

INTERIOR-POINT ALGORITHMS WITH FULL NEWTON STEPS FOR NONSYMMETRIC CONVEX CONIC OPTIMIZATION*

DÁVID PAPP[†] AND ANITA VARGA[‡]

Abstract. We design and analyze primal-dual, feasible interior-point algorithms (IPAs) employing full Newton steps to solve convex optimization problems in standard conic form. Unlike most nonsymmetric cone programming methods, the algorithms presented in this paper require only a logarithmically homogeneous self-concordant barrier (LHSCB) of the primal cone, but compute feasible and ε -optimal solutions to both the primal and dual problems in $\mathcal{O}(\sqrt{\nu} \log(1/\varepsilon))$ iterations, where ν is the barrier parameter of the LHSCB; this matches the best known theoretical iteration complexity of IPAs for both symmetric and nonsymmetric cone programming. The definition of the neighborhood of the central path and feasible starts ensure that the computed solutions are compatible with the dual certificates framework of (Davis and Papp, 2022). Several initialization strategies are discussed, including two-phase methods that can be applied if a strictly feasible primal solution is available, and one based on a homogeneous self-dual embedding that allows the rigorous detection of large feasible or optimal solutions. In a detailed study of a classic, notoriously difficult, polynomial optimization problem, we demonstrate that the methods are efficient and numerically reliable. Although the standard approach using semidefinite programming fails for this problem with the solvers we tried, the new IPAs compute highly accurate near-optimal solutions that can be certified to be near-optimal in exact arithmetic.

Key words. Nonsymmetric conic optimization, Interior point algorithms, Feasible method, Polynomial complexity

MSC codes. 90C51, 90C25, 90C23

1. Introduction. We propose and analyze new interior-point algorithms (IPAs) for convex conic optimization problems, that is, problems in the form

$$(1.1) \quad \left. \begin{array}{l} \inf \mathbf{c}^\top \mathbf{x} \\ \text{s.t. } A\mathbf{x} = \mathbf{b} \\ \mathbf{x} \in \mathcal{K} \end{array} \right\} \quad (\text{P}) \qquad \left. \begin{array}{l} \sup \mathbf{b}^\top \mathbf{y} \\ \text{s.t. } A^\top \mathbf{y} + \mathbf{s} = \mathbf{c} \\ \mathbf{s} \in \mathcal{K}^*, \mathbf{y} \in \mathbb{R}^m \end{array} \right\} \quad (\text{D})$$

where $\mathbf{c} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$ are given vectors, $A \in \mathbb{R}^{m \times n}$ is a given matrix with full row rank, and $\mathcal{K} \subseteq \mathbb{R}^n$ is a proper convex cone whose dual cone is denoted by \mathcal{K}^* . The only assumption we make about \mathcal{K} is that its interior, denoted by \mathcal{K}° , is the domain of a known logarithmically homogeneous self-concordant barrier function (LHSCB), see Appendix A, whose gradient and Hessian are computable efficiently. The cone need not have any other desirable properties; in particular, we do not assume that \mathcal{K} is symmetric, homogeneous, or hyperbolic. Nor do we make any assumptions about its dual.

The vast majority of the abundant literature on IPAs is focused on the special case when $\mathcal{K} = \mathcal{K}^*$ is a symmetric (self-dual and homogeneous) cone, also known in the earlier literature as a self-scaled cone [19]; many algorithms for these problems can be seen as broad generalizations of algorithms originally developed for linear programming.

***Funding:** This material is based upon work supported by the National Science Foundation under Grant No. DMS-1847865. This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-23-1-0370.

[†]North Carolina State University, Department of Mathematics. (ORCID: 0000-0003-4498-6417)
Email: dpapp@ncsu.edu.

[‡]North Carolina State University, Department of Mathematics. (ORCID: 0000-0001-7138-2921)
Email: avarga@ncsu.edu.

For example, in [34, 36], Nesterov and Todd proposed efficient IPAs for self-scaled cones, based on the existence of Nesterov-Todd scaling points; Schmieta and Alizadeh used the machinery of Jordan algebras to extend IPAs for linear optimization problems to second-order cone and semidefinite optimization problems [47]. Symmetric conic optimization has also enjoyed excellent software support for decades [51, 28, 53, 7, 5, 55], with novel applications (e.g., semidefinite programs with low-rank solutions) inspiring innovative approaches even today [20].

In recent years, there has been a growing interest in IPAs for more general, *nonsymmetric*, conic optimization [32]. This is partly due to models in contemporary applications that cannot be formulated as symmetric conic programs, such as the applications of the exponential cone and the relative entropy cone [8] in quantum information theory [21], entropy maximization problems, and risk-averse portfolio optimization [1]. It has also been established that some problems that are known to be representable as second-order cone or semidefinite optimization problems can be more efficiently handled when formulated and solved as nonsymmetric conic optimization problems; ℓ_p cones [49] and sums-of-squares cones [40] are but two examples; [9] is a compendium of additional ones.

The primary difficulty in further extending symmetric cone programming approaches, especially primal-dual IPAs, to nonsymmetric conic programs is that the previously identical primal and dual cones are now replaced by two different cones, one of which might be difficult to handle. For example, \mathcal{K}° might be the domain of an easily computable LHSCB, but the same might not hold for $(\mathcal{K}^*)^\circ$, or the known barrier for the dual cone may not be the (Fenchel) conjugate of the primal barrier, which precludes the use of certain methods. For instance, the IPAs proposed by Nesterov et al. in [36] require both the primal and dual LHSCBs, and the size of the linear systems to be solved is doubled compared to the symmetric case. Similarly, the method in [32] is a corrector-predictor type IPA for nonsymmetric conic optimization that uses information on the dual barrier function in its predictor steps.

Skajaa and Ye [49] proposed a primal-dual predictor-corrector IPA that solves a homogeneous self-dual embedding model for the simultaneous solution of (P) and (D), and it is one of the few methods that make no assumptions on the dual cone. Their original manuscript contained inaccuracies in certain proofs, which were later addressed by Papp and Yıldız [39]. This algorithm is the basis of the IPAs implemented in the Matlab package *alfonso* [41] and the Julia solver *Hypatia* [10].

A specialized IPA for sums-of-squares cones, which can be generalized word-for-word to spectrahedral cones, was proposed in [12]; the authors also show that for some spectrahedral cones this method is strictly more efficient than solving the conventional semidefinite programming formulations, although the iteration complexity of the method is somewhat higher than typical, $\mathcal{O}(\nu)$ instead of $\mathcal{O}(\sqrt{\nu})$, where ν is the barrier parameter of the LHSCB used in the algorithm.

Yet another recent approach to handling difficult dual cones is the “domain-driven” approach by Karimi and Tunçel [22].

Serrano introduced short-step potential reduction methods in his thesis [48], and these algorithms were implemented in the ECOS solver [16]; however, this is specialized to the second-order cone and the exponential cone. Mosek [28] is a highly capable commercial solver that can solve problems over the exponential cone and power cones (in addition to symmetric cones).

Tunçel et al. [54, 29] extended the concept of Nesterov-Todd directions from symmetric cones to the non-symmetric case. These ideas were combined with the analysis of Skajaa and Ye by Dahl and Andersen to propose an IPA for exponential

cone optimization in [11]. The authors did not provide the theoretical convergence and complexity results, however, they reported that the IPA performed well in practice compared to ECOS. Based on the main ideas of the IPA in [11], Badenbroek and Dahl [2] proposed an algorithm for nonsymmetric conic optimization. They used a homogeneous self-dual embedding model and proved that the IPA has polynomial complexity, although, once again, the iteration complexity is $\mathcal{O}(\nu)$ rather than $\mathcal{O}(\sqrt{\nu})$.

Our contributions. We introduce and analyze feasible, full-step, primal-dual IPAs that generate iterates in a narrow neighborhood of the central path. Unlike most of the methods cited above, they are applicable to general cones with a known LHSCB, and they generate ε -optimal primal-dual pairs without any assumptions on the dual cone. They have $\mathcal{O}(\sqrt{\nu} \log(1/\varepsilon))$ iteration complexity, where ν is the barrier parameter of the LHSCB of \mathcal{K} ; this matches the iteration complexity of commonly used IPAs for symmetric cones. (This is a strict improvement, for instance, over the algorithms in [12] and [2].)

To address cases where no feasible starting solution is known, we show that the proposed methods can be applied in conjunction with the homogeneous self-dual embedding technique found in, e.g., [14, 15, 27].

We also outline two different two-phase approaches that avoid homogenization for applications in which a primal feasible point is known, but we do not have a well-centered primal-dual pair to initialize our methods with. This is also desirable, since in most IPA literature relying on self-dual embeddings, it is an unaddressed (and seemingly fundamentally complicated) issue that the homogenizing variables may tend to very small positive values (not *a priori* bounded away from zero), meaning that there is a theoretical gap between computing an “ ε -feasible and ε -optimal” solution to the embedding problem and computing an “ ε -feasible and ε -optimal” solution to the original problem (1.1). (The former may be far from being the latter.)

Short-step methods and dual feasibility certificates. Our methods are based on neighborhoods constructed from Dikin ellipsoids (often considered “short-step” methods); see Eq. (2.1). Although for symmetric cone programming, primal-dual predictor-corrector methods operating in wider neighborhoods appear to be the most practical, and those are the ones used in most general-purpose solvers, direct generalizations of these approaches seem to be exceedingly difficult due to the aforementioned reasons.

Feasible methods operating in our small neighborhood have additional advantages beyond the good theoretical complexity and the relative simplicity of analysis. One of our motivating applications is the aforementioned optimization over low-dimensional linear images of “nice” (e.g., symmetric) high-dimensional cones such as sums-of-squares (SOS) cones, which are low-dimensional images of the positive semidefinite cone. (In other words, these are the dual cones of spectrahedral cones.) The concept of *dual certificates* introduced in [12] allows the rigorous certification of membership in such cones in lower time and substantially lower memory than what is required to construct an explicit SOS certificate. As we shall detail in Section 4, the methods proposed in this paper automatically compute dual feasibility certificates in the sense of [12], and they do so at a lower time complexity than the algorithm proposed in that paper. The same conclusion would not hold if we were working with a different neighborhood definition.

Our simple feasible IPA (Algorithm 2.1) is presented and analyzed in detail in Section 2, along with its adaptive update variants, Algorithm 2.2 and Algorithm 2.3. Some of the background material on barrier functions necessary for the analysis is summarized in Appendix A. We then discuss various initialization strategies in

Section 3. Finally, in Section 4, we present a detailed numerical example demonstrating the numerical robustness and efficiency of the method.

2. The full-step IPAs and their analysis. All variants of our algorithm are path-following IPAs that generate iterates in the same neighborhood. We start by introducing the necessary notation and then define the neighborhood of the central path.

Let \mathcal{F} denote the set of primal-dual feasible solutions:

$$\mathcal{F} = \{(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{K} \times \mathbb{R}^m \times \mathcal{K}^* \mid A\mathbf{x} = \mathbf{b}, A^\top \mathbf{y} + \mathbf{s} = \mathbf{c}\},$$

and let \mathcal{F}° be the set of strictly feasible solutions:

$$\mathcal{F}^\circ = \{(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{K}^\circ \times \mathbb{R}^m \times (\mathcal{K}^*)^\circ \mid A\mathbf{x} = \mathbf{b}, A^\top \mathbf{y} + \mathbf{s} = \mathbf{c}\}.$$

Suppose that the interior point (aka. Slater) condition is satisfied: $\mathcal{F}^\circ \neq \emptyset$. Then strong duality holds, and the optimality conditions can be formulated as

$$(OC) \quad \left. \begin{array}{l} (\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{F} \\ \mathbf{x}^\top \mathbf{s} = 0 \end{array} \right\}.$$

We assume that an LHSCB $f : \mathcal{K}^\circ \rightarrow \mathbb{R}$ with barrier parameter ν is given for the primal cone \mathcal{K} . Let $g(\mathbf{x})$ and $H(\mathbf{x})$ denote the gradient and Hessian of f at $\mathbf{x} \in \mathcal{K}^\circ$, respectively. With this notation in place, we define the central path problem, parametrized by the positive scalar τ , called the *central path parameter*, as follows:

$$(CP) \quad \left. \begin{array}{l} (\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{F}^\circ \\ \mathbf{s} + \tau g(\mathbf{x}) = \mathbf{0} \end{array} \right\}.$$

For every $\tau > 0$, the solution of (CP) is unique; these solutions form the *primal-dual central path*. As τ tends to 0, the central path approaches a solution of (OC).

In the analysis, we consider the following *neighborhood* of the central path:

$$(2.1) \quad \mathcal{N}(\eta, \tau) = \{(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{F}^\circ : \|\mathbf{s} + \tau g(\mathbf{x})\|_{\mathbf{x}}^* \leq \eta\tau\},$$

where η is the radius of the neighborhood, and τ is the central path parameter; the choice of η will be discussed later. This definition can be considered the generalization of the classic 2-norm neighborhood for linear optimization. Note the small but important departure from how a similar neighborhood is most commonly defined in the literature: it is customary to use the *normalized duality gap* $\mu := \frac{\mathbf{x}^\top \mathbf{s}}{\nu}$ in place of the parameter τ (for example [33, 48, 49]); thus, our variant generates iterates in a slightly wider neighborhood of the central path. The modified definition applied here was also examined in [35] in the context of self-scaled cones.

Given an $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{F}^\circ$, the *Newton system* associated with the central path (CP) is

$$(NS) \quad \left. \begin{array}{l} A\Delta\mathbf{x} = \mathbf{0} \\ A^\top \Delta\mathbf{y} + \Delta\mathbf{s} = \mathbf{0} \\ \tau H(\mathbf{x})\Delta\mathbf{x} + \Delta\mathbf{s} = -(\mathbf{s} + \tau g(\mathbf{x})) \end{array} \right\}.$$

Its solution $(\Delta\mathbf{x}, \Delta\mathbf{y}, \Delta\mathbf{s})$ is called the *Newton direction* at $(\mathbf{x}, \mathbf{y}, \mathbf{s})$.

We start with our basic method, which takes full Newton steps and reduces the central path parameter τ by a fixed fraction every iteration. The pseudo-code of the algorithm is shown in Algorithm 2.1. It requires an initial primal-dual iterate near the central path; Theorem 2.4 prescribes parameter values for which the convergence of the algorithm (with a suitable initial iterate) is guaranteed, and in Section 3 we discuss different strategies (two-phase variants, embeddings, etc.) for when a well-centered initial iterate is not readily available.

Algorithm 2.1 A full-step IPA for nonsymmetric cone programming

inputs: $A \in \mathbb{R}^{m \times n}$ with full row rank; $\mathbf{b} \in \mathbb{R}^m$; $\mathbf{c} \in \mathbb{R}^n$;
 an oracle for computing the Hessian H of the ν -LHSCB $f : \mathcal{K}^\circ \rightarrow \mathbb{R}$;
 an initial $\tau_0 > 0$ and $(\mathbf{x}_0, \mathbf{y}_0, \mathbf{s}_0) \in \mathcal{F}^\circ$ such that $\|\mathbf{s}_0 + \tau_0 g(\mathbf{x}_0)\|_{\mathbf{x}_0}^* \leq \eta \tau_0$;
 $\varepsilon > 0$ optimality tolerance.

parameters: neighborhood parameter $\eta \in (0, 1)$; update parameter ϑ .

output: a triplet $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{F}^\circ$ satisfying $\mathbf{x}^\top \mathbf{s} \leq \varepsilon$.

Initialize $\mathbf{x} := \mathbf{x}_0$, $\mathbf{y} := \mathbf{y}_0$, $\mathbf{s} := \mathbf{s}_0$, $\tau := \tau_0$.

while $\mathbf{x}^\top \mathbf{s} > \varepsilon$ **do**

 Solve (NS) for $(\Delta \mathbf{x}, \Delta \mathbf{y}, \Delta \mathbf{s})$.

 Set $\mathbf{x} := \mathbf{x} + \Delta \mathbf{x}$, $\mathbf{y} := \mathbf{y} + \Delta \mathbf{y}$, $\mathbf{s} := \mathbf{s} + \Delta \mathbf{s}$.

 Set $\tau := (1 - \vartheta)\tau$.

end while

return $(\mathbf{x}, \mathbf{y}, \mathbf{s})$.

The Newton system (NS) is identical to the system used for determining the corrector directions in [49] and [39], and parts of their analyses are also applicable here. Similar bounds were also used in [48]. To keep the paper self-contained, we restate these results briefly as part of Lemma 2.1 below. The main difference in our analysis (throughout the paper) is that we do not use the (normalized) duality gap μ as the central path parameter, as is customary in most short-step IPAs. Instead, a separate parameter τ is used, which in Algorithm 2.1 is prescribed to converge geometrically to zero. Therefore, we have to establish separately that the duality gap also converges to 0 and at what rate; we do so in Lemma 2.3.

The following lemma summarizes some elementary results from the literature regarding the Newton directions.

LEMMA 2.1 ([39, 48, 49]). *Let $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{N}(\eta, \tau)$ and let $(\Delta \mathbf{x}, \Delta \mathbf{y}, \Delta \mathbf{s})$ be the solution of (NS). Then*

(i) $\Delta \mathbf{x}^\top \Delta \mathbf{s} = 0$,

(ii) $\|\Delta \mathbf{x}\|_{\mathbf{x}} \leq \eta$ and $\|\Delta \mathbf{s}\|_{\mathbf{x}}^* \leq \eta \tau$,

(iii) $\|\mathbf{s} + \tau g(\mathbf{x}) + \Delta \mathbf{s}\|_{\mathbf{x}}^* \leq \eta \tau$.

Proof. (i) Using the first two equations of (NS), we get

$$0 = \Delta \mathbf{x}^\top (A^\top \Delta \mathbf{y} + \Delta \mathbf{s}) = \Delta \mathbf{y}^\top A \Delta \mathbf{x} + \Delta \mathbf{x}^\top \Delta \mathbf{s} = \Delta \mathbf{x}^\top \Delta \mathbf{s}.$$

(ii) Using the neighborhood definition, the orthogonality of $\Delta \mathbf{x}$ and $\Delta \mathbf{s}$, and the third equality of the Newton system, we can derive the upper bounds the

following way:

$$\begin{aligned}\eta^2 \tau^2 &\geq (\|\mathbf{s} + \tau g(\mathbf{x})\|_{\mathbf{x}}^*)^2 = (\|\tau H(\mathbf{x})\Delta\mathbf{x} + \Delta\mathbf{s}\|_{\mathbf{x}}^*)^2 \\ &= \tau^2 \Delta\mathbf{x}^\top H(\mathbf{x})\Delta\mathbf{x} + \Delta\mathbf{s}^\top H(\mathbf{x})^{-1}\Delta\mathbf{s} = \tau^2 \|\Delta\mathbf{x}\|_{\mathbf{x}}^2 + (\|\Delta\mathbf{s}\|_{\mathbf{x}}^*)^2.\end{aligned}$$

As both terms on the right-hand side are nonnegative, it follows that $\|\Delta\mathbf{x}\|_{\mathbf{x}} \leq \eta$ and $\|\Delta\mathbf{s}\|_{\mathbf{x}}^* \leq \eta\tau$.

(iii) Using the third equation of (NS) and the upper bound on $\|\Delta\mathbf{x}\|_{\mathbf{x}}$ from (ii), we get

$$\|\mathbf{s} + \tau g(\mathbf{x}) + \Delta\mathbf{s}\|_{\mathbf{x}}^* = \|\tau H(\mathbf{x})\Delta\mathbf{x}\|_{\mathbf{x}}^* = \tau \|\Delta\mathbf{x}\|_{\mathbf{x}} \leq \eta\tau. \quad \square$$

In our next Lemma, we show that with a suitable choice of parameter values, the iterate after a full Newton step remains in the required neighborhood $\mathcal{N}(\eta, \tau^+)$ for our chosen $\tau^+ = (1 - \vartheta)\tau$.

LEMMA 2.2. *Let $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{N}(\eta, \tau)$, let $(\Delta\mathbf{x}, \Delta\mathbf{y}, \Delta\mathbf{s})$ be the solution of (NS), and let $(\mathbf{x}^+, \mathbf{y}^+, \mathbf{s}^+) = (\mathbf{x}, \mathbf{y}, \mathbf{s}) + (\Delta\mathbf{x}, \Delta\mathbf{y}, \Delta\mathbf{s})$. If $0 < \eta \leq \frac{1}{4}$ and $\vartheta = \frac{\eta/2}{\sqrt{\nu}+1}$, then the new point satisfies $(\mathbf{x}^+, \mathbf{y}^+, \mathbf{s}^+) \in \mathcal{N}(\eta, \tau^+)$ with $\tau^+ = (1 - \vartheta)\tau$.*

Proof. Let us define the function $\varphi : \mathcal{K}^\circ \rightarrow \mathbb{R}$:

$$\varphi(\mathbf{v}) = \frac{1}{\tau}(\mathbf{s} + \Delta\mathbf{s})^\top \mathbf{v} + f(\mathbf{v}),$$

where f is the known LHSCB for \mathcal{K}° , with gradient g and Hessian H .

Then φ is self-concordant, and its Hessian is also H . Let $\mathbf{n}_\varphi(\mathbf{v})$ denote the (unconstrained) Newton step for φ at $\mathbf{v} \in \mathcal{K}^\circ$:

$$\mathbf{n}_\varphi(\mathbf{v}) = -\tau^{-1}H(\mathbf{v})^{-1}(\mathbf{s} + \Delta\mathbf{s} + \tau g(\mathbf{v})).$$

Thus,

$$\mathbf{n}_\varphi(\mathbf{x}) = -\tau^{-1}H(\mathbf{x})^{-1}(\mathbf{s} + \Delta\mathbf{s} + \tau g(\mathbf{x})) = \Delta\mathbf{x},$$

and therefore,

$$\mathbf{x}^+ = \mathbf{x} + \Delta\mathbf{x} = \mathbf{x} + \mathbf{n}_\varphi(\mathbf{x}).$$

For any $\mathbf{v} \in \mathcal{K}^\circ$,

$$\|\mathbf{n}_\varphi(\mathbf{v})\|_{\mathbf{v}} = \|\tau^{-1}H(\mathbf{v})^{-1}(\mathbf{s} + \Delta\mathbf{s} + \tau g(\mathbf{v}))\|_{\mathbf{v}} = \tau^{-1}\|\mathbf{s} + \Delta\mathbf{s} + \tau g(\mathbf{v})\|_{\mathbf{v}}^*.$$

For $\mathbf{v} = \mathbf{x}$, $\|\mathbf{n}_\varphi(\mathbf{x})\|_{\mathbf{x}} = \|\Delta\mathbf{x}\|_{\mathbf{x}} \leq \eta$, using Lemma 2.1, part (ii). As $\eta < 1$, we can use (A1):

$$(2.2) \quad \tau^{-1}\|\mathbf{s} + \Delta\mathbf{s} + \tau g(\mathbf{x}^+)\|_{\mathbf{x}^+}^* = \|\mathbf{n}_\varphi(\mathbf{x}^+)\|_{\mathbf{x}^+} \leq \left(\frac{\|\mathbf{n}_\varphi(\mathbf{x})\|_{\mathbf{x}}}{1 - \|\mathbf{n}_\varphi(\mathbf{x})\|_{\mathbf{x}}} \right)^2 \leq \frac{\eta^2}{(1 - \eta)^2}.$$

Thus,

$$\begin{aligned}\frac{1}{\tau^+} \|\mathbf{s}^+ + \tau^+ g(\mathbf{x}^+)\|_{\mathbf{x}^+}^* &= \frac{1}{1 - \vartheta} \tau^{-1} \|\mathbf{s} + \Delta\mathbf{s} + \tau^+ g(\mathbf{x}^+)\|_{\mathbf{x}^+}^* \\ &\leq \frac{1}{1 - \vartheta} \tau^{-1} \|\mathbf{s} + \Delta\mathbf{s} + \tau g(\mathbf{x}^+)\|_{\mathbf{x}^+}^* + \frac{\vartheta}{1 - \vartheta} \|g(\mathbf{x}^+)\|_{\mathbf{x}^+}^* \\ &\stackrel{(2.2), (A7)}{\leq} \frac{1}{1 - \vartheta} \frac{\eta^2}{(1 - \eta)^2} + \frac{\vartheta}{1 - \vartheta} \sqrt{\nu} \\ &= \frac{\sqrt{\nu} + 1}{\sqrt{\nu} + 1 - \eta/2} \cdot \frac{\eta^2}{(1 - \eta)^2} + \frac{\eta/2}{\sqrt{\nu} + 1 - \eta/2} \sqrt{\nu} \\ &\leq \frac{2}{2 - \eta/2} \cdot \frac{\eta^2}{(1 - \eta)^2} + \frac{\eta}{2}.\end{aligned}$$

The last expression is less than η if $\eta \leq \frac{1}{4}$ holds. \square

In our final Lemma of this section, we give bounds on the duality gap after each iteration to measure our progress toward optimality. This will help establish linear convergence and an $\mathcal{O}(\sqrt{\nu} \log(1/\varepsilon))$ iteration complexity.

LEMMA 2.3. *Let $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{N}(\eta, \tau)$ and let $(\Delta \mathbf{x}, \Delta \mathbf{y}, \Delta \mathbf{s})$ be the solution of (NS). We have the following lower and upper bounds on the duality gap after a full Newton step:*

$$\tau(\nu - \eta^2) \leq (\mathbf{x}^+)^{\top} \mathbf{s}^+ \leq \tau\nu.$$

Proof. Using (i) from Lemma 2.1,

$$(\mathbf{x}^+)^{\top} \mathbf{s}^+ = (\mathbf{x} + \Delta \mathbf{x})^{\top} (\mathbf{s} + \Delta \mathbf{s}) = \mathbf{x}^{\top} \mathbf{s} + \mathbf{s}^{\top} \Delta \mathbf{x} + \mathbf{x}^{\top} \Delta \mathbf{s}.$$

From the third equation of (NS), we have

$$\begin{aligned} \Delta \mathbf{x}^{\top} \mathbf{s} &= \Delta \mathbf{x}^{\top} \mathbf{s} + \Delta \mathbf{x}^{\top} \Delta \mathbf{s} \\ &= \Delta \mathbf{x}^{\top} \mathbf{s} - \Delta \mathbf{x}^{\top} (\mathbf{s} + \tau g(\mathbf{x}) + \tau H(\mathbf{x}) \Delta \mathbf{x}) \\ &= -\tau (\Delta \mathbf{x}^{\top} g(\mathbf{x}) + \|\Delta \mathbf{x}\|_{\mathbf{x}}^2). \end{aligned}$$

Similarly, we can express $\mathbf{x}^{\top} \Delta \mathbf{s}$ as

$$\begin{aligned} \mathbf{x}^{\top} \Delta \mathbf{s} &= -\mathbf{x}^{\top} (\mathbf{s} + \tau g(\mathbf{x}) + \tau H(\mathbf{x}) \Delta \mathbf{x}) \\ &\stackrel{(A4), (A5)}{=} -\mathbf{x}^{\top} \mathbf{s} + \tau\nu + \tau \Delta \mathbf{x}^{\top} g(\mathbf{x}). \end{aligned}$$

Combining the previous equations, we can write $(\mathbf{x}^+)^{\top} \mathbf{s}^+$ as

$$\begin{aligned} (\mathbf{x}^+)^{\top} \mathbf{s}^+ &= \mathbf{x}^{\top} \mathbf{s} - \tau \Delta \mathbf{x}^{\top} g(\mathbf{x}) - \tau \|\Delta \mathbf{x}\|_{\mathbf{x}}^2 - \mathbf{x}^{\top} \mathbf{s} + \tau\nu + \tau \Delta \mathbf{x}^{\top} g(\mathbf{x}) \\ &= \tau(\nu - \|\Delta \mathbf{x}\|_{\mathbf{x}}^2). \end{aligned}$$

Using part (ii) of Lemma 2.1, we get

$$\tau(\nu - \eta^2) \leq (\mathbf{x}^+)^{\top} \mathbf{s}^+ \leq \tau\nu. \quad \square$$

We are now ready to show that Algorithm 2.1 produces an ε -optimal solution in $\mathcal{O}(\sqrt{\nu} \log 1/\varepsilon)$ iterations.

THEOREM 2.4. *If $0 < \eta \leq \frac{1}{4}$ and $\vartheta = \frac{\eta/2}{\sqrt{\nu}+1}$, then Algorithm 2.1 produces a feasible solution to (1.1) with duality gap below $\varepsilon > 0$ in no more than $\frac{2}{\eta}(\sqrt{\nu}+1) \ln\left(\frac{\tau_0\nu}{\varepsilon}\right) + 1$ iterations.*

In particular, if the initial point satisfies $\tau_0 = \frac{\mathbf{x}_0^{\top} \mathbf{s}_0}{\nu}$, as in the initialization strategies discussed in Section 3, then the required number of iterations is

$$\mathcal{O}\left(\sqrt{\nu} \ln \frac{\mathbf{x}_0^{\top} \mathbf{s}_0}{\varepsilon}\right).$$

Proof. At the beginning of the k^{th} iteration, $\tau = \tau_k = (1 - \vartheta)^{k-1} \tau_0$. According to Lemma 2.3, we have

$$\mathbf{x}_k^{\top} \mathbf{s}_k \leq \tau_k \nu = (1 - \vartheta)^{k-1} \tau_0 \nu.$$

We would like to find the smallest k for which $(1 - \vartheta)^{k-1} \tau_0 \nu \leq \varepsilon$ holds, which is equivalent to

$$(k - 1) \ln(1 - \vartheta) + \ln(\tau_0 \nu) \leq \ln \varepsilon.$$

Using the inequality $-\ln(1 - \vartheta) \geq \vartheta$ for $\vartheta \in [0, 1)$, we can require the fulfillment of the stronger inequality

$$-(k - 1)\vartheta + \ln(\tau_0\nu) \leq \ln \varepsilon,$$

which is true if

$$k \geq \frac{1}{\vartheta} \ln \left(\frac{\tau_0\nu}{\varepsilon} \right) + 1 = \frac{2}{\eta} (\sqrt{\nu} + 1) \ln \left(\frac{\tau_0\nu}{\varepsilon} \right) + 1. \quad \square$$

2.1. Adaptive update strategies. In this section, we consider more efficient strategies for updating the parameter value τ .

After taking a full Newton step, we can find the smallest τ^+ for which the new point belongs to the neighborhood $\mathcal{N}(\eta, \tau^+)$, and set this τ^+ as the updated value of τ . (See Algorithm 2.2.) This update is only marginally more expensive than the simple update formula used in Algorithm 2.1: although it requires the computation of the Hessian of the barrier function at the new point, that Hessian is needed anyway for the next iteration. With this Hessian computed, the updated τ^+ can be calculated by solving a scalar quadratic equation; it can be explicitly expressed as

$$(2.3) \quad \tau^+ = \frac{1}{\nu - \eta^2} \left((\mathbf{x}^+)^{\top} \mathbf{s}^+ - \sqrt{((\mathbf{x}^+)^{\top} \mathbf{s}^+)^2 - (\nu - \eta^2) \left(\|\mathbf{s}^+\|_{\mathbf{x}^+}^* \right)^2} \right).$$

It follows automatically that the complexity of Algorithm 2.2 is no greater than the complexity of Algorithm 2.1, that is, Theorem 2.4 applies verbatim to our second algorithm as well.

In [12], the authors proposed an IPA (Algorithm 3.1) to determine (weighted) sums-of-squares certificates for positive polynomials with a similar update strategy. In our notation, the method of [12] defines the η -neighborhood of the central path as the set of strictly feasible solutions satisfying $\|\mathbf{s} + g(\mathbf{x})\|_{\mathbf{x}}^* \leq \eta$, reminiscent but different from both our neighborhood definition and the customary one in the literature. The different neighborhood definition considered in this paper makes it possible to improve the complexity of the algorithm: the iteration complexity of the algorithm in [12] is $\mathcal{O}(\nu)$, while we have seen that our algorithms have an $\mathcal{O}(\sqrt{\nu})$ iteration complexity.

Another known update strategy from the theory of IPAs for linear programming (LP) and linear complementarity problems (LCPs) is the *largest step method* [18, 44]. In our previous methods, τ is updated after taking the Newton step. In contrast, the largest update strategy updates $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ and τ simultaneously by finding the lowest value τ^+ for which the new point after the update remains in the neighborhood $\mathcal{N}(\eta, \tau^+)$. That is, denoting the solution of (NS) by $(\Delta \mathbf{x}_{\tau}, \Delta \mathbf{y}_{\tau}, \Delta \mathbf{s}_{\tau})$ to emphasize its dependence on τ ,

$$(2.4) \quad \tau^+ = \inf \left\{ \tau > 0 \mid \|\mathbf{s} + \Delta \mathbf{s}_{\tau} + \tau g(\mathbf{x} + \Delta \mathbf{x}_{\tau})\|_{\mathbf{x} + \Delta \mathbf{x}_{\tau}}^* \leq \eta \tau \right\}.$$

The complexity of computing or approximating the solution of (2.4) depends on the LHSCB at hand; the pseudo-code in Algorithm 2.3 is therefore formulated generically. For LPs, monotone LCPs, and weighted monotone LCPs, (2.4) can be solved in closed form for a very efficient update. For other barriers, one may solve (NS) parametrically, and if g and H have a convenient characterization, τ^+ may be approximated efficiently, without explicitly solving (NS) for different values of τ in each iteration.

In the most general case, we can use line search (starting with the theoretically guaranteed $(1 - \vartheta)\tau$ from Algorithm 2.1) to approximately solve (2.4), although this requires the repeated solution of (NS) for different values of τ .

Algorithm 2.2 Full-step IPA with a simple adaptive update strategy

inputs: $A \in \mathbb{R}^{m \times n}$ with full row rank; $\mathbf{b} \in \mathbb{R}^m$; $\mathbf{c} \in \mathbb{R}^n$;
 an oracle for computing the Hessian H of the ν -LHSCB $f : \mathcal{K}^\circ \rightarrow \mathbb{R}$;
 an initial $\tau_0 > 0$ and $(\mathbf{x}_0, \mathbf{y}_0, \mathbf{s}_0) \in \mathcal{F}^\circ$ such that $\|\mathbf{s}_0 + \tau_0 g(\mathbf{x}_0)\|_{\mathbf{x}_0}^* \leq \eta\tau_0$;
 $\varepsilon > 0$ optimality tolerance.

parameters: neighborhood parameter $\eta \in (0, 1)$; update parameter ϑ .

output: a triplet $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{F}^\circ$ satisfying $\mathbf{x}^\top \mathbf{s} \leq \varepsilon$.

Initialize $\mathbf{x} := \mathbf{x}_0$, $\mathbf{y} := \mathbf{y}_0$, $\mathbf{s} := \mathbf{s}_0$, $\tau := \tau_0$

while $\mathbf{x}^\top \mathbf{s} > \varepsilon$ **do**
 Solve (NS) for $(\Delta \mathbf{x}, \Delta \mathbf{y}, \Delta \mathbf{s})$.
 Set $\mathbf{x}^+ := \mathbf{x} + \Delta \mathbf{x}$, $\mathbf{y}^+ := \mathbf{y} + \Delta \mathbf{y}$, $\mathbf{s}^+ := \mathbf{s} + \Delta \mathbf{s}$.
 Set $\tau^+ := \min\{\tau \mid \|\mathbf{s}^+ + \tau g(\mathbf{x}^+)\|_{\mathbf{x}^+}^* \leq \eta\tau\}$.
 Set $\tau := \tau^+$, $\mathbf{x} := \mathbf{x}^+$, $\mathbf{y} := \mathbf{y}^+$, $\mathbf{s} := \mathbf{s}^+$.
end while
return $(\mathbf{x}, \mathbf{y}, \mathbf{s})$.

Each of these solution methods for (2.4) is theoretically sound, since all iterates remain in $\mathcal{N}(\eta, \tau)$, and the progress in each iteration is at least as large as in Algorithm 2.1. Therefore, the $\mathcal{O}(\sqrt{\nu})$ iteration complexity also holds, but if we have to resort to line search, that may come at an additional computational cost per iteration.

Algorithm 2.3 Full-step IPA with the largest update strategy

inputs: $A \in \mathbb{R}^{m \times n}$ with full row rank; $\mathbf{b} \in \mathbb{R}^m$; $\mathbf{c} \in \mathbb{R}^n$;
 an oracle for computing the Hessian H of the ν -LHSCB $f : \mathcal{K}^\circ \rightarrow \mathbb{R}$;
 an initial $\tau_0 > 0$ and $(\mathbf{x}_0, \mathbf{y}_0, \mathbf{s}_0) \in \mathcal{F}^\circ$ such that $\|\mathbf{s}_0 + \tau_0 g(\mathbf{x}_0)\|_{\mathbf{x}_0}^* \leq \eta\tau_0$;
 $\varepsilon > 0$ optimality tolerance.

parameters: neighborhood parameter $\eta \in (0, 1)$; update parameter ϑ .

output: a triplet $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{F}^\circ$ satisfying $\mathbf{x}^\top \mathbf{s} \leq \varepsilon$.

Initialize $\mathbf{x} := \mathbf{x}_0$, $\mathbf{y} := \mathbf{y}_0$, $\mathbf{s} := \mathbf{s}_0$, $\tau := \tau_0$

while $\mathbf{x}^\top \mathbf{s} > \varepsilon$ **do**
 Set τ^+ as the (approximate) solution of (2.4) and solve (NS) for $(\Delta \mathbf{x}, \Delta \mathbf{y}, \Delta \mathbf{s})$ using $\tau = \tau^+$.
 Set $\tau := \tau^+$, $\mathbf{x}^+ := \mathbf{x} + \Delta \mathbf{x}$, $\mathbf{y}^+ := \mathbf{y} + \Delta \mathbf{y}$, $\mathbf{s}^+ := \mathbf{s} + \Delta \mathbf{s}$.
end while
return $(\mathbf{x}, \mathbf{y}, \mathbf{s})$.

3. Initialization. Our algorithms are fundamentally *feasible IPAs*, meaning that they require an initial primal-dual strictly feasible pair in the appropriate neighborhood of the central path. In this section, we address the question of finding such an initial point. First, we consider an important special case that covers one of our motivating applications outlined in the introduction, in which initialization is trivial. Then we outline two different two-phase approaches for bounded problems in which only a

primal feasible solution is known, but not a primal-dual pair near the central path. Finally, we briefly demonstrate the application of a standard homogeneous self-dual embedding technique (see, e.g., [14] for semidefinite programming) in our setting, with no assumptions about the feasibility and boundedness of the problem.

3.1. Dual membership verification problems. Recall one of our motivating questions from the introduction: the certification of membership in the dual cone \mathcal{K}^* , without any assumption on \mathcal{K}^* other than it being the dual cone of a cone with an LHSCB. Given a vector $\mathbf{t} \in \mathbb{R}^n$, we can decide whether $\mathbf{t} \in (\mathcal{K}^*)^\circ$ by considering the following primal-dual pair of optimization problems:

$$(3.1) \quad \left. \begin{array}{ll} \min & \mathbf{t}^\top \mathbf{x} \\ \text{s.t.} & \mathbf{w}^\top \mathbf{x} = 1 \\ & \mathbf{x} \in \mathcal{K} \end{array} \right\} \quad \left. \begin{array}{ll} \max & y \\ \text{s.t.} & y\mathbf{w} + \mathbf{s} = \mathbf{t} \\ & \mathbf{s} \in \mathcal{K}^*, y \in \mathbb{R} \end{array} \right\},$$

where $\mathbf{w} \in (\mathcal{K}^*)^\circ$ is any vector of our choice from the interior of the dual cone. (Having access to such a vector \mathbf{w} is not a strong assumption; recall that given any $\mathbf{x}_0 \in \mathcal{K}^\circ$, $-g(\mathbf{x}_0) \in (\mathcal{K}^*)^\circ$ automatically.)

It is clear from the formulation that the dual problem is strictly feasible, since every sufficiently negative y satisfies $\mathbf{t} - y\mathbf{w} \in (\mathcal{K}^*)^\circ$. The primal problem is also strictly feasible if \mathcal{K} is full-dimensional, since $\mathbf{w} \in (\mathcal{K}^*)^\circ$, and therefore every $\mathbf{x}_0 \in \mathcal{K}$ can be rescaled to satisfy the only equality constraint. Hence, strong duality holds for the problems in (3.1), with attainment in both the primal and dual. Additionally, the optimal value is nonnegative if and only if $\mathbf{t} \in \mathcal{K}^*$.

Our next Lemma shows that if we have a strictly feasible primal solution \mathbf{x}_0 with $\|\nu\mathbf{w} + g(\mathbf{x}_0)\|_{\mathbf{x}_0}^* < \eta$, where $\eta < 1/4$ is the given neighborhood radius for initialization, and y_0 is small enough, then not only is $\mathbf{t} - y_0\mathbf{w} \in (\mathcal{K}^*)^\circ$, but the initial vector $(\mathbf{x}_0, y_0, \mathbf{s}_0 := \mathbf{t} - y_0\mathbf{w})$ is near the central path.

Since in this application we may choose $\mathbf{w} \in (\mathcal{K}^*)^\circ$ arbitrarily, we can initialize either variant of our IPA as follows: take any known $\mathbf{x}_0 \in \mathcal{K}^\circ$, and choose $\mathbf{w} := -g(\mathbf{x}_0)/\nu$. Now \mathbf{x}_0 is primal feasible by Eq. (A5) and $\|\nu\mathbf{w} + g(\mathbf{x}_0)\|_{\mathbf{x}_0}^* = 0 < \eta$.

LEMMA 3.1. *Let $\mathbf{x}_0 \in \mathcal{K}^\circ$ with $\mathbf{x}_0^\top \mathbf{w} = 1$ be given such that $\|\nu\mathbf{w} + g(\mathbf{x}_0)\|_{\mathbf{x}_0}^* < \eta$, where $\eta < \frac{1}{4}$ is the radius of our desired neighborhood for initialization. If*

$$y_0 \leq \mathbf{x}_0^\top \mathbf{t} - \nu \frac{\|\mathbf{t} - \mathbf{x}_0^\top \mathbf{t} \mathbf{w}\|_{\mathbf{x}_0}^*}{\eta - \|\nu\mathbf{w} + g(\mathbf{x}_0)\|_{\mathbf{x}_0}^*},$$

$\mathbf{s}_0 = \mathbf{t} - y_0\mathbf{w}$, and $\tau_0 := \frac{\mathbf{x}_0^\top \mathbf{s}_0}{\nu}$, then $(\mathbf{x}_0, y_0, \mathbf{s}_0)$ is in the $\mathcal{N}(\eta, \tau_0)$ neighborhood of the central path of (3.1), that is,

$$(3.2) \quad \|\mathbf{s}_0 + \tau_0 g(\mathbf{x}_0)\|_{\mathbf{x}_0}^* \leq \eta \tau_0,$$

and $(\mathbf{x}_0, y_0, \mathbf{s}_0) \in \mathcal{F}^\circ$.

Proof. First, observe that we can derive a lower bound on τ_0 using the given upper bound on y_0 :

$$\tau_0 = \frac{\mathbf{x}_0^\top \mathbf{s}_0}{\nu} = \frac{\mathbf{x}_0^\top (\mathbf{t} - y_0\mathbf{w})}{\nu} = \frac{\mathbf{x}_0^\top \mathbf{t} - y_0 \mathbf{x}_0^\top \mathbf{w}}{\nu} = \frac{\mathbf{x}_0^\top \mathbf{t} - y_0}{\nu} \geq \frac{\|\mathbf{t} - \mathbf{x}_0^\top \mathbf{t} \mathbf{w}\|_{\mathbf{x}_0}^*}{\eta - \|\nu\mathbf{w} + g(\mathbf{x}_0)\|_{\mathbf{x}_0}^*}.$$

Using the triangle inequality and this bound, the inequality (3.2) follows:

$$\begin{aligned}
\frac{1}{\tau_0} \|s_0 + \tau_0 g(\mathbf{x}_0)\|_{\mathbf{x}_0}^* &= \frac{1}{\tau_0} \|\mathbf{t} - y_0 \mathbf{w} + \tau_0 g(\mathbf{x}_0)\|_{\mathbf{x}_0}^* \\
&= \frac{1}{\tau_0} \|\mathbf{t} - (\mathbf{x}_0^\top \mathbf{t} - \nu \tau_0) \mathbf{w} + \tau_0 g(\mathbf{x}_0)\|_{\mathbf{x}_0}^* \\
&\leq \frac{1}{\tau_0} \|\mathbf{t} - \mathbf{x}_0^\top \mathbf{t} \mathbf{w}\|_{\mathbf{x}_0}^* + \|\nu \mathbf{w} + g(\mathbf{x}_0)\|_{\mathbf{x}_0}^* \\
&\leq \frac{\eta - \|\nu \mathbf{w} + g(\mathbf{x}_0)\|_{\mathbf{x}_0}^*}{\|\mathbf{t} - \mathbf{x}_0^\top \mathbf{t} \mathbf{w}\|_{\mathbf{x}_0}^*} \|\mathbf{t} - \mathbf{x}_0^\top \mathbf{t} \mathbf{w}\|_{\mathbf{x}_0}^* + \|\nu \mathbf{w} + g(\mathbf{x}_0)\|_{\mathbf{x}_0}^* \leq \eta.
\end{aligned}$$

This also means that $\frac{s_0}{\tau_0} \in \mathcal{B}^*(-g(\mathbf{x}_0), \eta)$. As $-g(\mathbf{x}_0) \in (\mathcal{K}^*)^\circ$ and $\eta < 1$, it follows that $s_0 \in (\mathcal{K}^*)^\circ$ holds. Therefore, $(\mathbf{x}_0, y_0, s_0) \in \mathcal{F}^\circ$. \square

3.2. A two-phase approach for problems with a single equality constraint.

We now turn to a more general family of problems, which admits a simple two-phase strategy with strong complexity guarantees. Assume that our problem at hand is of the form (3.1), as in the previous section, except that \mathbf{w} is a given vector from $(\mathcal{K}^*)^\circ$ that does not necessarily satisfy the norm inequality condition of Lemma 3.1. Also assume that we have a known strictly primal feasible solution \mathbf{x}_0 .

In this case, an initial primal-dual iterate near the central path can be found by applying Algorithm 2.1 to the following pair of auxiliary problems:

$$(3.3) \quad \left. \begin{array}{ll} \min & \nu \mathbf{w}^\top \mathbf{x} \\ \text{s.t.} & \mathbf{e}^\top \mathbf{x} = 1 \\ & \mathbf{x} \in \mathcal{K} \end{array} \right\} \quad \left. \begin{array}{ll} \max & y \\ \text{s.t.} & y \mathbf{e} + \mathbf{s} = \nu \mathbf{w} \\ & \mathbf{s} \in \mathcal{K}^*, y \in \mathbb{R} \end{array} \right\},$$

where $\mathbf{e} = -\frac{g(\mathbf{x}_0)}{\nu}$.

LEMMA 3.2. *The problems (3.3) have strong duality with attained optimal solutions. The optimal value is positive, and there are strictly feasible solutions $(\mathbf{x}, y, \mathbf{s})$ with $y = 0$.*

Proof. The strong duality claim follows from our assumptions on \mathbf{x}_0 and \mathbf{w} : \mathbf{x}_0 is a Slater point for the primal problem by definition, and $(0, \mathbf{s}) = (0, \nu \mathbf{w})$ is a Slater point for the dual problem because $\mathbf{w} \in (\mathcal{K}^*)^\circ$. This also shows that 0 is a suboptimal attainable value in the dual, therefore the optimal y is positive. \square

We can compute a strictly feasible solution $(\mathbf{x}, 0, \mathbf{s})$ to (3.3) as follows. By construction, (3.3) is an instance of (3.1) studied in the previous subsection, for which $(\mathbf{x}_0, 0, -g(\mathbf{x}_0))$ is a strictly feasible solution on the central path, hence we can apply Lemma 3.1 to initialize Algorithm 2.1 and proceed to compute an approximate optimal solution to (3.3), stopping at $y = 0$ along the way.

The only (small, but technical) missing detail is that when we apply either of our methods to (3.3), it is likely that we will never hit $y = 0$ exactly; instead, the last full Newton step will take us from $y < 0$ to $y > 0$. We can easily fix this by taking a shorter Newton step in the last iteration: if $y < 0$ but $y + \Delta y > 0$, update $(\mathbf{x}, y, \mathbf{s})$ to $(\mathbf{x} + \alpha \Delta \mathbf{x}, 0, \mathbf{s} + \alpha \Delta \mathbf{s})$ with the step size α that yields $y + \alpha \Delta y = 0$, and we expect that this will always work in practice. However, parts of our analysis do not apply if any of our Newton steps are damped. In the remainder of this subsection we show that if we solve the Phase 1 problem (3.3) by applying Algorithm 2.1 with a sufficiently small neighborhood $\tilde{\eta} < \eta$, then even with a damped final step, the computed feasible

solution $(\mathbf{x}, 0, \mathbf{s})$ yields a well-centered initial iterate for the solution of (3.1). This claim is made precise by our next Lemma.

LEMMA 3.3. *Assume that a strictly feasible solution to (3.3) of the form $(\mathbf{x}, y, \mathbf{s}) = (\mathbf{x}, 0, \mathbf{s})$ was computed with a modification of Algorithm 2.1, using a neighborhood radius of $\tilde{\eta} \leq 1/10$, where after a series of full Newton steps, the last iterate was arrived at with a damped Newton step of length $\alpha \in (0, 1]$. Then \mathbf{x}/μ (with $\mu = \mathbf{s}^\top \mathbf{x}/\nu$) is a primal feasible solution to (3.1) that also satisfies the inequality*

$$\|\nu \mathbf{w} + g(\mathbf{x}/\mu)\|_{\mathbf{x}/\mu}^* < 1/4.$$

Proof. Because our last step is damped, we can no longer assume that the algorithm returns a point within some $\mathcal{N}(\tilde{\eta}, \tau)$ neighborhood, but it can be shown that the final iterate is in some $\mathcal{N}(2\tilde{\eta}, \tau)$ neighborhood. (See Lemma B.1 in Appendix B.) That is, $(\mathbf{x}, 0, \mathbf{s})$ is a strictly feasible solution satisfying

$$\|\mathbf{s} + \tau g(\mathbf{x})\|_{\mathbf{x}}^* \leq 2\tilde{\eta}\tau.$$

Using the properties of LHSCBs and substituting $\mathbf{s} = \nu \mathbf{w}$, we can equivalently write this inequality as

$$\left\| \nu \mathbf{w} + \frac{\tau}{\mu} g(\mathbf{x}/\mu) \right\|_{\mathbf{x}/\mu}^* \leq \frac{2\tilde{\eta}\tau}{\mu}.$$

From this, we have the bound

$$(3.4) \quad \|\nu \mathbf{w} + g(\mathbf{x}/\mu)\|_{\mathbf{x}/\mu}^* \leq \frac{2\tau}{\mu} \tilde{\eta} + \left| 1 - \frac{\tau}{\mu} \right| \|g(\mathbf{x}/\mu)\|_{\mathbf{x}/\mu}^* \stackrel{(A7)}{=} \frac{2\tau}{\mu} \tilde{\eta} + \left| 1 - \frac{\tau}{\mu} \right| \sqrt{\nu},$$

and we need to show that the right-hand side is below $1/4$ for our choice of $\tilde{\eta}$, using suitable bounds on τ/μ . Lemma 2.3 is *not* applicable, as it assumes full Newton steps. An analogous argument, however, yields the following bounds for all step sizes (see Lemma B.2 in Appendix B):

$$(3.5) \quad \left(1 - \frac{\tilde{\eta}^2}{\nu} \right) \tau \leq \mu \leq \frac{1}{1 - \vartheta} \tau.$$

From (3.5) and $\nu \geq 1$, we get that for every $\tilde{\eta} < 1/4$, $\tau/\mu \leq \frac{\nu}{\nu - \tilde{\eta}^2}$ and $|1 - \tau/\mu| \leq \vartheta$. Thus, (3.4) can be further bounded as

$$(3.6) \quad \|\nu \mathbf{w} + g(\mathbf{x}/\mu)\|_{\mathbf{x}/\mu}^* \leq 2 \frac{\tilde{\eta}\nu}{\nu - \tilde{\eta}^2} + \vartheta \sqrt{\nu} = 2 \frac{\tilde{\eta}\nu}{\nu - \tilde{\eta}^2} + \frac{\tilde{\eta}/2}{\sqrt{\nu} + 1} \sqrt{\nu},$$

which in turn is strictly below $1/4$ for every $\nu \geq 1$ if $\tilde{\eta} \leq 1/10$, completing our proof. \square

In summary, after obtaining a solution $(\mathbf{x}, 0, \mathbf{s})$ to our Phase 1 problem (3.3) as prescribed in Lemma 3.3, we are ready to apply Lemma 3.1 to initialize either of our algorithms using the computed \mathbf{x}/μ and solve our original problem.

Aside from a single damped Newton step in its last iteration, Phase 1 is Algorithm 2.1 applied to the problem (3.3), meaning that the number of Phase 1 iterations is $\mathcal{O}(\sqrt{\nu})$. Therefore, the two-phase approach described in this section has a total iteration complexity of order $\mathcal{O}(\sqrt{\nu})$.

3.3. Transforming a bounded problem with a known feasible solution to one with a single inhomogeneous equation. So far we have assumed that our primal problem (P) in standard form has only a single linear equality constraint. Next, we consider solving a general conic program in the form of (1.1), under two additional assumptions that ensure a finite infimum: suppose that we are given a strictly (primal) feasible vector \mathbf{x}_0 and that the feasible region is *a priori* bounded. In this section, we show that under these assumptions, every problem can be reduced to an instance of the problem studied in the previous section.

Consider an instance of our general conic problem (1.1), suppose that a vector $\mathbf{z} \in (\mathcal{K}^*)^\circ$ is readily available, perhaps in the form of $\mathbf{z} = -g(\bar{\mathbf{x}})$ for some $\bar{\mathbf{x}} \in \mathcal{K}^\circ$, and that the feasible region of the primal problem, $\{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{b}, \mathbf{x} \in \mathcal{K}\}$, is *a priori* known to be bounded explicitly. The latter implies that we have a redundant inequality $(0 \leq) \mathbf{z}^\top \mathbf{x} < U$ for some known upper bound U . Now consider the following representation of the feasible region:

$$(3.7) \quad \begin{pmatrix} \mathbf{z}^\top & 1 & 1 \\ \mathbf{z}^\top & 1 & -U \\ A & \mathbf{0} & -\mathbf{b} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \xi \\ \zeta \end{pmatrix} = \begin{pmatrix} U+1 \\ 0 \\ \mathbf{0} \end{pmatrix}, \quad (\mathbf{x}, \xi, \zeta) \in \mathcal{K} \times \mathbb{R}_+ \times \mathbb{R}_+.$$

On the one hand, it is clear that if \mathbf{x} is a feasible (resp., strictly feasible) solution to the original problem, then there exists a feasible (resp., strictly feasible) solution of the form $(\mathbf{x}, \xi, 1)$ of (3.7). (Choose $\xi := U - \mathbf{z}^\top \mathbf{x}$.) The converse also holds: from the first two equations in (3.7), we see that every feasible solution to (3.7) satisfies $\zeta = 1$, which in turn implies $A\mathbf{x} = \mathbf{b}$ from the last equation. Therefore, the \mathbf{x} component of every (\mathbf{x}, ξ, ζ) that satisfies (3.7) is automatically a feasible solution to the original primal problem.

In this manner, every bounded primal problem can be rewritten in an equivalent form in which there is only a single inhomogeneous equality constraint. The homogeneous equations can then be incorporated into the cone constraint: defining

$$\mathcal{K}_{A,\mathbf{b},U} := \left\{ (\mathbf{x}, \xi, \zeta) \in \mathcal{K} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid \begin{pmatrix} \mathbf{z}^\top & 1 & -U \\ A & \mathbf{0} & -\mathbf{b} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \xi \\ \zeta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\},$$

we can write our original primal optimization problem as

$$(3.8) \quad \left. \begin{array}{l} \min \mathbf{c}^\top \mathbf{x} \\ \text{s.t. } \mathbf{z}^\top \mathbf{x} + \xi + \zeta = U + 1 \\ (\mathbf{x}, \xi, \zeta) \in \mathcal{K}_{A,\mathbf{b},U} \end{array} \right\}.$$

The reformulation (3.8) has two crucial properties: it has only a single, inhomogeneous, equality constraint besides the convex cone constraint, and the coefficient vector in that constraint belongs to the interior of the dual cone: $(\mathbf{z}, 1, 1) \in (\mathcal{K}^*)^\circ \times \mathbb{R}_{++} \times \mathbb{R}_{++} \subseteq (\mathcal{K}_{A,\mathbf{b},U}^*)^\circ$. Moreover, if we have an efficiently computable LHSCB for \mathcal{K} , say $f_{\mathcal{K}}$, then we also have one readily available for the new cone $\mathcal{K}_{A,\mathbf{b},U}$: a suitable LHSCB is simply the restriction of $f_{\mathcal{K}} + f_{\mathbb{R}_+^2}$ to the linear subspace in the definition of $\mathcal{K}_{A,\mathbf{b},U}$. Finally, any strictly feasible \mathbf{x}_0 for our original primal problem extends automatically to a strictly feasible $(\mathbf{x}, \xi, \zeta) = (\mathbf{x}_0, 1, U - \mathbf{z}^\top \mathbf{x}_0)$ of (3.8).

This completes our transformation of our problem to an instance that can be solved using the two-phase method of the previous section.

3.4. A direct two-phase approach for general equality constraints. In this subsection, we show that another well-known initialization technique from the theory of IPAs can also be applied if a strictly feasible primal solution is available and the feasible set is bounded. This technique requires no transformations, additional variables, or modifications to the underlying cone.

For a suitably defined Phase 1 problem, we will follow the central path “backwards” by iteratively increasing the parameter τ until we get close enough to the analytic center of the feasible region, at which point a strictly feasible primal-dual solution can be defined in the $\mathcal{N}(\eta, \tau)$ neighborhood of the central path of our original problem (1.1).

For this initialization strategy, the Phase 1 problem can be defined as follows:

$$(3.9) \quad \left. \begin{array}{ll} \min & -g(\mathbf{x}_0)^\top \mathbf{x} \\ \text{s.t.} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \in \mathcal{K} \end{array} \right\} \quad \left. \begin{array}{ll} \max & \mathbf{b}^\top \mathbf{y} \\ \text{s.t.} & A^\top \mathbf{y} + \mathbf{s} = -g(\mathbf{x}_0) \\ & \mathbf{s} \in \mathcal{K}^* \end{array} \right\},$$

where \mathbf{x}_0 is the known strictly feasible primal solution, that is, $A\mathbf{x}_0 = \mathbf{b}$ and $\mathbf{x}_0 \in \mathcal{K}^\circ$. It is immediate that the triplet $(\mathbf{x}, \mathbf{y}, \mathbf{s}) = (\mathbf{x}_0, \mathbf{0}, -g(\mathbf{x}_0))$ is a strictly feasible primal-dual solution to (3.9), moreover, this point is on the central path of this problem. The $\tilde{\eta}$ -neighborhood of the central path (recall Eq. (2.1)) is now defined as the set of strictly feasible points satisfying

$$(3.10) \quad \|-g(\mathbf{x}_0) - A^\top \mathbf{y} + \tau g(\mathbf{x})\|_{\mathbf{x}}^* \leq \tilde{\eta}\tau.$$

This inequality obviously holds for our starting triplet with $\tau = 1$ and every $\tilde{\eta} \geq 0$.

Thus, we may run Algorithm 2.1 with two modifications: first, instead of decreasing τ in every iteration, we increase it to move “backwards” on the central path. The second change is in the stopping criterion: we increase τ until we find ourselves in the neighborhood of the central path of our *original* problem; that is, until

$$\|\mathbf{c} - A^\top \mathbf{y} + \tau g(\mathbf{x})\|_{\mathbf{x}}^* \leq \eta\tau$$

is satisfied for some $\eta \leq 1/4$, meaning that we have a suitable starting point in the required neighborhood of the central path (CP) of our original problem. Intuitively, Phase 1 terminates because all central paths meet at the analytic center; the precise argument is below.

First, we need to show that the new iterates remain in the neighborhood of the Phase 1 central path, even if we increase the value of τ instead of decreasing it.

LEMMA 3.4. *Let $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{N}(\tilde{\eta}, \tau)$, let $(\Delta\mathbf{x}, \Delta\mathbf{y}, \Delta\mathbf{s})$ be the solution of (NS), and let $(\mathbf{x}^+, \mathbf{y}^+, \mathbf{s}^+) = (\mathbf{x}, \mathbf{y}, \mathbf{s}) + (\Delta\mathbf{x}, \Delta\mathbf{y}, \Delta\mathbf{s})$. If $0 < \tilde{\eta} \leq \frac{1}{3}$ and $\vartheta = \frac{\tilde{\eta}/2}{\sqrt{\nu+1}}$, then the new point satisfies $(\mathbf{x}^+, \mathbf{y}^+, \mathbf{s}^+) \in \mathcal{N}(\tilde{\eta}, \tau^+)$ with $\tau^+ = (1 + \vartheta)\tau$.*

The proof is essentially identical to the proof of Lemma 2.2; we omit the details.

As we assumed that the feasible set is bounded, the analytic \mathbf{x}_{ac} center of the feasible region exists, and the sequence of the vectors \mathbf{x} generated by increasing τ and taking full Newton steps converges to a neighborhood of \mathbf{x}_{ac} , inside of which the norm of the inverse of the barrier Hessian, $\|H(\mathbf{x})^{-1}\|$, is uniformly bounded for all \mathbf{x} . Therefore, using the results of Lemma 3.4, we can iterate our Phase 1 algorithm with a neighborhood radius $\tilde{\eta} < 1/4$ until τ is large enough to satisfy $\frac{1}{\tau}\|\mathbf{c} + g(\mathbf{x}_0)\|_{\mathbf{x}}^* < \eta - \tilde{\eta}$, meaning that

$$\|-g(\mathbf{x}_0) - A^\top \mathbf{y} + \tau g(\mathbf{x})\|_{\mathbf{x}}^* \leq \tilde{\eta}\tau \implies \|\mathbf{c} - A^\top \mathbf{y} + \tau g(\mathbf{x})\|_{\mathbf{x}}^* \leq 1/4\tau,$$

at which point we can switch to Phase 2, running either variant of our IPA with $\eta = 1/4$ as before.

To describe the path switching more precisely: Phase 1 terminates after finite many steps, and if it ends at $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{s}})$, with a central path parameter τ , then we can initialize Phase 2 with the triplet

$$(\mathbf{x}, \mathbf{y}, \mathbf{s}) = (\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \mathbf{c} - A^\top \tilde{\mathbf{y}})$$

and the same value of τ . This triplet is automatically a primal-dual strictly feasible solution, since by assumption, $\left\| \frac{\mathbf{c} - A^\top \tilde{\mathbf{y}}}{\tau} + g(\mathbf{x}) \right\|_{\mathbf{x}}^* \leq 1/4$, and therefore $\mathbf{s} = \mathbf{c} - A^\top \tilde{\mathbf{y}} \in (\mathcal{K}^*)^\circ$, and it is also in the $\mathcal{N}(\eta, \tau)$ neighborhood of the central path of the original problem, as required.

3.5. Homogeneous self-dual embedding. In the most general case, when the feasibility and boundedness of primal and dual problems are unclear, we may apply a homogeneous self-dual embedding technique [14, 15, 27], in conjunction with either variant of our algorithm.

Assume that $\bar{\mathbf{x}} \in \mathcal{K}^\circ$, $\bar{\mathbf{s}} \in (\mathcal{K}^*)^\circ$ and $\bar{\mathbf{y}} \in \mathbb{R}^m$ are given. (We denote these initial iterates with bars rather than the previously used $(\mathbf{x}_0, \mathbf{y}_0, \mathbf{s}_0)$ to emphasize that they are not feasible points, but arbitrary points in the interiors of their respective cones. We may choose, e.g., $\bar{\mathbf{y}} = \mathbf{0}$ and $\bar{\mathbf{s}} = -g(\bar{\mathbf{x}})$ with any $\bar{\mathbf{x}} \in \mathcal{K}^\circ$.) Let $\bar{\mathbf{b}} = \mathbf{b} - A\bar{\mathbf{x}}$, $\bar{\mathbf{c}} = \mathbf{c} - A^\top \bar{\mathbf{y}} - \bar{\mathbf{s}}$ and $\bar{\mathbf{z}} = \mathbf{c}^\top \bar{\mathbf{x}} - \mathbf{b}^\top \bar{\mathbf{y}} + 1$, and let

$$(3.11) \quad G = \begin{pmatrix} \mathbf{0} & A & -\mathbf{b} & \bar{\mathbf{b}} \\ -A^\top & \mathbf{0} & \mathbf{c} & -\bar{\mathbf{c}} \\ \mathbf{b}^\top & -\mathbf{c}^\top & 0 & \bar{\mathbf{z}} \\ -\bar{\mathbf{b}}^\top & \bar{\mathbf{c}}^\top & -\bar{\mathbf{z}} & 0 \end{pmatrix}.$$

With this notation in place, the homogeneous self-dual (HSD) model can be formulated as follows:

$$(3.12) \quad \left. \begin{aligned} & \min (\bar{\mathbf{x}}^\top \bar{\mathbf{s}} + 1)\theta \\ & \text{s.t. } G \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \\ \xi \\ \theta \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \mathbf{s} \\ \kappa \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ 0 \\ -\bar{\mathbf{x}}^\top \bar{\mathbf{s}} - 1 \end{pmatrix} \\ & (\mathbf{x}, \xi) \in \mathcal{K} \times \mathbb{R}_+, (\mathbf{s}, \kappa) \in \mathcal{K}^* \times \mathbb{R}_+, \mathbf{y} \in \mathbb{R}^m \end{aligned} \right\}.$$

The following theorem summarizes the well-known properties of the HSD model:

THEOREM 3.5. [14, 15, 27] *The HSD model (3.12) is self-dual, and the point $(\mathbf{x}, \mathbf{y}, \mathbf{s}, \xi, \theta, \kappa) = (\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}, 1, 1, 1)$ is a strictly feasible solution. The optimum value is 0, and if $\xi > 0$ at an optimal solution, then $(\mathbf{x}/\xi, \mathbf{y}/\xi, \mathbf{s}/\xi)$ is an optimal solution to the original primal-dual pair of problems. If $\xi = 0$, the problem is either infeasible, unbounded, or the optimal duality gap is not zero.*

The central path problem for (3.12), parametrized by the path parameter $\tau > 0$,

can be defined analogously to (CP) as follows:

$$(3.13) \quad \left. \begin{aligned} G \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \\ \xi \\ \theta \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \mathbf{s} \\ \kappa \\ 0 \end{pmatrix} &= \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ 0 \\ -\bar{\mathbf{x}}^\top \bar{\mathbf{s}} - 1 \end{pmatrix} \\ \mathbf{s} + \tau g(\mathbf{x}) &= \mathbf{0} \\ \kappa - \frac{\tau}{\xi} &= 0 \\ (\mathbf{x}, \xi) \in \mathcal{K}^\circ \times \mathbb{R}_{++}, (\mathbf{s}, \kappa) \in (\mathcal{K}^*)^\circ \times \mathbb{R}_{++}, \mathbf{y} \in \mathbb{R}^m \end{aligned} \right\}.$$

It can be assumed without loss of generality that $\bar{\mathbf{s}} = -g(\bar{\mathbf{x}})$. With this assumption, it is easy to show that the strictly feasible solution given in Theorem 3.5 is well-centered.

LEMMA 3.6. *Assume that $\bar{\mathbf{s}} = -g(\bar{\mathbf{x}})$ holds. Then the strictly feasible solution $(\mathbf{x}, \mathbf{y}, \mathbf{s}, \xi, \theta, \kappa) = (\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}, 1, 1, 1)$ is on the central path (3.13).*

Proof. The equations of (3.13) are satisfied for $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}, 1, 1, 1)$ and $\tau = 1$. \square

As the problem (3.12) is self-dual, we apply Newton's method directly to this problem (rather than explicitly formulating the dual instance with separate dual variables) to determine the search directions. The Newton system is the following:

$$(3.14) \quad \left. \begin{aligned} G \begin{pmatrix} \Delta \mathbf{y} \\ \Delta \mathbf{x} \\ \Delta \xi \\ \Delta \theta \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \Delta \mathbf{s} \\ \Delta \kappa \\ 0 \end{pmatrix} &= \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ 0 \\ 0 \end{pmatrix} \\ \tau H(\mathbf{x}) \Delta \mathbf{x} + \Delta \mathbf{s} &= -\mathbf{s} - \tau g(\mathbf{x}) \\ \frac{\tau}{\xi^2} \Delta \xi + \Delta \kappa &= -\kappa + \frac{\tau}{\xi} \end{aligned} \right\}.$$

Either variant of our IPA can be applied to solve (3.12), but instead of (NS), we need to solve (3.14) to determine the search directions. The analysis can be adapted easily for the new Newton system. Introduce the new variables

$$\tilde{\mathbf{x}} = \begin{pmatrix} \mathbf{x} \\ \xi \end{pmatrix} \text{ and } \tilde{\mathbf{s}} = \begin{pmatrix} \mathbf{s} \\ \kappa \end{pmatrix}$$

and their respective cones: $\tilde{\mathbf{x}} \in \tilde{\mathcal{K}} := \mathcal{K} \times \mathbb{R}_+$ and $\tilde{\mathbf{s}} \in \tilde{\mathcal{K}}^* := \mathcal{K}^* \times \mathbb{R}_+$. Given a ν -LHSCB f for \mathcal{K} , the extended cone $\tilde{\mathcal{K}}$ is the domain of the $(\nu + 1)$ -LHSCB $\tilde{f}(\tilde{\mathbf{x}}) = f(\mathbf{x}) - \ln \xi$. We denote the gradient of \tilde{f} by \tilde{g} .

It follows from the skew-symmetry of G

that $\Delta \mathbf{x}^\top \Delta \mathbf{s} + \Delta \xi \Delta \kappa = 0$, that is, $\Delta \tilde{\mathbf{x}}^\top \Delta \tilde{\mathbf{s}} = 0$, analogously to the first claim of Lemma 2.1. With this orthogonality result in place, the rest of the arguments in Section 2 leading up to Theorem 2.4 can be repeated verbatim, with $\tilde{\mathbf{x}}, \tilde{\mathbf{s}}, \tilde{g}$, etc. playing the roles of $\mathbf{x}, \mathbf{s}, g$, etc., to show that Algorithm 2.1 with the Newton system (3.14) finds an ε -optimal solution to (3.12) in $\mathcal{O}(\sqrt{\nu} \log \frac{\tilde{\mathbf{x}}^\top \tilde{\mathbf{s}}}{\varepsilon})$ iterations.

Algorithms 2.2 and 2.3 can be adapted similarly. In the definition of τ^+ , we need to use the new variables $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{s}}$, and the local norm induced by the Hessian of $\tilde{f}(\tilde{\mathbf{x}})$.

It is important to note that in this manner we can only expect to get ε -optimal solutions to the HSD-model (3.12), that is, a strictly feasible solution to (3.12) satisfying $\mathbf{x}^\top \mathbf{s} + \xi \kappa < \varepsilon$, which may not yield an $\hat{\varepsilon}$ -optimal solution to the original primal-dual

problem pair for an *a priori* known $\hat{\varepsilon}$ even in the case when the original problem has optimal solutions, in the absence of any *a priori* positive lower bound on the optimal value of the homogenizing parameter ξ .

Different stopping criteria for the HSD model (3.12) were analyzed in several papers, see [17, 30, 43, 36], to name a few. In our implementation we have used the criteria proposed by Pólik and Terlaky [43], which are theoretically sound stopping criteria based on a generalization of the “Approximate Farkas Lemma” of Todd and Ye [52] to conic optimization. To be able to apply their results here, we need to prove that their assumption [43, (3.1)]

$$(3.15) \quad \kappa\xi \geq (1 - \beta)\theta$$

is satisfied for all iterates of our algorithms with a fixed constant $\beta \in (0, 1)$.

LEMMA 3.7. *Assume that $\bar{\mathbf{s}} = -g(\bar{\mathbf{x}})$ holds. Then the Pólik-Terlaky condition (3.15) is satisfied throughout all algorithm variants with $\beta := 1 - (1 - \eta)\omega < 1$, where $\omega = \min_k \frac{\tau_k}{\tau_{k-1}}$, where τ_k is the value of τ at the end of the k^{th} iteration, and the minimum is taken over all iterations k .*

Proof. Let $(\mathbf{x}, \mathbf{y}, \mathbf{s}, \xi, \theta, \kappa)$ be the iterate in the k^{th} iteration (immediately after the Newton step) and $\mu = (\mathbf{x}^\top \mathbf{s} + \xi\kappa)/(\nu + 1)$ be the corresponding normalized duality gap. As the iterates generated by the IPA remain in the neighborhood $\mathcal{N}(\eta, \tau_k)$ in the k^{th} iteration, we have

$$\begin{aligned} \eta^2(\tau_k)^2 &\geq \|\tilde{\mathbf{s}} + \tau_k \tilde{g}(\tilde{\mathbf{x}})\|_{\tilde{\mathbf{x}}}^* = \\ &= \begin{pmatrix} \mathbf{s} + \tau_k g(\mathbf{x}) \\ \kappa - \tau_k/\xi \end{pmatrix}^\top \begin{pmatrix} H(\mathbf{x})^{-1} & 0 \\ 0 & \xi^2 \end{pmatrix} \begin{pmatrix} \mathbf{s} + \tau_k g(\mathbf{x}) \\ \kappa - \tau_k/\xi \end{pmatrix} = \\ &= (\|\mathbf{s} + \tau_k g(\mathbf{x})\|_{\mathbf{x}}^*)^2 + \xi^2(\kappa - \tau_k/\xi)^2. \end{aligned}$$

Therefore, $|\kappa\xi - \tau_k| \leq \eta\tau_k$, implying $(1 - \eta)\tau_k \leq \kappa\xi$.

From (3.12) and the skew-symmetry of G it follows that $\theta = \frac{\mathbf{x}^\top \mathbf{s} + \kappa\xi}{\bar{\mathbf{x}}^\top \bar{\mathbf{s}} + 1}$. Using our assumption $\bar{\mathbf{s}} = -g(\bar{\mathbf{x}})$ and (A6), we have $\bar{\mathbf{x}}^\top \bar{\mathbf{s}} = \nu$. Therefore, $\mu = \theta$, that is, the value of the variable θ equals to the normalized duality gap μ in every iteration.

Finally, from Lemma 2.3, it follows that $\mu \leq \tau_{k-1}$ immediately after the Newton step (but before the step that updates τ).

Combining these bounds, we obtain the following bound in every iteration:

$$\kappa\xi \geq (1 - \eta)\tau_k = (1 - \eta)\frac{\tau_k}{\tau_{k-1}}\tau_{k-1} \geq (1 - \eta)\omega\mu = (1 - \eta)\omega\theta = (1 - \beta)\theta. \quad \square$$

4. Application: Sums-of-squares lower bounds and dual nonnegativity certificates. We illustrate the efficiency and numerical stability of our algorithms in our motivating application from Section 1: computing moment-sums-of-squares lower bounds and exact sum-of-squares certificates for positive polynomials. We compare our IPAs applied directly to a nonsymmetric cone programming formulation of the problem to state-of-the-art semidefinite programming (SDP) solvers applied to an equivalent, but much larger, SDP formulation of the same problem.

Let us briefly recall the relevant theory: given a polynomial $f \in \mathbb{R}[\mathbf{x}]$ and a basic, closed, and bounded semialgebraic set Δ of the form

$$(4.1) \quad \Delta = \{\mathbf{x} \in \mathbb{R}^n \mid g_i(\mathbf{x}) \geq 0, i = 1, \dots, m\},$$

with $g_i \in \mathbb{R}[\mathbf{x}]$, the minimum of f over Δ is clearly certified to be at least γ if $f - \gamma$ can be written as follows:

$$(4.2) \quad f(\mathbf{x}) - \gamma = \sigma_0(\mathbf{x}) + \sum_{i=1}^m g_i(\mathbf{x})\sigma_i(\mathbf{x}),$$

where each $\sigma_i \in \mathbb{R}[\mathbf{x}]$ is a sum of perfect squares. Under an additional assumption on the representation (4.1) of Δ , called the *Archimedean property* [6, Sec. 3.4.4], the following converse also holds. Let SOS_d be the (finite dimensional) cone of polynomials of degree at most d that can be written as a sum of perfect squares, define the degree- d truncated quadratic module as

$$\mathcal{M}_d(\mathbf{g}) := \left\{ \sigma_0(\mathbf{x}) + \sum_{i=1}^m g_i(\mathbf{x})\sigma_i(\mathbf{x}) \mid \sigma_0 \in SOS_d, \sigma_i \in SOS_{d-\deg g_i} (i = 1, \dots, m) \right\},$$

and let

$$(4.3) \quad \gamma_d^* := \sup\{\gamma \mid f - \gamma \in \mathcal{M}_d(\mathbf{g})\}.$$

Putinar's Positivstellensatz [45] states that

$$\lim_{d \rightarrow \infty} \gamma_d^* = \min_{\mathbf{x} \in \Delta} f(\mathbf{x}).$$

This provides a recipe for computing a convergent series of lower bounds (called *SOS lower bounds*) on the global minimum of the (not necessarily convex) polynomial f by solving a series of increasingly large convex optimization problems (known as the *SOS hierarchy*).

Under the Archimedean assumption, the convergence of the sequence $(\gamma_d^*)_{d=1,2,\dots}$ is “generically” finite in a sense precisely defined in [37], but might be infinite even if f is univariate [50], and in practice, the degree necessary to certify a sharp bound may be much higher than the degrees of f and g_i [3]. On the other hand, using high-degree polynomials σ_i in the representation (4.2) may lead to numerical problems; this can be mitigated by representing the polynomials in an interpolant basis instead of the more commonly employed monomial basis [38].

The set $\mathcal{M}_d(\mathbf{g})$ is by definition a finite-dimensional convex cone, but optimization over it is not obviously tractable. In particular, no tractable LHSCB is known for this set. It is, however, a projected spectrahedron [6, Chapter 6], which means that linear optimization problems over it (such as the problem (4.3) defining γ_d^*) can be formulated as SDPs [42], but only at the cost of substantially increasing the number of optimization variables. For example, in the univariate case, SOS_d is a $(d+1)$ -dimensional cone, but its semidefinite representation involves $(d+1) \times (d+1)$ real symmetric matrices, effectively squaring the number of optimization variables. As another illustration, consider the case when m is large: the value of m has no bearing on the dimension of $\mathcal{M}_d(\mathbf{g})$, but the SDP representation grows linearly with m .

Despite the lack of tractable LHSCBs, we can employ certain nonsymmetric conic IPAs to compute SOS lower bounds and optimality certificates, because tractable LHSCBs are known for the dual cone $\mathcal{M}_d(\mathbf{g})^*$; see, e.g., [40]. The sequence of lower bounds obtained by optimizing over $\mathcal{M}_d(\mathbf{g})^*$, particularly in connection with SDP, is also referred to as Lasserre's *moment-SOS hierarchy* [23]. To summarize this approach, we need some additional notation. Let $\mathbb{R}_{n,d}$ be the space of n -variate polynomials of

degree at most d . The dual cone of $\mathcal{M}_d(\mathbf{g})$, sometimes called the *pseudo-moment cone*, is defined as

$$\mathcal{M}_d(\mathbf{g})^* = \{\ell \in \mathbb{R}_{n,d}^* \mid \ell(\sigma) \geq 0 \ \forall \sigma \in \mathcal{M}_d(\mathbf{g})\},$$

where $\mathbb{R}_{n,d}^*$ is the space of linear functionals on $\mathbb{R}_{n,d}$. (Since we are working with finite dimensional spaces, all linear functionals are automatically bounded.) If $\mathcal{M}_d(\mathbf{g})$ is a proper cone, then by strong duality, γ_d^* is also the optimal value of the dual problem of (4.3):

$$(4.4) \quad \inf\{\ell(f) \mid \ell(1) = 1, \ell \in \mathcal{M}_d(\mathbf{g})^*\}.$$

Because our cones are finite dimensional, the polynomials f and 1 can be represented by their coefficient vectors $\mathbf{f} \in \mathbb{R}^N$ and $\mathbf{1} \in \mathbb{R}^N$, respectively, in a fixed (ordered) basis; similarly, each $\ell \in \mathcal{M}_d(\mathbf{g})^*$ can be represented by a vector $\boldsymbol{\lambda} \in \mathbb{R}^N$ describing its action on the same ordered basis. In this manner, the dual definition (4.4) of γ_d^* can be written as follows:

$$(4.5) \quad \left. \begin{array}{l} \min \mathbf{f}^\top \boldsymbol{\lambda} \\ \text{s.t. } \mathbf{1}^\top \boldsymbol{\lambda} = 1 \\ \boldsymbol{\lambda} \in \mathcal{M}_d(\mathbf{g})^* \end{array} \right\}.$$

Conventionally, this problem would be solved as an SDP, using the semidefinite representation of $\mathcal{M}_d(\mathbf{g})^*$, a well-known spectrahedron [23, 24]. However, (4.5) is also an optimization problem of the form studied in Section 3.2: a standard form conic optimization problem with a single equality constraint, whose coefficient vector $\mathbf{1}$ belongs to $\mathcal{M}_d(\mathbf{g})^\circ$ for every d . Thus, we can circumvent the SDP and compute γ_d^* with our IPAs instead, solving (4.5) directly as written. And as we mentioned in Section 1, the numerical vectors $\boldsymbol{\lambda}$ computed with our IPAs are *dual nonnegativity certificates* in the sense of [12, 13], meaning that an exact, rational sums-of-squares representation of $f - \hat{\gamma}_d$ in the form (4.2) can be computed *in closed-form* from the computed numerical approximation $\hat{\gamma}_d$ of γ_d^* and the computed near-optimal $\boldsymbol{\lambda}$. The details of this certification process are beyond the scope of this short summary, but it suffices to say here that by [12, Thm. 2.4], if $(\boldsymbol{\lambda}, \gamma, \mathbf{f} - \gamma \mathbf{1}) \in \mathcal{N}(1, \tau)$ (for any $\tau > 0$), then $\boldsymbol{\lambda}$ is a certificate proving the nonnegativity of $f - \gamma$ (without having to construct an explicit SOS representation to verify this fact). Therefore, we may stop our IPAs at any time, and the current iterate $\boldsymbol{\lambda}$ can be used to certify the current bound, as long as the numerical errors remain small enough that the iterates (which theoretically stay within $\mathcal{N}(1/4, \tau)$) remain at least in $\mathcal{N}(1, \tau)$. We underscore that the same would not be true if we used a different neighborhood definition in our IPAs.

We shall illustrate all of the above using a seemingly innocuous, but surprisingly challenging instance.

EXAMPLE 4.1 ([50, 4]). *Let $m = n = 1$, $f(x) = 1 - x^2$, and $g_1(x) = (1 - x^2)^3$. Then $\Delta = [-1, 1]$ and $\min_{x \in \Delta} f(x) = 0$, the optimizers are $x = \pm 1$. The quadratic module associated with $\{g_1\}$ is Archimedean:*

$$(1 - x^2)^2 + (x(-2 + x^2))^2 + (1 - x^2)^3 = 2 - x^2 \in \mathcal{M}_6(\mathbf{g}).$$

The exact bounds γ_d^ from the Lasserre hierarchy are not known for this instance, but a simple degree-parity argument shows that this problem does not have finite convergence, meaning that although $\lim_{d \rightarrow \infty} \gamma_d^* = 0$, the value of every γ_d^* is strictly negative.*

More involved analysis reveals that to reach $\gamma_d^ \geq -\varepsilon$, we need a degree d of order $\Omega(\varepsilon^{-1/2})$ at least [50, Thm. 4]. Below, we present numerical evidence for a new conjecture: the exact values of the SOS lower bounds in this instance are given by $\gamma_{2d}^* = -1/(d(d-2))$.*

Table 1 shows numerical results comparing SOS lower bounds computed by solving (4.5) using our IPAs, along with the solutions of the equivalent SDP formulations computed with SeDuMi and Mosek. To improve the conditioning of the formulations, all polynomials were represented as Chebyshev interpolants, both in the nonsymmetric formulations and in the equivalent SDPs, as previously proposed in, e.g., [26, 38].

Both IPAs were implemented with the largest update strategy (Algorithm 2.3), using line search. Since (4.5) has a single equality constraint with the coefficient vector in the interior of the dual cone, the two-phase method proposed in Subsection 3.2 can be applied to determine a well-centered and strictly feasible starting point for the feasible IPA. For Example 4.1, this approach was highly effective and the first phase always produced a suitable starting point in 13–17 iterations, virtually independent of the problem size.

To avoid constructing a strictly feasible starting point, we also implemented the self-dual embedding technique. As shown by our numerical results, the feasible variant was more effective in both running time and the quality of the obtained lower bounds.

The experiments were carried out on a MacBook Pro with an Apple M3 CPU and 24 GB of unified memory. The algorithms were implemented in MATLAB R2024b. We formulated the SDP models using YALMIP version R20230622 [25], and solved with Mosek version 10.2.7 [28] and SeDuMi 1.3.7. [51]. The timing results refer to solver times (excluding the times YALMIP took to process the models and call the solvers).

Results. SeDuMi reports numerical errors from $d = 60$, and although Mosek and our IPAs all claim to have found near-optimal, near-feasible solutions, there is a large disagreement between the bounds from $d = 80$ onwards. As detailed below, the reported “optimal” values from both SDP solvers are incorrect from $d \geq 80$; in contrast, the feasible 2-phase IPA and its HSD variant both yield high-quality bounds up to approximately $d = 600$.

Note that in Table 1 we reported *dual* objective values, as these are supposed to be lower bounds on the optimal values of (4.3) and (4.5), that is, they ought to be valid global lower bounds on γ_d^* , and thus on f .

It is easy to verify (rigorously, in rational arithmetic) that the computed primal solutions λ are indeed strictly feasible solutions, after rescaling them to satisfy the only equality constraint exactly. This also yields a certified, rational *upper* bound on the true value of γ_d^* . For example, in the $d = 80$ case, it can be certified using the solution returned by our 2-phase feasible IPA that $-1/\gamma_d^* \leq 1520.000032$. This confirms that the SDP lower bounds cannot be sharp.

In Table 2, we compare the running times of each of the solvers. Although the incorrect bounds make the comparisons somewhat moot, it is worthwhile to point out that solving the nonsymmetric conic formulations of these problems scales better than their SDP counterparts, underscoring another key motivation behind the development of nonsymmetric conic solvers [9, 41].

5. Concluding remarks. Most convex cones \mathcal{K} whose membership problem is not known to be NP-hard have the property that either \mathcal{K} or its dual \mathcal{K}^* has a known LHSCB whose gradient and Hessian can be computed efficiently. This means that the algorithms presented in this paper are among the most generally applicable second-

	2-phase IPA		IPA with HSD		Mosek		SeDuMi	
d	dObj	-1/dObj	dObj	-1/dObj	dObj	-1/dObj	dObj	-1/dObj
20	-1.250000E-02	79.999838	-1.250000E-02	80.000021	-1.250001E-02	79.999954	-1.250001E-02	79.999964
40	-2.777800E-03	359.998865	-2.777780E-03	360.000020	-2.777784E-03	359.999181	-2.777830E-03	359.993258
60	-1.190500E-03	839.998546	-1.190480E-03	840.000042	-1.190548E-03	839.949616	-1.190552E-03	839.946371
80	-6.580000E-04	1,519.998800	-6.578950E-04	1,520.000021	-7.230688E-04	1,382.994260	-6.825633E-04	1,465.065584
100	-4.170000E-04	2,399.997850	-4.166670E-04	2,399.999972	-4.385459E-04	2,280.263024	-4.558204E-04	2,193.846524
120	-2.870000E-04	3,479.990060	-2.873560E-04	3,479.999717	-4.202670E-04	2,379.439737	-3.166395E-04	3,158.165674
140	-2.100000E-04	4,759.984770	-2.100840E-04	4,759.999279	-4.115541E-04	2,429.814209	-3.682441E-04	2,715.590012
160	-1.600000E-04	6,239.979070	-1.602560E-04	6,239.998055	-3.657808E-04	2,733.877776	-1.711278E-04	5,843.585905
180	-1.260000E-04	7,919.962130	-1.262630E-04	7,919.996061	-3.366643E-04	2,970.317910	-1.429270E-04	6,996.578673
200	-1.020000E-04	9,799.935480	-1.020410E-04	9,799.993284	-3.508363E-04	2,850.332192	-1.156350E-04	8,647.900722
400	-2.530000E-05	39,599.555200	-2.525260E-05	39,599.827255	-1.518160E-04	6,586.921010	-6.482170E-05	15,426.932648
600	-1.120000E-05	89,398.411500	-1.118580E-05	89,399.223922	-7.604590E-05	13,149.952857	-7.689500E-06	130,047.467326
800	-6.280000E-06	159,191.174000	-6.281400E-06	159,200.068639	-4.913750E-05	20,351.055711	1.306050E-05	-76566.747062
1000	-4.020000E-06	248,972.289000	-4.016060E-06	249,000.230609	-2.195120E-05	45,555.596049	9.723700E-06	-102841.510947
1200	-2.790000E-06	358,686.431000	-2.787040E-06	358,803.168207	—	—	6.180700E-06	-161793.971557
1400	-2.050000E-06	488,307.483000	-2.046640E-06	488,604.678218	—	—	2.964000E-06	-337381.916329
1600	-1.570000E-06	637,951.582000	-1.566220E-06	638,478.164902	—	—	—	—
1800	-1.240000E-06	807,073.298000	-1.236920E-06	808,458.887837	—	—	—	—
2000	-1.000000E-06	996,751.362000	-1.001130E-06	998,874.189340	—	—	—	—

TABLE 1

Near-optimal dual objective function values (numerical moment-SOS lower bounds) for the polynomial optimization problem in Example 4.1 in support of our conjecture that $\gamma_{2d}^* = -1/(d(d-2))$. The “2-phase IPA” method is Alg. 2.3 with the two-phase variant outlined in Sec. 3.2; “IPA with HSD” is Alg. 2.3 in conjunction with the self-dual embedding of Sec. 3.5. The Mosek and SeDuMi columns represent solutions computed using the standard SOS SDP formulation (representing all polynomials in the Chebyshev interpolant basis). Missing values indicate insufficient memory. Both IPAs show excellent agreement with our conjecture up to approximately $d = 600$ and reasonable agreement up to $d = 2000$; Mosek and SeDuMi are incorrect from $d = 80$ onwards.

	2-phase IPA		IPA with HSD		Mosek		SeDuMi	
d	Iterations	Time	Iterations	Time	Iterations	Time	Iterations	Time
20	85	0.19	74	0.12	14	0.01	21	0.48
40	126	0.28	134	0.19	20	0.02	28	0.23
60	149	0.42	146	0.30	20	0.05	33	0.39
80	160	0.49	151	0.56	21	0.13	31	0.43
100	170	0.62	153	0.52	28	0.33	32	0.93
120	163	0.82	169	0.61	28	0.54	35	2.26
140	164	0.87	175	0.75	23	0.76	32	3.14
160	170	1.08	188	1.00	28	1.36	41	6.33
180	171	1.92	191	1.37	22	1.71	41	10.58
200	171	2.19	209	2.02	25	2.30	41	14.65
400	314	15.19	349	16.95	16	16.34	32	140.51
600	353	36.84	383	41.19	24	108.39	31	664.20
800	357	64.71	405	127.67	22	324.22	37	2398.18
1000	397	118.35	465	225.34	24	869.12	46	7203.89
1200	422	192.70	490	335.51	—	—	42	13669.90
1400	436	302.84	527	532.86	—	—	40	24508.19
1600	442	447.44	505	732.90	—	—	—	—
1800	406	527.35	515	867.49	—	—	—	—
2000	464	815.18	499	1,004.33	—	—	—	—

TABLE 2

Total running times (in seconds) and numbers of iteration for the computations presented in Table 1. For the 2-phase IPA, the iterations of both phases are included. Our IPAs were run until failure due to numerical errors, resulting in a higher than usual number of iterations. Regardless, for larger values of d , the time and memory complexity of the nonsymmetric IPAs scale better than SDP, matching the theoretical predictions.

order methods for conic optimization. This does not come at a price of inefficiency either, as the iteration complexity of the methods matches those of the commonly used IPAs for symmetric cone programming.

The HSD variant of the method is the most general one, applicable to any problem whose primal or dual feasibility status is unknown. We opted for the less commonly used extended form of HSD, as it is more directly adaptable to feasible methods and comes with theoretically sound and quantitative stopping criteria for the detection of large optimal solutions. A drawback of using HSD is that its complexity can only be described in terms of the achieved tolerance for the embedded model, but in the absence of any *a priori* positive lower bound on the optimal value of the homogenizing parameter ξ , a near-feasible and near-optimal solution to the embedded model may not yield a similarly near-feasible and near-optimal solution to the original problem even in the ideal case, when the original problem has an optimal solution of small norm.

In contrast, the initialization approaches studied in Sections 3.1 and 3.2 come with polynomial-time performance guarantees, and yield near-optimal and very nearly feasible solutions. In our computations, the number of Phase 1 iterations was practically constant, which means that initialization was not a significant overhead. In addition, in our experiments, we found the feasible two-phase method to be numerically more reliable than HSD.

More detailed comparisons of the different variants of these IPAs, practical performance enhancements, and numerical comparisons to other nonsymmetric cone programming algorithms are the subject of a separate study.

REFERENCES

- [1] A. AHMADI-JAVID, *Entropic Value-at-Risk: A new coherent risk measure*, Journal of Optimization Theory and Applications, 155 (2012), pp. 1105–1123, <https://doi.org/10.1007/s10957-011-9968-2>.
- [2] R. BADENBROEK AND J. DAHL, *An algorithm for nonsymmetric conic optimization inspired by MOSEK*, Optimization Methods and Software, 37 (2022), pp. 1027–1064, <https://doi.org/10.1080/10556788.2021.1882457>.
- [3] L. BALDI AND B. MOURRAIN, *On the effective Putinar’s Positivstellensatz and moment approximation.*, Mathematical Programming, 200 (2023), pp. 71–103, <https://doi.org/10.1007/s10107-022-01877-6>.
- [4] L. BALDI AND L. SLOT, *Degree bounds for Putinar’s Positivstellensatz on the hypercube*, SIAM Journal on Applied Algebra and Geometry, 8 (2024), pp. 1–25, <https://doi.org/10.1137/23M1555430>.
- [5] S. J. BENSON AND Y. YE, *DSDP—software for semidefinite programming*, ACM Transactions on Mathematical Software, 34 (2008), p. 16, <https://doi.org/10.1145/1356052.1356057>.
- [6] G. BLEKHERMAN, P. A. PARRILO, AND R. R. THOMAS, eds., *Semidefinite optimization and convex algebraic geometry*, SIAM, Philadelphia, PA, 2013, <https://doi.org/10.1137/1.9781611972290>.
- [7] B. BORCHERS, *CSDP, a C library for semidefinite programming*, Optimization Methods and Software, 11-2 (1999), pp. 613–623, <https://doi.org/10.1080/10556789908805765>.
- [8] V. CHANDRASEKARAN AND P. SHAH, *Relative entropy relaxations for signomial optimization*, SIAM Journal on Optimization, 26 (2016), pp. 1147–1173, <https://doi.org/10.1137/140988978>.
- [9] C. COEY, L. KAPELEVICH, AND J. P. VIELMA, *Solving natural conic formulations with Hypatia.jl*, INFORMS Journal on Computing, 34 (2022), pp. 2686–2699, <https://doi.org/10.1287/ijoc.2022.1202>.
- [10] C. COEY, L. KAPELEVICH, AND J. P. VIELMA, *Performance enhancements for a generic conic interior point algorithm*, Mathematical Programming Computation, 15 (2023), pp. 53–101, <https://doi.org/10.1007/s12532-022-00226-0>.
- [11] J. DAHL AND E. D. ANDERSEN, *A primal-dual interior-point algorithm for nonsymmetric exponential-cone optimization*, Mathematical Programming, 194 (2022), pp. 341–370, <https://doi.org/10.1007/s101070220022>.

- [//doi.org/10.1007/s10107-021-01631-4](https://doi.org/10.1007/s10107-021-01631-4).
- [12] M. M. DAVIS AND D. PAPP, *Dual certificates and efficient rational sum-of-squares decompositions for polynomial optimization over compact sets*, SIAM Journal on Optimization, 32 (2022), pp. 2461–2492, <https://doi.org/10.1137/21M1422574>.
 - [13] M. M. DAVIS AND D. PAPP, *Rational dual certificates for weighted sums-of-squares polynomials with boundable bit size*, Journal of Symbolic Computation, 121 (2024), p. 102254, <https://doi.org/10.1016/j.jsc.2023.102254>.
 - [14] E. DE KLERK, C. ROOS, AND T. TERLAKY, *Initialization in semidefinite programming via a self-dual skew-symmetric embedding*, Operations Research Letters, 20 (1997), pp. 213–221, [https://doi.org/10.1016/S0167-6377\(97\)00011-4](https://doi.org/10.1016/S0167-6377(97)00011-4).
 - [15] E. DE KLERK, C. ROOS, AND T. TERLAKY, *Infeasible start semidefinite programming algorithms via self-dual embeddings*, in Topics in Semidefinite and Interior Point Methods, H. Wolkowicz and P. M. Pardalos, eds., vol. 18 of Fields Institute Communications, American Mathematical Society, Providence, RI, 1998, pp. 215–236, <https://doi.org/10.1090/fic/018/15>.
 - [16] A. DOMAHIDI, E. K.-W. CHU, AND S. P. BOYD, *ECOS: An SOCP solver for embedded systems*, in 2013 European Control Conference (ECC), IEEE, 2013, pp. 3071–3076, <https://doi.org/10.23919/ECC.2013.6669541>.
 - [17] R. M. FREUND, *On the behavior of the homogeneous self-dual model for conic convex optimization*, Mathematical Programming, 106 (2006), pp. 527–545, <https://doi.org/10.1007/s10107-005-0667-3>.
 - [18] C. C. GONZAGA, *The largest step path following algorithm for monotone linear complementarity problems*, Mathematical Programming, 76 (1997), pp. 309–332, <https://doi.org/10.1007/BF02614443>.
 - [19] O. GÜLER, *Barrier functions in interior point methods*, Mathematics of Operations Research, 21 (1996), pp. 860–885, <https://doi.org/10.1287/moor.21.4.860>.
 - [20] S. HABIBI, M. KOČVARA, AND M. STINGL, *Lorraine – an interior-point solver for low-rank semidefinite programming*, Optimization Methods and Software, 39 (2024), pp. 1185–1215, <https://doi.org/10.1080/10556788.2023.2250522>.
 - [21] K. HE, J. SAUNDERSON, AND H. FAWZI, *QICS: Quantum Information Conic Solver*, arXiv preprint arXiv:2410.17803, (2024), <https://doi.org/10.48550/arXiv.2410.17803>.
 - [22] M. KARIMI AND L. TUNÇEL, *Domain-Driven Solver (DDS) version 2.1: a MATLAB-based software package for convex optimization problems in domain-driven form*, Mathematical Programming Computation, 16 (2024), pp. 37–92, <https://doi.org/10.1007/s12532-023-00248-2>.
 - [23] J. B. LASSERRE, *Global optimization with polynomials and the problem of moments*, SIAM Journal on Optimization, 11 (2001), pp. 796–817, <https://doi.org/10.1137/S1052623400366802>.
 - [24] M. LAURENT, *Sums of squares, moment matrices, and optimization over polynomials*, in Emerging Applications of Algebraic Geometry, M. Putinar and S. Sullivant, eds., Springer, 2008, pp. 157–270, https://doi.org/10.1007/978-0-387-09686-5_7.
 - [25] J. LÖFBERG, *YALMIP : A Toolbox for Modeling and Optimization in MATLAB*, in In Proceedings of the CACSD Conference, Taipei, Taiwan, 2004, <https://doi.org/10.1109/CACSD.2004.1393890>.
 - [26] J. LÖFBERG AND P. A. PARRILO, *From coefficients to samples: a new approach to SOS optimization*, in 43rd IEEE Conference on Decision and Control, vol. 3, IEEE, 2004, pp. 3154–3159, <https://doi.org/10.1109/CDC.2004.1428957>.
 - [27] Z.-Q. LUO, J. F. STURM, AND S. ZHANG, *Conic convex programming and self-dual embedding*, Optimization Methods and Software, 14 (2000), pp. 169–218, <https://doi.org/10.1080/10556780008805800>.
 - [28] MOSEK APS, *MOSEK documentation and API reference (version 10.2)*, 2024, <https://www.mosek.com/documentation/>.
 - [29] T. MYKLEBUST AND L. TUNÇEL, *Interior-point algorithms for convex optimization based on primal-dual metrics*, arXiv preprint arXiv:1411.2129, (2014), <https://doi.org/10.48550/arXiv.1411.2129>.
 - [30] S. NALDI AND R. SINN, *Conic programming: Infeasibility certificates and projective geometry*, Journal of Pure and Applied Algebra, 225 (2021), p. 106605, <https://doi.org/10.1016/j.jpaa.2020.106605>.
 - [31] Y. NESTEROV, *Lectures on Convex Optimization*, Springer, 2nd ed., 2001, <https://doi.org/10.1007/978-3-319-91578-4>.
 - [32] Y. NESTEROV, *Towards non-symmetric conic optimization*, Optimization Methods and Software, 27 (2012), pp. 893–917, <https://doi.org/10.1080/10556788.2011.567270>.
 - [33] Y. NESTEROV AND A. NEMIROVSKII, *Interior-point polynomial algorithms in convex programming*, SIAM, 1994, <https://doi.org/10.1137/1.9781611970791>.
 - [34] Y. NESTEROV AND M. J. TODD, *Self-scaled barriers and interior-point methods for convex*

- programming*, Mathematics of Operations Research, 22 (1997), pp. 1–42, <https://doi.org/10.1287/moor.22.1.1>.
- [35] Y. NESTEROV AND M. J. TODD, *Primal-dual interior-point methods for self-scaled cones*, SIAM Journal on Optimization, 8 (1998), pp. 324–364, <https://doi.org/10.1137/S1052623495290209>.
 - [36] Y. NESTEROV, M. J. TODD, AND Y. YE, *Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems*, Mathematical Programming, 84 (1999), pp. 227–267, <https://doi.org/10.1007/s10107980009a>.
 - [37] J. NIE, *Optimality conditions and finite convergence of Lasserre’s hierarchy*, Mathematical Programming, Series A, 146 (2014), pp. 97–121, <https://doi.org/10.1007/s10107-013-0680-x>.
 - [38] D. PAPP, *Semi-infinite programming using high-degree polynomial interpolants and semidefinite programming*, SIAM Journal on Optimization, 27 (2017), pp. 1858–1879, <https://doi.org/10.1137/15M1053578>.
 - [39] D. PAPP AND S. YILDIZ, *On “A homogeneous interior-point algorithm for nonsymmetric convex conic optimization”*, arXiv preprint arXiv:1712.00492, (2017), <https://doi.org/10.48550/arXiv.1712.00492>.
 - [40] D. PAPP AND S. YILDIZ, *Sum-of-squares optimization without semidefinite programming*, SIAM Journal on Optimization, 29 (2019), pp. 822–851, <https://doi.org/10.1137/17M1160124>.
 - [41] D. PAPP AND S. YILDIZ, *Alfonso: Matlab package for nonsymmetric conic optimization*, INFORMS Journal on Computing, 34 (2022), pp. 11–19, <https://doi.org/10.1287/ijoc.2021.1058>.
 - [42] P. A. PARRILO, *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*, PhD thesis, California Institute of Technology, May 2000, <https://doi.org/10.7907/2K6Y-CH43>.
 - [43] I. PÓLIK AND T. TERLAKY, *New stopping criteria for detecting infeasibility in conic optimization*, Optimization Letters, 3 (2009), pp. 187–198, <https://doi.org/10.1007/s11590-008-0100-y>.
 - [44] F. A. POTRA, *Weighted complementarity problems—a new paradigm for computing equilibria*, SIAM Journal on Optimization, 22 (2012), pp. 1634–1654, <https://doi.org/10.1137/110837310>.
 - [45] M. PUTINAR, *Positive polynomials on compact semi-algebraic sets*, Indiana University Mathematics Journal, 42 (1993), pp. 969–984, <https://doi.org/10.1512/iumj.1993.42.42045>.
 - [46] J. RENEGAR, *A mathematical view of interior-point methods in convex optimization*, SIAM, 2001, <https://doi.org/10.1137/1.9780898718812>.
 - [47] S. H. SCHMIETA AND F. ALIZADEH, *Extension of primal-dual interior point algorithms to symmetric cones*, Mathematical Programming Ser. A, 96 (2003), pp. 409–438, <https://doi.org/10.1007/s10107-003-0380-z>.
 - [48] S. A. SERRANO, *Algorithms for unsymmetric cone optimization and an implementation for problems with the exponential cone*, Stanford University, 2015.
 - [49] A. SKAJAA AND Y. YE, *A homogeneous interior-point algorithm for nonsymmetric convex conic optimization*, Mathematical Programming, 150 (2015), pp. 391–422, <https://doi.org/10.1007/s10107-014-0773-1>.
 - [50] G. STENGLE, *Complexity estimates for the Schmüdgen Positivstellensatz*, Journal of Complexity, 12 (1996), pp. 167–174, <https://doi.org/10.1006/jcom.1996.0011>.
 - [51] J. F. STURM, *Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones*, Optimization Methods and Software, 11–12 (1999), pp. 625–653, <https://doi.org/10.1080/10556789908805766>.
 - [52] M. J. TODD AND Y. YE, *Approximate Farkas lemmas and stopping rules for iterative infeasible-point algorithms for linear programming*, Mathematical Programming, 81 (1998), pp. 1–21, <https://doi.org/10.1007/BF01584841>.
 - [53] K.-C. TOH, M. J. TODD, AND R. H. TÜTÜNCÜ, *SDPT3 — a Matlab software package for semidefinite programming, version 1.3*, Optimization Methods and Software, 11–12 (1999), pp. 545–581, <https://doi.org/10.1080/10556789908805762>.
 - [54] L. TUNÇEL, *Generalization of primal—dual interior-point methods to convex optimization problems in conic form*, Foundations of Computational Mathematics, 1 (2001), pp. 229–254, <https://doi.org/10.1007/s002080010009>.
 - [55] M. YAMASHITA, K. FUJISAWA, AND M. KOJIMA, *Implementation and evaluation of SDPA 6.0 (Semidefinite Programming Algorithm 6.0)*, Optimization Methods and Software, 18 (2003), pp. 491–505, <https://doi.org/10.1080/1055678031000118482>.

Appendix A. Self-concordant barrier functions and local norms. The notion of self-concordant barriers, originally introduced by Nesterov and Nemirovski [33, Chapter 2], are quite broadly known at least to experts of interior-point algorithms, but

for completeness and for the ease of referencing a few specific properties used throughout the paper, we have collected a few key results about them in this Appendix. For a more in-depth study, the reader is recommended to confer the excellent monographs [46] and [31, Chapter 5].

DEFINITION A.1 (Self-concordant barrier). *Let $\mathcal{K} \subseteq \mathbb{R}^n$ be a convex cone with a non-empty interior \mathcal{K}° . A function $f : \mathcal{K}^\circ \rightarrow \mathbb{R}$ is called a strongly nondegenerate, self-concordant barrier for \mathcal{K} if f is strictly convex, at least three times differentiable, and has the following two properties (with $\text{bd}(\mathcal{K})$ denoting the boundary of \mathcal{K} , and g and H denoting the gradient and Hessian of f):*

1. $f(\mathbf{x}) \rightarrow \infty$ as $\mathbf{x} \rightarrow \text{bd}(\mathcal{K})$,
2. $|D^3 f(\mathbf{x})[\mathbf{h}, \mathbf{h}, \mathbf{h}]| \leq 2 (D^2 f(\mathbf{x})[\mathbf{h}, \mathbf{h}])^{3/2}$,
3. $\nu := \sup_{\mathbf{x} \in \mathcal{K}^\circ} g(\mathbf{x})^\top H(\mathbf{x})^{-1} g(\mathbf{x}) < \infty$.

We refer to the parameter ν above as the barrier parameter of f ; we also use the shorthand ν -SCB to mean “an SCB with barrier parameter ν ”.

DEFINITION A.2 (Logarithmic homogeneity and LHSCBs). *Let $f : \mathcal{K}^\circ \rightarrow \mathbb{R}$ be a ν -SCB function for the proper convex cone \mathcal{K} . We say that f is a logarithmically homogeneous self-concordant barrier (LHSCB) with barrier parameter ν for \mathcal{K} , or a ν -LHSCB for short, if it satisfies the logarithmic homogeneity property*

$$f(t\mathbf{x}) = f(\mathbf{x}) - \nu \ln(t) \quad \text{for all } \mathbf{x} \in \mathcal{K}^\circ \text{ and } t > 0.$$

We associate with each $\mathbf{x} \in \mathcal{K}^\circ$ the local inner product $\langle \cdot, \cdot \rangle_{\mathbf{x}} : \mathcal{K} \times \mathcal{K} \rightarrow \mathbb{R}$ defined as $\langle \mathbf{u}, \mathbf{v} \rangle_{\mathbf{x}} := \mathbf{u}^\top H(\mathbf{x}) \mathbf{v}$ and the local norm $\|\cdot\|_{\mathbf{x}}$ induced by this local inner product. Thus, $\|\mathbf{v}\|_{\mathbf{x}} = \|H(\mathbf{x})^{1/2} \mathbf{v}\|$. We define the local (open) ball centered at \mathbf{x} with radius r by $\mathcal{B}(\mathbf{x}, r) := \{\mathbf{v} \in \mathbb{R}^n \mid \|\mathbf{v} - \mathbf{x}\|_{\mathbf{x}} < r\}$. When $r = 1$, these balls are also called the *Dikin ellipsoids*. Crucially, for every $\mathbf{x} \in \mathcal{K}^\circ$, we have $\mathcal{B}(\mathbf{x}, 1) \subset \mathcal{K}^\circ$.

Analogously, we define the dual local inner product $\langle \cdot, \cdot \rangle_{\mathbf{x}}^* : \mathcal{K}^* \times \mathcal{K}^* \rightarrow \mathbb{R}$ by $\langle \mathbf{s}, \mathbf{t} \rangle_{\mathbf{x}}^* := \mathbf{s}^\top H(\mathbf{x})^{-1} \mathbf{t}$, its induced dual local norm $\|\cdot\|_{\mathbf{x}}^*$, and the corresponding open balls $\mathcal{B}^*(\mathbf{x}, r)$. The dual local norm, by definition, satisfies the identity $\|\mathbf{t}\|_{\mathbf{x}}^* = \|H(\mathbf{x})^{-1/2} \mathbf{t}\|$.

The most important properties of SCBs and LHSCBs applied in the analysis of our algorithms are compiled in the next two Propositions.

PROPOSITION A.3 ([46, Chapters 2–3], [31, Chapter 5], and [32, Lemma 1]). *Let f be a ν -SCB for the proper convex cone \mathcal{K} , let $\mathbf{x} \in \mathcal{K}^\circ$, and let $\mathbf{n}(\mathbf{x}) = -H(\mathbf{x})^{-1} g(\mathbf{x})$ be the (unconstrained) Newton step for f at \mathbf{x} .*

(A1) *If $\|\mathbf{n}(\mathbf{x})\|_{\mathbf{x}} < 1$, then $\mathbf{x}^+ := \mathbf{x} + \mathbf{n}(\mathbf{x})$ satisfies*

$$\|\mathbf{n}(\mathbf{x}^+)\|_{\mathbf{x}^+} \leq \left(\frac{\|\mathbf{n}(\mathbf{x})\|_{\mathbf{x}}}{1 - \|\mathbf{n}(\mathbf{x})\|_{\mathbf{x}}} \right)^2.$$

(A2) *If $\mathbf{u} \in \mathcal{B}_{\mathbf{x}}(\mathbf{x}, 1)$, then $\mathbf{u} \in \mathcal{K}^\circ$, and for all $\mathbf{v} \neq \mathbf{0}$, one has*

$$\frac{\|\mathbf{v}\|_{\mathbf{u}}^*}{\|\mathbf{v}\|_{\mathbf{x}}^*} \leq \frac{1}{1 - \|\mathbf{u} - \mathbf{x}\|_{\mathbf{x}}}.$$

(A3) *Let $\mathbf{x}, \mathbf{u} \in \mathcal{K}^\circ$, and let $r := \|\mathbf{x} - \mathbf{u}\|_{\mathbf{u}} < 1$. Then*

$$\|g(\mathbf{x}) - g(\mathbf{u}) - H(\mathbf{u})(\mathbf{x} - \mathbf{u})\|_{\mathbf{u}}^* \leq \frac{r^2}{1 - r}.$$

PROPOSITION A.4 ([46, Chapters 2–3], [31, Chapter 5]). *Let f be a ν -LHSCB for the proper convex cone \mathcal{K} and let $\mathbf{x} \in \mathcal{K}^\circ$. Then:*

- (A4) $H(\mathbf{x})\mathbf{x} = -g(\mathbf{x})$;
- (A5) $g(\mathbf{x})^\top \mathbf{x} = -\nu$;
- (A6) $g(\alpha\mathbf{x}) = \alpha^{-1}g(\mathbf{x})$ and $H(\alpha\mathbf{x}) = \alpha^{-2}H(\mathbf{x})$ for all $\alpha > 0$;
- (A7) $\|g(\mathbf{x})\|_x^* = \sqrt{\nu}$.

Appendix B. Stopping Algorithm 2.1 with a damped Newton step.

Here we collect a few technical results about a variant of Algorithm 2.1 with damped Newton steps, used in the analysis of the Phase 1 method of Section 3.2.

LEMMA B.1. *Let $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{N}(\eta, \tau)$ and let $(\Delta\mathbf{x}, \Delta\mathbf{y}, \Delta\mathbf{s})$ be the solution of (NS). Let $\mathbf{x}^+ := \mathbf{x} + \alpha\Delta\mathbf{x}$, $\mathbf{s}^+ := \mathbf{s} + \alpha\Delta\mathbf{s}$, and $\mathbf{y}^+ := \mathbf{y} + \alpha\Delta\mathbf{y}$, where $\alpha \in (0, 1]$. Let $\tau^+ := (1 - \vartheta)\tau$, where $\vartheta = \frac{\eta/2}{\sqrt{\nu}+1}$ and $0 < \eta \leq 1/4$ as before in Algorithm 2.1. Then $\|\mathbf{s}^+ + \tau^+g(\mathbf{x}^+)\|_{\mathbf{x}^+}^* \leq 2\eta\tau^+$.*

Proof. Using (A2) and (ii) of Lemma 2.1, we have

$$(B.1) \quad \|\mathbf{s}^+ + \tau g(\mathbf{x}^+)\|_{\mathbf{x}^+}^* \leq \frac{\|\mathbf{s}^+ + \tau g(\mathbf{x}^+)\|_{\mathbf{x}}^*}{1 - \alpha\eta},$$

as $\|\mathbf{x}^+ - \mathbf{x}\|_{\mathbf{x}} = \alpha\|\Delta\mathbf{x}\|_{\mathbf{x}} \leq \alpha\eta < 1$ by part (ii) of Lemma 2.1 and $\eta \leq 1/4$.

Applying (A3) and (ii) of Lemma 2.1 again gives

$$(B.2) \quad \|g(\mathbf{x}^+) - g(\mathbf{x}) - \alpha H(\mathbf{x})\Delta\mathbf{x}\|_{\mathbf{x}}^* \leq \frac{\alpha^2\eta^2}{1 - \alpha\eta}.$$

Using (B.2), the third equation of (NS) and the neighborhood definition, we can derive the following upper bound:

$$\begin{aligned} \|\mathbf{s}^+ + \tau g(\mathbf{x}^+)\|_{\mathbf{x}}^* &\leq \|\mathbf{s} + \alpha\Delta\mathbf{s} + \tau(g(\mathbf{x}) + \alpha H(\mathbf{x})\Delta\mathbf{x})\|_{\mathbf{x}}^* \\ &\quad + \tau\|g(\mathbf{x}^+) - g(\mathbf{x}) - \alpha H(\mathbf{x})\Delta\mathbf{x}\|_{\mathbf{x}}^* \\ &= \|\mathbf{s} + \tau g(\mathbf{x}) + \alpha(\Delta\mathbf{s} + \tau H(\mathbf{x})\Delta\mathbf{x})\|_{\mathbf{x}}^* \\ &\quad + \tau\|g(\mathbf{x}^+) - g(\mathbf{x}) - \alpha H(\mathbf{x})\Delta\mathbf{x}\|_{\mathbf{x}}^* \\ &\stackrel{(NS)}{=} (1 - \alpha)\|\mathbf{s} + \tau g(\mathbf{x})\|_{\mathbf{x}}^* + \tau\|g(\mathbf{x}^+) - g(\mathbf{x}) - \alpha H(\mathbf{x})\Delta\mathbf{x}\|_{\mathbf{x}}^* \\ &\stackrel{(B.2)}{\leq} (1 - \alpha)\eta\tau + \tau \frac{\alpha^2\eta^2}{1 - \alpha\eta} = \tau \left((1 - \alpha)\eta + \frac{\alpha^2\eta^2}{1 - \alpha\eta} \right). \end{aligned}$$

From (B.1) and the previously derived upper bound it follows that

$$(B.3) \quad \|\mathbf{s}^+ + \tau g(\mathbf{x}^+)\|_{\mathbf{x}^+}^* \stackrel{(B.1)}{\leq} \frac{\|\mathbf{s}^+ + \tau g(\mathbf{x}^+)\|_{\mathbf{x}}^*}{1 - \alpha\eta} \leq \tau \left[\frac{(1 - \alpha)\eta}{1 - \alpha\eta} + \frac{\alpha^2\eta^2}{(1 - \alpha\eta)^2} \right].$$

Therefore,

$$(B.4) \quad \begin{aligned} \frac{1}{\tau^+} \|\mathbf{s}^+ + \tau^+g(\mathbf{x}^+)\|_{\mathbf{x}^+}^* &\leq \frac{1}{(1 - \vartheta)\tau} (\|\mathbf{s}^+ + \tau g(\mathbf{x}^+)\|_{\mathbf{x}^+}^* + \vartheta\tau\|g(\mathbf{x}^+)\|_{\mathbf{x}^+}^*) \\ &\stackrel{(B.3), (A7)}{\leq} \frac{1}{1 - \vartheta} \left(\frac{(1 - \alpha)\eta}{1 - \alpha\eta} + \frac{\alpha^2\eta^2}{(1 - \alpha\eta)^2} + \vartheta\sqrt{\nu} \right). \end{aligned}$$

In the analysis of Algorithm 2.1, we have $\eta \leq \frac{1}{4}$, and $\vartheta = \frac{\eta/2}{\sqrt{\nu}+1}$, thus,

$$\frac{1}{1 - \vartheta} = \frac{\sqrt{\nu} + 1}{\sqrt{\nu} + 1 - \eta/2} \leq \frac{2}{2 - \eta/2} \text{ and } \frac{\vartheta\sqrt{\nu}}{1 - \vartheta} = \frac{\frac{\eta}{2}\sqrt{\nu}}{\sqrt{\nu} + 1 - \eta/2} \leq \frac{\eta}{2}.$$

Using these inequalities and (B.4), we can derive the upper bound in the statement of the lemma:

$$\begin{aligned} \frac{1}{\tau^+} \|\mathbf{s}^+ + \tau^+ g(\mathbf{x}^+)\|_{\mathbf{x}^+}^* &\leq \frac{2}{2 - \eta/2} \left(\frac{(1 - \alpha)\eta}{1 - \alpha\eta} + \frac{\alpha^2 \eta^2}{(1 - \alpha\eta)^2} \right) + \frac{\eta}{2} \\ &\leq \frac{2\eta}{2 - \eta/2} + \frac{\eta}{2} \leq 2\eta, \end{aligned}$$

as $\frac{(1-\alpha)\eta}{1-\alpha\eta} + \frac{\alpha^2 \eta^2}{(1-\alpha\eta)^2}$ is strictly decreasing in α if $\eta \leq \frac{1}{4}$ holds. \square

LEMMA B.2. *Suppose Algorithm 2.1 is stopped after a damped Newton step of size $\alpha \in (0, 1]$, following a series of full Newton steps. Then the final iterate satisfies*

$$\left(1 - \frac{\eta^2}{\nu}\right) \tau \leq \mu \leq \frac{1}{1 - \vartheta} \tau.$$

Proof. Let us denote the penultimate iterate by $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ and the last iterate by

$$(\mathbf{x}^+, \mathbf{y}^+, \mathbf{s}^+) = (\mathbf{x}, \mathbf{y}, \mathbf{s}) + \alpha(\Delta \mathbf{x}, \Delta \mathbf{y}, \Delta \mathbf{s}).$$

The final duality gap is $\mu = \frac{(\mathbf{x}^+)^{\top} \mathbf{s}^+}{\nu}$. Proceeding analogously to the proof of Lemma 2.3, we can write

$$(\mathbf{x}^+)^{\top} \mathbf{s}^+ = \mathbf{x}^{\top} \mathbf{s} - \alpha\tau \|\Delta \mathbf{x}\|_{\mathbf{x}}^2 - \alpha \mathbf{x}^{\top} \mathbf{s} + \alpha\tau\nu = (1 - \alpha)\mathbf{x}^{\top} \mathbf{s} + \alpha\tau(\nu - \|\Delta \mathbf{x}\|_{\mathbf{x}}^2).$$

From Lemma 2.3, after the penultimate step we have

$$(B.5) \quad \frac{\tau}{1 - \vartheta}(\nu - \eta^2) \leq \mathbf{x}^{\top} \mathbf{s} \leq \frac{\tau}{1 - \vartheta}\nu.$$

Using (B.5) and (ii) from Lemma 2.1, we can derive the desired lower and an upper bounds on μ :

$$\mu = \frac{(\mathbf{x}^+)^{\top} \mathbf{s}^+}{\nu} \leq \frac{(1 - \alpha)\tau}{1 - \vartheta} + \alpha\tau \leq \frac{\tau}{1 - \vartheta},$$

and

$$\mu = \frac{(\mathbf{x}^+)^{\top} \mathbf{s}^+}{\nu} \geq \frac{(1 - \alpha)\tau \left(1 - \frac{\eta^2}{\nu}\right)}{1 - \vartheta} + \alpha\tau \left(1 - \frac{\eta^2}{\nu}\right) \geq \tau \left(1 - \frac{\eta^2}{\nu}\right). \quad \square$$