

INTERIOR-POINT ALGORITHMS FOR CONVEX OPTIMIZATION BASED ON PRIMAL-DUAL METRICS

TOR G. J. MYKLEBUST, LEVENT TUNÇEL

ABSTRACT. We propose and analyse primal-dual interior-point algorithms for convex optimization problems in conic form. The families of algorithms whose iteration complexity we analyse are so-called short-step algorithms. Our iteration complexity bounds match the current best iteration complexity bounds for primal-dual symmetric interior-point algorithm of Nesterov and Todd, for symmetric cone programming problems with given self-scaled barriers. Our results apply to any self-concordant barrier for any convex cone. We also prove that certain specializations of our algorithms to hyperbolicity cone programming problems (which lie strictly between symmetric cone programming and general convex optimization problems in terms of generality) can take advantage of the favourable special structure of hyperbolic barriers. We make new connections to Riemannian geometry, integrals over operator spaces, Gaussian quadrature, and strengthen the connection of our algorithms to quasi-Newton updates and hence first-order methods in general.

Date: November 2014, revised: April 6, 2016.

Key words and phrases. primal-dual interior-point methods, convex optimization, variable metric methods, local metric, self-concordant barriers, Hessian metric, polynomial-time complexity;

AMS subject classification (MSC): 90C51, 90C22, 90C25, 90C60, 90C05, 65Y20, 52A41, 49M37, 90C30.

Tor Myklebust: Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada (e-mail: tmyklebu@csclub.uwaterloo.ca). Research of this author was supported in part by an NSERC Doctoral Scholarship, ONR Research Grant N00014-12-10049, and a Discovery Grant from NSERC. Some of the results in this paper appeared in the PhD Thesis of the first author [19].

Levent Tunçel: Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada (e-mail: ltuncel@uwaterloo.ca). Research of this author was supported in part by ONR Research Grants N00014-12-10049, N00014-15-12171, and Discovery Grants from NSERC.

1. INTRODUCTION

Convex optimization problems (of minimizing a given convex function in a given convex set) form a beautiful research area with very powerful theory and algorithms and many far-reaching applications. Among the main algorithms to solve convex optimization problems are modern interior-point methods. The modern theory of interior-point methods have flourished since Karmarkar’s ground-breaking paper [11].

Every convex optimization problem can be paired with another convex optimization problem based on the same data, called its *dual*. Rigorous solution methods for convex optimization problems typically generate, together with solutions to the problem at hand (primal problem), solutions for its dual. A particularly successful line of research pursued methods that work “equally hard” at solving both primal and dual problems simultaneously, called *primal-dual symmetric methods* (for a rigorous definition, see [45]).

In the special cases of linear programming (LP) and semidefinite programming (SDP), these primal-dual symmetric methods, in addition to carrying certain elegance, led to improved results in theory, in computation, and in applications.

Part of the success of primal-dual symmetric methods for LP and SDP might stem from the fact that both classes admit convex conic formulations where the underlying cone is *self-dual* (the primal convex cone and the dual convex cone are linearly isomorphic) and *homogeneous* (the automorphism group of the cone acts transitively in its interior). Convex cones that are both homogeneous and self-dual are called *symmetric cones*. The success of primal-dual symmetric interior-point methods was further extended to the setting of symmetric cone programming, which led to deeper connections with other areas of mathematics.

In this paper, we will extend many of the underlying mathematical entities, analyses and iteration complexity bounds of these algorithms to a general convex optimization setting (with arbitrary convex cones).

The primal variable x lives in a finite dimensional vector space \mathbb{E} and the dual variables y and s live in the finite dimensional vector spaces \mathbb{Y} and \mathbb{E}^* respectively, where \mathbb{E}^* is the dual space of \mathbb{E} . Every convex optimization problem can be put into the following conic form (under some mild assumptions):

$$(P) \quad \begin{aligned} \inf \quad & \langle c, x \rangle \\ & \mathcal{A}x = b, \\ & x \in K, \end{aligned}$$

where $\mathcal{A} : \mathbb{E} \rightarrow \mathbb{Y}^*$ is a linear map, $b \in \mathbb{Y}^*$, $c \in \mathbb{E}^*$ and $K \subset \mathbb{E}$ is a pointed, closed, convex cone with nonempty interior. We assume that \mathcal{A}, b, c are given explicitly and K is described via $F : \text{int}(K) \rightarrow \mathbb{R}$, a *logarithmically homogeneous self-concordant barrier function for K* (defined in the next section). We assume without loss of generality that \mathcal{A} is surjective (i.e., $\mathcal{A}\mathbb{E} = \mathbb{Y}^*$).

We define the dual of (P) as

$$(D) \quad \sup \quad \langle b, y \rangle_D \\ \mathcal{A}^*y + s = c, \\ s \in K^*,$$

where K^* is the dual of cone K , namely

$$K^* := \{s \in \mathbb{E}^* : \langle s, x \rangle \geq 0, \forall x \in K\}.$$

We are using $\langle \cdot, \cdot \rangle$ to denote the dual pairing on $(\mathbb{E}, \mathbb{E}^*)$ and $\langle \cdot, \cdot \rangle_D$ to denote the dual pairing on $(\mathbb{Y}, \mathbb{Y}^*)$. $\mathcal{A}^* : \mathbb{Y} \rightarrow \mathbb{E}^*$ denotes the adjoint of \mathcal{A} defined by the equations:

$$\langle \mathcal{A}^*y, x \rangle = \langle \mathcal{A}x, y \rangle_D, \quad \forall x \in \mathbb{E}, y \in \mathbb{Y}.$$

For both of these spaces \mathbb{E} and \mathbb{Y} , we identify the dual of the dual space with the original space, i.e., $\mathbb{E}^{**} = \mathbb{E}$ and $\mathbb{Y}^{**} = \mathbb{Y}$. In this setting, under our assumptions, K^* is also a pointed, closed convex cone with nonempty interior and $K^{**} = K$. Now, it is not difficult to verify that under our definitions, the optimization problem dual to the problem (D) is indeed equivalent to (P) . Primal-dual symmetric conic formulations of convex optimization problems with self-concordant barrier functions were suggested by Nesterov and Nemirovski [28], and since their inception in the early 1990s, such formulations have had a significant influence on modern convex optimization.

In this paper, we utilize many of the fundamental techniques developed throughout the history of interior-point methods. One of the main algorithms we design and analyse is a predictor-corrector algorithm generalizing the algorithm of Mizuno, Todd and Ye [15] from the LP setting. However, even in the LP setting, some of our algorithms are new. Our algorithms use Newton directions as in Renegar’s algorithm for LP [35], and one of our main algorithms uses a similar predictor-corrector scheme, but both predictor and corrector parts in a primal-dual setting.

The modern theory of interior-point methods employed the concept of self-concordant barrier functions (see Nesterov and Nemirovski [29]). The Hessians of these nice convex functions induce local metrics with excellent properties. For instance, the unit ball induced by the Hessian at a point, called the *Dikin ellipsoid*, is contained in the cone.

The self-concordance property implies that, as we make local moves in the domain of the barrier function, the local metrics do not change fast unless their norms are large. More precisely, the speed of the change in the local metric can be bounded in terms of the local metric itself.

One of the indicators of how good a barrier function is (in terms of the local metrics it generates) can be measured by how well the Dikin ellipsoid approximates the domain of the function. This leads to the notion of *long-step Hessian estimation property* (defined in Section 2) of barriers. This property amounts to extending the controlled change of the Hessian of the barrier to the “best, norm-like local approximation” of the domain of the barrier. Various long-step strategies for interior-point methods have been investigated extensively in [21]. This long-step Hessian estimation property has been proven for self-scaled barriers [30, 31] (whose domains are

symmetric cones and these barriers have been completely classified [8, 9, 42]) and hyperbolic barriers [6] (whose domains are hyperbolicity cones; for results on the structure of these cones see [6, 1, 37, 38]), but there exist barriers for which this property fails.

For a related, but weaker notion of long-step, also see [26]. Indeed, we would like our algorithms to exploit such properties when they are present. (Beyond the set-up of symmetric cones and self-scaled barriers, exploitation of this long step property forces us to break the primal-dual symmetry of our algorithms in the favor of the problem with barrier function admitting this property.) Our general approach and many of our main technical results are primal-dual symmetric (in the sense of [45]); however, beyond symmetric cones and self-scaled barriers, there may be advantages to breaking the primal-dual symmetry in favor of the better behaved (or better posed) problem. A large part of these favorable structures of hyperbolic barriers that we understand, lies with the long-step Hessian-estimation property, another large part are the special algebraic structures given by these polynomials defining the self-concordant barriers (such structures created advantageous situations for breaking primal-dual symmetry, even in the case of convex optimization with univariate polynomials, see [22, 5, 34]). Our approach also provides useful tools for exploiting such structures when they are present.

Independent of this work, recently, simultaneously with an announcement of this work [18], Renegar announced a primal-dual affine-scaling method for hyperbolicity cone programming [39] (also see Renegar and Sondjaja [40]). Nesterov and Todd [30, 31] present very elegant primal-dual interior-point algorithms with outstanding mathematical properties in the setting of self-scaled barriers; however, beyond self-scaled barriers, there are many other primal-dual approaches that are not as symmetric, but retain some of the desired properties (see [47, 20, 2, 3, ?, 25]).

Some recent study of the interior-point methods and the central paths defined by self-concordant barriers led to connections with Riemannian geometry, see [32, 27]. We also make some additional connections to Riemannian geometry through the local primal-dual metrics that we utilize in this paper.

Hessians of self-concordant barriers, in addition to inducing local metrics, provide linear isomorphisms between primal and dual spaces \mathbb{E} and \mathbb{E}^* . In the special case of self-scaled barriers, we have $F''(w)\text{int}(K) = \text{int}(K^*)$ for every $w \in \text{int}(K)$. We focus on generalization of such behaviour and simulate it with a symmetric positive-definite bilinear form T^2 mapping \mathbb{E}^* to \mathbb{E} . Upon fixing an inner product (e.g., based on the Hessians of F and F_* evaluated at the starting pair $(x^{(0)}, s^{(0)})$ and considering realization of $T^2 : \mathbb{R}^n \rightarrow \mathbb{R}^n$, this leads to (via its unique self-adjoint, positive-definite square-root T) construction of a v -space as a space that is “half-way between” \mathbb{E} and \mathbb{E}^* , i.e., $\mathbb{E}^{(* / 2)}$.

The overall structure of the remainder of this paper is as follows:

- Section 2 presents some fundamental notation, definitions and properties of underlying primal-dual spaces and the class of convex barrier functions.

- Section 3 presents a new primal-dual scaling map based on integration of the barrier's Hessian.
- Section 4 presents some low-rank update formulae for the construction of local primal-dual metrics and connects these formulae to classical work on quasi-Newton updates.
- Section 5 delves deeper into investigating the relationship between the set of Hessians of self-concordant barriers and the set of local primal-dual metrics we use.
- Sections 6 and 7 combine a crude approximation of the scaling from Section 4 with the technique of Section 3 to derive and analyse some theoretically efficient interior-point algorithms for convex programming.
- Section 8 examines some of the special properties of our approach for hyperbolic cones that may make our techniques particularly effective.
- Some side issues, such as technical details of primal-dual symmetry of our approach (which is established via geodesic convexity of the underlying sets of local metrics), details of the complexity analysis in terms of individual tensor terms and with specific absolute constants, and the numerical evaluation of integrals of Hessians are relegated to Appendices A, B and C respectively.

Before moving into the technical development, we next summarize the major goals of the paper, highlight some of the major ideas, and describe how our results fit in the general framework of interior-point methods and in which way our results improve the state-of-the-art in interior-point-methods for general convex cones:

- Since the proposal of primal-dual symmetric polynomial-time interior-point algorithms in the late 1980's, the problems of determining which classes of convex optimization problems admit such algorithms and of determining what worst-case iteration complexity bounds are achievable for such algorithms have been of great interest to many researchers. One of our major goals is to prove that for all convex optimization problems in the conic form, there exists such a primal-dual symmetric interior-point algorithm attaining the current best, worst-case iteration complexity bound for interior-point algorithms. Another major goal is to construct these proofs and develop the underlying theory in such a manner that in addition to the usual theoretical goals, paves the way for numerical implementations and further inspires the design and analysis of better algorithms.
- Some of the highlights in the paper include the construction of primal-dual metrics with special properties via integration of Hessians of convex barrier functions, new decompositions for the existing rank-4 update formulae for interior-point methods [47], that are written as two consecutive rank-2 update formulas, with the second rank-2 formula exposing new, very meaningful terms in regard to proximity to the central path. Another feature of the paper is on the one hand, a high-level complexity analysis that reveals geometric and algebraic insights and emphasizes the generality of the framework, on the other hand, a very detailed term by term analysis (relegated to an appendix)

that is very useful in designing robust algorithms which perform numerical computations accurately.

- Even though the theoretical analysis with the polynomial iteration complexity bound only applies to short-step version of the algorithm, our development of the underlying theory does properly generalize Nesterov and Todd’s primal-dual symmetric, long-step algorithms and in more general settings is applicable to the analysis of long-step algorithms, provided one is willing to sacrifice at least some of the primal-dual symmetry when necessary (that is, beyond the symmetric cone case).
- We also provide short proofs of existence of integral scalings, and a relatively transparent proof of nonexistence of pointwise evaluated primal-dual Hessian metrics beyond symmetric cones. This proof utilizes an axiomatic characterization of Euclidean Jordan Algebras in an elementary way.
- We further provide deeper connections from primal-dual metrics for interior-point methods to quasi-Newton methods, Riemannian geometry (both via integration of Hessians, and via a primal-dual symmetry proof (Appendix A)).
- Algorithms that we analyse are new even for Linear Programming problems. In our computational experiments, we observed that the practical versions of the algorithms with low-rank updates save a few Cholesky decompositions on many of the test instances (see [19]). This is already a promising development indicating some computational promise of the approach. The algorithms allow for adjusting the amount of the second-order information utilized from one iteration to the next.

2. PRELIMINARIES

In this section, we introduce some of the fundamental concepts, definitions and properties that will be useful in the rest of the paper. For a more detailed exposure to these concepts, see the standard references [29, 30, 31, 36, 23] as well as [45, 47].

Definition 2.1. (*Self-concordance*) *Let $K \subset \mathbb{E}$ be a pointed, closed convex cone with nonempty interior and let $F : \text{int}(K) \rightarrow \mathbb{R}$ be a C^3 -smooth, closed convex function such that the domain of F is $\text{int}(K)$ and there exists $\vartheta \in \mathbb{R}$ such that, for every $t > 0$,*

$$F(tx) = F(x) - \vartheta \ln(t),$$

and

$$(1) \quad |D^3F(x)[h, h, h]| \leq 2 (D^2F(x)[h, h])^{3/2}$$

for all $x \in \text{int}(K)$ and for all $h \in \mathbb{E}$. Then F is called a ϑ -logarithmically homogeneous self-concordant barrier (ϑ -LHSCB) for K .

It follows from the above definition that $\vartheta \geq 1$ and that for every sequence in the interior of K , converging to a boundary point of K , the corresponding function values $F(x) \rightarrow +\infty$.

If F is a ϑ -LHSCB for K , then its Legendre-Fenchel conjugate

$$F_*(s) := \sup\{-\langle s, x \rangle - F(x) : x \in \text{int}(K)\},$$

is a ϑ -LHSCB for the dual cone K^* (Nesterov and Nemirovskii [29]). We refer to F_* simply as the *conjugate barrier*.

Once we have a ϑ -LHSCB F for K , at every point $x \in \text{int}(K)$, $F''(x)[\cdot, \cdot] : \mathbb{E} \oplus \mathbb{E} \rightarrow \mathbb{R}$ is a symmetric, bilinear, positive-definite form. Hence, the Hessian of F defines a local metric. For every $h \in \mathbb{E}$ the *local norm induced by F at x* is

$$\|h\|_x := \langle F''(x)h, h \rangle^{1/2}.$$

Moreover, for every $x \in \text{int}(K)$, we may think of $F''(x) : \mathbb{E} \rightarrow \mathbb{E}^*$ as a linear map from the primal space to the dual space. Therefore, $[F''(x)]^{-1} : \mathbb{E}^* \rightarrow \mathbb{E}$ is well-defined and $[F''(x)]^{-1}[\cdot, \cdot] : \mathbb{E}^* \oplus \mathbb{E}^* \rightarrow \mathbb{R}$ is a symmetric, bilinear, positive-definite form. For every $u \in \mathbb{E}^*$, we define

$$\|u\|_x^* := \left\langle u, [F''(x)]^{-1} u \right\rangle^{1/2}.$$

Proposition 2.2. *Let F be a ϑ -LHSCB for K . Then for every $x \in \text{int}(K)$, $\|\cdot\|_x^*$ is the norm dual to $\|\cdot\|_x$. I.e., for every $x \in \text{int}(K)$,*

$$\|u\|_x^* = \sup \{ \langle u, h \rangle : \|h\|_x \leq 1, h \in \mathbb{E} \}, \quad \forall u \in \mathbb{E}^*.$$

The above proposition implies:

$$|\langle u, h \rangle| \leq \|u\|_x^* \|h\|_x, \quad \forall u \in \mathbb{E}^*, h \in \mathbb{E}.$$

We use the above ‘‘Cauchy-Schwarz’’ inequality quite often. Note that

$$\|F'(x)\|_x^* = \|h\|_x,$$

where $h := [F''(x)]^{-1} F'(x)$, the Newton step at x for minimizing F .

Theorem 2.3. *Let F be a ϑ -LHSCB for K . Then, for every $x \in \text{int}(K)$, the open unit ellipsoid centered at x and defined by the positive-definite Hessian $F''(x)$ is contained in the interior of the cone. That is*

$$E(x; F''(x)) := \{z \in \mathbb{E} : \langle F''(x)(z - x), z - x \rangle < 1\} \subset \text{int}(K).$$

Moreover, for every $z \in \text{int}(K)$ such that $\alpha := \|x - z\|_x < 1$, we have

$$(1 - \alpha)^2 F''(x)[h, h] \leq F''(z)[h, h] \leq \frac{1}{(1 - \alpha)^2} F''(x)[h, h],$$

for all $h \in \mathbb{E}$.

We use (as above) $F''(x)[h^{(1)}, h^{(2)}]$ to denote the second derivative of F evaluated along the directions $h^{(1)}, h^{(2)} \in \mathbb{E}$. We also use the notation $\langle F''(x)h^{(1)}, h^{(2)} \rangle$ to denote the same quantity. In both expressions, $F''(x)$ is the Hessian of F . As we deal with the Hessians and other

symmetric bilinear forms in the same space, we sometimes utilize the Löwner order, we write $A \preceq B$ to mean $(B - A)$ is a symmetric, bilinear positive semidefinite form. With these clarifications, the above inequalities in the statement of the last theorem, can be equivalently written as:

$$(1 - \alpha)^2 F''(x) \preceq F''(z) \preceq \frac{1}{(1 - \alpha)^2} F''(x);$$

we refer to the above relations as the *Dikin ellipsoid bound*.

For every $x \in \text{int}(K)$ and every $h \in \mathbb{R}^n$, define

$$\sigma_x(h) := \frac{1}{\sup \{t : (x - th) \in K\}}.$$

We say that F has the *long-step Hessian estimation property* if

$$(2) \quad \frac{1}{[1 + t\sigma_x(-h)]^2} F''(x) \preceq F''(x - th) \preceq \frac{1}{[1 - t\sigma_x(h)]^2} F''(x),$$

for every $x \in \text{int}(K)$, $h \in \mathbb{R}^n$ and $t \in [0, 1/\sigma_x(h)]$. Nesterov and Todd [30] proved that every self-scaled barrier has this property. Güler [6] extended this property to *hyperbolic barriers*. However, Nesterov proved (see Theorem 7.2 in [6]) that the relation (2) can hold for a self-concordant barrier and its conjugate only if K is a symmetric cone. Essentially equivalent properties are expressed as the *convexity of $\langle -F'(x), u \rangle$* , (in x) for every $u \in K$, or, as sufficiently smooth F having *negative curvature*: $F'''(x)[u] \preceq 0$ for every $x \in \text{int}(K)$ and for every $u \in K$.

All of the properties listed in the next theorem can be derived directly from the logarithmic homogeneity property of F . We denote the k th derivative of F by $D^k F$.

Theorem 2.4. *Let F be a ϑ -LHSCB barrier for K . Then for all $x \in \text{int}(K)$ and $s \in \text{int}(K^*)$, F has the following properties:*

- (1) For all $k \geq 1$ integer and $t > 0$, if F is k times differentiable, then $D^k F(tx) = \frac{1}{t^k} D^k F(x)$;
- (2) $\langle -F'(x), x \rangle = \vartheta$;
- (3) $F''(x)x = -F'(x)$;
- (4) $\langle F''(x)x, x \rangle = \vartheta = \|x\|_x^2$, $\langle [F''(x)]^{-1} F'(x), F'(x) \rangle = \vartheta = (\|F'(x)\|_x^*)^2$;
- (5) $F'''(x)[x] = -2F''(x)$.

The LHSCB F and its conjugate barrier F_* interact very nicely due to the elegant and powerful analytic structure imposed by the Legendre-Fenchel conjugacy:

Theorem 2.5. *Let F be a ϑ -LHSCB barrier for K . Then for all $x \in \text{int}(K)$ and $s \in \text{int}(K^*)$, F and F_* satisfy the following properties:*

- (1) $F_*(-F'(x)) = -\vartheta - F(x)$ and $F(-F'_*(s)) = -\vartheta - F_*(s)$;
- (2) $-F'_*(-F'(x)) = x$ and $-F'(-F'_*(s)) = s$;

$$(3) \quad F_*''(-F'(x)) = [F''(x)]^{-1} \text{ and } F_*''(-F'(s)) = [F''(s)]^{-1}.$$

Maps between primal space \mathbb{E} and the dual space \mathbb{E}^* play very fundamental roles. Among such maps, those which take $\text{int}(K)$ to $\text{int}(K^*)$ bijectively are particularly useful. One such map is $-F'(\cdot)$ (i.e., the *duality mapping*). In the special case of symmetric cones and self-scaled barriers F , $F''(u)$ gives a linear isomorphism between $\text{int}(K)$ and $\text{int}(K^*)$ for every choice of $u \in \text{int}(K)$.

Next, we define three sets of such linear transformations: $\mathcal{T}_0^2, \mathcal{T}_1^2, \mathcal{T}_2^2$ [47]. We further require in the definitions that these linear transformations, just like Hessians of self-concordant barriers above, be also symmetric bilinear positive-definite forms. These sets are indexed based on their usage of information from the underlying self-concordant barrier functions (zeroth-order information only, up to first-order information only, and up to second-order information only, respectively). We denote by $\mathbb{S}_{++}(\mathbb{E}, \mathbb{E})$ the set of symmetric bilinear positive-definite forms $H[\cdot, \cdot] : \mathbb{E} \oplus \mathbb{E} \rightarrow \mathbb{R}$. Note that such forms are also interpreted in this paper as linear maps $H : \mathbb{E} \rightarrow \mathbb{E}^*$.

Sometimes it is suitable to fix some inner product on \mathbb{E} , and hence some linear isomorphism between \mathbb{E} and \mathbb{E}^* , and identify both \mathbb{E} and \mathbb{E}^* by \mathbb{R}^n ($n := \dim(\mathbb{E})$). We can then represent every $T^2 \in \mathbb{S}_{++}(\mathbb{E}^*, \mathbb{E}^*)$ by a symmetric positive-definite matrix over the reals. We use the same notation $T^2 \in \mathbb{S}_{++}^n$ for such representations. Then, we are able utilize the unique symmetric positive-definite square-root, $T \in \mathbb{S}_{++}^n$ of T^2 . Since $T^2 : \mathbb{E}^* \equiv \mathbb{R}^n \rightarrow \mathbb{E} \equiv \mathbb{R}^n$, one may conceptually think of $T : \mathbb{E}^* \rightarrow \mathbb{E}^{*/2}$, as we eluded to in the Introduction. Such representations are only necessary for our explicit discussions of the v -space constructions and for drawing closer parallels to the Linear Programming special case. In fact, the algorithms can be stated, implemented and analysed without any need to directly computing T or even referring to it (we can fully operate with T^2 and T^{-2}).

Definition 2.6. For every pair $(x, s) \in \text{int}(K) \oplus \text{int}(K^*)$, we define

$$\mathcal{T}_0^2(x, s) := \{T^2 \in \mathbb{S}_{++}(\mathbb{E}^*, \mathbb{E}^*) : T^2 s = x\}.$$

In words, $\mathcal{T}_0^2(x, s)$ is the set of all symmetric bilinear positive-definite forms which map s to x .

It may not be obvious, but this set is nonempty for every convex cone K and for every pair of points x and s in the interiors of respective cones. A proof of this fact was given, in the context of self-scaled barriers, by Nesterov and Todd [30] and two other proofs were given in [47] (Theorems 3.2, and 3.3). In fact, this set is convex and its dimension is $\Omega(n^2)$.

For a given pair of x, s , let us define

$$\mu(x, s) := \frac{\langle s, x \rangle}{\vartheta}.$$

Note that if x and s are feasible in their respective problems, then $\vartheta\mu(x, s)$ is their duality gap:

$$\vartheta\mu(x, s) = \langle s, x \rangle = \langle c, x \rangle - \langle b, y \rangle_D \geq 0.$$

We will often write simply μ for $\mu(x, s)$. We immediately further abuse this notation and consider for every $\mu > 0$, the unique minimizers of the problem

$$(P_\mu) \quad \min \quad \frac{1}{\mu} \langle c, x \rangle + F(x) \\ \mathcal{A}x = b, \\ (x \in \text{int}(K)).$$

Under the assumptions that both problems (P) and (D) have Slater points, and that \mathcal{A} is surjective, (P_μ) has a unique solution for every $\mu > 0$. The necessary and sufficient conditions for optimality applied to (P_μ) yield that for every $\mu > 0$, the unique solution $x(\mu)$ of the minimization problem above make up the x -part of the unique solution $(x(\mu), y(\mu), s(\mu))$ to the system:

$$\begin{aligned} \mathcal{A}x &= b, & x &\in \text{int}(K) \\ \mathcal{A}^*y + s &= c, \\ s &= -\mu F'(x). \end{aligned}$$

Moreover, these solutions $(x(\mu), y(\mu), s(\mu))$ define a smooth path, called *central path*. Now we can explain the latest abuse of notation; we have, for such points on the central path:

$$\langle s(\mu), x(\mu) \rangle = \vartheta \mu.$$

Therefore, our definition of $\mu(x, s)$ is consistent with this usage of μ as well.

To pass from the general set up of T^2 to the Euclidean structure requiring set up of T , we may use a “static” Hessian to (for example, assume $(x^{(0)}, s^{(0)})$ lies on the central path with parameter $\mu = 1$, and based on $F'_*(s^{(0)}) = [F''(x^{(0)})]^{-1}$) fix an inner product for primal and dual spaces as in [34].

Let $T^2 \in \mathcal{T}_0^2(x, s)$, and define T as above. Upon further defining $v := Ts = T^{-1}x$, we observe

$$\vartheta \mu = \langle s, x \rangle = (s^\top T) (T^{-1}x) = v^\top v.$$

These linear transformations allow generalizations of so-called *v-space* approaches to primal-dual interior-point methods from LP and SDP (see for instance, [17, 12, 10, 44, 24]) as well as symmetric cone programming settings [30, 31, 45, 7] to the general convex optimization setting [45, 47]. What perhaps started as a convenience of notation as well as simplicity and elegance of analysis in the 1980’s, by now turned into a solid theoretical framework in which deep problems can be posed and solved whether they relate to the mathematical structures or are directly motivated by a family of algorithms.

Note that the set \mathcal{T}_0^2 does not use any information about the most difficult constraints of the convex optimization problem. Indeed, in the worst-case, certain algorithms using only \mathcal{T}_0^2 need not even converge to an optimal solution (see [50]). Using first order information via $F'(x)$ and $F'_*(s)$, we can focus on a smaller, but more interesting and useful subset of \mathcal{T}_0^2 :

Definition 2.7. For every pair $(x, s) \in \text{int}(K) \oplus \text{int}(K^*)$, we define

$$\mathcal{T}_1^2(x, s) := \left\{ H \in \mathbb{S}_{++}(\mathbb{E}^*, \mathbb{E}^*) : Hs = x, H \begin{bmatrix} -F'(x) \end{bmatrix} = -F'_*(s) \right\}.$$

Just like \mathcal{T}_0^2 , \mathcal{T}_1^2 is also convex, nonempty and has dimension $\Omega(n^2)$. However, the nonemptiness may be less obvious (for a proof, see [47], or Sections 3 and 4 of this paper).

For convenience, we sometimes write $\tilde{x} := \tilde{x}(s) := -F'_*(s)$ and $\tilde{s} := \tilde{s}(x) := -F'(x)$. One can think of \tilde{x} and \tilde{s} as the *shadow* iterates, as $\tilde{x} \in \text{int}(K)$ and $\tilde{s} \in \text{int}(K^*)$ and if (x, s) is a feasible pair, then $\mu\tilde{x} = x$ iff $\mu\tilde{s} = s$ iff (x, s) lies on the central path. In this paper, s and decorated versions of it; \tilde{s} , \check{s} , \bar{s} , etc. all belong to the dual s -space \mathbb{E}^* ; similarly, x and all of its decorated versions belong to the x -space \mathbb{E} .

We also denote $\tilde{\mu} := \tilde{\mu}(x, s) := \langle \tilde{s}, \tilde{x} \rangle / \vartheta$. In this context, \tilde{x} is the *primal shadow of the dual solution* s . Analogous to the definition of v , for every T obtained from a $T^2 \in \mathcal{T}_1^2(x, s)$, we can now define

$$w := T\tilde{s} = T^{-1}\tilde{x}.$$

Note that $w^\top w = \vartheta\tilde{\mu}$.

Once we have such a transformation T , we can map the primal space using T^{-1} , and the dual space using T to arrive at the data of the underlying (equivalent) *scaled* primal-dual pair. We define $\bar{\mathcal{A}} := \mathcal{A}(T(\cdot))$ and, we have

$$(\bar{P}) \quad \inf \quad (Tc)^\top \bar{x} \\ \bar{\mathcal{A}}\bar{x} = b, \\ \bar{x} \in T^{-1}K,$$

$$(\bar{D}) \quad \sup \quad \langle b, y \rangle_D \\ \bar{\mathcal{A}}^*y + \bar{s} = Tc, \\ \bar{s} \in TK^*.$$

Then the search directions for the scaled primal-dual pair (\bar{P}) and (\bar{D}) are (respectively)

$$\bar{d}_x := T^{-1}d_x, \quad \bar{d}_s := Td_s,$$

where d_x, d_s denote the search directions in the original primal and dual spaces respectively.

Lemma 2.8. (Nesterov and Todd [31], also see [47, 48]) For every $(x, s) \in \text{int}(K) \oplus \text{int}(K^*)$,

$$\frac{\langle s, x \rangle}{\vartheta} \frac{\langle F'(x), F'_*(s) \rangle}{\vartheta} = \mu\tilde{\mu} \geq 1.$$

Equality holds above iff $x = -\mu F'_*(s)$ (and hence $s = -\mu F'(x)$).

Note that the equality above together with feasibility of x and s in their respective problems, define the *primal-dual central path*:

$$\{(x(\mu), s(\mu)) : \mu > 0\}.$$

Further note that for each $T \in \mathcal{T}_1$, we can equivalently express the centrality condition as $v = \mu w$ (of course, together with the requirement that the underlying x and s be feasible in their respective problems). Primal and dual deviations from the central path are:

$$\delta_P := x - \mu \tilde{x} \text{ and } \delta_D := s - \mu \tilde{s}.$$

In v -space, these deviations are both represented by

$$\delta_v := v - \mu w.$$

Corollary 2.9. *For every $(x, s) \in \text{int}(K) \oplus \text{int}(K^*)$,*

$$\langle s + \mu F'(x), x + \mu F'_*(s) \rangle = \langle \delta_D, \delta_P \rangle = \|\delta_v\|_2^2 \geq 0.$$

The equality holds above iff $x = -\mu F'_(s)$ (and hence $s = -\mu F'(x)$).*

The above quantities have their counterpart, called the *gradient proximity measure* and denoted by $\gamma_G(x, s)$, in the work of Nesterov and Todd [31]. In their paper, the corresponding measure is

$$\gamma_G(x, s) = \vartheta(\mu \tilde{\mu} - 1) = \frac{\|\delta_v\|_2^2}{\mu}.$$

Now, we consider the system of linear equations (we utilized Newton's method, to follow the central path, approximately):

$$(3) \quad \bar{A} \bar{d}_x = 0$$

$$(4) \quad \bar{A}^* \bar{d}_y + \bar{d}_s = 0.$$

$$(5) \quad \bar{d}_x + \bar{d}_s = -v + \gamma \mu w,$$

where $\gamma \in [0, 1]$ is a centering parameter.

Clearly, (\bar{d}_x, \bar{d}_s) solves the above linear system of equations iff \bar{d}_x and \bar{d}_s are the orthogonal projections of $(-v + \gamma \mu w)$ onto the kernel of \bar{A} and the image of \bar{A}^* respectively. Given a pair of search directions, we define

$$x(\alpha) := x + \alpha \bar{d}_x, \quad s(\alpha) := s + \alpha \bar{d}_s.$$

Using the above observations, the following result is immediate.

Lemma 2.10. *Let $\gamma \in [0, 1]$. Then*

$$\langle s(\alpha), x(\alpha) \rangle = [1 - \alpha(1 - \gamma)] \langle s, x \rangle.$$

For every $s \in \text{int}(K^*)$, we define the operator norm

$$\|M\|_s = \sup_{\|u\|_s \leq 1} \|Mu\|_s^*.$$

Then, by definition, $\|Mu\|_s^* \leq \|M\|_s \|u\|_s$. We will also utilize bilinear forms defined via tensors. Just like the Hessians and bilinear forms T^2 above, these new bilinear forms will also

be interpreted as linear transformations between two spaces: For $u \in \mathbb{E}_1$, $z \in \mathbb{E}_2$, we define $u \otimes z : \mathbb{E}_2^* \rightarrow \mathbb{E}_1$ by

$$(u \otimes z)h := \langle z, h \rangle u, \quad \forall h \in \mathbb{E}_2^*,$$

as well as a bilinear form, $u \otimes z : \mathbb{E}_2 \oplus \mathbb{E}_1^* \rightarrow \mathbb{R}$. This bilinear form acts as

$$(u \otimes z)[h_1, h_2] := \langle u, h_1 \rangle \langle z, h_2 \rangle, \quad \forall h_1 \in \mathbb{E}_1^*, h_2 \in \mathbb{E}_2^*.$$

Next, let us observe that for every pair $h, u \in \mathbb{E}$, we have

$$2(h \otimes h - u \otimes u) = [(h - u) \otimes (h + u)] + [(h + u) \otimes (h - u)].$$

We make considerable use of the following difference-of-squares bound for the operator norm of the difference of two tensors:

Lemma 2.11. *Let h and u lie in \mathbb{E} and $s \in \text{int}(K^*)$. Then*

$$\|(h \otimes h) - (u \otimes u)\|_s \leq \|h - u\|_s^* \|h + u\|_s^*.$$

Proof. By the triangle inequality,

$$\begin{aligned} 2\|(h \otimes h) - (u \otimes u)\|_s &= \|[(h - u) \otimes (h + u)] + [(h + u) \otimes (h - u)]\|_s \\ &\leq \|(h - u) \otimes (h + u)\|_s + \|(h + u) \otimes (h - u)\|_s. \end{aligned}$$

Now, we compute

$$\begin{aligned} \|(h - u) \otimes (h + u)\|_s &= \sup_{\|z\|_s \leq 1} \|(h - u) \langle z, h + u \rangle\|_s^* \\ &= \sup_{\|z\|_s \leq 1} \langle z, h + u \rangle \|h - u\|_s^* = \|h + u\|_s^* \|h - u\|_s^*; \end{aligned}$$

and similarly

$$\|(h + u) \otimes (h - u)\|_s = \|h + u\|_s^* \|h - u\|_s^*.$$

Adding these together gives the advertised result. \square

The next lemma is used many times in the following analysis.

Lemma 2.12. *Let $s \in \text{int}(K^*)$, $h \in \mathbb{E}^*$ such that $\|h\|_s < 1$. Then,*

$$(1 - \|h\|_s)^2 \leq \sup_{u \in \mathbb{E}^*: \|u\|_s \leq 1} \|F_*''(s + h)u\|_s^* \leq \frac{1}{(1 - \|h\|_s)^2}.$$

Proof. Let s and h be as in the statement of the lemma. Then $F_*''(s + h) \preceq \frac{1}{(1 - \|h\|_s)^2} F_*''(s)$ by the Dikin ellipsoid bound. Thus, the maximum eigenvalue of $[F_*''(s)]^{-1/2} F_*''(s + h) [F_*''(s)]^{-1/2}$ is bounded above by $\frac{1}{(1 - \|h\|_s)^2}$. The square of this quantity is an upper bound on the largest eigenvalue of $[F_*''(s)]^{-1/2} F_*''(s + h) [F_*''(s)]^{-1} F_*''(s + h) [F_*''(s)]^{-1/2}$. Therefore, the supremum

in the statement of the lemma is bounded above by $\frac{1}{(1-\|h\|_s)^2}$ as desired. The left-hand side inequality can be proved similarly. \square

Now, we are ready to define the set of local primal-dual metrics which will utilize local second-order information from the underlying pair of self-concordant barriers. For each F and each pair $(x, s) \in \text{int}(K) \oplus \text{int}(K^*)$, consider the optimization problem with variables $T^2 \in \mathbb{S}_{++}(\mathbb{E}^*, \mathbb{E}^*)$ and $\xi \in \mathbb{R}$:

$$(PD)^2 \quad \inf \quad \begin{array}{l} \xi \\ T^2 s \quad = \quad x, \\ T^2 [-F'(x)] \quad = \quad -F'_*(s), \\ \frac{\mu}{\xi^{[\vartheta(\mu\tilde{\mu}-1)+1]}} F''_*(s) \leq T^2 \leq \frac{\xi^{[\vartheta(\mu\tilde{\mu}-1)+1]}}{T^2} [F''(x)]^{-1}, \\ \xi \geq 1, \\ T^2 \in \mathbb{S}_{++}(\mathbb{E}^*, \mathbb{E}^*). \end{array}$$

This problem is indeed an SDP (a SemiDefinite Programming problem in variables T^2 and ξ). Let ξ^* be the optimum value of $(PD)^2$.

Definition 2.13. *For every pair $(x, s) \in \text{int}(K) \oplus \text{int}(K^*)$, we define*

$$\mathcal{T}_2^2(\eta; x, s) := \left\{ H \in \mathbb{S}_{++}(\mathbb{E}^*, \mathbb{E}^*) : \begin{array}{l} Hs = x, \\ H [-F'(x)] = -F'_*(s), \\ \frac{\mu}{\eta \xi^{[\vartheta(\mu\tilde{\mu}-1)+1]}} F''_*(s) \leq H \leq \frac{\eta \xi^{[\vartheta(\mu\tilde{\mu}-1)+1]}}{\mu} [F''(x)]^{-1} \end{array} \right\},$$

for $\eta \geq 1$.

Sometimes we find it convenient to refer to T directly. So, we define $\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2$ as the set of $T \in \mathbb{S}_{++}^n$ whose square, T^2 lie in $\mathcal{T}_0^2, \mathcal{T}_1^2, \mathcal{T}_2^2$ respectively.

When the underlying cone K is symmetric, we have (P) and (D) as symmetric cone programming problems. Symmetric cones (are homogeneous and self-dual, and) admit self-scaled barriers. For such problems, there is a very small upper bound on ξ^* for every pair of interior-points. The following theorem follows mainly from Theorem 5.2 and Corollary 4.1 of Nesterov and Todd [30].

Theorem 2.14. *(see [47]) Let K be a symmetric cone and F be a self-scaled barrier for K . Then, for every $(x, s) \in \text{int}(K) \oplus \text{int}(K^*)$, $\xi^* \leq 4/3$.*

Being able to establish a nice upper bound on ξ^* for all iterates, and then, for some small $\eta \geq 1$, finding an efficient way of computing an element of $\mathcal{T}_2^2(\eta; x, s)$, directly yield primal-dual interior-point algorithms with good properties. In particular, if the upper bounds on ξ^* and η are both $O(1)$, then such results directly yield primal-dual interior-point algorithms with iteration complexity bound $O\left(\sqrt{\vartheta} \ln(1/\epsilon)\right)$; see, [47]. Next, we consider various ways of constructing good symmetric positive-definite bilinear forms in \mathcal{T}_1^2 and \mathcal{T}_2^2 .

3. PRIMAL-DUAL METRICS VIA HESSIAN INTEGRATION

In the special case where cone K is symmetric and F is a self-scaled barrier for K , Nesterov and Todd [30, 31] identified a specific element of the sets \mathcal{T}_0 , \mathcal{T}_1 and \mathcal{T}_2 in terms of the Hessians of certain *scaling points* (i.e., for every pair (x, s) , there exists $w \in \text{int}(K^*)$ such that $T^2 = F_*''(w)$). There is also an explicit linear algebraic formula which expresses the Hessian at w in terms of the Hessians of F and the conjugate barrier F_* at x and s respectively. The next theorem achieves an analogous goal in the fully general case of an arbitrary convex cone K and an arbitrary self-concordant barrier F for K by expressing a primal-dual scaling in \mathcal{T}_1^2 as an integral of Hessians along the line segment joining the dual iterate to the dual shadow of the primal iterate. Later in Section 5, we prove that beyond symmetric cones and self-scaled barriers, one should *not* expect in general to find, for every pair of primal dual interior points (x, s) , a $w \in \text{int}(K^*)$ such that $F_*''(w) \in \mathcal{T}_1^2(x, s)$. Perhaps surprisingly, we prove next that “an average of the Hessians” works!

Theorem 3.1. *Let F be a LHSCB for K and $(x, s) \in \text{int}(K) \oplus \text{int}(K^*)$. Then, the linear transformation*

$$T_D^2 := \mu \int_0^1 F_*''(s - t\delta_D) dt$$

maps s to x , and maps \tilde{s} to \tilde{x} , and $T_D^2[\cdot, \cdot]$ is a symmetric positive-definite bilinear form. Therefore, T_D^2 is in $\mathcal{T}_1^2(x, s)$.

Proof. Using the fundamental theorem of calculus (for the second equation below) followed by the property $-F_*'(-F'(x)) = x$ (for the third equation below), we obtain

$$T_D^2 \delta_D = \mu \int_0^1 F_*''(s - t\delta_D) \delta_D dt = \mu [-F_*'(s - \delta_D) + F_*'(s)] = \mu (x/\mu - \tilde{x}) = \delta_P.$$

We next compute, using the substitution $\tilde{t} = 1/t$,

$$\begin{aligned} T_D^2 s &= \mu \int_0^1 F_*''(s - t\delta_D) s dt \\ &= \mu \int_0^1 \frac{1}{t^2} F_*''(s/t - \delta_D) s dt \\ &= \mu \int_1^\infty F_*''(\tilde{t}s - \delta_D) s d\tilde{t} \\ &= -\mu F_*'(s - \delta_D) = x. \end{aligned}$$

Further, T_D^2 is the mean of some symmetric positive-definite bilinear forms, so T_D^2 itself is a symmetric positive-definite bilinear form. \square

We call this scaling operator the *dual integral scaling*. Note that the above theorem holds under weaker assumptions (we only used the facts that F is logarithmically homogeneous and

F, F_* are Legendre-type functions in the sense of Rockafellar [41]). The dual integral scaling is expected to inherit many of the nice properties of the Hessians. Thus, if F_* is well-behaved, then one can prove nice bounds on the deviation of dual integral scaling from the dual Hessian at s and $\mu\tilde{s}$:

Theorem 3.2. *If $\sigma < 1$ is such that, for any $t \in [0, 1]$,*

$$(1 - t\sigma)^2 F_*''(s) \preceq F_*''(s - t\delta_D) \preceq \frac{1}{(1 - t\sigma)^2} F_*''(s),$$

then

$$(1 - \sigma)\mu F_*''(s) \preceq T_D^2 \preceq \frac{1}{1 - \sigma}\mu F_*''(s).$$

Proof. This follows directly from the definition of T_D^2 . □

Interestingly, the dual integral scaling (the mean of Hessians along the line segment joining s and $\mu\tilde{s}$) is not as “canonical” as the Nesterov–Todd scaling (the geodesic mean of the Hessians joining the same two points in the interior of K^*) in terms of primal-dual symmetry properties. For the remainder of this section, we elaborate on this and related issues. Also, in Section 8 when we specialize on *Hyperbolicity Cone Programming* problems, we show that the integral scaling can have advantages (when one of the primal and dual problems is *more tractable* or *nicer* for the approach at hand, as a benefit of breaking this primal-dual symmetry well, and properties like those given by Theorem 3.2).

Notice that the dual integral scaling

$$\int_0^1 \mu F_*''(ts + (1 - t)\mu\tilde{s}) dt$$

and the primal integral scaling

$$T_P^2 := \left(\int_0^1 \mu F''(tx + (1 - t)\mu\tilde{x}) dt \right)^{-1}$$

are both scalings that map s to x and \tilde{s} to \tilde{x} . These are not in general the same, though they do coincide with the usual scaling $\text{Diag}(x) [\text{Diag}(s)]^{-1}$ in the case of Linear Programming.

Example 3.3. (A comparison of primal-dual local metrics for the positive semidefinite cone) We work out the integral scaling for the positive semidefinite cone \mathbb{S}_+^n . If X is the primal iterate and \tilde{X} is the primal shadow of the dual iterate, then we see that

$$(T_P^2)^{-1} [H, H] = \mu \int_0^1 \left\langle (tX + (1 - t)\mu\tilde{X})^{-1} H (tX + (1 - t)\mu\tilde{X})^{-1}, H \right\rangle dt.$$

One can make this slightly more explicit. There always exists a $U \in GL(n)$ such that $UXU^\top = I$ and $U\mu\tilde{X}U^\top$ is diagonal; one can compose a Q that orthogonally diagonalises X , an S that

scales $Q^\top X Q$ to the identity matrix, and a Q' that orthogonally diagonalises $SQ^\top \mu \tilde{X} Q S$. Say $U \mu \tilde{X} U^\top = D$. Then we can compute

$$T_D^2[UHU^\top, UHU^\top] = \mu \int_0^1 \langle (tI + (1-t)D)^{-1} H (tI + (1-t)D)^{-1}, H \rangle dt.$$

In particular, if H is E_{ij} , we have

$$T_D^2[UE_{ij}U^\top, UE_{ij}U^\top] = \mu \int_0^1 (t + (1-t)D_i)^{-1} (t + (1-t)D_j)^{-1} dt = \mu \frac{\ln D_j - \ln D_i}{D_j - D_i}.$$

Special attention needs to be given to the case when $D_i = D_j$; here, the integral evaluates to μ/D_i .

If $H = E_{ij} + E_{kl}$ (with $(i, j) \neq (k, l)$), we have $T_D^2[U(E_{ij} + E_{kl})U^\top, U(E_{ij} + E_{kl})U^\top]$ given by

$$\mu \int_0^1 (t + (1-t)D_i)^{-1} (t + (1-t)D_j)^{-1} + (t + (1-t)D_k)^{-1} (t + (1-t)D_l)^{-1} dt.$$

This is the sum of $T_D^2[UE_{ij}U^\top, UE_{ij}U^\top]$ with $T_D^2[UE_{kl}U^\top, UE_{kl}U^\top]$, meaning that

$$T_D^2[UE_{ij}U^\top, UE_{kl}U^\top] = \delta_{ik} \delta_{jl} T_D^2[UE_{ij}U^\top, UE_{ij}U^\top].$$

Put another way, the operator $C_{U^\top} T_D^2 C_U$ is diagonal where C_U is the conjugation operator given by $C_U(Z) = UZU^\top$. For every nonsingular U , the map C_U preserves operator geometric means; that is,

$$\begin{aligned} UA^{1/2} (A^{-1/2} B A^{-1/2})^{1/2} A^{1/2} U^\top \\ = (UAU^\top)^{1/2} ((UAU^\top)^{-1/2} U B U^\top (UAU^\top)^{-1/2})^{1/2} (UAU^\top)^{1/2}. \end{aligned}$$

One can show this as follows: The geometric mean of $X \succ 0$ and $Y \succ 0$ is the unique positive-definite $G := X^{1/2} (X^{1/2} Y^{-1} X^{1/2})^{-1/2} X^{1/2}$ such that $GX^{-1}G = Y$. Taking $H = UGU^\top$, we have

$$HU^{-\top} X^{-1} U^{-1} H = UGU^\top U^{-\top} X^{-1} U^{-1} UGU^\top = UGX^{-1}GU^\top = UYU^\top.$$

Interestingly, Molnár [16] proved that *every* linear automorphism of the semidefinite cone over a complex Hilbert space that preserves geometric means is a conjugation operator. The converse is also true, since the set of automorphisms of \mathbb{S}_+^n is the set of conjugations given by the nonsingular U (see a discussion in [46] of Güler's proof utilizing the proof technique of Waterhouse [49]), and as we observed above, it is easy to verify that the conjugation operator preserves operator geometric means. Thus, we can make a natural comparison with the Nesterov–Todd scaling given by

$$N[H, H] = \langle (VD^{1/2}V^\top)^{-1} H (VD^{1/2}V^\top)^{-1}, H \rangle,$$

where $VD'V^\top$ is the spectral decomposition of operator geometric mean of X and $\mu\tilde{X}$. Notice that

$$N[VE_{ij}V^\top, VE_{ij}V^\top] = \langle D'^{-1/2}E_{ij}D'^{-1/2}, E_{ij} \rangle = \frac{1}{\sqrt{D'_i D'_j}}$$

and that N is similarly diagonal with respect to that basis. (The conjugation $C_U N C_U^\top$ does not in general result in a diagonal matrix, so we need to take this different V instead in order to diagonalise N .) Notice that

$$N[UE_{ij}U^\top, UE_{ij}U^\top] = e_i^\top U^\top G U e_i e_j^\top U^\top G U e_j$$

whatever U is. Thus, when we form the matrix whose ij entry is $N[UE_{ij}U^\top, UE_{ij}U^\top]$, we always obtain a rank-one matrix. However, such a matrix formed from the integral scaling T_D^2 can have full rank. We proceed with an example. Consider $D = (\epsilon, \epsilon^2, \dots, \epsilon^n)$. Notice that $\ln D_i - \ln D_j = (i - j) \ln \epsilon$, while $D_i - D_j = \epsilon^i - \epsilon^j$. Thus, the (i, j) entry of this matrix is on the order of $\epsilon^{-\min(i, j)}$. For sufficiently small ϵ , then, the determinant of this matrix is dominated by the product along the diagonal, which is positive—it follows that this matrix is nonsingular.

4. LOCAL PRIMAL-DUAL METRICS EXPRESSED AS LOW RANK UPDATES

It may be impractical in many cases to evaluate the integral scaling from Section 2 exactly or to high enough accuracy. Due to this, or perhaps due to numerical instabilities arising in computations, we may have to make do with an approximation H that does not necessarily satisfy the equations $HS = x$ and $H\tilde{s} = \tilde{x}$.

The second author [47] constructed the following low-rank update which “fixes” such problems encountered by *any* symmetric, positive-definite H :

$$(6) \quad T_H^2 := H + a_1 x \otimes x + g_1 Hs \otimes Hs + \tilde{a}_1 \tilde{x} \otimes \tilde{x} + \tilde{g}_1 H\tilde{s} \otimes H\tilde{s} + a_2 [(x \otimes \tilde{x}) + (\tilde{x} \otimes x)] \\ + g_2 [(Hs \otimes H\tilde{s}) + (H\tilde{s} \otimes Hs)],$$

where

$$a_1 = \frac{\tilde{\mu}}{\vartheta(\mu\tilde{\mu} - 1)}, \tilde{a}_1 = \frac{\mu}{\vartheta(\mu\tilde{\mu} - 1)}, a_2 = \frac{-1}{\vartheta(\mu\tilde{\mu} - 1)}, \\ g_1 = \frac{H[\tilde{s}, \tilde{s}]}{H[s, s]H[\tilde{s}, \tilde{s}] - (H[\tilde{s}, s])^2}, \tilde{g}_1 = \frac{H[s, s]}{H[s, s]H[\tilde{s}, \tilde{s}] - (H[\tilde{s}, s])^2}, g_2 = \frac{H[s, \tilde{s}]}{H[s, s]H[\tilde{s}, \tilde{s}] - (H[\tilde{s}, s])^2}.$$

The second author [47] proved that, as long as H is positive-definite, T_H^2 is positive-definite, maps s to x , and maps \tilde{s} to \tilde{x} .

As written, Equation (6) is somewhat unwieldy for purposes of analysis. It is not immediately obvious that, in the case that the pair (x, s) satisfy the centrality condition $x = \mu\tilde{x}$ (or, equivalently, $\mu\tilde{\mu} = 1$), the formula collapses to a rank-two update. (Indeed, it has a singularity there.) We prove that the above given complicated operator has an equivalent form as two, simple, consecutive updates due to the special structure of our set-up:

Theorem 4.1. *Let $x, \tilde{x} \in \mathbb{E}$ and $s, \tilde{s} \in \mathbb{E}^*$ satisfy the following conditions (under the definitions $\delta_P := x - \mu\tilde{x}$ and $\delta_D := s - \mu\tilde{s}$):*

- $0 < \langle \tilde{s}, x \rangle = \langle s, \tilde{x} \rangle =: \vartheta$
- $0 < \langle s, x \rangle =: \vartheta\mu$
- $0 < \langle \tilde{s}, \tilde{x} \rangle =: \vartheta\tilde{\mu}$
- $\langle \delta_D, \delta_P \rangle > 0$.

Further let $H \in \mathbb{S}_{++}(\mathbb{E}^, \mathbb{E}^*)$. Then, H_2 in the following formula is a symmetric, positive-definite bilinear form and it maps s to x , and maps \tilde{s} to \tilde{x} :*

$$(7) \quad \begin{aligned} H_1 &:= H + \frac{1}{\langle s, x \rangle} x \otimes x - \frac{1}{\langle s, Hs \rangle} Hs \otimes Hs \\ H_2 &:= H_1 + \frac{1}{\langle \delta_D, \delta_P \rangle} \delta_P \otimes \delta_P - \frac{1}{\langle \delta_D, H_1 \delta_D \rangle} H_1 \delta_D \otimes H_1 \delta_D. \end{aligned}$$

Proof. Notice that $H_1 s = Hs + x - Hs = x$. Notice further that $\langle s, \delta_P \rangle = \langle \delta_D, x \rangle = 0$ by expanding the dual pairing conditions, so $H_2 s = H_1 s$. Thus, H_2 maps s to x . Next, note that $H_2 \delta_D = H_1 \delta_D + \delta_P - H_1 \delta_D = \delta_P$. Thus, H_2 also maps δ_D to δ_P . Hence, H_2 maps \tilde{s} to \tilde{x} . Clearly H_2 is a linear combination of symmetric bilinear forms and hence itself is a symmetric bilinear form. We recall from the theory of quasi-Newton updates (see for instance Lemma 9.2.1 in [4]) that the ‘‘curvature condition’’ $\langle s, x \rangle > 0$ is necessary and sufficient to guarantee that H_1 is positive-definite, and, given that H_1 is positive-definite, the curvature condition $\langle \delta_D, \delta_P \rangle > 0$ is necessary and sufficient for H_2 to be positive-definite. Therefore, H_2 is positive-definite as well. \square

Note that the positivity of the scalar products $\langle s, x \rangle$ and $\langle \delta_D, \delta_P \rangle$, together with the orthogonality conditions $\langle s, \delta_P \rangle = \langle \delta_D, x \rangle = 0$ suffice for the above theorem to hold. Thus, there may be a potential use of these formulae in classical quasi-Newton approaches. Such considerations are left for future work.

We remark that we can apply the above formulas after switching x and s and then inverting the resulting T^2 to obtain the following low-rank updates (under the same conditions):

Theorem 4.2. *Let H, x, \tilde{x}, s , and \tilde{s} be as in Theorem 4.1. Then H_2 in the following formula is symmetric, positive-definite, maps s to x , and maps \tilde{s} to \tilde{x} :*

$$(8) \quad \begin{aligned} H_1 &:= \left(I - \frac{x \otimes s}{\langle s, x \rangle} \right) H \left(I - \frac{s \otimes x}{\langle s, x \rangle} \right) + \frac{x \otimes x}{\langle s, x \rangle} \\ H_2 &:= \left(I - \frac{\delta_P \otimes \delta_D}{\langle \delta_D, \delta_P \rangle} \right) H_1 \left(I - \frac{\delta_D \otimes \delta_P}{\langle \delta_D, \delta_P \rangle} \right) + \frac{\delta_P \otimes \delta_P}{\langle \delta_D, \delta_P \rangle}. \end{aligned}$$

Proof. We again see that $H_1s = x$ and $H_2\delta_D = \delta_P$ by correctness of the BFGS/DFP update. We compute

$$H_2s = \left(I - \frac{\delta_P \otimes \delta_D}{\langle \delta_D, \delta_P \rangle} \right) H_1 \left(I - \frac{\delta_D \otimes \delta_P}{\langle \delta_D, \delta_P \rangle} \right) s + \frac{\delta_P \otimes \delta_P}{\langle \delta_D, \delta_P \rangle} s = \left(I - \frac{\delta_P \otimes \delta_D}{\langle \delta_D, \delta_P \rangle} \right) x + 0 = x.$$

It is easy to verify that both H_1 and H_2 are symmetric bilinear forms. H_1 is positive-definite because the curvature condition $\langle s, x \rangle > 0$ is satisfied. H_2 is positive-definite because the curvature condition $\langle \delta_D, \delta_P \rangle > 0$ is satisfied. \square

Due to the variational interpretations of quasi-Newton update formulae, we have the corresponding interpretations in our set-up (i.e., H_1 is the closest—minimum distance—symmetric bilinear form to H satisfying $H_1s = x$; similarly for H_2 and H_1). Moreover, as in [47], the above formulae can be used in convex combinations (analogous to *Broyden's convex class* in the classical quasi-Newton context) due to convexity of \mathcal{T}_1 . Of course, we may also use the formula for H_1 in Theorem 4.1 together with the formula for H_2 in Theorem 4.2, etc. In Appendix B, we will look at these properties more deeply from a strict primal-dual symmetry viewpoint.

5. PRIMAL-DUAL HESSIAN LOCAL METRICS BASED ON A SINGLE SCALING POINT

In this section, we ask and partially answer the following question: For which ϑ -LHSCBs F_* does there exist a unique scaling point $w \in \text{int}(K^*)$ such that $F_*''(w) \in \mathcal{T}_1^2(x, s)$ for every $x \in \text{int}(K)$ and $s \in \text{int}(K^*)$?

Note that, on the one hand, for every ϑ -LHSCB F_* , there exists a scaling point $w \in \text{int}(K^*)$ such that $F_*''(w) \in \mathcal{T}_0^2(x, s)$ for every $x \in \text{int}(K)$ and $s \in \text{int}(K^*)$ (as it was already proved by Nesterov and Todd [30]; see [47], Theorem 3.1). On the other hand, the Nesterov–Todd scaling point given by the geodesic mean of s and $-F'(x)$ provides an example of such a w in the symmetric cone case (in this special case, $F_*''(w) \in \mathcal{T}_1^2(x, s)$). We show that this property does not generalise to slices of a symmetric cone that are not themselves symmetric cones. Every symmetric cone is a slice of a positive semidefinite cone. Therefore, for the sake of simplicity of notation we will stick with the slices of positive semidefinite cones in the following.

Lemma 5.1. *Let L be a linear subspace of \mathbb{S}^n that contains I but is not closed under matrix squaring. Then there exists a $B \in L$ such that $\text{Tr } B = 0$ and $B^2 \notin L$.*

Proof. Select a C such that $C^2 \notin L$. Let $t := \text{Tr } C/n$ and $B := C - tI$. Then, $\text{Tr } B = 0$ and $B^2 = C^2 - 2tC + t^2I$. The latter two terms of this expansion lie in L while C^2 does not, so $B^2 \notin L$. \square

Given a linear subspace L , let Π_L denote the orthogonal projection onto L .

Proposition 5.2. *Let L be a linear subspace of \mathbb{S}^n that contains I but is not closed under matrix squaring. Choose the barrier $F(X) = -\ln \det X$ for the cone $K := \mathbb{S}_+^n \cap L$. (The dual cone of K*

is $K^* = \{S \in L : \forall X \in K, \text{Tr}(SX) \geq 0\} = \Pi_L(\mathbb{S}_+^n)$. There exist $X \in \text{int}(K)$ and $S \in \text{int}(K^*)$ such that, for all $W \in \text{int}(K)$, either $F''(W)[X] \neq S$ or $F''(W)[-F'_*(S)] \neq -F'(X)$.

Proof. For the claim about K^* , note that

$$K^* = (\mathbb{S}_+^n \cap L)^* \cap L = (\mathbb{S}_+^n + L^\perp) \cap L = \Pi_L(\mathbb{S}_+^n),$$

where the second equation above uses the facts that $L \cap \mathbb{S}_{++}^n \neq \emptyset$, $(\mathbb{S}_+^n)^* = \mathbb{S}_+^n$ and $L^* = L^\perp$. Assume for the purpose of contradiction that there are n and L such that, for every choice of $X \in \text{int}(K)$ and $S \in \text{int}(K^*)$, there exists a $W \in \text{int}(K)$ for which $F''(W)[X] = S$ and $F''(W)[-F'_*(S)] = -F'(X)$.

For $Z \in \text{int}(K)$, we compute (e.g., by using Taylor's formula for $F(x + td)$, for $x \in \text{int}(K)$, $d \in L$)

$$F'(Z) = -\Pi_L(Z^{-1})$$

and

$$F''(Z)[H] = \lim_{h \rightarrow 0^+} -\frac{1}{h} \Pi_L((Z + hH)^{-1} - Z^{-1}) = \Pi_L(Z^{-1}HZ^{-1}).$$

Select $B \in L$ such that $\text{Tr} B = 0$ and $B^2 \notin L$. Let $S := X := I + \epsilon B$; we shall choose $\epsilon > 0$ later. (For sufficiently small ϵ , X lies in the interior of K and S in the interior of K^* .) Notice that this choice of S and X implies that $S = F''(W)[X] = \Pi_L(W^{-1}XW^{-1})$.

Next, we check that $W = I$ is the only solution to $S = F''(W)[X]$. Suppose W is such that $S = F''(W)[X]$. Consider the problem

$$\begin{aligned} \min \quad & \text{Tr}(Z^{-1}X) \\ \text{subject to} \quad & \text{Tr}(ZX) \leq \text{Tr}(WX) \\ & Z \in L \\ & Z \succeq 0. \end{aligned}$$

Notice that the objective is strictly convex (this is indeed related to the fact that the long-step Hessian estimation property holds for $-\ln \det(\cdot)$) and that the feasible region is nonempty, compact and convex (since X is positive-definite). Thus this problem has a unique optimal solution. This optimal solution does not lie in the boundary of the positive semidefinite cone since the objective is $+\infty$ there and finite elsewhere. Thus, we may drop the constraint $Z \succeq 0$. $Z := W/2$ is a feasible solution satisfying the positive semidefiniteness constraint and the linear inequality strictly, so Slater's condition holds. Notice that the gradient of the objective is $-Z^{-1}XZ^{-1}$.

By the Karush-Kuhn-Tucker theorem applied to the problem without the explicit constraint $Z \succeq 0$, every point Z for which $\text{Tr}(ZX) = \text{Tr}(WX)$ and there exists a $\lambda \geq 0$ and $\ell^\perp \in L^\perp$ such that $Z^{-1}XZ^{-1} - \lambda X + \ell^\perp = 0$ is optimal. Note that $Z = W$ is a solution with $\lambda = 1$. Note that $Z := \text{Tr}(WX)I / \text{Tr} X$ is also a solution with $\lambda = [\text{Tr}(X) / \text{Tr}(WX)]^2$. Thus W must be a scalar multiple of I ; since $\Pi_L(W^{-1}XW^{-1}) = X$, that scalar must be 1.

Define $\tilde{X} := -F'_*(S)$. Then, we must have $F''(W)[\tilde{X}] = -F'(X)$; $\tilde{X} = \Pi_L(X^{-1}) = X^{-1} - P$ for some $P \in L^\perp$. Observe that, for sufficiently small ϵ ,

$$X^{-1} = I - \epsilon B + \epsilon^2 B^2 + O(\epsilon^3).$$

Both I and B lie in L , so

$$P = \Pi_{L^\perp}(\epsilon^2 B^2) + O(\epsilon^3).$$

We compute, for sufficiently small ϵ ,

$$\begin{aligned} n &= \operatorname{Tr} S = \operatorname{Tr} \left[-F'(\tilde{X}) \right] = \operatorname{Tr} \left[(X^{-1} - P)^{-1} \right] = \operatorname{Tr} \left[X + XPX + XPXPX + O(\epsilon^6) \right] \\ &= \operatorname{Tr} \left[I + \epsilon B + P + \epsilon BP + \epsilon PB + \epsilon^2 BPB + P^2 + O(\epsilon^5) \right] \\ &= n + 0 + 0 + 0 + 0 + \epsilon^2 \operatorname{Tr}(B^2 P) + \operatorname{Tr}(P^2) + O(\epsilon^5). \end{aligned}$$

For the last equation above, we used the fact that $\operatorname{Tr}(P) = 0$ (since $P \in L^\perp$ and $I \in L$). Noting that $P = \Pi_{L^\perp}(\epsilon^2 B^2) + O(\epsilon^3)$, we see that $\epsilon^2 \operatorname{Tr}(B^2 P) = \operatorname{Tr}(P^2)$, which is the squared Frobenius norm of P . By our choice of B , this is positive. Thus, for sufficiently small ϵ , $\operatorname{Tr} X = \operatorname{Tr} S = \operatorname{Tr} \left[-F'(\tilde{X}) \right] > \operatorname{Tr} X$, a contradiction. \square

There is a five-dimensional semidefinite cone slice where a \mathcal{T}_1 scaling defined by a single point does not exist for all x and s , namely the cone of symmetric, positive semidefinite matrices with the sparsity pattern

$$\begin{pmatrix} * & * & * \\ * & * & 0 \\ * & 0 & * \end{pmatrix}.$$

This cone (described above as a slice of 3-by-3 positive semidefinite cone) is also known as the *Vinberg cone*. (Here we are working with its SDP representation). It is the smallest-dimensional homogeneous cone that is not self-dual, whence not a symmetric cone. This particular subspace L contains I , and it is not closed under matrix squaring.

Let us note that if L is a linear subspace of \mathbb{S}^n that contains I and is closed under matrix squaring, then

$$\forall U, V \in L, \quad UV + VU = (U + V)^2 - U^2 - V^2 \in L.$$

Moreover, since for every pair $U, V \in L$, the equation

$$U(U^2 V + VU^2) + (U^2 V + VU^2)U = U^2(UV + VU) + (UV + VU)U^2$$

is self-evident, we have an Euclidean Jordan algebra over L . Hence, $(\mathbb{S}_+^n \cap L)$ is an SDP representation of a symmetric cone and indeed the function $-\ln \det(\cdot) : \operatorname{int}(K) \rightarrow \mathbb{R}$ is a self-scaled barrier for $(\mathbb{S}_+^n \cap L)$. Therefore, Proposition 5.2 proves that among all SDP-representable cones (as a slice of \mathbb{S}_+^n), symmetric cones are the only ones for which

$$\{F''_*(w) : w \in \operatorname{int}(K^*)\} \cap \mathcal{T}_1^2(x, s) \neq \emptyset, \quad \forall (x, s) \in \operatorname{int}(K) \oplus \operatorname{int}(K^*),$$

where $F(X) := -\ln \det(X)$ (over the representing cone \mathbb{S}_+^n).

The proof technique of Proposition 5.2 is more generally applicable. As a result, a more general and sharper characterization of the underlying behaviour is possible. This will be addressed, in detail, in future work.

6. THE NORM OF THE LOW-RANK UPDATES NEAR THE CENTRAL PATH

We know that if we can compute $T^2 \in \mathcal{T}_2^2(\eta; x, s)$ efficiently, with ensuring $\xi^* = O(1)$ and $\eta = O(1)$ for all iterates, then we will have one of the most important ingredients of a general family of primal-dual interior-point algorithms with iteration complexity $O\left(\sqrt{\vartheta} \ln(1/\epsilon)\right)$.

Up to this point, we have seen many ways of constructing $T^2 \in \mathcal{T}_1^2(x, s)$. However, we also discovered that we cannot expect to have a $z \in \text{int}(K^*)$ such that $F_*''(z) \in \mathcal{T}_1^2(x, s)$, in general.

We will later present and analyse an algorithm for convex programming based on Mizuno, Todd, and Ye's predictor-corrector approach [15]. We will first assume explicit access to oracles computing a ϑ -self-concordant primal barrier F for the primal cone and the conjugate barrier F_* for the dual cone. We will argue later that the algorithms can be modified to work without an explicit F_* oracle.

These oracles may be expensive; it may be unreasonable (or simply unnecessary) to try to compute the dual integral scaling directly to high precision at every iteration. We thus consider computing an approximation H to T_D^2 and then using a low-rank update to H to get a scaling in $T_H^2 \in \mathcal{T}_1(x, s)$. Specifically, we approximate the operator integral by evaluating it at the midpoint of the line segment joining s and $\mu\tilde{s}$, corresponding to the two extremes of the set of symmetric bilinear positive-definite forms $\{F_*''(s - t\delta_D) : t \in [0, 1]\}$. Then, we use the low-rank updates of Section 4 to restore membership in $\mathcal{T}_1^2(x, s)$.

Define

$$\check{s} := \frac{s + \mu\tilde{s}}{2} \text{ and } H := \mu F_*''(\check{s}),$$

and take H_1 and $T_H^2 := H_2$ as in (7).

Our analysis hinges on the resulting scaling (local metric T_H^2) being close to $\mu F_*''(s)$ and $[\mu F''(x)]^{-1}$ (in the sense of Definition 2.13) in every iteration of the algorithm. We therefore devote the remainder of this section to computing bounds, somehow dependent on the error in approximating T_D^2 by H , of the additional error introduced by the low-rank updates. We will prove in this section that for every pair of interior points $(x, s) \in \text{int}(K) \oplus \text{int}(K^*)$ for which x is close to $\mu\tilde{x}$, and hence s is close to $\mu\tilde{s}$ (in the sense that e.g., $\|\delta_D\|_s < 1/64$), ξ^* is $O(1)$ (in fact, less than $4/3$), and T_H^2 is a $4/3$ -approximate solution to the SDP defining $\mathcal{T}_2^2(1; x, s)$. We made a particular choice of $1/64$ for the neighbourhood parameter; indeed, a

continuous parametrization of the following analysis with respect to the neighbourhood parameter is possible (and implicit). More importantly, we chose a measure of proximity in the dual space for the analysis (rather than, for example, choosing a primal-dual symmetric measure of proximity). For the order of the iteration bounds this does not seem to make a difference, moreover, we wanted state at least some of the analysis in a way that would be useful in implementations of the algorithms in this framework that break the symmetry in favor of the dual problem.

For $x \in \text{int } K$, define $\|h\|_{K,x}$ to be the smallest positive λ such that $x + h/\lambda$ and $x - h/\lambda$ both lie in K . Then,

$$\|h\|_{K,x} = \min \{ \sigma_x(h), \sigma_x(-h) \}.$$

One can check that $\|\cdot\|_{K,x}$ is a norm. Let $\|h\|_{K,x}^*$ be its dual norm. If $H : \mathbb{E}^* \oplus \mathbb{E}^* \rightarrow \mathbb{R}$ is a symmetric bilinear form, define

$$\|H\|_{K,x} := \max_{\|u\|_{K,x}^* \leq 1} \max_{\|\ell\|_{K,x}^* \leq 1} H[u, \ell];$$

this is the operator norm induced by the norm $\|\cdot\|_{K,x}$.

Good bounds on the difference between two vectors in $\|\cdot\|_{K,x}$ imply some very useful conic inequalities:

Proposition 6.1. *If $\|u - \ell\|_{K,\ell} \leq \delta < 1$, then there are x and y in K such that $u = (1 - \delta)\ell + x = (1 + \delta)\ell - y$.*

Proof. Let $s \in K^*$. Then $\langle s, \ell \rangle \geq 0$ and $\langle s, \ell \pm (u - \ell)/\delta \rangle \geq 0$. Thus $\langle s, (\delta - 1)\ell + u \rangle \geq 0$ and $\langle s, (1 + \delta)\ell - u \rangle \geq 0$. This is true for every $s \in K^*$, so $(\delta - 1)\ell + u$ and $(1 + \delta)\ell - u$ both lie in K , as desired. \square

The operator norm still behaves as expected on rank-one matrices:

Proposition 6.2. *Let $u, \ell \in \mathbb{E}$. Then the operator norm of the bilinear form $u \otimes \ell$ is*

$$\|u \otimes \ell\|_{K,x} = \|u\|_{K,x} \|\ell\|_{K,x}.$$

Proof. We use the definitions, to deduce:

$$\|u \otimes \ell\|_{K,x} = \max_{\|y\|_{K,x}^* \leq 1} \max_{\|z\|_{K,x}^* \leq 1} \langle u, y \rangle \langle \ell, z \rangle = \left(\max_{\|y\|_{K,x}^* \leq 1} \langle u, y \rangle \right) \left(\max_{\|z\|_{K,x}^* \leq 1} \langle \ell, z \rangle \right) = \|u\|_{K,x} \|\ell\|_{K,x}.$$

\square

The difference-of-squares bound also still holds:

Proposition 6.3. *Let $u, \ell \in \mathbb{E}$. Then*

$$\|(u \otimes u) - (\ell \otimes \ell)\|_{K,x} \leq \|u - \ell\|_{K,x} \|u + \ell\|_{K,x}$$

Proof. We apply the definitions:

$$(u \otimes u) - (\ell \otimes \ell) = \frac{1}{2} [(u - \ell) \otimes (u + \ell)] + [(u + \ell) \otimes (u - \ell)].$$

Now, we apply the triangle inequality and the previous proposition. \square

We can control the $\|\cdot\|_{K,x}$ norm of the quasi-Newton updates from Section 4 in terms of the $\|\cdot\|_{K,x}$ norm of the correction to be made. The following theorem makes special use of the fact that s lies in K^* and x in K to get a strong bound:

Theorem 6.4. *Let $x \in \text{int}(K)$ and $s \in \text{int}(K^*)$. Let $H : \mathbb{E}^* \oplus \mathbb{E}^* \rightarrow \mathbb{R}$ be a symmetric bilinear positive-definite form. Suppose $\|Hs - x\|_{K,x} \leq \epsilon_1 < 1$. Then,*

$$\left\| \frac{Hs \otimes Hs}{\langle s, Hs \rangle} - \frac{x \otimes x}{\langle s, x \rangle} \right\|_{K,x} \leq \epsilon_1 \frac{3 + \epsilon_1}{1 - \epsilon_1}.$$

Proof. By the triangle inequality,

$$(9) \quad \left\| \frac{Hs \otimes Hs}{\langle s, Hs \rangle} - \frac{x \otimes x}{\langle s, x \rangle} \right\|_{K,x} \leq \left| \frac{\langle s, x - Hs \rangle}{\langle s, x \rangle \langle s, Hs \rangle} \right| \|Hs \otimes Hs\|_{K,x} + \frac{1}{\langle s, x \rangle} \|(x \otimes x) - (Hs \otimes Hs)\|_{K,x}.$$

We use Proposition 6.1 to write $Hs = (1 - \epsilon_1)x + y = (1 + \epsilon_1)x - z$ for y and z in K . Then, $\langle s, x - Hs \rangle = \langle s, \epsilon_1 x - y \rangle \leq \epsilon_1 \langle s, x \rangle$ and $\langle s, x - Hs \rangle = \langle s, z - \epsilon_1 x \rangle \geq -\epsilon_1 \langle s, x \rangle$. Thus,

$$|\langle s, x - Hs \rangle| \leq \epsilon_1 \langle s, x \rangle.$$

Hence,

$$\left| \frac{\langle s, x - Hs \rangle}{\langle s, x \rangle \langle s, Hs \rangle} \right| \leq \left| \frac{\epsilon_1 \langle s, x \rangle}{\langle s, x \rangle (\langle s, x \rangle + \langle s, Hs - x \rangle)} \right| \leq \frac{\epsilon_1}{(1 - \epsilon_1) \langle s, x \rangle}.$$

Since $\|Hs\|_{K,x} \leq \|x\|_{K,x} + \|Hs - x\|_{K,x} \leq 1 + \epsilon_1$, it follows that the first term on the right-hand side of (9) is bounded above by

$$\frac{\epsilon_1 \|Hs\|_{K,x}^2}{(1 - \epsilon_1) \langle s, x \rangle} \leq \frac{\epsilon_1 (1 + \epsilon_1)^2}{(1 - \epsilon_1) \langle s, x \rangle}.$$

The second term on the right-hand side of (9) can be bounded above by

$$\frac{(2 + \epsilon_1)\epsilon_1}{\langle s, x \rangle}.$$

This implies the desired bound. \square

The following theorem applies more generally, but the bound is weaker:

Theorem 6.5. *Let $x \in \text{int}(K)$. Further let $\delta_P \in \mathbb{E}$ and $\delta_D \in \mathbb{E}^*$ and H and $T^2 : \mathbb{E}^* \oplus \mathbb{E}^* \rightarrow \mathbb{R}$ be a pair of symmetric, positive-definite bilinear forms such that*

- $\langle \delta_D, \delta_P \rangle > 0$;
- $T^2 \delta_D = \delta_P$;
- there exists $\sigma < 1$ such that, for every $z \in \mathbb{E}^*$,

$$(1 - \sigma)^2 T^2[z, z] \leq H[z, z] \leq \frac{T^2[z, z]}{(1 - \sigma)^2};$$

- there exists $\epsilon_2 < 1$ such that $\|H\delta_D - \delta_P\|_{K,x} \leq \epsilon_2 \|H\delta_D\|_{K,x}$.

Then,

$$\left\| \frac{\delta_P \otimes \delta_P}{\langle \delta_D, \delta_P \rangle} - \frac{H\delta_D \otimes H\delta_D}{\langle \delta_D, H\delta_D \rangle} \right\|_{K,x} \leq \left(\frac{\sigma(2 + \sigma)}{1 - 4\sigma} + \epsilon_2(2 + \epsilon_2) \right) \frac{\|H\delta_D\|_{K,x}^2}{\langle \delta_D, \delta_P \rangle}.$$

Proof. Note that

$$\begin{aligned} & \left\| \frac{\delta_P \otimes \delta_P}{\langle \delta_D, \delta_P \rangle} - \frac{H\delta_D \otimes H\delta_D}{\langle \delta_D, H\delta_D \rangle} \right\|_{K,x} \\ & \leq \left| \frac{\langle \delta_D, H\delta_D - \delta_P \rangle}{\langle \delta_D, \delta_P \rangle \langle \delta_D, H\delta_D \rangle} \right| \|H\delta_D \otimes H\delta_D\|_{K,x} + \frac{1}{\langle \delta_D, \delta_P \rangle} \|(\delta_P \otimes \delta_P) - (H\delta_D \otimes H\delta_D)\| \\ & \leq \left| \frac{\langle \delta_D, H\delta_D - \delta_P \rangle}{\langle \delta_D, H\delta_D \rangle} \right| \frac{\|H\delta_D\|_{K,x}^2}{\langle \delta_D, \delta_P \rangle} + \frac{\epsilon_2(2 + \epsilon_2) \|H\delta_D\|_{K,x}^2}{\langle \delta_D, \delta_P \rangle}. \end{aligned}$$

We write $\langle \delta_D, H\delta_D - \delta_P \rangle = \langle \delta_D, (H - T^2)\delta_D \rangle$. Then,

$$((1 - \sigma)^2 - 1) \langle \delta_D, T^2 \delta_D \rangle \leq \langle \delta_D, (H - T^2)\delta_D \rangle \leq \left(\frac{1}{(1 - \sigma)^2} - 1 \right) \langle \delta_D, T^2 \delta_D \rangle.$$

Thus,

$$|\langle \delta_D, H\delta_D - \delta_P \rangle| \leq \frac{2\sigma + \sigma^2}{(1 - \sigma)^2} \langle \delta_D, \delta_P \rangle.$$

Writing $\langle \delta_D, H\delta_D \rangle = \langle \delta_D, \delta_P \rangle + \langle \delta_D, H\delta_D - \delta_P \rangle$, it follows that $\left| \frac{\langle \delta_D, H\delta_D - \delta_P \rangle}{\langle \delta_D, H\delta_D \rangle} \right| \leq \frac{2\sigma + \sigma^2}{1 - 4\sigma}$. Hence,

$$\left| \frac{\langle \delta_D, H\delta_D - \delta_P \rangle}{\langle \delta_D, \delta_P \rangle \langle \delta_D, H\delta_D \rangle} \right| \|H\delta_D \otimes H\delta_D\|_{K,x} \leq \frac{2\sigma + \sigma^2}{1 - 4\sigma} \frac{\|H\delta_D\|_{K,x}^2}{\langle \delta_D, \delta_P \rangle}.$$

The desired result follows. \square

Now that we have seen a high-level analysis on bounding the norms of low rank updates in a neighbourhood of the central path, we see that providing upper bounds on $\|Hs - x\|_s^*$, $\|H\delta_D - \delta_P\|_s^*$, and lower bounds on $\langle \delta_D, \delta_P \rangle$, $\langle \delta_D, H\delta_D \rangle$, and $\langle \delta_D, H_1\delta_D \rangle$ will be useful in the complexity analysis as well as in designing robust implementations of the algorithms. For the rest of the analysis, we will increase the use of explicit absolute constants, for the sake of concreteness. We either write these constants as ratios of two integers or as decimals which

represent rationals with denominator: 10^6 (whence, we are able to express the latter constants *exactly* as decimals with six digits after the point).

Theorem 6.6. *Assume $\|\delta_D\|_s \leq 1/64$, and take $H := \mu F''(\check{s})$ and*

$$H_1 := H + \frac{x \otimes x}{\langle s, x \rangle} - \frac{Hs \otimes Hs}{\langle s, Hs \rangle}.$$

Then,

$$\left\| \frac{\delta_P \otimes \delta_P}{\langle \delta_D, \delta_P \rangle} - \frac{H_1 \delta_D \otimes H_1 \delta_D}{\langle \delta_D, H_1 \delta_D \rangle} \right\|_s^* \leq 3.502346 \mu \|\delta_D\|_s.$$

Proof. We write

$$\begin{aligned} \frac{\delta_P \otimes \delta_P}{\langle \delta_D, \delta_P \rangle} - \frac{H_1 \delta_D \otimes H_1 \delta_D}{\langle \delta_D, H_1 \delta_D \rangle} &= \frac{1}{\langle \delta_D, \delta_P \rangle} [(\delta_P \otimes \delta_P) - (H_1 \delta_D \otimes H_1 \delta_D)] \\ &\quad + \left(\frac{1}{\langle \delta_D, H_1 \delta_D \rangle} - \frac{1}{\langle \delta_D, \delta_P \rangle} \right) H_1 \delta_D \otimes H_1 \delta_D. \end{aligned}$$

Notice that

$$\begin{aligned} \left| \frac{1}{\langle \delta_D, H_1 \delta_D \rangle} - \frac{1}{\langle \delta_D, \delta_P \rangle} \right| &= \left| \frac{\langle \delta_D, H_1 \delta_D - \delta_P \rangle}{\langle \delta_D, H_1 \delta_D \rangle \langle \delta_D, \delta_P \rangle} \right| \leq \frac{\|\delta_D\|_s \|H_1 \delta_D - \delta_P\|_s^*}{|\langle \delta_D, H_1 \delta_D \rangle \langle \delta_D, \delta_P \rangle|} \\ &\leq (1065087000000/907152721283) \frac{1}{\mu \|\delta_D\|_s}. \end{aligned}$$

Further, recall that $\|H_1 \delta_D\|_s^* \leq 1.024263 \mu \|\delta_D\|_s$. Thus, the second term's norm is bounded above by $1.231765 \mu \|\delta_D\|_s$. Using the lower bound on $\langle \delta_D, \delta_P \rangle$ and the upper bounds on $\|H_1 \delta_D - \delta_P\|_s^*$ and $\|H_1 \delta_D + \delta_P\|_s^*$, we get a bound on the first term's norm of

$$\frac{(1.065087) \cdot (2.048265)}{0.960803} \mu \|\delta_D\|_s \leq 2.270581 \mu \|\delta_D\|_s.$$

Adding the bounds on the two terms together gives the advertised bound. \square

Theorem 6.7. *Assume $\|\delta_D\|_s \leq 1/64$, and take $H := \mu F''(\check{s})$,*

$$\begin{aligned} H_1 &:= H + \frac{x \otimes x}{\langle s, x \rangle} - \frac{Hs \otimes Hs}{\langle s, Hs \rangle}, \text{ and} \\ T^2 &:= H_1 + \frac{\delta_P \otimes \delta_P}{\langle \delta_D, \delta_P \rangle} - \frac{H_1 \delta_D \otimes H_1 \delta_D}{\langle \delta_D, H_1 \delta_D \rangle}. \end{aligned}$$

Then, $\|T^2 - H\|_s^* \leq 9.859785 \mu \|\delta_D\|_s \leq 0.154060 \mu$.

Proof. We consider the two rank-two updates separately; Theorem B.3 controls the size of the first rank-two update and Theorem 6.6 controls the size of the second update. We simply add the two bounds together. \square

Theorem 6.8. *If $\|\delta_D\|_s \leq 1/64$, then there exists a $T^2 \in \mathbb{S}_{++}(\mathbb{E}^*, \mathbb{E}^*)$ satisfying the following properties:*

- $T^2 s = x$;
- $T^2 \tilde{s} = \tilde{x}$;
- $0.8303376\mu F''_*(s) \preceq T^2 \preceq 1.169871\mu F''_*(s)$;
- $\frac{0.825446}{\mu} [F''(x)]^{-1} \preceq T^2 \preceq \frac{1.174800}{\mu} [F''(x)]^{-1}$.

That is, $T^2 \in \mathcal{T}_2^2(1.211468; x, s)$.

Proof. We use Theorem 6.7 which gives a $T^2 \in \mathcal{T}_1^2(x, s)$ such that $\|T^2 - \mu F''_*(\check{s})\| \leq 0.154060\mu$. Now, we use the Dikin ellipsoid bound, the fact that $\frac{1}{(1-\|\delta_D\|_s)^2} \leq 1.031998$ and Theorem 2.5 to obtain the last two operator relations. \square

Note that in the above analysis, we did not utilize the additional flexibility provided by the term $(\mu\tilde{\mu} - 1)$. This establishes, in the language of [47], that ξ^* is $O(1)$ within a particular neighbourhood of the central path. Moreover, our specific choice T_H is in $\mathcal{T}_2(\eta; x, s)$ for $\eta = O(1)$, for every pair (x, s) that is in the same neighbourhood.

Therefore, Theorem 5.1 of [47] implies that a wide range of potential reduction algorithms (whose iterates are restricted in a neighbourhood of the central path) have the iteration complexity of $O\left(\sqrt{\vartheta} \ln(1/\epsilon)\right)$.

6.1. Bounds in v -space. In this subsection and the next section, we assume that some suitable bases (and the underlying inner product) have been chosen for the underlying spaces (and we identify both \mathbb{E} and \mathbb{E}^* with \mathbb{R}^n) and for the sake of concreteness, we write A for the underlying matrix representation of \mathcal{A} etc.

The following lemma is useful to the convergence analysis in the next section.

Lemma 6.9. *Suppose $x \in \text{int}(K)$ and $s \in \text{int}(K^*)$ are such that $\|\delta_D\|_s < 1/64$. Take T^2 as in Theorem 6.8 and take $T \in \mathbb{S}_{++}^n$ to be its unique symmetric positive-definite square root. Let $v := Ts = T^{-1}x$ and $\delta_v := T\delta_D$. Let z be an arbitrary vector in v -space. Let $x' \in \text{int}(K)$ and $s' \in \text{int}(K^*)$; define $\delta'_D := s' + \mu F'(x')$ and $\delta'_v := T\delta'_D$. Then,*

- (1) $\|Tz\| \leq 1.081606\sqrt{\mu}\|z\|_s$;
- (2) $\|z\|_s \leq 1.097395\|Tz\|/\sqrt{\mu}$;
- (3) $\|\delta_v\| \leq 0.016901\sqrt{\mu}$;
- (4) $\|\delta'_D\|_{s'} \leq \frac{\|\delta'_D\|_s}{1-\|s-s'\|_s}$;
- (5) if $\|\delta'_v\| \leq 0.003703\sqrt{\mu}$ and $\|s-s'\|_s \leq 1/25$, then $\|\delta'_D\|_{s'} \leq 0.004234$;
- (6) if $\|\delta'_v\| \leq 0.010519\sqrt{\mu}$ and $\|s-s'\|_s \leq 1/25$, then $\|\delta'_D\|_{s'} \leq 1/64$.

Proof. Recall from Theorem 6.8 that

$$(10) \quad 0.830376\mu\|z\|_s^2 \leq \|Tz\|^2 \leq 1.1169871\mu\|z\|_s^2.$$

- (1) This is the square root of $\|Tz\|^2 \leq 1.1169871\mu\|z\|_s^2$ with a constant rounded up.
- (2) This is the square root of $\mu\|z\|_s^2 \leq \frac{1}{0.830376}\|Tz\|^2$ with a constant rounded up.
- (3) Take $z = \delta_D$ in part (1) and then use $\|\delta_D\|_s \leq 1/64$; we see

$$\|\delta_v\| < 1.081606\sqrt{\mu}\|\delta_D\|_s \leq 0.016901\sqrt{\mu},$$

as desired.

- (4) This is the Dikin ellipsoid bound for comparing the s' -norm with the s -norm.
- (5) If $\|\delta'_v\| \leq 0.003703\sqrt{\mu}$, then by part (2) $\|\delta'_D\|_s \leq 0.004064$. By part (4), then, $\|\delta'_D\|_{s'} \leq 0.004234$, as desired.
- (6) If $\|\delta'_v\| \leq 0.010519\sqrt{\mu}$, then by part (2) $\|\delta'_D\|_s \leq 0.011544$. By part (4), then, $\|\delta'_D\|_{s'} \leq 0.012025$, which implies the desired result.

□

7. ALGORITHMS

In this section, we complete our proof of $O(\sqrt{\vartheta} \ln \frac{1}{\epsilon})$ iteration complexity bounds on variants of the following feasible-start primal-dual interior point algorithm with different choices of α and γ :

Take $k := 0$ and $(x^{(0)}, y^{(0)}, s^{(0)})$ to be feasible and central (we can also accept approximately central points).

while $(s^{(k)})^\top x^{(k)} > \epsilon\vartheta$ **do**

 Let $\mu_k := (s^{(k)})^\top x^{(k)} / \vartheta$.

 Take T^2 as in Theorem 6.7.

 Select $\gamma \in [0, 1]$.

 Solve

$$\begin{pmatrix} 0 & A^\top & I \\ A & 0 & 0 \\ I & 0 & T^2 \end{pmatrix} \begin{pmatrix} d_x \\ d_y \\ d_s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -x^{(k)} - \gamma\mu_k F'_*(s) \end{pmatrix}.$$

 Select $\alpha_k \in [0, \infty)$.

$(x^{(k+1)}, y^{(k+1)}, s^{(k+1)}) \leftarrow (x^{(k)}, y^{(k)}, s^{(k)}) + \alpha_k(d_x, d_y, d_s)$.

$k \leftarrow k + 1$.

end while

Lemma 7.1. *Let $r_v := -v + \gamma\mu w$. Then, the system of equations in Line 3 of the above algorithm imply*

$$T^{-1}d_x = \Pi_{\ker(AT)}r_v, \quad Td_s = \Pi_{\text{im}((AT)^\top)}r_v.$$

In particular, $\|T^{-1}d_x\| \leq \|r_v\|$ and $\|Td_s\| \leq \|r_v\|$.

Proof. The third equation ensures that $T^{-1}d_x + Td_s = r_v$. The first two equations imply that $T^{-1}d_x$ must lie in $\ker(AT)$ while Td_s must lie in $\text{im}((AT)^\top)$. Since these two linear spaces are orthogonal, the result follows. \square

The following result does not hint at quadratic convergence. However, a tighter analysis of the low-rank updates showing that the approximation error is linear in $\|\delta_D\|_s$ within the $\frac{1}{64}$ -neighbourhood would suffice to establish quadratic convergence. This is not hard to do, since the ingredients are already given above. We do not do this here, because quadratic convergence of centering is not needed to establish the desired complexity result.

Lemma 7.2. *Suppose $x^{(k)} \in \text{int}(K)$ and $s^{(k)} \in \text{int}(K^*)$ define a feasible solution. If $\gamma_k = 1$ and $\alpha_k = 1$ and $\|\delta_D^k\|_{s^{(k)}} \leq \frac{1}{64}$, then*

- $Ax^{(k+1)} = b$ and $A^\top y^{(k+1)} + s^{(k+1)} = c$.
- $x^{(k+1)} \in \text{int}(K)$ and $s^{(k+1)} \in \text{int}(K^*)$.
- $\|\delta_D^{(k+1)}\|_{s^{(k+1)}} \leq 0.004234$.
- $\mu_{k+1} = \mu_k$.

Proof. The system of linear equations that determine d_x , d_y , and d_s guarantee that $d_x \in \ker A$ and $d_s = -A^\top d_y$; since $Ax^{(k)} = b$ and $A^\top y^{(k)} + s^{(k)} = c$, it follows that $Ax^{(k+1)} = b$ and $A^\top y^{(k+1)} + s^{(k+1)} = c$. We drop the superscript k when speaking of the k th iterate in this proof.

Notice that, with this choice of γ ,

$$T^{-1}d_x + Td_s = -\delta_v.$$

Since

$$\|d_x\|_x \leq \frac{\sqrt{1.174800}}{\sqrt{\mu}} \|T^{-1}d_x\| \leq \frac{\sqrt{1.174800}}{\sqrt{\mu}} \|\delta_v\| \leq 0.018319 < 1,$$

strict primal feasibility is retained (for the last inequality above, we used Lemma 6.9 part (3)). A similar argument shows that strict dual feasibility is retained.

For the first equation below, we use the identity $F'(x + d_x) = F'(x) + \int_0^1 F''(x + \tau d_x) d_x d\tau$ and obtain

$$\begin{aligned} \|T(s + d_s + \mu F'(x + d_x))\| &= \left\| T \left[\delta_D + d_s + \mu \left(\int_0^1 F''(x + \tau d_x) d\tau \right) d_x \right] \right\| \\ &\leq \|\delta_v + Td_s + T^{-1}d_x\| + \left\| T \left[T^{-2} - \mu \left(\int_0^1 F''(x + \tau d_x) d\tau \right) \right] d_x \right\|. \end{aligned}$$

The first term is, of course, zero. However, notice that, by the Dikin ellipsoid bound and Theorem 6.8

$$\mu \left(\int_0^1 F''(x + \tau d_x) d\tau \right) \leq 1.219055 T^{-2}$$

and, similarly,

$$\mu \left(\int_0^1 F''(x + \tau d_x) d\tau \right) \geq 0.795480 T^{-2}$$

We therefore bound, using Lemma 6.9 part (3),

$$\left\| T \left[T^{-2} - \mu \left(\int_0^1 F''(x + \tau d_x) d\tau \right) \right] d_x \right\| \leq 0.219055 \|T^{-1} d_x\| = 0.219055 \|\delta_v\| < 0.003703 \sqrt{\mu},$$

which implies, by Lemma 6.9 part (5), the advertised bound on the new $\|\delta_D\|_s$.

As we observed in Section 2, μ is unchanged by a centering iteration. \square

Lemma 7.3. *Suppose $\|\delta_D^{(k)}\|_{s^{(k)}} \leq 0.004234$, let $\gamma_k := 0$ and $\alpha_k := \frac{0.005}{\sqrt{\vartheta}}$. Then,*

- $Ax^{(k+1)} = b$ and $A^\top y^{(k+1)} + s^{(k+1)} = c$;
- $x^{(k+1)} \in \text{int}(K)$ and $s^{(k+1)} \in \text{int}(K^*)$;
- $\mu_{k+1} \leq (1 - \alpha_k)\mu_k$;
- $\|\delta_D^{(k+1)}\|_{s^{(k+1)}} \leq \frac{1}{64}$.

Proof. We recall that $\|\delta_D\|_s \leq 0.004234$ means that, by Lemma 6.9 part (1), $\|\delta_v\| \leq 0.004580\sqrt{\mu}$. Notice that by Theorem ??, we have

$$\|d_x\|_x \leq \sqrt{\frac{1.174800}{\mu_k}} \|T^{-1} d_x\| \leq \sqrt{\frac{1.174800}{\mu_k}} \|v\| \leq 1.0839\sqrt{\vartheta}.$$

Consequently, any step with $\alpha < \frac{1}{2\sqrt{\vartheta}}$ retains strict primal feasibility. A similar analysis (due to the primal-dual symmetry of our set-up) reveals that $\|d_s\|_s \leq 1.097393\sqrt{\vartheta}$ and hence the dual step retains strict dual feasibility for α similarly bounded. Notice that $\|\alpha_k d_s\|_s \leq 1/25$; this permits us to use Lemma 6.9 part (6) later.

As we observed in Section 2, $\langle s(\alpha), x(\alpha) \rangle = (1 - \alpha)\vartheta\mu$. This establishes the desired reduction in μ .

We compute

$$\begin{aligned}
\left\| T\delta_D^{(k+1)} \right\| &= \left\| T[s + \alpha d_s + \mu F'(x + \alpha d_x)] \right\| \\
&= \left\| T \left[s + \alpha d_s + \mu F'(x) + \alpha \mu \left(\int_0^1 F''(x + \tau \alpha d_x) d\tau \right) d_x \right] \right\| \\
&\leq \left\| \delta_v \right\| + \alpha \left\| T d_s + \mu T \left(\int_0^1 F''(x + \tau \alpha d_x) d\tau \right) d_x \right\|,
\end{aligned}$$

where for the second equation above, we used the identity

$$F'(x + \alpha d_x) = F'(x) + \alpha \left(\int_0^1 F''(x + \tau \alpha d_x) d\tau \right) d_x. \text{ Let us write}$$

$$E := \mu \left(\int_0^1 F''(x + \tau \alpha d_x) d\tau \right) - T^{-2}.$$

Then, by Theorem 6.8, and the Dikin ellipsoid bound, we obtain

$$-0.183477T^{-2} \preceq E \preceq 0.1876638T^{-2}.$$

We thus get an upper bound of

$$\begin{aligned}
\left\| T\delta_D^{(k+1)} \right\| &\leq \left\| \delta_v \right\| + \alpha_k \left\| T d_s + T(E + T^{-2}) d_x \right\| \\
&\leq \left\| \delta_v \right\| + \alpha_k \left\| T d_s + T^{-1} d_x \right\| + \alpha_k \left\| T E d_x \right\| \\
&\leq \left\| \delta_v \right\| + \alpha_k \left\| v \right\| + \alpha_k \left\| T E T \right\| \left\| T^{-1} d_x \right\| \\
&\leq \left\| \delta_v \right\| + \alpha_k (1 + \left\| T E T \right\|) \left\| v \right\| \\
&\leq 0.004580\sqrt{\mu} + (0.005) \cdot (1.187638)\sqrt{\mu} \\
&< 0.010519\sqrt{\mu}.
\end{aligned}$$

The norm in the expression $\left\| T E T \right\|$ above is the spectral (operator) 2-norm. This implies, by Lemma 6.9 part (6), that $\left\| \delta_D^{(k+1)} \right\|_{s^{(k+1)}} \leq \frac{1}{64}$, as desired. \square

We immediately have

Corollary 7.4. *Starting from an initial, primal-dual pair of feasible and central points, one can alternately apply the predictor and corrector steps outlined from the last two lemmata and recover an algorithm that takes at most $278\sqrt{\vartheta}$ iterations to reduce μ by a factor of two. In particular, this gives an $O\left(\sqrt{\vartheta} \ln(1/\epsilon)\right)$ bound on the iteration complexity of the algorithm using this choice of γ .*

Proof. By Lemma 7.3, μ is reduced by a factor of $\left(1 - \frac{0.005}{\sqrt{\vartheta}}\right)$ in each iteration performing a predictor step. By Lemma 7.2 every corrector step (which follows every predictor step) gets us in a neighbourhood of the central path given by $\left\| \delta_D \right\|_s \leq 1/64$ without changing μ . Therefore,

to provide an upperbound on the number of predictor steps (to reduce μ by a factor of two), it suffices to solve the inequality

$$\left(1 - \frac{0.005}{\sqrt{\vartheta}}\right)^k \leq \frac{1}{2},$$

for a (small) positive integer k . We obtain (using the fact that $\ln(1 - \kappa) \leq -\kappa$, for all $\kappa < 1$), choosing

$$k \geq \frac{\ln(2)}{0.005} \sqrt{\vartheta}$$

suffices. Doubling such a feasible k (to account for the corrector steps) yields the claimed bound. \square

8. HYPERBOLICITY CONE PROGRAMMING: THE HYPERBOLIC BARRIERS SPECIAL CASE

In this section, we assume that we are given access to a hyperbolic polynomial p and K is a corresponding hyperbolicity cone. As we mentioned earlier, on the one hand, $F(x) := -\ln(p(x))$ is a self-concordant barrier for K with long-step Hessian estimation property. On the other hand, we do not necessarily have explicit and efficient access to F_* or its derivatives; moreover, F_* will *not* have the long-step Hessian estimation property, unless K is a symmetric cone. We will discuss two issues:

- How do we evaluate F_* ?
- Can we compute the primal integral scaling?

8.1. Evaluating the dual barrier. Given an oracle returning $F(x)$, $F'(x)$, and $F''(x)$ on input x , we describe an algorithm for approximating F 's Legendre-Fenchel conjugate, F_* , and discuss its convergence.

```

loop
   $r \leftarrow F'(\hat{x}) + s$ 
  if  $\|r\|_{\hat{x}}^* < \epsilon$  then return  $\hat{x}$ 
  end if
   $N \leftarrow -[F''(\hat{x})]^{-1} r$ 
   $\hat{x} \leftarrow \hat{x} + N$ 
end loop

```

Intuitively, this is steepest descent in the local \hat{x} -norm (i.e., Newton's method for solving $F'(x) + s = 0$). This algorithm is locally quadratically convergent, since the dual s -norm is well-approximated by the dual \hat{x} -norm when $-F'(\hat{x})$ is "close to" s . In particular, one can show that if $\|r\|_{\hat{x}}^* \leq \frac{1}{4}$, then the \hat{x} -norm of the new residual is at most $\frac{16}{27}$ of that of the old. This implies that the dual s -norm of the new residual is at most $\frac{64}{81}$ of that of the old residual,

ensuring descent. The dual s -norm is scaled by a factor of at most

$$\|r\|_{\hat{x}}/(1 - \|r\|_{\hat{x}})^4 \leq 3.2\|r\|_{\hat{x}}$$

in each iteration, establishing local quadratic convergence.

Note that replacing s with $-F'(\hat{x})$, for some sufficiently good approximation \hat{x} , can degrade the complementarity gap and the measure of centrality $\|\delta_P\|_x$ and ruins the equation $A^\top y + s = c$. Thus one needs to work with an infeasible interior-point method or a self-dual embedding technique in the absence of an exactly-evaluated dual barrier. However, it is straightforward to bound the local s -norm of the increase in residual by $\|s + F'(\hat{x})\|_s$.

We can use various integration techniques to evaluate the primal integral scaling as we outline in Appendix C.

9. CONCLUSIONS AND FUTURE WORK

We presented a new primal-dual scaling map based on a line integral for convex programming where only a ϑ -LHSCB is supplied. We derived some properties of this scaling, notably that it points to the richness of potential primal-dual local metrics, enriches the connection of primal-dual interior-point methods to Riemannian geometry. We presented a new analysis of low-rank updates of [47] showing that, if one is close to the central path and one begins with a certain approximation to the integral scaling, the low-rank update has small norm close to the central path. Some steps of this analysis were peculiar to the particular approximation chosen; we leave it to future work to generalise the analysis to something depending more directly on the approximation error. We presented a generalization of the Mizuno-Todd-Ye predictor-corrector scheme that uses the above tools and showed that it matches the current best, worst-case iteration complexity of $O(\sqrt{\vartheta} \ln(1/\epsilon))$, of the special case of symmetric cone programming.

We presented an algorithm for computing an approximation to the conjugate barrier given an oracle that computes the primal barrier and discussed some bounds that, within the context of an infeasible-start interior-point method or a self-dual embedding technique (see [51, 33]), do not degrade the worst-case iteration complexity.

We presented two techniques based on Gaussian quadrature for evaluating the new primal-dual scaling map for hyperbolic barrier functions; one exact, and one with bounded approximation error. Again, we leave to future work the problem of tying such an approximation error bound to a bound on the magnitude of the necessary low-rank update to the approximation.

REFERENCES

- [1] Heinz H. Bauschke, Osman Güler, Adrian S. Lewis, and Hristo S. Sendov. Hyperbolic polynomials and convex analysis. *Canad. J. Math.*, 53(3):470–488, 2001.
- [2] Chek Beng Chua. The primal-dual second-order cone approximations algorithm for symmetric cone programming. *Found. Comput. Math.*, 7(3):271–302, 2007.

- [3] Chek Beng Chua. A T -algebraic approach to primal-dual interior-point algorithms. *SIAM J. Optim.*, 20(1):503–523, 2009.
- [4] John E. Dennis, Jr. and Robert B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. Prentice Hall Series in Computational Mathematics. Prentice Hall, Inc., Englewood Cliffs, NJ, 1983.
- [5] Yves Genin, Yvan Hachez, Yu Nesterov, and Paul Van Dooren. Optimization problems over positive pseudopolynomial matrices. *SIAM Journal on Matrix Analysis and Applications*, 25(1):57–79, 2003.
- [6] Osman Güler. Hyperbolic polynomials and interior point methods for convex programming. *Math. Oper. Res.*, 22(2):350–377, 1997.
- [7] Raphael Hauser. The Nesterov-Todd direction and its relation to weighted analytic centers. *Found. Comput. Math.*, 4(1):1–40, 2004.
- [8] Raphael A. Hauser and Osman Güler. Self-scaled barrier functions on symmetric cones and their classification. *Found. Comput. Math.*, 2(2):121–143, 2002.
- [9] Raphael A. Hauser and Yongdo Lim. Self-scaled barriers for irreducible symmetric cones. *SIAM J. Optim.*, 12(3):715–723, 2002.
- [10] B. Jansen, C. Roos, and T. Terlaky. A polynomial primal-dual Dikin-type algorithm for linear programming. *Math. Oper. Res.*, 21(2):341–353, 1996.
- [11] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.
- [12] M. Kojima, N. Megiddo, T. Noma, and A. Yoshise. *A unified approach to interior point algorithms for linear complementarity problems*, volume 538 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 1991.
- [13] Masakazu Kojima, Shinji Mizuno, and Akiko Yoshise. An $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems. *Math. Programming*, 50(3, (Ser. A)):331–342, 1991.
- [14] Yongdo Lim. Maximum-volume symmetric gauge ball problem on the convex cone of positive definite matrices and convexity of optimal sets. *SIAM J. Optim.*, 21(4):1275–1288, 2011.
- [15] Shinji Mizuno, Michael J. Todd, and Yinyu Ye. On adaptive-step primal-dual interior-point algorithms for linear programming. *Math. Oper. Res.*, 18(4):964–981, 1993.
- [16] Lajos Molnár. Maps preserving the geometric mean of positive operators. *Proc. Amer. Math. Soc.*, 137(5):1763–1770, 2009.
- [17] Renato D. C. Monteiro, Ilan Adler, and Mauricio G. C. Resende. A polynomial-time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series extension. *Math. Oper. Res.*, 15(2):191–214, 1990.
- [18] Tor Myklebust and Levent Tunçel. Hyperbolic cone programming: Structure and interior-point algorithms. Talk presented at the SIAM Applied Algebraic Geometry Conference, Fort Collins, CO, USA, 2013.
- [19] Tor Gunnar Josefsson Myklebust. *On primal-dual interior-point algorithms for convex optimisation*. PhD thesis, Department of Combinatorics and Optimization, University of Waterloo, 2015.
- [20] Arkadi Nemirovski and Levent Tunçel. “Cone-free” primal-dual path-following and potential-reduction polynomial time interior-point methods. *Math. Program.*, 102(2, Ser. A):261–294, 2005.
- [21] Yu. Nesterov. Long-step strategies in interior-point primal-dual methods. *Math. Programming*, 76(1, Ser. B):47–94, 1997. Interior point methods in theory and practice (Iowa City, IA, 1994).
- [22] Yurii Nesterov. Squared functional systems and optimization problems. In *High performance optimization*, pages 405–440. Springer, 2000.
- [23] Yurii Nesterov. *Introductory lectures on convex optimization*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, Boston, MA, 2004. A basic course.
- [24] Yurii Nesterov. Parabolic target space and primal-dual interior-point methods. *Discrete Appl. Math.*, 156(11):2079–2100, 2008.
- [25] Yurii Nesterov. Towards non-symmetric conic optimization. *Optim. Methods Softw.*, 27(4-5):893–917, 2012.

- [26] Yurii Nesterov and Arkadi Nemirovski. Multi-parameter surfaces of analytic centers and long-step surface-following interior point methods. *Math. Oper. Res.*, 23(1):1–38, 1998.
- [27] Yurii Nesterov and Arkadi Nemirovski. Primal central paths and Riemannian distances for convex sets. *Found. Comput. Math.*, 8(5):533–560, 2008.
- [28] Yurii Nesterov and Arkadii Nemirovskii. Conic duality and its applications in convex programming. *Optim. Methods Softw.*, 1:95–115, 1992.
- [29] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*, volume 13 of *SIAM Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.
- [30] Yurii Nesterov and Michael J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Math. Oper. Res.*, 22(1):1–42, 1997.
- [31] Yurii Nesterov and Michael J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM J. Optim.*, 8(2):324–364, 1998.
- [32] Yurii Nesterov and Michael J. Todd. On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Found. Comput. Math.*, 2(4):333–361, 2002.
- [33] Yurii Nesterov, Michael J. Todd, and Yinyu Ye. Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems. *Math. Program.*, 84:227–267, 1999.
- [34] Yurii Nesterov and Levent Tunçel. Local superlinear convergence of polynomial-time interior-point methods for hyperbolicity cone optimization problems. *SIAM Journal on Optimization*, 26(1):139–170, 2016.
- [35] James Renegar. A polynomial-time algorithm, based on Newton’s method, for linear programming. *Math. Programming*, 40(1, (Ser. A)):59–93, 1988.
- [36] James Renegar. *A mathematical view of interior-point methods in convex optimization*. MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Programming Society (MPS), Philadelphia, PA, 2001.
- [37] James Renegar. Hyperbolic programs, and their derivative relaxations. *Found. Comput. Math.*, 6(1):59–79, 2006.
- [38] James Renegar. Central swaths: a generalization of the central path. *Found. Comput. Math.*, 13(3):405–454, 2013.
- [39] James Renegar. Primal-dual algorithms for optimization over hyperbolicity cones. Talk presented at the SIAM Applied Algebraic Geometry Conference, Fort Collins, CO, USA, 2013.
- [40] James Renegar and Mutiara Sondjaja. A polynomial-time affine-scaling method for semidefinite and hyperbolic programming. arXiv:1410.6734, 2014.
- [41] R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.
- [42] Stefan H. Schmieta. Complete classification of self-scaled barrier functions. Tech. Report, Dept. of IEOR, Columbia Univ., NY, USA, 2000.
- [43] Josef Stoer, Roland Bulirsch, Richard H. Bartels, Walter Gautschi, and Christoph Witzgall. *Introduction to numerical analysis*. Texts in applied mathematics. Springer, New York, 2002.
- [44] Jos F. Sturm and Shuzhong Zhang. Symmetric primal-dual path-following algorithms for semidefinite programming. In *Proceedings of the Stieltjes Workshop on High Performance Optimization Techniques (HPOPT ’96) (Delft)*, volume 29, pages 301–315, 1999.
- [45] Levent Tunçel. Primal-dual symmetry and scale invariance of interior-point algorithms for convex optimization. *Math. Oper. Res.*, 23(3):708–718, 1998.
- [46] Levent Tunçel. Potential reduction and primal-dual methods. In *Handbook of semidefinite programming*, volume 27 of *Internat. Ser. Oper. Res. Management Sci.*, pages 235–265. Kluwer Acad. Publ., Boston, MA, 2000.

- [47] Levent Tunçel. Generalization of primal-dual interior-point methods to convex optimization problems in conic form. *Found. Comput. Math.*, 1(3):229–254, 2001.
- [48] Levent Tunçel. *Polyhedral and semidefinite programming methods in combinatorial optimization*, volume 27 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2010.
- [49] William C. Waterhouse. Linear transformations preserving symmetric rank one matrices. *J. Algebra*, 125(2):502–518, 1989.
- [50] Hua Wei. *Convergence analysis of generalized primal-dual interior-point algorithms for linear optimization*. Department of Combinatorics and Optimization, Faculty of Mathematics, Waterloo, Ontario, Canada, 2002. Thesis (M.Math.)—University of Waterloo.
- [51] Yinyu Ye, Michael J. Todd, and Shinji Mizuno. An $O(\sqrt{n}L)$ -iteration homogeneous and self-dual linear programming algorithm. *Math. Oper. Res.*, 19(1):53–67, 1994.

APPENDIX A. PRIMAL-DUAL SYMMETRY BASED ON THE LOCAL METRICS

In [47] a very general framework for primal-dual symmetric algorithms were provided. In this section, we make the richness for the choice of such algorithms within the framework provided by the sets $\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2$, more explicit. Basically, *every* consistent choice of a local primal-dual metric T^2 from any of the sets $\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2$ can be used to design a primal-dual symmetric interior-point algorithm as we prove below.

Proposition A.1. *Let $(x, s) \in \text{int}(K) \oplus \text{int}(K^*)$. Then, for every pair $H, T \in \mathcal{T}_0^2(x, s)$,*

$$\frac{1}{2}(H + T), \left(\frac{H^{-1} + T^{-1}}{2} \right)^{-1} \in \mathcal{T}_0^2(x, s).$$

The same property holds for $\mathcal{T}_1^2(x, s)$ and for $\mathcal{T}_2^2(\eta; x, s)$ (for every $\eta \geq 1$).

Proof. Follows from the definitions and convexity of $\mathcal{T}_0^2, \mathcal{T}_1^2, \mathcal{T}_2^2$. □

Corollary A.2. *For every convex cone $K \subset \mathbb{E}$ and for every pair $(x, s) \in \text{int}(K) \oplus \text{int}(K^*)$, the sets $\mathcal{T}_0^2(x, s)$, $\mathcal{T}_1^2(x, s)$ and $\mathcal{T}_2^2(\eta; x, s)$ (for every $\eta \geq 1$) are geodesically convex.*

Proof. Since $\mathcal{T}_2^2(\eta; x, s)$ is a closed subset of \mathbb{S}_{++}^n , employing Proposition A.1 above and Lemma 2.3 of [14], we conclude that $\mathcal{T}_2^2(\eta; x, s)$ is geodesically convex for all η . For \mathcal{T}_0 and \mathcal{T}_1 we can adapt the proof of the same lemma (even though our sets are not closed, we can argue that all limits of all mean iterations on the elements of \mathcal{T}_0^2 and \mathcal{T}_1^2 are positive-definite and hence stay in the corresponding set). □

The above corollary indicates a way to convert *any* consistent choice of T^2 to a scaling for a primal-dual symmetric interior-point algorithm in the sense of [45]. (Simply take the operator geometric mean of the consistent choice of T^2 with the inverse of the same formula for T^2 applied to (P) and (D) switched.)

APPENDIX B. DETAILS OF THE COMPLEXITY AND ERROR ANALYSIS

B.1. Zeroth-order low-rank update. We bound the operator norm of the zeroth-order low-rank update in a small neighbourhood of the central path defined by the condition $\|\delta_D\|_s \leq \frac{1}{64}$.

Theorem B.1. *Assume $\|\delta_D\|_s \leq 1/64$. Then*

- (1) $\|F_*''(s)\delta_D\|_s^* = \|\delta_D\|_s$;
- (2) $\|F_*''(\check{s})\delta_D\|_s^* \leq 1.031998 \|\delta_D\|_s$;
- (3) for every $v \in \mathbb{E}^*$,

$$|\langle v, \mu F_*''(\check{s})\delta_D - \delta_P \rangle| \leq 0.524190\mu \|v\|_s \|\delta_D\|_s^2;$$

- (4) $\|\mu F_*''(\check{s})\delta_D - \delta_P\|_s^* \leq 0.524190\mu \|\delta_D\|_s^2$.

Proof. (1) This follows straightforwardly from the definitions.

- (2) Recall that $F_*''(\bar{s}) \preceq \frac{1}{(1-\|\delta_D\|_s)^2} F_*''(s)$. Now we apply the previous part. Next, we notice that $\frac{1}{(1-\|\delta_D\|_s)^2} \leq 1.031998$.

- (3) Let $f(t) = \langle u, F_*'(\check{s} + t\delta_D) \rangle$. We consider an order-two Taylor expansion of f around zero; we see that, for every $t \in [-1/2, 1/2]$, there exists a $\bar{t} \in [\min(0, t), \max(0, t)]$ such that

$$f(t) = f(0) + tf'(0) + \frac{1}{2}t^2 f''(\bar{t}).$$

Notice that

$$|f''(\bar{t})| = |F_*'''(\check{s} + \bar{t}\delta_D)[u, \delta_D, \delta_D]| \leq 2\|u\| \|\delta_D\|_s^2,$$

where both norms are the local $(\check{s} + \bar{t}\delta_D)$ norms (we used self-concordance property of F_*). We then apply the Dikin ellipsoid bound to these norms to relate

$$\|u\|_{\check{s} + \bar{t}\delta_D} \leq \frac{1}{1 - \|\delta_D\|_s} \|u\|_s$$

and similarly for $\|\delta_D\|_s$. Consequently, using $1/(1 - \|\delta_D\|_s) \leq 1.015874$, we see that

$$|f''(\bar{t})| \leq 1.048378 \|u\|_s \|\delta_D\|_s^2.$$

Thus, for some $\bar{t}_1 \in [-1/2, 0]$ and $\bar{t}_2 \in [0, 1/2]$, we have

$$f(1/2) - f(-1/2) = f'(0) + \frac{1}{8}(f''(\bar{t}_1) - f''(\bar{t}_2)).$$

Consequently,

$$|f'(0) - f(1/2) + f(-1/2)| \leq 0.262094 \|u\|_s \|\delta_D\|_s^2.$$

Notice that, by substitution and the chain rule,

- $f'(0) = \langle u, F_*''(\check{s})\delta_D \rangle$;
- $f(-1/2) = \langle u, F_*'(\mu\check{s}) \rangle = -\langle u, x/\mu \rangle$;
- $f(1/2) = \langle u, F_*'(s) \rangle = -\langle u, \tilde{x} \rangle$.

The claimed bound now follows.

(4) We use the definition of a dual norm:

$$\begin{aligned} \|F_*''(\check{s})\delta_D - x/\mu + \tilde{x}\|_s^* &= \sup_{\|u\|_s=1} \langle u, F_*''(\check{s})\delta_D - x/\mu + \tilde{x} \rangle \\ &\leq \sup_{\|u\|_s=1} 0.524190\|u\|_s\|\delta_D\|_s^2 = 0.524190\|\delta_D\|_s^2. \end{aligned}$$

□

Lemma B.2. *Assume $\|\delta_D\|_s \leq 1/64$. Then,*

(1) *for every $v \in \mathbb{E}^*$,*

$$|\langle v, F_*''(\check{s})[\delta_D] \rangle| \leq 1.015811 \|v\|_s \|\delta_D\|_s;$$

(2) *for every $v \in \mathbb{E}^*$ and $\bar{s} \in [s, \mu\tilde{s}]$,*

$$|F_*'''(\bar{s})[v, \delta_D, \mu\tilde{s}]| \leq 2.096758 \|v\|_s \|\delta_D\|_s \sqrt{\vartheta};$$

(3) $\|Hs - x\|_s^* \leq 2.064190\sqrt{\vartheta}\mu \|\delta_D\|_s$;

(4) $\|x\|_s^* \leq 1.015874\sqrt{\vartheta}\mu$;

(5) $\|Hs\|_s^* \leq 1.015811\sqrt{\vartheta}\mu$;

(6) $\|Hs + x\|_s^* \leq 2.031685\sqrt{\vartheta}\mu$;

(7) $\langle s, Hs \rangle \geq 0.984436\vartheta\mu$.

Proof. (1) We compute, using Cauchy-Schwarz and the Dikin ellipsoid bound,

$$|\langle v, F_*''(\check{s})[\delta_D] \rangle| = |F_*''(\check{s})[\delta_D, v]| \leq \|\delta_D\|_{\check{s}} \|v\|_{\check{s}} \leq 1.015810 \|\delta_D\|_s \|v\|_s.$$

(2) We compute

$$|F_*'''(\bar{s})[v, \delta_D, \mu\tilde{s}]| \leq 2 \|v\|_{\bar{s}} \|\delta_D\|_{\bar{s}} \|\mu\tilde{s}\|_{\bar{s}} \leq 2.096758 \|v\|_s \|\delta_D\|_s \|\mu\tilde{s}\|_{\mu\tilde{s}}.$$

(3) We write

$$Hs = \mu F_*''(\check{s})(\mu\tilde{s}) + \mu F_*''(\check{s})\delta_D.$$

On the first term, we perform a Taylor expansion around $\mu\tilde{s}$; for every v there is a \bar{s} on the line segment between s and $\mu\tilde{s}$ such that

$$\begin{aligned} F_*''(\check{s})[\mu\tilde{s}, v] &= F_*''(s - \delta_D)[\mu\tilde{s}, v] + \frac{1}{2}F_*'''(\bar{s})[\mu\tilde{s}, \delta_D, v] \\ &= \langle v, x \rangle / \mu + \frac{1}{2}F_*'''(\bar{s})[\mu\tilde{s}, \delta_D, v]. \\ &\leq \langle v, x \rangle / \mu + 1.048379 \|v\|_s \|\delta_D\|_s \sqrt{\vartheta}. \end{aligned}$$

We also bound (using the Dikin ellipsoid bound first, followed by Lemma 7.2)

$$\langle v, F_*''(\check{s})\delta_D \rangle \leq 1.015811 \|v\|_s \|\delta_D\|_s.$$

Adding these bounds and taking a supremum over all v such that $\|v\|_s = 1$, since $\vartheta \geq 1$, yields the bound

$$\|Hs - x\|_s^* \leq 2.064190\sqrt{\vartheta}\mu\|\delta_D\|_s,$$

as desired.

(4) This follows directly from the Dikin ellipsoid bound.

(5) Notice that

$$\|Hs\|_s^* = \mu \langle s, F_*''(\check{s})(F_*''(s))^{-1}F_*''(\check{s})s \rangle^{1/2} \leq \mu \frac{\langle s, F_*''(s)s \rangle^{1/2}}{(1 - \|\delta_D\|_s/2)^2} \leq 1.015810\sqrt{\vartheta}\mu.$$

(6) We apply the triangle inequality to the last two parts.

(7) Note that $\langle s, Hs \rangle = \mu F''(\check{s})[s, s] \geq \mu(1 - \|\delta_D\|_s/2)^2 F''(s)[s, s] \geq (16129/16384)\vartheta\mu$.

□

Theorem B.3. *Assume $\|\delta_D\|_s \leq 1/64$. Then, the zeroth-order low rank update has a small norm:*

$$\left\| \frac{x \otimes x}{\langle s, x \rangle} - \frac{Hs \otimes Hs}{\langle s, Hs \rangle} \right\|_s \leq 6.357439\mu\|\delta_D\|_s.$$

Proof. We write the first low-rank update as

$$\frac{(x \otimes x) - (Hs \otimes Hs)}{\langle s, x \rangle} + \left(\frac{1}{\langle s, x \rangle} - \frac{1}{\langle s, Hs \rangle} \right) Hs \otimes Hs.$$

Then, using the triangle inequality, Lemma 2.11, and the equation $\|h \otimes h\|_s = \|h\|_s^{*2}$ we bound its norm above by

$$\frac{1}{\langle s, x \rangle} \|x - Hs\|_s^* \|x + Hs\|_s^* + \left| \frac{1}{\langle s, x \rangle} - \frac{1}{\langle s, Hs \rangle} \right| \|Hs\|_s^{*2}.$$

The first term is bounded above by

$$(83875677203/20000000000)\mu\|\delta_D\|_s.$$

To bound the second term, note that

$$\left| \frac{1}{\langle s, x \rangle} - \frac{1}{\langle s, Hs \rangle} \right| = \left| \frac{\langle s, Hs - x \rangle}{\vartheta\mu \langle s, Hs \rangle} \right| \leq \frac{\|s\|_s \|Hs - x\|_s^*}{\vartheta\mu \langle s, Hs \rangle} \leq 2.064190 \frac{\|\delta_D\|_s}{\langle s, Hs \rangle}.$$

Now, we bound $\langle s, Hs \rangle$ below by $0.984436\vartheta\mu$ to get a bound of

$$(105686528/50403125) \frac{\|\delta_D\|_s}{\vartheta\mu}.$$

The bound $\|Hs\|_s^* \leq 1.015811\sqrt{\vartheta}\mu$ then gives an overall bound on the second term of

$$(2129979833381099/98443603515625000)\mu\|\delta_D\|_s$$

Adding fractions gives (something slightly stronger than) the desired bound. □

B.2. First-order low-rank update. Next, we bound the operator norm of the first-order low-rank update in a small neighbourhood of the central path defined by the condition $\|\delta_D\|_s \leq \frac{1}{64}$.

Lemma B.4. *Assume $\|\delta_D\|_s \leq 1/64$. Let $H_1 := H + \frac{x \otimes x}{\langle s, x \rangle} - \frac{Hs \otimes Hs}{\langle s, Hs \rangle}$. Then,*

- (1) $\|H\delta_D - \delta_P\|_s^* \leq 0.524190\mu \|\delta_D\|_s^2$;
- (2) $\langle \delta_D, H\delta_D \rangle \geq 0.968994\mu \|\delta_D\|_s^2$;
- (3) $\langle \delta_D, \delta_P \rangle \geq 0.960803\mu \|\delta_D\|_s^2$;
- (4) $\|H\delta_D\|_s^* \leq 1.015811\mu \|\delta_D\|_s$;
- (5) $\|\delta_P\|_s^* \leq 1.024002\mu \|\delta_D\|_s$;
- (6) $\|H\delta_D + \delta_P\|_s^* \leq 2.039813\mu \|\delta_D\|_s$;
- (7) $\|H_1\delta_D - H\delta_D\|_s^* \leq 0.540897\mu \|\delta_D\|_s^2$;
- (8) $\|H_1\delta_D - \delta_P\|_s^* \leq 1.065087\mu \|\delta_D\|_s^2$;
- (9) $\|H_1\delta_D\|_s^* \leq 1.024263\mu \|\delta_D\|_s$;
- (10) $\|H_1\delta_D + \delta_P\|_s^* \leq 2.048265\mu \|\delta_D\|_s$;
- (11) $\langle \delta_D, H_1\delta_D \rangle \geq 0.944161\mu \|\delta_D\|_s^2$.

Proof. (1) This was proven in Theorem B.1, part (4).

(2) This follows from the Dikin ellipsoid bound; $H \succeq 0.968994\mu F''_*(s)$.

(3) Notice that

$$\langle \delta_D, \delta_P \rangle = \langle \delta_D, H\delta_D \rangle + \langle \delta_D, \delta_P - H\delta_D \rangle.$$

We bound the second term by Cauchy-Schwarz:

$$\langle \delta_D, \delta_P - H\delta_D \rangle \leq \|\delta_D\|_s \|H\delta_D - \delta_P\|_s^* \leq 0.524190\mu \|\delta_D\|_s^3.$$

Using this with the bound from the previous part gives the advertised inequality.

(4) We compute, using Lemma 7.2

$$\begin{aligned} \|H\delta_D\|_s^* &= \mu \left\langle \delta_D, F''_*(\check{s}) (F''_*(s))^{-1} F''_*(\check{s}) \delta_D \right\rangle^{1/2} \leq 1.015811\mu \langle \delta_D, F''_*(s) \delta_D \rangle^{1/2} \\ &= 1.015811\mu \|\delta_D\|_s, \end{aligned}$$

as advertised.

(5) We use the triangle inequality followed by parts (1) and (4):

$$\|\delta_P\|_s^* \leq \|H\delta_D\|_s^* + \|\delta_P - H\delta_D\|_s^* \leq 1.015811\mu \|\delta_D\|_s + 0.524190\mu \|\delta_D\|_s^2 \leq 1.024002\mu \|\delta_D\|_s.$$

(6) We use the triangle inequality, part (4) and the bound $\|\delta_D\|_s \leq 1/64$:

$$\begin{aligned} \|H\delta_D + \delta_P\|_s^* &\leq 2\|H\delta_D\|_s^* + \|H\delta_D - \delta_P\|_s^* \leq 2.031622\mu \|\delta_D\|_s + 0.524190\mu \|\delta_D\|_s^2 \\ &\leq 2.031622\mu \|\delta_D\|_s + (52419/6400000)\mu \|\delta_D\|_s, \end{aligned}$$

which is the claimed bound.

(7) Recall that $\langle \delta_D, x \rangle = 0$, so

$$H\delta_D - H_1\delta_D = \frac{\langle s, H\delta_D \rangle}{\langle s, Hs \rangle} Hs.$$

Now, we bound using $\langle s, \delta_P \rangle = 0$, triangle inequality and part (1):

$$|\langle s, H\delta_D \rangle| = |\langle s, \delta_P \rangle + \langle s, H\delta_D - \delta_P \rangle| \leq 0 + \|s\|_s \|H\delta_D - \delta_P\|_s^* \leq 0.524190\sqrt{\vartheta}\mu\|\delta_D\|_s^2$$

and recall (Lemma B.2 part (7))

$$\langle s, Hs \rangle \geq 0.984436\vartheta\mu$$

and (Lemma B.2 part (5))

$$\|Hs\|_s^* \leq 1.015811\sqrt{\vartheta}\mu.$$

Thus,

$$\|H_1\delta_D - H\delta_D\|_s^* \leq 0.540897\mu\|\delta_D\|_s^2.$$

(8) We use the triangle inequality followed by parts (1) and (7).

(9) We use the triangle inequality followed by parts (4), (7) and the fact that $\|\delta_D\|_s \leq 1/64$.

(10) We use the triangle inequality and parts (5) and (9).

(11) We compute, using previous parts of this lemma,

$$\begin{aligned} \langle \delta_D, H_1\delta_D \rangle &= \langle \delta_D, \delta_P \rangle + \langle \delta_D, H_1\delta_D - \delta_P \rangle \\ &\geq 0.960803\mu\|\delta_D\|_s^2 - \|\delta_D\|_s \|H_1\delta_D - \delta_P\|_s^* \\ &\geq 0.960803\mu\|\delta_D\|_s^2 - 1.065087\mu\|\delta_D\|_s^3 \\ &\geq 0.944161\mu\|\delta_D\|_s^2. \end{aligned}$$

□

APPENDIX C. EVALUATING THE PRIMAL INTEGRAL SCALING

Given an oracle that can:

- Evaluate $F(x)$, $F'(x)$, and $F''(x)$ for any $x \in \text{int}(K)$, and
- Compute, in some explicit form, the univariate polynomial $t \mapsto \exp(-F(x + td))$ for any $x \in \text{int}(K)$ and $d \in \mathbb{R}^n$,

we describe how to compute the primal integral scaling

$$\left(\mu \int_0^1 F''(x - t\delta_P) dt \right)^{-1}$$

exactly. We do not claim that this method is practical or useful; in particular, it requires ϑ evaluations of F'' . However, we later describe a slightly more practical variant that admits concrete bounds on approximation error.

The method is a straightforward application of the theory of Gaussian quadrature. A reader unfamiliar with Gaussian quadrature might consult Section 3.6 of the excellent book by Stoer and Bulirsch [43].

Notice that, with $F = -\ln p$, we have

$$F'' = \frac{p'(p')^\top - pp''}{p^2}.$$

The denominator is a polynomial of degree 2ϑ and the numerator is a matrix whose entries are polynomials of degree $2\vartheta - 2$.

Theorem C.1. *There exist ϑ points r_1, \dots, r_ϑ in $[0, 1]$ and associated weights w_1, \dots, w_ϑ such that, if q is a univariate polynomial of degree less than 2ϑ , then*

$$\int_0^1 \frac{q(t)}{p^2(x - t\delta_P)} dt = \sum_{i=1}^{\vartheta} w_i q(r_i).$$

Proof. Notice that $1/p^2(x - t\delta_P)$ is positive and bounded for t in $[0, 1]$. In particular, it is measurable, all moments exist and are finite, and any nonnegative polynomial q such that $\int_0^1 q(t)/p^2(x - t\delta_P) dt = 0$ is itself zero. Thus a Gaussian quadrature rule with weight function $1/p^2(x - t\delta_P)$ exists. That is, there exist ϑ points r_1, \dots, r_ϑ in $[0, 1]$ and associated weights w_1, \dots, w_ϑ such that, for any $C^{2\vartheta}$ function f ,

$$\left| \int_0^1 \frac{f(t)}{p^2(x - t\delta_P)} dt - \sum_{i=1}^{\vartheta} w_i f(r_i) \right| \leq \frac{f^{(2\vartheta)}(\tau)}{(2\vartheta)!} C$$

for some $0 \leq \tau \leq 1$ and some constant $C \geq 0$ dependent on ϑ .

Take $f = q$; the derivative of order 2ϑ vanishes and hence the difference must be zero. \square

Indeed, the entries of $p'(p')^\top - pp''$ have degree $2\vartheta - 2$; the primal integral scaling is exactly

$$\left(\mu \sum_{i=1}^{\vartheta} w_i (p'(r_i)(p'(r_i))^\top - p(r_i)p''(r_i)) \right)^{-1}.$$

It may be practical to use a well-known Gaussian quadrature rule instead of the one arising from $1/p^2$. The following theorem considers Gauss-Legendre quadrature of fixed order:

Theorem C.2. *If $1 \leq k$ is an integer and $\|\delta_P\|_x \leq 1$, then*

$$\left\| T_P^{-2} - \mu \sum_{i=1}^k w_i^L F''(x - r_i^L \delta_P) \right\|_x \leq \frac{1}{(1 - \|\delta_P\|_x)^2} \max_{t \in [0,1]} 2 \|\delta_P\|_{x-t\delta_P}^{2k} \|F''(x - t\delta_P)\|_x^{k/2}$$

where r_i^L are the order- k Gauss-Legendre nodes and w_i^L are the associated weights.

Proof. Note that

$$\left\| T_P^{-2} - \mu \sum_{i=1}^k w_i^L F''(x - r_i^L \delta_P) \right\|_x = \max_{\|h\| \leq 1} \left| T_P^{-2}[h, h] - \mu \sum_{i=1}^k w_i^L F''(x - r_i^L \delta_P)[h, h] \right|.$$

Thus let h be the point at which this maximum is attained. By the error bound for Gaussian quadrature ([43], Theorem 3.6.24),

$$(11) \quad \left| T_P^{-2}[h, h] - \mu \sum_{i=1}^k w_i^L F''(x - r_i^L \delta_P)[h, h] \right| \leq \max_{t \in [0,1]} \left| \frac{F^{(2k+2)}(x - t\delta_P)[\delta_P, \dots, \delta_P, h, h] \langle p_k, p_k \rangle}{(2k)!} \right|,$$

where p_k is the k th Legendre polynomial.

A theorem of Güler ([6], Theorem 4.2), together with the results from Appendix 1 of Nesterov and Nemirovskii's book [29], shows that

$$(12) \quad |F^{(2k+2)}(x - t\delta_P)[\delta_P, \dots, \delta_P, h, h]| \leq (2k+1)! \|\delta_P\|_{x-t\delta_P}^{2k} \|h\|_{x-t\delta_P}^2.$$

Using self-concordance, we bound

$$\|h\|_{x-t\delta_P} \leq \frac{1}{1 - t\|\delta_P\|_x} \|h\|_x = \frac{1}{1 - t\|\delta_P\|_x}.$$

The squared L_2 norm of the k th Legendre polynomial is $2/(2k+1)$. Substituting these and (12) into (11), we get the advertised bound. \square