arXiv:1804.06925v1 [math.OC] 18 Apr 2018

PRIMAL-DUAL INTERIOR-POINT METHODS FOR DOMAIN-DRIVEN FORMULATIONS: ALGORITHMS*

MEHDI KARIMI AND LEVENT TUNÇEL

Abstract. We study infeasible-start primal-dual interior-point methods for convex optimization problems given in a typically natural form we denote as Domain-Driven formulation. Our algorithms extend many advantages of primal-dual interior-point techniques available for conic formulations, such as the current best complexity bounds, and more robust certificates of approximate optimality, unboundedness, and infeasibility, to Domain-Driven formulations. The complexity results are new for the infeasible-start setup used, even in the case of linear programming. In addition to complexity results, our algorithms aim for expanding the applications of, and software for interior-point methods to wider classes of problems beyond optimization over symmetric cones.

1. Introduction

In this article, a *convex optimization* problem is minimizing a *convex function* over a *convex set* in a finite dimensional Euclidean space. Convex optimization's powerful and elegant theory has been coupled with faster and more reliable numerical linear algebra software and powerful computers to spread its applications over many fields such as (1) data science: machine learning, compressed sensing (see [9, 21, 13, 2]), (2) engineering: control theory, signal processing, circuit design (see [8, 4, 6, 1]), (3) relaxation and randomization: provable bounds and robust heuristics for hard nonconvex problems (see [46]), and (4) robust optimization (see [5, 3]). Development of modern *interior-point methods* has had a huge impact on the popularity of convex optimization. Modern theory of interior-point methods, with polynomial iteration complexity, started with Karmarkar's revolutionary paper [25] in 1984 and then extended from linear optimization to general convex optimization problems by Nesterov and Nemirovskii [37] in the late 1980's. The literature on this topic has become extensive and many different approaches have been proposed since then. In this article, we are interested in the modern *primal-dual* interior-point techniques.

* Some of the material in this manuscript appeared in a preliminary form in Karimi's PhD thesis [24].

Mehdi Karimi (m7karimi@uwaterloo.ca) and Levent Tunçel (ltuncel@math.uwaterloo.ca): Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. Research of the authors was supported in part by Discovery Grants from NSERC and by U.S. Office of Naval Research under award numbers: N00014-12-1-0049 and N00014-15-1-2171.

[33] has a detailed discussion about the advantages of primal-dual techniques over purely primal ones, for example, in designing long-step algorithms.

The focus of research for primal-dual algorithms has been mostly on *conic formulations* (where minimization is over the intersection of an affine subspace and a *convex cone*), see for example [36, 34, 41, 39, 40, 47, 30, 45]. Following this research, in many settings, the state-of-the-art for utilizing primal-dual interior-point methods is to reformulate the given convex optimization problem as a conic optimization problem (see [33] or [37]-Section 5.1). This usually requires the introduction of additional variables and constraints (that are artificial in the context of the original problem). However, the applications and software for conic optimization itself have not gone much beyond optimization over *symmetric (self-scaled) cones*; more specifically linear programming (LP), second-order cone programming (SOCP), and semidefinite programming (SDP). Some of the desired properties of optimization over symmetric cones have been extended to general conic optimization [45, 35, 44, 31]. While the conic reformulation implies that, under reasonable assumptions, all convex optimization problems enjoy the same iteration complexity bounds, there is a gap (remained unchanged for many years) between the efficiency and robustness of the software we have for optimization over symmetric cones and many other classes of problems. In the *feasible-start*[1] setup, [33] demonstrated that not all advantages of the primal-dual interior-point techniques are intrinsically related to conic formulation. In this article, we expand this conclusion to the more challenging and practical *infeasible-start* scenario. Specifically, we design and analyze infeasible-start primal-dual algorithms for problems given in a typically natural form (can be a conic formulation or not or an arbitrary mixture of both) that not only have comparable theoretical performance to the current best algorithms for conic formulations, but also have been used to create practical software. Let us define our setup:

Definition 1.1. *A convex optimization problem is said to be in the* Domain-Driven *setup if it is in the form*

$$\inf_x \{\langle c, x\rangle : Ax \in D\}, \tag{1}$$

*where $x \mapsto Ax : \mathbb{R}^n \to \mathbb{R}^m$ is a linear embedding, with A and $c \in \mathbb{R}^n$ are given, and $D \subset \mathbb{R}^m$ is a convex set given as the closure of the domain of a ϑ-*self-concordant (s.c.) barrier Φ.

A s.c. barrier (rigorously defined in Appendix A) is a convex function whose second derivative regulates its third and first derivatives. Every open convex set is the domain of a s.c. barrier [37]. Thus, in principle, every convex optimization problem can be treated in the Domain-Driven setup. In applications, the restrictive part of Definition 1.1 is that a "computable"[2] s.c. barrier is not necessarily available for a general convex set. However, for many interesting convex sets (each

[1]Where a pair of points in the relative interior of the primal and dual feasible regions are given.

[2]Computable means we can evaluate the function and its first and second derivatives at a reasonable cost.

of which allows us to handle a class of convex optimization problems), we know how to construct an efficient s.c. barrier. Specifically, the feasible region of many classes of problems that arise in practice is the direct sum of small dimensional convex sets with known, computable s.c. barriers. In the case of linear programming, for example, consider the 1-dimensional set $\{z \in \mathbb{R} : z \geq \beta\}$ for $\beta \in \mathbb{R}$. It is well-known that $-\ln(z-\beta)$ is a s.c. barrier for this set. Using this simple function and the fact that if convex sets D_1 and D_2 have s.c. barriers f_1 and f_2, respectively, then f_1+f_2 is a s.c. barrier for the direct sum of D_1 and D_2, we can construct a s.c. barrier for any polyhedron; for $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, a s.c. barrier for

$$\{x \in \mathbb{R}^n : Ax \leq b\} = \{x \in \mathbb{R}^n : Ax \in D\},$$

where $D := b - \mathbb{R}^m_+$, is $-\sum_{i=1}^m \ln(b_i - a_i^\top x)$, where $a_i^\top$ is the ith row of A. This discussion for LP exemplifies the fact that knowing a s.c. barrier for small dimensional convex sets combined with the direct sum operator lets us solve problems with an arbitrarily large number of variables and constraints (of the same type).

The power of the Domain-Driven setup is further accentuated when we consider the possibility of direct summing (or alternatively, intersecting) convex sets of different types. In the following, we show many set constraints/functions as the building blocks of a problem in the Domain-Driven setup. We start by showing that the Domain-Driven setup covers the popular optimization over symmetric cones. Many of these s.c. functions can be found in Nesterov and Nemirovski's seminal book [37].

LP, SOCP, and SDP: optimization over symmetric cones is a special case of the Domain-Driven setup. Table 1 shows the constraints that specify D and a s.c. barrier associated with the convex set defined by the constraint. For example, if our problem has a constraint of the form $a^\top x \leq \beta$

TABLE 1. LP, SOCP, and SDP constraints and the corresponding s.c. barriers. $\mathbb{S}^n$ is the set of n-by-n symmetric matrices and $A \preceq B$ for $A, B \in \mathbb{S}^n$ means $B - A$ is positive semidefinite.

	constraint	s.c. barrier Φ
LP	$z \leq \beta, \quad z, \beta \in \mathbb{R},$	$-\ln(\beta - z)$
SOCP	$\|z\| \leq t, \quad z \in \mathbb{R}^n, \quad t \in \mathbb{R},$	$-\ln(t^2 - z^\top z)$
SDP	$Z \preceq B, \quad Z, B \in \mathbb{S}^n$	$-\ln(\det(B - Z))$

for $a \in \mathbb{R}^n, \beta \in \mathbb{R}$, the convex set defined by this constraint is the set of $x \in \mathbb{R}^n$ such that $a^\top x \in \{z : z \leq \beta\}$.

Direct sum of 2-dimensional sets: The Domain-Driven setup allows inequalities of the form

$$\sum_{i=1}^{\ell} \alpha_i f_i(a_i^\top x + \beta_i) + g^\top x + \gamma \leq 0, \quad a_i, g \in \mathbb{R}^n, \quad \beta_i, \gamma \in \mathbb{R}, \quad i \in \{1, \ldots, \ell\}, \tag{2}$$

where $\alpha_i \geq 0$ and $f_i(x)$, $i \in \{1, \ldots, \ell\}$, can be any univariate convex function whose epigraph is a 2-dimensional set equipped with a known s.c. barrier. Three popular examples are given in Table 2, and several more can be found in [37]. The fact that constraints of the form (2) fit into the Domain-Driven setup is implied by the following relation:

$$\begin{aligned} & \left\{x : \textstyle\sum_{i=1}^{\ell} \alpha_i f_i(a_i^\top x + \beta_i) + g^\top x + \gamma \leq 0\right\} \\ = \; & \left\{x : \exists u \in \mathbb{R}^\ell \text{ such that } \textstyle\sum_{i=1}^{\ell} \alpha_i u_i + g^\top x + \gamma \leq 0, \quad f_i(a_i^\top x + \beta_i) \leq u_i, \ \forall i\right\}. \end{aligned} \tag{3}$$

Note that Geometric Programming [7] and Entropy Programming [14] with vast applications in engineering are constructed with constraints of the form (2) when $f_i(z) = e^z$ for $i \in \{1, \ldots, \ell\}$ and $f_i(z) = z\ln(z)$ for $i \in \{1, \ldots, \ell\}$, respectively.

TABLE 2. Some 2-dimensional convex sets and their s.c. barriers.

	set (z, t)	s.c. barrier $\Phi(z, t)$
1	$e^z \leq t$	$-\ln(\ln(t) - z) - \ln(t)$
2	$z\ln(z) \leq t, \ z > 0$	$-\ln(t - z\ln(z)) - \ln(z)$
3	$\lvert z\rvert^p \leq t, \ p \geq 1$	$-\ln(t^{\frac{2}{p}} - z^2) - 2\ln(t)$

Epigraph of matrix norm, minimizing nuclear norm: Assume that we have constraints of the form

$$Z - UU^\top \succeq 0, \quad \text{where} \quad Z = Z_0 + \sum_{i=1}^{\ell} x_i Z_i, \quad U = U_0 + \sum_{i=1}^{\ell} x_i U_i. \tag{4}$$

Z_i, $i \in \{0, \ldots, \ell\}$, are m-by-m symmetric matrices, and U_i, $i \in \{0, \ldots, \ell\}$, are m-by-n matrices. Using the Schur complement theorem, we can reformulate (4) as an SDP constraint with size $m + n$. However, the set $\{(Z, U) : Z - UU^\top \succeq 0\}$ accepts the following s.c. barrier:

$$\Phi(Z, U) := -\ln(\det(Z - UU^\top)). \tag{5}$$

In the cases that $m \ll n$, the parameter of the s.c. barrier (responsible for worst-case iteration complexity bounds (see Appendix A)) for (5) is much smaller than the one we need for the SDP reformulation, which can make a huge difference both in theory and applications.

A special application for constraints of the form (4) arises in minimizing the *nuclear norm*. The nuclear norm of a matrix Z is $\|Z\|_* := \mathrm{Tr}\left((ZZ^\top)^{1/2}\right)$. The dual norm of $\|\cdot\|_*$ is the 2-norm

$\|\cdot\|$ of a matrix. It can be shown that the following optimization problems are a primal-dual pair [43].

$$(6)\qquad \begin{array}{llll} (P_N) & \min_X & \|X\|_* & \\ & s.t. & AX = b. & \end{array} \qquad \begin{array}{llll} (D_N) & \max_z & \langle b, z\rangle & \\ & s.t. & \|A^* z\| \le 1, & \end{array}$$

where A is a linear transformation on matrices and A^* is its adjoint. In machine learning and compressed sensing, (P_N) is a very popular relaxation of the problem of minimizing $\mathrm{rank}(X)$ subject to $AX = b$. The dual problem (D_N) is a special case of (4) where $Z = I$ and $U = A^* z$. It can be shown that solving (D_N) by our primal-dual techniques immediately gives us a solution for (P_N).

Compatibility of s.c. barriers, epigraph of quantum entropy and quantum relative entropy: Another useful theoretical tool for constructing s.c. functions and barriers is the *compatibility result*, see Chapter 5 of [37] and Theorem 9.1.1 of [32]. Recently, such an approach was used [18, 17] to construct a s.c. barrier for the epigraph of *quantum entropy*. Consider a function $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ and let $X \in \mathbb{H}^n$ be a Hermitian matrix (with entries from $\mathbb{C}$) with a spectral decomposition $X = U\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)U^*$, where Diag returns a diagonal matrix with the given entries on its diagonal and U^* is the conjugate transpose of a unitary matrix U. Then, $F : \mathbb{H}^n \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$F(X) := \mathrm{Tr}(U\mathrm{Diag}(f(\lambda_1), \ldots, f(\lambda_n))U^*).$$

Study of such matrix functions go back to the work of Löwner as well as Von-Neumann (see [12], [27], and the references therein). It is proved in [18] (and it follows from the above-mentioned compatibility result) that if f is continuously differentiable with a *matrix monotone* derivative on $\mathbb{R}_+$, then the function

$$\Phi(t, X) := -\ln(t - F(X)) - \ln\det(X)$$

is a s.c. barrier for the epigraph of $F(X)$ in $\mathbb{S}^n_+$. For $f(x) := x\ln(x)$, the function $F(X)$ is called *quantum entropy*. In this case, $\Phi(t, X)$ can be seen as a lift for the s.c. barrier we gave in Table 2 for the entropy function. Optimization of quantum entropy and its extension *relative quantum entropy* have many recent applications [10, 11]. The authors in [15, 16] approximate these problems by SDP. We can handle convex optimization problems involving quantum entropy in the Domain-Driven setup by the above s.c. barrier.

Combination of all the above examples: Assume that we have ℓ problems in the Domain-Driven setup, with corresponding sets $D_1, \ldots, D_\ell$, and corresponding s.c. barriers $\Phi_1, \ldots, \Phi_\ell$. Now, let $D := D_1 \oplus \cdots \oplus D_\ell$. Then, $\Phi := \Phi_1 + \cdots + \Phi_\ell$ is a s.c. barrier for D (see Subsection A.2), and (1) for D is also in the Domain-Driven setup.

1.1. **Contributions of this paper.** Although the terminology Domain-Driven is new, the concept was proposed in [33], then named *cone-free*. The underlying algorithms were feasible-start

primal-dual algorithms for problems in the Domain-Driven setup. In theory of convex optimization, having a theory of feasible-start algorithms is sufficient for many purposes. In applications of convex optimization as well as in software, infeasible-start algorithms are essential. For the infeasible-start setup, the most common approach is (1) conic reformulation, (2) using homogeneous self-dual embedding type algorithms (see for example [41]). However, software and applications of modern conic optimization itself has not gone much beyond optimization over symmetric cones. There are other types of algorithms such as Nesterov and Nemirovski's which approximately follow multi-parameter surfaces of analytic centers [38]. These algorithms seem too complicated to directly result in a practical code. [20, 22] are some recent developments on infeasible-start interior-point algorithms for convex and nonconvex optimization that are related to our approach. For infeasible-start algorithms that solve a Newton system at every iteration, we can consider two extremes based on the number of artificial variables. At one extreme (see [28, 29, 26, 48, 49]), there is no artificial variable and the systems we solve at every iteration are similar to the ones we solve in the feasible-start case with a perturbed right-hand-side. At the other extreme are the homogeneous self-dual embedding type algorithms [47, 41] where we have artificial variables and homogenization variables. Our infeasible-start approach is in the middle, closer to the first group as we add only one artificial variable, but do not impose an explicit homogenization (moreover, we tie our artificial variable to our central path parameter). Our complexity results here are new for this approach, even in the case of LP.

We introduce a notion of duality gap for the Domain-Driven setup and define an infeasible-start primal-dual central path (Section 2). Then, in Section 3 we design our path-following algorithms and in Section 4 we give the analysis that yields the current best iteration complexity bounds for solving the problem. By solving, we mean determining the status of a given problem (as being unbounded, infeasible, having optimal solutions, etc.) and providing suitable approximate certificates for the status. Several cases of ill-conditioning can happen for a given problem. In order to evaluate the performance of any algorithm in determining the status of a problem in the Domain-Driven setup, we need to carefully categorize these statuses. This has been done preliminarily in [24] and will be presented in subsequent works. In this paper, we briefly discuss how to interpret the outcome of the algorithms and elaborate on the case of strict primal and dual feasibility. The different patterns that can be detected by our algorithms and the iteration complexity bounds for them are comparable to the current best results available for infeasible-start conic optimization, which to the best of our knowledge is mostly in the work of Nesterov-Todd-Ye [41]. The algorithms we design make up the foundation of a new code DDS (Domain-Driven Solver).

Part of the strength and elegance of the interior-point machinery for conic optimization comes from the fact that convex cones accept s.c. barriers that are *logarithmically-homogenous* (LH). Figure 1 shows the relation between various classes of s.c. functions. LF conjugate of a LH s.c.

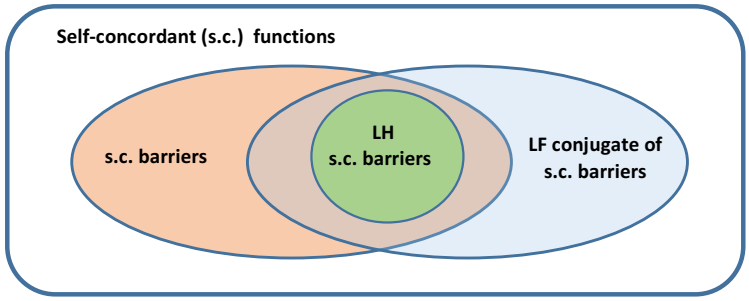

FIGURE 1. A diagram that shows the relationships among various classes of self-concordant functions (see Appendix B for various examples).

barrier is also a LH s.c. barrier; an important property that we loose for a general s.c. barrier. However, importantly, the LF conjugate of a s.c. barrier has more properties than an arbitrary s.c. function. Another contribution of this article is that in the design and analysis of our algorithms, we vastly exploit this property, which has not been considered at this level of detail in the literature.

1.2. **Assumptions and notations.** To design our primal-dual algorithms for the Domain-Driven setup, we make some assumptions. First, we assume that the kernel of A is $\{0\}$ in Definition 1.1; otherwise we can update A to a matrix $\bar{A}$ whose columns form a basis for $\mathrm{img}A$ (image of A) and then update c and Φ accordingly (see subsection A.2 for stability of s.c. barriers under affine maps). We also assume that the Legendre-Fenchel (LF) conjugate Φ_* of Φ is given. Even though restricting, such assumptions are unavoidable in the context of primal-dual techniques. Also, for many classes of problems, including the above examples, Φ_* is computable. The domain of Φ_* is the interior of a cone D_* defined as (see (142)):

$$D_* = \{y : \langle y, h\rangle \leq 0, \quad \forall h \in \mathrm{rec}(D)\}, \tag{7}$$

where $\mathrm{rec}(D)$ is the recession cone of D defined in (143). Consider an Euclidean vector space $\mathbb{E}$ with dual space $\mathbb{E}^*$ and a scalar product $\langle\cdot,\cdot\rangle$. For a self-adjoint positive definite linear transformation $B : \mathbb{E} \to \mathbb{E}^*$, we define a conjugate pair of Euclidean norms as:

$$\begin{aligned} \|x\|_B &:= [\langle Bx, x\rangle]^{1/2}, \\ \|s\|_B^* &:= \max\{\langle s, y\rangle : \ \|y\|_B \leq 1\} = \|s\|_{B^{-1}} = \left[\langle s, B^{-1}s\rangle\right]^{1/2}. \end{aligned} \tag{8}$$

Note that (8) immediately gives us a general Cauchy-Schwarz (CS) inequality:

$$\langle s, x\rangle \leq \|x\|_B \|s\|_B^*, \quad \forall x \in \mathbb{E}, \forall s \in \mathbb{E}^*. \tag{9}$$

For simplicity, we use abbreviations RHS and LHS for right-hand-side and left-hand-side, respectively.

2. Duality gap for Domain-Driven setup and central path

Considering the *support function* of D,

$$\delta_*(y|D) := \sup\{\langle y, z\rangle : z \in D\}, \tag{10}$$

we define the duality gap as:

Definition 2.1. *For every point $x \in \mathbb{R}^n$ such that $Ax \in D$ and every point $y \in D_*$ such that $A^\top y = -c$,* the duality gap *is defined as:*

$$\langle c, x\rangle + \delta_*(y|D). \tag{11}$$

Duality gap must be easily computable and support function is not generally easy to calculate. However, the following theorem shows that we can estimate the support function within any desired accuracy using the fact that Φ_* is the LF conjugate of a s.c. barrier.

Theorem 2.1 (Theorem 2.4.2 of [37])**.** *Assume that Φ is a ϑ-s.c. barrier on D and let Φ_* be the LF conjugate of Φ with domain D_*. Then, the support function of D satisfies*

$$\delta_*(y|D) - \frac{\vartheta}{k} \le \langle \Phi_*'(ky), y\rangle \le \delta_*(y|D), \quad y \in D_*, \quad \forall k > 0. \tag{12}$$

The following lemma shows that duality gap is well-defined and zero duality gap is a guarantee for optimality:

Lemma 2.1. *For every point $x \in \mathbb{R}^n$ such that $Ax \in D$ and every point $y \in D_*$ such that $A^\top y = -c$, we have*

$$\langle c, x\rangle + \delta_*(y|D) \ge 0. \tag{13}$$

Moreover, if the equality holds above for a pair $(\hat{x}, \hat{y})$ with $A\hat{x} \in D$ and $\hat{y} \in D_*$, $A^\top \hat{y} = -c$, then $\hat{x}$ is an optimal solution of (1)*.*

Proof. Let x and y be as above. Then,

$$\langle c, x\rangle \underbrace{=}_{A^\top y = -c} -\langle A^\top y, x\rangle = -\langle y, Ax\rangle \underbrace{\ge}_{Ax \in D,\ y \in D_*} -\delta_*(y|D).$$

Thus, $\langle c, x\rangle + \delta_*(y|D) \ge 0$, as desired. If equality holds for $(\hat{x}, \hat{y})$, then for every x such that $Ax \in D$, we have

$$\langle c, \hat{x}\rangle \underbrace{=}_{\text{(13) holds with equality}} -\delta_*(\hat{y}|D) \underbrace{\le}_{(10)} -\langle \hat{y}, Ax\rangle = \langle -A^\top \hat{y}, x\rangle \underbrace{=}_{A^\top \hat{y} = -c} \langle c, x\rangle.$$

Therefore, $\hat{x}$ is an optimal solution for (1). □

Corollary 2.1. *Assume that there exist a sequence* $\{z^k\} \in \mathrm{int}D$ *such that* $z^k \to A\hat{x} \in D$, *and a sequence* $\{y^k\} \in \mathrm{int}D_*$ *such that* $y^k \to \hat{y} \in D_*$ *and* $A^\top \hat{y} = -c$. *If*

$$\lim_k \Big(\langle c, x^k \rangle + \langle y^k, \Phi'_*(ky^k) \rangle \Big) = 0,$$

then $\hat{x}$ *is an optimal solution of* (1).

Proof. We use Theorem 2.1 to approximate the support function and then apply Lemma 2.1. □

2.1. **Primal-dual infeasible-start central path.** Our algorithms are *infeasible-start*, which means we do not require a feasible point from the user to start the algorithm. To introduce our infeasible-start central path, we start with a feasible start central path, called cone-free in [33], which is defined by the set of solutions to:

$$\begin{aligned} &(a) \quad Ax \in \mathrm{int}D, \\ &(b) \quad A^\top y = -\tau c, \quad y \in \mathrm{int}D_*, \\ &(c) \quad y = \Phi'(Ax), \end{aligned} \tag{14}$$

where $\tau > 0$ is the parameter of the path. It is proved in [33] that under *strict primal-dual feasibility* (there exists $\hat{x}$ such that $A\hat{x} \in \mathrm{int}D$ and $\hat{y} \in \mathrm{int}D_*$ such that $A^\top \hat{y} = -c$), the system (14) has a unique solution $(x(\tau), y(\tau))$ for every $\tau > 0$ and $x(\tau)$ converges to a solution of (1) when $\tau \to +\infty$. Note that we can also prove this by our notion of duality gap and using Theorem 2.1.

Let us see how to modify (14) for an infeasible-start algorithm. We assume that we can choose a point $z^0 \in \mathrm{int}D$ and then we define $y^0 := \Phi'(z^0) \in \mathrm{int}D_*$. We modify the primal and dual feasibility parts of (14) as follows:

$$\begin{aligned} &(a) \quad Ax + \tfrac{1}{\tau} z^0 \in \mathrm{int}D, && \tau > 0, \\ &(b) \quad A^\top y = A^\top y^0 - (\tau - 1)c, && y \in \mathrm{int}D_*, \end{aligned} \tag{15}$$

where $(x^0 := 0, \tau_0 := 1, y^0)$, is feasible for this system, and when $\tau \to +\infty$, we get a pair of primal-dual feasible points in the limit. Let us give a name to the set of points that satisfy (15):

$$Q_{DD} := \left\{ (x, \tau, y) : Ax + \frac{1}{\tau} z^0 \in \mathrm{int}D,\ \tau > 0,\ \ A^\top y - A^\top y^0 = -(\tau - 1)c,\ y \in \mathrm{int}D_* \right\}. \tag{16}$$

Our goal is to design infeasible-start primal-dual algorithms as robust as the best ones for the conic setup, which as far as we know, are the homogenous self-dual embedding type algorithms proposed in [41]. For the primal-dual conic setup, the duality gap for the modified problem in [41] has the following two crucial properties when the parameter of the path tends to $+\infty$:
(1) it tends to zero if the problem is solvable,

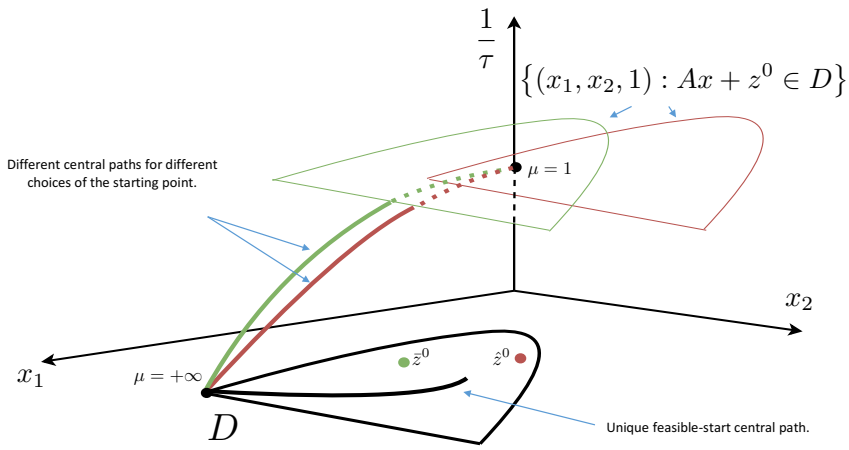

FIGURE 2. A problem in the Domain-Driven setup with $D \subset \mathbb{R}^2$, $A := I$. The infeasible-start and the unique feasible-start central paths projected onto the primal space are shown.

(2) it tends to $+\infty$ if primal or dual is infeasible.
To enforce such a property for the Domain-Driven setup, we treat τ as a variable (artificial variable) and add another parameter μ which plays the role of the parameter for the central path. Figure 2 schematically shows the primal-dual central paths.

Let us fix $\xi > 1$ and define:

$$z^0 := \text{any vector in } \mathrm{int}(D), \quad y^0 := \Phi'(z^0), \quad y_{\tau,0} := -\langle y^0, z^0\rangle - \xi\vartheta. \tag{17}$$

The following theorem defines our central path.

Theorem 2.2. *Consider the convex set $D \subset \mathbb{R}^m$ equipped with a ϑ-s.c. barrier Φ and let Φ_* be its LF conjugate with domain $\mathrm{int}D_*$. Then, for every set of starting points defined in* (17)*, the system*

$$\begin{aligned}
&(a) \quad Ax + \tfrac{1}{\tau}z^0 \in \mathrm{int}D, \quad \tau > 0,\\
&(b) \quad A^\top y - A^\top y^0 = -(\tau - 1)c, \quad y \in \mathrm{int}D_*,\\
&(c) \quad y = \tfrac{\mu}{\tau}\Phi'\left(Ax + \tfrac{1}{\tau}z^0\right),\\
&(d) \quad \langle c, x\rangle + \tfrac{1}{\tau}\langle y, Ax + \tfrac{1}{\tau}z^0\rangle = -\tfrac{\vartheta\xi\mu}{\tau^2} + \tfrac{-y_{\tau,0}}{\tau},
\end{aligned} \tag{18}$$

has a unique solution $(x(\mu), \tau(\mu), y(\mu))$ for every $\mu > 0$.

We denote the solution set of (18) for $\mu > 0$ by the *Domain-Driven primal-dual central path.* Note that for $\mu_0 = 1$, the point $(x, \tau, y) = (0, 1, y^0)$ satisfies all the equations in (18). In view of

the definition of the central path, for all the points $(x, \tau, y) \in Q_{DD}$, we define

$$
\begin{aligned}
\mu(x,\tau,y) &:= \tfrac{\tau}{\xi\vartheta}[-y_{\tau,0} - \tau\langle c, x\rangle - \langle y, Ax + \tfrac{1}{\tau}z^0\rangle], \\
&= -\tfrac{1}{\xi\vartheta}\left[\langle y, z^0\rangle + \tau(y_{\tau,0} + \langle y, Ax\rangle) + \tau^2\langle c, x\rangle\right] \\
&= -\tfrac{1}{\xi\vartheta}\left[\langle y, z^0\rangle + \tau(y_{\tau,0} + \langle c, x\rangle + \langle y^0, Ax\rangle)\right], \quad \text{using (18)-(b).}
\end{aligned} \tag{19}
$$

The formula in the second line is a quadratic in terms of τ. However, when we use the dual feasibility condition, we get the third formula that is linear in τ. In other words, the dual feasibility condition removes one of the roots. Assume that both the primal and dual are strictly feasible and we choose $z^0 = 0$, $Ax^0 \in \text{int}D$, and y^0 such that $A^\top y^0 = -c$. Then, the last equation of (19) reduces to $\mu = \tau$ and (18) reduces to the cone-free setup in (14).

Proof of Theorem 2.2. Consider the function $\Phi(\frac{z}{\tau}) - \xi\vartheta\ln(\tau)$ that we prove in Lemma A.2 is a s.c. function. The LF conjugate of this function, as a function of (y, y_τ), is also a s.c. function (Appendix A) and is calculated from the following formula:

$$
\max_{\gamma>0}\left[\Phi_*(\gamma y) + y_\tau\gamma + \xi\vartheta\ln\gamma\right]. \tag{20}
$$

The gradient of $\Phi(\frac{z}{\tau}) - \xi\vartheta\ln(\tau)$ is

$$
\begin{bmatrix} \frac{1}{\tau}\Phi'(\frac{z}{\tau}) \\ -\frac{1}{\tau^2}\langle\Phi'(\frac{z}{\tau}), z\rangle - \frac{\xi\vartheta}{\tau} \end{bmatrix}. \tag{21}
$$

By substituting (18)-(c) in (18)-(d) and reordering the terms, we can show that for every $\mu > 0$, the solution set of (18) corresponds to the solution set of the following system

$$
\begin{aligned}
\begin{bmatrix} y \\ y_\tau \end{bmatrix} &= \mu\begin{bmatrix} \frac{1}{\tau}\Phi'(\frac{z}{\tau}) \\ -\frac{1}{\tau^2}\langle\Phi'(\frac{z}{\tau}), z\rangle - \frac{\xi\vartheta}{\tau} \end{bmatrix}, \\
z &= \tau Ax + z^0, \\
A^\top y &= A^\top y^0 - (\tau - 1)c, \\
y_\tau &= y_{\tau,0} + \tau\langle c, x\rangle.
\end{aligned} \tag{22}
$$

Consider the following function:

$$
\begin{aligned}
&\Phi\left(\tfrac{z}{\tau}\right) - \xi\vartheta\ln(\tau) + \max_{\gamma>0}\left[\Phi_*(\gamma y) + y_\tau\gamma + \xi\vartheta\ln\gamma\right] - \tfrac{1}{\mu}\left(\langle y, z\rangle + \tau y_\tau\right) \\
\geq\ &\Phi\left(\tfrac{z}{\tau}\right) - \xi\vartheta\ln\tau + \Phi_*(\gamma y) + y_\tau\gamma + \xi\vartheta\ln\gamma - \tfrac{1}{\mu}\left(\langle y, z\rangle + \tau y_\tau\right), \quad \forall\gamma > 0,
\end{aligned} \tag{23}
$$

where the last inequality trivially holds because of the max function. Let us substitute $\gamma := \frac{\tau}{\mu}$, then by using the Fenchel-Young inequality (Theorem A.1)

$$
\Phi\left(\frac{z}{\tau}\right) + \Phi_*\left(\frac{\tau y}{\mu}\right) \geq \langle\frac{\tau y}{\mu}, \frac{z}{\tau}\rangle = \frac{1}{\mu}\langle y, z\rangle,
$$

we can continue (23) as

$$\geq \frac{1}{\mu}\langle y, z\rangle + \frac{1}{\mu}\tau y_\tau - \xi\vartheta \ln \mu - \frac{1}{\mu}\left(\langle y, z\rangle + \tau y_\tau\right) \ = -\xi\vartheta \ln \mu. \tag{24}$$

Hence, the function is bounded from below for every $\mu > 0$. Fix $\mu > 0$ and consider the optimization problem

$$\begin{array}{ll} \min & \Phi\left(\frac{z}{\tau}\right) - \xi\vartheta \ln(\tau) + \max_{\gamma>0}\left[\Phi_*(\gamma y) + y_\tau \gamma + \xi\vartheta \ln \gamma\right] - \frac{1}{\mu}\left(\langle y, z\rangle + \tau y_\tau\right) \\ \text{s.t.} & Fz = Fz^0, \\ & A^\top y = A^\top y^0 - (\tau - 1)c, \\ & y_\tau = y_{\tau,0} + \langle c_A, z - z^0\rangle, \end{array} \tag{25}$$

where F is a matrix whose rows form a basis for the kernel of $A^\top$ and c_A is any vector such that $A^\top c_A = c$. We have $A^\top y = A^\top y^0 - (\tau - 1)c$ iff there exists a vector v such that $y = y^0 - (\tau - 1)c_A - F^\top v$, so over the feasible region of (25) we have

$$\begin{aligned} \langle y, z\rangle + \tau y_\tau &= \langle y^0 - (\tau - 1)c_A - F^\top v, z\rangle + \tau y_{\tau,0} + \tau\langle c_A, z - z^0\rangle \\ &= \langle y^0 + c_A, z\rangle - \langle v, Fz^0\rangle + \tau y_{\tau,0} - \tau\langle c_A, z^0\rangle, \\ &= \langle y^0 + c_A, z\rangle + \langle y, z^0\rangle + \tau y_{\tau,0} - \langle y^0 + c_A, z^0\rangle, \end{aligned} \tag{26}$$

which is linear in (z, τ, y, y_τ). Therefore, the objective function in (25) is the summation of a s.c. function, its LF conjugate and another term that we showed is linear on the feasible region. Hence, by property SC-1 of s.c. functions in Appendix A, the objective function is a s.c. function. Therefore, (25) is minimizing a non-degenerate s.c. function that is bounded from below and so attains its unique minimizer $(\bar{z}, \bar{\tau}, \bar{y}, \bar{y}_\tau)$ by property SC-4 of s.c. functions in Appendix A. We claim that $(\bar{z}, \bar{\tau}, \bar{y}, \bar{y}_\tau)$ satisfies the first equality of (22). Assume that $\frac{1}{\mu}(\hat{y}, \hat{y}_\tau)$ is the image of $(\bar{z}, \bar{\tau})$ under the map (21). Then, we can check that $(\bar{z}, \bar{\tau}, \hat{y}, \hat{y}_\tau)$ also satisfies the optimality conditions and by uniqueness, we have $(\hat{y}, \hat{y}_\tau) = (\bar{y}, \bar{y}_\tau)$. To conclude the proof, $F\bar{z} = Fz^0$ implies that there exists a unique $\bar{x}$ such that $\bar{z} = \bar{\tau}A\bar{x} + z^0$. Therefore, $(\bar{x}, \bar{\tau}, \bar{y})$ is a solution of (22) and so (18) for the fixed μ. Uniqueness follows from the fact that, by using Fenchel-Young inequality (Theorem A.1), the system (22) implies optimality for (25). □

Let us finish this section by a discussion that why following the central path defined above solves the problem for us. First we prove the following key lemma:

Lemma 2.2. *Let $(x, \tau, y) \in Q_{DD}$, $\mu = \mu(x, \tau, y)$, and $\hat{y} := \frac{y}{\tau}$. Then,*

$$\frac{-y_{\tau,0}}{\tau} - \frac{\xi\mu\vartheta + \mu\kappa\sqrt{\vartheta}}{\tau^2} \leq \langle c, x\rangle + \delta_*(\hat{y}|D) \leq \frac{-y_{\tau,0}}{\tau} - \frac{(\xi - 1)\mu\vartheta - \mu\kappa\sqrt{\vartheta}}{\tau^2}, \tag{27}$$

where

$$\kappa := \left\| Ax + \frac{1}{\tau}z^0 - \Phi_*'\left(\frac{\tau}{\mu}y\right)\right\|_{[\Phi_*''(\frac{\tau}{\mu}y)]^{-1}}. \tag{28}$$

Proof. By applying Theorem 2.1 to $k := \frac{\tau^2}{\mu}$ and $\hat{y}$ we get

$$\langle \hat{y}, \Phi_*'\left(\frac{\tau y}{\mu}\right)\rangle \leq \delta_*(\hat{y}|D) \leq \langle \hat{y}, \Phi_*'\left(\frac{\tau y}{\mu}\right)\rangle + \frac{\mu\vartheta}{\tau^2}. \tag{29}$$

Note that by adding and subtracting a term we have

$$\langle \hat{y}, \Phi_*'\left(\frac{\tau y}{\mu}\right)\rangle = \langle \hat{y}, \Phi_*'\left(\frac{\tau y}{\mu}\right) - \left(Ax + \frac{1}{\tau}z^0\right)\rangle + \langle \hat{y}, Ax + \frac{1}{\tau}z^0\rangle. \tag{30}$$

Now by using the fact that $\|\frac{\tau y}{\mu}\|_{\Phi_*''(\frac{\tau}{\mu}y)} \leq \sqrt{\vartheta}$ (property (144)), definition of κ, and using CS inequality (9) we get

$$-\frac{\mu\kappa}{\tau^2}\sqrt{\vartheta} \leq \langle \hat{y}, \Phi_*'\left(\frac{\tau y}{\mu}\right) - \left(Ax + \frac{1}{\tau}z^0\right)\rangle \leq \frac{\mu\kappa}{\tau^2}\sqrt{\vartheta}. \tag{31}$$

By substituting (31) in (30) and the result in (29) we get

$$\langle \hat{y}, Ax + \frac{1}{\tau}z^0\rangle - \frac{\mu\kappa}{\tau^2}\sqrt{\vartheta} \leq \delta_*(\hat{y}|D) \leq \langle \hat{y}, Ax + \frac{1}{\tau}z^0\rangle + \frac{\mu\kappa}{\tau^2}\sqrt{\vartheta} + \frac{\mu\vartheta}{\tau^2}. \tag{32}$$

The result of the lemma follows if we substitute for $\langle \hat{y}, Ax + \frac{1}{\tau}z^0\rangle$ from the definition of μ in the first line of (19). □

2.2. **Brief interpretation of outcomes of the algorithm.** A given problem may have one of several possible statuses. Just in terms of primal feasibility, when the problem is feasible, we can have strict feasibility ($\text{img}A \cap \text{int}D \neq \emptyset$) or otherwise weak feasibility. When the problem is infeasible, we can have weak infeasibility (an arbitrarily small perturbation makes it feasible), or otherwise strict infeasibility. The same analysis also applies to dual feasibility. Next, we discuss what can be said in two of the possible cases about the problem based on the value of τ, using Lemma 2.2:

(i) $\lim_{\mu\to+\infty}\tau = +\infty$ **such that** $\frac{\mu}{\tau^2}$ **tends to zero:** Then x converges to a point $\hat{x}$ that satisfies $A\hat{x} \in D$ and $\frac{y}{\tau}$ converges to a point $\hat{y} \in D_*$ that satisfies $A^\top\hat{y} = -c$. Moreover, Lemma 2.2 implies that the duality gap $\langle c, \hat{x}\rangle + \delta_*(\hat{y}|D)$ is zero. Therefore, by Lemma 2.1, $\hat{x}$ is an optimal solution of the problem.

(ii) τ **stays bounded when** $\mu \to +\infty$**:** In this case, Lemma 2.1 shows that $\langle c, x\rangle + \delta_*(y/\tau|D)$ tends to $-\infty$. If $\langle c, x\rangle$ tends to $-\infty$, ignoring the pathological cases, we can argue that the problem is unbounded. If $\langle c, x\rangle$ stays bounded, then Lemma 2.1 implies that $\bar{y} := \lim_{\mu\to+\infty}\frac{\tau y}{\mu}$ satisfies $\delta_*(\bar{y}|D) < 0$ and we also have $A^\top\bar{y} = 0$. Such a $\bar{y}$ implies primal infeasibility; otherwise, if there exists $A\bar{x} \in D$, then we have the following contradiction.

$$0 > \delta_*(\bar{y}|D) \underbrace{\geq}_{\text{definition of } \delta_*(\bar{y}|D)} \langle \bar{y}, A\bar{x}\rangle = \langle A^\top\bar{y}, \bar{x}\rangle = 0.$$

Now the question is which statuses make the above cases happen? and what is the behavior of (x, τ, y) when μ tends to $+\infty$? Answering these questions requires scrutinizing the geometry of the primal and dual problems and a careful categorization of the possible statuses. Part of this has been done in [24] and a complete study comes in a subsequent article. For example, it is proved in [24] that if the problem is strictly primal and dual feasible (and in a weaker sense if it is just solvable), there exists a parameter $\omega > 0$ (depending on the geometry of the problem) such that $\tau \geq \omega\mu$ for all the points close to the central path. This implies that the first of the above cases happens when $\mu \to +\infty$ and our algorithms return an optimal solution. Also different infeasibility and unboundedness statuses are defined in [24] and it is shown that, ignoring pathological scenarios, the second of the above cases happens and we can extract (approximate and under some conditions exact) certificates of infeasibility or unboundedness out of (x, τ, y) when $\mu \to +\infty$.

3. Algorithms

In the previous section, we defined our infeasible-start primal-dual central path, parameterized with μ. In this section, we express a predictor-corrector path-following algorithm that efficiently follows the path to $\mu = +\infty$. To define neighborhoods of the central path, we need a notion of proximity. For a point $(x, \tau, y) \in Q_{DD}$, defined in (16), we define a proximity measure as

$$\begin{aligned} \Omega_\mu(x,\tau,y) &:= \Phi\left(Ax + \frac{1}{\tau}z^0\right) + \Phi_*\left(\frac{\tau y}{\mu}\right) - \frac{\tau}{\mu}\langle y, Ax + \frac{1}{\tau}z^0\rangle, \\ \mu &:= \mu(x,\tau,y), \quad \text{as defined in (19).} \end{aligned} \tag{33}$$

Throughout the paper, we may drop the arguments of Φ and Φ_* (and also their gradients and Hessians) for simplicity, i.e., $\Phi := \Phi\left(Ax + \frac{1}{\tau}z^0\right)$ and $\Phi_* := \Phi_*\left(\frac{\tau}{\mu}y\right)$.

Remark 3.1. *The proximity measure used for the feasible-start case* [33] *is*

$$\Phi(Ax) + \Phi_*(y) - \langle y, Ax\rangle. \tag{34}$$

Even though this proximity measure and (33) *have similar structures (indeed for* $z^0 = 0$ *and* $\tau = \mu$*, we recover* (34)*),* τ *and* μ *bring nonlinearity into the arguments of* Φ *and* Φ_* *in* (33)*.*

Theorem 3.1. *For every* $(x, \tau, y) \in Q_{DD}$ *and* $\mu > 0$ *we have* $\Omega_\mu(x, \tau, y) \geq 0$*. Moreover,* $\Omega_\mu(x, \tau, y) = 0$ *with* $\mu = \mu(x, \tau, y)$ *iff* (x, τ, y) *is on the central path for parameter* $\mu(x, \tau, y)$*.*

Proof. Both parts of the theorem are implied by Fenchel-Young inequality (Theorem A.1) and the definition of the central path. □

Now, we can state a predictor-corrector algorithm. Note that we choose different step sizes for x and for (τ, y), i.e., for a search direction (d_x, d_τ, d_y), the updates are

$$x^+ := x + \alpha_1 d_x, \quad \tau^+ := \tau + \alpha_2 d_\tau, \quad y^+ := y + \alpha_2 d_y. \tag{35}$$

Framework for Predictor-Corrector Algorithms

Input: $A \in \mathbb{R}^{m\times n}$, $c \in \mathbb{R}^n$, neighborhood parameters $\delta_1, \delta_2 \in (0,1)$ such that $\delta_1 < \delta_2$, desired tolerance $tol \in (0,1)$. Access to gradient and Hessian oracles for a ϑ-s.c. barrier Φ such that $\vartheta \geq 1$ and $\mathrm{dom}\Phi = \mathrm{int}D$, and its LF conjugate Φ_*, $z^0 \in \mathrm{int}D$.
Initialization: $k := 0$, $y^0 := \Phi'(z^0)$, $x^0 := 0$, $\tau_0 := 1$, and $\mu_0 := \mu(x^0, \tau_0, y^0)$.
while (the stopping criteria are not met)

if $(\Omega_{\mu_k}(x^k, \tau_k, y^k) > \delta_1)$

Calculate the corrector search direction (d_x, d_τ, d_y), choose $(\alpha_1, \alpha_2) \in \mathbb{R}^2_+$, and apply the update in (35) to get $(x^{k+1}, \tau_{k+1}, y^{k+1})$, such that $\Omega_{\mu_k}(x^{k+1}, \tau_{k+1}, y^{k+1})$ is smaller than $\Omega_{\mu_k}(x^k, \tau_k, y^k)$ by a "large enough" amount. Define $\mu_{k+1} := \mu_k$.

if $(\Omega_{\mu_k}(x^k, \tau_k, y^k) \leq \delta_1)$

Calculate the predictor search direction (d_x, d_τ, d_y), choose $(\alpha_1, \alpha_2) \in \mathbb{R}^2_+$, and apply the update in (35) to get $(x^{k+1}, \tau_{k+1}, y^{k+1})$, such that $\mu_{k+1} := \mu(x^{k+1}, \tau_{k+1}, y^{k+1})$ is larger than μ_k by a "large enough" amount, while $\Omega_{\mu_{k+1}}(x^{k+1}, \tau_{k+1}, y^{k+1}) \leq \delta_2$.

$k \leftarrow k + 1$.

end while

The best choices for (α_1, α_2) are achieved by a plane search. However, for simplicity of the analysis, it is enough to choose $\alpha_1 := \frac{\alpha_2}{\tau + \alpha_2 d_\tau}$, where α_2 is chosen such that $\tau + \alpha_2 d_\tau > 0$. Then, our search space becomes 1-dimensional and we can choose α_2 large enough to get the desired complexity bounds.

Next, we discuss how to calculate the search directions and choose the step lengths. The Dikin ellipsoid property SC-2 in Appendix A is perhaps the most fundamental property of s.c. functions. This elegant property implies that we can move all the way to the boundary of Dikin ellipsoid and stay feasible. A challenge in our Domain-Driven setup is the nonlinear way that τ is combined with x and y, for example in the proximity measure (33). What typically appears in a primal-dual proximity measure in the literature is that the summation of the s.c. barrier and its LF conjugate is composed with an affine function of the variables, which makes the algorithm and analysis easier. The positive definite matrix that defines the Dikin ellipsoid for our algorithm

has a special form that controls the nonlinear displacements in the arguments of Φ and Φ_* in the proximity measure.

Let us define $\bar{H}(x,\tau)$ as follows (with $u := Ax + \frac{1}{\tau}z^0$)

$$\bar{H}(x,\tau) := \begin{bmatrix} \underbrace{\frac{1}{\tau^2}\Phi''(u)}_{=:H} & \underbrace{\frac{-1}{\tau^2}\Phi''(u)\,u - \frac{1}{\tau^2}\Phi'(u)}_{=:h} \\ \left[\frac{-1}{\tau^2}\Phi''(u)\,u - \frac{1}{\tau^2}\Phi'(u)\right]^\top & \underbrace{\frac{2}{\tau^2}\langle \Phi'(u), u\rangle + \frac{1}{\tau^2}\langle u, \Phi''(u)\,u\rangle + \frac{\xi\vartheta}{\tau^2}}_{=:\zeta} \end{bmatrix}. \tag{36}$$

One can easily verify that for every $(d, d_\tau) \in \mathbb{R}^m \oplus \mathbb{R}$ we have

$$\bar{H}(x,\tau)[(d,d_\tau),(d,d_\tau)] = \left\| \frac{d}{\tau} - \frac{d_\tau}{\tau}u \right\|^2_{\Phi''(u)} - \frac{2d_\tau}{\tau}\left[\frac{d}{\tau} - \frac{d_\tau}{\tau}u\right]^\top \Phi'(u) + \xi\frac{d_\tau^2}{\tau^2}\vartheta. \tag{37}$$

By using the definition of s.c barriers (133) for the second term in the RHS of (37), we have

$$\left[\left\| \frac{d}{\tau} - \frac{d_\tau}{\tau}u \right\|_{\Phi''(u)} - \left|\frac{d_\tau}{\tau}\right|\sqrt{\vartheta}\right]^2 + (\xi - 1)\frac{d_\tau^2}{\tau^2}\vartheta \le \bar{H}(x,\tau)[(d,d_\tau),(d,d_\tau)], \tag{38}$$

which shows that $\bar{H}(x,\tau)$ is a positive definite matrix for every $\xi > 1$.

Remark 3.2. *If we replace u with $\frac{z}{\tau}$ in (36), we get the Hessian for the function $\Phi\left(\frac{z}{\tau}\right) + \xi\vartheta\ln(\tau)$, which is a s.c. function (see Lemma A.2).*

$\bar{H}(x,\tau)$ is positive definite for $\xi > 1$ and so invertible. Considering the definition of H, h, and ζ in (36), by substitution, one can directly verify that for every $(w, w_\tau) \in \mathbb{R}^m \oplus \mathbb{R}$, we have

$$\begin{bmatrix} w \\ w_\tau \end{bmatrix}^\top \left(\begin{bmatrix} H & h \\ h^\top & \zeta \end{bmatrix}\right)^{-1} \begin{bmatrix} w \\ w_\tau \end{bmatrix} = \langle w, H^{-1}w\rangle + \eta\left(\langle w, H^{-1}h\rangle - w_\tau\right)^2,$$

$$H^{-1}h = -u - [\Phi'']^{-1}\Phi', \qquad \eta = \frac{\tau^2}{\xi\vartheta - \langle \Phi', [\Phi'']^{-1}\Phi'\rangle}. \tag{39}$$

Note that $\eta \ge 0$ by using $\xi > 1$ and property (134). The following key lemma, which we prove later, shows that when we are close to the central path, $\bar{H}(x,\tau)$ is under control.

Lemma 3.1. *For every $\bar{\epsilon} \in (0,1)$, there exist $\epsilon > 0$ depending on ξ such that for every pair $(x,\tau,y) \in Q_{DD}$ and $\mu > 0$ with $\Omega_\mu(x,\tau,y) \le \epsilon$, we have*

$$(1-\bar{\epsilon})^2\bar{H}(x(\mu),\tau(\mu)) \preceq \bar{H}(x,\tau) \preceq \frac{1}{(1-\bar{\epsilon})^2}\bar{H}(x(\mu),\tau(\mu)). \tag{40}$$

Let F be a matrix whose rows give a basis for the kernel of $A^\top$ and let c_A be any vector such that $A^\top c_A = c$. We define a block matrix U that comes up frequently in our discussion and

contains the linear transformations we need, and also the vector r^0 that is used in the RHS of our systems:

$$U := \begin{bmatrix} A & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -c_A & -F^\top \\ c^\top & 0 & 0 \end{bmatrix}, \quad r^0 := \begin{bmatrix} -A^\top y^0 - c \\ -y_{\tau,0} + \langle c_A, z^0 \rangle \\ F z^0 \end{bmatrix}. \tag{41}$$

At a current point (x, τ, y), both the predictor and corrector steps are derived by solving the system

$$\underbrace{U^\top \begin{bmatrix} \bar{H}(x,\tau) & 0 \\ 0 & \left[\hat{H}(x,\tau,y)\right]^{-1} \end{bmatrix} U}_{\mathcal{H}(\bar{H},\hat{H})} \begin{bmatrix} \bar{d}_x \\ d_\tau \\ d_v \end{bmatrix} = r_{RHS},$$

$$d_x := \bar{d}_x - d_\tau x, \qquad d_y := -d_\tau c_A - F^\top d_v, \tag{42}$$

where $\hat{H}(x, \tau, y)$ is a positive definite matrix that we elaborate more on later. For both the predictor and corrector steps, we discuss the choice of $\hat{H}(x, \tau, y)$ and r_{RHS} in (42).

3.1. **Predictor step.** An efficient predictor search direction must increase μ by a large rate and at the same time let us take a long enough step. We first give the choices of $\hat{H}(x, \tau, y)$ and r_{RHS} for the system in (42) and then justify them. For the RHS vector we choose $r_{RHS} := r^0/\mu^2$, where r_0 defined in (41). We have different choices for $\hat{H}(x, \tau, y)$ to attain our desired properties (such as a low complexity bound). We express a sufficient condition and discuss two choices that satisfy the condition. We will see that to achieve enough increase in μ at every predictor step, it is sufficient that for every $\bar{\epsilon} \in (0, 1)$, there exists a choice of ϵ in Lemma 3.1 such that

$$(1-\bar{\epsilon})^2 \left[\bar{H}(x(\mu),\tau(\mu))\right]^{-1} \preceq \mu^2 \left[\hat{H}(x,\tau,y)\right]^{-1} \preceq \frac{1}{(1-\bar{\epsilon})^2} \left[\bar{H}(x(\mu),\tau(\mu))\right]^{-1}, \tag{43}$$

for every point $(x, \tau, y) \in Q_{DD}$ with $\Omega_\mu(x, \tau, y) \leq \epsilon$.

Remark 3.3. *In view of* (40)*, one obvious choice for* $\hat{H}(x, \tau, y)$ *is* $\hat{H}(x, \tau, y) := \mu^2 \bar{H}(x, \tau)$*. Another choice is one that yields the predictor direction for the primal-dual conic setup given in* [34, 41]*. More explicitly, if we reformulate our problem in the Domain-Driven setup as a conic optimization problem (see* [33] *or* [37]*-Section 5.1), then the predictor step calculated in* [34, 41] *can be achieved by* (41) *for a special choice of* $\hat{H}$*:*

$$\left[\hat{H}(x,\tau,y)\right]^{-1} := \begin{bmatrix} G + \eta_* h_* h_*^\top & -\eta_* h_* \\ -\eta_* h_*^\top & \eta_* \end{bmatrix},$$

$$G := \bar{\tau}^2 \Phi_*''(\bar{\tau} y), \quad h_* := -\Phi_*'(\bar{\tau} y) - \bar{\tau}\Phi_*''(\bar{\tau} y) y, \quad 1/\eta_* := \frac{\xi \vartheta}{\bar{\tau}^2} - \langle y, \Phi_*''(\bar{\tau} y) y \rangle, \tag{44}$$

where

$$\bar{\tau} := \operatorname{argmax}_\tau \left\{ \Phi_*(\tau y) + y_\tau \tau + \xi\vartheta \ln \tau \right\}. \tag{45}$$

To check that (44) *satisfies condition* (43) *for every* $\bar{\epsilon} \in (0,1)$ *for a right choice of* ϵ*, we can use the arguments in* [34, 41]*, or the fact that* $\Phi\left(\frac{z}{\tau}\right) + \xi\vartheta \ln(\tau)$ *is a s.c. function, Lemma 3.1, and the properties of LF conjugates. Calculating* $\bar{\tau}$ *can be done efficiently, since evaluating the RHS of* (45) *is equivalent to minimizing a s.c. function.*

Let us justify our predictor step. If we choose $\alpha_1 = \frac{\alpha_2}{\tau+\alpha_2 d_\tau}$ (assuming $\tau + \alpha_2 d_\tau > 0$) for the updates in (35), then by using the third line of (19) for μ, we have

$$\begin{aligned}
& \mu(x^+, \tau^+, y^+) - \mu(x, \tau, y) \\
= \; & \tfrac{-1}{\xi\vartheta}\left[\alpha_2 \langle d_y, z^0\rangle + \alpha_2 d_\tau y_{\tau,0} + \langle c + A^\top y^0, \alpha_2 d_\tau x + (\tau + \alpha_2 d_\tau)\alpha_1 d_x\rangle\right] \\
= \; & \tfrac{-\alpha_2}{\xi\vartheta}\left(\langle d_y, z^0\rangle + d_\tau y_{\tau,0} + \langle c + A^\top y^0, d_\tau x + d_x\rangle\right), && \text{substituting } \alpha_1 = \tfrac{\alpha_2}{\tau+\alpha_2 d_\tau}, \\
= \; & \tfrac{\alpha_2}{\xi\vartheta}\left(\langle d_v, F z^0\rangle + d_\tau(\langle c_A, z^0\rangle - y_{\tau,0}) - \langle c + A^\top y^0, \bar{d}_x\rangle\right), && \text{substituting } d_x \text{ and } d_y \text{ from (42)}, \\
= \; & \tfrac{\alpha_2}{\xi\vartheta}[\bar{d}_x^\top \quad d_\tau \quad d_v^\top] r^0, && \text{for } r^0 \text{ defined in (41)}.
\end{aligned} \tag{46}$$

Let $d^\top := [\bar{d}_x^\top \;\; d_\tau \;\; d_v^\top]$, then we see that the Dikin ellipsoid type constraint $d^\top U^\top \mathcal{H}(\bar{H}, \hat{H}) U d \le 1$ guarantees the feasibility of new iterates with respect to the domains of the underlying s.c. functions. The search direction in (42) is, up to some scaling, the solution of the following optimization problem

$$\begin{aligned}
\max \quad & \langle d, r^0\rangle \\
s.t. \quad & d^\top U^\top \mathcal{H}(\bar{H}, \hat{H}) U d \le 1,
\end{aligned} \tag{47}$$

which can be seen as maximizing the linear function of (46) in a trust region.

3.2. **Corrector step.** After doing a predictor step to increase μ, we need to perform corrector steps to come back into the small neighborhood. Note that our proximity measure $\Omega_\mu(x, \tau, y)$ is not a convex function and to decrease it we use a quasi-Newton like step. In most of the literature on this topic, for example papers [34, 41, 33], the corrector step is simply minimizing a s.c. function that can be done efficiently by taking damped Newton steps (Appendix A, SC-3). Even though our proximity measure is not a s.c. function and we cannot directly use damped Newton steps, Φ and Φ_* are 1-s.c. functions and we can exploit their properties. We first define the corrector step and then explain our choice. The corrector search direction is the solution of (42) with

$$\hat{H} := \mu^2 \bar{H}, \quad r_{RHS} := -(U^\top \psi^c + \beta r^0), \quad \beta := -\frac{\langle r^0, [U^\top \mathcal{H}(\bar{H}, \mu^2 \bar{H}) U]^{-1} U^\top \psi^c\rangle}{\langle r^0, [U^\top \mathcal{H}(\bar{H}, \mu^2 \bar{H}) U]^{-1} r^0\rangle}, \tag{48}$$

where r_0 is defined in (41) and

$$\psi^c := \begin{bmatrix} \frac{1}{\tau}\Phi' \\ -\frac{1}{\tau}\langle \Phi', Ax + \frac{1}{\tau}z^0\rangle + \frac{1}{\mu}\langle y, \Phi'_*\rangle + \frac{1}{\mu}(y_{\tau,0} + \tau\langle c, x\rangle) \\ \frac{\tau}{\mu}\Phi'_* \\ \frac{\tau}{\mu} \end{bmatrix}. \tag{49}$$

Remark 3.4. *If we choose $\alpha_1 = \frac{\alpha_2}{\tau+\alpha_2 d_\tau}$ (assuming $\tau+\alpha_2 d_\tau > 0$) for the updates in* (35)*, then* (46) *holds. The parameters in* (48) *are chosen so that the solution of* (42) *satisfies $[\bar{d}_x^\top \;\; d_\tau \;\; d_v^\top]r^0 = 0$ and thus, we automatically have $\mu(x^+, \tau^+, y^+) = \mu(x, \tau, y)$ in the corrector step.*

The following lemma justifies our corrector search direction.

Lemma 3.2. *Consider a choice of r_{RHS} and $\hat{H}(x, \tau, y)$ such that for the solution of* (42) *and the updates in* (35) *with $\alpha_1 = \frac{\alpha_2}{\tau+\alpha_2 d_\tau}$ we have $\mu(x^+, \tau^+, y^+) = \mu(x, \tau, y)$. Then,*

$$\begin{aligned} \rho(D(\alpha_2)) &\leq \Omega_\mu(x^+, \tau^+, y^+) - \Omega_\mu(x, \tau, y) - \alpha_2 \left[\bar{d}_x^\top \;\; d_\tau \;\; d_v^\top\right] U^\top \psi^c \\ &\quad + \tfrac{\alpha_2^2 d_\tau}{\tau(\tau+\alpha_2 d_\tau)}\langle \Phi', A\bar{d}_x - d_\tau\left(Ax + \tfrac{1}{\tau}z^0\right)\rangle - \tfrac{\alpha_2^2 d_\tau}{\mu}\left(\langle d_y, \Phi'_*\rangle + \langle c, \bar{d}_x\rangle\right) \\ &\leq \rho\left(-D(\alpha_2)\right), \\ D(\alpha_2) &:= \tfrac{\alpha_2}{\tau+\alpha_2 d_\tau}\left\|A\bar{d}_x - d_\tau\left(Ax + \tfrac{1}{\tau}z^0\right)\right\|_{\Phi''} + \alpha_2\left\|\tfrac{d_\tau y + (\tau+\alpha_2 d_\tau)d_y}{\mu}\right\|_{\Phi''_*}, \end{aligned} \tag{50}$$

where ψ^c is defined in (49) *and ρ is defined in* (118)*.*

Proof. (50) is derived by substituting for Ω_μ from (33) and then using both bounds in (125) for the 1-s.c. function $f(u, w) = \Phi(u) + \Phi_*(w)$. We just need to explicitly calculate the displacements in the arguments of Φ and Φ_*. By the hypothesis, $\mu^+ := \mu(x^+, \tau^+, y^+) = \mu$. First we have

$$\begin{aligned} Ax^+ + \tfrac{1}{\tau^+}z^0 - Ax - \tfrac{1}{\tau}z^0 &= \alpha_1 A d_x - \tfrac{\alpha_2 d_\tau}{\tau(\tau+\alpha_2 d_\tau)}z^0 \\ &= \tfrac{\alpha_2}{\tau+\alpha_2 d_\tau}\left[Ad_x - \tfrac{d_\tau}{\tau}z^0\right], && \text{using } \alpha_1 = \tfrac{\alpha_2}{\tau+\alpha_2 d_\tau}, \\ &= \tfrac{\alpha_2}{\tau+\alpha_2 d_\tau}\left[A\bar{d}_x - d_\tau\left(Ax + \tfrac{1}{\tau}z^0\right)\right], && \text{using } d_x = \bar{d}_x - d_\tau x. \end{aligned} \tag{51}$$

For displacement in the argument of Φ_*, we have

$$\frac{\tau^+ y^+}{\mu^+} - \frac{\tau y}{\mu} = \frac{\alpha_2 d_\tau y + \alpha_2 \tau d_y + \alpha_2^2 d_\tau d_y}{\mu}.$$

As an intermediate step, similar to (51), by substituting $\alpha_1 = \frac{\alpha_2}{\tau+\alpha_2 d_\tau}$ and $d_x = \bar{d}_x - d_\tau x$, we have

$$\tau^+ x^+ = (\tau + \alpha_2 d_\tau)(x + \alpha_1 d_x) = \tau x + \alpha_2 \bar{d}_x. \tag{52}$$

Then, by using $\mu^+ = \mu$ and the first line of (19), and then substituting (52), we have

$$\begin{aligned} -\langle \tfrac{\tau^+ y^+}{\mu^+}, Ax^+ + \tfrac{1}{\tau^+}z^0\rangle + \langle \tfrac{\tau y}{\mu}, Ax + \tfrac{1}{\tau}z^0\rangle &= -\tfrac{\tau^+}{\mu}[-y_{\tau,0} - \tau^+\langle c, x^+\rangle] + \tfrac{\tau}{\mu}[-y_{\tau,0} - \tau\langle c, x\rangle] \\ &= \tfrac{\alpha_2}{\mu}[d_\tau y_{\tau,0} + \tau\langle c, \bar{d}_x\rangle + d_\tau\langle c, \tau x + \alpha_2 \bar{d}_x\rangle]. \end{aligned}$$

We can verify by direct substitution that

$$U^\top \psi^c = \begin{bmatrix} \frac{1}{\tau}A^\top \Phi' + \frac{1}{\mu}\tau c \\ -\frac{1}{\tau}\langle \Phi', Ax + \frac{1}{\tau}z^0\rangle + \frac{1}{\mu}\langle y - \tau c_A, \Phi'_*\rangle + \frac{1}{\mu}(y_{\tau,0} + \tau\langle c, x\rangle) \\ -\frac{\tau}{\mu}F\Phi'_* \end{bmatrix}. \tag{53}$$

If we also use the equality $F^\top d_v = -d_y - d_\tau c_A$, then we have

$$\begin{aligned} \begin{bmatrix} \bar{d}_x^\top & d_\tau & d_v^\top \end{bmatrix} U^\top \psi^c &= \frac{1}{\tau}\langle \Phi', A\bar{d}_x - d_\tau\left(Ax + \frac{1}{\tau}z^0\right)\rangle + \frac{\tau}{\mu}\langle c, \bar{d}_x\rangle \\ &\quad + \frac{d_\tau}{\mu}\langle y, \Phi'_*\rangle + \frac{\tau}{\mu}\langle d_y, \Phi'_*\rangle + \frac{d_\tau}{\mu}(y_{\tau,0} + \tau\langle c, x\rangle). \end{aligned} \tag{54}$$

By substituting all the above equations we get (50). □

In view of (46), $\mu^+ = \mu$ is equivalent to $\langle d, r^0\rangle = 0$ for $d^\top := \begin{bmatrix} \bar{d}_x^\top & d_\tau & d_v^\top \end{bmatrix}$. The corrector search direction in (48) is, up to some scaling, the optimal solution of

$$\begin{aligned} \min \quad & \langle d, U^\top \psi^c\rangle \\ s.t. \quad & \langle d, r^0\rangle = 0 \\ & d^\top U^\top \mathcal{H}(\bar{H}, \mu^2\bar{H}) U d \le 1. \end{aligned} \tag{55}$$

Before a concrete analysis, to intuitively justify this search direction using (50), note that our goal is to minimize $\Omega_\mu(x^+, \tau^+, y^+) - \Omega_\mu(x, \tau, y)$. The coefficient of $\langle d, U^\top \psi^c\rangle$ is α_2, whereas all the other terms are (almost) proportional to α_2^2. Therefore, we can look at $\alpha_2\langle d, U^\top \psi^c\rangle$ as the first order approximation of $\Omega_\mu(x^+, \tau^+, y^+) - \Omega_\mu(x, \tau, y)$ that we minimize in (55) in a trust region.

Remark 3.5. *What we prove for the corrector step above is enough for the purposes of obtaining the desired complexity results. However, corrector steps in most of the other papers in this context (such as* [34, 41, 33]*) is simply minimizing a s.c. function and have the stronger property of quadratic convergence for the points close enough to the central path (see property* (127)*-(c)). Proving asymptotic quadratic convergence for a suitable variant of our algorithm is a future goal.*

4. Analysis of the algorithms

In this section, we analyze the predictor and corrector steps we defined in the previous section. This analysis lets us modify the framework for primal-dual algorithms in Section 3 to achieve the current best iteration complexity bounds. This modification and the main theorem about it come in Section 4.3. The following lemma shows how to bound the proximity measure (33) based on the local norm defined by the current primal and dual iterates:

Lemma 4.1. *(a) Assume that $f(x)$ is an a-s.c. function and let $f_*(y)$ be its LF conjugate. Then, for every x and y in the domains of f and f_* we have*

$$\rho\left(r\right) \leq f(x) + f_*(y) - \langle y, x\rangle \leq \rho\left(-r\right), \tag{56}$$

where $r := a^{-1/2}\|y - f'(x)\|_{[f''(x)]^{-1}}$ and ρ is defined in (118).
(b) Moreover, assume that there exist $\hat{x}$ and $\hat{y}$ in the domains of f and f_* respectively such that $\hat{y} = f'(\hat{x})$ and $\langle x - \hat{x}, y - \hat{y}\rangle = 0$. Then,

$$\rho(r) + \rho(s) \leq f(x) + f_*(y) - \langle y, x\rangle \leq \rho(-r) + \rho(-s), \tag{57}$$

where $r := a^{-1/2}\|x - \hat{x}\|_{f''(\hat{x})}$ and $s := a^{-1/2}\|y - \hat{y}\|_{f_*''(\hat{y})}$.

Proof. (a) By writing the second inequality in (125) for f_* at two points y and $f'(x)$, we have

$$f_*(y) \leq f_*(f'(x)) + \langle f_*'(f'(x)), y - f'(x)\rangle + \rho(-a^{-1/2}\|y - f'(x)\|_{f_*''(f'(x))}).$$

To get the RHS inequality in (56), we substitute $f_*'(f'(x)) = x$ and $f_*''(f'(x)) = [f''(x)]^{-1}$ from (131), and $f_*(f'(x)) + f(x) = \langle f'(x), x\rangle$ from Theorem A.1. The LHS inequality can be similarly proved by using the first inequality in (125).
(b) We write the property (125) for f at x and $\hat{x}$ and for f_* at y and $\hat{y}$, and add them together. □

Corollary 4.1. *For every $(x, \tau, y) \in Q_{DD}$, we have*

$$\rho\left(\left\|\frac{\tau y}{\mu} - \Phi'\left(u\right)\right\|_{[\Phi''(u)]^{-1}}\right) \leq \Omega_\mu(x, \tau, y) \leq \rho\left(-\left\|\frac{\tau y}{\mu} - \Phi'\left(u\right)\right\|_{[\Phi''(u)]^{-1}}\right), \tag{58}$$

where $\mu := \mu(x, \tau, y)$ and $u := Ax + \frac{1}{\tau}z^0$.

As we explained before, matrix $\bar{H}$ in (36) defines Dikin ellipsoid type properties that are crucial in our analysis. To elaborate more, in both the predictor and corrector steps, we have a vector $d^\top := [\bar{d}_x^\top \;\; d_\tau \;\; d_v^\top]$ as the solution of (42) which satisfies $d^\top U^\top \mathcal{H}(\bar{H}, \mu^2\bar{H})Ud \leq q$ for $\mathcal{H}$ defined in (42) and a scalar q. Let us define

$$f := Ud = U\begin{bmatrix} \bar{d}_x \\ d_\tau \\ d_v \end{bmatrix} \underbrace{=}_{(41)} \begin{bmatrix} A\bar{d}_x \\ d_\tau \\ -d_\tau c_A - F^\top d_v \\ \langle c, \bar{d}_x\rangle \end{bmatrix} \underbrace{=}_{(42)} \begin{bmatrix} A\bar{d}_x \\ d_\tau \\ d_y \\ \langle c, \bar{d}_x\rangle \end{bmatrix}.$$

Using (38) and (39), $f^\top \mathcal{H}(\bar{H}, \mu^2\bar{H})f \leq q$ yields

$$\begin{aligned} &\left[\left\|\frac{A\bar{d}_x}{\tau} - \frac{d_\tau}{\tau}\left(Ax + \frac{1}{\tau}z^0\right)\right\|_{\Phi''} - \left|\frac{d_\tau}{\tau}\right|\sqrt{\vartheta}\right]^2 + (\xi - 1)\frac{d_\tau^2}{\tau^2}\vartheta \\ &+ \frac{\tau^2}{\mu^2}\langle d_y, [\Phi'']^{-1}d_y\rangle + \frac{\left[\langle \frac{\tau d_y}{\mu}, [\Phi'']^{-1}\Phi'\rangle + \frac{\tau}{\mu}(\langle d_y, Ax + \frac{1}{\tau}z^0\rangle + \langle c, \bar{d}_x\rangle)\right]^2}{\xi\vartheta - \langle \Phi', [\Phi'']^{-1}\Phi'\rangle} \leq q. \end{aligned} \tag{59}$$

We can break down (59) into several useful bounds for our analysis. First, clearly

$$(\xi-1)\frac{d_\tau^2}{\tau^2}\vartheta \le q \quad \Rightarrow \quad \left(\frac{d_\tau}{\tau}\right)^2 \le \frac{q}{(\xi-1)\vartheta}. \tag{60}$$

Using (59) and (60), we get

$$\frac{1}{\tau}\left\|A\bar{d}_x - d_\tau\left(Ax+\frac{1}{\tau}z^0\right)\right\|_{\Phi''} \le \sqrt{q} + \left|\frac{d_\tau}{\tau}\right|\sqrt{\vartheta} \underbrace{\le}_{(60)} \left(1+\sqrt{\frac{1}{\xi-1}}\right)\sqrt{q}. \tag{61}$$

(61) gives a bound on the displacement in $Ax+\frac{1}{\tau}z^0$ as shown in (51). Also from (59) we get

$$\frac{\tau^2}{\mu^2}\langle d_y, [\Phi'']^{-1}d_y\rangle \le q. \tag{62}$$

Let us see how to use these bounds in the analysis of the predictor and corrector steps.

4.1. **Predictor step.** Let us first show how the predictor step increases μ. For analyzing this, we prove a result about the structure of U defined in (41). We start with a lemma:

Lemma 4.2. *Assume that $\mathcal{H}$ is a symmetric positive definite matrix and U is a matrix of proper size with linearly independent columns. Then, for any given vector f of proper size, we have*

$$f^\top U\left(U^\top\mathcal{H}U\right)^{-1}U^\top f = f^\top\mathcal{H}^{-1}f - f^\top\mathcal{H}^{-1}U^{\perp\top}\left(U^\perp\mathcal{H}^{-1}U^{\perp\top}\right)^{-1}U^\perp\mathcal{H}^{-1}f, \tag{63}$$

where $U^\perp$ is a matrix whose rows form a basis for the kernel of $U^\top$.

Proof. As $\mathcal{H}$ is symmetric positive definite and U has linearly independent columns, the system $U^\top\mathcal{H}Ug = U^\top f$ has a unique solution g. By definition of $U^\perp$, there exists w such that $\mathcal{H}Ug = f + U^{\perp\top}w$. Multiplying both sides by $\mathcal{H}^{-1}$ gives us $Ug = \mathcal{H}^{-1}f + \mathcal{H}^{-1}U^{\perp\top}w$. To calculate w, we multiply both sides of the last equation from the left by $U^\perp$. Note that $U^\perp U = 0$ and $U^\perp\mathcal{H}^{-1}U^{\perp\top}$ is invertible. If we solve for w and substitute it in $Ug = \mathcal{H}^{-1}f + \mathcal{H}^{-1}U^{\perp\top}w$, we get

$$Ug = \mathcal{H}^{-1}f - \mathcal{H}^{-1}U^{\perp\top}\left(U^\perp\mathcal{H}^{-1}U^{\perp\top}\right)^{-1}U^\perp\mathcal{H}^{-1}f. \tag{64}$$

If we multiply both sides of (64) from the left by $f^\top$ and substitute $g = (U^\top\mathcal{H}U)^{-1}U^\top f$, we get (63). □

We are interested in matrix $U \in \mathbb{R}^{(2m+2)\times(m+1)}$ we defined in (41), which has a very special structure. For this U, one option for $U^\perp$, defined in Lemma 4.2, is

$$U^\perp = \begin{bmatrix} 0 & c & A^\top & 0 \\ -c_A & 0 & 0 & 1 \\ -F & 0 & 0 & 0 \end{bmatrix}. \tag{65}$$

If we compare U and $U^{\perp}$, we see that the rows of U is a permutation of the columns of $U^{\perp}$. Explicitly

$$U^{\perp} = U^{\top} P, \qquad P := \begin{bmatrix} 0 & I_{m+1} \\ I_{m+1} & 0 \end{bmatrix}. \tag{66}$$

We have the following lemma:

Lemma 4.3. *Let $\bar{H}$ be a symmetric positive definite matrix and $\mu > 0$. Assume the setup of Lemma 4.2 where $\mathcal{H}$ and f have the form*

$$\mathcal{H} := \begin{bmatrix} \bar{H} & 0 \\ 0 & \frac{1}{\mu^2}\bar{H}^{-1} \end{bmatrix}, \quad f := \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}, \tag{67}$$

such that f_1 and f_2 further satisfy $f_1 = \mu \bar{H} f_2$ or $f_1 = -\mu \bar{H} f_2$. Also assume that (66) holds for U and $U^{\perp}$. Then,

$$f^{\top} U \left(U^{\top} \mathcal{H} U \right)^{-1} U^{\top} f = \frac{1}{2} f^{\top} \mathcal{H}^{-1} f. \tag{68}$$

Proof. We can verify that $\mathcal{H}^{-1} = \mu^2 P \mathcal{H} P$ for P defined in (66). Using this and (66), for the second term in the RHS of (63) we have

$$f^{\top} \mathcal{H}^{-1} U^{\perp\top} \left(U^{\perp} \mathcal{H}^{-1} U^{\perp\top} \right)^{-1} U^{\perp} \mathcal{H}^{-1} f = \mu^2 \begin{bmatrix} f_2 \\ f_1 \end{bmatrix}^{\top} \mathcal{H} U \left(U^{\top} \mathcal{H} U \right)^{-1} U^{\top} \mathcal{H} \begin{bmatrix} f_2 \\ f_1 \end{bmatrix}. \tag{69}$$

Using $f_1 = \mu \bar{H} f_2$ or $f_1 = -\mu \bar{H} f_2$, (69) equals $f^{\top} U \left(U^{\top} \mathcal{H} U \right)^{-1} U^{\top} f$ and so (63) reduces to (68). □

Let us see how Lemma 4.3 is useful for our setup. We define

$$\psi^p := \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}, \quad f_1 := \begin{bmatrix} \frac{1}{\tau}\Phi' \\ -\frac{1}{\tau}\langle \Phi', Ax + \frac{1}{\tau} z^0 \rangle - \frac{\xi\vartheta}{\tau} \end{bmatrix}, \quad f_2 := \begin{bmatrix} \frac{\tau}{\mu}\left(Ax + \frac{1}{\tau} z^0 \right) \\ \frac{\tau}{\mu} \end{bmatrix}. \tag{70}$$

For matrix $\bar{H}$ defined in (36), we can directly verify

$$\frac{1}{\mu} \begin{bmatrix} \frac{1}{\tau}\Phi' \\ -\frac{1}{\tau}\langle \Phi', Ax + \frac{1}{\tau} z^0 \rangle - \frac{\xi\vartheta}{\tau} \end{bmatrix} = -\bar{H} \begin{bmatrix} \frac{\tau}{\mu}\left(Ax + \frac{1}{\tau} z^0 \right) \\ \frac{\tau}{\mu} \end{bmatrix}. \tag{71}$$

Therefore, $f_1 = -\mu \bar{H} f_2$ and so (68) holds for our setup. Now, we prove the following lemma:

Lemma 4.4. *Consider $\mathcal{H}$ defined in (42) and ψ^p defined in (70) for a point $(x, \tau, y) \in Q_{DD}$. Then, we have*

$$\langle U^{\top} \psi^p, \left[U^{\top} \mathcal{H}\left(\bar{H}, \mu^2 \bar{H} \right) U \right]^{-1} U^{\top} \psi^p \rangle = \xi\vartheta. \tag{72}$$

Proof. (71) confirms that $f_1 = -\mu\bar{H} f_2$, so we have equation (68). Hence, we need to show that $(\psi^p)^\top \mathcal{H}^{-1}\psi^p = 2\xi\vartheta$ to get our result. This holds since by direct verification we have

$$-\mu \begin{bmatrix} \frac{1}{\tau}\Phi' \\ -\frac{1}{\tau}\langle \Phi', Ax + \frac{1}{\tau}z^0\rangle - \frac{\xi\vartheta}{\tau} \end{bmatrix}^\top \begin{bmatrix} \frac{\tau}{\mu}\left(Ax + \frac{1}{\tau}z^0\right) \\ \frac{\tau}{\mu} \end{bmatrix} = \xi\vartheta, \tag{73}$$

and $(\psi^p)^\top \mathcal{H}^{-1}\psi^p$, by using (71), is exactly the summation of two terms like (73). □

Now we are ready to prove the following main proposition about how the predictor step increases μ.

Proposition 4.1. *Assume that $(x, \tau, y) \in Q_{DD}$ and conditions (40) and (43) hold. Let our search direction be the solution of (42) with $r_{RHS} = r^0/\mu^2$ and any $\hat{H}$ that satisfies (43). Let $\alpha_2 > 0$ be such that $\tau + \alpha_2 d_\tau > 0$ and choose $\alpha_1 = \frac{\alpha_2}{\tau + \alpha_2 d_\tau}$. Then, for the updates in (35) we have*

$$(1-\bar{\epsilon})^2\alpha_2 \le \mu(x^+, \tau^+, y^+) - \mu(x, \tau, y) \le \frac{\alpha_2}{(1-\bar{\epsilon})^2}. \tag{74}$$

Proof. A key to the proof is that on the central path we have $U^\top \psi^p(\mu) = -\frac{1}{\mu}r^0$, where ψ^p is defined in (70) and r^0 is defined in (42). This can be directly verified by using (18) and (19) for the points on the central path. By starting from (46) for $\mu(x^+, \tau^+, y^+) - \mu(x, \tau, y)$, we can continue

$$\begin{aligned} & \mu(x^+, \tau^+, y^+) - \mu(x, \tau, y) \\ = \; & \tfrac{\alpha_2}{\xi\vartheta}[\bar{d}_x^\top \quad d_\tau \quad d_v^\top] r^0 \\ = \; & \tfrac{\alpha_2}{\xi\vartheta}\tfrac{1}{\mu^2}\langle r^0, [U^\top \mathcal{H}(\bar{H}, \hat{H})U]^{-1} r^0\rangle, && \text{using (42)}, \\ = \; & \tfrac{\alpha_2}{\xi\vartheta}\langle U^\top\psi^p(\mu), [U^\top \mathcal{H}(\bar{H}, \hat{H})U]^{-1} U^\top \psi^p(\mu)\rangle, && \text{using } U^\top\psi^p(\mu) = -\tfrac{1}{\mu}r^0. \end{aligned} \tag{75}$$

We get the desired result by using conditions (40) and (43) and then utilizing Lemma 4.4 for the points on the central path. □

Proposition 4.1 implies that the amount of increase in μ depends directly on α_2. Therefore, we need to show how large α_2 can be chosen in the predictor step.

Lemma 4.5. *Assume that $(x, \tau, y) \in Q_{DD}$ and conditions (40) and (43) hold. Then, (59) holds with $q := \frac{1}{(1-\bar{\epsilon})^6}\frac{\xi\vartheta}{\mu^2}$ for the solution of (42) with $r_{RHS} = r^0/\mu^2$ and any $\hat{H}$ that satisfies (43).*

Proof. Let us define $f = Ud$ for d the solution of (42). Then, by using (40) and (43), we have

$$
\begin{array}{rcll}
f^\top \mathcal{H}(\bar{H}, \mu^2 \bar{H}) f & = & \frac{1}{\mu^4} \langle \left[U^\top \mathcal{H}(\bar{H}, \hat{H}) U\right]^{-1} r^0, (U^\top \mathcal{H}(\bar{H}, \mu^2 \bar{H}) U) \left[U^\top \mathcal{H}(\bar{H}, \hat{H}) U\right]^{-1} r^0 \rangle & \\
& \leq & \frac{1}{(1-\bar{\epsilon})^4 \mu^4} \langle r^0, \left[U^\top \mathcal{H}(\bar{H}, \hat{H}) U\right]^{-1} r^0 \rangle, & \text{using (40) and (43),} \\
& \leq & \frac{1}{(1-\bar{\epsilon})^6 \mu^4} \langle r^0, \left[U^\top \mathcal{H}\left(\bar{H}(\mu), \mu^2 \bar{H}(\mu)\right) U\right]^{-1} r^0 \rangle, & \text{using (40),} \\
& = & \frac{\langle U^\top \psi^p(\mu), \left[U^\top \mathcal{H}\left(\bar{H}(\mu), \mu^2 \bar{H}(\mu)\right) U\right]^{-1} U^\top \psi^p(\mu) \rangle}{(1-\bar{\epsilon})^6 \mu^2}, & \text{using } U^\top \psi^p(\mu) = -\frac{1}{\mu} r^0, \\
& = & \frac{1}{(1-\bar{\epsilon})^6 \mu^2} \xi \vartheta, & \text{using Lemma 4.4.}
\end{array} \tag{76}
$$

□

We want to control the change in $\Omega_\mu(x, \tau, y)$ by using Corollary 4.1. In view of this, by adding and subtracting some terms, we have (with $\mu^+ := \mu(x^+, \tau^+, y^+)$)

$$
\left(\frac{\tau^+ y^+}{\mu^+} - \Phi'\left(u^+\right)\right) - \left(\frac{\tau y}{\mu} - \Phi'\left(u\right)\right) = \left(\frac{\tau^+}{\mu^+} - \frac{\tau}{\mu}\right) y + \frac{\tau^+}{\mu^+} \alpha_2 d_y - \left(\Phi'\left(u^+\right) - \Phi'\left(u\right)\right). \tag{77}
$$

Let us give a bound on the local norm defined by Φ'' on the three terms in (77). Note that using Proposition 4.1, we have

$$
\left|\frac{\tau^+}{\mu^+} - \frac{\tau}{\mu}\right| = \left|\frac{\tau + \alpha_2 d_\tau}{\mu^+} - \frac{\tau}{\mu}\right| = \left|\frac{\alpha_2 \mu d_\tau - \tau(\mu^+ - \mu)}{\mu \mu^+}\right| \leq \alpha_2 \left(\left|\frac{d_\tau}{\tau}\right| + \left|\frac{1}{\mu(1-\bar{\epsilon})^2}\right|\right) \frac{\tau}{\mu}. \tag{78}
$$

$\Omega_\mu(x, \tau, y) \leq \delta_1$ and (58) imply that $\left\|\frac{\tau y}{\mu} - \Phi'\left(u\right)\right\|_{[\Phi''(u)]^{-1}} \leq \sigma(\delta_1)$, where σ is defined in (119). Then, by using (131) and property (124) for Φ_*, assuming $\sigma(\delta_1) < 1$ we have

$$
[\Phi''(u)]^{-1} = \Phi_*''(\Phi'(u)) \preceq \frac{1}{(1-\sigma(\delta_1))^2} \Phi_*''\left(\frac{\tau y}{\mu}\right). \tag{79}
$$

Using (78) and (79), we can bound the local norm of the first term in the RHS of (77) as

$$
\begin{array}{rcll}
\left|\frac{\tau^+}{\mu^+} - \frac{\tau}{\mu}\right| \|y\|_{[\Phi''(u)]^{-1}} & \leq & \frac{\alpha_2}{1-\sigma(\delta_1)} \left(\left|\frac{d_\tau}{\tau}\right| + \left|\frac{1}{\mu(1-\bar{\epsilon})^2}\right|\right) \left\|\frac{\tau}{\mu} y\right\|_{\Phi_*''} & \\
& \leq & \left(\left|\frac{d_\tau}{\tau}\right| + \left|\frac{1}{\mu(1-\bar{\epsilon})^2}\right|\right) \frac{\alpha_2}{1-\sigma(\delta_1)} \sqrt{\vartheta}, & \text{using (144).}
\end{array} \tag{80}
$$

For the second term in the RHS of (77) we have

$$
\begin{array}{rcll}
\frac{\tau^+}{\mu^+} \alpha_2 \|d_y\|_{[\Phi''(u)]^{-1}} & \leq & \left[1 + \alpha_2 \left(\left|\frac{d_\tau}{\tau}\right| + \left|\frac{1}{\mu(1-\bar{\epsilon})^2}\right|\right)\right] \alpha_2 \left\|\frac{\tau}{\mu} d_y\right\|_{[\Phi''(u)]^{-1}}, & \text{using (78),} \\
& \leq & \left[1 + \alpha_2 \left(\left|\frac{d_\tau}{\tau}\right| + \left|\frac{1}{\mu(1-\bar{\epsilon})^2}\right|\right)\right] \alpha_2 \sqrt{q}, & \text{using (62).}
\end{array} \tag{81}
$$

For the third term, first by using (51) and substituting the bound in (61) we have

$$
\left\|u^+ - u\right\|_{\Phi''} \leq \underbrace{\frac{1}{1 + \alpha_2 (d_\tau/\tau)} \left(1 + \sqrt{\frac{1}{\xi - 1}}\right) \alpha_2 \sqrt{q}}_{=:\bar{\delta}}. \tag{82}
$$

If we choose α_2 such that $\bar{\delta} < 1$, then, by Lemma A.1, we have

$$\left\|\Phi'\left(u^+\right) - \Phi'\left(u\right)\right\|_{[\Phi''(u)]^{-1}} \leq \frac{\bar{\delta}}{1-\bar{\delta}}. \tag{83}$$

Putting together the above bounds, we can prove the following main result:

Proposition 4.2. *Assume that* $0.2 > \delta_2 > 4\delta_1 > 0$ *and for a point* $(x, \tau, y) \in Q_{DD}$ *we have* $\Omega_\mu(x, \tau, y) \leq \delta_1$. *Let the predictor step be calculated from* (42) *with* $r_{RHS} = r^0/\mu^2$ *and any* $\hat{H}$ *that satisfies* (43). *Then, there exists a positive constant* κ_1 *depending on* δ_1, δ_2, *and* ξ *such that we can choose* α_2 *large enough to satisfy*

$$\alpha_2 \geq \frac{\kappa_1}{\sqrt{\vartheta}}\mu, \tag{84}$$

and $\alpha_1 := \frac{\alpha_2}{\tau + \alpha_2 d_\tau}$ *for the update of* (35) *while* $\Omega_\mu(x^+, \tau^+, y^+) \leq \delta_2$.

Proof. We choose α_2 to make sure that $\bar{\delta}$ defined in (82) satisfies $\bar{\delta} \leq 1/4$. To achieve this, we first assume that $\alpha_2|d_\tau/\tau| \leq 1/2$, and then in view of (82) we choose $2(1+1/\sqrt{\xi-1})\alpha_2\sqrt{q} \leq 1/4$. If we substitute the value of $q = \frac{1}{(1-\bar{\epsilon})^6}\frac{\xi\vartheta}{\mu^2}$ defined in Lemma 4.5 and also use the bound in (60), the following inequality guarantees $\bar{\delta} \leq 1/4$:

$$\frac{\alpha_2\sqrt{\vartheta}}{\mu} \leq \underbrace{\min\left\{\sqrt{\frac{\xi-1}{\xi}}\frac{(1-\bar{\epsilon})^3}{2}, \frac{1}{8(1+1/\sqrt{\xi-1})\sqrt{\xi}}(1-\bar{\epsilon})^3\right\}}_{=:\kappa_{1,1}}. \tag{85}$$

Consider the bound we have for the proximity measure in Corollary 4.1. Assuming that $\bar{\delta}$ defined in (82) satisfies $\bar{\delta} \leq 1/4$, by using property (124), we have

$$\begin{aligned}
\left\|\tfrac{\tau^+y^+}{\mu^+} - \Phi'\left(u^+\right)\right\|^*_{\Phi''(u^+)} &\leq \tfrac{4}{3}\left\|\tfrac{\tau^+y^+}{\mu^+} - \Phi'\left(u^+\right)\right\|^*_{\Phi''(u)} \\
&\leq \tfrac{4}{3}\left\|\tfrac{\tau^+y^+}{\mu^+} - \Phi'\left(u^+\right) - \tfrac{\tau y}{\mu} + \Phi'\left(u\right)\right\|^*_{\Phi''(u)} + \tfrac{4}{3}\left\|\tfrac{\tau y}{\mu} - \Phi'\left(u\right)\right\|^*_{\Phi''(u)} \\
&\leq \tfrac{4}{3}\left\|\tfrac{\tau^+y^+}{\mu^+} - \Phi'\left(u^+\right) - \tfrac{\tau y}{\mu} + \Phi'\left(u\right)\right\|^*_{\Phi''(u)} + \tfrac{4}{3}\sigma(\delta_1),
\end{aligned} \tag{86}$$

where $\sigma(\cdot)$ is the inverse of $\rho(\cdot)$ defined in (119). Similarly, we define the inverse of $\rho(-\cdot)$ as $\bar{\sigma}(\cdot)$. To satisfy $\Omega_\mu(x^+, \tau^+, y^+) \leq \delta_2$, in view of Corollary 4.1 and using (86), a sufficient condition is

$$\left\|\frac{\tau^+y^+}{\mu^+} - \Phi'\left(u^+\right) - \frac{\tau y}{\mu} + \Phi'\left(u\right)\right\|_{[\Phi''(u)]^{-1}} \leq \frac{3}{4}\bar{\sigma}(\delta_2) - \sigma(\delta_1). \tag{87}$$

For this analysis, we need to choose δ_1 and δ_2 such that $\frac{3}{4}\bar{\sigma}(\delta_2) > \sigma(\delta_1)$. One way to force this is choosing $0.2 > \delta_2 > 4\delta_1$. To verify this, in view of inequality (120), we can check that $\delta_2 \geq \rho(-\frac{4}{3}(\sqrt{\delta_2/2} + \delta_2/4))$ for $\delta_2 \in (0, 0.2)$ and then apply $\bar{\sigma}$ to both sides. We have split the term inside the norm in the LHS of (87) into three terms in (77) and bounded the local norm for each of them. We add the bounds in (80), (81), and (83). Then, by substituting $q = \frac{1}{(1-\bar{\epsilon})^6}\frac{\xi\vartheta}{\mu^2}$

and the bound in (60), and considering $\bar{\delta} \leq 1/4$ and $\alpha_2|d_\tau/\tau| \leq 1/2$, we can bound the LHS of (87) from above by

$$\underbrace{\left[\frac{\left(\sqrt{\frac{\xi}{\xi-1}}\frac{1}{(1-\bar{\epsilon})^3}+\frac{1}{(1-\bar{\epsilon})^2}\right)}{1-\sigma(\delta_1)}+2\frac{\sqrt{\xi}}{(1-\bar{\epsilon})^3}+\frac{8}{3}\left(1+\sqrt{\frac{1}{\xi-1}}\right)\frac{\sqrt{\xi}}{(1-\bar{\epsilon})^3}\right]}_{=:1/\kappa_{1,2}}\frac{\alpha_2\sqrt{\vartheta}}{\mu}. \tag{88}$$

Note that for (81), the term inside the bracket is bounded from above by 2 using the fact that we force (80) to be smaller than 1. Therefore, if we choose

$$\frac{\alpha_2\sqrt{\vartheta}}{\mu}=\kappa_1:=\min\left\{\kappa_{1,1},\kappa_{1,2}\left(\frac{3}{4}\bar{\sigma}(\delta_2)-\sigma(\delta_1)\right)\right\},$$

then $\Omega_\mu(x^+,\tau^+,y^+)\leq\delta_2$ holds, which concludes the proof. □

To complete the whole discussion, we need to prove Lemma 3.1. Let us start with the following lemma:

Lemma 4.6. *For every set of points (z,τ,y,y_τ,μ) such that $u:=\frac{z}{\tau}\in D$, $y\in D_*$, $\mu>0$, and $y_\tau+\frac{1}{\tau}\langle y,z\rangle+\frac{\mu\xi\vartheta}{\tau}=0$, we have*

$$\left\|\frac{\tau y}{\mu}-\Phi'(u)\right\|_{[\Phi''(u)]^{-1}}\leq\beta\leq\sqrt{\frac{\xi}{\xi-1}}\left\|\frac{\tau y}{\mu}-\Phi'(u)\right\|_{[\Phi''(u)]^{-1}}, \tag{89}$$

where

$$\beta(z,\tau,y,y_\tau,\mu):=\left\|\frac{1}{\mu}\begin{bmatrix}y\\ y_\tau\end{bmatrix}-\begin{bmatrix}\frac{1}{\tau}\Phi'(u)\\ -\frac{1}{\tau}\langle\Phi'(u),u\rangle-\frac{\xi\vartheta}{\tau}\end{bmatrix}\right\|_{[\bar{H}(u,\tau)]^{-1}}, \tag{90}$$

for $\bar{H}(u,\tau)$ defined in (36) as a function of u and τ.

Proof. Consider the definition of $\bar{H}$ in (36) and the formula for its inverse in (39). We want to substitute $w:=\frac{y}{\mu}-\frac{1}{\tau}\Phi'(u)$ and $w_\tau:=\frac{y_\tau}{\mu}+\frac{1}{\tau}\langle\Phi'(u),u\rangle+\frac{\xi\vartheta}{\tau}$ in (39). Note that by using the hypothesis of the lemma, we have

$$w_\tau=\frac{y_\tau}{\mu}+\frac{1}{\tau}\langle\Phi'(u),u\rangle+\frac{\xi\vartheta}{\tau}=-\langle\frac{y}{\mu}-\frac{1}{\tau}\Phi',\frac{z}{\tau}\rangle.$$

Hence, by substituting this formula for w_τ and also w in (39), we get

$$\begin{aligned}\beta^2 &= \left\|\frac{\tau y}{\mu}-\Phi'(u)\right\|^2_{[\Phi''(u)]^{-1}}+\frac{\left[\langle\frac{\tau y}{\mu}-\Phi',[\Phi'']^{-1}\Phi'\rangle\right]^2}{\xi\vartheta-\langle\Phi',[\Phi'']^{-1}\Phi'\rangle}\\ &\leq \left\|\frac{\tau y}{\mu}-\Phi'(u)\right\|^2_{[\Phi''(u)]^{-1}}+\frac{\left\|\frac{\tau y}{\mu}-\Phi'(u)\right\|^2_{[\Phi''(u)]^{-1}}\vartheta}{(\xi-1)\vartheta}=\frac{\xi}{\xi-1}\left\|\frac{\tau y}{\mu}-\Phi'(u)\right\|^2_{[\Phi''(u)]^{-1}},\end{aligned} \tag{91}$$

where for the inequality we used CS inequality and property (134) of ϑ-s.c. barriers. (91) immediately gives us (89). □

Proof of Lemma 3.1. Assume that $\Omega_\mu(x,\tau,y) \le \epsilon < 1$, by Corollary 4.1, we have

$$\rho\left(\left\|\frac{\tau y}{\mu} - \Phi'(u)\right\|_{[\Phi''(u)]^{-1}}\right) \le \epsilon \;\Rightarrow\; \left\|\frac{\tau y}{\mu} - \Phi'(u)\right\|_{[\Phi''(u)]^{-1}} \le \sigma(\epsilon),$$

where $\sigma(\cdot)$, defined in (119), is the inverse of $\rho(\cdot)$ for nonnegative values. If we define $(z,\tau) := (\tau Ax + z^0, \tau)$ and $\frac{1}{\mu}(y, y_\tau) := \frac{1}{\mu}(y, y_{\tau,0} + \tau\langle c, x\rangle)$, the hypotheses of Lemma 4.6 are satisfied. Then, we have $\beta \le \sqrt{\frac{\xi}{\xi-1}}\sigma(\epsilon)$. In Remark 3.2, we mentioned that $\bar{H}(x,\tau)$, with some change of variables, is the Hessian of $f := \Phi\left(\frac{z}{\tau}\right) - \xi\vartheta\ln(\tau)$, which we proved in Lemma A.2 that is a $\bar{\xi}$-s.c. function for an absolute constant $\bar{\xi}$ depending on ξ. We want to use Lemma 4.1 for f and its conjugate at the points (z,τ) and $\frac{1}{\mu}(y, y_\tau)$, and the corresponding points with the same μ on the central path. One can verify that condition of Lemma 4.1-(b) holds for these points, i.e.,

$$\begin{aligned} & \langle y - y(\mu), z - z(\mu)\rangle + (y_\tau - y_\tau(\mu))(\tau - \tau(\mu)) \\ = \;& \langle A^\top(y - y(\mu)), \tau x - \tau(\mu)x(\mu)\rangle + (\tau - \tau(\mu))\langle c, \tau x - \tau(\mu)x(\mu)\rangle \\ = \;& -(\tau - \tau(\mu))\langle c, \tau x - \tau(\mu)x(\mu)\rangle + (\tau - \tau(\mu))\langle c, \tau x - \tau(\mu)x(\mu)\rangle = 0. \end{aligned} \tag{92}$$

Note that the terms in the middle of both parts (a) and (b) of Lemma 4.1 are the same. If we use the upper bound from (56) and the lower bound from (57) and ignore one term in the LHS, we get

$$\begin{aligned} & \rho\left(\frac{1}{\sqrt{\bar{\xi}}}\left(\bar{H}(x(\mu),\tau(\mu))[z - z(\mu), \tau - \tau(\mu)]\right)^{1/2}\right) \le \rho\left(-\frac{\beta}{\sqrt{\bar{\xi}}}\right) \\ \Rightarrow \;& \left(\bar{H}(x(\mu),\tau(\mu))[z - z(\mu), \tau - \tau(\mu)]\right)^{1/2} \le \sqrt{\bar{\xi}}\sigma\left(\rho\left(-\frac{1}{\sqrt{\bar{\xi}}}\sqrt{\frac{\xi}{\xi-1}}\sigma(\epsilon)\right)\right). \end{aligned} \tag{93}$$

We have $\sigma(\epsilon) \le \sqrt{2\epsilon} + \epsilon$ by (120), and for $\epsilon \le 0.1$ we can easily verify that $\sqrt{2\epsilon} + \epsilon \le \sqrt{3\epsilon}$. Also we can verify that for $t \le 0.6$, we have $\rho(-t) \le t^2$. Assume that $\sigma(\epsilon)$ is small enough to have $\sqrt{\frac{\xi}{\xi(\xi-1)}}\sigma(\epsilon) \le 0.6$. Then, the RHS of (93) becomes

$$\le \sqrt{\bar{\xi}}\sigma\left(\frac{\xi}{\bar{\xi}(\xi-1)}\sigma^2(\epsilon)\right) \le 3\sqrt{\frac{\xi}{\xi-1}}\sqrt{\epsilon}. \tag{94}$$

Now we just need to use property (124) of s.c. functions for $f = \Phi\left(\frac{z}{\tau}\right) - \xi\vartheta\ln(\tau)$ to get the result of the lemma. □

Before analyzing the corrector step, let us elaborate more on the above proof. For a point $(x,\tau,y) \in Q_{DD}$ with parameter μ, let us define

$$d := \begin{bmatrix} \tau(\mu)x(\mu) - \tau x \\ \tau(\mu) - \tau \\ v(\mu) - v \end{bmatrix}. \tag{95}$$

We can easily verify that (using $y = y^0 - (\tau - 1)c_A - F^\top v$):

$$Ud = \begin{bmatrix} \tau(\mu)Ax(\mu) + z^0 \\ \tau(\mu) \\ y(\mu) \\ y_{\tau,0} + \tau(\mu)\langle c, x(\mu)\rangle \end{bmatrix} - \begin{bmatrix} \tau Ax + z^0 \\ \tau \\ y \\ y_{\tau,0} + \tau\langle c, x\rangle \end{bmatrix}. \tag{96}$$

We want to use property (124) for $r = 1/4$ to change the local norm in (93); it suffices to force $3\sqrt{\frac{\xi}{\bar{\xi}(\xi-1)}}\sqrt{\epsilon} \le \frac{1}{4}$ in view of (94). Consider the proof of Lemma 3.1 and also the term for y that we ignored in (93). Then, using (94) and the above discussion, we have

Corollary 4.2. *If for a point $(x, \tau, y) \in Q_{DD}$ we have $3\sqrt{\frac{\xi}{\bar{\xi}(\xi-1)}}\sqrt{\Omega_\mu(x,\tau,y)} \le \frac{1}{4}$, then for d defined in* (95) *we have*

$$\|d\|_{U^\top \mathcal{H}(\bar{H}(x,\tau),\mu^2\bar{H}(x,\tau))U} \le 2 \cdot \frac{4}{3}\left(3\sqrt{\frac{\xi}{\xi-1}}\sqrt{\Omega_\mu(x,\tau,y)}\right) = \underbrace{8\sqrt{\frac{\xi}{\xi-1}}}_{=:\bar{\xi}_1}\sqrt{\Omega_\mu(x,\tau,y)}. \tag{97}$$

This inequality gives us (59) for $q = \bar{\xi}_1^2\Omega_\mu(x,\tau,y)$ that we break down to get the bounds we need for the analysis of the corrector step.

4.2. **Corrector step.** We focus on the case that $\alpha_1 = \frac{\alpha_2}{\tau+\alpha_2 d_\tau}$ (assuming $\tau + \alpha_2 d_\tau > 0$) in the updates of (35). By Remark 3.4, $\mu^+ = \mu$ for every α_2 and so we just need to show that α_2 can be chosen to get enough reduction in the proximity measure. Let d^c be the corrector step derived by solving (42) with parameters defined in (48). We argued by using (50) that the value of $\langle d^c, U^\top\psi^c\rangle$ represents the first order reduction in Ω_μ. On the other hand, by using (42) and (48), we can verify

$$-\langle d^c, U^\top\psi^c\rangle = \left\|U^\top\psi^c + \beta r^0\right\|^2_{(U^\top\mathcal{H}U)^{-1}}. \tag{98}$$

The following key lemma shows that this quantity has a large enough value:

Lemma 4.7. *Let $(x,\tau,y) \in Q_{DD}$. If*

$$\Omega_\mu(x,\tau,y) \le \frac{1}{100\left((\bar{\xi}_2\bar{\xi}_1)^3 + \bar{\xi}_3\bar{\xi}_1^3\right)^2}, \tag{99}$$

where $\bar{\xi}_1 = 8\sqrt{\xi}/\sqrt{\xi-1}$ is defined in (97) *and*

$$\begin{aligned} \bar{\xi}_2 &:= 3\sqrt{\tfrac{1}{\xi-1}} + \tfrac{7}{2}, \\ \bar{\xi}_3 &:= \tfrac{1}{2\sqrt{\xi-1}}\left(\tfrac{11}{2} + \tfrac{5}{\sqrt{\xi-1}}\right)\left(3 + \tfrac{2}{\sqrt{\xi-1}}\right) + \tfrac{2}{\xi-1}\left(1 + \tfrac{1}{\sqrt{\xi-1}}\right), \end{aligned} \tag{100}$$

then,

$$\left\|U^\top \psi^c + \beta r^0\right\|_{(U^\top \mathcal{H} U)^{-1}} \geq \frac{1}{4\bar{\xi}_1}\sqrt{\Omega_\mu(x,\tau,y)},$$

where β is defined in (48) *and ψ^c is defined in* (49)*.*

Proof. $-\|U^\top\psi^c+\beta r^0\|_{(U^\top\mathcal{H}U)^{-1}}$ is the optimal objective value of (55) and we find an upper bound for it by using a specific feasible solution. Our feasible solution is

$$\frac{d}{\|d\|_{U^\top \mathcal{H}(\bar{H},\mu^2\bar{H})U}}, \tag{101}$$

where d is defined in (95) and we have Corollary 4.2 for a bound on its local norm. We can verify that (101) satisfies all the constraints. Now, we need to prove that $-\langle d, U^\top\psi^c\rangle$ is large enough. The idea of the proof is that we consider the bounds in (50) at $\alpha_2 = 1$ and $\alpha_2 = 2$, and if $-\langle d, U^\top\psi^c\rangle$ is not large enough, we get a contradiction.

For simplicity, let $\Omega_\mu := \Omega_\mu(x,\tau,y)$ and define $\sqrt{q} = \bar{\xi}_1\sqrt{\Omega_\mu}$ for $\bar{\xi}_1$ defined in (97). Then (59) becomes the expansion of (97) and we have all the inequalities we extracted after Lemma 4.5, which we use to find bounds for the terms we have in (50). For the first term of $D(\alpha_2)$ we can use (61). For the second term of $D(\alpha_2)$ we use triangle inequality and we have

$$\left\|\frac{d_\tau y}{\mu}\right\|_{\Phi''_*} = \frac{d_\tau}{\tau}\left\|\frac{\tau y}{\mu}\right\|_{\Phi''_*} \leq \sqrt{\frac{q}{\xi-1}}, \quad \text{using (60) and (144)}, \tag{102}$$

and using (60) and (62), we have

$$\left\|\frac{(\tau+\alpha_2 d_\tau)d_y}{\mu}\right\|_{\Phi''_*} = \frac{(\tau+\alpha_2 d_\tau)}{\tau}\left\|\frac{\tau d_y}{\mu}\right\|_{\Phi''_*} \leq \left(1+\alpha_2\sqrt{\frac{q}{(\xi-1)\vartheta}}\right)\sqrt{q}. \tag{103}$$

If we use the CS inequality (9) for $B = \Phi''$ and use $\|\Phi'\|_{[\Phi'']^{-1}} \leq \sqrt{\vartheta}$ (see (134)), then (60) and (61) imply

$$\left|\frac{d_\tau}{\tau(\tau+\alpha_2 d_\tau)}\langle \Phi', A\bar{d}_x - d_\tau\left(Ax+\frac{1}{\tau}z^0\right)\rangle\right| \leq \frac{\sqrt{\frac{q}{\xi-1}}}{1-\alpha_2\sqrt{\frac{q}{(\xi-1)\vartheta}}}\left(1+\sqrt{\frac{1}{\xi-1}}\right)\sqrt{q}. \tag{104}$$

We want to make the second line of the term in the middle of inequalities in (50) a quadratic in terms of α_2, while the upper and lower bounds are proportional to α_2^3. To do this, we modify

(50) by adding and subtracting some terms to all sides as:

$$
\begin{aligned}
& \rho(D(\alpha_2)) - \tfrac{1}{2}(D(\alpha_2))^2 + \hat{D}(\alpha_2) \\
\leq\ & \Omega_\mu(x^+,\tau^+,y^+) - \Omega_\mu(x,\tau,y) - \alpha_2 \begin{bmatrix} \bar{d}_x^\top & d_\tau & d_v^\top \end{bmatrix} U^\top \psi^c \\
& + \tfrac{\alpha_2^2 d_\tau}{\tau^2}\langle \Phi', A\bar{d}_x - d_\tau\left(Ax + \tfrac{1}{\tau}z^0\right)\rangle - \tfrac{\alpha_2^2 d_\tau}{\mu}\left(\langle d_y, \Phi'_*\rangle + \langle c, \bar{d}_x\rangle\right) - \tfrac{1}{2}(\bar{D}(\alpha_2))^2 \\
\leq\ & \rho\left(-D(\alpha_2)\right) - \tfrac{1}{2}(D(\alpha_2))^2 + \hat{D}(\alpha_2), \\
\bar{D}(\alpha_2) \ :=\ & \tfrac{\alpha_2}{\tau}\left\|A\bar{d}_x - d_\tau\left(Ax + \tfrac{1}{\tau}z^0\right)\right\|_{\Phi''} + \alpha_2\left\|\tfrac{d_\tau y + \tau d_y}{\mu}\right\|_{\Phi''_*}, \\
\hat{D}(\alpha_2) \ :=\ & \tfrac{1}{2}\left((D(\alpha_2))^2 - (\bar{D}(\alpha_2))^2\right) + \tfrac{\alpha_3^2 d_\tau^2}{\tau^2(\tau+\alpha_2 d_\tau)}\langle \Phi', A\bar{d}_x - d_\tau\left(Ax + \tfrac{1}{\tau}z^0\right)\rangle.
\end{aligned} \tag{105}
$$

Note that by definition (118), we can verify that

$$
\rho(-t) - \frac{t^2}{2} \leq t^3, \qquad \frac{t^2}{2} - \rho(t) \leq t^3, \quad \forall t \in (0, 0.8). \tag{106}
$$

Let us assume that $2\sqrt{\frac{q}{(\xi-1)\vartheta}} \leq \frac{1}{2}$, then (61), (102), (103), and (104) yield that for $\alpha_2 \in (0,2)$ we have

$$
\begin{aligned}
|D(\alpha_2)| &\leq \alpha_2 \bar{\xi}_2 \sqrt{q} = \alpha_2 \bar{\xi}_2 \bar{\xi}_1 \sqrt{\Omega_\mu}, \\
|\bar{D}(\alpha_2)| &\leq \alpha_2 \left(2\sqrt{\frac{1}{\xi-1}} + 2\right)\bar{\xi}_1\sqrt{\Omega_\mu}, \\
|\hat{D}(\alpha_2)| &\leq \alpha_2^3 \bar{\xi}_3 \bar{\xi}_1^3 \Omega_\mu^{3/2},
\end{aligned} \tag{107}
$$

where $\bar{\xi}_2$ nd $\bar{\xi}_3$ are defined in (100). For the bound on $|\hat{D}(\alpha_2)|$, we also used the fact that

$$
|D(\alpha_2) - \bar{D}(\alpha_2)| \leq \frac{\alpha_2^2 |d_\tau|}{(\tau + \alpha_2 d_\tau)\tau}\left\|A\bar{d}_x - d_\tau\left(Ax + \frac{1}{\tau}z^0\right)\right\|_{\Phi''} + \alpha_2^2\left\|\frac{d_\tau d_y}{\mu}\right\|_{\Phi''_*}.
$$

If we have $\bar{\xi}_2\bar{\xi}_1\sqrt{\Omega_\mu} \leq 0.8$, by using (106) and (107), the middle term of (105) is squeezed between $\pm\alpha_2^3\left((\bar{\xi}_2\bar{\xi}_1)^3 + \bar{\xi}_3\bar{\xi}_1^3\right)\Omega_\mu^{3/2}$ for $\alpha_2 \in (0,2)$. We want to choose Ω_μ small enough to make the term in the middle of (105) be squeezed between $\pm\frac{1}{10}\Omega_\mu$ for $\alpha_2 = 1$; it suffices to have

$$
\left((\bar{\xi}_2\bar{\xi}_1)^3 + \bar{\xi}_3\bar{\xi}_1^3\right)\Omega_\mu^{3/2} \leq \frac{1}{10}\Omega_\mu \quad \underbrace{\Leftrightarrow}_{\text{for } \Omega_\mu > 0} \quad \Omega_\mu \leq \frac{1}{100\left((\bar{\xi}_2\bar{\xi}_1)^3 + \bar{\xi}_3\bar{\xi}_1^3\right)^2}. \tag{108}
$$

We claim that in this case, $-\langle d, U^\top\psi^c\rangle \geq \frac{1}{4}\Omega_\mu$. If we substitute $\alpha_2 = 1$, then $\Omega_\mu(x^+,\tau^+,y^+) = 0$ as we can verify that the point lays on the central path. Suppose for the sake of reaching a contradiction $-\langle d, U^\top\psi^c\rangle < \frac{1}{4}\Omega_\mu$. Then, in view of (105), we must have

$$
\frac{d_\tau}{\tau^2}\langle \Phi', A\bar{d}_x - d_\tau\left(Ax + \frac{1}{\tau}z^0\right)\rangle - \frac{d_\tau}{\mu}\left(\langle \Phi'_*, d_y\rangle + \langle c, \bar{d}_x\rangle\right) - \frac{1}{2}(\bar{D}(1))^2 \geq \left(\frac{3}{4} - \frac{1}{10}\right)\Omega_\mu.
$$

We reach our contradiction when we consider $\alpha_2 = 2$. For $\alpha_2 = 2$ we have $\Omega_\mu(x^+,\tau^+,y^+) \geq 0$. The term in the second line of (50) is degree 2 of α_2 and so becomes at least $\left(\frac{12}{4} - \frac{4}{10}\right)\Omega_\mu$ for

$\alpha_2 = 2$. Then, at $\alpha_2 = 2$, (105) implies

$$-\Omega_\mu(x,\tau,y) + \left(\frac{12}{4} - \frac{4}{10}\right)\Omega_\mu(x,\tau,y) \le \frac{8}{10}\Omega_\mu(x,\tau,y),$$

which is a contradiction.

Now, if we consider the feasible solution (101) for the optimization problem (55) and putting together the bounds $-\langle d, U^\top \psi^c\rangle \ge \frac{1}{4}\Omega_\mu$ and $\|d\|_{U^\top \mathcal{H}(\bar{H},\mu^2\bar{H})U} \le \bar{\xi}_1\sqrt{\Omega_\mu}$ from (97), we get the result of the lemma. We can verify that for $\xi > 1$, (99) implies the other bounds we used for Ω_μ in the proof, including the hypothesis of Corollary 4.2, $\bar{\xi}_2\bar{\xi}_1\sqrt{\Omega_\mu} \le 0.8$, and $2\sqrt{\frac{q}{(\xi-1)\vartheta}} \le \frac{1}{2}$. □

Now we are ready to prove the main proposition for the corrector step.

Proposition 4.3. *Let $(x,\tau,y) \in Q_{DD}$ satisfy (99). Assume that the corrector step d^c is calculated by solving (42) with parameters defined in (48), and we choose $\alpha_1 = \frac{\alpha_2}{\tau+\alpha_2 d_\tau}$ for the updates of (35). Consider $\bar{\xi}_1$ and $\bar{\xi}_2$ defined in (97) and (100), respectively. Then, for*

$$\alpha_2 := \frac{1}{2(\bar{\xi}_4 + \bar{\xi}_2^2)}, \quad \bar{\xi}_4 := 2\sqrt{\frac{1}{\xi-1}}\left(1+\sqrt{\frac{1}{\xi-1}}\right) + \frac{\sqrt{\xi}+2}{\sqrt{\xi-1}}, \tag{109}$$

we have

$$\Omega_\mu(x^+,\tau^+,y^+) - \Omega_\mu(x,\tau,y) \le -\frac{\alpha_2}{32\bar{\xi}_1^2}. \tag{110}$$

Proof. Assume that $d^c = [\bar{d}_x^\top \ d_\tau \ d_v^\top]$ is the corrector search direction. Then, by (42) and (48) we have

$$(d^c)^\top U^\top \mathcal{H} U d^c = \|U^\top \psi^c + \beta r^0\|^2_{(U^\top \mathcal{H} U)^{-1}}. \tag{111}$$

Hence, we have inequality (59) with $q := \|U^\top \psi^c + \beta r^0\|^2_{(U^\top \mathcal{H} U)^{-1}}$, and we already have the bounds (61), (102), (103), and (104). Here, we use (59) to get another bound; if we consider the last term in the LHS of (59), we get

$$\frac{\tau}{\mu}\left|\langle d_y, Ax + \frac{1}{\tau}z^0\rangle + \langle c, \bar{d}_x\rangle\right| \le (\sqrt{\xi}+1)\sqrt{\vartheta q}. \tag{112}$$

Note that from Corollary 4.1, we have $\left\|\frac{\tau y}{\mu}-\Phi'\right\|_{[\Phi'']^{-1}}\le\sigma(\Omega_\mu)$. Using this and (112), we have

$$\begin{aligned}
&\left|\tfrac{d_\tau}{\mu}\left(\langle d_y,\Phi'_*\rangle+\langle c,\bar{d}_x\rangle\right)\right|\\
=&\left|\tfrac{d_\tau}{\mu}\left(\langle d_y,Ax+\tfrac{1}{\tau}z^0\rangle+\langle c,\bar{d}_x\rangle+\langle d_y,\Phi'_*-Ax+\tfrac{1}{\tau}z^0\rangle\right)\right|\\
\le&\left|\tfrac{d_\tau}{\tau}\right|\left(\tfrac{\tau}{\mu}\left|\langle d_y,Ax+\tfrac{1}{\tau}z^0\rangle+\langle c,\bar{d}_x\rangle\right|+\tfrac{\tau}{\mu}\|d_y\|_{[\Phi'']^{-1}}\left\|\Phi'_*-Ax+\tfrac{1}{\tau}z^0\right\|_{\Phi''}\right)\\
(113)\le&\sqrt{\tfrac{q}{(\xi-1)\vartheta}}\left((\sqrt{\xi}+1)\sqrt{\vartheta q}+\tfrac{\tau}{\mu}\|d_y\|_{[\Phi'']^{-1}}\left\|\Phi'_*-Ax+\tfrac{1}{\tau}z^0\right\|_{\Phi''}\right),\quad\text{by (60) and (112),}\\
\le&\sqrt{\tfrac{q}{(\xi-1)\vartheta}}\left((\sqrt{\xi}+1)\sqrt{\vartheta q}+\sqrt{q}\tfrac{\sigma(\Omega_\mu)}{1-\sigma(\Omega_\mu)}\right),\qquad\text{by (62) and Lemma A.1,}\\
=&\sqrt{\tfrac{1}{\xi-1}}\left((\sqrt{\xi}+1)+\tfrac{\sigma(\Omega_\mu)}{\sqrt{\vartheta}(1-\sigma(\Omega_\mu))}\right)q\\
\le&\tfrac{\sqrt{\xi}+2}{\sqrt{\xi}-1}q,\qquad\text{for the case }\sigma(\Omega_\mu)\le 0.5.
\end{aligned}$$

We want to work with the second inequality in (50). We already have a bound for $D(\alpha_2)$ in (107) and we also have $\rho(-t)\le t^2$ for $t\in(0,0.6)$. By substituting (104) and (113), we get

$$\Omega_\mu(x^+,\tau^+,y^+)-\Omega_\mu(x,\tau,y)\le(-\alpha_2+(\bar{\xi}_4+\bar{\xi}_2^2)\alpha_2^2)\|U^\top\psi^c+\beta r^0\|^2_{(U^\top\mathcal{H}U)^{-1}},\tag{114}$$

where $\bar{\xi}_4$ is defined in (109). If we choose $\alpha_2\le\frac{1}{2(\bar{\xi}_4+\bar{\xi}_2^2)}$, then for the RHS we have

$$\le-\frac{1}{2}\alpha_2\|U^\top\psi^c+\beta r^0\|^2_{(U^\top\mathcal{H}U)^{-1}}\le-\frac{\alpha_2}{32\bar{\xi}_1^2}\Omega_\mu,\tag{115}$$

where we used the bound for $\|U^\top\psi^c+\beta r^0\|^2_{(U^\top\mathcal{H}U)^{-1}}$ by Lemma 4.7. □

4.3. **Complexity of following the path to $\mu=+\infty$.** We have analyzed the predictor and corrector search directions in Section 4. Now we can modify the statement of our predictor-corrector algorithm to one that provably follows the path in polynomial time.

Polynomial-time Predictor-Corrector Algorithm (PtPCA)

Initialization: Choose $z^0\in\text{int}D$ and set $y^0:=\Phi'(z^0)$. Set $x^0:=0$, $\tau_0:=1$, $\mu_0:=\mu(x^0,\tau_0,y^0)$, and $k=0$. Choose a constant $\xi>1$ and constants $0<4\delta_1<\delta_2\le\frac{1}{100\left((\bar{\xi}_2\bar{\xi}_1)^3+\bar{\xi}_3\bar{\xi}_1^3\right)^2}$, where $\bar{\xi}_1$, $\bar{\xi}_2$, and $\bar{\xi}_3$ are functions of ξ defined in (97) and (100).

while (the stopping criteria are not met)

  if $(\Omega_{\mu_k}(x^k,\tau_k,y^k)>\delta_1)$

    Calculate the corrector search direction (d_x,d_τ,d_y) by (42) with r_{RHS} and $\hat{H}$ defined in (48), and choose α_2 as in (109) and $\alpha_1:=\frac{\alpha_2}{\tau+\alpha_2 d_\tau}$. Apply the update in (35) to get $(x^{k+1},\tau_{k+1},y^{k+1})$, and define $\mu_{k+1}:=\mu_k$.

  if $(\Omega_{\mu_k}(x^k,\tau_k,y^k)\le\delta_1)$

Calculate the predictor search direction (d_x, d_τ, d_y) by (42) with $r_{RHS} = r^0/\mu_k^2$ and any $\hat{H}$ that satisfies (43), and choose $\alpha_2 = \frac{\kappa_1}{\sqrt{\vartheta}}\mu_k$ for κ_1 defined in the proof of Proposition 4.2, and $\alpha_1 := \frac{\alpha_2}{\tau+\alpha_2 d_\tau}$. Apply the update in (35) to get $(x^{k+1}, \tau_{k+1}, y^{k+1})$, and define $\mu_{k+1} := \mu(x^{k+1}, \tau_{k+1}, y^{k+1})$.

$k \leftarrow k+1$.

end while

Note that even though the choices of δ_1 and δ_2 in the PtPCA, as we show in the following, gives us the desired iteration complexity bounds, these choices are too small for practical purposes. In practice, as we have done in the DDS code, δ_1 and δ_2 are chosen large enough to guarantee long steps. To achieve long steps in practice, we should not restrict the algorithm to Dikin ellipsoids. There are properties for classes of s.c. barriers that strengthen the Dikin ellipsoid property to anywhere in the interior of the domain. We mention *negative curvature* [39, 19, 42] and *α-regularity* [38] here. Negative curvature is a property for many interesting LH s.c. barriers (see [39], [19], and [42]-Section 9.2) that lets us extend a Hessian estimation property like (124) to effectively the whole domain of the s.c. barrier. A s.c. function is additionally α-regular if the second derivative also controls the fourth derivative in a proper way [38]. It was shown in [38] that many useful s.c. barriers are α-regular, such as the ones in Table 1 for LP, SOCP, and SDP, and the ones we built for Geometric Programing and Entropy Programming. If all the s.c. barriers given in a problem instance have one of these properties, the practical version of our algorithm is theoretically guaranteed to take long steps (a large portion of the distance between the current iterate and the boundary). If even one of these barriers does not have any long-step property, this theoretical guarantee may not hold. It is possible to construct some pathological examples on which the algorithm has to take a short step in every iteration; however, the practical version of the algorithm generally has a chance to take long steps in most of the iterations.

Our analysis of the predictor and corrector steps implies the following theorem:

Theorem 4.1. *For the polynomial-time predictor-corrector algorithm, there exists a positive constant κ_2 depending on ξ such that after N iterations, we get a point $(x, \tau, y) \in Q_{DD}$ such that*

$$\mu(x, \tau, y) \geq \exp\left(\frac{\kappa_2}{\sqrt{\vartheta}} N\right). \tag{116}$$

Proof. By Proposition 4.3, after each predictor step, we have to do at most

$$64(\bar{\xi}_4 + \bar{\xi}_2^2)\bar{\xi}_1^2(\delta_2 - \delta_1),$$

number of corrector steps to satisfy $\Omega_\mu(x,\tau,y) \le \delta_1$. Also, by Propositions 4.1 and 4.2, after $\bar{N}$ cycles of predictor-corrector steps, we have

$$\mu \ge \left(1 + \frac{\kappa_1}{\sqrt{\vartheta}}\right)^{\bar{N}}.$$

Therefore, we have (116) for $\kappa_2 = O(1)\kappa_1$. □

Theorem 4.1 is the core of several consequences about determining the statuses of the problem in polynomial time (see [24]-Chapter 7). In this article, we briefly discuss the case where the problem and its dual both are strictly feasible. In this case, we can define a *feasibility measure* σ_f (which is a complexity measure) that represents how good the geometry of the feasible regions are and the proximity of z^0 and y^0 to the boundaries of their respective domains (see [24]-Chapter 7 for the exact definition). Then we can prove the following theorem about the connection between τ and μ.

Theorem 4.2 ([24])**.** *Assume that both primal and dual are strictly feasible and for a point $(x,\tau,y) \in Q_{DD}$ we have the additional property that $\delta_*(y|D) + y_{\tau,0} + \tau\langle c,x\rangle \le 0$. Then,*

$$\tau - 1 \ge \sigma_f \mu - \frac{1}{\sigma_f}, \tag{117}$$

where σ_f is the feasibility measure.

Note that by Lemma 2.2, the hypothesis of the above theorem holds for the points close to the central path. Putting together the discussion we had in Subsection 2.2 and Theorem 4.1, we conclude that when we have strict primal and dual feasibility, in $O\left(\sqrt{\vartheta}\ln\left(\frac{\vartheta}{\epsilon}\right)\right)$ number of iterations, we obtain an ϵ-solution of the problem.

5. Conclusions

After introducing the Domain-Driven setup, we defined an infeasible-start primal-dual central path and designed and analyzed algorithms that can follow this path efficiently (Theorem 4.1). Following our discussion in Subsection 2.2, the important question is: for different statuses of the problem, what is the behavior of (x,τ,y) when $\mu \to +\infty$, and for which values of μ we can determine the status of the problem with ϵ accuracy using (x,τ,y)? We answered this question for the case of strict primal and dual feasibility, for which our algorithm can return an ϵ-solution in $O\left(\sqrt{\vartheta}\ln\left(\frac{\vartheta}{\epsilon}\right)\right)$ number of iterations. This bound is the current best and is new for the type of formulations we used for handling infeasibility, even in the special case of SDP.

Different statuses are defined in [24] for a problem in the Domain-Driven setup. Theorem 4.1 and the preliminary discussion in [24]-Chapter 7 show that our algorithms return certificates

(heavily relying on duality) for each of these statuses in polynomial time. The iteration complexity bounds are comparable to the current best ones we have for the conic formulations (to the best of our knowledge mostly in [41]). The discussion on the geometry of the problem, different statuses, and the performance of the algorithm for each of them will appear in a subsequent article. The algorithms of this article are the base of a code, called DDS (Domain-Driven Solver), that solves many classes of problems, including those listed in Section 1, and the list is expanding.

References

[1] B. Alkire and L. Vandenberghe, *Convex optimization problems involving finite autocorrelation sequences*, Mathematical Programming, 93 (2002), pp. 331–359.

[2] B. P. W. Ames and S. A. Vavasis, *Nuclear norm minimization for the planted clique and biclique problems*, Mathematical Programming, 129 (2011), pp. 69–89.

[3] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*, Princeton University Press, 2009.

[4] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, vol. 2, SIAM, 2001.

[5] ———, *Robust optimization–methodology and applications*, Mathematical Programming, 92 (2002), pp. 453–480.

[6] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM, 1994.

[7] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, *A tutorial on geometric programming*, Optimization and Engineering, 8 (2007), pp. 67–127.

[8] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[9] E. Candes and B. Recht, *Exact matrix completion via convex optimization*, Communications of the ACM, 55 (2012), pp. 111–119.

[10] V. Chandrasekaran and P. Shah, *Relative entropy relaxations for signomial optimization*, SIAM Journal on Optimization, 26 (2016), pp. 1147–1173.

[11] ———, *Relative entropy optimization and its applications*, Mathematical Programming, 161 (2017), pp. 1–32.

[12] C. Davis, *All convex invariant functions of hermitian matrices*, Archiv der Mathematik, 8 (1957), pp. 276–278.

[13] D. L. Donoho, *Compressed sensing*, IEEE Transactions on information theory, 52 (2006), pp. 1289–1306.

[14] S.-C. Fang, J. R. Rajasekera, and H.-S. J. Tsao, *Entropy optimization and Mathematical Programming*, vol. 8, Springer Science & Business Media, 1997.

[15] H. Fawzi and O. Fawzi, *Relative entropy optimization in quantum information theory via semidefinite programming approximations*, arXiv preprint arXiv:1705.06671, (2017).

[16] H. Fawzi, J. Saunderson, and P. A. Parrilo, *Semidefinite approximations of the matrix logarithm*, arXiv preprint arXiv:1705.00812, (2017).

[17] L. Faybusovich, *Primal-dual potential reduction algorithm for symmetric programming problems with nonlinear objective functions*, Linear Algebra and its Applications, 536 (2018), pp. 228–249.

[18] L. Faybusovich and T. Tsuchiya, *Matrix monotonicity and self-concordance: how to handle quantum entropy in optimization problems*, Optimization Letters, (2017), pp. 1513–1526.

[19] O. Güler, *Hyperbolic polynomials and interior point methods for convex programming*, Mathematics of Operations Research, 22 (1997), pp. 350–377.

[20] G. Haeser, O. Hinder, and Y. Ye, *On the behavior of lagrange multipliers in convex and non-convex infeasible interior point methods*, arXiv preprint arXiv:1707.07327, (2017).

[21] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*, CRC press, 2015.

[22] O. Hinder and Y. Ye, *A one-phase interior point method for nonconvex optimization*, arXiv preprint arXiv:1801.03072, (2018).

[23] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*, Springer Science & Business Media, 2001.

[24] M. Karimi, *Convex Optimization via Domain-Driven Barriers and Primal-Dual Interior-Point Methods*, PhD thesis, University of Waterloo, 2017.

[25] N. Karmarkar, *A new polynomial-time algorithm for linear programming*, Combinatorica, 4 (1984), pp. 373–395.

[26] M. Kojima, N. Megiddo, and S. Mizuno, *A primal-dual infeasible-interior-point algorithm for linear programming*, Mathematical Programming, 61 (1993), pp. 263–280.

[27] A. S. Lewis, *The mathematics of eigenvalue optimization*, Mathematical Programming, 97 (2003), pp. 155–176.

[28] I. J. Lustig, *Feasibility issues in a primal-dual interior-point method for linear programming*, Mathematical Programming, 49 (1990), pp. 145–162.

[29] I. J. Lustig, R. E. Marsten, and D. F. Shanno, *Computational experience with a primal-dual interior point method for linear programming*, Linear Algebra and its Applications, 152 (1991), pp. 191–222.

[30] S. Mizuno, M. J. Todd, and Y. Ye, *On adaptive-step primal-dual interior-point algorithms for linear programming*, Mathematics of Operations Research, 18 (1993), pp. 964–981.

[31] T. Myklebust and L. Tunçel, *Interior-point algorithms for convex optimization based on primal-dual metrics*, arXiv preprint arXiv:1411.2129, (2014).

[32] A. Nemirovski, *Interior point polynomial time methods in convex programming*, Lecture Notes, (2004).

[33] A. Nemirovski and L. Tunçel, *Cone-free primal-dual path-following and potential reduction polynomial time interior-point methods*, Mathematical Programming, 102 (2005), pp. 261–294.

[34] Y. Nesterov, *Infeasible-start interior-point primal-dual methods in nonlinear programming*, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE), (1995).

[35] Y. Nesterov, *Towards non-symmetric conic optimization*, Optimization Methods and Software, 27 (2012), pp. 893–917.

[36] Y. Nesterov and A. Nemirovski, *Conic duality and its applications in convex programming*, Optimization Methods and Software, 1 (1992), pp. 95–115.

[37] Y. Nesterov and A. Nemirovski, *Interior point polynomial methods in Convex Programming*, SIAM Series in Applied Mathematics, SIAM: Philadelphia, 1994.

[38] ———, *Multi-parameter surfaces of analytic centers and long-step surface-following interior-point methods*, Mathematics of Operations Research, 23 (1998), pp. 1–38.

[39] 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.

[40] ———, *Primal-dual interior-point methods for self-scaled cones*, SIAM Journal on Optimization, 8 (1998), pp. 324–364.

[41] 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.

[42] Y. Nesterov and L. Tunçel, *Local superlinear convergence of polynomial-time interior-point methods for hyperbolicity cone optimization problems*, SIAM Journal on Optimization, 26 (2016), pp. 139–170.

[43] B. Recht, M. Fazel, and P. A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Review, 52 (2010), pp. 471–501.

[44] A. Skajaa and Y. Ye, *A homogeneous interior-point algorithm for nonsymmetric convex conic optimization*, Mathematical Programming, 150 (2015), pp. 391–422.

[45] 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.

[46] ———, *Polyhedral and Semidefinite Programming Methods in Combinatorial Optimization*, American Mathematical Soc., 2010.

[47] Y. Ye, M. J. Todd, and S. Mizuno, *An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm*, Mathematics of Operations Research, 19 (1994), pp. 53–67.

[48] Y. Zhang, *On the convergence of a class of infeasible interior-point methods for the horizontal linear complementarity problem*, SIAM Journal on Optimization, 4 (1994), pp. 208–227.

[49] Y. Zhang, *On extending some primal–dual interior-point algorithms from linear programming to semidefinite programming*, SIAM Journal on Optimization, 8 (1998), pp. 365–386.

Appendix A. Self-Concordant Functions

In this section, we summarize the properties of self-concordant (s.c.) functions that we use in this paper. The properties are divided into three parts; 1) general properties of s.c. functions, 2) s.c. barriers, and 3) Legendre-Fenchel (LF) conjugate of s.c. barriers.

We define the following function which is frequently used in the context of self-concordant functions.

$$\rho(t) := \begin{cases} t - \ln(1+t) = \frac{t^2}{2} - \frac{t^3}{3} + \frac{t^4}{4} + \cdots, & t \geq -1, \\ +\infty, & t < -1. \end{cases} \tag{118}$$

We also need, in some sense, the inverse of this function

$$\sigma(s) := \max\{t : \rho(t) \leq s\}, \ s \geq 0. \tag{119}$$

By [33]-Lemma 2.1, we have

$$\sigma(s) \leq \sqrt{2s} + s, \quad \forall s \geq 0. \tag{120}$$

A.1. **Self-concordant (s.c.) functions.** A convex function $f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ is called a-s.c. function if its domain Q is open, f is $\mathcal{C}^3$ on Q and

(i) $f(x_i) \to +\infty$ for every sequence $\{x_i\} \subset Q$ that converges to a point on the boundary of Q.
(ii) There exists a positive real constant a such that

$$|f'''(x)[h,h,h]| \leq 2a^{-1/2}(f''(x)[h,h])^{3/2} = 2a^{-1/2}\|h\|^3_{f''(x)}, \quad \forall (x \in Q, h \in \mathbb{E}), \tag{121}$$

where $f^k(x)[h_1, \ldots, h_k]$ henceforth is the value of the kth differential of f along directions $h_1, \ldots, h_k \in \mathbb{E}$.

We say that f is *non-degenerate* if its Hessian $f''(x)$ is positive definite at some point (and then it can be proved to be positive definite at all points) in Q.

For a a-s.c. function f and any point x in its domain, we define an important concept of the *Newton decrement* of f at x as

$$\lambda(f,x) := a^{-1/2} \max\{f'(x)[h] : h \in \mathbb{E}, f''(x)[h,h] \leq 1\}. \tag{122}$$

When f is non-degenerate, it can be shown that we have

$$\lambda(f,x) = a^{-1/2}\|f'(x)\|^*_{f''(x)}. \tag{123}$$

In the following, we list some of the important properties of s.c. functions and s.c. barriers. Properties are labeled with SC for future reference:

SC-1 **(Stability under intersections, direct sums, and affine maps)** [37]-Proposition 2.1.1:

(a) Let f_i, $i \in \{1, \dots, m\}$, be an a_i-s.c. function on $\mathbb{E}$ with domains Q_i. Then, for real coefficients $\gamma_i \geq 1$, if $Q := \cap_{i=1}^m Q_i$ is not empty, $f := \sum_{i=1}^m \gamma_i f_i$ is an a-s.c. function with domain Q, where $a := \min\{\gamma_i a_i : i \in \{1, \dots, m\}\}$.
(b) Let f_i, $i \in \{1, \dots, m\}$, be an a-s.c. function on $\mathbb{E}_i$ with domains Q_i. Then, the function $f(x^1, \dots, x^m) := \sum_{i=1}^m f_i(x^i)$, defined on $Q_1 \oplus \cdots \oplus Q_m$, is an a-s.c. function.
(c) Let f be a s.c. function with domain Q and $x = Ay + b$ be an affine mapping with image intersecting Q, then $f(Ay + b)$ is also a s.c. function on $\{y : Ay + b \in Q\}$.

From now on, we assume that f is a s.c. function with domain Q.

SC-2 **(Behaviour in Dikin ellipsoid and some basic inequalities):**

(a) For every point $x \in Q$, we define the *Dikin ellipsoid* centered at x as

$$W_1(x) := \left\{ y \in \mathbb{E} : \frac{1}{\sqrt{a}} \|y - x\|_{f''(x)} \leq 1 \right\}.$$

Then we have $W_1(x) \subset Q$ and for every point $y \in W_1(x)$ we can estimate the Hessian of f at y in term of the Hessian of f at x as

$$(1 - r)^2 f''(x) \preceq f''(y) \preceq \frac{1}{(1 - r)^2} f''(x), \tag{124}$$

where $r := \frac{1}{\sqrt{a}} \|y - x\|_{f''(x)}$. For a proof see [37]-Theorem 2.1.1.
(b) For every point $x, y \in Q$ and for $r := \frac{1}{\sqrt{a}} \|y - x\|_{f''(x)}$, we have

$$\begin{aligned} f(y) &\geq f(x) + \langle f'(x), y - x \rangle + \rho(r), \\ f(y) &\leq f(x) + \langle f'(x), y - x \rangle + \rho(-r), \end{aligned} \tag{125}$$

where $\rho(\cdot)$ is defined in (118). For a proof of this for $a = 1$, see [32] or [37]. The proof for general a is similar.

SC-3 **(Newton iterate):** For every point x, we define the *Newton direction* as

$$\text{Newton}(x) := \text{argmin}_h \left\{ f(x) + f'(x)[h] + \frac{1}{2} f''(x)[h, h] \right\}.$$

Then, we define the *damped Newton iterate* of x as

$$x^+ = x + \frac{1}{1 + \lambda(f, x)} \text{Newton}(x). \tag{126}$$

We have the following properties for a damped Newton step

$$\begin{aligned} &\text{(a)} \quad x^+ \in Q, \\ &\text{(b)} \quad f(x^+) \leq f(x) - a\rho(\lambda(f,x)), \\ &\text{(c)} \quad \lambda(f,x^+) \leq 2\lambda^2(f,x). \end{aligned} \tag{127}$$

For parts (a) and (b), see [37]-Proposition 2.2.2. For part (c), plug in $s = \frac{1}{1+\lambda}$ in [37]-Theorem 2.2.1.

SC-4 **(Existence of minimizer):** f attains its minimizer on Q if and only if f is bounded below on Q, and if and only if there exists $x \in Q$ such that $\lambda(f,x) < 1$. For an arbitrary minimizer x_f, we have

$$\lambda(f,x) < 1 \;\Rightarrow\; f(x) - f(x_f) \leq \rho(-\lambda(f,x)). \tag{128}$$

If f is non-degenerate, the minimizer is unique. The proof of $a = 1$ is in [32], and the proof for general a is similar.

SC-6 **(LF conjugate of a s.c. function):** Let $f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ be convex. The *Legendre-Fenchel (LF) conjugate* of f is defined as

$$f_*(y) := \sup_x \{\langle y, x\rangle - f(x)\}. \tag{129}$$

f_* is always a convex function and its domain is all the points that (129) has a bounded solution. For a proper convex function, we have $(f_*)_* = f$ if and only if the epigraph of f is closed (f is a closed convex function), see for example [23]. We use the following inequality frequently in this paper.

Theorem A.1. (Fenchel-Young inequality) *Let $f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ be a convex function and f_* be its LF conjugate. For every point x in the domain of f and every y in the domain of f_*, we have*

$$f(x) + f_*(y) \geq \langle y, x\rangle. \tag{130}$$

Equality holds if and only if $y \in \partial f(x)$.

Assume that $f(x)$ is differentiable and the optimal value of (129) for $\bar{y}$ is attained at $\bar{x}$, then we must have $\bar{y} = f'(\bar{x})$. By Theorem A.1, if both f and f_* are twice differentiable, for every point x in the domain of f we have

$$x = f_*'(f'(x)) \quad \Rightarrow \quad f_*''(f'(x)) = [f''(x)]^{-1}. \tag{131}$$

Let Q_* be the domain of f_*; the set of all points for which the right hand side of (129) is finite. We mentioned that Q_* is convex and f_* is a convex function on Q_*. It is shown in [37]-

Section 2.4 that $Q_* = f'(Q)$, f_* is a non-degenerate s.c. function and the LF conjugate of f_* is exactly f.

The following lemma is also very useful:

Lemma A.1. *Let f be a 1-s.c. function and x and y in its domain such that $r := \|x-y\|_{f''(x)} < 1$. Then*

$$\|f'(x) - f'(y)\|^*_{f''(x)} \leq \frac{r}{1-r}. \tag{132}$$

Proof. Let us define $q := y - x$. Starting with the fundamental theorem of calculus, we have:

$$\begin{aligned}
\|f'(x) - f'(y)\|^*_{f''(x)} &= \left\| \int_0^1 f''(x+tq)q dt \right\|^*_{f''(x)} \leq \int_0^1 \|f''(x+tq)q\|^*_{f''(x)} dt \\
&\underbrace{\leq}_{(124)} \int_0^1 \tfrac{1}{1-\|tq\|_{f''(x)}} \|f''(x+tq)q\|^*_{f''(x+tq)} dt \\
&= \int_0^1 \tfrac{1}{1-\|tq\|_{f''(x)}} \|q\|_{f''(x+tq)} dt \underbrace{\leq}_{(124)} \left(\int_0^1 \tfrac{1}{(1-tr)^2} dt \right) r = \tfrac{r}{1-r}.
\end{aligned}$$

□

A.2. **Self-concordant (s.c.) barriers.** For a $\vartheta \geq 1$, we say that a 1-s.c. function is a ϑ-s.c. barrier for $\mathrm{cl}(Q)$ if we have

$$|f'(x)[h]| \leq \sqrt{\vartheta} \|h\|_{f''(x)}, \quad \forall (x \in Q, h \in \mathbb{E}). \tag{133}$$

In view of definition (122), a non-degenerate s.c. function f is a ϑ-s.c. barrier if and only if

$$\lambda(f, x) = \|f'(x)\|_{[f''(x)]^{-1}} \leq \sqrt{\vartheta}, \quad \forall x \in Q. \tag{134}$$

If Q is a convex cone, we say f is ϑ-logarithmically-homogenous if for every $x \in Q$, we have

$$f(tx) = f(x) - \vartheta \ln(t), \quad \forall (t > 0). \tag{135}$$

SCB-1 **(Stability under intersections, direct sums, and affine maps)** [37]-Proposition 2.3.1:

(a) Assume that for each $i \in \{1, \ldots, m\}$, f_i is a ϑ_i-s.c. barrier on $\mathbb{E}$ with domains Q_i, and consider real coefficients $\gamma_i \geq 1$. If $Q := \cap_{i=1}^m Q_i$ is not empty, then $f := \sum_{i=1}^m \gamma_i f_i$ is a $(\sum_{i=1}^m \gamma_i \vartheta_i)$-s.c. barrier on Q.

(b) Let f_i, $i \in \{1, \ldots, m\}$, be a ϑ_i-s.c. barrier on $\mathbb{E}_i$ with domains Q_i. Then, the function $f(x^1, \ldots, x^m) := \sum_{i=1}^m f_i(x^i)$, defined on $Q := Q_1 \oplus \cdots \oplus Q_m$, is a $(\sum_{i=1}^m \vartheta_i)$-s.c. barrier on Q.

(c) Let f be a ϑ-s.c. barrier with domain Q and $x = Ay + b$ be an affine mapping with image intersecting Q, then $f(Ay + b)$ is also a ϑ-s.c. barrier on $\{y : Ay + b \in Q\}$.

SCB-2 **(Basic properties of s.c. barrier's):** Let f be a ϑ-s.c. barrier, then the following inequalities hold for every pair $x, y \in Q$ (see [37]-Proposition 2.3.2 and [32]-Chapter 3):

$$f'(x)[y-x] \leq \vartheta; \tag{136}$$

where, as before, $f'(x)[h]$ is the first order differential of f taken at x along the direction h. SCB-3 **(Non-degeneracy, attaining minimizer):** f is non-degenerate if and only if Q does not contain lines. f is bounded below if and only if Q is bounded. Then, f is non-degenerate and attains its unique minimizer x_f on Q, and we have

$$\left\{y : \|y - x_f\|_{f''(x_f)} < 1\right\} \subseteq Q \subseteq \left\{y : \|y - x_f\|_{f''(x_f)} < \vartheta + 2\sqrt{\vartheta}\right\}. \tag{137}$$

Let us prove the following lemma that is useful in this paper.

Lemma A.2. *Let Φ be a ϑ-s.c. barrier with domain $\mathrm{int}D \subset \mathbb{E}$, and $\xi > 1$. Then, the function $\Phi\left(\frac{z}{\tau}\right) - \xi\vartheta\ln(\tau)$ with domain $\{(z,\tau) : \tau > 0, \frac{z}{\tau} \in \mathrm{int}D\}$ is a $\bar{\xi}$-s.c. function for an absolute constant $\bar{\xi}$ depending on ξ. Moreover, its LF conjugate and also the summation of $\Phi\left(\frac{z}{\tau}\right) - \xi\vartheta\ln(\tau)$ with its LF conjugate are also $\bar{\xi}$-s.c. functions.*

Proof. Consider the function $\Phi(\frac{z}{\tau}) - \xi\vartheta\ln(\tau)$. First we show that the function is convex. Let us define

$$g(\alpha) := \Phi\left(\frac{z + \alpha d_z}{\tau + \alpha d_\tau}\right) - \xi\vartheta\ln(\tau + \alpha d_\tau).$$

Then, we have

$$g''(0) \;=\; \frac{1}{\tau^2}\left[\langle d_z - \frac{d_\tau}{\tau}z, \Phi''\left(\frac{z}{\tau}\right)\left(d_z - \frac{d_\tau}{\tau}z\right)\rangle - 2d_\tau\langle\Phi'\left(\frac{z}{\tau}\right), d_z - \frac{d_\tau}{\tau}z\rangle + \xi\vartheta d_\tau^2\right].$$

By using inequality (133) for the middle term and doing some simple algebra we get

$$g''(0) \geq \frac{1}{\tau^2}\left[\left\|d_z - \frac{d_\tau}{\tau}z\right\|_{\Phi''} - |d_\tau|\sqrt{\vartheta}\right]^2 + (\xi - 1)\frac{d_\tau^2}{\tau^2}\vartheta. \tag{138}$$

(138) shows that $\Phi(\frac{z}{\tau}) - \xi\vartheta\ln(\tau)$ is strictly convex for every $\xi > 1$.

To prove that it is a s.c. function, we show that there exists an absolute constant $\bar{\xi}$ depending on ξ such that

$$|g'''(0)| \leq 2\bar{\xi}^{-1/2}(g''(0))^{3/2}.$$

For simplicity, let us define $h := \frac{1}{\tau}\left(d_z - \frac{d_\tau}{\tau}z\right)$. First, note that from (138) we have

$$\begin{aligned} \left|\frac{d_\tau}{\tau}\sqrt{\vartheta}\right| &\leq \frac{\sqrt{g''(0)}}{\sqrt{\xi - 1}}, \\ \|h\|_{\Phi''} &\leq \sqrt{g''(0)} + \left|\frac{d_\tau}{\tau}\sqrt{\vartheta}\right| \leq \underbrace{\left(1 + \frac{1}{\sqrt{\xi - 1}}\right)}_{=:\gamma}\sqrt{g''(0)}. \end{aligned} \tag{139}$$

By expanding the expression for $g'''(0)$, we have

$$g'''(0) = \Phi'''[h,h,h] - 6\Phi''[h,h]\left(\frac{d_\tau}{\tau}\right) + 6\Phi'[h]\left(\frac{d_\tau}{\tau}\right)^2 - 2\xi\vartheta\left(\frac{d_\tau}{\tau}\right)^3. \tag{140}$$

Because Φ is a 1-s.c. function, by definition in (121), we have $|\Phi'''[h,h,h]| \leq 2(\Phi''[h,h])^{3/2} = 2(\|h\|_{\Phi''})^3$, and because Φ is a ϑ-s.c barrier, by definition (133), we have $|\Phi'[h]| \leq \sqrt{\vartheta}\|h\|_{\Phi''}$. Substituting these in (140), using the inequalities in (139) and the fact that $\vartheta \geq 1$, we have:

$$g'''(0) \leq \left(2\gamma^3 + \frac{6\gamma^2}{\sqrt{\xi-1}} + \frac{6\gamma}{\xi-1} + \frac{2\xi}{(\xi-1)^{3/2}}\right)(g''(0))^{3/2}, \tag{141}$$

where γ is defined in (139).

For the second part of the lemma for the conjugate function, see the proof of Theorem 2.4.1 in [37]. □

A.3. **LF conjugate of s.c. barriers.** If f is a ϑ-s.c. barrier, then f_* is a s.c. function, but it is **not** necessarily a s.c. barrier. Q_* is either the entire $\mathbb{E}^*$ if Q is bounded, or the open cone

$$\mathrm{rec}_*(Q) := \{s \in \mathbb{E}^* : \langle s, h\rangle < 0, \forall h \in \mathrm{rec}(Q)\}, \tag{142}$$

where $\mathrm{rec}(Q)$ is the *recession cone* of Q defined as

$$\mathrm{rec}(Q) := \{h \in \mathbb{E} : x + th \in Q, \quad \forall x \in Q, \ \forall t \geq 0\}. \tag{143}$$

In this article, we frequently use the fact that f_* has some useful properties beyond those of an arbitrary s.c. function, such as Theorem 2.1 and the following lemma:

Lemma A.3 (Theorem 2.4.2 of [37])**.** *Assume that f is a ϑ-s.c. barrier on D and let f_* be the LF conjugate of f with domain D_*. Then, for every point $y \in D_*$, we have*

$$f_*''(y)[y,y] \leq \vartheta. \tag{144}$$

Appendix B. Examples of s.c. functions to clarify Figure 1

It is well-known that $-\ln(x)$ is a 1-LH s.c. barrier for the cone $\mathbb{R}_{++}$ and its LF conjugate $-1-\ln(-y)$ is also a 1-LH s.c. barrier. Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function with the LF conjugate f_*. Then, we can easily verify that for every $b \in \mathbb{R}^n$, the LF conjugate of $f(x-b)$ is $\langle b, y\rangle + f_*(y)$. Consider the following univariate function and its LF conjugate:

$$f(x) := -\ln(x-1), \quad f_*(y) = -1 + y - \ln(-y).$$

$f(x)$ is a 1-s.c. barrier. $f_*(y)$ is a s.c. function, but is not a s.c. barrier.

As it is shown in Figure 1, if a function is LH s.c. barrier, its LF conjugate is also a LH s.c. barrier [37]. A question is: does there exist a s.c. barrier f that is not LH, while its LF conjugate

f_* is also a s.c. barrier, as pictorially shown in Figure 1? Note that by Subsection A.3, both $\mathrm{dom} f$ and $\mathrm{dom} f_*$ must be convex cones. The following theorem shows that the answer is yes:

Theorem B.1. *Assume that $f(x)$ is a non-degenerate ϑ-LH s.c. barrier with domain $K \subset \mathbb{R}^n$ and let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation such that $AK \subseteq K$. Then, for every $b \in K \backslash \{0\}$, the function $g(x) := f(Ax + b) + f(x)$ is a 2ϑ-s.c. barrier, and its LF conjugate g_* is also a s.c. barrier.*

Proof. By the material of subsection A.2, we know that f is a 2ϑ-s.c. barrier. To show g_* is also a s.c. barrier, we need to prove that $\langle g_*'(y), [g_*''(y)]^{-1} g_*'(y)\rangle$ is bounded by an absolute constant for every $y \in K_*$. For a given y, let $x := g_*'(y)$, then by the properties of LF conjugate, we have

$$\begin{aligned} \langle g_*'(y), [g_*''(y)]^{-1} g_*'(y)\rangle &= \langle x, g''(x)x\rangle \\ &= \langle Ax, f''(Ax+b)Ax\rangle + \langle x, f''(x)x\rangle \\ &\leq (\vartheta + 2\sqrt{\vartheta})^2 \langle Ax, f''(Ax)Ax\rangle + \langle x, f''(x)x\rangle \\ &= (\vartheta + 2\sqrt{\vartheta})^2 \vartheta + \vartheta, \quad \text{[37]-eq (2.3.14).} \end{aligned} \tag{145}$$

For the inequality above, we used equation [32]-(3.16) and also the fact that $Ax + \alpha b \in K$ for all $\alpha \in \mathbb{R}_+$ and so $\pi_{Ax}(Ax + b) = 0$, where π is the Minkowski function of K (defined in [37]-Subsection 2.3.2 or [32]). Inequality (145) confirms that g_* is a $((\vartheta + 2\sqrt{\vartheta})^2\vartheta + \vartheta)$-s.c. barrier. $\square$

As an example, let $f(x) := -\sum_{i=1}^m \ln(a_i^\top x)$ for $a_i \in \mathbb{R}^n$, $i \in \{1, \ldots, m\}$, which is a m-LH s.c. barrier. Then, the function $g(x) := -\sum_{i=1}^m \ln(a_i^\top x) - \sum_{i=1}^m \ln(a_i^\top x + 1)$ is a $2m$-s.c. barrier that is not LH and g_* is also a s.c. barrier.