"CONE-FREE" PRIMAL-DUAL PATH-FOLLOWING AND POTENTIAL REDUCTION POLYNOMIAL TIME INTERIOR-POINT METHODS

ARKADI NEMIROVSKI* AND LEVENT TUNÇEL †

Abstract. We present a framework for designing and analyzing primal-dual interior-point methods for convex optimization. We assume that a self-concordant barrier for the convex domain of interest and the Legendre transformation of the barrier are both available to us. We directly apply the theory and techniques of interior-point methods to the given good formulation of the problem (as is, without a conic reformulation) using the very usual primal central path concept and a less usual version of a dual path concept. We show that many of the advantages of the primal-dual interior-point techniques are available to us in this framework and therefore, they are not intrinsically tied to the conic reformulation and the logarithmic homogeneity of the underlying barrier function.

Key words. convex optimization, interior-point methods, primal-dual algorithms, self-concordant barriers

AMS subject classifications. 90C51, 90C25, 65Y20, 90C28, 49D49

1. Introduction.

In what follows, we are interested in solving the optimization problem

$$c_* \equiv \inf_{x \in D} \langle c, x \rangle_{\mathcal{E}}, \tag{1.1}$$

where D is an open convex domain in an Euclidean space $\mathcal{E}$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{E}}$. What we intend to use, is a kind of a primal-dual interior-point method. With the traditional conic approach, in order to solve (1.1) by a primal-dual *path-following* method, we would act as follows.

1) We represent the feasible domain D of the problem as the inverse image of the interior of a closed pointed cone $K \subset \mathcal{F}$ under the affine embedding $x \mapsto Ax - b$ of $\mathcal{E}$ into an Euclidean space $\mathcal{F}$:

$$D = \{x : (Ax - b) \in \operatorname{int} K\}, \tag{1.2}$$

thus reformulating (1.1) as the conic problem

$$\min_{\xi} \{\langle d, \xi \rangle_{\mathcal{F}} : \xi \in (\mathcal{L} - b) \cap K\}, \tag{1.3}$$

where $\mathcal{L} = \operatorname{Im} A$ and d is such that $A^* d = c$;

2) we associate with (1.3) the dual problem

$$\max_{y} \left\{\langle b, y \rangle_{\mathcal{F}} : y \in (\mathcal{L}^\perp + d) \cap K_*\right\}, \tag{1.4}$$

where K_* is the cone dual to K (without loss of generality, we can assume $b \in \mathcal{L}^\perp$ and $d \in \mathcal{L}$);

3) we equip K with a ϑ-self-concordant logarithmically homogeneous barrier $H(\cdot)$ with known Legendre transformation $H_*(\cdot)$; the function $H^*(y) = H_*(-y)$ is a ϑ-self-concordant logarithmically homogeneous barrier for K_*;

4) we trace, as $t \to \infty$, the *primal-dual central path* $(\xi_*(t), y_*(t))$ defined by the requirements

$$\begin{array}{rcl} \xi_*(t) & = & \operatorname*{argmin}_{\xi} \{t\langle d, \xi \rangle_{\mathcal{F}} + H(\xi) : \xi \in (\mathcal{L} - b) \cap \operatorname{int} K\}, \\ y_*(t) & = & \operatorname*{argmin}_{y} \left\{-t\langle b, y \rangle_{\mathcal{F}} + H^*(y) : y \in (\mathcal{L}^\perp + d) \cap \operatorname{int} K_*\right\}. \end{array} \tag{1.5}$$

When primal-dual *potential reduction* methods are used, at step 4) we, rather than tracing the primal-dual central path, reduce step by step the *primal-dual potential*

$$S(\xi, y) = H(\xi) + H^*(y) + (\vartheta + \sqrt{\vartheta}) \ln(\langle y, \xi \rangle_{\mathcal{F}}), \tag{1.6}$$

*Faculty of IE&M, Technion, Haifa, Israel (nemirovs@ie.technion.ac.il). Part of the research was done while the author was a Visiting Professor at the University of Waterloo.

†Dept. of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Waterloo, Canada (ltuncel@math.waterloo.ca). Research of this author is supported in part by a PREA from Ontario and by a NSERC Discovery Grant.

keeping ξ and y feasible for the respective problems (1.3), (1.4).

Note that *all we are interested in is the original problem* (1.1), *not the primal-dual pair* (1.3), (1.4), and *in principle* we could solve (1.1) by interior-point methods "as the problem is", provided that we can equip D with a ϑ-self-concordant barrier $F(x)$. Indeed, given such a barrier, we could trace as $t \to \infty$ the *primal central path*

$$x_*(t) = \operatorname*{argmin}_x \{F_t(x) \equiv t\langle c, x\rangle_{\mathcal{E}} + F(x) : x \in D\} \tag{1.7}$$

or reduce step by step the "primal potential"

$$s(x,t) = [F_t(x) - \min_{z\in D} F_t(z)] - \sqrt{\vartheta}\ln t. \tag{1.8}$$

In fact, the primal-dual techniques can be interpreted as no more than some particular cases of the latter "straight-forward" approach. Indeed, given a "primal-dual frame" $A, b, K, \vartheta, H(\cdot)$, we can set $F(x) = H(Ax - b)$, thus getting a ϑ-self-concordant barrier for $\mathrm{cl}D$. With this F, the path (1.7) exists if and only if the primal-dual central path exists, and the latter path is readily given by the former one:

$$\xi_*(t) = Ax_*(t) - b; \quad y_*(t) = -t^{-1}H'(Ax_*(t) - b),$$

so that it, basically, is the same — to trace the primal central path, or the primal-dual one. One important advantage of the primal-dual path-following framework, at least in its theoretical aspects, comes partly from the fact that in this framework it is easy to realize whether a given primal-dual pair (ξ, y) is close to a given "target pair" $(\xi_*(t), y_*(t))$ on the primal-dual central path, which allows for theoretically valid long-step path-following policies (see [13]). In contrast to this, in the "purely primal" framework it seems to be impossible to realize, at a low computational cost, whether a given primal solution x is close to a given target point $x_*(t)$ on the path; as a result, all known theoretically efficient purely primal path-following methods are enforced to use the worst-case-oriented short-step policy. The situation with potential reduction techniques is similar. Indeed, given a primal-dual frame $A, b, K, \vartheta, H(\cdot)$, let us equip $\mathrm{cl}D$ with the ϑ-self-concordant barrier $F(x) = H(Ax - b)$, and consider the function

$$P(x, y, t) = H(Ax - b) + H^*(y) + t\langle y, Ax - b\rangle_{\mathcal{F}} - (\vartheta + \sqrt{\vartheta})\ln t,$$

where y is restricted to satisfy the relation $A^*y = c$. This function in a way "contains" both the primal-dual potential (1.6) and the primal potential (1.8); it is easily seen that

$$S(Ax - b, y) = \min_{t>0} P(x, y, t) + \mathrm{const}, \quad s(x,t) = \min_{y:A^*y=c} P(x, y, t).$$

Thus, we can say that both in the primal-dual and in the (conceptual) primal potential reduction methods we are pushing the potential $P(\cdot)$ to $-\infty$, keeping x feasible for (1.1) and y feasible for (1.4). Here, the advantages of the primal-dual framework become even more apparent than in the path-following case: the primal-dual potential S is explicitly computable, while this is not so for the primal potential $s(x,t)$ (this is why the "primal potential reduction" method is a conceptual, not a computational one).

The goal of this paper is to demonstrate that *the outlined advantages of the primal-dual interior-point techniques are not intrinsically related to conic reformulation of the original problem and logarithmic homogeneity of the barriers underlying the interior-point methods.* Specifically, it turns out that we can build "good analogies" of the path-following and the potential reduction primal-dual interior-point techniques in the following

> <u>"Good Case":</u> We can equip the domain of the problem of interest (1.1) with a self-concordant barrier
>
> $$F(x) = \Phi(Ax - b)$$
>
> which is obtained, via affine substitution of argument, from a ϑ-self-concordant barrier $\Phi(\cdot)$ *with known Legendre transformation* $\Phi_*(\cdot)$.

The difference with the traditional primal-dual framework is that we do *not* require Φ to be logarithmically homogeneous self-concordant barrier for a cone, and this indeed is a not negligible difference. As an example, consider a Geometric Programming problem:

$$\begin{gathered}\min_x \left\{ c^T x : f_i(x) \leq 0,\ i = 1, ..., m,\ Bx \leq b \right\}, \\ f_i(x) \equiv \ln\left(\sum_{\ell=1}^{L} \alpha_{i\ell} \exp\{d_{i\ell}^T x\}\right) + e_i^T x + \beta_i\end{gathered} \tag{1.9}$$

where $\alpha_{i\ell} > 0$ for all i, ℓ. Assuming the problem strictly feasible, the interior D of the feasible set can be easily represented in the form

$$D = \{x : (Ax - b) \in \mathcal{D}\}, \quad \mathcal{D} = \{(t, y, s) \in \mathbf{R}^p \times \mathbf{R}^q \times \mathbf{R}^q : t_i > 0,\ i = 1, ..., p, \exp\{y_i\} < s_i,\ i = 1, ..., q\}.$$

Now, the set cl$\mathcal{D}$ admits $(p + 2q)$-self-concordant barrier

$$\Phi(t, y, s) = -\sum_{i=1}^{p} \ln t_i - \sum_{i=1}^{q} [\ln(\ln(s_i) - y_i) + \ln s_i] \tag{1.10}$$

(see [14], Section 5.3.2). One can immediately compute the Legendre transformation of Φ:

$$\Phi_*(\tau, \eta, \sigma) = -(p + 2q) - \sum_{i=1}^{p} \ln(-\tau_i) - \sum_{i=1}^{q} \left[(\eta_i + 1) \ln\left(\frac{-\sigma_i}{\eta_i + 1}\right) + \ln \eta_i + \eta_i \right] \tag{1.11}$$

(from now on, if otherwise is not stated, all functions are $+\infty$ outside of their natural domains). Note that Φ is *not* a logarithmically homogeneous self-concordant barrier for a cone.

It should be mentioned that *in principle* the possibility to get rid of cones and logarithmical homogeneity is not *that* big of a deal: it is known (see [14], Proposition 5.1.4) that a ϑ-self-concordant barrier $\Phi(x)$ can be associated with a *$(\kappa\vartheta)$-logarithmically homogeneous* barrier $\Phi^+(x, t) = \kappa\left[\Phi(x/t) - \vartheta \ln(t)\right]$ (κ is an appropriate absolute constant — for instance, $25\left[\Phi(x/t) - 7\vartheta \ln(t)\right]$ works for every Φ, see [6]); note that the original barrier Φ, up to absolute constant factor, can be obtained from Φ_+ by an affine substitution of the argument. Further, if $\Phi_*(\cdot)$ is available, then it is not that difficult to compute $\Phi_*^+(\xi, \tau)$:

$$\Phi_*^+(\xi, \tau) = \max_{t > 0} \left[\kappa \Phi_*(t\xi/\kappa) + \tau t + \kappa\vartheta \ln t\right]. \tag{1.12}$$

In particular, in the Geometric Programming case we could, *in principle*, associate with the barrier (1.10) a logarithmically homogeneous barrier, thus getting a $\kappa(p+2q)$-logarithmically homogeneous barrier Φ_+ with "nearly explicitly computable" Legendre transformation and such that $\Phi_+(A_+x - b_+)$ is a barrier for the feasible set of (1.9). With this barrier, we can solve (1.9) by the standard conic, primal-dual technique. In light of these observations, a question arises: what could be the advantages of new methods we intend to propose, given that the applications covered by these methods can be covered by the standard conic, primal-dual techniques as well? Our answer to this question is that "to enforce" the standard conic framework, when the problem in the original form does not fit this framework, can be computationally costly: one-dimensional minimization in (1.12) is perhaps not too expensive, but for sure is not costless. And it is absolutely unclear in advance why the primal-dual techniques we intend to develop should be that inferior as compared to the standard conic ones to justify "enforcement" of the standard techniques. It should be added that, at the time of this writing, there is neither clear theoretical reasons (perhaps with the exception of [18]), nor computational experience in favour of the standard primal-dual interior-point techniques *beyond the scope of problems on self-scaled cones*, i.e., beyond the scope of Linear, Conic Quadratic and Semidefinite Programming.

Note that the above "good case" was already considered in [15], where the long-step path-following (in fact, "surface-following") interior-point methods for this case were proposed. Below, we investigate in much more detail the primal-dual framework associated with Good Case, with emphasis on developing the associated potential reduction techniques.

The rest of the paper is organized as follows. In Section 2, we introduce some notation and outline a number of basic facts on self-concordance which will be frequently used in the sequel. In Section 3, we describe our "cone-free" primal-dual framework and introduce and investigate the main ingredients of our approach – primal-dual path, proximity measure and potential. In Section 4, we analyze centering and path-tracing directions. In Sections 5 and 6 we use the preceding results to develop path-following, resp., potential reduction "cone-free" primal-dual methods and to analyze their complexity. Section 7 contains a discussion of possible applications and extensions.

2. Preliminaries on self-concordant functions. We start by summarizing the properties of self-concordant functions and barriers we will frequently use in the sequel; for the proofs, see [14].

2.1. Notation. In what follows letters like $\mathcal{E}$, $\mathcal{F}$, etc., denote Euclidean linear spaces; corresponding inner products are denoted $\langle\cdot,\cdot\rangle_{\mathcal{E}}$, $\langle\cdot,\cdot\rangle_{\mathcal{F}}$. We skip subscripts in $\langle\cdot,\cdot\rangle$, when it is clear from the context what the Euclidean space in question is.

For a linear operator $x \mapsto Bx : \mathcal{F} \to \mathcal{E}$, B^* stands for the conjugate operator: $\langle y, Bx\rangle_{\mathcal{E}} = \langle B^* y, x\rangle_{\mathcal{F}}$. We write $B \succeq 0$ ($B \succ 0$) to express that B is a symmetric and positive semidefinite (resp., positive definite) operator on E, with evident interpretation of relations like $A \succeq B$ or $B \prec A$.

We associate with an operator $B \succ 0$ on $\mathcal{E}$ a conjugate pair of Euclidean norms on $\mathcal{E}$:

$$\begin{array}{rcl} \|x\|_B &=& \langle x, Bx\rangle^{1/2}, \\ \|x\|_B^* &=& \max\{\langle x, y\rangle : \|y\|_B \le 1\} = \|x\|_{B^{-1}}. \end{array}$$

From now on, we set

$$\begin{array}{rcl} \rho(t) &=& t - \ln(1+t) \\ \omega(t) &=& \rho(-t) - \frac{t^2}{2} \end{array} \quad \begin{bmatrix} = \frac{t^2}{2}(1+o(t)),\ t \to 0 \\ = \frac{t^3}{3}(1+o(t)),\ t \to 0 \end{bmatrix},$$

and

$$\sigma(s) = \max\{t : \rho(t) \le s\},\ s \ge 0;$$

it is easily seen that

$$\sigma(s) \le \sqrt{2s} + s. \tag{2.1}$$

For a convex C^2 on its domain and nondegenerate ($f'' \succ 0$) function $f : \mathcal{E} \mapsto \mathbf{R} \cup \{+\infty\}$ and $x \in \mathrm{Dom} f$, we define the *Newton decrement of f at x* as

$$\lambda(f, x) = \|f'(x)\|^*_{f''(x)}.$$

2.2. Self-concordant functions and barriers: definitions. A convex function $f : \mathcal{E} \to \mathbf{R} \cup \{+\infty\}$ is called *self-concordant* (s.c.), if the domain Q of f is open, f is C^3 on Q, satisfies the differential inequality

$$\left| \frac{d^3}{dt^3}\bigg|_{t=0} f(x+th) \right| \le 2 \left(\frac{d^2}{dt^2}\bigg|_{t=0} f(x+th) \right)^{3/2} \quad \forall (x \in Q, h \in \mathcal{E}) \tag{2.2}$$

and is a barrier for Q: $f(x_i) \to \infty$ along every sequence $\{x_i\} \subset Q$ converging to a boundary point of Q.

A s.c. function f is called *nondegenerate*, if its Hessian $f''(x)$ is nondegenerate at some (and then automatically at every) point $x \in \mathrm{Dom} f$.

Let $\vartheta \ge 1$. Function f is called *ϑ-self-concordant barrier* (ϑ-s.c.b.) *for* $\mathrm{cl}\,\mathrm{Dom} f$, if f is self-concordant and

$$\left| \frac{d}{dt}\bigg|_{t=0} f(x+th) \right| \le \sqrt{\vartheta} \left(\frac{d^2}{dt^2}\bigg|_{t=0} f(x+th) \right)^{1/2} \quad \forall (x \in \mathrm{Dom} f, h \in \mathcal{E}). \tag{2.3}$$

A nondegenerate s.c. function f is ϑ-s.c.b. if and only if $\lambda(f, x) \le \sqrt{\vartheta}$ for all $x \in \mathrm{Dom} f$.

2.3. Basic properties of self-concordant functions. We summarize these properties in the following list.

SC.I. [Stability w.r.t. linear operations]

1) *Let f_i, $i = 1,..,m$, be s.c. functions on $\mathcal{E}$, and let $\lambda_i \geq 1$. Then the function $f = \sum_i \lambda_i f_i$ is s.c. If every f_i is ϑ_i-s.c.b., then f is $(\sum_i \lambda_i \vartheta_i)$-s.c.b.*

2) *Let f be s.c. on $\mathcal{E}$, and let $y \mapsto Ay + b$ be an affine mapping from Euclidean space $\mathcal{F}$ to $\mathcal{E}$ with image intersecting* $\mathrm{Dom} f$. *Then the function $g(y) = f(Ay + b)$ is s.c. If f is ϑ-s.c.b., then so is g.*

SC.II. [Local behaviour and damped Newton step] *Let f be a nondegenerate s.c. function with $Q = \mathrm{Dom} f$. Then*

1) *For every $x \in Q$, the ellipsoid $\{y : \|y - x\|_{f''(x)} < 1\}$ is contained in Q. Besides this,*

$$\begin{array}{rcll} r \equiv \|y - x\|_{f''(x)} < 1 & \Rightarrow & (1-r)^2 f''(x) \preceq f''(y) \preceq (1-r)^{-2} f''(x) & (a) \\ r \equiv \|y - x\|_{f''(x)} < 1 & \Rightarrow & f(y) \leq f(x) + \langle f'(x), y - x\rangle + \rho(-r) & (b.1) \\ y \in Q, r \equiv \|y - x\|_{f''(x)} & \Rightarrow & f(y) \geq f(x) + \langle f'(x), y - x\rangle + \rho(r). & (b.2) \end{array} \tag{2.4}$$

2) *For $x \in Q$, we define the damped Newton iterate of x as*

$$x_+ = x - \frac{1}{1 + \lambda(f, x)} [f''(x)]^{-1} f'(x).$$

For every $x \in Q$ we have

$$\begin{array}{rcll} x_+ & \in & Q & (a) \\ f(x_+) & \leq & f(x) - \rho(\lambda(f, x)) & (b) \\ \lambda(f, x_+) & \leq & 2\lambda^2(f, x). & (c) \end{array} \tag{2.5}$$

SC.III. [Minima of s.c. functions] *Let f be a nondegenerate s.c. function. f attains minimum on* $\mathrm{Dom} f$ *if and only if f is bounded below, and if and only if there exists $x \in \mathrm{Dom} f$ with $\lambda(f, x) < 1$. The minimizer x_f of f, if it exists, is unique, and*

$$\lambda(f, x) < 1 \Rightarrow f(x) - f(x_f) \leq \rho(-\lambda(f, x)). \tag{2.6}$$

SC.IV. [Additional properties of s.c.b.'s] *Let f be a nondegenerate ϑ-s.c.b., and let $Q = \mathrm{Dom} f$. Then*

1) *one has*

$$\begin{array}{rl} \forall (x, y \in Q): \quad \langle y - x, f'(x)\rangle \leq \vartheta & (a) \\ \forall (x, y \in Q): \langle y - x, f'(x)\rangle \geq 0 \Rightarrow \|y - x\|_{f''(x)} \leq \vartheta + 2\sqrt{\vartheta} & (b) \end{array} \tag{2.7}$$

2) *f is bounded below on Q if and only if Q is bounded, and in this case*

$$\{y : \|y - x_f\|_{f''(x_f)} < 1\} \subset Q \subset \{y : \|y - x_f\|_{f''(x_f)} < \vartheta + 2\sqrt{\vartheta}\}. \tag{2.8}$$

SC.V. [Legendre transformation of a s.c. function] *Let f be a nondegenerate s.c. function on $\mathcal{E}$.*

1) *The domain of the Legendre transformation*

$$f_*(\xi) = \sup_x [\langle \xi, x\rangle - f(x)]$$

is exactly the image of $\mathrm{Dom} f$ *under the mapping $x \mapsto f'(x)$, f_* is nondegenerate s.c. function, and the Legendre transformation of f_* is f.*

2) *If f is nondegenerate ϑ-s.c.b, then* $\mathrm{Dom} f_*$ *is either the entire $\mathcal{E}$ – this is the case if and only if* $\mathrm{Dom} f$ *is bounded – or the open cone*

$$\{\xi : \langle \xi, h\rangle < 0, \quad \forall (h \in R, h \neq 0)\},$$

where R is the recession cone of $\mathrm{Dom} f$.

3) *If f is ϑ-self-concordant logarithmically homogeneous barrier, i.e.,* Dom$f$ *is the interior of a pointed closed convex cone K and*

$$f(tx) = f(x) - \vartheta \ln t \quad \forall (x \in \mathrm{Dom} f, t > 0),$$

then f_* is ϑ-s.c. logarithmically homogeneous barrier with

$$\mathrm{Dom} f_* = -\mathrm{int}\, K_*,$$

where K_* is the cone dual to K:

$$K_* \equiv \{\xi : \langle \xi, x \rangle \geq 0, \quad \forall x \in K\}.$$

3. Path, proximity measure and potential.

3.1. The setup. As was indicated in Introduction, we intend to consider the situation as follows. We are given

- a nondegenerate ϑ-s.c.b. Φ with a domain $D^+ \subset \mathcal{F}$ and the Legendre transformation Φ_* of Φ (which is a nondegenerate s.c. function, SC.V.1)); the domain of Φ_* is denoted D_*^+. By SC.V.2), D_*^+ is a conic set:

$$(3.1) \qquad y \in D_*^+ \Rightarrow \tau y \in D_*^+ \quad \forall \tau > 0$$

- a linear embedding $x \mapsto Ax : \mathcal{E} \to \mathcal{F}$ ($\mathrm{Ker} A = \{0\}$) with the image intersecting D^+;
- a vector $c \in \mathcal{E}$, $c \neq 0$.

These data define

- the optimization problem

$$(3.2) \qquad c_* = \inf_x \{\langle c, x \rangle : x \in D\}, \quad D = \{x : Ax \in D^+\},$$

  we are interested in solving;
- the function $F(x) = \Phi(Ax)$ which is a nondegenerate ϑ-self-concordant barrier for cl$D$ (SC.I.2)).

REMARK 3.1. *1. In Introduction, we spoke about the affine mapping $x \mapsto (Ax - b)$ rather than about the linear mapping $x \mapsto Ax$; of course, this does not restrict generality, since a shift in the mapping is equivalent to translating the barrier Φ.*

2. To compare our constructions to follow, with the standard primal-dual interior-point constructions, let us specify the Standard case as the one where

$$\Phi(\xi) = H(\xi - b)$$

for a ϑ-logarithmically homogeneous s.c.b. $H(\cdot)$. Note that in this case

$$(3.3) \qquad \Phi_*(y) = H_*(y) + \langle y, b \rangle = H^*(-y) + \langle y, b \rangle.$$

3.2. Primal and dual paths. The major entity of our interest is the *primal path*

$$(3.4) \qquad x_*(t) = \operatorname*{argmin}_x F_t(x), \quad F_t(x) = F(x) + t\langle c, x \rangle,$$

and we are interested in this path to be well-defined for all $t > 0$. By SC.III, this is the case if and only if $F_t(\cdot)$ is below bounded for every $t > 0$. The corresponding condition can be stated as follows.

LEMMA 3.1. *Let $t > 0$. The function $F_t(x) = F(x) + t\langle c, x \rangle$ is bounded below if and only if $\exists y \in D_*^+ : A^* y = -c$. In particular,*

– either (case A) $F_t(\cdot)$ is bounded below for every $t > 0$,

– or (case B) $F_t(\cdot)$ is unbounded below for every $t > 0$.

Proof. If $F_t(x)$ is bounded below, then the function attains its minimum at a unique point $x_*(t)$ (SC.III). We have $A^*\Phi'(Ax) = F'(x_*(t)) = -tc$ and $z = \Phi'(Ax) \in D_*^+$, whence $y = t^{-1}z \in D_*^+$ by (3.1); thus, $\exists y \in D_*^+ : A^*y = -c$. Vice versa, let $y \in D_*^+$ be such that $A^*y = -c$, and let $t > 0$. Setting $z = ty$ and applying (3.1), we get $z \in D_*^+$ and $A^*z = -tc$. We now have

$$F_t(x) = \Phi(Ax) + t\langle c, x\rangle = \Phi(Ax) - \langle z, Ax\rangle \geq -\Phi_*(z),$$

so that $F_t(x)$ is bounded below. ■

From now on, we assume that case A takes place, so that the primal central path (3.4) is well-defined for all $t > 0$.

REMARK 3.2. *In the Standard case, the assumptions that $D \neq \emptyset$ and that case A takes place are equivalent to strict primal-dual feasibility of the primal-dual pair* (1.3), (1.4) *associated with* (3.2).

We associate with the primal path $x_*(t)$ the *dual path*

$$y_*(t) = \Phi'(Ax_*(t)),\ t > 0. \tag{3.5}$$

LEMMA 3.2. *For $t > 0$, the "primal-dual pair" $(x, y) = (x_*(t), y_*(t))$ is uniquely defined by the relations*

$$\begin{array}{ll} (a) & y \in D_*^+, \quad x \in D \\ (b) & A^*y = -tc \\ (c) & \Phi_*'(y) = Ax \quad [\Leftrightarrow y = \Phi'(Ax)]\,. \end{array} \tag{3.6}$$

Moreover,

$$y_*(t) = \operatorname*{argmin}_y\{\Phi_*(y) : A^*y = -tc\}. \tag{3.7}$$

Proof. Let $x = x_*(t)$, $y = y_*(t)$. Then (x, y) clearly satisfies (a) and (c); besides this, $-tc = F'(x) = A^*\Phi'(Ax) = A^*y$, so that (x, y) satisfies (b).

Now let (x, y) satisfy (3.6). Then $F'(x) = A^*\Phi'(Ax) = A^*y = -tc$ (we have used (c) and (b)), i.e., $x = x_*(t)$. Now from (c) it follows that $y = y_*(t)$.

To prove (3.7), note that, as we already know, $A^*y_*(t) = -ct$ and $\Phi_*'(y_*(t)) = Ax_*(t)$, i.e., $\Phi_*'(y_*(t)) = Ax_*(t)$ is orthogonal to the kernel of A^*. ■

REMARK 3.3. *It is immediately seen that in the Standard case (see Remark 3.1), $x_*(t)$ and $(Ax_*(t) - b, -t^{-1}y_*(t))$ are exactly what was called in Introduction "primal central path" and "primal-dual central path", respectively.*

3.2.1. Optimality gap. The role of the standard expression for the duality gap is now played by the following statement:

LEMMA 3.3. *Let $y \in D_*^+$ be such that $A^*y = -tc$. Then*

$$c_* \equiv \inf_{x' \in D}\langle c, x'\rangle \geq -\frac{\vartheta + \langle y, \Phi_*'(y)\rangle}{t} \tag{3.8}$$

and therefore

$$\forall (x \in D): \quad \langle c, x\rangle - c_* \leq t^{-1}\left[\vartheta + \langle y, \Phi_*'(y)\rangle - \langle y, Ax\rangle\right]. \tag{3.9}$$

REMARK 3.4. *In the Standard case (see Remark 3.1), it is immediately seen that vectors $y \in D_*^+$ such that $A^*y = -tc$ are exactly the vectors of the form $-t\widehat{y}$, where $\widehat{y}$ is a feasible solution to the conic dual* (1.4) *of our problem of interest* $\min\limits_x\{\langle c, x\rangle : (Ax - b) \in \mathrm{Dom}H\}$. *Besides this, in the Standard case*

$$\langle y, \Phi_*'(y)\rangle = \langle y, b\rangle + \langle y, H_*'(y)\rangle = \langle y, b\rangle - \vartheta.$$

Thus, in the Standard case (3.8) *reads*

$$\forall (x \in D, \widehat{y} \in \mathrm{Dom} H^*, A^*\widehat{y} = c): \quad \langle c, x\rangle - c_* \le \langle \hat{y}, Ax - b\rangle$$

which is the standard result on the duality gap in Conic Duality.
Proof of Lemma 3.3. Let $z = \Phi_*'(y)$, so that $y = \Phi'(z)$. For $x' \in D$ we have

$$\begin{array}{rcl} -t\langle c, x'\rangle & = & \langle y, Ax'\rangle = \langle \Phi'(z), Ax'\rangle = \langle \Phi'(z), Ax' - z\rangle + \langle \Phi'(z), z\rangle \\ & \le & \vartheta + \langle \Phi'(z), z\rangle \\ & & \qquad\qquad [\text{by } (2.7.a)] \\ & = & \vartheta + \langle y, \Phi_*'(y)\rangle, \end{array}$$

whence

$$\inf_{x' \in D} \langle c, x'\rangle \ge -\frac{\vartheta + \langle y, \Phi_*'(y)\rangle}{t}$$

and therefore, in view of $\langle c, x\rangle = -t^{-1}\langle A^* y, x\rangle = -t^{-1}\langle y, Ax\rangle$,

$$\langle c, x\rangle - \inf_{x' \in D} \langle c, x'\rangle \le t^{-1}\left[\vartheta + \langle y, \Phi_*'(y)\rangle - \langle y, Ax\rangle\right],$$

as claimed. ∎

Note that at the primal-dual path $\Phi_*'(y) = Ax$, and (3.9) gives the standard accuracy bound

$$\langle c, x_*(t)\rangle - c_* \le t^{-1}\vartheta. \tag{3.10}$$

3.3. Proximity measure. Let us define *proximity measure* as the function

$$\Psi(x, y) = \Phi(Ax) + \Phi_*(y) - \langle y, Ax\rangle : D \times D_*^+ \to \mathbf{R}.$$

Notice that for every $x \in D$ and every $y \in D_*^+$, we have

$$\begin{array}{rcl} \Psi(x, y) & = & \Phi(Ax) + \sup\limits_{z \in D^+} \{\langle y, z\rangle - \Phi(z)\} - \langle y, Ax\rangle \\ & \ge & \sup\limits_{x' \in D} \{\langle y, Ax'\rangle - \Phi(Ax')\} - [\langle y, Ax\rangle - \Phi(Ax)]. \end{array}$$

Clearly, the last expression is always nonnegative. Also note that for such a pair (x, y) we have $\Psi(x, y) = 0$ iff $y = \Phi'(Ax)$. We elaborate on the properties of this proximity measure in the next proposition.

PROPOSITION 3.4. *Let $x \in D$, $t > 0$, and let $y \in D_*^+$ be such that*

$$A^* y = -tc. \tag{3.11}$$

Then

(i) *One has*

$$\begin{array}{rcl} \Psi(x, y) & = & F_t(x) + \Phi_*(y) = \left[F_t(x) - \min\limits_{u \in D} F_t(u)\right] + \left[\Phi_*(y) - \min\limits_{v \in D_*^+, A^* v = -tc} \Phi_*(v)\right] \\ & = & \left[F_t(x) - F_t(x_*(t))\right] + \left[\Phi_*(y) - \Phi_*(y_*(t))\right]. \end{array} \tag{3.12}$$

(ii) *Let*

$$\begin{array}{rcl} r & = & \|x - x_*(t)\|_{F''(x_*(t))}, \\ s & = & \|y - y_*(t)\|_{\Phi_*''(y_*(t))}, \\ \lambda_*(y) & = & \max\{h^* \Phi_*'(y) : A^* h = 0, \langle h, \Phi_*''(y) h\rangle \le 1\} \end{array} \tag{3.13}$$

(note that $\lambda_*(y)$ is the Newton decrement, taken at y, of the restriction of $\Phi_*(\cdot)$ to the affine subspace $\{z : A^* z = -tc\}$). Then

$$\rho(r) + \rho(s) \le \Psi(x, y) \le \rho(-r) + \rho(-s) \tag{3.14}$$

and

$$\rho(\lambda(F_t, x)) + \rho(\lambda_*(y)) \le \Psi(x, y) \le \rho(-\lambda(F_t, x)) + \rho(-\lambda_*(y)). \tag{3.15}$$

Proof. (i): The first equality in (3.12) follows from the definition of Ψ combined with (3.11). To prove the second equality, it suffices to verify that

$$\min_{u\in D} F_t(u) + \min_{v\in D_*^+, A^*v=-tc} \Phi_*(v) = 0,$$

or, which is the same in view of Lemma 3.2, that

$$\Phi(Ax_*) + t\langle c, x_*\rangle + \Phi_*(y_*) = 0, \tag{3.16}$$

where $x_* = x_*(t)$, $y_* = y_*(t)$. Since $\Phi_*'(y_*) = Ax_*$ and $A^*y_* = -tc$ by Lemma 3.2, we have

$$\Phi_*(y_*) = \langle y_*, Ax_*\rangle - \Phi(Ax_*) = -t\langle c, x_*\rangle - \Phi(Ax_*),$$

and (3.16) follows.

The third equality in (3.12) is readily given by (3.4) and (3.7).

(ii): Setting $x_* = x_*(t)$, $y_* = y_*(t)$, we have by (2.4.b.2):

$$\begin{array}{rcl} F(x) & \ge & F(x_*) + \langle x - x_*, F'(x_*)\rangle + \rho(r) \\ & = & F(x_*) + \langle Ax - Ax_*, \Phi'(Ax_*)\rangle + \rho(r) \\ & = & F(x_*) + \langle Ax - Ax_*, y_*\rangle + \rho(r), \\ \Phi_*(y) & \ge & \Phi_*(y_*) + \langle y - y_*, \Phi_*'(y_*)\rangle + \rho(s) \\ & = & \Phi_*(y_*) + \langle y - y_*, Ax_*\rangle + \rho(s) \\ & = & \Phi_*(y_*) + \rho(s) \\ & & \qquad [\text{since } A^*y = A^*y_* = -tc] \end{array}$$

whence, taking into account that

$$F(x_*) + \Phi_*(y_*) = \Phi(Ax_*) + \Phi_*(\underbrace{\Phi'(Ax_*)}_{y_*}) = \langle y_*, Ax_*\rangle, \tag{3.17}$$

we get

$$\begin{array}{rcl} F(x) + \Phi_*(y) & \ge & F(x_*) + \Phi_*(y_*) + \langle Ax - Ax_*, y_*\rangle + [\rho(r) + \rho(s)] \\ & = & \langle y_*, Ax_*\rangle + \langle Ax - Ax_*, y_*\rangle + [\rho(r) + \rho(s)] \\ & = & \langle y_*, Ax\rangle + [\rho(r) + \rho(s)] \\ & = & \langle y, Ax\rangle + [\rho(r) + \rho(s)], \\ & & \qquad [\text{since } A^*y = A^*y_*] \end{array}$$

and we arrive at

$$\rho(r) + \rho(s) \le \Psi(x, y),$$

as required in the first inequality in (3.14). The second inequality in (3.14) is trivial when $\max[s, r] \ge 1$; assuming $\max[s, r] < 1$, we have by (2.4.b.1):

$$\begin{array}{rcl} F(x) & \le & F(x_*) + \langle x - x_*, F'(x_*)\rangle + \rho(-r) \\ & = & F(x_*) + \langle Ax - Ax_*, \Phi'(Ax_*)\rangle + \rho(-r) \\ & = & F(x_*) + \langle Ax - Ax_*, y_*\rangle + \rho(-r), \\ \Phi_*(y) & \le & \Phi_*(y_*) + \langle y - y_*, \Phi_*'(y_*)\rangle + \rho(-s) \\ & = & \Phi_*(y_*) + \langle y - y_*, Ax_*\rangle + \rho(-s) \\ & = & \Phi_*(y_*) + \rho(-s) \\ & & \qquad [\text{since } A^*y = A^*y_* = -tc] \end{array}$$

whence, taking into account (3.17),

$$\begin{array}{rcl} F(x)+\Phi_*(y) & \le & F(x_*)+\Phi_*(y_*)+\langle Ax-Ax_*,y_*\rangle+[\rho(-r)+\rho(-s)] \\ & = & \langle y_*,Ax_*\rangle+\langle Ax-Ax_*,y_*\rangle+[\rho(-r)+\rho(-s)] \\ & = & \langle Ax,y_*\rangle+[\rho(-r)+\rho(-s)] \\ & = & \langle Ax,y\rangle+[\rho(-r)+\rho(-s)], \end{array}$$

and we arrive at

$$\Psi(x,y)\le\rho(-r)+\rho(-s),$$

as required in the second inequality in (3.14).

Finally, since $F_t(\cdot)$ is self-concordant, we have

$$\rho(\lambda(F_t,x))\le F_t(x)-\min F_t(\cdot)=F_t(x)-F_t(x_*)\le\rho(-\lambda(F_t,x))$$

by (2.5.b) and (2.6). The same arguments as applied to the self-concordant function $\Phi_*|_{\{z:A^*z=-tc\}}$ result in

$$\rho(\lambda_*(y))\le\Phi_*(y)-\Phi_*(y_*)\le\rho(-\lambda_*(y)).$$

These relations, in view of (3.12), lead to (3.15). ∎

3.4. Potential. For $x\in D$, $y\in D_*^+$, $t>0$ let

$$\Theta(x,y,t)=\Psi(x,y)-\sqrt{\vartheta}\ln t.$$

Note that by (3.12) we have

$$A^*y=-tc\Rightarrow\Theta(x,y,t)\ \begin{array}{l} =F_t(x)+\Phi_*(y)-\sqrt{\vartheta}\ln t \\ =[F_t(x)-\min\limits_{u\in D}F_t(u)]+[\Phi_*(y)-\min\limits_{v\in D_*^+,A^*v=-tc}\Phi_*(v)]-\sqrt{\vartheta}\ln t. \end{array}\tag{3.18}$$

PROPOSITION 3.5. *Let $x\in D$, $t>0$, and let $y\in D_*^+$ be such that*

$$A^*y=-tc.$$

Then

$$\langle c,x\rangle-\inf_{u\in D}\langle c,u\rangle\le 2\vartheta\exp\left\{\frac{\sqrt{\vartheta}-\vartheta}{2\vartheta}\right\}\exp\left\{\frac{\Theta(x,y,t)}{\sqrt{\vartheta}}\right\}.\tag{3.19}$$

REMARK 3.5. *We will see in a while that the standard Newton-type techniques allow, given a initial triple (x_0,y_0,t_0) such that $x_0\in D$, $y_0\in D_*^+$, $t_0>0$ and $A^*y_0=-t_0c$, to build a sequence of iterates (x_i,y_i,t_i) such that $A^*y_i=-t_ic$ and $\Theta_i\equiv\Theta(x_i,y_i,t_i)\le\Theta_{i-1}-\kappa$ with an absolute constant $\kappa>0$. Relation (3.19) demonstrates that the resulting procedure obeys the standard $\sqrt{\vartheta}$-complexity bound.*

REMARK 3.6. *In the Standard case (see Remark 3.1), the points $y\in\mathrm{Dom}\Phi_*$ satisfying $A^*y=-tc$ are exactly the points of the form $y=-t\widehat{y}$, where $\widehat{y}$ is a strictly feasible solution to the dual problem (1.4). Expressing Θ in terms of $(x,\widehat{y},t)$, we arrive at the function*

$$\begin{array}{rcl} \widehat{\Theta}(x,\widehat{y},t) & \equiv & \Theta(x,-t\widehat{y},t)=H^*(t\widehat{y})+H(Ax-b)+t\langle\widehat{y},Ax-b\rangle-\sqrt{\vartheta}\ln t \\ & = & H^*(\widehat{y})+H(Ax-b)+t\langle\widehat{y},Ax-b\rangle-(\vartheta+\sqrt{\vartheta})\ln t. \end{array}$$

In the potential reduction scheme, we want to iterate on (x,y,t) in order to reduce step by step the potential Θ. In the Standard case, we can simplify this task by eliminating the variable t — by minimizing $\widehat{\Theta}$ in t analytically. The "optimal" t is $t=\frac{\vartheta+\sqrt{\vartheta}}{\langle\widehat{y},Ax-b\rangle}$, and the "optimized" potential is

$$\Xi(x,\widehat{y})=H^*(\widehat{y})+H(Ax-b)+(\vartheta+\sqrt{\vartheta})\ln\left(\langle\widehat{y},Ax-b\rangle\right)+\mathrm{const},$$

which is nothing but the usual primal-dual potential of the Standard case.

Proof of Proposition. Let $x_* = x_*(t)$, so that $F'(x_*) = -tc$, let $y_* = y_*(t)$, and let $\gamma = \Theta(x,y,t)$. Since $A^*y = -tc$, Proposition 3.4 implies the first statement in the following chain:

$$\gamma = -\sqrt{\vartheta}\ln t + [F_t(x) - F_t(x_*)] + \underbrace{[\Phi_*(y) - \Phi_*(y_*)]}_{\geq 0}$$
$$\Downarrow$$
$$F_t(x) - F_t(x_*) \leq \gamma + \sqrt{\vartheta}\ln t$$
$$\Downarrow$$
$$(*) \quad \rho(\|x - x_*\|_{F''(x_*)}) \leq \gamma + \sqrt{\vartheta}\ln t \quad \text{[using (2.4.}b\text{.2), } F_t'(x_*) = 0].$$

Observe that from $(*)$ it follows that

$$\|x - x_*\|_{F''(x_*)} \leq \sigma\left(\gamma + \sqrt{\vartheta}\ln t\right). \tag{3.20}$$

On the other hand,

$$\|tc\|_{[F''(x_*)]^{-1}} = \|F'(x_*)\|_{[F''(x_*)]^{-1}} = \|F'(x_*)\|^*_{F''(x_*)} \leq \sqrt{\vartheta}$$

(the concluding inequality comes from the fact that F is ϑ-s.c.b.), whence in view of (3.20) and the fact that $\langle c, x - x_*\rangle \leq \|c\|^*_{F''(x_*)}\|x - x_*\|_{F''(x_*)}$, one has

$$\langle c, x\rangle \leq \langle c, x_*\rangle + \sqrt{\vartheta}t^{-1}\sigma\left(\gamma + \sqrt{\vartheta}\ln t\right). \tag{3.21}$$

Recalling that $x_* = x_*(t)$ and invoking (3.10), we come to

$$\epsilon(x) \equiv \langle c, x\rangle - \inf_{u\in D}\langle c, u\rangle \leq \vartheta t^{-1} + \sqrt{\vartheta}t^{-1}\sigma\left(\gamma + \sqrt{\vartheta}\ln t\right). \tag{3.22}$$

From (2.1) it follows that $\sigma(s) \leq 1 + 2s$ for all $s \geq 0$. Now (3.22) implies that

$$\epsilon(x) \leq (\vartheta + \sqrt{\vartheta})t^{-1} + 2\vartheta t^{-1}\ln t + 2\sqrt{\vartheta}t^{-1}\gamma. \tag{3.23}$$

Consequently,

$$\epsilon(x) \leq \max_{\tau>0}\left\{(\vartheta + \sqrt{\vartheta})\tau^{-1} + 2\vartheta\tau^{-1}\ln\tau + 2\sqrt{\vartheta}\tau^{-1}\gamma\right\},$$

and the maximum in the right hand side, as it is easily seen, is exactly the right hand side in (3.19). ■

4. How to reduce the potential. Consider the following situation: We are given a triple $(x \in D, y \in D_*^+, t > 0)$ with

$$A^*y = -tc, \tag{4.1}$$

and we intend to update this triple into a triple (x_+, y_+, t_+) such that

$$\begin{array}{ll} (a) & x_+ \in D,\ y_+ \in D_*^+ \\ (b) & A^*y_+ = -t_+c \\ (c) & \Theta(x_+, y_+, t_+) \leq \Theta(x, y, t) - \Omega(1). \end{array} \tag{4.2}$$

The options we have are *at least* as follows:

4.1. Centering, damped Newton step in x. Here

$$\begin{array}{rcl} y_+ &=& y, \\ t_+ &=& t, \\ x_+ &=& x - \frac{1}{1+\lambda(F_t,x)}[F''(x)]^{-1}F_t'(x). \end{array} \tag{4.3}$$

This updating clearly satisfies $(4.2.a-b)$. Since $A^*y = -tc$, we have

$$\begin{array}{rcll} \Theta(x_+, y_+, t_+) - \Theta(x, y, t) &=& \Theta(x_+, y, t) - \Theta(x, y, t) & \\ &=& F_t(x_+) - F_t(x) & \text{[see (3.18)]} \\ &\leq& -\rho(\lambda(F_t, x)) & \text{[see (2.5.}b\text{)].} \end{array} \tag{4.4}$$

4.2. Centering, damped Newton step in y. Here

$$\begin{array}{rcl} x_+ &=& x, \\ t_+ &=& t, \\ y_+ &=& y - \frac{1}{1+\lambda_*(y)} e(y), \end{array} \tag{4.5}$$

where

$$\begin{array}{rcl} e(y) &\equiv& \underset{h}{\operatorname{argmax}}\{\langle h, \Phi'_*(y)\rangle : A^*h = 0, \langle h, \Phi''_*(y)h\rangle \le 1\} \\ &=& [\Phi''_*(y)]^{-1}\left[I - A[A^*[\Phi''_*(y)]^{-1}A]^{-1}A^*[\Phi''_*(y)]^{-1}\right]\Phi'_*(y), \\ \lambda_*(y) &\equiv& \max\{h^*\Phi'_*(y) : A^*h = 0, \langle h, \Phi''_*(y)h\rangle \le 1\} \\ &=& \|e(y)\|_{\Phi''_*(y)} \end{array} \tag{4.6}$$

are, respectively, the Newton direction and the Newton decrement, taken at y, of the function $\Phi_*|_{\{z:A^*z=-tc\}}$.

Updating (4.5) clearly satisfies $(4.2.a-b)$. Since $A^*y = A^*y_+ = -tc$, we have

$$\begin{array}{rcll} \Theta(x_+, y_+, t_+) - \Theta(x, y, t) &=& \Theta(x, y_+, t) - \Theta(x, y, t) & \\ &=& \Phi_*(y_+) - \Phi_*(y) & \\ &\le& -\rho(\lambda_*(y)) & [(2.5.b) \text{ as applied to } \Phi_*|_{\{z:A^*z=-tc\}}]. \end{array} \tag{4.7}$$

4.3. Primal path-tracing. A generic primal path-tracing step is as follows:

$$\begin{array}{rcl} t_+ &=& t + \Delta t \quad [\Delta t > 0], \\ x_+ &=& x - [F''(x)]^{-1}F'_{t_+}(x), \\ y_+ &=& \Phi'(Ax) + \Phi''(Ax)A(x_+ - x). \end{array} \tag{4.8}$$

The motivation behind this construction is clear: our ideal goal would be to update (x, y, t) into the triple (x_*^+, y_*^+, t_+) with $t_+ > t$ and x_*^+, y_*^+ on the primal-dual path:

$$\begin{array}{rcl} F'_{t_+}(x_*^+) &=& 0, \\ \Phi'(Ax_*^+) - y_*^+ &=& 0. \end{array} \tag{4.9}$$

x_+, y_+ as given by (4.8) solve the linearization of the system (4.9) at x.

We are about to analyze the primal path-tracing step.

LEMMA 4.1. *Let a triple $(x \in D, y \in D_*^+, t > 0)$ satisfy (4.1), and let (x_+, y_+, t_+) be obtained from (x, y, t) by a primal path-tracing step (4.8). Then*

(i) *One has*

$$A^*y_+ = -t_+c. \tag{4.10}$$

(ii) *Let $z = \Phi'(Ax)$. Then*

$$\|y_+ - z\|_{\Phi''_*(z)} = \|x_+ - x\|_{F''(x)} = \|F'_t(x) + \Delta tc\|^*_{F''(x)}. \tag{4.11}$$

(iii) *The relation*

$$\|x_+ - x\|^*_{F''(x)} < 1 \tag{4.12}$$

is a sufficient condition for the inclusions

$$x_+ \in D, \quad y_+ \in D_*^+.$$

(iv) *One has*

$$\|x_+ - x\|_{F''(x)} \le \lambda(F_t, x) + \frac{|\Delta t|}{t}(\lambda(F_t, x) + \sqrt{\vartheta}). \tag{4.13}$$

Proof. (i): We have

$$\begin{array}{rcl} A^*y_+ &=& A^*\Phi'(Ax) + A^*\Phi''(Ax)A(x_+ - x) = F'(x) + F''(x)(x_+ - x) \\ &=& F'(x) - F'_t(x) - \Delta t c = -(t+\Delta t)c = -t_+ c, \end{array}$$

which proves (i).

(ii): The second equality in (4.11) is evident. We have

$$\begin{array}{rcl} \|x_+ - x\|^2_{F''(x)} &=& \langle x_+ - x, A^*\Phi''(Ax)A(x_+ - x)\rangle \\ &=& \langle \Phi''(Ax)A(x_+ - x), \underbrace{[\Phi''(Ax)]^{-1}}_{\Phi''_*(z)} \underbrace{\Phi''(Ax)A(x_+ - x)}_{y_+ - z}\rangle = \langle y_+ - z, \Phi''_*(z)(y_+ - z)\rangle. \end{array}$$

(ii) is proved.

(iii): By (4.11), in the case of (4.12) one has

$$\|x - x_+\|_{F''(x)} = \|y_+ - z\|_{\Phi''_*(z)} < 1,$$

whence, by SC.II.1), $x_+ \in D$ and $y_+ \in D_*^+$.

(iv): By (ii),

$$\begin{array}{rcl} \|x_+ - x\|_{F''(x)} &=& \|F'_t(x) + \Delta t c\|^*_{F''(x)} \le \|F'_t(x)\|^*_{F''(x)} + |\Delta t| \|c\|^*_{F''(x)} \\ &=& \lambda(F_t, x) + |\Delta t| \|c\|^*_{F''(x)} \end{array} \tag{4.14}$$

and

$$\begin{array}{rcl} \|F'_t(x)\|^*_{F''(x)} &=& \|F'(x) + tc\|^*_{F''(x)} \ge t\|c\|^*_{F''(x)} - \|F'(x)\|^*_{F''(x)} \\ &\ge& t\|c\|^*_{F''(x)} - \sqrt{\vartheta}, \end{array}$$

whence

$$\|c\|^*_{F''(x)} \le t^{-1}\left[\lambda(F_t, x) + \sqrt{\vartheta}\right],$$

which combines with (4.14) to yield (4.13). ∎

LEMMA 4.2. *Let a triple* $(x \in D, y \in D_*^+, t > 0)$ *satisfy* (4.1), *and let* (x_+, y_+, t_+) *be obtained from* (x, y, t) *by a primal path-tracing step* (4.8). *Assume that*

$$\gamma \equiv \|x_+ - x\|_{F''(x)} < 1.$$

Then

$$\begin{array}{rcll} \Psi(x_+, y_+) &\le& 2\omega(\gamma), & (a) \\ \Theta(x_+, y_+, t_+) - \Theta(x, y, t) &\le& 2\omega(\gamma) - \sqrt{\vartheta}\ln\frac{t_+}{t}. & (b) \end{array} \tag{4.15}$$

Proof. Let $z = \Phi'(Ax)$, $\Phi'' = \Phi''(Ax)$, $\Delta x = x_+ - x$. Since $\|y_+ - z\|_{\Phi''_*(z)} = \gamma$ by (4.11) and $\gamma < 1$, relation (2.4) implies that

$$\begin{array}{rcl} \Phi_*(y_+) &\le& \Phi_*(z) + \langle y_+ - z, \Phi'_*(z)\rangle + \rho(-\gamma) \\ &=& \Phi_*(z) + \langle \Delta x, A^*\Phi'' Ax\rangle + \rho(-\gamma), \end{array} \tag{4.16}$$

and similarly

$$\begin{array}{rcl} \Phi(Ax_+) &\le& \Phi(Ax) + \langle \Delta x, A^*\Phi'(Ax)\rangle + \rho(-\gamma) \\ &=& \Phi(Ax) + \langle \Delta x, A^* z\rangle + \rho(-\gamma) \end{array} \tag{4.17}$$

whence, due to $\Phi_*(z) + \Phi(Ax) = \langle z, Ax\rangle$ in view of $z = \Phi'(Ax)$,

$$
\begin{array}{rl}
\Phi(Ax_+) + \Phi_*(y_+) - \langle y_+, Ax_+\rangle & \\
& \le \underbrace{[\Phi_*(z) + \Phi(Ax)]}_{\langle z, Ax\rangle} + \langle \Delta x, A^*\Phi'' Ax\rangle + \langle \Delta x, A^* z\rangle - \langle y_+, Ax_+\rangle + 2\rho(-\gamma) \\
& = \langle z, Ax\rangle + \langle \Delta x, A^*\Phi'' Ax\rangle + \langle \Delta x, A^* z\rangle - \langle z + \Phi'' A\Delta x, A(x+\Delta x)\rangle + 2\rho(-\gamma) \\
& = -\langle \Delta x, A^*\Phi'' A\Delta x\rangle + 2\rho(-\gamma) \\
& = -\gamma^2 + 2\rho(-\gamma) \\
& = 2\omega(\gamma),
\end{array}
\tag{4.18}
$$

as required in (4.15.a). We now have

$$
\begin{array}{rl}
\Theta(x_+, y_+, t_+) - \Theta(x, y, t) & \\
& = \underbrace{[\Phi(Ax_+) + \Phi_*(y_+) - \langle y_+, Ax_+\rangle]}_{\le 2\omega(\gamma) \text{ by } (4.18)} - \underbrace{[\Phi(Ax) + \Phi_*(y) - \langle y, Ax\rangle]}_{\ge 0} - \sqrt{\vartheta}\ln\frac{t_+}{t} \\
& \le 2\omega(\gamma) - \sqrt{\vartheta}\ln\frac{t_+}{t}.
\end{array}
$$

■

Corollary 4.3. *Let $t > 0$ and x be such that $\lambda(F_t, x) \le 0.1$. Then with $\frac{\Delta t}{t} = \frac{0.25}{\sqrt{\vartheta}}$ the primal path-tracing step is feasible (i.e., $x_+ \in D$, $y_+ \in D_*^+$) and*

$$\Theta(x_+, y_+, t_+) - \Theta(x, y, t) \le -0.17.$$

Proof. This is an immediate consequence of the previous two lemmas, in particular, the bounds (4.13) and (4.15). ■

4.4. Dual path-tracing. A generic dual path-tracing step is as follows:

$$
\begin{array}{rcl}
t_+ & = & t + \Delta t \quad [\Delta t > 0], \\
y_+ & = & y + \Delta y : \quad A^* y = -t_+ c, \ \ \Phi_*'(y) + \Phi_*''(y)\Delta y \in \mathrm{Im} A, \\
x_+ & : & \Phi_*'(y) + \Phi_*''(y)\Delta y = Ax_+.
\end{array}
\tag{4.19}
$$

The motivation behind the construction is completely similar to the one in Section 4.3, up to the fact that now we linearize an alternative to (4.9) description of the triple (x_*^+, y_*^+, t_+) on the primal-dual path, specifically, the description

$$
\begin{array}{rcl}
A^* y_*^+ + t_+ c & = & 0, \\
\Phi_*'(y_*^+) - Ax_*^+ & = & 0.
\end{array}
\tag{4.20}
$$

We are about to analyze a dual path-tracing step. Although the results to follow are completely similar to those for the primal path-tracing step, the analysis is slightly different — we do not have enough primal-dual symmetry!

Lemma 4.4. *Let a triple $(x \in D, y \in D_*^+, t > 0)$ satisfy (4.1), and let*

$$\xi = \Phi_*'(y), \quad \Phi'' = \Phi''(\xi). \tag{4.21}$$

Then

(i) *The triple (x_+, y_+, t_+) in (4.19) is well-defined and is explicitly given by the relations*

$$
\begin{array}{rcl}
x_+ & = & [A^*\Phi'' A]^{-1} A^* \left[\frac{\Delta t}{t} y + \Phi''\xi\right], \\
\Delta y & = & \Phi''\left[Ax_+ - \xi\right] \\
& = & \Phi''\Big[\frac{\Delta t}{t} \underbrace{A[A^*\Phi'' A]^{-1} A^* y}_{\delta_1} - \underbrace{\left[I - A[A^*\Phi'' A]^{-1} A^*\Phi''\right]\xi}_{\delta_2}\Big], \\
y_+ & = & y + \Delta y.
\end{array}
\tag{4.22}
$$

(ii) *One has*

$$\|Ax_+ - \xi\|_{\Phi''(\xi)} = \|\Delta y\|_{\Phi_*''(y)}. \tag{4.23}$$

(iii) *The relation*

$$\|\Delta y\|_{\Phi_*''(y)} < 1 \tag{4.24}$$

is a sufficient condition for the inclusions

$$x_+ \in D, \quad y_+ \in D_*^+.$$

(iv) *One has*

$$\|\Delta y\|_{\Phi_*''(y)} \le \sqrt{\lambda_*^2(y) + \vartheta\frac{(\Delta t)^2}{t^2}}. \tag{4.25}$$

Proof. (i): This is given by a straightforward computation, where one should take into account that $\Phi'' = \Phi''(\xi) = [\Phi_*''(y)]^{-1}$ due to $\xi = \Phi_*'(y)$ and that $A^*y = -tc$ by (4.1).

(ii): This is an immediate consequence of the relations $\Delta y = \Phi''(\xi)[Ax_+ - \xi]$ (see (4.22)) and $\Phi_*''(y) = [\Phi''(\xi)]^{-1}$ (recall that $\xi = \Phi_*'(y)$).

(iii): This is an immediate consequence of (4.23) and SC.II.1).

(iv): By (4.22) and in view of $\Phi_*''(y) = [\Phi'']^{-1}$ we have

$$\begin{aligned} \|\Delta y\|^2_{\Phi_*''(y)} &= \|\tfrac{\Delta t}{t}\delta_1 - \delta_2\|^2_{\Phi''} \\ &= \tfrac{(\Delta t)^2}{t^2}\|\delta_1\|^2_{\Phi''} + \|\delta_2\|^2_{\Phi''} \\ &\quad \text{[direct computation]}. \end{aligned} \tag{4.26}$$

Taking into account that $\xi = \Phi_*'(y)$ and $\Phi_*''(y) = [\Phi'']^{-1}$, by (4.6) we have

$$\|\delta_2\|^2_{\Phi''} = \lambda_*^2(y). \tag{4.27}$$

Finally, $y = \Phi'(\xi)$ due to $\xi = \Phi_*'(y)$, and we have

$$\begin{aligned} \|\delta_1\|^2_{\Phi''} &= \langle y, A[A^*\Phi''A]^{-1}A^*y\rangle && \text{[direct computation]} \\ &\le \langle y, [\Phi'']^{-1}y\rangle && \text{[Linear Algebra (projection of } [\Phi'']^{-1/2}y)] \\ &= \langle \Phi'(\xi), [\Phi''(\xi)]^{-1}\Phi'(\xi)\rangle = \left(\|\Phi'(\xi)\|^*_{\Phi''(\xi)}\right)^2 \\ &\le \vartheta && \text{[since } \Phi \text{ is } \vartheta\text{-s.c.b.]}. \end{aligned} \tag{4.28}$$

Combining (4.26) – (4.28), we arrive at (4.25). ∎

LEMMA 4.5. *Let a triple $(x \in D, y \in D_*^+, t > 0)$ satisfy* (4.1)*, and let (x_+, y_+, t_+) be obtained from (x, y, t) by a dual path-tracing step* (4.19)*. Assume that*

$$\gamma \equiv \|y_+ - y\|_{\Phi_*''(y)} < 1.$$

Then

$$\begin{aligned} \Psi(x_+, y_+) &\le 2\omega(\gamma), && (a) \\ \Theta(x_+, y_+, t_+) - \Theta(x, y, t) &\le 2\omega(\gamma) - \sqrt{\vartheta}\ln\tfrac{t_+}{t}. && (b) \end{aligned} \tag{4.29}$$

Proof. Let $\xi = \Phi_*'(y)$, $\Phi'' = \Phi''(\xi)$, $\Delta y = y_+ - y$. Since $\|\Delta y\|_{\Phi_*''(y)} = \gamma < 1$, relation (2.4) implies that

$$\Phi_*(y_+) \le \Phi_*(y) + \langle \Delta y, \underbrace{\Phi_*'(y)}_{\xi}\rangle + \rho(-\gamma). \tag{4.30}$$

Similarly, in view of $\|Ax_+ - \xi\|_{\Phi''(\xi)} = \gamma$ (see (4.23)), we have

$$\Phi(Ax_+) \leq \Phi(\xi) + \langle Ax_+ - \xi, \underbrace{\Phi'(\xi)}_{y} \rangle + \rho(-\gamma), \tag{4.31}$$

whence, due to $\Phi_*(y) + \Phi(\xi) = \langle y, \xi \rangle$ in view of $\xi = \Phi'_*(y)$,

$$\begin{aligned} &\Phi(Ax_+) + \Phi_*(y_+) - \langle y_+, Ax_+ \rangle \\ &\quad \leq \underbrace{[\Phi_*(y) + \Phi(\xi)]}_{\langle y, \xi \rangle} + \langle \Delta y, \xi \rangle + \langle Ax_+ - \xi, y \rangle - \langle y_+, Ax_+ \rangle + 2\rho(-\gamma) \\ &\quad = \langle \Delta y, \xi - Ax_+ \rangle + 2\rho(-\gamma) \\ &\quad = -\langle \Delta y, [\Phi'']^{-1} \Delta y \rangle + 2\rho(-\gamma) && \text{[see (4.22)]} \\ &\quad = -\gamma^2 + 2\rho(-\gamma) && [\text{since } [\Phi'']^{-1} = \Phi''_*(y)] \\ &\quad = 2\omega(\gamma), \end{aligned} \tag{4.32}$$

as required in (4.29.a). We now have

$$\begin{aligned} &\Theta(x_+, y_+, t_+) - \Theta(x, y, t) \\ &\quad = \underbrace{[\Phi(Ax_+) + \Phi_*(y_+) - \langle y_+, Ax_+ \rangle]}_{\leq 2\omega(\gamma) \text{ by } (4.32)} - \underbrace{[\Phi(Ax) + \Phi_*(y) - \langle y, Ax \rangle]}_{\geq 0} - \sqrt{\vartheta} \ln \tfrac{t_+}{t} \\ &\quad \leq 2\omega(\gamma) - \sqrt{\vartheta} \ln \tfrac{t_+}{t}. \end{aligned}$$

■

COROLLARY 4.6. *Let $t > 0$ and y be such that (4.1) takes place and $\lambda_*(y) \leq 0.1$. Then with $\frac{\Delta t}{t} = \frac{0.25}{\sqrt{\vartheta}}$ the dual path-tracing step is feasible (i.e., $x_+ \in D$, $y_+ \in D_*^+$) and*

$$\Theta(x_+, y_+, t_+) - \Theta(x, y, t) \leq -0.17.$$

Proof. This is an immediate consequence of bounds (4.25) and (4.29). ■

5. Primal-dual path-following methods. Now we are ready to describe primal-dual path-following methods for solving (3.2). The construction to follow reproduces in our "Good case" setting the construction developed in [13] for the Standard case (and in fact was investigated, even in a more general "surface-following" form, in [15]).

Let us say that a triple $(x \in D, y \in D_*^+, t > 0)$ is close to the primal-dual path, if

$$A^* y = -tc \ \& \ \max[\lambda(F_t, x), \lambda_*(y)] \leq 0.1. \tag{5.1}$$

Assume that we are given a close to the path starting triple (x_0, t_0, y_0) [1]. Starting with this point, we trace the primal-dual path using a predictor-corrector scheme. Specifically, at step i of the scheme we act as follows:

1. [predictor step] Given a close to the path triple $(x_{i-1}, y_{i-1}, t_{i-1})$, we
   (a) specify a *search direction* (dx_i, dy_i) in such a way that

   $$A^* dy_i = -c; \tag{5.2}$$

   (b) find a *stepsize* $\Delta t_i > 0$ in such a way that

   $$\Psi(\underbrace{x_{i-1} + \Delta t_i dx_i}_{x_i^+}, \underbrace{y_{i-1} + \Delta t_i dy_i}_{y_i^+}) \leq \kappa \tag{5.3}$$

   ($\kappa \geq 1$ is a parameter of the method) and set

   $$t_i = t_{i-1} + \Delta t_i.$$

[1] Such a triple can be found by every one of the well-known interior-point initialization routines.

2. [corrector step] Starting with (x_i^+, y_i^+, t_i), we apply the damped Newton updatings

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x^+ = x - \frac{1}{1+\lambda(F_{t_i},x)}[F''(x)]^{-1}F'_{t_i}(x) \\ y^+ = y - \frac{1}{1+\lambda_*(y)}e(y) \end{bmatrix} \tag{5.4}$$

(see (4.5)) until a pair (x, y) with

$$\max[\lambda(F_{t_i}, x), \lambda_*(y)] \leq 0.1 \tag{5.5}$$

is built, and set

$$x_i = x, \quad y_i = y,$$

thus obtaining a close to the path triple (x_i, y_i, t_i).

Note that with this approach, the number of damped Newton updatings at a corrector step is $O(1)(\kappa+1)$. Indeed, in view of SC.II and (3.12), updating (5.4) ensures that

$$\Psi(x^+, y^+) \leq \Psi(x, y) - \rho(\lambda(F_{t_i}, x)) - \rho(\lambda_*(y));$$

since Ψ is nonnegative and $\Psi \leq \kappa$ at the beginning of the corrector step by (5.3), the number of updatings (5.4) before the termination criterion (5.5) is met is at most $O(1)(\kappa + 1)$.

Note also that the subsequent iterates x_i satisfy the condition $\lambda(F_{t_i}, x_i) \leq 0.1$; by standard interior-point argument (see [14], Chapter 3), it follows that

$$\langle c, x_i \rangle - c_* \leq \frac{2\vartheta}{t_i}. \tag{5.6}$$

Note that we can easily build search directions (dx_i, dy_i) which, along with the stepsizes

$$\Delta t_i = \frac{0.6 t_{i-1}}{\sqrt{\vartheta}}, \tag{5.7}$$

ensure (5.3). Indeed, let us specify (x, y, t) and Δt in the primal path-tracing step (4.8)[2)] as $(x_{i-1}, y_{i-1}, t_{i-1})$ and Δt_i as given by (5.7), respectively, compute the corresponding x_+, y_+, t_+ and then set

$$dx_i = \frac{x_+ - x}{\Delta t_i}, \quad dy_i = \frac{y_+ - y}{\Delta t_i},$$

thus getting $(x_i, y_i, t_i) = (x_+, y_+, t_+)$. Since $(x, y, t) \equiv (x_{i-1}, y_{i-1}, t_{i-1})$ is close to the path, (4.13) ensures that

$$\gamma \equiv \|x_+ - x\|_{F''(x)} \leq \lambda(F_t, x) + \frac{|\Delta t|}{t}(\lambda(F_t, x) + \sqrt{\vartheta}) \leq 0.1 + \frac{0.6}{\sqrt{\vartheta}}(0.1 + \sqrt{\vartheta}) \leq 0.76;$$

consequently, Lemma 4.2 implies that

$$\Psi(x_i, y_i) \leq 2\omega(0.76) < 1 \leq \kappa,$$

and (5.3) follows. Note that (5.7), same as every "more aggressive" stepsize rule compatible with (5.3), guarantees the standard $\sqrt{\vartheta}$-complexity bounds for the resulting algorithm.

The major advantage of the primal-dual path-following framework we have developed (same as of the one of the Standard-case-oriented primal-dual framework developed in [13]) is that we have no reason to restrict ourselves with the worst-case-oriented short-step policies like (5.7). The proximity measure $\Psi(x, y)$ usually is easy to compute, which allows to implement various policies for on-line adjustment of the stepsizes (for theoretical results on the "power" of these adjustments, see [15]).

[2)] We could use the dual path-tracing step as well.

6. Primal-dual potential reduction methods. Proposition 3.5 combines with the results of Section 4 to yield primal-dual potential reduction methods obeying the standard $\sqrt{\vartheta}$-complexity bounds. A generic method of this type is as follows.

We generate a sequence of triples $(x_i \in D, y_i \in D_*^+, t_i > 0)$ such that

$$A^* y_i = -t_i c \tag{6.1}$$

in a way which ensures that

$$\Theta(x_i, y_i, t_i) \leq \Theta(x_{i-1}, y_{i-1}, t_{i-1}) - \kappa, \tag{6.2}$$

where $\kappa > 0$ is a parameter of the method. Specifically, given $(x_{i-1}, y_{i-1}, t_{i-1})$ satisfying (6.1), we build somehow a *search direction* (dx_i, dy_i, dt_i) satisfying the requirement

$$A^* dy_i = -dt_i c$$

and a stepsize τ_i in such a way that the point

$$(x_i, y_i, t_i) = (x_{i-1}, y_{i-1}, t_{i-1}) + \tau_i(dx_i, dy_i, dt_i)$$

satisfies (6.2).

The results of Section 4 suggest rules for choosing the search directions and the stepsizes which ensure (6.2) for an appropriate *absolute constant* κ. For example, if $\lambda(F_{t_{i-1}}, x_{i-1}) > 0.1$, then the centering step in x reduces the potential by at least $\rho(\lambda(F_{t_{i-1}}, x_{i-1})) \geq \rho(0.1)$ (Section 4.1), and if $\lambda(F_{t_{i-1}}, x_{i-1}) \leq 0.1$, then a primal path-tracing step with $t_i - t_{i-1} = 0.25 t_{i-1}/\sqrt{\vartheta}$ reduces the potential by at least 0.17 (Section 4.3). (Of course, we can utilize, in the same fashion, the centering in y and the dual path-tracing step.) Needless to say, a reasonable implementation should include a line-search in the chosen direction in order to get as large reduction in the potential as possible, or even a multi-dimensional search (e.g., 4-dimensional search along the linear span of four search directions described in Section 4). A deeper investigation of possible variants and implementations of potential reduction methods goes beyond the scope of this paper; what matters theoretically, is that whenever we ensure (6.2) and thus – the relation $\Theta(x_i, y_i, t_i) \leq \Theta(x_0, y_0, t_0) - i\kappa$, Proposition 3.5 implies that

$$c^T x_i - \inf_{u \in D} c^T u \leq \left[2\vartheta \exp\left\{ \frac{\sqrt{\vartheta} - \vartheta}{2\vartheta} \right\} \exp\left\{ \frac{\Theta(x_0, y_0, t_0)}{\sqrt{\vartheta}} \right\} \right] \exp\left\{ -\frac{i\kappa}{\sqrt{\vartheta}} \right\},$$

i.e., we get a polynomial time method with the standard $\sqrt{\vartheta}$-complexity bound (provided, of course, that $\kappa = O(1)$).

7. Possible Applications and Extensions. When working on polynomial-time interior-point methods, among others, four important issues arise.

1. Are there interesting classes of problems which are covered by the new method in an effective manner?
2. How provably long are the primal and/or dual steps?
3. How much dual information is utilized (and generated) by the method and how effectively?
4. How can we initiate the method for an arbitrary input in a way that preserves 1., 2. and 3. above?

In this last section, we comment on the above issues.

7.1. Potential applications. Geometric Programming provides an interesting class of applications (for a survey, see [5]; for a set of test problems see [4]; interesting recent applications in Engineering are presented in [2]). We have seen in Introduction that this problem class fits our primal-dual framework; at the same time, it is not directly covered by the existing primal-dual polynomial time algorithms, and the only previous primal-dual interior point method for Geometric Programming [10], although globally converging, is *not* known to be a polynomial time one.

Note that, essentially, the only feature of Geometric Programming which was responsible for the possibility to process this class within our framework, is the fact that the "underlying entity" – the epigraph of the exponential function $f(y) = \exp\{y\}$ – admits an explicit self-concordant barrier with explicit Legendre transformation. Now, constructing a self-concordant barrier for the epigraph of a *univariate* convex function f usually is a routine task. As a rule, it is not that difficult to obtain, along with such a barrier, its Legendre transformation, either in an explicit analytical form, as in the case of $f(y) = \exp\{y\}$, or "semi-explicitly" – via a real parameter which should satisfy a "well-posed" equation. As an instructive example, consider the entropy function $f(y) = y \ln y$. The 2-self-concordant barrier for the epigraph of f is given by

$$G(s, y) = -\left[\ln(s - y\ln(y)) + \ln(y)\right]$$

(see [14], Section 5.3.1), and the Legendre transformation of this barrier is

$$G_*(\sigma, \eta) = -\ln(-\sigma) + \theta\left(1 + \frac{\eta}{\sigma} - \ln(-\sigma)\right) - \frac{\eta}{\sigma} + \frac{1}{\theta\left(1 + \frac{\eta}{\sigma} - \ln(-\sigma)\right)} - 3,$$

where $\theta(r)$ is the unique root of the equation

$$\frac{1}{\theta} - \ln\theta = r. \tag{7.1}$$

It is not that difficult to write a dedicated code which computes $\theta(r), \theta'(r), \theta''(r)$ in time comparable with the one required to compute a standard elementary function, like $\mathrm{acos}(\cdot)$[3)], so that it is not a great sin to say that $G_*(\cdot,\cdot)$ is as easily computable as, say, the Legendre transformation of the barrier for the epigraph of the exponent. Note that $G(\cdot,\cdot)$ is *not* a logarithmically homogeneous barrier for a cone. Now, with $G(\cdot)$ and $G_*(\cdot)$ in our disposal, we can process an Entropy Optimization problem

$$\begin{array}{c} \min\limits_x \left\{c^T x : f_i(x) \le 0,\ i = 1, ..., m,\ Bx \le b\right\}, \\ f_i(x) = \sum\limits_{\ell=1}^{L} \alpha_{i\ell}(\delta_\ell + d_\ell^T x)\ln(\delta_\ell + d_\ell^T x) + e_i^T x + \beta_i, \end{array} \tag{7.2}$$

with $\alpha_{i\ell} \ge 0$ in the same fashion as a Geometric Programming problem: assuming (7.2) strictly feasible, the interior D of the feasible set can be easily represented in the form

$$\begin{array}{rcl} D &=& \{x : (Ax - b) \in \mathcal{D}\}, \\ \mathcal{D} &=& \{(t, y, s) \in \mathbf{R}^p \times \mathbf{R}^q \times \mathbf{R}^q : t_i > 0,\ i = 1, ..., p,\ y_i\ln(y_i) < s_i,\ i = 1, ..., q\}. \end{array}$$

The set $\mathrm{cl}\mathcal{D}$ admits the explicit $(p + 2q)$-self-concordant barrier

$$\Phi(t, y, s) = -\sum_{i=1}^{p} \ln t_i + \sum_{i=1}^{q} G(s_i, y_i), \tag{7.3}$$

with the Legendre transformation

$$\Phi_*(\tau, \eta, \sigma) = -p - \sum_{i=1}^{p} \ln(-\tau_i) + \sum_{i=1}^{q} G_*(\sigma_i, \eta_i), \tag{7.4}$$

[3)] The Newton iteration

$$\theta_t = \frac{\theta_{t-1}^2}{\theta_{t-1} + 1}\left[1 + \frac{2}{\theta_{t-1}} - \ln\theta_{t-1} - r\right], \quad \theta_0 \equiv \begin{cases} \exp\{-r\}, & r \le 1 \\ \frac{1}{r - \ln(r - \ln r)}, & r > 1 \end{cases}$$

converges to $\theta(r)$ quadratically, and it takes at most 6 steps to compute $\theta(r)$ within relative accuracy 10^{-15} in the entire range of values of r where $10^{-400} \le \theta(r) \le 10^{400}$. With $\theta(r)$ computed, the derivatives of the function are readily available: $\theta'(r) = -\frac{\theta^2(r)}{\theta(r)+1}$, $\theta''(r) = \frac{\theta^2(r) - 2\theta(r)}{[\theta(r)+1]^2}\theta'(r)$.

and we can apply the primal-dual machinery we have developed to get new families of polynomial time interior-point methods for Entropy Optimization, an important generic problem which, in particular, possesses very interesting applications in graph theory (see [3, 9]). (To the moment, there exists just one dedicated polynomial time algorithm for Entropy Minimization [20]).

We can handle similarly many other convex programs where the feasible set can be represented as the inverse image, under affine mapping, of a direct product of sets of the form $f_i(y_i) \leq s_i$ with univariate f_i. In fact, the family of problems we can handle is pretty rich. Indeed, let us say that an "essentially open" ($Q = \operatorname{rint} Q$) convex domain $Q \subset \mathcal{E}$ is *well-representable*, if it admits a representation

$$Q = \{x \in \mathcal{E} \mid \exists u \in \mathcal{E}' : (Ax + Bu + b) \in \mathrm{Dom}\Phi\}, \tag{7.5}$$

where Φ is a self-concordant barrier with known Legendre transformation. Whenever the relative interior Q of the feasible set of a convex program $\min\limits_{x \in \mathrm{cl} Q} c^T x$ is well-representable and we are given a representation (7.5) for Q, we can rewrite our problem equivalently as $\inf\limits_{x,u} \left\{c^T x : (Ax + Bu + b) \in \mathrm{Dom}\Phi\right\}$, thus arriving at a problem which fits our framework. On the other hand, it is easily seen that the family $\mathcal{F}$ of well-representable domains is closed w.r.t. basic convexity-preserving operations, specifically, taking direct products, intersections and images/inverse images under affine mappings (cf. "calculus of coverings" in [14] or "calculus of Conic Quadratic/Semidefinite Representable sets in [1]). Note that $\mathcal{F}$ is much wider than the family of all domains over which we can minimize by the existing primal-dual interior point techniques (these are exactly the domains which can be well-represented via logarithmically homogeneous barriers for cones) and contains, e.g., domains given by semidefinite *and* Geometric Programming constraints.

We conclude this discussion with one more example which demonstrates that our framework may have (at least theoretical) advantages even in the case where an excellent conic formulation is readily available. Assume that our decision vector is an $m \times n$ matrix u, $m \leq n$, which should satisfy the norm bound $\|u\| \leq 1$, where $\|\cdot\|$ is the standard matrix norm (maximum singular value); for the sake of definiteness, let there be no other constraints (the conclusion to follow remains intact when allowing for no more than m "simple" – linear or quadratic – additional constraints on u). The standard way to process our problem is to express the norm bound by the LMI

$$\begin{bmatrix} I_{m\times m} & u \\ u^T & I_{n\times n} \end{bmatrix} \succeq 0$$

and to treat the problem as a semidefinite program; with this approach, the theoretical iteration count per given accuracy will be proportional to $\sqrt{m+n}$. At the same time, the domain $\mathcal{U} = \{u : \|u\| < 1\}$ of our problem admits the representation

$$\mathcal{U} = \{u : (I, u) \in \mathrm{Dom}\Phi,\ \Phi(x,u) = -\ln \mathrm{Det}(x - uu^T)\},$$

where x belongs to the space $\mathbf{S}^m$ of $m \times m$ symmetric matrices. Let us use the inner product

$$\langle (x,u), (y,v) \rangle \equiv \mathrm{Tr}(xy) + \mathrm{Tr}(v^T u)$$

on $\mathbf{S}^m \times \mathbf{R}^{m\times n}$. Note that Φ is m-self-concordant barrier (see [14]) with the explicit Legendre transformation

$$\Phi_*(y,v) = -\ln \mathrm{Det}(-y) - \frac{1}{4}\mathrm{Tr}(vy^{-1}v^T) - m \qquad [\mathrm{Dom}\Phi_* = \{(y,v) : y \prec 0\}]$$

so that the problem fits our framework *with the parameter of self-concordance of the barrier equal to m*. Consequently, the complexity bound for the primal-dual methods we have developed is proportional to $\sqrt{m}$, which, for $m << n$, is much better than the "standard" $O(\sqrt{m+n})$-complexity bound.

7.2. Long steps. We consider three related viewpoints:

(a) α-regularity of s.c.b. [15];

(b) convexity of the "gradient product" $\langle -H'(x), y \rangle$ [16, 17];

(c) β-normality of s.c.b. [11].

All of these properties are strengthenings of the fundamental property of the self-concordant barriers which says that the Hessian of a s.c.b. behaves very well inside the Dikin ellipsoid (see SC.II), anywhere in the interior of the domain. Each of the three notions tries to make this property valid in a wider region than the Dikin ellipsoid, with the ultimate goal to understand "how long are the long steps" yielded by path-following (or potential reduction) methods with on-line stepsize policies.

(a) Let f be a s.c. function with $Q = \text{Dom} f \subset \mathcal{E}$. f is called *α-regular* if

$$\left| \frac{d^4}{dt^4}\bigg|_{t=0} f(x+th) \right| \leq \alpha(\alpha+1) \left(\frac{d^2}{dt^2}\bigg|_{t=0} f(x+th) \right) \left[\pi_{Q,x}(h)\right]^2, \ \forall x \in Q, h \in \mathcal{E},$$

where

$$\pi_{Q,x}(h) \equiv \inf \left\{ \frac{1}{\mu} : \mu > 0, (x \pm \mu h) \in Q \right\}.$$

It was shown in [15] that many useful s.c.b.'s are α-regular with a quite moderate value of α. The examples include: the standard s.c.b.'s for the Lorentz and the semidefinite cone (both are 2-regular), the aforementioned barrier for Geometric Programming (and its Legendre transformation) and the barrier for Entropy (all are 6-regular). Besides this, α-regularity is preserved under summation of barriers and affine substitution of argument, see [15]. The fact that the universal barrier for a convex set is $O\left(\vartheta^2\right)$-regular was shown in [7]. We note that the barrier $\left[-\ln \text{Det}\left(x - uu^T\right)\right]$ with the domain

$$\left\{(x,u) \in \mathcal{E} \times \mathbf{R}^{n \times m} : \ \left(x - uu^T\right) \succeq 0\right\}.$$

(see above) is also 2-regular. Indeed, we have

$$\begin{pmatrix} I & 0 \\ 0 & x - uu^T \end{pmatrix} = \begin{pmatrix} I & 0 \\ -u & I \end{pmatrix} \begin{pmatrix} I & u^T \\ u & x \end{pmatrix} \begin{pmatrix} I & -u^T \\ 0 & I \end{pmatrix}.$$

Therefore,

$$\text{Det} \begin{pmatrix} I & u^T \\ u & x \end{pmatrix} = \text{Det} \begin{pmatrix} I & 0 \\ 0 & x - uu^T \end{pmatrix} = \text{Det}\left(x - uu^T\right).$$

Since $-\ln \text{Det} \begin{pmatrix} I & u^T \\ u & x \end{pmatrix}$ is 2-regular by the results of [15], it follows that $-\ln \text{Det}(x - uu^T)$ is also 2-regular for its domain. Actually, it is now known that all hyperbolic barriers are 2-regular (see Theorem 4.2 of [8]). The above fact can also be easily obtained using an affine restriction of this theorem. As a final remark on α-regularity, we note that this property behaves very nice under the symmetries of the domain of the s.c.b. For instance, if Q is a cone and A is an automorphism of it such that for a self-concordant barrier f for Q, we have the $f(x)$ and $f(Ax)$ differing only by a constant depending only on A, then the kth derivative of f at Ax along the direction Ah coincides with the kth derivative of f at x along h. Moreover, as it is easily seen, $\pi_{Q,Ax}(Ah) = \pi_{Q,x}(h)$. Therefore, if the automorphism group $\text{Aut}(Q)$ of Q acts transitively on Q and the barrier f in question is "semi-invariant" ($f(Ax) = f(x) + \text{constant}(A)$ for every $A \in \text{Aut}(Q)$), then it suffices to check the α-regularity condition at a single point of Q (but along every direction).

(b) Let H be a self-scaled barrier for K (so K is a symmetric cone). Define

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

Then

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

for every $x \in \operatorname{int} K$, $h \in \mathcal{E}$ and $t \in [0, 1/\sigma_x(h))$. This property was proven via establishing the convexity of the function $\langle -H'(x), y\rangle : \operatorname{int} K \to \mathbf{R}$, for every $y \in K$ [16]. Later, this property was extended to all hyperbolic barriers [8].

(c) f is β-normal if for every $x, z \in Q$, $r \equiv \pi_{Q,x}(z-x) < 1$ implies

$$(1-r)^\beta \left(\frac{d^2}{dt^2}\bigg|_{t=0} f(x+th)\right) \leq \left(\frac{d^2}{dt^2}\bigg|_{t=0} f(z+th)\right) \leq \frac{1}{(1-r)^\beta}\left(\frac{d^2}{dt^2}\bigg|_{t=0} f(x+th)\right), \ \forall h \in \mathcal{E}.$$

It is known that all specific examples discussed here are β-normal for moderate values of β (see [11]).

Our approach is very flexible to take advantage of any of the aforementioned good properties of special self-concordant barriers (for the related results in the context of predictor-corrector path-following methods, see [15, 11]).

7.3. Primal-dual symmetry and dual information. The setting of self-scaled barriers is ideal for the strongest use of primal-dual symmetry in interior-point algorithms. However, taking all of these nice properties beyond symmetric cones is not possible (see, for instance [21]).

In most applications, the importance of generating good bounds (via good dual feasible solutions) on the optimal objective value of the problem at hand cannot be denied. In the self-scaled case, the dual is proven to be even more powerful in that good dual solutions are also used (via so-called "primal-dual joint scaling") to generate excellent search directions for both primal and the dual problems. Some properties of primal-dual joint scaling interior-point methods have been generalized and extended to all convex optimization problems in conic form (see [22]). We can use analogous search directions in our set-up as well.

An important advantage of the current set-up is that when we are in the "Good Case", the primal and the dual paths are "asymmetric": the primal path is comprised of minimizers of the penalized objective $t\langle c, x\rangle + \Phi(Ax)$, while the dual path is comprised of minimizers of Φ_* on "shifted affine planes" $A^*y = -tc$; unless Φ_* is logarithmically homogeneous, the dual path is *not* of the same nature as the primal one.[4] This asymmetry may make the task of tracing one of the paths more relevant and/or easier for the interior-point approach. In such a case, the flexibility of our approach allows us to focus on the problem which has the s.c.b. with better long-step properties (we can also switch the focus of the algorithm from one problem to the other dynamically depending on the progress of the algorithm). Moreover, we still use the dual problem to generate improved lower bounds on c_* and guide the search directions.

7.4. Infeasible-start. As we already commented, the standard initialization techniques as given in [14] can be applied. We could also apply the surface-following idea developed in [15]. However, a particularly attractive choice would be an effective analogue of the approach of [19]. Such analogues seem possible and the development of such techniques is left for future work.

REFERENCES

[1] Ben-Tal, A., Nemirovski, A., *Lectures on Modern Convex Optimization*, MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2001.

[2] Dawson, J. L., Boyd, S. P., Hershenson, M., Lee, T. H. "Optimal allocation of local feedback in multistage amplifiers via geometric programming", *IEEE Transactions on Circuits and Systems I* **v.** 48 (2001), 1–11.

[3] Csiszár, I., Körner, J., Lovász, L., Marton, K., Simonyi, G. "Entropy splitting for antiblocking corners and perfect graphs", *Combinatorica* **v.** 10 (1990), 27–40.

[4] Dembo, R. S. "A set of geometric programming test problems and their solutions", *Math. Prog. A* **v.** 10 (1976), 192–213.

[5] Ecker, J. G. "Geometric programming: methods, computations and applications", *SIAM Review* **v.** 22 (1980), 338–362.

[6] Freund, R. W., Jarre, F., Schaible, S. "On self-concordant barrier functions for conic hulls and fractional programming", *Math. Prog. A* **v.** 74 (1996), 237–246.

[4] The idea to solve the problem by tracing the primal path is, of course, a common place. The idea to trace what we call here the dual path is not new either (it originates from Nesterov [12]; for a more general treatment, see [14], Section 3.4). What is seemingly new (beyond the scope of the Standard case, of course), is the idea to work with both of these paths simultaneously.

[7] Güler, O. "On the self-concordance of the universal barrier function", *SIAM J. Optim.* **v.** 7 (1997), 295–303.

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

[9] Kahn, J., Kim, J. H. "Entropy and sorting", 24th Annual ACM Symposium on the Theory of Computing (Victoria, BC, 1992), *J. Comput. System Sci.* **v.** 51 (1995), 390–399.

[10] Kortanek, K. O., Xu, X., Ye, Y. "An infeasible interior-point algorithm for solving primal and dual geometric programs", *Math. Prog. B* **v.** 76 (1997), 155–181.

[11] Nemirovski, A. (1997). *On normal self-concordant barriers and long-step interior-point methods*, Report, Faculty of IE&M, Technion, Haifa, Israel.

[12] Nesterov, Yu., "The method for Linear Programming which requires $O(n^3L)$ operations", *Ekonomika i Matem. Metody* **v.** 24 (1988), 174-176 (in Russian; English translation in *Matekon: Translations of Russian and East European Math. Economics*).

[13] Nesterov, Yu., "Long-step strategies in interior point primal-dual methods", *Math. Prog. B* **v.** 76 (1997), 47-94.

[14] Nesterov, Yu., Nemirovskii, A. *Interior point polynomial methods in Convex Programming.- SIAM Series in Applied Mathematics*, SIAM: Philadelphia, 1994

[15] Nesterov, Yu., Nemirovski, A. "Multi-parameter surfaces of analytic centers and long-step surface-following interior point methods", *Mathematics of Operations Research* **v.** 23 (1998), 1–38.

[16] Nesterov, Yu., Todd, M. J. "Self-scaled barriers and interior-point methods for convex programming", *Mathematics of Operations Research* **v.** 22 (1997), 1–46.

[17] Nesterov, Yu. E., Todd, M. J. "Primal-dual interior-point methods for self-scaled cones", *SIAM J. Optim.* **v.** 8 (1998), 324–364.

[18] Nesterov, Yu. E., Todd, M. J. "On the Riemannian geometry defined by self-concordant barriers and interior-point methods", *Foundations of Computational Mathematics*, to appear.

[19] Nesterov, Yu., Todd, M. J., Ye, Y. "Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems", *Math. Prog. A* **v.** 84 (1999), 227–267.

[20] Potra, F., Ye, Y. "A quadratically convergent polynomial algorithm for solving entropy optimization problem", *SIAM. J. Optim.* **v.** 3 (1993), 843–860.

[21] Tunçel, L. "Primal-dual symmetry and scale-invariance of interior-point algorithms for convex optimization", *Mathematics of Operations Research* **v.** 23 (1998), 708–718.

[22] Tunçel, L. "Generalization of primal-dual interior-point methods to convex optimization problems in conic form", *Foundations of Computational Mathematics* **v.** 1 (2001), 229–254.