)}, x_{I-1}^0$) outputs a 2ε -KKT point for problem (Opt) in the sense of Definition 4. Summing inequalities (41) for all the performed restarts $i = 0, \dots, I - 1$ and rearranging the terms, we obtain

$$\begin{aligned} \sum_{i=0}^{I-1} K_i \frac{\varepsilon_i^2}{4\theta^2(\bar{M}_i + \varepsilon_i/\theta)} &\leq F_{\mu_0}(x_0^0) - F_{\mu_I}(x_I^0) + (\mu_I - \mu_0)(h(x_0^0) - \theta) \\ &\stackrel{(2)}{=} f(x_0^0) + \mu_0h(x_0^0) - f(x_I^0) - \mu_Ih(x_I^0) + (\mu_I - \mu_0)(h(x_0^0) - \theta) \\ &\stackrel{(21)}{\leq} f(x_0^0) - f_{\min}(\mathbf{X}) + \mu_0h(x_0^0) - \mu_Ih(x_0^0) + \mu_I\theta + (\mu_I - \mu_0)(h(x_0^0) - \theta) \\ &\leq f(x_0^0) - f_{\min}(\mathbf{X}) + \mu_0\theta = f(x_0^0) - f_{\min}(\mathbf{X}) + \varepsilon_0. \end{aligned}$$

Moreover, based on our updating choice $L_0^{(i+1)} = \hat{M}_i/2$, it holds that

$$\begin{aligned} \bar{M}_i &= \max\{M, L_0^{(i)}\} = \max\{M, \hat{M}_{i-1}/2\} \\ &= \max\{M, L_{K_{i-1}}^{(i-1)}/2\} \leq \max\{M, \bar{M}_{i-1}\} \leq \dots \leq \max\{M, \bar{M}_0\} \leq \max\{M, L_0^{(0)}\}. \end{aligned}$$

Hence,

$$K_i \leq 4(f(x^0) - f_{\min}(\mathbf{X}) + \varepsilon_0) \cdot \frac{\theta^2(\bar{M}_i + \varepsilon_i/\theta)}{\varepsilon_i^2} \leq \frac{C}{\varepsilon_i^2},$$

where $C \equiv 4(f(x^0) - f_{\min}(\mathbf{X}) + \varepsilon_0)\theta^2(\max\{M, L_0^{(0)}\} + \varepsilon_0/\theta)$. Finally, we obtain that the total number of iterations of procedures **FAHBA**($\mu_i, \varepsilon_i, L_0^{(i)}, x_i^0$), $0 \leq i \leq I - 1$,

to reach accuracy ε is at most

$$\begin{aligned} \sum_{i=0}^{l-1} K_i &\leq \sum_{i=0}^{l-1} \frac{C}{\varepsilon_i^2} \leq \frac{C}{\varepsilon_0^2} \sum_{i=0}^{l-1} (2^2)^i \leq \frac{C}{3\varepsilon_0^2} \cdot \left(4^{1+\log_2(\frac{\varepsilon_0}{\varepsilon})}\right) = \frac{4C}{3\varepsilon^2} \\ &= \frac{16(f(x^0) - f_{\min}(X) + \varepsilon_0)\theta^2(\max\{M, L_0^{(0)}\} + \varepsilon_0/\theta)}{3\varepsilon^2}. \end{aligned}$$

□

5 A second-order Hessian barrier algorithm

We now present a second-order potential reduction method for (Opt) under the assumption that the second-order Taylor expansion of f on the set of feasible directions \mathcal{T}_x defined in (16) is sufficiently accurate in the geometry induced by $h \in \mathcal{H}_\theta(K)$.

Assumption 4 (Local second-order smoothness) $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{+\infty\}$ is twice continuously differentiable on X and there exists a constant $M > 0$ such that, for all $x \in X$ and $v \in \mathcal{T}_x$, where \mathcal{T}_x is defined in (16), we have

$$\|\nabla f(x + v) - \nabla f(x) - \nabla^2 f(x)v\|_x^* \leq \frac{M}{2} \|v\|_x^2. \tag{42}$$

A sufficient condition for (42) is the following local counterpart of the global Lipschitz condition on the Hessian of f :

$$(\forall x \in X)(\forall u, v \in \mathcal{T}_x) : \|\nabla^2 f(x + u) - \nabla^2 f(x + v)\|_{\text{op},x} \leq M\|u - v\|_x,$$

where $\|\mathbf{B}\|_{\text{op},x} \triangleq \sup_{\|u\|_x \leq 1} \left\{ \frac{\|\mathbf{B}u\|_x^*}{\|u\|_x} \right\}$ is the induced operator norm for a linear operator $\mathbf{B} : \mathbb{E} \rightarrow \mathbb{E}^*$. Indeed, this condition implies (42):

$$\begin{aligned} \|\nabla f(x + v) - \nabla f(x) - \nabla^2 f(x)v\|_x^* &= \left\| \int_0^1 (\nabla^2 f(x + tv) - \nabla^2 f(x))v \, dt \right\|_x^* \\ &\leq \int_0^1 \|\nabla^2 f(x + tv) - \nabla^2 f(x)\|_{\text{op},x} \cdot \|v\|_x \, dt \leq \frac{M}{2} \|v\|_x^2. \end{aligned}$$

Further, Eq. (42) implies

$$f(x + v) - \left[f(x) + \langle \nabla f(x), v \rangle + \frac{1}{2} \langle \nabla^2 f(x)v, v \rangle \right] \leq \frac{M}{6} \|v\|_x^3. \tag{43}$$

To obtain this, observe that for all $x \in X$ and $v \in \mathcal{T}_x$, we have

$$\begin{aligned} & |f(x + v) - f(x) - \langle \nabla f(x), v \rangle - \frac{1}{2} \langle \nabla^2 f(x)v, v \rangle| \\ &= \left| \int_0^1 \langle \nabla f(x + tv) - \nabla f(x) - \frac{1}{2} \nabla^2 f(x)v, v \rangle dt \right| \\ &\leq \int_0^1 \|\nabla f(x + tv) - \nabla f(x) - \frac{1}{2} \nabla^2 f(x)v\|_x^* dt \cdot \|v\|_x \leq \frac{M}{6} \|v\|_x^3. \end{aligned}$$

Remark 7 Assumption 4 subsumes, when \bar{X} is bounded, the standard Lipschitz-Hessian setting. If the Hessian of f is Lipschitz with modulus M with respect to the standard Euclidean norm, we have by [76, Lemma 1.2.4] that

$$\|\nabla f(x + v) - \nabla f(x) - \nabla^2 f(x)v\| \leq \frac{M}{2} \|v\|^2.$$

Since \bar{X} is bounded, one can observe that $\lambda_{\max}([H(x)]^{-1})^{-1} = \lambda_{\min}(H(x)) \geq \sigma$ for some $\sigma > 0$, and (42) holds. Indeed, denoting $g = \nabla f(x + v) - \nabla f(x) - \nabla^2 f(x)v$, we obtain

$$(\|g\|_x^*)^2 \leq \lambda_{\max}([H(x)]^{-1}) \|g\|^2 \leq \frac{M^2}{4\lambda_{\min}(H(x))} \|v\|^4 \leq \frac{M^2}{4\sigma^3} \|v\|_x^4.$$

◇

Remark 8 The cubic overestimation of the objective function in (43) does not rely on global second order differentiability assumptions. To illustrate this, we revisit the structured composite optimization problem (1), assuming that the data fidelity function ℓ is twice continuously differentiable on an open neighborhood containing X , with Lipschitz continuous Hessian $\nabla^2 \ell$ with modulus γ w.r.t. the Euclidean norm. On the domain K_{NN} we employ the canonical barrier $h(x) = -\sum_{i=1}^n \ln(x_i)$, having $H(x) = \text{diag}[x_1^{-2}, \dots, x_n^{-2}] = X^{-2}$. This means, for all $x, x^+ \in X$, we have

$$\ell(x^+) \leq \ell(x) + \langle \nabla \ell(x), x^+ - x \rangle + \frac{1}{2} \langle \nabla^2 \ell(x)(x^+ - x), x^+ - x \rangle + \frac{\gamma}{6} \|x^+ - x\|^3.$$

As penalty function, we again consider the L_p -regularizer with $p \in (0, 1)$. For any $t, s > 0$, one has

$$t^p \leq s^p + ps^{p-1}(t - s) + \frac{p(p - 1)}{2} s^{p-2}(t - s)^2 + \frac{p(p - 1)(p - 2)}{6} s^{p-3}(t - s)^3.$$

Further, $v \in \mathcal{T}_x$ if and only if $v = [H(x)]^{-1/2}d = Xd$ for some $d \in \mathbb{R}^n$ satisfying $AXd = 0$ and $\|d\| < 1$. Since $p(1 - p) \leq 1/4$, it follows that $p(1 - p)(2 - p) \leq 1/2$. Thus, using $x^+ = x + v = x + Xd$, we get

$$\begin{aligned}
 f(x^+) - \left((f(x) + \langle \nabla f(x), \mathbf{X}d \rangle + \frac{1}{2} \langle \nabla^2 f(x) \mathbf{X}d, \mathbf{X}d \rangle) \right) &\leq \frac{\gamma}{6} \|\mathbf{X}d\|^3 + \frac{1}{12} \sum_{i=1}^n x_i^p d_i^3 \\
 &\leq \frac{\gamma}{6} \|\mathbf{X}d\|^3 + \frac{1}{12} \|x\|_\infty^p \sum_{i=1}^n d_i^3 \leq \frac{1}{6} \left(\gamma \|x\|_\infty^3 + \frac{1}{2} \|x\|_\infty^p \right) \|d\|^3.
 \end{aligned}$$

Since we assume that \bar{X} is bounded, there exists a universal constant $M > 0$ such that $\gamma \|x\|_\infty^2 + \frac{1}{2} \|x\|_\infty^p \leq M$. Combining this with Remark 7, we obtain a cubic overestimation as in Eq. (43). Importantly, $f(x)$ is not differentiable for $x \in \text{bd}(\mathbb{K}_{\text{NN}}) = \{x \in \mathbb{R}_+^n \mid x_i = 0 \text{ for some } i\}$. \diamond

We emphasize that in Assumption 4, the Lipschitz modulus M is hardly known exactly in practice, and it is also not an easy task to obtain universal upper bounds that can be used in implementations. Therefore, adaptive techniques should be used to estimate it and are likely to improve the practical performance of the method.

Assumption 4 also implies, via (43) and (7) (with $d = v$ and $t = 1 < \frac{1}{\|v\|_x} \stackrel{(5)}{\leq} \frac{1}{\zeta(x, v)}$), the following upper bound for the potential function F_μ .

Lemma 3 (Cubic overestimation) *For all $x \in X, v \in \mathcal{X}_x$ and $L \geq M$, we have*

$$F_\mu(x + v) \leq F_\mu(x) + \langle \nabla F_\mu(x), v \rangle + \frac{1}{2} \langle \nabla^2 f(x)v, v \rangle + \frac{L}{6} \|v\|_x^3 + \mu \|v\|_x^2 \omega(\zeta(x, v)).$$

5.1 Algorithm description and its complexity

5.1.1 Defining the step direction

Let $x \in X$ be given. In order to find a search direction, we choose a parameter $L > 0$, construct a cubic-regularized model of the potential F_μ (2), and minimize it on the linear subspace L_0 :

$$v_{\mu,L}(x) \in \underset{v \in \mathbb{E}: \mathbf{A}v=0}{\text{Argmin}} \left\{ Q_{\mu,L}^{(2)}(x, v) \triangleq F_\mu(x) + \langle \nabla F_\mu(x), v \rangle + \frac{1}{2} \langle \nabla^2 f(x)v, v \rangle + \frac{L}{6} \|v\|_x^3 \right\}, \tag{44}$$

where by Argmin we denote the set of global minimizers. The model consists of three parts: linear approximation of h , quadratic approximation of f , and a cubic regularizer with penalty parameter $L > 0$. Since this model and our algorithm use the second derivative of f , we call it a second-order method. Our further derivations rely on the first-order optimality conditions for the problem (44), which say that there exists $y_{\mu,L}(x) \in \mathbb{R}^m$ such that $v_{\mu,L}(x)$ satisfies

$$\nabla F_\mu(x) + \nabla^2 f(x)v_{\mu,L}(x) + \frac{L}{2} \|v_{\mu,L}(x)\|_x H(x)v_{\mu,L}(x) - \mathbf{A}^* y_{\mu,L}(x) = 0, \tag{45}$$

$$-\mathbf{A}v_{\mu,L}(x) = 0. \tag{46}$$

We also use the following extension of [79, Prop. 1] with the local norm induced by $H(x)$.

Proposition 5 *For all $x \in X$ it holds*

$$\nabla^2 f(x) + \frac{L}{2} \|v_{\mu,L}(x)\|_x H(x) \geq 0 \quad \text{on } L_0. \tag{47}$$

Proof The proof follows the same strategy as Lemma 3.2 in [29]. Let $\{z_1, \dots, z_p\}$ be an orthonormal basis of L_0 and the linear operator $\mathbf{Z} : \mathbb{R}^p \rightarrow L_0$ be defined by $\mathbf{Z}w = \sum_{i=1}^p z_i w^i$ for all $w = [w^1; \dots; w^p]^\top \in \mathbb{R}^p$. With the help of this linear map, we can absorb the nullspace restriction, and formulate the search-direction finding problem (44) using the projected data

$$\mathbf{g} \triangleq \mathbf{Z}^* \nabla F_\mu(x), \quad \mathbf{J} \triangleq \mathbf{Z}^* \nabla^2 f(x) \mathbf{Z}, \quad \mathbf{H} \triangleq \mathbf{Z}^* H(x) \mathbf{Z} \succ 0. \tag{48}$$

We then arrive at the cubic-regularized subproblem to find $u_L \in \mathbb{R}^p$ s.t.

$$u_L \in \underset{u \in \mathbb{R}^p}{\text{Argmin}} \left\{ \langle \mathbf{g}, u \rangle + \frac{1}{2} \langle \mathbf{J}u, u \rangle + \frac{L}{6} \|u\|_{\mathbf{H}}^3 \right\}, \tag{49}$$

where $\|\cdot\|_{\mathbf{H}}$ is the norm induced by the operator \mathbf{H} . From [79, Thm. 10] we deduce

$$\mathbf{J} + \frac{L\|u_L\|_{\mathbf{H}}}{2} \mathbf{H} \geq 0.$$

Denoting $v_{\mu,L}(x) = \mathbf{Z}u_L$, we see

$$\|u_L\|_{\mathbf{H}} = \langle \mathbf{Z}^* H(x) \mathbf{Z}u_L, u_L \rangle^{1/2} = \langle H(x)(\mathbf{Z}u_L), \mathbf{Z}u_L \rangle^{1/2} = \|v_{\mu,L}(x)\|_x, \text{ and}$$

$$\mathbf{Z}^* \left(\nabla^2 f(x) + \frac{L}{2} \|v_{\mu,L}(x)\|_x H(x) \right) \mathbf{Z} \geq 0,$$

which implies $\nabla^2 f(x) + \frac{L}{2} \|v_{\mu,L}(x)\|_x H(x) \succ 0$ over the nullspace $L_0 = \{v \in \mathbb{E} : \mathbf{A}v = 0\}$. □

The above proposition gives some ideas on how one could numerically solve problem (44) in practice. In a preprocessing step, we once calculate matrix \mathbf{Z} and use it during the whole algorithm execution. At each iteration we calculate the new data using (48), leaving us with a standard *unconstrained* cubic subproblem (49). Nesterov and Polyak [79] show how such problems can be transformed to a *convex* problem to which fast convex programming methods could in principle be applied. However, we can also solve it via recent efficient methods based on Lanczos' method [29, 68]. Whatever numerical tool is employed, we can recover our search direction $v_{\mu,L}(x)$ by the matrix vector product $\mathbf{Z}u_L$ in which u_L denotes the solution obtained from this subroutine.

5.1.2 Defining the step-size

Our next goal is to construct an admissible step-size policy, given the search direction $v_{\mu,L}(x)$. Let $x \in X$ be the current position of the algorithm. Define the parameterized family of arcs $x^+(t) \triangleq x + tv_{\mu,L}(x)$, where $t \geq 0$ is a step-size. By (6) and since $v_{\mu,L}(x) \in L_0$ by (46), we know that $x^+(t)$ is in X provided that $t \in I_{x,\mu,L} \triangleq [0, \frac{1}{\zeta(x,v_{\mu,L}(x))})$. For all such t , Lemma 3 yields

$$F_{\mu}(x^+(t)) \leq F_{\mu}(x) + t\langle \nabla F_{\mu}(x), v_{\mu,L}(x) \rangle + \frac{t^2}{2}\langle \nabla^2 f(x)v_{\mu,L}(x), v_{\mu,L}(x) \rangle + \frac{Mt^3}{6}\|v_{\mu,L}(x)\|_x^3 + \mu t^2\omega(t\zeta(x, v_{\mu,L}(x))). \tag{50}$$

Since $v_{\mu,L}(x) \in L_0$, multiplying (47) with $v_{\mu,L}(x)$ from the left and the right, and multiplying (45) by $v_{\mu,L}(x)$ and combining with (46), we obtain

$$\langle \nabla^2 f(x)v_{\mu,L}(x), v_{\mu,L}(x) \rangle \geq -\frac{L}{2}\|v_{\mu,L}(x)\|_x^3, \tag{51}$$

$$\langle \nabla F_{\mu}(x), v_{\mu,L}(x) \rangle + \langle \nabla^2 f(x)v_{\mu,L}(x), v_{\mu,L}(x) \rangle + \frac{L}{2}\|v_{\mu,L}(x)\|_x^3 = 0. \tag{52}$$

Under the additional assumption that $t \leq 2$ and $L \geq M$, we obtain

$$\begin{aligned} & t\langle \nabla F_{\mu}(x), v_{\mu,L}(x) \rangle + \frac{t^2}{2}\langle \nabla^2 f(x)v_{\mu,L}(x), v_{\mu,L}(x) \rangle + \frac{Mt^3}{6}\|v_{\mu,L}(x)\|_x^3 \\ & \stackrel{(52)}{=} -t\left(\langle \nabla^2 f(x)v_{\mu,L}(x), v_{\mu,L}(x) \rangle + \frac{L}{2}\|v_{\mu,L}(x)\|_x^3\right) \\ & \quad + \frac{t^2}{2}\langle \nabla^2 f(x)v_{\mu,L}(x), v_{\mu,L}(x) \rangle + \frac{Mt^3}{6}\|v_{\mu,L}(x)\|_x^3 \\ & = \left(\frac{t^2}{2} - t\right)\langle \nabla^2 f(x)v_{\mu,L}(x), v_{\mu,L}(x) \rangle - \frac{Lt}{2}\|v_{\mu,L}(x)\|_x^3 + \frac{Mt^3}{6}\|v_{\mu,L}(x)\|_x^3 \\ & \stackrel{(51), t \leq 2}{\leq} \left(\frac{t^2}{2} - t\right)\left(-\frac{L}{2}\|v_{\mu,L}(x)\|_x^3\right) - \frac{Lt}{2}\|v_{\mu,L}(x)\|_x^3 + \frac{Mt^3}{6}\|v_{\mu,L}(x)\|_x^3 \\ & = -\|v_{\mu,L}(x)\|_x^3\left(\frac{Lt^2}{4} - \frac{Mt^3}{6}\right) \stackrel{L \geq M}{\leq} -\|v_{\mu,L}(x)\|_x^3\frac{Lt^2}{12}(3 - 2t). \end{aligned}$$

Substituting this into (50), we arrive at

$$F_{\mu}(x^+(t)) \leq F_{\mu}(x) - \|v_{\mu,L}(x)\|_x^3\frac{Lt^2}{12}(3 - 2t) + \mu t^2\omega(t\zeta(x, v_{\mu,L}(x))) \stackrel{(3)}{\leq} F_{\mu}(x) - \|v_{\mu,L}(x)\|_x^3\frac{Lt^2}{12}(3 - 2t) + \mu\frac{t^2\|v_{\mu,L}(x)\|_x^2}{2(1 - t\zeta(x, v_{\mu,L}(x)))}.$$

for all $t \in I_{x,\mu,L}$. Therefore, if $t\zeta(x, v_{\mu,L}(x)) \leq 1/2$, we readily see

$$\begin{aligned} F_\mu(x^+(t)) - F_\mu(x) &\leq -\frac{Lt^2\|v_{\mu,L}(x)\|_x^3}{12} (3 - 2t) + \mu t^2\|v_{\mu,L}(x)\|_x^2 \\ &= -\|v_{\mu,L}(x)\|_x^3 \frac{Lt^2}{12} \left(3 - 2t - \frac{12\mu}{L\|v_{\mu,L}(x)\|_x} \right) \triangleq -\eta_x(t). \end{aligned} \tag{53}$$

Maximizing the function $\eta_x(t)$ explicitly and carrying the resulting minimizer t^* through the analysis by finding an upper bound for the corresponding per-iteration decrease $-\eta_x(t^*)$ is technically quite challenging. Instead, we adopt the following step-size rule

$$\tau_{\mu,L}(x) \triangleq \frac{1}{\max\{1, 2\zeta(x, v_{\mu,L}(x))\}} = \min \left\{ 1, \frac{1}{2\zeta(x, v_{\mu,L}(x))} \right\}. \tag{54}$$

Note that $\tau_{\mu,L}(x) \leq 1$ and $\tau_{\mu,L}(x)\zeta(x, v_{\mu,L}(x)) \leq 1/2$. Thus, this choice of the step-size is feasible to derive (53).

5.1.3 Backtracking on the Lipschitz modulus

Just like Algorithm 1, our second-order method employs a line-search procedure to estimate the Lipschitz constant M in (42), (43) in the spirit of [27, 79]. More specifically, suppose that $x^k \in X$ is the current position of the algorithm with the corresponding initial local Lipschitz estimate M_k . To determine the next iterate x^{k+1} , we solve problem (44) with $L = L_k = 2^{i_k} M_k$ starting with $i_k = 0$, find the corresponding search direction $v^k = v_{\mu,L_k}(x^k)$ and the new point $x^{k+1} = x^k + \tau_{\mu,L_k}(x^k)v^k$. Then, we check whether the inequalities (42) and (43) hold with $M = L_k$, $x = x^k$, $v = \tau_{\mu,L_k}(x^k)v^k$, see (58) and (57). If they hold, we make a step to x^{k+1} . Otherwise, we increase i_k by 1 and repeat the procedure. Obviously, when $L_k = 2^{i_k} M_k \geq M$, both inequalities (42) and (43) with M changed to L_k , i.e., (58) and (57), are satisfied and the line-search procedure ends. For the next iteration we set $M_{k+1} = \max\{2^{i_k-1} M_k, \underline{L}\} = \max\{L_k/2, \underline{L}\}$, so that the estimate for the local Lipschitz constant on the one hand can decrease allowing larger step-sizes, and on the other hand is bounded from below.

5.1.4 Second-order algorithm and its complexity result

The resulting procedure gives rise to a Second-order Adaptive Hessian Barrier Algorithm (**SAHBA**, Algorithm 3).

Our main result on the iteration complexity of Algorithm 3 is the following Theorem, whose proof is given in Sect. 5.2.

Theorem 3 *Let Assumptions 1, 2, and 4 hold. Fix the error tolerance $\varepsilon > 0$, the regularization parameter $\mu = \frac{\varepsilon}{4\theta}$, and some initial guess $M_0 > 144\varepsilon$ for the Lipschitz constant in (42). Let $(x^k)_{k \geq 0}$ be the trajectory generated by **SAHBA**($\mu, \varepsilon, M_0, x^0$),*

Algorithm 3: Second-order Adaptive Hessian Barrier Algorithm - **SAHBA**($\mu, \varepsilon, M_0, x^0$)

Data: $h \in \mathcal{H}_\theta(\mathbb{K}), \mu > 0, \varepsilon > 0, M_0 \geq 144\varepsilon, x^0 \in X$.

Result: $(x^k, y^{k-1}, s^k, M_k) \in X \times \mathbb{R}^m \times \mathbb{K}^* \times \mathbb{R}_+$, where $s^k = \nabla f(x^k) - A^*y^{k-1}$, and M_k is the last estimate of the Lipschitz constant in (42).

Set $\underline{L} \triangleq 144\varepsilon, k = 0$;

repeat

Set $i_k = 0$. **repeat**

Set $L_k = 2^{i_k} M_k$. Find $v^k \triangleq v_{\mu, L_k}(x^k)$ and $y^k \triangleq y_{\mu, L_k}(x^k)$ as a global solution to

$$\min_{v: Av=0} \left\{ F_{\mu}(x^k) + \langle \nabla F_{\mu}(x^k), v \rangle + \frac{1}{2} \langle \nabla^2 f(x^k)v, v \rangle + \frac{L_k}{6} \|v\|_{x^k}^3 \right\}, \quad (55)$$

$$\text{Set } \alpha_k \triangleq \min \left\{ 1, \frac{1}{2\zeta(x^k, v^k)} \right\}, \text{ where } \zeta(\cdot, \cdot) \text{ as in (4)}. \quad (56)$$

Set $z^k = x^k + \alpha_k v^k, i_k = i_k + 1$;

until

$$f(z^k) \leq f(x^k) + \langle \nabla f(x^k), z^k - x^k \rangle + \frac{1}{2} \langle \nabla^2 f(x^k)(z^k - x^k), z^k - x^k \rangle + \frac{L_k}{6} \|z^k - x^k\|_{x^k}^3, \text{ and} \quad (57)$$

$$\|\nabla f(z^k) - \nabla f(x^k) - \nabla^2 f(x^k)(z^k - x^k)\|_{x^k}^* \leq \frac{L_k}{2} \|z^k - x^k\|_{x^k}^2. \quad (58)$$

;

Set $M_{k+1} = \max\{\frac{L_k}{2}, \underline{L}\}, x^{k+1} = z^k, k = k + 1$

until $\|v^{k-1}\|_{x^{k-1}} < \Delta_{k-1} \triangleq \sqrt{\frac{\varepsilon}{4L_{k-1}\theta}}$ and $\|v^k\|_{x^k} < \Delta_k \triangleq \sqrt{\frac{\varepsilon}{4L_k\theta}}$;

where x^0 is a 4θ -analytic center satisfying (21). Then the algorithm stops in no more than

$$\mathbb{K}_{II}(\varepsilon, x^0) = \left\lceil \frac{192\theta^{3/2} \sqrt{2 \max\{M, M_0\}} (f(x^0) - f_{\min}(X) + \varepsilon)}{\varepsilon^{3/2}} \right\rceil \quad (59)$$

outer iterations, and the number of inner iterations is no more than $2(\mathbb{K}_{II}(\varepsilon, x^0) + 1) + 2 \max\{\log_2(2M/M_0), 1\}$. Moreover, the output of **SAHBA**($\mu, \varepsilon, M_0, x^0$) constitutes an $(\varepsilon, \frac{\max\{M, M_0\}\varepsilon}{8\theta})$ -2KKT point for problem (Opt) in the sense of Definition 5.

Remark 9 Since $f(x^0) - f_{\min}(X)$ is expected to be larger than ε , and the constant M is potentially large, we see that the main term in the complexity bound (59) is $O\left(\frac{\theta^{3/2} \sqrt{M(f(x^0) - f_{\min}(X))}}{\varepsilon^{3/2}}\right) = O\left(\left(\frac{\theta}{\varepsilon}\right)^{3/2}\right)$. Note that the complexity result $O(\max\{\varepsilon_1^{-3/2}, \varepsilon_2^{-3/2}\})$ reported in [24, 25] to find an $(\varepsilon_1, \varepsilon_2)$ -2KKT point for arbitrary $\varepsilon_1, \varepsilon_2 > 0$, is known to be optimal for unconstrained smooth non-convex optimization by second-order methods under the standard Lipschitz–Hessian assumption, subsumed

on bounded sets by our Assumption 4. A similar dependence on arbitrary $\varepsilon_1, \varepsilon_2 > 0$ can be easily obtained from our theorem by setting $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$. \diamond

5.2 Proof of Theorem 3

The main steps of the proof are similar to the analysis of Algorithm 1. We start by showing the feasibility of the iterates and correctness of the line-search process. Next, we analyze the per-iteration decrease of F_μ and f and show that if the stopping criterion does not hold at iteration k , then the objective function is decreased by the value $O(\varepsilon^{3/2})$. From this, since the objective is globally lower bounded, we conclude that the algorithm stops in $O(\varepsilon^{-3/2})$ iterations. Finally, we show that when the stopping criterion holds, the primal–dual pair (x^k, y^{k-1}) resulting from solving the cubic subproblem (55) yields a dual slack variable s^k such that this triple constitutes an approximate second-order KKT point.

5.2.1 Interior point property of the iterates

By construction $x^0 \in X$. Proceeding inductively, let $x^k \in X$ be the k -th iterate of the algorithm, with the search direction $v^k \triangleq v_{\mu, L}(x^k)$. By Eq. (56), the step-size α_k satisfies $\alpha_k \leq \frac{1}{2\zeta(x^k, v^k)}$. Consequently, $\alpha_k \zeta(x^k, v^k) \leq 1/2$ for all $k \geq 0$, and using (6) as well as $\mathbf{A}v^k = 0$ (see Eq. (55)), we have that $x^{k+1} = x^k + \alpha_k v^k \in X$. By induction, it follows that $x^k \in X$ for all $k \geq 0$.

5.2.2 Bounding the number of backtracking steps

To bound the number of cycles involved in the line-search process for finding appropriate constants L_k , we proceed as in Sect. 4.3.2. Let us fix an iteration k . The sequence $L_k = 2^{i_k} M_k$ is increasing as i_k is increasing, and Assumption 4 holds. This implies (43), and thus when $L_k = 2^{i_k} M_k \geq \max\{M, M_k\}$, the line-search process for sure stops since inequalities (57) and (58) hold. Hence, $L_k = 2^{i_k} M_k \leq 2 \max\{M, M_k\}$ must be the case, and, consequently, $M_{k+1} = \max\{L_k/2, \underline{L}\} \leq \max\{\max\{M, M_k\}, \underline{L}\} = \max\{M, M_k\}$, which, by induction, gives $M_k \leq \bar{M} \triangleq \max\{M, M_0\}$ and $L_k \leq 2\bar{M}$. At the same time, by construction, $M_{k+1} = \max\{2^{i_{k+1}-1} M_k, \underline{L}\} = \max\{L_k/2, \underline{L}\} \geq L_k/2$. Hence, $L_{k+1} = 2^{i_{k+1}} M_{k+1} \geq 2^{i_{k+1}-1} L_k$ and therefore $\log_2 \left(\frac{L_{k+1}}{L_k} \right) \geq i_{k+1} - 1, \forall k \geq 0$. At the same time, at iteration 0 we have $L_0 = 2^{i_0} M_0 \leq 2\bar{M}$, whence, $i_0 \leq \log_2 \left(\frac{2\bar{M}}{M_0} \right)$. Let $N(k)$ denote the number of inner line-search iterations up to iteration k of SAHBA. Then,

$$\begin{aligned} N(k) &= \sum_{j=0}^k (i_j + 1) \leq i_0 + 1 + \sum_{j=1}^k \left(\log_2 \left(\frac{L_j}{L_{j-1}} \right) + 2 \right) \\ &\leq 2(k + 1) + 2 \log_2 \left(\frac{2\bar{M}}{M_0} \right), \end{aligned}$$

since $L_k \leq 2\bar{M} = 2 \max\{M, M_0\}$ in the last step. Thus, on average, the inner loop ends after two trials.

5.2.3 Per-iteration analysis and a bound for the number of iterations

Let us fix iteration counter k . The main assumption of this subsection is that the stopping criterion is not satisfied, i.e., either $\|v^k\|_{x^k} \geq \Delta_k$ or $\|v^{k-1}\|_{x^{k-1}} \geq \Delta_{k-1}$. Without loss of generality, we assume that the first inequality holds, i.e., $\|v^k\|_{x^k} \geq \Delta_k$, and consider iteration k . Otherwise, if the second inequality holds, the same derivations can be made considering the iteration $k - 1$ and using the second inequality $\|v^{k-1}\|_{x^{k-1}} \geq \Delta_{k-1}$. Thus, at the end of the k -th iteration

$$\|v^k\|_{x^k} \geq \Delta_k = \sqrt{\frac{\varepsilon}{4L_k\theta}}. \tag{60}$$

Since the step-size $\alpha_k = \min\{1, \frac{1}{2\zeta(x^k, v^k)}\} = \tau_{\mu, L_k}(x^k)$ in (56) satisfies $\alpha_k \leq 1$ and $\alpha_k \zeta(x^k, v^k) \leq 1/2$ (cf. (54) and a remark after it), we can repeat the derivations of Sect. 5.1, changing (43) to (57). In this way we obtain the following counterpart of (53) with $t = \alpha_k$, $L = L_k$, $x = x^k$, $v_{\mu, L_k}(x^k) \triangleq v^k$:

$$\begin{aligned} F_\mu(x^{k+1}) - F_\mu(x^k) &\leq -\|v^k\|_{x^k}^3 \frac{L_k \alpha_k^2}{12} \left(3 - 2\alpha_k - \frac{12\mu}{L_k \|v^k\|_{x^k}}\right) \\ &\leq -\|v^k\|_{x^k}^3 \frac{L_k \alpha_k^2}{12} \left(1 - \frac{12\mu}{L_k \|v^k\|_{x^k}}\right), \end{aligned} \tag{61}$$

where in the last inequality we used that $\alpha_k \leq 1$ by construction. Substituting $\mu = \frac{\varepsilon}{4\theta}$, and using (60), we obtain

$$\begin{aligned} 1 - \frac{12\mu}{L_k \|v^k\|_{x^k}} &= 1 - \frac{12\varepsilon}{4\theta L_k \|v^k\|_{x^k}} \stackrel{(60)}{\geq} 1 - \frac{3\varepsilon}{\theta L_k \sqrt{\frac{\varepsilon}{4L_k\theta}}} \\ &= 1 - \frac{6\sqrt{\varepsilon}}{\sqrt{\theta} L_k} \geq 1 - \frac{6\sqrt{\varepsilon}}{\sqrt{144\theta\varepsilon}} \geq \frac{1}{2}, \end{aligned}$$

using that, by construction, $L_k = 2^i M_k \geq \underline{L} = 144\varepsilon$ and that $\theta \geq 1$. Hence, from (61),

$$F_\mu(x^{k+1}) - F_\mu(x^k) \leq -\|v^k\|_{x^k}^3 \frac{L_k \alpha_k^2}{24}. \tag{62}$$

Substituting into (62) the two possible values of the step-size α_k in (56) gives

$$F_\mu(x^{k+1}) - F_\mu(x^k) \leq \begin{cases} -\|v^k\|_{x^k}^3 \frac{L_k}{24}, & \text{if } \alpha_k = 1, \\ -\frac{L_k \|v^k\|_{x^k}^3}{96(\zeta(x^k, v^k))^2} \stackrel{(5)}{\leq} -\frac{L_k \|v^k\|_{x^k}}{96} & \text{if } \alpha_k = \frac{1}{2\zeta(x^k, v^k)}. \end{cases} \tag{63}$$

This implies

$$F_\mu(x^{k+1}) - F_\mu(x^k) \leq -\frac{L_k \|v^k\|_{x^k}}{96} \min \left\{ 1, 4\|v^k\|_{x^k}^2 \right\} \triangleq -\delta_k. \tag{64}$$

Rearranging and summing these inequalities for k from 0 to $K - 1$, and using that $L_k \geq \underline{L}$, we obtain

$$\begin{aligned} K \min_{k=0, \dots, K-1} \frac{\underline{L} \|v^k\|_{x^k}}{96} \min \left\{ 1, 4\|v^k\|_{x^k}^2 \right\} &\leq \sum_{k=0}^{K-1} \delta_k \leq F_\mu(x^0) - F_\mu(x^K) \\ &\stackrel{(2)}{=} f(x^0) - f(x^K) + \mu(h(x^0) - h(x^K)) \leq f(x^0) - f_{\min}(X) + \varepsilon, \end{aligned} \tag{65}$$

where we used that, by the assumptions of Theorem 3, x^0 is a 4θ -analytic center defined in (21) and $\mu = \frac{\varepsilon}{4\theta}$, implying that $h(x^0) - h(x^K) \leq 4\theta = \varepsilon/\mu$. Thus, up to passing to a subsequence, we have $\|v^k\|_{x^k} \rightarrow 0$ as $k \rightarrow \infty$, which makes the stopping criterion in Algorithm 3 achievable.

Assume now that the stopping criterion does not hold for K iterations of SAHBA. Then, for all $k = 0, \dots, K - 1$, it holds that

$$\begin{aligned} \delta_k &= \frac{L_k}{96} \min \left\{ \|v^k\|_{x^k}, 4\|v^k\|_{x^k}^3 \right\} \stackrel{(60)}{\geq} \frac{L_k}{96} \min \left\{ \sqrt{\frac{\varepsilon}{4L_k\theta}}, \frac{4\varepsilon^{3/2}}{4^{3/2}L_k^{3/2}\theta^{3/2}} \right\} \\ &\stackrel{L_k \leq 2\bar{M}, \theta \geq 1}{\geq} \frac{1}{96} \min \left\{ \frac{L_k\sqrt{\varepsilon}}{\sqrt{8\bar{M}}\theta^{3/2}}, \frac{\varepsilon^{3/2}}{2L_k^{1/2}\theta^{3/2}} \right\} \\ &\stackrel{L_k \leq 2\bar{M}, L_k \geq 144\varepsilon}{\geq} \frac{1}{96} \min \left\{ \frac{(144\varepsilon) \cdot \sqrt{\varepsilon}}{\sqrt{8\bar{M}}\theta^{3/2}}, \frac{\varepsilon^{3/2}}{\sqrt{8\bar{M}}\theta^{3/2}} \right\} = \frac{\varepsilon^{3/2}}{192\theta^{3/2}\sqrt{2\bar{M}}}. \end{aligned}$$

Thus, from (65)

$$K \frac{\varepsilon^{3/2}}{192\theta^{3/2}\sqrt{2\bar{M}}} \leq f(x^0) - f_{\min}(X) + \varepsilon.$$

Hence, recalling that $\bar{M} = \max\{M_0, M\}$, we obtain $K \leq \frac{192\theta^{3/2}\sqrt{2\max\{M_0, M\}}(f(x^0) - f_{\min}(X) + \varepsilon)}{\varepsilon^{3/2}}$, i.e., the algorithm stops for sure after no more than this number of iterations. This, combined with the bound for the number of inner steps in Sect. 5.2.2, proves the first statement of Theorem 3.

5.2.4 Generating $(\varepsilon_1, \varepsilon_2)$ -2KKT point

In this section, to finish the proof of Theorem 3, we show that if the stopping criterion in Algorithm 3 holds, i.e., $\|v^{k-1}\|_{x^{k-1}} < \Delta_{k-1}$ and $\|v^k\|_{x^k} < \Delta_k$, then the algorithm has generated an $(\varepsilon_1, \varepsilon_2)$ -2KKT point of (Opt) according to Definition 5, with $\varepsilon_1 = \varepsilon$ and $\varepsilon_2 = \frac{\max\{M_0, M\}\varepsilon}{8\theta}$.

Let the stopping criterion hold at iteration k . Using the first-order optimality condition (45) for the subproblem (55) solved at iteration $k - 1$, there exists a Lagrange multiplier $y^{k-1} \in \mathbb{R}^m$ such that (45) holds. Now, expanding the definition of the potential (2) and adding $\nabla f(x^k)$ to both sides, we obtain

$$\begin{aligned} \nabla f(x^k) - \mathbf{A}^*y^{k-1} + \mu \nabla h(x^{k-1}) &= \nabla f(x^k) - \nabla f(x^{k-1}) - \nabla^2 f(x^{k-1})v^{k-1} \\ &\quad - \frac{L_{k-1}}{2} \|v^{k-1}\|_{x^{k-1}} H(x^{k-1})v^{k-1}. \end{aligned}$$

Setting $s^k \triangleq \nabla f(x^k) - \mathbf{A}^*y^{k-1} \in \mathbb{E}^*$ and $g^{k-1} \triangleq -\mu \nabla h(x^{k-1})$, after multiplication by $[H(x^{k-1})]^{-1}$, this is equivalent to

$$\begin{aligned} [H(x^{k-1})]^{-1} (s^k - g^{k-1}) &= [H(x^{k-1})]^{-1} (\nabla f(x^k) - \nabla f(x^{k-1}) - \nabla^2 f(x^{k-1})v^{k-1}) \\ &\quad - \frac{L_{k-1}}{2} \|v^{k-1}\|_{x^{k-1}} v^{k-1}. \end{aligned}$$

Multiplying both of the above equalities, we arrive at

$$\left(\|s^k - g^{k-1}\|_{x^{k-1}}^* \right)^2 = \left(\left\| \nabla f(x^k) - \nabla f(x^{k-1}) - \nabla^2 f(x^{k-1})v^{k-1} - \frac{L_{k-1}}{2} \|v^{k-1}\|_{x^{k-1}} H(x^{k-1})v^{k-1} \right\|_{x^{k-1}}^* \right)^2.$$

Taking the square root and applying the triangle inequality, we obtain

$$\begin{aligned} \|s^k - g^{k-1}\|_{x^{k-1}}^* &\leq \|\nabla f(x^k) - \nabla f(x^{k-1}) - \nabla^2 f(x^{k-1})v^{k-1}\|_{x^{k-1}}^* + \frac{L_{k-1}}{2} \|v^{k-1}\|_{x^{k-1}}^2 \\ &\stackrel{(58)}{\leq} \frac{L_{k-1}}{2} \|\alpha_{k-1} v^{k-1}\|_{x^{k-1}}^2 + \frac{L_{k-1}}{2} \|v^{k-1}\|_{x^{k-1}}^2. \end{aligned} \tag{66}$$

Since the stopping criterion holds, at iteration $k - 1$ we have

$$\zeta(x^{k-1}, v^{k-1}) \stackrel{(5)}{\leq} \|v^{k-1}\|_{x^{k-1}} < \Delta_{k-1} = \sqrt{\frac{\varepsilon}{4L_{k-1}\theta}} \leq \sqrt{\frac{\varepsilon}{4 \cdot 144\varepsilon\theta}} < \frac{1}{2}, \tag{67}$$

where we used that, by construction, $L_{k-1} \geq \underline{L} = 144\varepsilon$ and that $\theta \geq 1$. Hence, by (56), we have that $\alpha_{k-1} = 1$ and $x^k = x^{k-1} + v^{k-1}$. This, in turn, implies that

$$\|s^k - g^{k-1}\|_{x^{k-1}}^* \stackrel{(66)}{\leq} L_{k-1} \|v^{k-1}\|_{x^{k-1}}^2. \tag{68}$$

As in the analysis of the first-order method, we note that $\|s^k - g^{k-1}\|_{x^{k-1}}^* = \mu \|s^k - g^{k-1}\|_{\nabla^2 h_*(g^{k-1})}$ and $\mu = \frac{\varepsilon}{4\theta}$, which implies

$$\|s^k - g^{k-1}\|_{\nabla^2 h_*(g^{k-1})} \leq \frac{L_{k-1}}{\mu} \|v^{k-1}\|_{x^{k-1}}^2 < \frac{L_{k-1}}{\mu} \Delta_{k-1}^2 = \frac{L_{k-1}}{\frac{\varepsilon}{4\theta}} \cdot \frac{\varepsilon}{4L_{k-1}\theta} = 1. \tag{69}$$

Thus, since, by (78), $g^{k-1} = -\mu \nabla h(x^{k-1}) \in \text{int}(K^*)$, applying Lemma 1, we deduce that $\nabla f(x^k) - A^* y^{k-1} = s^k \in \text{int}(K^*)$. By construction, $x^k \in K$ and $Ax^k = b$. Therefore, conditions (10) and (11) both hold. We now check for the complementarity condition (14). We have

$$\langle s^k, x^k \rangle = \langle s^k, x^{k-1} + v^{k-1} \rangle = \langle s^k, x^{k-1} \rangle + \langle s^k, v^{k-1} \rangle.$$

We estimate each of the two terms in the r.h.s. separately. First,

$$\begin{aligned} 0 \leq \langle s^k, x^{k-1} \rangle &= \langle s^k - g^{k-1}, x^{k-1} \rangle + \langle g^{k-1}, x^{k-1} \rangle \\ &\leq \|s^k - g^{k-1}\|_{x^{k-1}}^* \cdot \|x^{k-1}\|_{x^{k-1}} - \mu \langle \nabla h(x^{k-1}), x^{k-1} \rangle \\ &\stackrel{(68), (82), (81)}{\leq} L_{k-1} \|v^{k-1}\|_{x^{k-1}}^2 \sqrt{\theta} + \mu\theta. \end{aligned}$$

Second,

$$\begin{aligned} \langle s^k, v^{k-1} \rangle &\leq \|s^k\|_{x^{k-1}}^* \cdot \|v^{k-1}\|_{x^{k-1}} \leq \left(\|s^k - g^{k-1}\|_{x^{k-1}}^* + \|g^{k-1}\|_{x^{k-1}}^* \right) \cdot \|v^{k-1}\|_{x^{k-1}} \\ &\stackrel{(68), (82), (81)}{\leq} \left(L_{k-1} \|v^{k-1}\|_{x^{k-1}}^2 + \mu\sqrt{\theta} \right) \Delta_{k-1}. \end{aligned}$$

Summing up, using the stopping criterion $\|v^{k-1}\|_{x^{k-1}} < \Delta_{k-1}$ and that, by (67), $\Delta_{k-1} \leq 1 \leq \sqrt{\theta}$, we obtain

$$\begin{aligned} 0 \leq \langle s^k, x^k \rangle &= \langle s^k, x^{k-1} + v^{k-1} \rangle \leq 2L_{k-1} \Delta_{k-1}^2 \sqrt{\theta} + 2\mu\theta \\ &= 2L_{k-1} \frac{\varepsilon}{4L_{k-1}\theta} \sqrt{\theta} + 2\frac{\varepsilon}{4\theta} \theta \leq \varepsilon, \end{aligned} \tag{70}$$

i.e., (14) holds with $\varepsilon_1 = \varepsilon$.

Finally, we show the second-order condition (15). By inequality (47) for subproblem (55) solved at iteration k , we obtain on L_0

$$\begin{aligned} \nabla^2 f(x^k) &\geq -\frac{L_k \|v^k\|_{x^k}}{2} H(x^k) \geq -\frac{L_k \Delta_k}{2} H(x^k) \\ &= -\frac{L_k}{2} \sqrt{\frac{\varepsilon}{4L_k\theta}} H(x^k) = -\frac{\sqrt{L_k \varepsilon}}{4\theta^{1/2}} H(x^k) \geq -\frac{\sqrt{2\bar{M}\varepsilon}}{4\theta^{1/2}} H(x^k), \end{aligned} \tag{71}$$

where we used the second part of the stopping criterion, i.e., $\|v^k\|_{x^k} < \Delta_k$ and that $L_k \leq 2\bar{M} = 2 \max\{M, M_0\}$ (see Sect. 5.2.2). Thus, (15) holds with $\varepsilon_2 = \frac{\max\{M, M_0\}\varepsilon}{8\theta}$, which finishes the proof of Theorem 3.

5.3 Discussion

5.3.1 Special case of non-negativity constraints

As in Sect. 4.4, our aim in this section is to compare our result with those available in the contemporary literature. [58, 82] focus exclusively on the domain $\bar{K} = \bar{K}_{NN} = \mathbb{R}_+^n$, i.e., a particular case covered by our general problem template (Opt). A second-order algorithm and a first-order implementation of a second-order algorithm are proposed respectively. To compare their results with ours, consider the cone \bar{K}_{NN} , endowed with the standard log-barrier $h(x) = -\sum_{i=1}^n \ln(x_i)$. Recall that for this barrier setup we have $\nabla h(x) = [-x_1^{-1}, \dots, -x_n^{-1}]^\top$ and $H(x) = \text{diag}[x_1^{-2}, \dots, x_n^{-2}] = \mathbf{X}^{-2}$. Assume that the stopping criterion applies at iteration k . Using the first-order optimality condition (45) for the subproblem (55) solved at iteration $k - 1$ and expanding the definition of the potential (2), there exists a dual variable $y^{k-1} \in \mathbb{R}^m$ such that (45) holds, i.e.,

$$\begin{aligned} &\nabla f(x^{k-1}) + \mu \nabla h(x^{k-1}) + \nabla^2 f(x^{k-1})v^{k-1} - \mathbf{A}^* y^{k-1} \\ &= -\frac{L^{k-1}}{2} \|v^{k-1}\|_{x^{k-1}} H(x^{k-1})v^{k-1}. \end{aligned}$$

Multiplying both sides by $[H(x^{k-1})]^{-1/2} = \mathbf{X}^{k-1}$, using the stopping criterion $\|v^{k-1}\|_{x^{k-1}} < \sqrt{\frac{\varepsilon}{4\theta L^{k-1}}}$, since $\mathbf{X}^{k-1} \nabla h(x^{k-1}) = [H(x^{k-1})]^{-1/2} \nabla h(x^{k-1}) = -\mathbf{1}_n$ and $\theta = n$, we obtain

$$\begin{aligned} &\|\mathbf{X}^{k-1}(\nabla^2 f(x^{k-1})v^{k-1} + \nabla f(x^{k-1}) - \mathbf{A}^* y^{k-1}) - \mu \mathbf{1}_n\|_\infty \\ &\leq \|\mathbf{X}^{k-1}(\nabla^2 f(x^{k-1})v^{k-1} + \nabla f(x^{k-1}) - \mathbf{A}^* y^{k-1}) \\ &\quad - \mu \mathbf{1}_n\| = \frac{L^{k-1}}{2} \|-H(x^{k-1})^{\frac{1}{2}} v^{k-1}\|^2 = \frac{L^{k-1}}{2} \|v^{k-1}\|_{x^{k-1}}^2 \\ &< \frac{\varepsilon}{8n}. \end{aligned} \tag{72}$$

Whence, since $\mu = \frac{\varepsilon}{4n}$, the above bound (72) combined with the triangle inequality yields

$$\begin{aligned} &\|\mathbf{X}^{k-1}(\nabla^2 f(x^{k-1})v^{k-1} + \nabla f(x^{k-1}) - \mathbf{A}^* y^{k-1})\|_\infty \\ &\leq \|\mathbf{X}^{k-1}(\nabla^2 f(x^{k-1})v^{k-1} + \nabla f(x^{k-1}) - \mathbf{A}^* y^{k-1}) - \mu \mathbf{1}_n\|_\infty + \|\mu \mathbf{1}_n\|_\infty < \frac{3\varepsilon}{8n}. \end{aligned} \tag{73}$$

Let $\mathbf{V}^{k-1} = \text{diag}[v_1^{k-1}, \dots, v_n^{k-1}] = \text{diag}(v^{k-1})$. Using the fact that $x^k = x^{k-1} + v^{k-1}$ shown after (67), we obtain

$$\begin{aligned} &\|\mathbf{X}^k(\nabla f(x^k) - \mathbf{A}^* y^{k-1})\|_\infty \\ &= \|(\mathbf{X}^{k-1} + \mathbf{V}^{k-1})(\nabla^2 f(x^{k-1})v^{k-1} + \nabla f(x^{k-1}) - \mathbf{A}^* y^{k-1} + \nabla f(x^k))\|_\infty \end{aligned}$$

$$\begin{aligned}
 & -\nabla f(x^{k-1}) - \nabla^2 f(x^{k-1})v^{k-1} \|\infty \\
 \leq & \|\mathbf{X}^{k-1}(\nabla^2 f(x^{k-1})v^{k-1} + \nabla f(x^{k-1}) - \mathbf{A}^*y^{k-1})\|\infty \\
 & + \|\mathbf{X}^{k-1}(\nabla f(x^k) - \nabla f(x^{k-1}) - \nabla^2 f(x^{k-1})v^{k-1})\|\infty \\
 & + \|\mathbf{V}^{k-1}(\nabla^2 f(x^{k-1})v^{k-1} + \nabla f(x^{k-1}) - \mathbf{A}^*y^{k-1})\|\infty \\
 & + \|\mathbf{V}^{k-1}(\nabla f(x^k) - \nabla f(x^{k-1}) - \nabla^2 f(x^{k-1})v^{k-1})\|\infty \\
 = & I + II + III + IV.
 \end{aligned}$$

Let us estimate each of the four terms $I - IV$ using two technical facts (87) and (88) proved in Appendix C. We have:

$$\begin{aligned}
 I & \stackrel{(73)}{<} \frac{3\varepsilon}{8n}, \\
 II & \leq \|\mathbf{X}^{k-1}(\nabla f(x^k) - \nabla f(x^{k-1}) - \nabla^2 f(x^{k-1})v^{k-1})\| \\
 & = \|\nabla f(x^k) - \nabla f(x^{k-1}) - \nabla^2 f(x^{k-1})v^{k-1}\|_{x^{k-1}}^* \stackrel{(58)}{\leq} \frac{L_{k-1}}{2} \|v^{k-1}\|_{x^{k-1}}^2 < \frac{\varepsilon}{8n}, \\
 III & \stackrel{(88)}{\leq} \|v^{k-1}\|_{x^{k-1}} \cdot \|\mathbf{X}^{k-1}(\nabla^2 f(x^{k-1})v^{k-1} + \nabla f(x^{k-1}) - \mathbf{A}^*y^{k-1})\|\infty \stackrel{(67),(73)}{<} \frac{3\varepsilon}{8n},
 \end{aligned}$$

where we used $x^k = z^{k-1} = x^{k-1} + v^{k-1}$ in bounding II , and the last bound for expression III uses $\|v^{k-1}\|_{x^{k-1}} < 1$, which is implied by Eq. (67). Finally, we also obtain

$$\begin{aligned}
 IV & \stackrel{(87)}{\leq} \|v^{k-1}\|_{x^{k-1}} \cdot \|\nabla f(x^k) - \nabla f(x^{k-1}) - \nabla^2 f(x^{k-1})v^{k-1}\|_{x^{k-1}}^* \\
 & \stackrel{(67),(58)}{\leq} \frac{L_{k-1}}{2} \|v^{k-1}\|_{x^{k-1}}^2 < \frac{\varepsilon}{8n}.
 \end{aligned}$$

Summarizing, we arrive at

$$\|\mathbf{X}^k(\nabla f(x^k) - \mathbf{A}^*y^{k-1})\|\infty \leq \frac{\varepsilon}{n}. \tag{74}$$

Further, by Theorem 3, we have that $\nabla f(x^k) - \mathbf{A}^*y^{k-1} = s^k \in \text{int}(\mathbf{K}_{\mathbb{N}\mathbb{N}}^*) = \mathbb{R}_{++}^n$, and

$$\nabla^2 f(x^k) + H(x^k)\sqrt{\frac{M\varepsilon}{n}} \geq 0 \text{ on } \mathbf{L}_0.$$

By Remark 9, these inequalities are achieved after $O\left(\frac{\sqrt{M}n^{3/2}(f(x^0)-f_{\min}(X))}{\varepsilon^{3/2}}\right)$ iterations of SAHBA. Assuming that $M \geq 1$, if we change $\varepsilon \rightarrow \tilde{\varepsilon} = \min\{n\varepsilon, n\varepsilon/M\}$, we obtain from these inequalities that in $O\left(\frac{\sqrt{M}n^{3/2}(f(x^0)-f_{\min}(X))}{\tilde{\varepsilon}^{3/2}}\right) = O\left(\frac{M^2(f(x^0)-f_{\min}(X))}{\varepsilon^{3/2}}\right)$ iterations SAHBA guarantees

$$\begin{aligned}
 &x^k > 0, \quad \nabla f(x^k) - \mathbf{A}^* y^{k-1} > 0 \\
 &\|\mathbf{X}^k (\nabla f(x^k) - \mathbf{A}^* y^{k-1})\|_\infty \leq \frac{\tilde{\varepsilon}}{n} \leq \varepsilon, \\
 &\nabla^2 f(x^k) + H(x^k) \sqrt{\varepsilon} \succeq \nabla^2 f(x^k) + H(x^k) \sqrt{\frac{M\tilde{\varepsilon}}{n}} \succeq 0 \text{ on } L_0.
 \end{aligned}$$

In contrast, the second-order algorithm of [58] gives a slightly worse guarantee $\nabla f(x^k) - \mathbf{A}^* y^{k-1} > -\varepsilon$, and requires a larger number of iterations $O\left(\frac{\max\{M, R\}^{7/2}(f(x^0) - f_{\min}(X))}{\varepsilon^{3/2}}\right)$ (R denoting the L_∞ upper bound on the diameter of the level set corresponding to x^0), albeit making similar assumptions to ours. We also can repeat the remarks from Sect. 4.4, arguing that our measure of complementarity $0 \leq \langle s^k, x^k \rangle \leq \varepsilon$ is stronger than $\max_{1 \leq i \leq n} |x_i^k s_i^k|$ used in [58, 82]. Also, the works [65, 82] consider problem (Opt) without linear equality constraints.

Our complexity of $O(\varepsilon^{-3/2})$ is better than the complexity bound $O(\varepsilon^{-7/4})$ in [65] when their algorithm is specified to problem with linear constraint $x \in \bar{K}_{\text{NN}}$ instead of nonlinear inequality constraint $a(x) \geq 0$ with an appropriately smooth function $a : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Furthermore, our algorithm is applicable to general cones admitting an efficient barrier setup, rather than only for \bar{K}_{NN} as in the discussed previous works [58, 65, 82]. For more general cones we can not use the coupling $H(x)^{-\frac{1}{2}} = \mathbf{X}$, which was seen to be very helpful in the derivations of the bound (74) above. Thus, to deal with general cones, we had to find and exploit suitable properties of the barrier class $\mathcal{H}_\theta(K)$ and develop a new analysis technique that works for general, potentially non-symmetric, cones. Finally, our method does not rely on the trust-region techniques as in [58] that may slow down the convergence in practice since the radius of the trust region is no greater than $O(\sqrt{\varepsilon})$ leading to short steps.

5.3.2 Exploiting the structure of symmetric cones

We note that in (63) we can clearly observe the benefit of the use of θ -SSB in our algorithm. When $\alpha_k = \frac{1}{2\zeta(x^k, v^k)}$, the per-iteration decrease of the potential is $\frac{L_k \|v^k\|_{x^k}^3}{96(\zeta(x^k, v^k))^2} \geq \frac{\sqrt{\varepsilon} L_k \|v^k\|_{x^k}^2}{96\sqrt{4\theta}(\zeta(x^k, v^k))^2}$ which may be large if $\zeta(x^k, v^k) = \sigma_{x^k}(-v^k) \ll \|v^k\|_{x^k}$.

5.3.3 The role of the penalty parameter

Next, we discuss more explicitly, how the algorithm and complexity bounds depend on the parameter μ . The first observation is that from (69), to guarantee that $s^k \in K^*$, we need the stopping criterion to be $\|v^{k-1}\|_{x^{k-1}} < \Delta_{k-1} = \sqrt{\mu/L_{k-1}}$, which by (70) leads to the error $4\mu\theta$ in the complementarity conditions and by (71) leads to the error $\sqrt{\mu/M}$ in the second-order condition. From the analysis following equation (63), we have that

$$K \frac{\mu^{3/2}}{24\sqrt{M}} \leq f(x^0) - f_{\min}(X) + \mu\theta.$$

Whence, recalling that $\bar{M} = \max\{M, M_0\}$,

$$K \leq 24(f(x^0) - f_{\min}(X) + \mu\theta) \cdot \frac{\sqrt{2 \max\{M, M_0\}}}{\mu^{3/2}}.$$

Thus, we see that after $O(\mu^{-3/2})$ iterations the algorithm finds a $(4\mu\theta, \mu/\bar{M})$ -2KKT point, and if $\mu \rightarrow 0$, we have convergence to a KKT point, but the complexity bound tends to infinity and becomes non-informative. At the same time, as it is seen from (55), when $\mu \rightarrow 0$, the algorithm resembles a cubic-regularized Newton method, but with the regularization with the cube of the local norm. We also see from the above explicit expressions in terms of μ that the design of the algorithm requires careful balance between the desired accuracy of the approximate KKT point expressed mainly by the complementarity conditions, stopping criterion, and complexity. Moreover, the step-size must be selected carefully to ensure the feasibility of the iterates. This is also in contrast to the cubic-regularized Newton's method where one can always take the step-size 1.

5.4 Anytime convergence via restarting SAHBA

Similarly to the restarted FAHBA (Algorithm 2), we can obtain anytime convergence with a similar complexity as that of SAHBA by invoking a restarted method that uses SAHBA as an inner procedure. We fix $\varepsilon_0 > 0$ and select the starting point x_0^0 as a 4θ -analytic center of X in the sense of Eq. (21). In epoch $i \geq 0$ we generate a sequence $\{x_i^k\}_{k=0}^{K_i}$ by calling SAHBA($\mu_i, \varepsilon_i, M_0^{(i)}, x_i^0$) with $\mu_i = \frac{\varepsilon_i}{4\theta}$ until the stopping condition is reached. We know that this inner procedure terminates after at most $\mathbb{K}_{II}(\varepsilon_i, x_i^0)$ iterations. We store the values $x_i^{K_i}$ and $M_{K_i}^{(i)}$, and set $x_{i+1}^0 \equiv x_i^{K_i}$, as well as $M_0^{(i+1)} \equiv M_{K_i}^{(i)}/2$. Updating the parameters to μ_{i+1} and ε_{i+1} , we restart by calling procedure SAHBA($\mu_{i+1}, \varepsilon_{i+1}, M_0^{(i+1)}, x_{i+1}^0$) anew. This is formalized in Algorithm 4.

Algorithm 4: Restarting SAHBA

Data: $h \in \mathcal{H}_\theta(K), \varepsilon_0 > 0, x_0^0 \in X - 4\theta$ -analytic center, $M_0^{(0)} \geq 144\varepsilon_0$.
Result: Point \hat{x}_i , dual variables $\hat{y}_i, \hat{s}_i = \nabla f(\hat{x}_i) - \mathbf{A}^* \hat{y}_i$.
for $i = 0, 1, \dots$ **do**
 Set $\varepsilon_i = 2^{-i} \varepsilon_0, \mu_i = \frac{\varepsilon_i}{4\theta}$;
 Obtain $(\hat{x}_i, \hat{y}_i, \hat{s}_i, \hat{M}_i)$ from SAHBA($\mu_i, \varepsilon_i, M_0^{(i)}, x_i^0$);
 Set $x_{i+1}^0 = \hat{x}_i$ and $M_0^{(i+1)} = \hat{M}_i/2$.
end

Theorem 4 *Let Assumptions 1, 2, 4 hold. Then, for any $\varepsilon \in (0, \varepsilon_0)$, Algorithm 4 finds an $(\varepsilon, \frac{\max\{M, M_0^{(0)}\}\varepsilon}{8\theta})$ -2KKT point for problem (Opt) in the sense of Definition 5 after no more than $I(\varepsilon) = \lceil \log_2(\varepsilon_0/\varepsilon) \rceil + 1$ restarts and at most $\lceil 300(f(x^0) - f_{\min}(X) + \varepsilon_0)\theta^{3/2}\varepsilon^{-3/2}\sqrt{2 \max\{M, M_0^{(0)}\}} \rceil$ iterations of SAHBA.*

Proof Let us consider a restart $i \geq 0$ and mimic the proof of Theorem 3 with the substitution $\varepsilon \rightarrow \varepsilon_i, \mu \rightarrow \mu_i = \varepsilon_i/(4\theta), M_0 \rightarrow M_0^{(i)} = \hat{M}_{i-1}/2, \underline{L} = 144\varepsilon \rightarrow \underline{L}_i = 144\varepsilon_i, \bar{M} = \max\{M, M_0\} \rightarrow \bar{M}_i = \max\{M, M_0^{(i)}\}, x^0 \rightarrow x_i^0 = \hat{x}_{i-1}$. Note that $M_0^{(i)} \geq 144\varepsilon_i = \underline{L}_i$ for $i \geq 0$. We verify this via induction. By construction $M_0^{(0)} \geq 144\varepsilon_0$. Assume the bound holds for some $i \geq 1$. Then, $M_0^{(i+1)} = M_{K_i}^{(i)}/2 = \max\{L_{K_i-1}^{(i)}/2, \underline{L}_i\}/2 \geq 144\varepsilon_i/2 = 144\varepsilon_{i+1}$, where we used the induction hypothesis and the definition of the sequence ε_i .

Let K_i be the last iteration of SAHBA($\mu_i, \varepsilon_i, M_0^{(i)}, x_i^0$) meaning that the stopping criterion does not hold at the inner K_i iterations $k = 0, \dots, K_i - 1$. From the analysis following Eq. (64), we obtain

$$K_i \frac{\varepsilon_i^{3/2}}{192\theta^{3/2}\sqrt{2\bar{M}_i}} \leq F_{\mu_i}(x_i^0) - F_{\mu_i}(x_i^{K_i}). \tag{75}$$

Using that μ_i is a decreasing sequence and x_0^0 is a 4θ -analytic center, we see

$$\begin{aligned} F_{\mu_{i+1}}(x_{i+1}^0) &= F_{\mu_{i+1}}(x_i^{K_i}) = f(x_i^{K_i}) + \mu_{i+1}h(x_i^{K_i}) = F_{\mu_i}(x_i^{K_i}) + (\mu_{i+1} - \mu_i)h(x_i^{K_i}) \\ &\stackrel{(21)}{\leq} F_{\mu_i}(x_i^{K_i}) + (\mu_{i+1} - \mu_i)(h(x_0^0) - 4\theta) \\ &\stackrel{(75)}{\leq} F_{\mu_i}(x_i^0) - K_i \frac{\varepsilon_i^{3/2}}{192\theta^{3/2}\sqrt{2\bar{M}_i}} + (\mu_{i+1} - \mu_i)(h(x_0^0) - 4\theta). \end{aligned} \tag{76}$$

Let $I \equiv I(\varepsilon) = \lceil \log_2 \frac{\varepsilon_0}{\varepsilon} \rceil + 1$. By Theorem 3 applied to the restart $I - 1$, we see that SAHBA($\mu_{I-1}, \varepsilon_{I-1}, M_0^{(I-1)}, x_{I-1}^0$) outputs an $(\varepsilon_{I-1}, \frac{\bar{M}_{I-1}\varepsilon_{I-1}}{8\theta})$ -2KKT point for problem (Opt) in the sense of Definition 5. Since $\varepsilon_{I-1} \leq \varepsilon$ and, for all $i \geq 1$,

$$\begin{aligned} \bar{M}_i &= \max\{M, M_0^{(i)}\} = \max\{M, \hat{M}_{i-1}/2\} = \max\{M, M_{K_i-1}^{(i-1)}/2\} \\ &= \max\{M, \max\{L_{K_i-1}^{(i-1)}/2, \underline{L}_{i-1}\}/2\} \\ &\leq \max\{M, \max\{\bar{M}_{i-1}, M_0^{(i-1)}\}/2\} \leq \max\{M, \bar{M}_{i-1}\} \\ &\leq \dots \leq \max\{M, \bar{M}_0\} \leq \max\{M, M_0^{(0)}\}, \end{aligned} \tag{77}$$

it follows that actually we generate an $(\varepsilon, \frac{\max\{M, M_0^{(0)}\}\varepsilon}{8\theta})$ -2KKT point. Summing inequalities (76) for all the performed restarts $i = 0, \dots, I - 1$ and rearranging the terms, we obtain

$$\begin{aligned} \sum_{i=0}^{I-1} K_i \frac{\varepsilon_i^{3/2}}{192\theta^{3/2}\sqrt{2\bar{M}_i}} &\leq F_{\mu_0}(x_0^0) - F_{\mu_I}(x_I^0) + (\mu_I - \mu_0)(h(x_0^0) - 4\theta) \\ &= f(x_0^0) + \mu_0h(x_0^0) - f(x_I^0) - \mu_Ih(x_I^0) + (\mu_I - \mu_0)(h(x_0^0) - 4\theta) \\ &\stackrel{(21)}{\leq} f(x_0^0) - f_{\min}(X) + \mu_0h(x_0^0) - \mu_Ih(x_0^0) + 4\mu_I\theta + (\mu_I - \mu_0)(h(x_0^0) - 4\theta) \\ &\leq f(x_0^0) - f_{\min}(X) + 4\mu_0\theta = f(x_0^0) - f_{\min}(X) + \varepsilon_0, \end{aligned}$$

where in the last steps we used the coupling $\mu_0 = \varepsilon_0/\theta$. From this inequality, using (77), we obtain

$$K_i \leq (f(x^0) - f_{\min}(X) + \varepsilon_0) \cdot \frac{192\theta^{3/2}\sqrt{2\bar{M}_i}}{\varepsilon_i^{3/2}} \leq \frac{C}{\varepsilon_i^{3/2}},$$

where $C \equiv 192(f(x^0) - f_{\min}(X) + \varepsilon_0)\theta^{3/2}\sqrt{2 \max\{M, M_0^{(0)}\}}$. Finally, we obtain that the total number of iterations of procedures **SAHBA**($\mu_i, \varepsilon_i, M_0^{(i)}, x_i^0$), $0 \leq i \leq I - 1$, to reach accuracy ε is at most

$$\begin{aligned} \sum_{i=0}^{I-1} K_i &\leq \sum_{i=0}^{I-1} \frac{C}{\varepsilon_i^{3/2}} \leq \frac{C}{\varepsilon_0^{3/2}} \sum_{i=0}^{I-1} (2^{3/2})^i \\ &\leq \frac{C}{\varepsilon_0^{3/2}} \cdot \frac{2^{3/2 \cdot (1 + \log_2(\frac{\varepsilon_0}{\varepsilon}))} - 1}{2^{3/2} - 1} \leq \frac{2^{3/2} C}{(\sqrt{8} - 1)\varepsilon^{3/2}} \\ &< \frac{300(f(x^0) - f_{\min}(X) + \varepsilon_0)\theta^{3/2}\sqrt{2 \max\{M, M_0^{(0)}\}}}{\varepsilon^{3/2}}. \end{aligned}$$

□

6 Conclusion

We derived Hessian barrier algorithms based on first- and second-order information on the objective f . We performed a detailed analysis of their worst-case iteration complexity in order to find a suitably defined approximate KKT point. Under weak regularity assumptions and in the presence of general conic constraints, our Hessian barrier algorithms share the best known complexity rates in the literature for first- and second-order approximate KKT points. Our methods are characterized by a decomposition approach of the feasible set which leads to numerically efficient subproblems at each their iteration. Several open questions for the future remain. First, our iterations assume that the subproblems are solved exactly, and for practical reasons, this should be relaxed. Second, we mentioned that **FAHBA** can be interpreted as a discretization of the Hessian-barrier gradient system [4], but the exact relationship is not explored yet. This, however, could be an important step towards understanding acceleration techniques of **FAHBA**, akin to an accelerated version of the cubic regularized Newton method. Furthermore, the cubic-regularized version has no corresponding continuous-time version yet. It will be very interesting to investigate this question further. Additionally, the question of convergence of the trajectory $(x^k)_{k \geq 0}$ generated by either scheme is open. Another interesting direction for future research would be to allow for higher-order Taylor expansions in the subproblems in order to boost convergence speed further, similar to [28].

Acknowledgements We would like to thank Coralia Cartis, Yurii Nesterov, Jerome Bolte, Anton Rodomanov, Nikita Doikov, Giovanni Grapiglia and Radu Dragomir for fruitful discussions on this topic. We also would like to thank the three referees and the associate editor for the excellent remarks that helped us tremendously to improve the quality of the paper.

Funding Open Access funding enabled and organized by Projekt DEAL. M. Staudigl acknowledges support from the COST Action CA16228 “European Network for Game Theory”. The work by P. Dvurechensky was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy—The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

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

Appendix A: More results on self-concordant barriers

In this appendix we collect important auxiliary results about logarithmically homogeneous barriers and their limit properties along primal sequences that converge to the boundary.

We denote by $\mathcal{L}(\mathbb{E}, \mathbb{E}^*)$ the set of linear operators from \mathbb{E} to \mathbb{E}^* . Given $h \in \mathcal{H}_\theta(K)$, we define the operator $D^2h : K \rightarrow \mathcal{L}(\mathbb{E}, \mathbb{E}^*)$, by the bilinear form $D^2h(x)[u, v]$ for all $x \in K$ and for all $(u, v) \in \mathbb{E}$. Fixing the coordinate system \mathbf{B} , we can define a matrix representation of the operator $D^2h(x)$ as follows. First, by the Riesz–Fréchet representation theorem, there exists a unique element in \mathbb{E}^* , denoted by $H(x)u$ such that $D^2h(x)[u, v] = \langle H(x)u, v \rangle$ for all $v \in \mathbb{E}$. $H(x)$ is a linear operator, so that $H(x)u \in \mathbb{E}^*$ for all $u \in \mathbb{E}$. Let $u, v \in \mathbb{E}$ with $u = \mathbf{B}\xi$ and $v = \mathbf{B}\alpha$ and $\xi, \alpha \in \mathbb{R}^p$. Then, using the bilinearity, we see that for all $x \in K$

$$\begin{aligned} D^2h(x)[u, v] &= D^2h(x)[u, \sum_{j=1}^n b_j \alpha_j] = \sum_{j=1}^n \alpha_j D^2h(x)[u, b_j] \\ &= \sum_{i,j=1}^n \xi_i \alpha_j D^2h(x)[b_i, b_j]. \end{aligned}$$

Hence, $\langle H(x)b_i, b_j \rangle = D^2h(x)[b_i, b_j]$ for all $i, j = 1, \dots, p$. Therefore,

$$D^2h(x)[u, v] = \langle H(x)u, v \rangle = \langle H(x)\mathbf{B}\xi, \mathbf{B}\alpha \rangle = \langle \mathbf{B}^* H(x)\mathbf{B}\xi, \alpha \rangle_{\mathbb{R}^p}.$$

This gives $\mathbf{H}_x^{\mathbf{B}} \triangleq \mathbf{B}^*H(x)\mathbf{B} : \mathbb{R}^p \rightarrow \mathbb{R}^p$ as a matrix representation of the linear operator $H(x)$ in terms of the primal basis \mathbf{B} on \mathbb{E} and the dual basis \mathbf{B}^* on \mathbb{E}^* . From now on we identify \mathbb{E}^* with \mathbb{E} so that $H(x) \in \mathcal{L}(\mathbb{E}; \mathbb{E}) \triangleq \mathcal{L}(\mathbb{E})$. Since $H(x)$ is a positive definite operator, there exists an inverse $[H(x)]^{-1} : \mathbb{E} \rightarrow \mathbb{E}$ so that $H(x) \circ [H(x)]^{-1} = [H(x)]^{-1} \circ H(x) = \text{Id}_{\mathbb{E}}$. $[H(x)]^{-1}$ is again a positive definite linear operator and thus admits a unique *square root* $R_x : \mathbb{E} \rightarrow \mathbb{E}$ satisfying $[H(x)]^{-1} = R_x \circ R_x$. We denote this linear operator as $R_x = [H(x)]^{-1/2} \in \mathcal{L}(\mathbb{E})$.

The *dual cone* \bar{K}^* is defined as $\bar{K}^* \triangleq \{s \in \mathbb{E}^* \mid \langle s, x \rangle \geq 0 \ \forall x \in \bar{K}\}$, and the *dual barrier* is defined as $h_*(s) \triangleq \sup_{x \in K} \{-s, x\} - h(x)$ for $s \in \bar{K}^*$. From [83, Theorem 3.3.1] we know that if $h \in \mathcal{H}_\theta(K)$, then $h_* \in \mathcal{H}_\theta(K^*)$. Moreover,

$$x \in K \Rightarrow -\nabla h(x) \in \text{int}(K^*), \tag{78}$$

$$s = -\nabla h(x) \iff \nabla h_*(s) = -x \Rightarrow \nabla^2 h_*(s) = [\nabla^2 h(x)]^{-1}. \tag{79}$$

We also use the following properties listed in [76, Lemma 5.4.3, Theorem 5.3.7].

Proposition 6 *Let $h \in \mathcal{H}_\theta(K)$, $x \in K$, $t > 0$ and $H(x) = \nabla^2 h(x)$. Then,*

$$\nabla^2 h(tx) = t^{-2} \nabla^2 h(x), \tag{80}$$

$$-\langle \nabla h(x), x \rangle = \theta, \tag{81}$$

$$\|x\|_x^2 = \langle H(x)x, x \rangle = \theta, \quad \langle \nabla h(x), [H(x)]^{-1} \nabla h(x) \rangle = \theta. \tag{82}$$

$$\langle \nabla h(x), y - x \rangle < \theta \quad \forall y \in K \tag{83}$$

The next fact is essentially [4, Lemma 4.2]. We give the simple proof here to be self-contained.

Lemma 4 *Let $h \in \mathcal{H}_\theta(K)$. Consider a sequence $(x_k)_{k \geq 1} \subset K$ such that $x_k \rightarrow \bar{x} \in \text{bd}(K)$. Then all accumulation points of $\left(\frac{\nabla h(x_k)}{\|\nabla h(x_k)\|}\right)_{k \geq 1}$ are contained in $\text{NC}_{\bar{K}}(\bar{x})$.*

Proof By convexity of h , we have $\langle \nabla h(x_k) - \nabla h(x), x_k - x \rangle \geq 0$ for all $x \in K$. The sequence $(\|\nabla h(x_k)\|^{-1} \nabla h(x_k))_{k \geq 1}$ is bounded and thus converging subsequences exist. Let ξ denote an accumulation point. After eventual relabeling, let us assume $\|\nabla h(x_k)\|^{-1} \nabla h(x_k) \rightarrow \xi$. Then, dividing the last inner product by $\|\nabla h(x_k)\|$ and letting $k \rightarrow \infty$, we get $\langle \xi, \bar{x} - x \rangle \geq 0$ for all $x \in K$. Hence, $\xi \in \text{NC}_{\bar{K}}(\bar{x})$. Since ξ is an arbitrary accumulation point, the claim follows. \square

Lemma 5 *Define the mapping $e_* : \mathbb{E} \rightarrow [0, \infty]$ by*

$$e_*(v) \triangleq \sup_{x \in K: \|x\|=1} \|v\|_x^*.$$

Then $\text{dom}(e_) = \mathbb{E}$.*

Proof Let $x \in K$ satisfy $\|x\| = 1$ and $y \in K$ be arbitrary. Denote $r_1 \triangleq \|x - y\|$. There exists $r_2 > 0$ such that the open ball with center y and radius r_2 is contained in K .

By convexity, all points on the line $x + \alpha(y - x)$ are therefore in K , provided that $\alpha \in [0, \frac{r_1+r_2}{r_1}]$. Define the Minkowski functional of \bar{K} with pole y as

$$\pi_y(x) \triangleq \inf \left\{ t \geq 0 \mid y + \frac{1}{t}(x - y) \in \bar{K} \right\}.$$

It follows

$$\pi_x(y) \leq \frac{r_1}{r_1 + r_2} \leq \frac{1 + \|y\|}{1 + \|y\| + r_2} \triangleq a(y).$$

This, together with [78, Proposition 2.3.2], yields the bound

$$\|v\|_y \leq \frac{1 + 3\theta}{1 - \pi_x(y)} \|v\|_x \leq \frac{1 + 3\theta}{1 - a(y)} \|v\|_x.$$

Hence, $H(x) \left(\frac{1+3\theta}{1-a(y)} \right)^2 \geq H(y)$, and therefore $[H(y)]^{-1} \left(\frac{1+3\theta}{1-a(y)} \right)^2 \geq [H(x)]^{-1}$. We conclude

$$\|v\|_x^* \leq \frac{1 + 3\theta}{1 - a(y)} \|v\|_y^* \quad \forall y \in K.$$

Therefore, fixing some $y \in K$, we have $e_*(v) = \sup_{x \in K: \|x\|=1} \|v\|_x^* < \frac{1+3\theta}{1-a(y)} \|v\|_y^* < \infty$. □

Lemma 6 *Let $h \in \mathcal{H}_\theta(K)$. For all $(x, v) \in K \times \mathbb{E}$ we have*

$$\|v\|_x^* \leq \|x\| \cdot e_*(v).$$

Proof If $h \in \mathcal{H}_\theta(K)$, then by (80)

$$t^2 H(tx) = H(x) \quad \forall x \in K, t > 0.$$

This gives $t^2 H(tx) \circ H(x)^{-1} = H(x)^{-1} \circ (t^2 H(tx)) = \text{Id}_{\mathbb{E}}$. Hence,

$$[t^2 H(tx)]^{-1} = t^{-2} [H(tx)]^{-1} = [H(x)]^{-1}.$$

Choosing $t = 1/\|x\|$ gives $\|x\|^2 [H(x/\|x\|)]^{-1} = [H(x)]^{-1}$ for all $x \in K$. Hence, for all $v \in \mathbb{E}$, we get

$$\|v\|_x^* = \|x\| \cdot \|v\|_{x/\|x\|}^* \leq \|x\| \sup_{d:\|d\|=1} \|v\|_d^* \equiv \|x\| e_*(v) \quad \forall v \in \mathbb{E}.$$

□

Note that

$$\|R_x v\|_x^2 = \langle H(x)R_x v, R_x v \rangle = \langle v, v \rangle = \|v\|^2 \quad \forall (x, v) \in K \times \mathbb{E}.$$

We use the operator R_x to define a mapping

$$\tau : K \times \mathbb{E} \rightarrow \mathbb{E}, (x, v) \mapsto \tau(x, v) \triangleq x - \frac{1}{\|v\| + 1} R_x v.$$

In geometric terms, we can think of $\tau(x, v)$ as a *retraction* acting on the Riemannian manifold $(K, \|\cdot\|_x)$. This is the content of the next result.

Lemma 7 For all $(x, v) \in K \times \mathbb{E}$ we have $\tau(x, v) \in K$.

Proof We use the Dikin ellipsoid condition (Lemma 1) for viability:

$$\|x - \frac{1}{\|v\| + 1} R_x v - x\|_x = \frac{1}{\|v\| + 1} \|R_x v\|_x = \frac{\|v\|}{\|v\| + 1} < 1.$$

□

We next show that the operator R_x can be extended to the boundary of the cone K . Recall that we identify the dual space \mathbb{E}^* with \mathbb{E} . Lemma 6 states that $\|R_x v\| \leq \|x\| \cdot e_*(v)$ for all $x \in K$ and $v \in \mathbb{E}$. This shows that the mapping $K \ni x \mapsto R_x v \in \mathbb{E}$ is locally bounded. This allows us to define an extension of the domain of $x \mapsto R_x$ to $\bar{K} = K \cup \text{bd}(K)$ via a limiting procedure. Let $(x_n)_{n \geq 1} \subset K$ with $x_n \rightarrow \bar{x} \in \text{bd}(K)$, and $v \in \mathbb{E}$ arbitrary. Let $\{b_1, \dots, b_p\}$ be a basis of \mathbb{E} , so that any vector $v \in \mathbb{E}$ admits a unique representation $v = \sum_i \alpha_i b_i$ for some scalars $\alpha_i \in \mathbb{R}$. Then, by linearity, we have

$$R_{x_n} v = \sum_{i=1}^p \alpha_i R_{x_n} b_i.$$

Since $R_{x_n} b_i \leq \|x_n\| e_*(b_i)$ for all i , we can extract a converging subsequence $R_{x_{n_k}} b_i \rightarrow R b_i$ as $k \rightarrow \infty$. Via a diagonal procedure, we obtain another subsequence $(x_{n_j})_{j \geq 1}$ in order to construct a linear operator $R : \mathbb{E} \rightarrow \mathbb{E}$ satisfying

$$R_{x_{n_j}} v = \sum_i \alpha_i R_{x_{n_j}} b_i \rightarrow \sum_i \alpha_i R b_i = R v$$

as $j \rightarrow \infty$ for all $v \in \mathbb{E}$.

Remark 10 If K is a symmetric cone, it is well-known that $\lim_{n \rightarrow \infty} R_{x_n}$ exists for $x_n \rightarrow \bar{x} \in \text{bd}(K)$ and is a well-defined linear operator [73].

The next result is concerned with the semi-continuity of the multi-valued mapping $x \mapsto \mathcal{T}_x$, defined in (16). The discussion following that equation showed that this set can be equivalently represented as $\mathcal{T}_x = \{R_x u \mid \mathbf{A} R_x u = 0, \|u\| < 1\}$.

Proposition 7 *Let $h \in \mathcal{H}_\theta(K)$, $\bar{x} \in \text{bd}(K)$, and $\mathbf{A}\bar{x} = b$. Let $(x_n)_{n \geq 1} \subset K$ such that $x_n \rightarrow \bar{x}$ and $v_n \in \mathcal{T}_{x_n}$ be a corresponding sequence of direction vectors. Then, the sequence $p_n \triangleq x_n + v_n$ is contained in K and converges to a limit point $p \in \bar{K}$. Moreover, $\mathbf{A}p = b$.*

Proof For all $n \geq 1$, we have $v_n = R_{x_n}u_n$ for some sequence $(u_n)_{n \geq 1} \subset \mathbb{B}_\mathbb{E} \triangleq \{u \in \mathbb{E} \mid \|u\| < 1\}$, and $\mathbf{A}v_n = 0$ for all $n \geq 1$. Therefore, $\|p_n - x_n\|_{x_n} = \|v_n\|_{x_n} = \|u_n\| < 1$, which shows $p_n \in K$ for all $n \geq 1$ (cf. Lemma 1). Let $(u_{n_j})_{j \geq 1}$ be a converging subsequence and denote $v_{n_j} = R_{x_{n_j}}u_{n_j}$ the corresponding subsequence in $\mathcal{T}_{x_{n_j}}$. Since $\|v_{n_j}\| = \|R_{x_{n_j}}u_{n_j}\| \leq \|x_{n_j}\|e_*(u_{n_j})$, taking eventually a further subsequence (we omit the relabeling), we can assume that $(v_{n_j})_{j \geq 1}$ converges to a limit v . This limit point must obviously satisfy the condition $\mathbf{A}v = 0$ and $\lim_{j \rightarrow \infty} \|v_{n_j}\|_{x_{n_j}} = \lim_{j \rightarrow \infty} \|u_{n_j}\| \leq 1$. Hence, $\lim_{j \rightarrow \infty} p_{n_j} = p \triangleq \bar{x} + v$ has the properties $\mathbf{A}p = \mathbf{A}\bar{x} = b$ and $p \in \bar{K}$. \square

Deriving second-order optimality conditions for (Opt) is based on a perturbation argument. To define admissible perturbations at boundary points $\bar{x} \in \text{bd}(K)$, we can use Proposition 7 that motivates the definition

$$\begin{aligned} \mathcal{T}_{\bar{x}} \triangleq \{v \in \mathbb{E} \mid \exists x_n \xrightarrow{X} \bar{x}, (u_n)_{n \geq 1} \subset \mathbb{B}_\mathbb{E} \text{ s.t. } v = \lim_{n \rightarrow \infty} R_{x_n}u_n \text{ and } \mathbf{A}R_{x_n}u_n \\ = 0 \text{ for all } n \geq 1\}. \end{aligned} \tag{84}$$

This definition is a natural extension of the corresponding set already defined at interior points in (16). Indeed, if $\bar{x} \in K \cap L = X$ is taken as the base point in the construction of the variations in $\mathcal{T}_{\bar{x}}$, we can take the constant sequences $x_n \equiv \bar{x}$ and $u_n \equiv u \in \mathbb{B}_\mathbb{E}$ in (84), and obtain the definition (16).

Appendix B: Optimality conditions

Suppose x^* is a local minimizer of problem (Opt) and f is continuously differentiable at x^* . Assumption 1 guarantees that there exists a Lagrange multiplier $y^* \in \mathbb{R}^m$ such that the inclusion (8) holds. The following exact optimality conditions for the model problem (Opt) are proved in [61].

Theorem 5 [61, Theorem 3] *Let x^* be a local solution of problem (Opt) and let R be a limit operator of the sequence $(R_{x_n})_{n \geq 1}$ obtained by $x_n \rightarrow x^*$. Suppose that f is continuously differentiable at x^* . Then there exists a Lagrangian multiplier $y^* \in \mathbb{R}^m$ such that*

$$\nabla f(x^*) - \mathbf{A}^*y^* \in \bar{K}^*, \tag{85}$$

$$R(\nabla f(x^*) - \mathbf{A}^*y^*) = 0. \tag{86}$$

[61, Proposition 1] proves that the conditions in Theorem 5 are equivalent to $\nabla f(x^*) - \mathbf{A}^*y^* \in \bar{K}^*$, and $\langle \nabla f(x^*) - \mathbf{A}^*y^*, x^* \rangle = 0$.

Theorem 6 [61, Theorem 4] *Let x^* be a local minimizer of problem (Opt) and R be a limit operator of the sequence $(R_{x_n})_{n \geq 1}$ obtained by $x_n \rightarrow x^*$. Suppose that f is twice continuously differentiable at x^* . Then there exists a Lagrangian multiplier $y^* \in \mathbb{R}^m$ such that (85) and (86) hold and additionally*

$$\langle \nabla^2 f(x^*)v, v \rangle \geq 0 \quad \forall v \in \mathcal{T}_{x^*}.$$

The set \mathcal{T}_{x^*} for $x^* \in \text{bd}(\bar{X})$ is defined in (84).

Appendix C: Useful inequalities

Consider the cone K_{NN} with the standard log-barrier $h(x) = -\sum_{i=1}^n \ln(x_i)$ which has Hessian $H(x) = \text{diag}[x_1^{-2}, \dots, x_n^{-2}] = \mathbf{X}^{-2}$. Let $\mathbf{V} = \text{diag}[v_1, \dots, v_n] = \text{diag}(v)$, $z \in \mathbb{R}^n$, and $x \in K_{NN}$. Then,

$$\|\mathbf{V}z\|_\infty \leq \|\mathbf{V}z\| \leq \|v\|_x \cdot \|z\|_x^*, \tag{87}$$

$$\|\mathbf{V}z\|_\infty \leq \|v\|_x \cdot \|\mathbf{X}z\|_\infty. \tag{88}$$

The first inequality in (87) is trivial. Let us prove the second inequality. Indeed, we have

$$\begin{aligned} \|\mathbf{V}z\|^2 &= \sum_{i=1}^n (v_i z_i)^2 = \sum_{i=1}^n (v_i/x_i)^2 \cdot (x_i z_i)^2 \leq \left(\sum_{i=1}^n (v_i/x_i)^2 \right) \\ &\cdot \left(\sum_{i=1}^n (x_i z_i)^2 \right) = \langle H(x)v, v \rangle \cdot \langle [H(x)]^{-1}z, z \rangle = \|v\|_x^2 \cdot (\|z\|_x^*)^2, \end{aligned}$$

which finishes the proof of (87). For the inequality (88), we have, denoting by v/x the componentwise division of v by x ,

$$\begin{aligned} \|\mathbf{V}z\|_\infty &= \|\mathbf{V}\mathbf{X}^{-1}\mathbf{X}z\|_\infty \leq \|v/x\|_\infty \cdot \|\mathbf{X}z\|_\infty \leq \|v/x\| \\ &\cdot \|\mathbf{X}z\|_\infty = \|\mathbf{X}^{-1}v\| \cdot \|\mathbf{X}z\|_\infty = \|H(x)^{1/2}v\| \cdot \|\mathbf{X}z\|_\infty = \|v\|_x \cdot \|\mathbf{X}z\|_\infty. \end{aligned}$$

References

1. Adler, I., Monteiro, R.D.: Limiting behavior of the affine scaling continuous trajectories for linear programming problems. *Math. Program.* **50**(1–3), 29–51 (1991)
2. Agarwal, N., Allen-Zhu, Z., Bullins, B., Hazan, E., Ma, T.: Finding approximate local minima faster than gradient descent. In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1195–1199. ACM (2017)
3. Alizadeh, F., Goldfarb, D.: Second-order cone programming. *Math. program.* **95**(1), 3–51 (2003)
4. Alvarez, F., Bolte, J., Brahic, O.: Hessian Riemannian gradient flows in convex programming. *SIAM J. Control. Optim.* **43**(2), 477–501 (2004)
5. Andreani, R., Fukuda, E.H., Haeser, G., Santos, D.O., Secchin, L.D.: Optimality conditions for non-linear second-order cone programming and symmetric cone programming. *Optimization* (2019)

6. Andreani, R., Fukuda, E.H., Haeser, G., Santos, D.O., Secchin, L.D.: On the use of Jordan algebras for improving global convergence of an augmented Lagrangian method in nonlinear semidefinite programming. *Comput. Optim. Appl.* **79**(3), 633–648 (2021). <https://doi.org/10.1007/s10589-021-00281-8>
7. Andreani, R., Gómez, W., Haeser, G., Mito, L.M., Ramos, A.: On optimality conditions for nonlinear conic programming. *Math. Oper. Res.* **47**(3), 2160–2185 (2021). <https://doi.org/10.1287/moor.2021.1203>
8. Andreani, R., Haeser, G., Ramos, A., Silva, P.J.: A second-order sequential optimality condition associated to the convergence of optimization algorithms. *IMA J. Numer. Anal.* **37**(4), 1902–1929 (2017)
9. Andreani, R., Haeser, G., Viana, D.S.: Optimality conditions and global convergence for nonlinear semidefinite programming. *Math. Program.* **180**(1), 203–235 (2020). <https://doi.org/10.1007/s10107-018-1354-5>
10. Andreani, R., Secchin, L.D.: A note on the convergence of an augmented Lagrangian algorithm to second-order stationary points. *Proc. Ser. Braz. Soc. Comput. Appl. Math.* **6**(1), 1 (2018)
11. Bayer, D., Lagarias, J.: The nonlinear geometry of linear programming, ii: Legendre transform coordinates and central trajectories. *Trans. Am. Math. Soc.* **314**, 527–581 (1989)
12. Bayer, D.A., Lagarias, J.C.: The nonlinear geometry of linear programming, i: affine and projective scaling trajectories. *Trans. Am. Math. Soc.* **314**, 499–526 (1989)
13. Ben-Tal, A., Nemirovski, A.: Lectures on modern convex optimization. *Soc. Ind. Appl. Math.* (2001). <https://doi.org/10.1137/1.9780898718829>
14. Bian, W., Chen, X.: Worst-case complexity of smoothing quadratic regularization methods for non-Lipschitzian optimization. *SIAM J. Optim.* **23**(3), 1718–1741 (2013)
15. Bian, W., Chen, X.: Linearly constrained non-Lipschitz optimization for image restoration. *SIAM J. Imaging Sci.* **8**(4), 2294–2322 (2015). <https://doi.org/10.1137/140985639>
16. Bian, W., Chen, X., Ye, Y.: Complexity analysis of interior point algorithms for non-Lipschitz and nonconvex minimization. *Math. Program.* **149**(1), 301–327 (2015). <https://doi.org/10.1007/s10107-014-0753-5>
17. Birgin, E.G., Gardenghi, J.L., Martínez, J.M., Santos, S.A., Toint, P.L.: Evaluation complexity for nonlinear constrained optimization using unscaled KKT conditions and high-order models. *SIAM J. Optim.* **26**(2), 951–967 (2016). <https://doi.org/10.1137/15M1031631>
18. Birgin, E.G., Martínez, J.M.: On regularization and active-set methods with complexity for constrained optimization. *SIAM J. Optim.* **28**(2), 1367–1395 (2018). <https://doi.org/10.1137/17M1127107>
19. Birgin, E.G., Martínez, J.M.: Complexity and performance of an augmented Lagrangian algorithm. *Optim. Methods Softw.* **35**(5), 885–920 (2020). <https://doi.org/10.1080/10556788.2020.1746962>
20. Bogolubsky, L., Dvurechensky, P., Gasnikov, A., Gusev, G., Nesterov, Y., Raigorodskii, A.M., Tikhonov, A., Zhukovskii, M.: Learning supervised Pagerank with gradient-based and gradient-free optimization methods. In: Lee, D.D., Sugiyama, M., Luxburg, U.V., Guyon, I., Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, vol. 29, pp. 4914–4922. Curran Associates, Inc. (2016). [arXiv:1603.00717](https://arxiv.org/abs/1603.00717)
21. Bolte, J., Teboulle, M.: Barrier operators and associated gradient-like dynamical systems for constrained minimization problems. *SIAM J. Control. Optim.* **42**(4), 1266–1292 (2003). <https://doi.org/10.1137/S0363012902410861>
22. Bomze, I.M., Mertikopoulos, P., Schachinger, W., Staudigl, M.: Hessian barrier algorithms for linearly constrained optimization problems. *SIAM J. Optim.* **29**(3), 2100–2127 (2019)
23. Carmon, Y., Duchi, J.C., Hinder, O., Sidford, A.: Convex Until Proven Guilty: Dimension-free Acceleration of Gradient Descent on Non-convex Functions, pp. 654–663 (2017)
24. Carmon, Y., Duchi, J.C., Hinder, O., Sidford, A.: Lower bounds for finding stationary points i. *Math. Program.* (2019). <https://doi.org/10.1007/s10107-019-01406-y>
25. Carmon, Y., Duchi, J.C., Hinder, O., Sidford, A.: Lower bounds for finding stationary points ii: first-order methods. *Math. Program.* (2019). <https://doi.org/10.1007/s10107-019-01431-x>
26. Cartis, C., Gould, N., Toint, P.: Complexity bounds for second-order optimality in unconstrained optimization. *J. Complex.* **28**(1), 93–108 (2012)
27. Cartis, C., Gould, N.I., Toint, P.L.: An adaptive cubic regularization algorithm for nonconvex optimization with convex constraints and its function-evaluation complexity. *IMA J. Numer. Anal.* **32**(4), 1662–1695 (2012)
28. Cartis, C., Gould, N.I., Toint, P.L.: Universal regularization methods: varying the power, the smoothness and the accuracy. *SIAM J. Optim.* **29**(1), 595–615 (2019). <https://doi.org/10.1137/16M1106316>

29. Cartis, C., Gould, N.I.M., Toint, P.L.: Adaptive cubic regularisation methods for unconstrained optimization, part i: motivation, convergence and numerical results. *Math. Program.* **127**(2), 245–295 (2011). <https://doi.org/10.1007/s10107-009-0286-5>
30. Cartis, C., Gould, N.I.M., Toint, P.L.: Adaptive cubic regularisation methods for unconstrained optimization, part ii: worst-case function- and derivative-evaluation complexity. *Math. Program.* **130**(2), 295–319 (2011). <https://doi.org/10.1007/s10107-009-0337-y>
31. Cartis, C., Gould, N.I.M., Toint, P.L.: Second-order optimality and beyond: characterization and evaluation complexity in convexly constrained nonlinear optimization. *Found. Comput. Math.* **18**(5), 1073–1107 (2018). <https://doi.org/10.1007/s10208-017-9363-y>
32. Cartis, C., Gould, N.I.M., Toint, P.L.: Optimality of orders one to three and beyond: characterization and evaluation complexity in constrained nonconvex optimization. *J. Complex.* **53**, 68–94 (2019)
33. Cartis, C., Gould, N.I.M., Toint, P.L.: Sharp worst-case evaluation complexity bounds for arbitrary-order nonconvex optimization with inexpensive constraints. *SIAM J. Optim.* **30**(1), 513–541 (2020). <https://doi.org/10.1137/17M1144854>
34. Chandrasekaran, V., Shah, P.: Relative entropy optimization and its applications. *Math. Program.* **161**(1), 1–32 (2017). <https://doi.org/10.1007/s10107-016-0998-2>
35. Chares, R.: Cones and interior-point algorithms for structured convex optimization involving powers and exponentials. Ph.D. Thesis, UCL-Université Catholique de Louvain Louvain-la-Neuve, Belgium (2009)
36. Chen, X., Ge, D., Wang, Z., Ye, Y.: Complexity of unconstrained l_2 - l_p minimization. *Math. Program.* **143**(1), 371–383 (2014). <https://doi.org/10.1007/s10107-012-0613-0>
37. Conn, A., Gould, N., Toint, P.: Trust region methods. Society for Industrial and Applied Mathematics (2000)
38. Curtis, F.E., Robinson, D.P., Samadi, M.: A trust region algorithm with a worst-case iteration complexity of $\mathcal{O}(\epsilon^{-3/2})$ for nonconvex optimization. *Math. Program.* **162**(1–2), 1–32 (2017)
39. Curtis, F.E., Robinson, D.P., Samadi, M.: Complexity analysis of a trust funnel algorithm for equality constrained optimization. *SIAM J. Optim.* **28**(2), 1533–1563 (2018). <https://doi.org/10.1137/16M1108650>
40. Danilova, M., Dvurechensky, P., Gasnikov, A., Gorbunov, E., Guminov, S., Kamzolov, D., Shibaev, I.: Recent Theoretical Advances in Non-convex Optimization, pp. 79–163. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-00832-0_3
41. De Klerk, E.: Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications, vol. 65. Springer (2006)
42. Doikov, N., Nesterov, Y.: Minimizing uniformly convex functions by cubic regularization of newton method. *J. Optim. Theory Appl.* **189**(1), 317–339 (2021). <https://doi.org/10.1007/s10957-021-01838-7>
43. Dvurechensky, P., Safin, K., Shtern, S., Staudigl, M.: Generalized self-concordant analysis of Frank–Wolfe algorithms. In: *Mathematical Programming*, pp. 1–69 (2022)
44. Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **96**(456), 1348–1360 (2001)
45. Faraut, J., Koranyi, A.: Analysis on symmetric cones. In: , Oxford Mathematical Monographs. Oxford University Press (1994)
46. Faybusovich, L.: Several Jordan-algebraic aspects of optimization. *Optimization* **57**(3), 379–393 (2008). <https://doi.org/10.1080/02331930701523510>
47. Faybusovich, L., Lu, Y.: Jordan-algebraic aspects of nonconvex optimization over symmetric cones. *Appl. Math. Optim.* **53**(1), 67–77 (2006). <https://doi.org/10.1007/s00245-005-0835-0>
48. Fiacco, A.V., McCormick, G.P.: Nonlinear Programming: Sequential Unconstrained Minimization Techniques (reprinted by SIAM Publications in 1990). Wiley, New York (1968)
49. Foucart, S., Lai, M.J.: Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q \leq 1$. *Appl. Comput. Harmon. Anal.* **26**(3), 395–407 (2009). <https://doi.org/10.1016/j.acha.2008.09.001>
50. Ge, D., Jiang, X., Ye, Y.: A note on the complexity of l_p minimization. *Math. Program.* **129**(2), 285–299 (2011). <https://doi.org/10.1007/s10107-011-0470-2>
51. Ghadimi, S., Lan, G.: Accelerated gradient methods for nonconvex nonlinear and stochastic programming. *Math. Program.* **156**(1), 59–99 (2016). <https://doi.org/10.1007/s10107-015-0871-8>

52. Gill, P.E., Murray, W., Wright, M.H.: Practical optimization. In: Classics in Applied Mathematics. SIAM, Society for Industrial and Applied Mathematics (2019). <https://books.google.de/books?id=G-vEDwAAQBAJ>
53. Gould, N.I.M., Toint, P.L.: A note on the convergence of barrier algorithms to second-order necessary points. *Math. Program.* **85**(2), 433–438 (1999). <https://doi.org/10.1007/s101070050066>
54. Grapiglia, G.N., Yuan, Y.X.: On the complexity of an augmented Lagrangian method for nonconvex optimization. *IMA J. Numer. Anal.* **41**(2), 1546–1568 (2020). <https://doi.org/10.1093/imanum/draa021>
55. Griewank, A.: The modification of newton's method for unconstrained optimization by bounding cubic terms. Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Technical Report (1981)
56. Güler, O., Tunçel, L.: Characterization of the barrier parameter of homogeneous convex cones. *Math. Program.* **81**(1), 55–76 (1998). <https://doi.org/10.1007/BF01584844>
57. Guminov, S., Dvurechensky, P., Tupitsa, N., Gasnikov, A.: On a combination of alternating minimization and Nesterov's momentum. In: Proceedings of the 38th International Conference on Machine Learning, PMLR, vol. 139, pp. 3886–3898. PMLR (2021). <http://proceedings.mlr.press/v139/guminov21a.html>
58. Haeser, G., Liu, H., Ye, Y.: Optimality condition and complexity analysis for linearly-constrained optimization without differentiability on the boundary. *Math. Program.* **178**(1), 263–299 (2019). <https://doi.org/10.1007/s10107-018-1290-4>
59. Harmany, Z.T., Marcia, R.F., Willett, R.M.: This is spiral-tap: sparse Poisson intensity reconstruction algorithms—theory and practice. *IEEE Trans. Image Process.* **21**(3), 1084–1096 (2011)
60. Hauser, R.A., Güler, O.: Self-scaled barrier functions on symmetric cones and their classification. *Found. Comput. Math.* **2**(2), 121–143 (2002). <https://doi.org/10.1007/s102080010022>
61. He, C., Lu, Z.: A Newton-CG based barrier method for finding a second-order stationary point of nonconvex conic optimization with complexity guarantees. *SIAM J. Optim.* (2022). [arXiv:2207.05697](https://arxiv.org/abs/2207.05697)
62. Helmke, U., Moore, J.B.: Optimization and Dynamical Systems: Communications and Control Engineering. Springer, Berlin (1996)
63. Hildebrand, R.: Canonical barriers on convex cones. *Math. Oper. Res.* **39**(3), 841–850 (2014). <https://doi.org/10.1287/moor.2013.0640>
64. Hildebrand, R.: Centro-affine hypersurface immersions with parallel cubic form. *Beitr. Algebra Geom. Contrib. Algebra Geom.* **56**(2), 593–640 (2015). <https://doi.org/10.1007/s13366-014-0216-4>
65. Hinder, O., Ye, Y.: Worst-case iteration bounds for log barrier methods for problems with nonconvex constraints (2018). [arXiv:1807.00404](https://arxiv.org/abs/1807.00404)
66. Huang, J., Ma, S., Xie, H., Zhang, C.H.: A group bridge approach for variable selection. *Biometrika* **96**(2), 339–355 (2009)
67. Ji, S., Sze, K., Zhou, Z., So, A.M., Ye, Y.: Beyond convex relaxation: A polynomial-time non-convex optimization approach to network localization. In: 2013 Proceedings IEEE INFOCOM, pp. 2499–2507 (2013). <https://doi.org/10.1109/INFOCOM.2013.6567056>
68. Jia, X., Liang, X., Shen, C., Zhang, L.H.: Solving the cubic regularization model by a nested restarting Lanczos method. *SIAM J. Matrix Anal. Appl.* **43**(2), 812–839 (2022). <https://doi.org/10.1137/21M1436324>
69. Lan, G.: First-Order and Stochastic Optimization Methods for Machine Learning. Springer, Berlin (2020)
70. Laurent, M., Rendl, F.: Semidefinite programming and integer programming. *Handb. Oper. Res. Manag. Sci.* **12**, 393–514 (2005)
71. Liu, H., Yao, T., Li, R., Ye, Y.: Folded concave penalized sparse linear regression: sparsity, statistical performance, and algorithmic theory for local solutions. *Math. Program.* **166**(1), 207–240 (2017). <https://doi.org/10.1007/s10107-017-1114-y>
72. Lourenço, B.F., Fukuda, E.H., Fukushima, M.: Optimality conditions for problems over symmetric cones and a simple augmented Lagrangian method. *Math. Oper. Res.* **43**(4), 1233–1251 (2018). <https://doi.org/10.1287/moor.2017.0901>
73. Lu, Y., Yuan, Y.: An interior-point trust-region algorithm for general symmetric cone programming. *SIAM J. Optim.* **18**(1), 65–86 (2007). <https://doi.org/10.1137/040611756>
74. Martínez, J.M.: On high-order model regularization for constrained optimization. *SIAM J. Optim.* **27**(4), 2447–2458 (2017). <https://doi.org/10.1137/17M1115472>
75. Molzahn, D.K., Hiskens, I.A.: A survey of relaxations and approximations of the power flow equations. *Found. Trends Electric Energy Syst.* **4**(1–2), 1–221 (2019). <https://doi.org/10.1561/3100000012>

76. Nesterov, Y.: Lectures on Convex Optimization, Springer Optimization and Its Applications, vol. 137. Springer, Berlin (2018)
77. Nesterov, Y., Gasnikov, A., Guminov, S., Dvurechensky, P.: Primal–dual accelerated gradient methods with small-dimensional relaxation oracle. In: Optimization Methods and Software, pp. 1–28 (2020). <https://doi.org/10.1080/10556788.2020.1731747>
78. Nesterov, Y., Nemirovski, A.: Interior Point Polynomial Methods in Convex Programming. SIAM Publications, Philadelphia (1994)
79. Nesterov, Y., Polyak, B.: Cubic regularization of newton method and its global performance. Math. Program. **108**(1), 177–205 (2006). <https://doi.org/10.1007/s10107-006-0706-8>
80. Nesterov, Y.E., Todd, M.J.: Self-scaled barriers and interior-point methods for convex programming. Math. Oper. Res. **22**(1), 1–42 (1997). <https://doi.org/10.1287/moor.22.1.1>
81. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer, Berlin (2000)
82. O’Neill, M., Wright, S.J.: A log-barrier Newton–CG method for bound constrained optimization with complexity guarantees. IMA J. Numer. Anal. (2020). <https://doi.org/10.1093/imanum/drz074>
83. Renegar, J.: A Mathematical View of Interior-Point Methods in Convex Optimization. Society for Industrial and Applied Mathematics (2001). <https://doi.org/10.1137/1.9780898718812>
84. Tseng, P.: Second-order cone programming relaxation of sensor network localization. SIAM J. Optim. **18**(1), 156–185 (2007)
85. Tseng, P., Bomze, I.M., Schachinger, W.: A first-order interior-point method for linearly constrained smooth optimization. Math. Program. **127**(2), 399–424 (2011). <https://doi.org/10.1007/s10107-009-0292-7>
86. Wen, F., Chu, L., Liu, P., Qiu, R.C.: A survey on nonconvex regularization-based sparse and low-rank recovery in signal processing, statistics, and machine learning. IEEE Access **6**, 69883–69906 (2018). <https://doi.org/10.1109/ACCESS.2018.2880454>
87. Ye, Y.: On affine scaling algorithms for nonconvex quadratic programming. Math. Program. **56**(1), 285–300 (1992). <https://doi.org/10.1007/BF01580903>

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