MATRIX SCALING DUALITIES IN CONVEX PROGRAMMING*

BAHMAN KALANTARI†

Dedicated to the memory of Leonid Khachiyan

Abstract. We consider convex programming problems in a canonical homogeneous format, a very general form of Karmarkar's canonical linear programming problem. More specifically, by *homogeneous programming* we shall refer to the problem of testing if a homogeneous convex function has a nontrivial zero over a subspace and its intersection with a pointed convex cone. To this canonical problem, endowed with a normal barrier for the underlying cone, we associate dual problems and prove several *matrix scaling dualities.* We make use of these scaling dualities to derive new and conceptually simple potential-reduction and path-following algorithms, applicable to self-concordant homogeneous programming, as well as three dual problems defined as: the *scaling problem*, the *homogeneous scaling problem*, and the *algebraic scaling problem.* The simplest of the scaling dualities is the following equivalent of the classic separation theorem of Gordan: a positive semidefinite symmetric matrix Q either has a nontrivial nonnegative zero, or there exists a positive definite diagonal matrix D such that $DQDe > 0$, where e is the vector of ones. This duality is a key ingredient in the very simple path-following algorithm of Khachiyan and Kalantari for linear programming, as well as for quasi doubly stochastic scaling of Q, i.e. computing D such that $DQDe = e$. Our general results here give nontrivial extensions of our previous work on the role of matrix scaling in linear or semidefinite programming, when formulated as a homogeneous program. To establish the general results we associate a cone of linear operators induced by the normal barrier, called *operator-cone* and use it to reveal the intrinsic nature of scaling dualities corresponding to homogeneous programming formulation of convex programs. Our general algorithms although make use of some basic properties from the self-concordance theory of Nesterov and Nemirovskii, offer new algorithms for linear programming, quadratic programming, semidefinite programming, self-concordant programming itself, as well as for the corresponding scaling problems. Scaling dualities also result in generalizations of the classic arithmetic-geometric mean and the trace-determinant inequalities. Our collective results reveal the ubiquitous nature of matrix scaling in convex programming.

Key words. convex programming, matrix scaling, convex cones, homogeneous programming, interior-point algorithms, self-concordance, Newton's method, duality

AMS subject classifications. 90C05, 90C20, 90C25, 90C26, 90C27, 90C30, 68Q25, 15A15, 15A48, 15A60, 46A20, 65F35

*This article is based on two unpublished technical reports by the author, "Scaling dualities and self-concordant homogeneous programming in finite dimensional spaces," Technical Report DCS-TR-39, and "A note on a boundedness property of normal barriers on convex sets," Technical Report LCSR-TR-398.

†Department of Computer Science, Rutgers University, New Brunswick, NJ 08903 (kalantar@cs.rutgers.edu)

1. Introduction. In this paper we consider four fundamental problems in finite dimensional spaces defined as: *homogeneous programming* (HP), *scaling problem* (SP), *homogeneous scaling problem* (HSP), and *algebraic scaling problem* (ASP). HP is the problem of computing a nontrivial zero of a homogeneous convex function over a pointed closed convex cone and its intersection with a subspace, or proving the nonexistence of such zero. HP is a canonical formulation of convex programming problems (possible via known conic linear programming dualities). The significance of HP in linear programming was established by the pioneering work of Karmarkar in 1984. SP is the minimization of an associated *logarithmic potential* function, or proving it is unbounded. HSP is analogous to SP, replacing the logarithmic potential by a *homogeneous potential* function. The classic arithmetic-geometric mean inequality can be viewed as an HSP: the minimization of the ratio of the two means over the interior of the cone of nonnegative vectors. Although it is well known that this minimum is attained at the center of the cone, the vector of ones, in the presence of a subspace not containing the center, the new minimization problem becomes nontrivial. ASP is to test the solvability of an algebraic equation, inherited from homogeneity, to be referred as the *scaling equation* (SE). It is a generalization of the problem of testing the solvability of the diagonal matrix scaling equation for a given symmetric positive semidefinite matrix.

There are intrinsic theorems of the alternatives relating exact and approximate solutions of these four problems. We will prove several distinct such theorems, and will refer to these theorems as *matrix scaling dualities*, or simply *scaling dualities*. These dualities surpass the ordinary dualities of mathematical programming, be it for linear programming, or general convex programming problems.

The HP formulation of convex programming problems offers new insight into the theory of this optimization problem in the sense that the three scaling problems, SP, HSP, and ASP come to life. In view of the scaling dualities, given an HP formulation of a convex program, the corresponding three scaling problems are all more general and fundamental problems than HP itself. This is because scaling dualities allow us to view any of the corresponding scaling problems as a problem *dual* to HP. This duality is in the sense that HP is solvable if and only if the other three are not. Moreover, the dual problems can be solved via interior-point algorithms that exploit homogeneity, convexity, scaling dualities, and the conic structure of the domain of the optimization problem.

On the one hand, scaling dualities give rise to novel algorithms for convex programming problems. These algorithms attempt to solve SP, HSP, or ASP, and if any of these problems is found to be unsolvable, then as a by-product, HP is solvable. For instance, using one of our scaling dualities, we describe a potential-reduction algorithm that simplifies, strengthens, and enhances Nesterov and Nemirovskii's corresponding algorithm for solving a homogeneous conic LP, a canonical homogeneous formulation of convex programs, and a generalization of Karmarkar's canonical LP. This particular scaling duality makes it possible to abandon the unnecessary assumption that the minimum value of this conic LP is zero. Such assumption has often been made in the interior-point literature, even in the textbook description of Karmarkar's canonical LP. On the other hand, scaling dualities give rise to new inequalities that generalize the classic arithmetic-geometric mean, trace-determinant, and Hadamard inequalities. Moreover, scaling dualities give solvability theorems on the scaling equation generalizing known diagonal matrix scaling theorems, also theorems of the alternative generalizing the classic separation theorem of Gordan.

Scaling dualities relate several seemingly unrelated problems. Consider for instance the minimization of the arithmetic-geometric mean ratio of positive numbers over an arbitrary subspace of the Euclidean space. If in this ratio we replace the arithmetic mean by an arbitrary linear function, then via scaling dualities the new minimization problem becomes an important dual problem to a canonical formulation of linear programming. Likewise, semidefinite programming can be viewed as a dual problem to a problem that is a generalization of the minimization of the classic trace-determinant ratio, over the semidefinite cone and its intersection with an arbitrary subspace of the Hilbert space of symmetric matrices. More generally, an analogous relationship can be stated for any self-concordant programming.

Despite the tremendous literature on interior-point algorithms for linear, quadratic, linear complementarity, semidefinite programming, or self-concordant programming, emerged since Karmarkar's LP algorithm, few papers have dealt with HP, and even fewer with the other three scaling problems, or the dualities relating them. Even in the self-concordance theory of Nesterov and Nemirovskii, despite the development of a new conic LP duality which makes it possible to state the HP formulation of convex programs, there is no mention of scaling dualities. In fact, since the introduction of Karmarkar's canonical linear programming in 1984, there has been somewhat of a divergence from viewing optimization problems and convex programming in the context of a corresponding HP. This divergence resulted in several breakthroughs on convex programming, such as Renegar's work on the complexity of the path-following method of centers for linear programming, and Nesterov and Nemirovskii's theory on the class of self-concordant convex programming. However, as we shall establish in this paper, the HP formulation of convex programs, initiated by Karmarkar for LP, has many advantages of its own as it gives rise to the scaling problems, the scaling dualities, and new algorithms. For instance, we will see that it is possible to state a conceptually simple path-following algorithm for Karmarkar's canonical LP (hence, for linear programming), requiring the minor modification of squaring its linear objective function.

On the one hand, from the theoretical point of view there is no loss of generality in restricting oneself to the HP formulation of convex programs. On the other hand, the HP formulation of convex programs is not necessarily a mere theoretical formulation. In many practical instances, a convex program can directly be formulated as an HP. For instance, the feasibility problem in linear programming can often be trivially reduced to the geometric problem of testing if the convex-hull of a given set of points contains the origin. In fact this simple geometric problem, considered by Gordan more than a century ago, is equivalent to several HP formulation of linear programming, as well as Karmarkar's canonical LP.

In this paper we employ basic but important properties of homogeneous functions to prove scaling dualities, derive several significant bounds, and by using these and by building on Nesterov and Nemirovskii's machinery of self-concordance, we obtain novel polynomial-time potential-reduction and path-following algorithms for the four problems stated above. In particular, we obtain new polynomial-time algorithms for quadratic programming, semidefinite programming, self-concordant programming, and even linear programming itself. Furthermore, our potential-reduction and path-following algorithms are both more powerful than their existing counterparts, since they also establish the polynomial-time solvability of the corresponding SP, HSP, and ASP. Our result extend the polynomial-time solvability of matrix scaling equation (even in the presence of a subspace) to very general cases of the scaling equation, SE, over the nonnegative cone, the semidefinite cone, the second-order cone, and more generally any homogeneous programming with a beta-compatible objective convex function. The polynomial-time solvable cases of HSP, a nonconvex program, includes the deceptively simple, but nontrivial problem of computing the minimum of the arithmetic-geometric mean ratio of positive numbers over an arbitrary subspace of the Euclidean space, or the problem of computing the minimum of the trace-determinant ratio of positive definite symmetric matrices over an arbitrary subspace of symmetric matrices. We emphasize that the polynomial-time solvability of several of the problems considered in this paper has neither been proved, nor even addressed in the previous interior-point literature. Although some of our work builds on the theory of self-concordance, many of the results proved in this paper, including the definition and existence of the scaling problems, the operator-cones, the proof of scaling dualities and bounds, are new and indispensable ingredients in establishing the polynomial-time solvability of the scaling problems via algorithms which themselves are very novel. Moreover, from the point of view of the theory of convex programming, our scaling results are both natural and fundamental. They exist and are true whether or not one may wish to use them in practice.

2. Mathematical preliminaries. In this section we first formally define the homogeneous programming problem (HP), the scaling problem (SP), the homogeneous scaling problem (HSP), the algebraic scaling problem (ASP), and their ϵ-approximate versions. The notation and symbols to be used to define these problems will remain unchanged throughout the paper. We summarize some basic but essential properties of homogeneous functions, and introduce the notion of operator-cone needed to define the scaling equation (SE). We give the precise statement of several distinct scaling dualities that are to be proved in subsequent sections. We summarize the essential needed results from Nesterov and Nemirovskii's theory of self-concordance in the form of a single theorem (Theorem 2.24). This theorem will be repeatedly utilized throughout the paper. In this section we also describe the steps of two basic algorithms, a potential-reduction algorithm, and a path-following algorithm. In the subsequent sections we will establish the applicability of these algorithms in solving ϵ-approximate versions of HP, SP, HSP, and ASP. Having formally stated all the necessary ingredients, we then give a summary of the precise results to be proved later.

2.1. HP, SP, HSP, and ASP in finite dimensional spaces. Let E be a finite dimensional normed vector space over the reals. In particular, since E is necessarily complete, it is a Banach space. Let K be a closed convex pointed cone in E. In particular, K contains all nonnegative scalar multiples of its elements, and does not contain a line, i.e., $K \cap -K = \{0\}$. The interior of K will be denoted by K°, and is assumed to be nonempty. We assume that we are given a subspace of E intersecting K°, described as $W = \{x \in E : \langle a_i, x\rangle = 0,\ i = 1, \dots, m\}$, where a_i lies in E^*, the dual space of E, and $\langle a_i, x\rangle$ denotes the value of the functional a_i at x. We may assume a_i's are linearly independent. Assuming that the norm in E is induced by an inner product $\langle\cdot,\cdot\rangle$, then via the theorem of Riesz, E^* can be identified with E itself, and in particular each a_i can be taken to be an element of E with $\langle a_i, x\rangle$ representing the given inner product. Given an orthonormal basis $\{e_1, \dots, e_n\}$, for each $i = 1, \dots, m$, a_i can be represented as $\sum_{j=1}^n \alpha_{ij} e_j$. Thus, W can be written as $W = \{x \in E : Ax = 0\}$, where A is an $m \times n$ matrix of rank m. The orthogonal projection onto W is the matrix $P = I - A^T(AA^T)^{-1}A$. If $W = E$, then $P = I$, the identity operator.

A cone such as K is also called an *order cone* since the following relation defines a partial order on E: given $u, v \in E$ we write

$$v \geq u \iff v - u \in K \cap W, \qquad v > u \iff v - u \in K^\circ \cap W. \tag{2.1}$$

From (2.1), we can write

$$x \geq 0 \iff x \in K \cap W, \qquad x > 0 \iff x \in K^\circ \cap W. \tag{2.2}$$

A point $x \in E$ is said to be *positive* (*nonnegative*), if $x > 0$ ($x \geq 0$). Throughout the paper we will use the notation $x > 0$ ($x \geq 0$) in the sense of (2.2).

Let $\phi(x)$ be a real-valued function, twice continuously differentiable over $K^\circ \cap W$, also homogeneous of degree $p > 0$, i.e., for each scalar $\alpha > 0$, and each $x > 0$,

$$\phi(\alpha x) = \alpha^p \phi(x). \tag{2.3}$$

For simplicity, we will assume that ϕ has a continuous extension to the boundary of K. Throughout we will also assume ϕ is convex over $K \cap W$.

Definition 2.1. *Homogeneous programming (HP) is to of test if the quantity*

$$\mu = \min\{\phi(x) : x \geq 0,\ \|x\| = 1\},$$

is nonnegative, and if $\mu \leq 0$, to compute a point $d \geq 0$ such that $\phi(d) \leq 0$, in which case HP is said to be "solvable", or to prove that such point does not exist. We will assume that a positive point x^0 is available. Without loss of generality we assume $\|x^0\| = 1$, and $\mu \leq 1$. For a given $\epsilon \in (0,1)$, ϵ-HP is to compute $d > 0$ satisfying

$$\phi\left(\frac{d}{\|d\|}\right) \leq \epsilon,$$

in which case ϵ-HP is said to be "solvable", or to prove that such a point does not exist.

Definition 2.2. *Let F be continuously differentiable, and strictly convex over K°. F is said to be a θ-logarithmically homogeneous barrier for K, if $\theta \geq 1$, and for each $x \in K^\circ$, and each scalar $t > 0$, we have*

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

and $F(x)$ approaches ∞, as x approaches a boundary point of K. The latter two conditions are equivalent to homogeneity of $exp(-F(x))$ (of degree θ), and that for each α the level set $K_\alpha(F) = \{x \in K^\circ : F(x) \leq \alpha\}$ is closed. F is said to be a θ-normal barrier, if in addition it is three times continuously differentiable, and for all $x \in K^\circ$, and $h \in E$, we have

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

where $D^k F(x)[h_1, \ldots, h_k]$ is the k-th Fréchet-differential of F at x in the direction of $h_1, \ldots, h_k$.

Definition 2.3. *Suppose $F(x)$ is a θ-normal barrier for K, and $\phi(x)$ is C^3 on K°. Then ϕ is said to be β-compatible (with F), $\beta \geq 0$, if it is convex, and for all $x \in K^\circ$, $h \in E$, we have*

$$|D^3\phi(x)[h,h,h]| \leq \beta \left(3D^2\phi(x)[h,h]\right)\left(3D^2F(x)[h,h]\right)^{1/2}.$$

Given a basis for E, any symmetric positive semidefinite matrix Q induces an inner product on E defined as $\langle x, y\rangle \equiv x^T Qy$. This inner product turns E into a Hilbert space. Thus we can identify $D^k F(x)$ (or $D^k \phi(x)$) with the multidimensional symmetric matrix $\nabla^k F(x) = (\partial^k F(x)/\partial x_{i_1} \cdots \partial x_{i_k})$ (or $\nabla^k \phi(x) = (\partial^k \phi(x)/\partial x_{i_1} \cdots \partial x_{i_k})$). In particular, ∇ and ∇^2 denote the gradient and the Hessian with respect to a basis induced by this inner product. Since E is finite dimensional, all norms are equivalent and convergence of sequences remain invariant. Thus, throughout the rest of the paper, we will assume that E is a Hilbert space. If E is already a Hilbert space, $\nabla^k F(x)$ is defined with respect to the original inner product.

DEFINITION 2.4. *The "logarithmic potential", and the "homogeneous potential" are*

$$\psi(x) = \phi(x) + F(x), \quad X(x) = \frac{\phi(x)}{exp(-\frac{p}{\theta}F(x))},$$

respectively. Let

$$\psi^* = \inf\{\psi(x) : x > 0\}, \quad X^* = \inf\{X(x) : x > 0\}.$$

DEFINITION 2.5. *The scaling problem (SP) is to determine if $\psi^* = -\infty$, and if $\psi^* > -\infty$, to compute its value, together with a corresponding minimizer d^*, called the "scaling vector". If d^* exists, SP is said to be "solvable". Given $\epsilon \in (0, 1)$, ϵ-SP is to determine if $\psi^* > -\infty$, and if so, to compute a point $d > 0$, satisfying $\psi(d) - \psi^* \leq \epsilon$. If such a point exists, ϵ-SP is said to be "solvable".*

DEFINITION 2.6. *The homogeneous scaling problem (HSP) is to determine if $X^* = -\infty$, and if $X^* > -\infty$, to compute its value, together with a corresponding minimizer $\bar{d}^*$, called the "homogeneous scaling vector". If $\bar{d}^*$ exists, HSP is said to be "solvable". Given $\epsilon \in (0, 1)$, ϵ-SP is to determine if $X^* > -\infty$, and if so, to compute a point $d > 0$, satisfying $X(d)/X^* \leq exp(\epsilon)$. If such a point exists, ϵ-HSP is said to be "solvable".*

Note that since ϕ is convex, SP is a convex programming problem. However, HSP is a nonconvex program. An important property of HSP is that $X(x)$ is homogeneous of degree zero. If d^* exists, from strict convexity of ψ, it is necessarily unique and from the Lagrange multiplier condition, it must satisfy

$$P\nabla\psi(d^*) = 0,$$

where P is the orthogonal projection operator onto W.

DEFINITION 2.7. *Assume that F is θ-normal. The "parameterized family of logarithmic potentials" is defined as*

$$f^t(x) = t\psi(x) + t\langle u, x\rangle + F(x), \quad t \in (0, \infty),$$

where $u = -P\nabla\psi(x^0)$. The "central-path" is defined as

$$\Pi = \{d_t^* : t \in (0, \infty)\}, \quad d_t^* = \operatorname{argmin}\{f^t(x) : x > 0\}.$$

The "spherical central-path", and the "homogeneous central-path" are, respectively

$$\Pi_S = \{\frac{d_t^*}{\|d_t^*\|} : t \in (0, \infty)\}, \quad \widehat{\Pi} = \{\hat{d}_t^* = t^{1/p} d_t^* : t \in (0, \infty)\}.$$

If ϕ is β-compatible with a θ-normal barrier $F(x)$, then ψ is strongly self-concordant with parameter $a = [1/(1+\beta)^2]$. The notion of β-compatibility and self-concordance are defined by Nesterov and Nemirovskii [17]. A function f is said to be *strongly self-concordant* over an open set G, with parameter a, if it is three times continuously differentiable over G; for each t, $G_t = \{x \in G : f(x) \leq t\}$ is closed; and for all $x \in G$, and $h \in E$, we have

$$D^3 f(x)[h,h,h] \leq \frac{2}{\sqrt{a}} (D^2 f(x)[h,h])^{3/2}. \tag{2.4}$$

With regard to our four conic problems, from the algorithmic point of view we are interested in strong self-concordance over cones. The significance of strong self-concordance in the context of β-compatibility is summarized as Theorem 2.24. This theorem is an adaptation of several significant result of Nesterov and Nemirovskii. It is a fundamental theorem instrumental in solving HP/SP/HSP, and ASP, which is the problem of testing the solvability of the scaling equation (SE), an algebraic equation induced from homogeneity and differentiability. Theorem 2.24 on self-concordance does not by itself solve these four problems. Additionally, we need the development of several essential results. In particular, the scaling dualities, and significant bounds.

Nesterov and Nemirovskii proved the existence of a universal strongly self-concordant barrier for any convex set with nonempty interior (see [17], Theorem 2.5.1), which for convex cones reduces to a θ-logarithmically homogeneous barrier. Since by multiplication of this barrier by an appropriate constant the parameter of self-concordance can be made to be one, this ensures the existence of a θ-normal barrier for the cone K of HP.

2.2. The scaling equation: algebraic characterization of the solvability of SP. In this section we will give the algebraic characterization of the solvability of SP. We will derive an equation that the scaling vector d^*, if it exists, must necessarily satisfy. We shall also obtain an analogous equation for d_t^*.

We shall first state the following proposition without proof, describing several essential properties of homogeneous functions. These properties will be used throughout the paper.

PROPOSITION 2.8. (Homogeneous properties) *Let K, F and ϕ be as before, and $\Gamma(x)$ an arbitrary twice continuously differentiable real-valued function, defined over K°, also homogeneous of degree κ. Given any $x \in K^\circ$, and α a positive real, we have*

(P1) $x^T \nabla \phi(x) = p\phi(x)$ (Euler's Equation).
(P2) $\nabla \Gamma(\alpha x) = \alpha^{\kappa-1} \nabla \Gamma(x)$.
(P3) $\nabla^2 \Gamma(\alpha x) = \alpha^{\kappa-2} \nabla^2 \Gamma(x)$.
(P4) $\nabla^2 \Gamma(x) x = (\kappa - 1) \nabla \Gamma(x)$.
(P5) $x^T \nabla F(x) = -\theta$.
(P6) $\nabla F(\alpha x) = \alpha^{-1} \nabla F(x)$.
(P7) $\nabla^2 F(\alpha x) = \alpha^{-2} \nabla^2 F(x)$.
(P8) $\nabla^2 F(x) x = -\nabla F(x)$.

Remark. Properties (P5) and (P6) of F are stated in Nesterov and Nemirovskii [17]. However, these are actually a consequence of the first two well-known properties, as applied to the function $\Gamma(x) = exp(-F(x))$.

DEFINITION 2.9. *Given $d > 0$, let D be any linear operator in $L(E,E)$, the space of continuous linear operators from E into itself, satisfying $D^T \nabla^2 F(d) D = I$,* *the identity operator. The "induced" center, subspace, and cone are respectively, the images of d, W, and K under the change of variable $x \leftarrow Dx$, i.e.,*

$$e_d = D^{-1}d, \quad W_d = D^{-1}W, \quad K_d = D^{-1}K.$$

The induced homogeneous function, θ-logarithmically homogeneous barrier function, and logarithmic potential function are, respectively

$$\phi_d(x) = \phi(Dx), \quad F_d(x) = F(Dx), \quad \psi_d(x) = \psi(Dx).$$

Note that each $d > 0$ induces a new HP, SP, and HSP, where the homogeneous degrees p and θ remain invariant. Moreover, an induced positivity can be defined:

$$x >_d 0 \quad \Longleftrightarrow \quad x \in W_d \cap K_d^{\circ}. \tag{2.5}$$

Definition 2.10. *The algebraic scaling problem (ASP) is to test the solvability of the "scaling equation" (SE) defined as:*

$$P_d \nabla \phi_d(e_d) = e_d, \quad d > 0,$$

where D is an operator satisfying $D^T \nabla^2 F(d) D = I$, and P_d is the orthogonal projection onto W_d, i.e., $P_d = I - D^T A^T (ADD^T A^T)^{-1} AD$.

Proposition 2.11. *If SP is solvable, then SE is solvable. Moreover, if $p \neq 1$, then SE is equivalent to the equation*

$$P_d D^T \nabla^2 \phi(d) D e_d = \frac{1}{p-1} e_d, \quad d > 0.$$

Proof. Let D be an operator satisfying $D^T \nabla^2 F(d) D = I$. If $g(x)$ represents either $\phi(x)$, or $F(x)$, or $\psi(x)$, and $g_d(x) = g(Dx)$, then from the chain rule we have

$$\nabla g_d(x) = D^T \nabla g(Dx), \tag{2.6}$$

$$\nabla^2 g_d(x) = D^T \nabla^2 g(Dx) D. \tag{2.7}$$

Now assume SP is solvable. Then, d^* (the minimizer of ψ) exists. For simplicity of notation, in the remaining of the proof we denote d^* by d. In particular, d is a stationary point of ψ over $W \cap K^{\circ}$. Thus,

$$\nabla \psi(d) = \nabla \phi(d) + \nabla F(d) = A^T v, \tag{2.8}$$

where v is the vector of Lagrange multipliers. Let $e_d = D^{-1}d$. From property (P8) of Proposition 2.8, $\nabla F(d) = -\nabla^2 F(d) d$. Using this and by applying the operator D^T to (2.8), together with the chain rule, see (2.6), we get

$$\nabla \psi_d(e_d) = \nabla \phi_d(e_d) - e_d = D^T A^T v, \tag{2.9}$$

Applying the operator AD to the above, using the fact that $ADe_d = Ad = 0$, and the invertibility of $ADD^T A^T$, we can solve for v to get

$$v = (ADD^T A^T)^{-1} AD \nabla \psi_d(e_d). \tag{2.10}$$

Substituting the above in (2.9), we get the scaling equation. The equivalence of the scaling equation to the claimed equation on the Hessian is a consequence of property (P4) of Proposition 2.8. ∎

DEFINITION 2.12. *Given $\epsilon > 0$, ϵ-ASP is to compute $d > 0$, such that*

$$\|P_d \nabla \psi_d(e_d)\| < \epsilon,$$

or to prove that such d does not exist. Also, given $t \in (0, \infty)$, ϵ-ASP(t) is to compute $d > 0$ satisfying

$$\|P_d \nabla f_d^t(e_d)\| < \epsilon.$$

The problem ϵ-ASP(t) corresponds to the approximation of the minimizer of $f^t(x)$. Note that ϵ-ASP, and ϵ-ASP(t) are defined with respect to an operator D induced by d, satisfying $D^T \nabla^2 F(d) D = I$. However, these problems can be defined in more generality as will be seen next.

2.3. Operator-cones. The following definition describes ASP, ϵ-ASP, and ϵ-ASP(t) in more generality and at the same time reveals some very important properties of the operators induced by the Hessian of the θ-logarithmically homogeneous barrier F. In subsequent sections these properties will be used heavily.

DEFINITION 2.13. *Given an HP, a subset of invertible operators in $L(E, E)$ is said to be an operator-cone, denoted by $T(K)$, if for each $d \in K^\circ$ there exists an operator D, also denoted by $T(d)$, such that*

(1) *If $d' = \alpha d$, α a positive scalar, then $D' = \alpha D$.*
(2) *For all $d \in K^\circ$, $\|e_d\|^2 \leq N < \infty$, where $e_d = D^{-1} d$.*
(3) *If $d^k > 0$ converges to $d > 0$, then D^k converges to D.*
(4) *There exist positive constants m and M such that for each $d > 0$, we have*

$$m\|x\|^2 \leq x^T \nabla^2 F_d(e_d) x \leq M\|x\|^2.$$

Without loss of generality we may assume that $m = 1$.

DEFINITION 2.14. *An operator-cone $T(K)$ is said to be bounded if there exists a fixed constant ρ such that*

$$\|D\| \leq \rho \|d\|, \quad \forall\ d \in K^\circ.$$

Remark. The operator-cone $T(K)$ is indeed a cone in $L(E, E)$, the space of continuous linear operators from E into itself, and it can be viewed as the image of a continuous homogeneous operator, T, of degree one from K into $L(E, E)$. In particular, if T is linear, then the number ρ is its operator norm.

PROPOSITION 2.15. *Let $T_F = \{D = \nabla^2 F(d)^{-1/2} : d \in K^\circ\}$. Then, T_F is an operator-cone with parameters $m = M = 1$, $N = \theta$ (see Definition 2.13).*

Proof. From property (P7) of Proposition 2.8, we have $\nabla^2 F(\alpha d) = \alpha^{-2} \nabla^2 F(d)$. Thus, $\nabla^2 F(\alpha d)^{1/2} = \alpha^{-1} \nabla^2 F(d)^{1/2}$. But this implies the validity of condition (1) of Definition 2.13. We have $\|e_d\|^2 = d^T D^{-2} d = d^T \nabla^2 F(d) d$. But from properties (P8) and (P5) of Proposition 2.8, we have $d^T \nabla^2 F(d) d = -d^T \nabla F(d) = \theta$. Hence, the validity of condition (2) of Definition 2.13. Clearly, condition (3) is valid. The validity of (4) is a consequence of the chain rule, see (2.7), and the definition of D, i.e., since $\nabla^2 F_d(e_d) = D^T \nabla F(d) D = I$. Hence, $m = M = 1$. ∎

Given any operator-cone $T(K)$, we can define ϵ-ASP and ϵ-ASP(t), analogous to the case of $T(K) = T_F = \{D = \nabla^2 F(d)^{-1/2} : d \in K^\circ\}$. We shall prove that when

ϕ is β-compatible, we can solve ϵ-HP, ϵ-SP, and ϵ-HSP in polynomial time. Given a bounded operator-cone, we can also solve the corresponding ϵ-ASP in polynomial time.

To understand further what it means for an operator-cone to be bounded, consider the following result.

PROPOSITION 2.16. *The operator-cone T_F is a bounded operator-cone with parameter ρ if and only if for all $d \in K^\circ \cap \{x : \|x\| = 1\}$, the minimum eigenvalue of $\nabla^2 F(d)$ is bounded below by $1/\rho^2$.*

Proof. Let $d \in K$, and assume that $\|D\| \leq \rho \|d\|$. Equivalently, this implies for all $x \in E$, we have

$$\frac{x^T D^2 x}{x^T x} = \frac{x^T \nabla^2 F(d)^{-1} x}{x^T x} \leq \rho^2 \|d\|^2.$$

Letting $x = \nabla^2 F(d) y$, for all $y \in E$ we must have

$$\frac{y^T \nabla^2 F(d) y \|d\|^2}{y^T y} \geq \frac{1}{\rho^2}.$$

But from property (P7) of Proposition 2.8, replacing d by αd in the above does not change the ratio. Thus, we may assume $\|d\| = 1$. The converse trivially follows. □

The next theorem is a significant result showing that T_F is a bounded operator-cone. For the three important cones, the nonnegative orthant $K = \Re^n_+ = \{x \in \Re^n : x \geq 0\}$, the cone of positive semidefinite symmetric matrices $K = S_n^+$, and the second-order cone (or Lorentz cone) $K = SO^{n+1} = \{x = (\tau, z) \in \Re^{n+1} : \tau \geq \|z\|\}$, it is easy to directly prove this boundedness property with respect to their natural barriers (Kalantari [9]). For these cones it can be shown that the quantity ρ is $1, 1$, and $\sqrt{2}, *respectively*$. Next we state the general theorem but omit its proof here.

THEOREM 2.17. *(Kalantari [10]) For the cone K define*

$$r = \inf \big\{ \frac{\langle x, y \rangle}{\|x\| \|y\|} : x, y \in K, \quad \|x\| \neq 0, \|y\| \neq 0 \big\}.$$

Then T_F is a bounded operator-cone with parameter

$$\rho \leq 1 + \sqrt{\frac{2}{1+r}}.$$

Remark. Clearly the quantity r is at least -1. However since K is a closed convex pointed cone, it follows that $r > -1$. The quantity r is a measure of obtuseness of K.

The theorem has an immediate corollary:

COROLLARY 2.18. *Assume K is acute, i.e. $\langle x, y \rangle \geq 0$ for all $x, y \in K$. Then T_F is a bounded operator-cone with parameter $\rho \leq 1 + \sqrt{2}$. In particular this is the case whenever K is self-dual, i.e. $K = \{x \in E : \langle x, y \rangle \geq 0, \forall y \in K\}$. .*

As we shall see in this paper, the boundedness of the underlying operator-cone will play a significant algorithmic role in all the four problems HP, SP, HSP, and ASP.

2.4. Copositive-plusness: an algebraic characterization of the solvability of HP. In order to characterize the solvability of HP, we first define a notion of copositive plusness.

DEFINITION 2.19. *Assume that in a given HP the homogeneous function ϕ is defined over the entire subspace W. Then, ϕ is said to be "copositive plus", if whenever $\phi(x) = 0$, $x \in W \cap K$, then $P\nabla\phi(x) = 0$.*

Note that in the very special case of HP where $\phi = \frac{1}{2}x^T Qx$, Q a symmetric matrix, $W = \Re^n$, and $K = \Re^n_+$, the above definition which coincides with the usual definition of copositive plusness, despite its generality, is the most informative notion of copositive plusness. It simply implies that a zero of ϕ is a stationary point over $W \cap K$. The following result characterizes the solvability of HP under copositive plusness.

PROPOSITION 2.20. *Assume that ϕ is copositive plus. Then, given $x \in W \cap K$, $P\nabla\phi(x) = 0$ if and only if $\phi(x) = 0$.*

Proof. Clearly, one direction follows from the definition. Suppose $P\nabla\phi(x) = 0$. Since $Px = x$, from Euler's equation, i.e., property (P1) of Proposition (2.8), it follows that $x^T P\nabla\phi(x) = p\phi(x)$. Since $p > 0$, $\phi(x) = 0$. □

The following result characterizes the solvability of HP when homogeneous degree $p > 1$.

PROPOSITION 2.21. *Suppose in a given HP the homogeneous convex function ϕ is defined over W and homogeneous degree $p > 1$. Then, ϕ is copositive plus. In particular, from convexity, $\mu = 0$ if and only if there exists $x \geq 0$, $x \neq 0$, such that $P\nabla\phi(x) = 0$.*

Proof. Fix $x \in W \cap K$, $x \neq 0$. We have $\phi(0) = \lim_{\alpha\to 0} \phi(\alpha x) = \phi(x) \lim_{\alpha\to 0} \alpha^p = 0$. Also, $\nabla\phi(0) = \lim_{\alpha\to 0} \nabla\phi(\alpha x) = \nabla\phi(x) \lim_{\alpha\to 0} \alpha^{p-1} = 0$. In particular, the origin is a global minimizer of ϕ over W, and the minimum value is zero. Thus, any zero of ϕ over W is also a minimizer, hence a stationary point. □

Proposition 2.21 is not true if $p = 1$, e.g. when ϕ is linear. From this proposition and additional results we can obtain a theorem that generalizes Gordan's Theorem (see Theorem 7.2 in [9]), and a matrix scaling duality for positive semidefinite symmetric matrices (see Theorem 5.7 in [9]). According to this generalization of Gordan's Theorem, for any convex HP with $p > 1$, defined over the cones $\Re^n_+$, S^+_n, or SO^{n+1}, either there exists a nonnegative nontrivial point whose projected gradient is zero, or there exists a positive point whose scaled projected gradient is positive.

2.5. Definition of scaling dualities. The following definition classifies some important dualities to be proved and utilized later in the article.

DEFINITION 2.22. (Scaling dualities) *Let $T(K)$ be a given operator-cone.*

The Weak Scaling Duality *holds if : either $\mu \leq 0$, or $P_d\nabla\phi_d(e_d) = e_d$ for some $d > 0$.*

The Scaling Duality *holds if : $\mu > 0$ if and only if $P_d\nabla\phi_d(e_d) = e_d$ for some $d > 0$.*

The Uniform Scaling Duality *holds if: $\mu \leq 0$ if and only if there exists $\gamma^* > 0$ such that for all $d > 0$, $\|P_d\nabla\psi_d(e_d)\| \geq \gamma^*$.*

The Scaling Separation Duality *holds if : $\mu > 0$ if and only if there exists $d > 0$ such that $P_d\nabla\phi_d(e_d) >_d 0$. In particular, if $K_d = K$ for all $d > 0$ and $e_d = e$, a fixed point of K°, then $\mu > 0$ if and only if there exists $d > 0$ such that $P_d\nabla\phi_d(e) > 0$.*

2.6. Previous applications of some scaling dualities. All the above scaling dualities hold for the case where $K = \Re^n_+$ and $F(x) = -\sum_{i=1}^n \ln x_i$. These dualities have given rise to various projective algorithms, see Kalantari [3]-[7]. Consider HP over this cone where $\phi(x) = \frac{1}{2}x^T Qx$, Q an $n \times n$ symmetric matrix, $W = \Re^n$, but where ϕ is not necessarily assumed to be convex. We shall refer to this HP as *Gordan's HP*, since it is closely related to Gordan's Theorem. On the one hand, when

Q is indefinite, Gordan's HP is NP-complete (see Kalantari [4]). On the other hand, linear programming can be formulated as Gordan's HP with a positive semidefinite Q (see Chvátal [1], Jin and Kalantari [2]). Jin and Kalantari [2] show that matrix scaling gives rise to a natural algorithm for computing an interior point of a linear system of strict inequalities, $Ax < b$.

Karmarkar's caninical linear programming problem, [14], is a very special case of the general HP defined in this article, where $K = \Re^n_+$ where $\phi(x) = c^T x$. Karmarkar's canonical LP can also be formulated as Gordan's HP with a positive semidefinite matrix Q, and conversely.

Most literature on interior-point algorithms, when referring to Karmarkar's canonical LP, automatically assumes that the corresponding μ is zero. This assumption is unnecessary and ignores the fundamental and intrinsic dualities that relate HP to SP, HSP, or ASP. Indeed the problem is exactly to test if $\mu = 0$. This assumption perhaps stems from the fact that in Karmarkar's first version of the projective algorithm, [13], the minimum objective value is assumed to be zero. However, in the later version, [14], the assumption is removed.

In the case of Gordan's HP the corresponding scaling equation reduces to $DQDe = e$, where $D = \text{diag}(d)$. The polynomial-time solvability of ASP for positive semidefinite matrices over the rationals was established in Khachiyan and Kalantari [15]. In fact via a single simple path-following algorithm, shown in [15], one can test if $\mu = 0$ or if the scaling equation $DQDe = e$ is solvable. In particular linear programming can be solved this way. For a rederivation of this algorithm which emphasizes the role of Gordan's Theorem, see Kalantari [12]. Indeed the approach in [12] is the spirit of the path-following algorithm in this article.

Extension of the algorithm of [15] for solving linear programming or the diagonal scaling equation over the algebraic numbers is given by Kalantari and Emamy-K [8]. When Q is an arbitrary symmetric not only Gordan's HP is NP-hard (see [4]), so is the corresponding ASP (see Khachiyan [16]). Kalantari [11], extends the scaling dualities and the path-following algorithm of [15] to the semidefinite programming problem.

2.7. Self-concordance theory and SP. In this section we review some results from self-concordance theory in the context of our conic problems. We also state two conceptually simple algorithms, a potential-reduction algorithm and a path-following algorithm. In subsequent sections we will examine the application of these algorithms in solving HP, SP, HSP, and ASP for β-compatible ϕ.

DEFINITION 2.23. *Assume ϕ is β-compatible. Given $d > 0$, let $P_2(y) = \psi(d) + \nabla\psi(d)^T y + \frac{1}{2} y^T \nabla^2 \psi(d) y$ (i.e. the quadratic approximation to ψ at d). Then, Newton direction is defined as*

$$y(d) = \text{argmin}\{P_2(y) : y \in W\},$$

Newton decrement as $\lambda(d) = (1+\beta)\sqrt{\Delta(d)}$, where $\Delta(d) = y(d)^T \nabla^2 \psi(d) y(d)$, and Newton iterate as $d' = NEW(\psi, d) = d + \sigma(\lambda(d)) y(d)$, where

$$\sigma(\lambda(d)) = \begin{cases} \frac{1}{1+\lambda(d)}, & \text{if } \lambda(d) > 1; \\ \frac{1-\lambda(d)}{\lambda(d)(3-\lambda(d))}, & \text{if } \lambda(d) \in [\lambda_*, 1), \quad \lambda_* = 2 - \sqrt{3} = 0.2679; \\ 1, & \text{if } \lambda(d) < \lambda_*. \end{cases}$$

Corresponding to f^t, $y_t(d)$, $\lambda_t(d)$, and $d'_t = NEW(f^t, d)$ are defined analogously.

The following theorem is an adaptation of three significant results of Nesterov and Nemirovskii in [17] combining them into a format suitable for our conic problem. It combines a theorem characterizing basic properties of Newton's method under self-concordance ([17], Theorem 2.2.2), a theorem on the main property of self-concordant families ([17], Theorem 3.1.1), and its consequence under β-compatibility ([17], Proposition 3.2.2). Let $\bar{\lambda}$ be a number in $[\lambda_*, 1]$.

THEOREM 2.24. (Nesterov and Nemirovskii's Theorem) *Let ϕ be β-compatible. Then, $d' = NEW(\psi, d) > 0$, $d'_t = NEW(f^t, d) > 0$, and*

$$\begin{cases} \psi(d') - \psi(d) \le -\frac{\lambda(d) - \ln(1+\lambda(d))}{(1+\beta)^2}, & \text{if } \lambda(d) > \bar{\lambda}; \\ \lambda(d') \le \frac{1}{4}(6\lambda(d) - \lambda^2(d) - 1), & \text{if } \lambda(d) \in [\lambda_*, \bar{\lambda}]; \\ \lambda(d') \le \frac{\lambda^2(d)}{(1-\lambda(d))^2}, & \text{if } \lambda(d) < \lambda_*. \end{cases} \tag{2.11}$$

If $\lambda(d) < \frac{1}{3}$, then $\psi^* > -\infty$ and

$$\psi(d) - \psi^* \le \frac{\omega^2(\lambda(d))(1+\omega(\lambda(d))}{2(1+\beta)^2(1-\omega(\lambda(d))}, \quad \omega(\lambda(d)) = 1 - (1 - 3\lambda(d))^{1/3}. \tag{2.12}$$

All the above applies to f^t. Moreover, given $t \in (0, \infty)$, suppose that

$$\lambda_t(d) \le \kappa < \lambda_*. \tag{2.13}$$

Then, for any t' satisfying

$$|\ln \frac{t'}{t}\, | \left(1 + (1+\beta)\frac{\sqrt{\theta}}{\kappa}\right) \le 1 - \frac{\kappa}{(1-\kappa)^2}, \tag{2.14}$$

we have

$$\lambda_{t'}(d) \le \kappa. \tag{2.15}$$

In particular, if we let $\kappa = 1/4$, the above holds for t' satisfying

$$t' = r_* t, \quad r_* = exp\left(\frac{-c}{(1+\beta)\sqrt{\theta}}\right), \quad c = \frac{5}{9}\left(\frac{4(1+\beta)\sqrt{\theta}}{1 + 4(1+\beta)\sqrt{\theta}}\right). \tag{2.16}$$

Thus, given any $t \in (0, 1]$, since $\lambda_1(x^0) = 0$ (see Definition 2.7 and Definition 2.23) and we can choose $c = 1/9$, the number of iterations, k_t, to obtain $d > 0$ such that $\lambda_t(d) < \lambda_*$, satisfies the inequality $r_*^{k_t} \le t$. Equivalently, it suffices to choose

$$k_t = \lceil 9(1+\beta)\sqrt{\theta} \ln\left(\frac{1}{t}\right)\rceil. \tag{2.17}$$

It also follows that the central-path $\Pi = \{d_t^* : t \in (0, \infty)\}$ is well-defined. □

We now state two conceptually simple algorithms and a theorem to be proved much later in the article, revealing the significance of the algorithms.

Potential-Reduction Algorithm:
Initialization. Let $d = x^0$.
Iterative Step. Replace d with $d' = NEW(\psi, d)$ and repeat.

Path-Following Algeorithm:
Initialization. Let $t = 1$, $d = x^0$, $t_* \in (0, 1)$.

Phase I. While $t > t_*$, replace (d, t) with (d'_t, t'), $d'_t = NEW(f^t, d)$, $t' = r_* t$.

Phase II. Replace d with $d' = NEW(f^t, d)$ and repeat.

THEOREM 2.25. *Consider any HP/SP/HSP/ASP over a pointed convex cone K where ϕ is of homogeneous degree p and β-compatible with the corresponding θ-normal barrier F. Assume that $T(K) = T_F = \{T(d) = D = \nabla^2 F(d)^{-1/2} : d \in K^\circ\}$. Given $\epsilon \in (0,1)$, the solvability of ϵ-HP ϵ-SP, ϵ-HSP, or ϵ-ASP can all be tested in polynomial-time via the Potential-Reduction algorithm and if $p > 1$ via the Path-Following algorithm. In particular, these algorithms apply to the case where ϕ is linear, or quadratic. To apply the Path-Following algorithm to a linear ϕ we simply need to square it.*

2.8. Summary of the results. In § 3, we prove the first scaling duality (see Definition 2.22), the Weak Scaling Duality (Theorem 3.2). This theorem also gives an inequality on homogeneous potentials, more general than the arithmetic-geometric mean, trace-determinant, and Hadamard inequalities. In fact we do not assume the convexity of ϕ in this theorem. In § 4, we prove the Conic Convex Programming Duality (Theorem 4.3), a new duality for convex programming. Using this duality we prove the Scaling Duality Theorem (Theorem 5.3). In particular, this gives a generalization of diagonal matrix scaling theorems. The Conic Convex Programming Duality also implies the existence of Π, the central-path of f^t, regardless of β-compatibility of ϕ. In § 5, we study Newton's method and its convergence analysis in the context of the scaling problem, SP. In this section we also prove the Uniform Scaling Duality (Theorem 5.3), given any operator-cone $T(K)$, or the specific operator-cone $T_F = \{D = \nabla^2 F(d)^{-1/2} : d \in K^\circ\}$, where ϕ is assumed to be β-compatible. In § 6, assuming that $\mu > 0$, we derive an upper bound on the norm of the scaling vector, d^*, and the norm of parameterized scaling vectors, d^*_t, all in terms of μ. Assuming $\mu > 0$, given $d \in S(\bar{\lambda}_*) = \{x > 0 : \lambda(x) < \bar{\lambda}_*\}$, where $\bar{\lambda}_* = \lambda_*(1+\beta)/(1+\beta+\rho)$, we bound the norm of d by a constant multiple of the norm of d^*. Given a point $d_t \in S_t(\bar{\lambda}_*) = \{x > 0 : \lambda_t(x) < \bar{\lambda}_*\}$, computed via Phase I of the Path-Following algorithm, we bound its norm as a function of t. We also obtain crucial bounds on the norm of the scaled gradient projection $P_d \nabla \psi_d(e_d)$ and $P_d \nabla f^t_d(e_d)$. We use these bounds in § 7 to prove a complexity theorem, the Potential-Reduction Complexity Theorem (Theorem 7.3). This theorem analyzes the application of the Potential-Reduction algorithm for solving HP and the three scaling problems, when ϕ is β-compatible. The theorem employs scaling dualities, the bounds, and Theorem 2.24 of Nesterov and Nemirovskii. Next in § 8, we prove the Path-Following Theorem (Theorem 8.3), for solving HP and the three scaling problems. We prove a complexity theorem, the Path-Following Complexity Theorem (Theorem 8.4), for solving β-compatible cases of the four problems, applicable when the homogeneous degree satisfies $p > 1$. In § 9, we consider applications of the two algorithms.

3. Proof of the Weak Scaling Duality and an inequality. In this section we do not assume ϕ to be convex. In the following lemma we establish a relationship between the stationary points of ψ and those of X.

LEMMA 3.1. *If $d \in K^\circ \cap W$ is a stationary point of ψ, then it is a stationary point of X, and $\phi(d) = \theta/p$. Conversely, if $d \in K^\circ \cap W$ is a stationary point of X, and $\phi(d) > 0$, then $\alpha_0 d$ is a stationary point of ψ, where $\alpha_0 = (\theta/p\phi(d))^{1/p}$.*

Proof. Differentiating $X(x)$, we have

$$\nabla X(x) = \left(\nabla \phi(x) + \frac{p}{\theta}\phi(x)\nabla F(x)\right) exp(\frac{p}{\theta}F(x)). \tag{3.1}$$

Since d is a stationary point of ψ, we have

$$P\nabla\psi(d) = P(\nabla\phi(d) + \nabla F(d)) = 0. \tag{3.2}$$

Multiplying the above by d^T, and applying properties (P1) and (P5) of Proposition 2.8 we get

$$\phi(d) = \frac{\theta}{p}. \tag{3.3}$$

Substituting (3.3) in (3.1), we conclude that d is a stationary point of X. Conversely, suppose that for some $d > 0$, $P\nabla X(d) = 0$, and $\phi(d) > 0$. Since X is homogeneous of degree 0, from property (P2) of Proposition 2.8, for all $\alpha > 0$, $P\nabla X(\alpha d) = 0$. In particular, if we choose α so that $\phi(\alpha d) = \theta/p$, for this choice of α, αd is a stationary point of ψ. But this coincides with α_0. □

THEOREM 3.2. (Weak Scaling Duality/Homogeneous Potential Inequality Theorem) *Either $\mu \leq 0$, or $P_d\nabla\phi_d(e_d) = e_d$, for some $d > 0$, where $D = \nabla^2F(d)^{1/2}$. If $p > 1$, then either $\mu \leq 0$, or $P_d\nabla^2\phi_d(e_d)P_de_d = e_d/(p-1)$, for some $d > 0$. If $\mu > 0$, and there exists a unique $d > 0$ satisfying the scaling equation, then for all $x > 0$,*

$$X(x) \geq X(d).$$

If $\mu > 0$, and ϕ is convex, then there exists a unique $d > 0$ satisfying the scaling equation. Moreover, for all $x > 0$,

$$\psi(x) \geq \psi(d), \quad X(x) \geq X(d).$$

Proof. Suppose $\mu > 0$. Consider the set $K^\circ \cap W \cap S$, where $S = \{x \in E : \|x\| = 1\}$, the unit sphere. Since $X(x)$ approaches infinity as x approaches a boundary point of $K^\circ \cap W \cap S$, its infimum is attained at some x^* in this intersection. We claim that x^* must be the minimum of $X(x)$ over $K^\circ \cap W$. Otherwise, there exists $\bar{x} \in K^\circ \cap W$ such that $X(\bar{x}) < X(x^*)$. But as $X(x)$ is homogeneous of degree zero, we get $X(\bar{x}/\|\bar{x}\|) < X(x^*)$, a contradiction. Also, we must have $\phi(x^*) > 0$. From Lemma 3.1, this implies that for some $\alpha > 0$, $d = \alpha x^*$ is a stationary point of ψ. In particular, d satisfies the scaling equation. Now suppose that $\mu > 0$, and $d > 0$ is a unique solution to the scaling equation. Then, it is also a unique stationary point of ψ. On the other hand, since $\mu > 0$, x^* exists. From Lemma 3.1, and uniqueness of d, x^* must be a scalar multiple of d. But since $X(x)$ is homogeneous of degree 0, d must be its global minimizer over $K^\circ \cap W$. If $\mu > 0$, and ϕ is convex, then from strict convexity of ψ, there exists a unique $d > 0$ satisfying the scaling equation, and it is necessarily the minimizer of ψ, as well as X. □

Remark. The theorem for ϕ a quadratic form over $K = \Re^n_+$, was established previously in Kalantari [4], and for general ϕ over this cone in [7]. The theorem also characterizes an optimality condition on the homogeneous potential function X, which is nonconvex. In particular, the theorem gives a generalization of the arithmetic-geometric mean, trace-determinant, and Hadamard inequalities (see Kalantari [9]). Note that it also implies the diagonal matrix scaling result, $DQDe = e$, for symmetric matrices of positive entries (positive matrices), or more generally, $P_dDQDP_de = e$.

4. The Conic Convex Programming Duality and its applications. In this section we prove a duality for conic convex programming (Theorem 4.3). The theorem will be applied to get three different results. Firstly, we apply this theorem to prove the Scaling Duality Theorem (Theorem 4.5). Secondly, we use it to prove the existence of Π, the central-path of the family $f^t(x)$, $t \in (0, \infty)$, independent of strong self-concordance (Theorem 4.8). The latter theorem will be used within an important theorem, called the Path-Following Theorem (Theorem 8.3). Thirdly, given that ϕ is β-compatible, we use the theorem to establish the convergence of Newton iterates, when computing the minimum of ψ, or the minimum of f^t for a fixed t (Theorem 4.9). The latter theorem will be invoked several times in subsequent sections (in particular, in § 6).

DEFINITION 4.1. *Let E, K, K°, and W be as defined in HP. A continuously differentiable strictly convex function $\bar{F}(x)$ is said to be a "recessive barrier" for K if the following conditions are satisfied:*

(1) *For each real number α, $K_\alpha(\bar{F}) = \{x \in K^\circ : \bar{F}(x) \leq \alpha\}$ is closed.*

(2) *For each $x \in K^\circ$, $v \in K$, $v \neq 0$, $q(\alpha) = \bar{F}(x + \alpha v)$ is decreasing in $(0, \infty)$.*

(3) $\lim_{\alpha \to \infty} q'(\alpha) = 0$.

Example. Let $K = \Re^n_+$. Then, $\bar{F}(x) = -\sum_{i=1}^n \ln x_i$, or $\bar{F}(x) = \sum_{i=1}^n 1/{x_i}^k$, k any natural number, are recessive barriers for K.

PROPOSITION 4.2. *Let $\bar{F}(x)$ be a recessive barrier for K. Let $g(x)$ be any continuously differentiable convex function defined over K°, and assume it has a continuous extension to K. Let $f(x) = g(x) + \bar{F}(x)$. The infimum of f over $W \cap K^\circ$ exists if and only if for each real α, $K_\alpha(f) = \{x \in W \cap K^\circ : f(x) \leq \alpha\}$ is compact. Equivalently, the infimum of f is not attained if and only if f has a recession direction $v \in W \cap K$, $v \neq 0$.*

Proof. The set $K_\alpha(f)$ is closed and convex. Thus, if unbounded, it has a recession direction. This implies the equivalence of the two statements of the theorem. Now suppose the infimum is attained at $x^* > 0$, but $K_\alpha(f)$ is not compact for some α, i.e., it has a recession direction, say v. In particular, v is a recession direction of f at x^*. But, since f is strictly convex, and x^* the minimizer, v cannot be a recession direction at x^*, a contradiction.

To prove the converse, suppose that the infimum of f over $W \cap K^\circ$ is not attained. This together with the fact that f approaches infinity as x approaches a nonnegative finite boundary point, implies that for each natural number k, if

$$\alpha_k = \inf\{f(x) : x > 0, \|x\| = k\}, \quad \beta_k = \inf\{f(x) : x > 0, \|x\| \leq k\},$$

then $\alpha_k = \beta_k$. Thus, the sequence of α_k's is nonincreasing, and $K_{\alpha_1}(f)$ is unbounded. This completes the proof. ∎

THEOREM 4.3. (Conic Convex Programming Duality) *Let $\bar{F}(x)$ be a recessive barrier for K. Then, either $g(x)$ has a recession direction in $W \cap K$, or the infimum of $f(x) = g(x) + \bar{F}(x)$ over $W \cap K^\circ$ is attained. Moreover, precisely one of the two conditions is satisfied. In particular, if g is homogeneous, then either there exists $x \geq 0$, $x \neq 0$ such that $g(x) \leq 0$, or the infimum of f over $W \cap K^\circ$ is attained.*

Proof. Without loss of generality we assume $W = E$. Suppose $g(x)$ has a recession direction $v \in K$, $v \neq 0$. Then, for any $x > 0$, $g(x + \alpha v)$ is nonincreasing on $(0, \infty)$. This implies $f(x + \alpha v)$ is decreasing on $(0, \infty)$, so that the infimum of f is not attained. This implies the two conditions of the theorem are exclusive. It remains to show that one of the two must be satisfied. Suppose g has no recession direction, but f does not attain its infimum. From Proposition 4.2, f has a recession direction $v \in K$, $v \neq 0$. Given a fixed $x > 0$, let $p(\alpha) = g(x + \alpha v)$, $\alpha \in [0, \infty)$. Since g has no recession direction, its convexity implies there exists some α_0 such that $p'(\alpha_0) = \epsilon > 0$. Again from convexity, for all $\alpha \geq \alpha_0$, we have

$$p(\alpha) \geq p(\alpha_0) + (\alpha - \alpha_0)p'(\alpha_0) \geq c_0 + \epsilon\alpha, \tag{4.1}$$

for some constant c_0. Thus,

$$f(x + \alpha v) \geq L(\alpha) = c_0 + \epsilon\alpha + \bar{F}(x + \alpha v). \tag{4.2}$$

Since $\bar{F}$ is a recessive barrier, $L'(\alpha) = \epsilon + q'(\alpha)$ approaches ϵ, as α approaches infinity. This implies $L(\alpha)$, and hence $f(x + \alpha v)$ must approach infinity, contradicting that v is a recession direction of f. ∎

We will now state three important applications of the theorem. Before doing so we state the following immediate consequence, already generalizing a theorem on matrix scaling.

COROLLARY 4.4. *Let Q be an $n \times n$ positive semidefinite symmetric matrix such that $x^T Qx$ has no nontrivial zero in $\Re^n_+$. Then, given any natural number k, there exists a positive diagonal matrix D such that*

$$D^k QDe = e.$$

Proof. Let $g(x) = \frac{1}{2}x^T Qx$, and $\bar{F}(x) = -\sum_{i=1}^n \ln x_i$, if $k = 1$; and $\bar{F}(x) = \sum_{i=1}^n 1/{x_i}^{k-1}$, if $k > 1$. Now apply Theorem 4.3. ∎

4.1. Proof of the Scaling Duality Theorem. The goal of this section is to prove the Scaling Duality Theorem:

THEOREM 4.5. (Scaling Duality Theorem) *Assume that ϕ is convex, and F is a θ-normal barrier. Then, either HP is solvable or SP is solvable, but not both. Equivalently, $\mu > 0$ if and only if $P_d \nabla \phi_d(e_d) = e_d$ for some $d > 0$, where $D = \nabla^2 F(d)^{-1/2}$ (if $p > 1$, the scaling equation is equivalent to $P_d \nabla^2 \phi_d(e) P_d e = e/(p-1)$). Moreover, for all $x > 0$,*

$$\psi(x) \geq \psi(d), \quad X(x) \geq X(d),$$

i.e., $d = d^*$, the minimizer of ψ, as well as the minimizer of X.

The proof of the Scaling Duality Theorem is an immediate consequence of Theorem 4.3, the Weak Scaling Duality (Theorem 3.2), and the following result.

LEMMA 4.6. *If $F(x)$ is a θ-normal barrier for K, then it is a recessive barrier.*

Proof. We have to verify that it satisfies the three conditions of Definition 4.1. Firstly, the set $K_\alpha(F) = \{x \in K^\circ : F(x) \leq \alpha\}$ is closed. To prove the second condition, from Corollary 2.3.1 in Nesterov and Nemirovskii [17] (page 39), we have

$$-\nabla^2 F(x)[h, h]^{1/2} \geq \nabla F(x)[h]. \tag{4.3}$$

If $q(\alpha) = F(x + \alpha v)$, then $q'(\alpha) = \nabla F(x + \alpha v)[v]$. Since F is strictly convex, $\nabla^2 F(x)[h, h] > 0$. Thus, $q'(\alpha) < 0$. To prove the third condition, from Proposition 2.3.4 of Nesterov and Nemirovskii [17] (page 34), for any $x, y \in K^\circ$, we have

$$\nabla F(x)[y - x] \leq 1. \tag{4.4}$$

In particular,

$$q'(\alpha) = \nabla F(x + \alpha v)[v] = \frac{1}{\alpha} \nabla F(x + \alpha v)[\alpha v] \geq \frac{-1}{\alpha}. \tag{4.5}$$

Thus, $-1/\alpha \leq q'(\alpha) < 0$. Taking the limit as α approaches infinity, we see that F satisfies the third condition of Definition 4.1. ∎

We now state a theorem that is a consequence of the Scaling Duality Theorem and Proposition 2.21 on copositive plusness. This theorem is the generalization of a matrix scaling duality for positive semidefinite symmetric matrices.

THEOREM 4.7. (Generalization of Positive Semidefinite Matrix Scaling Duality) *Suppose in a given HP the homogeneous function ϕ is defined over W, it is convex, $p > 1$, F is a θ-normal barrier, and $T(K)$ a given operator-cone. Then, either $P\nabla\phi(x) = 0$ for some $x \geq 0$, $x \neq 0$, or $P_d\nabla\phi_d(e_d) = e_d$ for some $d > 0$. Moreover, precisely one of the two conditions hold.* ∎

4.2. Existence of central-path. Let $\bar{F}(x)$, and $g(x)$ be as in Theorem 4.3. Let x^0 be in $W \cap K^\circ$ (a given positive point). For each $t \in (0, \infty)$, consider the family

$$\bar{f}^t(x) = tg(x) + t\langle \bar{u}, x\rangle + \bar{F}(x), \quad \bar{u} = -P\nabla g(x^0).$$

THEOREM 4.8. *For each $t \in (0, \infty)$, the (unique) minimum, $\bar{d}_t^*$ of $\bar{f}^t(x)$ over $W \cap K^\circ$ exists. In particular, given an operator-cone $T(K)$, if $\bar{f}^t_{\bar{d}_t^*}(x) \equiv \bar{f}^t(\bar{D}_t^* x)$, then*

$$P_{\bar{d}_t^*}\nabla \bar{f}^t_{\bar{d}_t^*}(e_{\bar{d}_t^*}) = 0.$$

Proof. From Theorem 4.3, $g(x) + \langle \bar{u}, x\rangle$ has no recession direction over K. This implies $t(g(x) + \langle \bar{u}, x\rangle)$ has no recession direction. Again by Theorem 4.3, $\bar{f}^t(x)$ must have a stationary point. ∎

Theorem 4.8 will be used to prove the Path-Following Theorem (Theorem 8.3). Theorem 4.8 can also be used to prove that if $g(x)$ has no recession direction, then

$$\lim_{t\to\infty} \bar{d}_t^* = \mathrm{argmin}\{g(x) : x \in W \cap K\}.$$

Thus, when minimizing a convex function over a cone we have a lot more information available to us than a convex program over a general convex set.

4.3. Convergence of Newton iterates.

THEOREM 4.9. *Assume ϕ is β-compatible. For a given $\alpha \in [0, \frac{1}{3}]$, the sets $S(\alpha) = \{d > 0 : \lambda(d) < \alpha\}$ and $S_t(\alpha) = \{d > 0, \lambda_t(d) < \alpha\}$, $t \in (0, 1]$, are bounded. In particular, suppose there exists a sequence $\{d^k > 0\}_{k=0}^\infty$ such that*

$$\lim_{k\to\infty} \lambda(d^k) = 0, \quad \lambda(d^0) < \lambda_*,$$

then all the iterates stay within $S(\lambda_*)$, and

$$\lim_{k\to\infty} d^k = d^* = \mathrm{argmin}\{\psi(x) : x > 0\}.$$

Also, given $t \in (0, 1]$, if there exists a sequence $\{d_t^k > 0\}_{k=0}^\infty$ such that

$$\lim_{k\to\infty} \lambda_t(d_t^k) = 0, \quad \lambda_t(d_t^0) < \lambda_*,$$

then $d_t^k \in S_t(\lambda_*)$, for all k, and

$$\lim_{k\to\infty} d_t^k = d_t^* = \mathrm{argmin}\{f^t(x) : x > 0\}.$$

Proof. From Proposition 4.2, the minimizer d^* of ψ exists if and only if for each $\alpha \in \Re$, the level set $K_\alpha(\psi) = \{x \in W \cap K^\circ : \psi(x) \leq \alpha\}$ is compact. From this and Theorem 2.24, see (2.12), it follows that if $S(\alpha)$ is nonempty, it must be bounded. Moreover, any accumulation of $\{d^k\}_{k=0}^{\infty}$ is a minimizer of ψ. But, by strict convexity of ψ, its minimizer is unique. Hence, this is a convergent sequence. Also, as shown in Theorem 4.8, d_t^* exists. In particular, from Proposition 4.2, $K_\alpha(f^t) = \{x \in W \cap K^\circ : \psi(x) \leq \alpha\}$ is compact. Now using an analogous argument as those used for ψ, the convergence of d_t^k to d_t^* can be established. □

5. Newton's method and proof of the Uniform Scaling Duality. In this section we consider the application of Newton's method in solving SP. The iterative step of Newton's method consists of the minimization of the quadratic approximation of ψ at a given $d > 0$. Given that ϕ is β-compatible, we first obtain upper and lower bounds on the Newton decrements (see Definition 2.23), namely $\lambda(d)$ and $\lambda_t(d)$ (Lemma 5.2). The first upper bound will be used to prove the Uniform Scaling Duality under the assumption of β-compatibility (Theorem 5.3). The lower bounds will be used to derive bounds on the norm of points within the quadratic regions of convergence (Lemma 6.3).

Assume that ψ is twice continuously differentiable over $W \cap K^\circ$. The quadratic approximation at d is given by

$$\Phi_d(x) = \psi(d) + \nabla\psi(d)[x-d] + \frac{1}{2}\nabla^2\psi(d)[x-d, x-d]. \tag{5.1}$$

Let $\bar{x}$ be the minimizer of $\Phi_d(x)$ over W. Then, the *Newton decrement* $\Delta(d)$, the *Newton direction* y, and the *Newton iterate* are defined, respectively as

$$\frac{\Delta(d)}{2} = \psi(d) - \Phi_d(\bar{x}), \quad y = \bar{x} - d, \quad \bar{x} = d + y. \tag{5.2}$$

From the optimality condition as applied to $\bar{x}$, the Newton direction and the Newton decrement must satisfy

$$P\nabla^2\psi(d)y = -P\nabla\psi(d), \quad \Delta(d) = y^T\nabla^2\psi(d)y, \tag{5.3}$$

where as before P is the orthogonal projection onto $W = \{x : Ax = 0\}$. Since $Py = y$, $P\nabla^2\psi(d)y = P\nabla^2\psi(d)Py$. Let $Q = \nabla^2\psi(d)$, and $c = \nabla\psi(d)$. It is easy to show that y satisfies

$$PQy = PQPy = -Pc, \quad y = -Q^{-1}c + Q^{-1}A^T(AQ^{-1}A^T)^{-1}AQ^{-1}c. \tag{5.4}$$

It is worth noticing that if for a given matrix M we define $P_M \equiv I - M^T(MM^T)^{-1}M$, then by substituting $Q = LL^T$, the Cholesky factorization, or by writing $Q = (Q^{1/2})^2$, the square of its square root, we get the following alternative formulas describing y

$$y = -Q^{-1/2}P_{AQ^{-1/2}}Q^{-1/2}c = -L^{-T}P_{AL^{-T}}L^{-1}c. \tag{5.5}$$

In particular, the Newton decrement satisfies

$$\Delta(d) = y^TQy = \|P_{AQ^{-1/2}}Q^{-1/2}c\|^2 = \|P_{AL^{-T}}L^{-1}c\|^2 \leq \|Q^{-1/2}c\|^2. \tag{5.6}$$

Using the spectral decomposition, the following can trivially be proved.

PROPOSITION 5.1. *Let H be a self-adjoint (symmetric) operator on a finite dimensional Hilbert space E. If for all $x \in E$, $\langle Hx, x \rangle \geq \|x\|^2$, then $\|Hx\|^2 \geq \langle Hx, x \rangle$.* □

LEMMA 5.2. *Assume ϕ is β-compatible, and $T(K)$ a given operator-cone. Then, for all $d \in S(\lambda_*) = \{x > 0 : \lambda(x) < \lambda_*\}$, we have*

$$\|z\| \leq \frac{1}{(1+\beta)} \lambda(d) \leq \|P_d \nabla \psi_d(e_d)\|,$$

where $z = D^{-1}y$, and y is the Newton direction with respect to ψ at d. Also, given any $t \in (0, 1]$, for all $d \in S_t(\lambda_*) = \{x : x > 0, \lambda_t(x) < \lambda_*\}$, we have

$$\|z_t\| \leq \frac{1}{(1+\beta)} \lambda_t(d) \leq \|P_d \nabla f^t_d(e_d)\|,$$

where $z_t = D^{-1}y_t$, and y_t is the Newton direction with respect to f^t at d.

Proof. We will prove the desired inequality for ψ. Analogous proof follows for f^t. Note that y must satisfy $\nabla\psi(d) + \nabla^2\psi(d)y = A^T v$, for some vector of Lagrange multipliers, v. From this, chain rule, and the definition of z it is easy to show that

$$P_d \nabla^2 \psi_d(e_d) z = -P_d \nabla \psi_d(e_d). \tag{5.7}$$

Thus,

$$\|P_d \nabla \psi_d(e_d)\| = \|P_d \nabla^2 \psi_d(e_d) P_d z\| = \|Hz\|, \tag{5.8}$$

where

$$H = P_d \nabla^2 \psi_d(e_d) P_d = P_d \nabla^2 \phi_d(e_d) P_d + P_d \nabla^2 F_d(e_d) P_d. \tag{5.9}$$

Since $\nabla^2 \phi_d(e_d)$ is positive semidefinite and $Pz = z$, we have

$$z^T H z \geq z^T \nabla^2 F_d(e_d) z. \tag{5.10}$$

Now from (5.10) and property (4) of operator-cones (see Definition 2.13), together with Proposition 5.1, we get

$$\|z\|^2 \leq z^T H z \leq \|Hz\|^2. \tag{5.11}$$

We have

$$\Delta(d) = y^T \nabla \psi(d) y = z^T \nabla \psi_d(e_d) z = z^T H z. \tag{5.12}$$

The proof now follows from (5.11), (5.12), and since $\lambda(d) = (1+\beta)\sqrt{\Delta(d)}$ (see Definition 2.23). □

THEOREM 5.3. (Uniform Scaling Duality Theorem) *Assume ϕ is β-compatible, and $T(K)$ a given operator-cone. Then, $\mu \leq 0$ if and only if*

$$\forall\ d > 0, \quad \|P_d \nabla \psi_d(e_d)\| \geq \gamma^* = \frac{1}{(1+\beta)}. \tag{5.13}$$

Proof. From the Scaling Duality Theorem (Theorem 4.5), $\mu > 0$ if and only if the scaling equation is solvable. Thus, if $\mu > 0$, (5.13) is violated for some $d > 0$. Conversely, suppose that (5.13) is violated for some $d > 0$. Then, from Lemma 5.2, $\lambda(d) < 1$. From Theorem 2.24, it follows that $S(\lambda_*) = \{x > 0 : \lambda(x) < \lambda_*\}$ is nonempty. Then, from Theorem 4.9, starting from any point in this set, Newton's iterates converge to a point d satisfying the scaling equation. Hence, $\mu > 0$. □

6. Bounds. In this section we derive several significant bounds. First, we bound the norm of the minimizer d^* of ψ, if it exits, in terms of the homogeneous degree p, θ, and μ. Next, assuming that ϕ is β-compatible, $\mu > 0$, and given a bounded operator-cone, we bound the norm of any d in $S(\bar{\lambda}_*) = \{x > 0 : \lambda(x) < \bar{\lambda}_*\}$, where $\bar{\lambda}_* = \lambda_*(1+\beta)/(1+\beta+\rho)$, as a function of p, β, θ, μ, and ρ. Moreover, in this case we also show that the homogeneous central-path when restricted to $t \in (0,1]$, i.e., $\widehat{\Pi} = \{\hat{d}_t^* = t^{1/p} d_t^* : t \in (0,1]\}$, is a bounded set. Whether or not $\mu > 0$, assuming β-compatibility of ϕ, and given a bounded operator-cone, if for a given $t \in (0,1]$, $d \in S_t(\bar{\lambda}_*) = \{x > 0 : \lambda_t(x) < \bar{\lambda}_*\}$ is obtained via Phase I of the Path-Following algorithm (see § 2.7), then we bound $\|d\|$ in terms of t, θ, β, and ρ. These bounds together with results from the previous sections imply bounds on the norm of the corresponding scaled gradient projections. In subsequent sections the latter bounds will be used to give polynomial-time algorithms for HP/SP/HSP/ASP. We mention here that the derived bounds on the norm of $d \in S(\bar{\lambda}_*)$, or $d \in S_t(\bar{\lambda}_*)$ are important bounds that may be extendible to non-conic self-concordant programming problems.

6.1. Bounds on the norm of scaling vector.

THEOREM 6.1. *Suppose $\mu > 0$. Then,*

$$\|d^*\| \le \left(\frac{\theta}{p\mu}\right)^{1/p}.$$

Proof. The point d^* satisfies

$$P\nabla\psi(d^*) = P\nabla\phi(d^*) + P\nabla F(d^*) = 0. \tag{6.1}$$

Taking the inner product of (6.1) with d^*, and using properties (P1) and (P5) of Proposition 2.8, we get $p\phi(d^*) = \theta$. Dividing the latter by $\|d^*\|^p$, using homogeneity of ϕ, and the definition of μ, it follows that

$$p\mu \le p\phi\left(\frac{d^*}{\|d^*\|}\right) = \frac{\theta}{\|d^*\|^p}. \square \tag{6.2}$$

The following result reveals a boundedness property of the homogeneous central-path. Recall the definition of d_t^* and $\hat{d}_t^* = t^{1/p} d_t^*$.

THEOREM 6.2. *Assume that ϕ is convex, $\mu > 0$, and $p > 1$. Then, for all $t \in (0,1]$, we have*

$$\|\hat{d}_t^*\| \le \max\left\{\theta, \left[\frac{1+\|u\|}{p\mu}\right]^{\frac{1}{p-1}}\right\}.$$

In particular, $\|\hat{d}_t^*\|$ is bounded, independent of t.

Proof. We have

$$P\nabla f^t(d_t^*) = tP\nabla\phi(d_t^*) + tPu + P\nabla F(d_t^*) = 0. \tag{6.3}$$

Taking the inner product of the above with d_t^*, and using properties (P1) and (P5) of Proposition 2.8, we get

$$pt\phi(d_t^*) + tu^T d_t^* - \theta = 0. \tag{6.4}$$

From the definition of μ and the Cauchy-Schwarz inequality, we obtain the inequalities

$$\phi\left(\frac{d_t^*}{\|d_t^*\|}\right) = \frac{1}{\|d_t^*\|^p}\phi(d_t^*) \ge \mu, \quad tu^T d_t^* \ge -t\|u\|\,\|d_t^*\|. \tag{6.5}$$

Using the inequalities of (6.5) in (6.4) we get

$$tp\mu\|d_t^*\|^p - t\|u\|\ \|d_t^*\| - \theta \le 0. \tag{6.6}$$

Since $t \in (0,1]$, implies $t \le t^{1/p}$, the above inequality gives

$$p\mu\|t^{1/p}d_t^*\|^p - \|u\|\ \|t^{1/p}d_t^*\| - \theta \le 0. \tag{6.7}$$

The function

$$w(r) = p\mu r^p - \|u\|r - \theta \tag{6.8}$$

is convex. Thus, its positive roots are bounded by any point $\bar{r}$ for which $w(\bar{r}) \ge 0$, and $w'(\bar{r}) \ge 0$. It is easy to check that $\bar{r} = \max\{\theta, ((1+\|u\|)/p\mu)^{1/p-1}\}$ satisfies these conditions. □

Remark. In § 8 (Corollary 8.2), we show that if $\mu \le 0$, then both $\|d_t^*\|$ and $\|\hat{d}_t^*\|$ approach infinity as t goes to zero. Simple example can be constructed for which $\|d_t^*\|$ diverges, even if d^* exists.

6.2. Bounds on the region of quadratic convergence and the central-path.

LEMMA 6.3. *Assume that ϕ is β-compatible, and $T(K)$ a bounded operator-cone. Let*

$$\bar{\lambda}_* = \frac{\lambda_*(1+\beta)}{(1+\beta+\rho)}, \quad \bar{\alpha} = \frac{\rho}{(1+\beta)}.$$

If $d \in S(\bar{\lambda}_*) = \{x > 0 : \lambda(x) < \bar{\lambda}_*\}$, then $d' = NEW(\psi, d) = d + y(d)$ satisfies

$$\|d\| \le \frac{\|d'\|}{1 - \bar{\alpha}\lambda(d)}.$$

Also, if $d \in S_t(\bar{\lambda}_*) = \{x > 0 : \lambda_t(x) < \bar{\lambda}_*\}$, then $d'_t = NEW(f^t, d)$ satisfies

$$\|d\| \le \frac{\|d'_t\|}{1 - \bar{\alpha}\lambda_t(d)}.$$

Proof. We will prove the inequality for $d' = NEW(\psi, d)$ since the other follows analogously. We have $d' = d + y(d) = D(e_d + z) = d + Dz$. Thus, $d' - d = Dz$. Using this together with the triangle inequality, the bound on $\|z\|$ from Lemma 5.2, Cauchy-Schwarz inequality, and the fact that $\|D\| \le \rho\|d\|$, we get

$$\|d\| - \|d'\| \le \|d' - d\| \le \|Dz\| \le \|D\|\ \|z\| \le \rho\|d\|\frac{\lambda(d)}{(1+\beta)}. \tag{6.9}$$

Rearranging the above, we get

$$\|d\|[1 - \frac{\rho}{1+\beta}\lambda(d)] \le \|d'\|. \tag{6.10}$$

We have

$$(1 - \frac{\rho}{1+\beta}\lambda(d)) = (1 - \bar{\alpha}\lambda(d)) \ge (1 - \bar{\alpha}\bar{\lambda}_*) \ge (1 - \lambda_*) > 0. \tag{6.11}$$

The desired inequality now follows from (6.10) and (6.11). □

THEOREM 6.4. *Assume ϕ is β-compatible, and $T(K)$ a bounded operator-cone. Suppose $\mu > 0$. Let*

$$r = \frac{\bar{\lambda}_*}{(1-\lambda_*)^2}, \quad \bar{\alpha} = \frac{\rho}{(1+\beta)}.$$

If $\bar{\alpha} \le 1$, let h be the constant defined as

$$h = \frac{1}{(1-\lambda_*)(1-2r^2)},$$

and if $\bar{\alpha} > 1$, let h be the constant defined as

$$h = \frac{2}{(1-\lambda_*)^{i-1}}, \quad i = \max\{\lceil \log_2 \big(\frac{\log_2 2\bar{\alpha}}{\log_2(1/r)}\big)\rceil, \quad 1\}.$$

Then, if $d \in S(\bar{\lambda}_*) = \{x > 0 : \lambda(x) < \bar{\lambda}_*\}$, we have

$$\|d\| \le h\|d^*\|,$$

and if $d \in S_t(\bar{\lambda}_*) = \{x > 0 : \lambda_t(x) < \bar{\lambda}_*\}$, we have

$$\|d\| \le h\|d_t^*\|.$$

In particular, $\|t^{1/p}d\| \le h\|\hat{d}_t^*\|$.

Proof. We only prove the desired result for ψ. Analogous proof holds for f^t. Let $d \in S(\bar{\lambda}_*)$. Let $d^0 = d$. For each natural number k, let $d^{k+1} = NEW(\psi, d^k)$. Letting $\lambda_k = \lambda(d^k)$, repeated application of the first inequality in Lemma 6.3 as applied to d^k gives

$$\|d\| \le \frac{\|d^{k+1}\|}{\prod_{j=0}^{k}(1-\bar{\alpha}\lambda_j)}. \tag{6.12}$$

From Nesterov and Nemirovskii's Theorem (Theorem 2.24), we have $\lambda_{k+1} < \lambda_k^2/(1-\lambda_*)^2$. This and a simple induction gives

$$\lambda_j \le \bar{\lambda}_* r^{2^j-1}, \quad j = 0, \ldots, k. \tag{6.13}$$

Since $\bar{\lambda}_* \le r$, we have

$$\lambda_j \le r^{2^j}. \tag{6.14}$$

From (6.11), $(1-\bar{\alpha}\lambda_0) \ge (1-\lambda_*)$. Suppose $\bar{\alpha} \le 1$. From this and (6.14) it follows that $(1-\bar{\alpha}\lambda_j) \ge (1-r^{2^j})$. Thus,

$$\prod_{j=0}^{k}(1-\bar{\alpha}\lambda_j) \ge (1-\lambda_*)\prod_{j=1}^{k}(1-r^{2^j}) \tag{6.15}$$

We note that by expanding $\prod_{j=1}^{k}(1-r^{2^j})$ and replacing all positive signs with negative ones we get the following inequality

$$\prod_{j=1}^{k}(1-r^{2^j}) \ge (1-r^2\sum_{i=0}^{\infty}r^{2i}) = (1-\frac{r^2}{1-r^2}).$$

Thus

$$\prod_{j=0}^{k}(1-\bar{\alpha}\lambda_j) \geq (1-\lambda_*)\frac{1-2r^2}{1-r^2} \geq (1-\lambda_*)(1-2r^2) > 0, \tag{6.16}$$

where the positivity follows from the fact that $r < \lambda_*/(1-\lambda_*)^2$ and that $\lambda_* = 2-\sqrt{3}$.

Substituting (6.16) in (6.12), it follows that $\|d\| \leq h(\beta,\rho)\|d^k\|$, for all k. Since from Theorem 4.9, d^k converges to d^*, the proof of the desired inequality follows for $\bar{\alpha} \leq 1$.

Now suppose that $\bar{\alpha} > 1$. Since from (6.11) for all j we have $(1-\bar{\alpha}\lambda_j) \geq (1-\lambda_*)$, then for any $0 \leq i \leq k$ we have

$$\prod_{j=0}^{k}(1-\bar{\alpha}\lambda_j) \geq (1-\lambda_*)^{i-1}\prod_{j=i}^{k}(1-\bar{\alpha}\lambda_j) \geq (1-\lambda_*)^{i-1}\prod_{j=i}^{k}(1-\bar{\alpha}r^{2^j}) \tag{6.17}$$

Note that by expanding $\prod_{j=i}^{k}(1-\bar{\alpha}r^{2^j})$ and by replacing all positive signs with negative ones, and by increasing exponents of $\bar{\alpha}$ in each term of this expansion, and using that $\bar{\alpha} > 1$, we get the following inequality

$$\prod_{j=i}^{k}(1-\bar{\alpha}r^{2^j}) \geq 1-\sum_{j=1}^{\infty}(\bar{\alpha}r^{2^i})^{2j} = 1-\frac{(\bar{\alpha}r^{2^i})^2}{1-(\bar{\alpha}r^{2^i})^2} = \frac{1-2(\bar{\alpha}r^{2^i})^2}{1-(\bar{\alpha}r^{2^i})^2}.$$

Thus we have

$$\prod_{j=0}^{k}(1-\bar{\alpha}\lambda_j) \geq (1-\lambda_*)^{i-1}\frac{1-2(\bar{\alpha}r^{2^i})^2}{1-(\bar{\alpha}r^{2^i})^2} \tag{6.18}$$

To make the right-hand-side of the above positive it suffices to have $1-2(\bar{\alpha}r^{2^i})^2 > 0$. It is enough to have $\bar{\alpha}r^{2^i} < 1/2$. If $\bar{\alpha}r < 1$ then it suffices to choose $i = 1$. Otherwise, we set

$$i = \lceil \log_2 \frac{\log_2 2\bar{\alpha}}{\log_2(1/r)} \rceil.$$

Note that with the bound on $\bar{\alpha}r^{2^i}$, $1-2(\bar{\alpha}r^2)^i > 1/2$. Thus we get

$$\prod_{j=0}^{k}(1-\bar{\alpha}\lambda_j) \geq \frac{1}{2}(1-\lambda_*)^{i-1}.$$

This proves the bound for $\bar{\alpha} > 1$. □

THEOREM 6.5. *Assume ϕ is β-compatible, and $T(K)$ a bounded operator-cone. For a given $t \in (0,1]$, let $d_t \in S_t(\bar{\lambda}_*)$ be a point obtained as the result of applying k_t iterations of Phase I of the Path-Following algorithm (see § 2.7) starting at $x^0 > 0$, $\|x^0\| = 1$. Then,*

$$\|d_t\| \leq \left(1+\frac{\rho}{1+\beta}\right)\left(\frac{1}{t}\right)^{9\rho\sqrt{\theta}}.$$

Proof. From Nesterov and Nemirovskii's Theorem (Theorem 2.24), the number of iteration k_t to obtain the point d_t satisfies $k_t \leq [9(1+\beta)\sqrt{\theta}\ln(1/t)+1]$. Let $\{d^k\}_{k=0}^{k_t}$ be the sequence of points generated via Phase I of the Path-Following algorithm, where $d^0 = x^0$, $d^{k_t} = d_t$. Since $d^{k+1} = NEW(f^{(t_k)}, d^k)$, for some appropriate t_k, we have

$$d^{k+1} = d^k + y_{t_k}(d^k) = d^k + D^k z_{t_k}, \tag{6.19}$$

where D^k is the operator corresponding to d^k, and $y_{t_k}(d^k)$ is the Newton direction corresponding to d^k. Thus, using the boundedness of the operator-cone, the bound on $\|z_{t_k}\|$ given in Lemma 5.2, and (6.19), we get

$$\|d^{k+1}\| \leq \|d^k\| + \|D^k z_{t_k}\| \leq \|d^k\| + \rho\|d^k\|\ \|z_{t_k}\| \leq \|d^k\|(1+\frac{\rho}{1+\beta}). \tag{6.20}$$

From the repeated application of (6.20), and since $\|x^0\| = 1$, we get

$$\|d^k\| \leq (1+\frac{\rho}{1+\beta})^k. \tag{6.21}$$

By setting $k = k_t \leq [9(1+\beta)\sqrt{\theta}\ln(1/t)+1]$ in (6.21), and $G = (1+\rho/(1+\beta))^k$, we have

$$\ln G \leq \left(\ln(\frac{1}{t})^{9(1+\beta)\sqrt{\theta}} + 1\right)\ln(1+\frac{\rho}{1+\beta}). \tag{6.22}$$

From the above we get

$$G \leq (1+\frac{\rho}{1+\beta})(\frac{1}{t})^{9(1+\beta)\sqrt{\theta}\ln(1+\frac{\rho}{1+\beta})}. \tag{6.23}$$

Since given $\delta > 0$, $1+\delta \leq e^{\delta}$, we have $\ln(1+\delta) \leq \delta$. Letting $\delta = \rho/(1+\beta)$, we have

$$(1+\beta)\sqrt{\theta}\ln(1+\frac{\rho}{1+\beta}) = \rho\sqrt{\theta}\frac{(1+\beta)}{\rho}\ln(1+\frac{\rho}{1+\beta}) \leq \rho\sqrt{\theta}. \tag{6.24}$$

From (6.24), (6.21), the bound on G in (6.23), and the fact that $t \in (0,1]$, the proof is immediate. □

Remark. It is interesting to note that if we replace $F(x)$ by $\alpha F(x)$, where α is a positive scalar, the new barrier will be $(\alpha\theta)$-logarithmically homogeneous, but the new ρ will be replaced by $\rho/\sqrt{\alpha}$. This implies that the product $\rho\sqrt{\theta}$ is invariant under scalar multiplication.

6.3. Bound on the norm of scaled gradient projections. The following can be proved (see [8], Lemma 3):

PROPOSITION 6.6. *Let H be a self-adjoint (symmetric) operator on a finite dimensional Hilbert space E. For any $x \in E$*

$$\|Hx\| \leq \|H^{1/2}\|\ \|H^{1/2}x\| \leq \max\{1, \|H\|\}\sqrt{x^T Hx}. \square$$

The following is an important implication of the above proposition.

LEMMA 6.7. *Assume that ϕ is convex. Given $d > 0$, let $\Delta(d) = y^T\nabla^2\psi(d)y$, where y satisfies $P\nabla^2\psi(d)y = -P\nabla\psi(d)$. Then,*

$$\|P\nabla\psi(d)\| \leq \max\{1, \|\nabla^2\psi(d)\|\}\sqrt{\Delta(d)}.$$

Proof. Let $H = P\nabla^2\psi(d)P$. Since $P\nabla^2\psi(d)Py = -P\nabla\psi(d)$, and $\|H\| \leq \|P\|\ \|\nabla^2\psi(d)\|\ \|P\| \leq \|\nabla^2\psi(d)\|$, Proposition 6.6 implies the desired result. □

The significance of the above bound becomes more apparent when we consider scaled gradient projections.

LEMMA 6.8. *Assume ϕ is β-compatible, and $T(K)$ a bounded operator-cone. Let M be the bound as in (4) of definition of operator-cone (Definition 2.13). Let $q = \sup\{\|\nabla^2\phi(d)\| : d > 0, \|d\| = 1\}$. Then, $d \in S(\lambda_*)$ implies*

$$\|P_d\nabla\psi_d(e_d)\| \leq \frac{1}{(1+\beta)}\lambda(d)\Big(q\rho^2\|d\|^p + M\Big).$$

Moreover, given any $t \in (0,1]$, if $d \in S_t(\lambda_*)$ then

$$\|P_d\nabla f_d^t(e_d)\| \leq \frac{1}{(1+\beta)}\lambda_t(d)\Big(tq\rho^2\|d\|^p + M\Big).$$

Proof. We prove the inequality for ψ_d. Analogous proof follows for f_d^t. Let $H = P_d\nabla^2\psi_d(e_d)P_d = P_d\nabla^2\phi_d(e_d)P_d + P_d\nabla^2F_d(e_d)P_d$. Let z be the solution to $Hz = -P_d\nabla\psi_d(e_d)$ (the scaled Newton direction). From Proposition 6.6, $\|Hz\| \leq \|H\|\sqrt{z^THz}$. On the one hand, $\lambda(d) = (1+\beta)\sqrt{z^THz}$. On the other hand, we can bound $\|H\|$. Firstly, recalling the definition of an operator-cone (Definition 2.13) we have $\|P_d\nabla^2F_d(e_d)P_d\| \leq \|P_d\|\ \|\nabla^2F_d(e_d)\|\ \|P_d\| \leq \|\nabla^2F_d(e_d)\| \leq M$. Also

$$\|P_d\nabla^2\phi_d(e_d)P_d\| = \|P_dD^T\nabla^2\phi(d)DP_d\| \leq \tag{6.25}$$

$$\|D^T\|\ \|\nabla^2\phi(d)\|\ \|D\| \leq \rho^2\|d\|^2\|\nabla^2\phi(d)\|.$$

But, from property (P3) of Proposition 2.8 and definition of q, we have

$$\|\nabla^2\phi(d)\| = \|\nabla^2\phi(\|d\|\frac{d}{\|d\|})\| = \|d\|^{p-2}\|\nabla^2\phi(\frac{d}{\|d\|})\| \leq q\|d\|^{p-2}. \tag{6.26}$$

Substituting (6.26) in (6.25), we get $\|H\| \leq (q\rho^2\|d\|^p + M)$. Hence the proof. □

We now state the following important theorem. It is a combination of Lemma 5.2 and Lemma 6.8, but replacing the norm of d in the upper bounds by its corresponding upper bounds. This theorem will be used to prove the two complexity theorems on the Potential-Reduction and Path-Following algorithms.

THEOREM 6.9. *Assume ϕ is β-compatible, and $T(K)$ a bounded operator-cone. Let $\bar{\alpha} = \rho/(1+\beta)$, $\bar{\lambda}_* = \lambda_*/(1+\bar{\alpha})$, and $q = \sup\{\|\nabla^2\phi(d)\| : d > 0, \|d\| = 1\}$. Let h be the constant defined in Theorem 6.4. If $\mu > 0$, then given $d \in S(\bar{\lambda}_*) = \{x > 0 : \lambda(x) < \bar{\lambda}_*\}$, we have*

$$\|z\| = \|D^{-1}y(d)\| \leq \frac{\lambda(d)}{(1+\beta)} \leq \|P_d\nabla\psi_d(e_d)\| \leq \tag{6.27}$$

$$\left(\frac{h^pq\rho^2\theta p}{\mu} + M\right)\frac{\lambda(d)}{(1+\beta)} = O\Big(\frac{\theta q\lambda(d)}{\mu}\Big).$$

In either case ($\mu > 0$, or $\mu \leq 0$), given any $t \in (0,1]$, suppose that $d \in S_t(\bar{\lambda}_*) = \{x : x > 0, \lambda_t(x) < \bar{\lambda}_*\}$ is a point obtained via Phase I of the Path-Following algorithm (see § 2.7). Then,

$$\|z_t\| = \|D^{-1}y_t(d)\| \leq \frac{\lambda_t(d)}{(1+\beta)} \leq \|P_d \nabla f_d^t(e_d)\| \leq \tag{6.28}$$

$$\left([1+\frac{\rho}{1+\beta}]^p q\rho^2 \left(\frac{1}{t}\right)^{9p\rho\sqrt{\theta}-1} + M\right)\frac{\lambda_t(d)}{(1+\beta)} = O\left(q\lambda_t(d)(\frac{1}{t})^{9p\rho\sqrt{\theta}}\right).$$

If $\mu > 0$, and $p > 1$, then given any $t \in (0,1]$, for all $d \in S_t(\lambda_*)$, we have

$$\|P_d \nabla f_d^t(e_d)\| \leq \tag{6.29}$$

$$\left(q\rho^2 h^p \max\left\{\theta, \left(\frac{1+\|u\|}{p\mu}\right)^{\frac{1}{p-1}}\right\}^p + M\right)\frac{\lambda_t(d)}{(1+\beta)} = O\left(\frac{\theta q\|u\|\lambda_t(d)}{\mu}\right).$$

Proof. The first two inequalities in (6.27) have already been proved in Lemma 5.2. The next inequality follows by bounding $\|d\|$, $d \in S(\bar{\lambda}_*)$ in Lemma 6.8, using Theorem 6.1, and Theorem 6.4. Again the first two inequalities in (6.28) have already been proved in Lemma 5.2. The next inequality follows by bounding $d \in S_t(\bar{\lambda}_*)$ in Lemma 6.8, using Theorem 6.5. The bound in (6.29) follows from bounding $t^{1/p}d$, in Lemma 6.8, using Theorem 6.2 and Theorem 6.4. □

7. The Potential-Reduction Complexity Theorem. In this section we will examine the application of the Potential-Reduction algorithm, described in § 2.7 for solving ϵ-HP, ϵ-SP, ϵ-HSP, and ϵ-ASP, given a bounded operator-cone. While the algorithm relies on some basic properties of self-concordance given in Nesterov and Nemirovskii's Theorem (Theorem 2.24), it in fact gives a simpler and more general algorithm than their corresponding algorithm for solving the HP formulation of conic LP (see Chapter 4 of Nesterov and Nemirovskii [17]). Given $d > 0$, the algorithm simply replaces d with $d' = NEW(\psi, d)$ and repeats. By making use of the Scaling Duality Theorem (Theorem 4.5), we will see that this simple algorithm solves the desired problems. We need some preliminary results.

LEMMA 7.1. *Assume $x \in W \cap K^\circ$ satisfies $\phi(x) > 0$. Let*

$$\tau_x = \left(\frac{\theta}{p\phi(x)}\right)^{1/p}.$$

Then,

$$\psi(x) \geq \psi(\tau_x\, x) = \frac{\theta}{p}(1 - \ln\frac{\theta}{p} + \ln X(x)).$$

Proof. We have

$$\psi(x) \geq \min\{\psi(tx) : t \in (0,\infty)\} = \tag{7.1}$$

$$\min\{t^p\phi(x) + F(x) - \theta \ln t : t \in (0,\infty)\} = \psi(\tau x).□$$

Let

$$R = \sup\{exp(-\frac{p}{\theta}F(x)) : x \in W \cap K^\circ, \|x\| = 1\}. \tag{7.2}$$

Since $F(x)$ approaches infinity as x approaches a boundary point of K, and in a finite-dimensional Banach space the boundary of the unit ball is compact, it follows that R is finite.

COROLLARY 7.2. *Suppose $\mu > 0$. Then,*

$$\psi^* \geq \frac{\theta}{p}(1 - \ln\frac{\theta}{p} + \ln\frac{\mu}{R}),$$

and for all $x \in W \cap K^\circ$ we have

$$\ln\left(\frac{X(x)}{X^*}\right)^{\frac{\theta}{p}} \leq \psi(x) - \psi^*.$$

Proof. The first inequality is an immediate consequence of the inequality in Lemma 7.1, definitions of R, μ, and that $X(x)$ is homogeneous of degree zero. To prove the second result, from Theorem 3.2 if d^* (the minimizer of ψ) exists, then it is also the minimizer of $X(x)$. From this, and the inequality in Lemma 7.1, we have

$$\psi^* = \frac{\theta}{p}(1 - \ln\frac{\theta}{p} + \ln X^*). \tag{7.3}$$

Subtracting the above from the inequality in Lemma 7.1, the desired results follow. □

THEOREM 7.3. (Potential-Reduction Complexity Theorem) *Assume ϕ is β-compatible, and $\epsilon \in (0,1)$. Consider the Potential-Reduction algorithm, and let*

$$\sigma = exp[\frac{p}{\theta}\psi(x^0) - 1 + \ln\frac{\theta}{p}], \quad q = \sup\{\|\nabla^2\phi(d)\| : d > 0, \|d\| = 1\}.$$

If $\mu \leq 0$, the number of iterations to solve ϵ-HP is

$$O\left(\theta \ln\frac{R\sigma}{\epsilon}\right). \tag{7.4}$$

If $\mu > 0$, the number of iterations to solve ϵ-SP or ϵ-HSP is

$$O\left(\theta \ln\frac{R\sigma}{\mu} + \ln\ln\frac{1}{\epsilon}\right). \tag{7.5}$$

If $T(K)$ is a given bounded operator-cone, the number of iterations to solve ϵ-ASP is

$$O\left(\theta \ln\frac{R\sigma}{\mu} + \ln\ln(1 + \frac{\rho}{1+\beta}) + \ln\ln\frac{q}{\epsilon}\right). \tag{7.6}$$

Proof. Let the k-th iterate of the algorithm be denoted by x^k. Let $\lambda_k = \lambda(x^k)$. We claim that if $\lambda_k \geq \bar{\lambda}$, then

$$\phi\left(\frac{x^k}{\|x^k\|}\right) \leq R\sigma\ exp(-\frac{k\delta p}{\theta}), \quad \delta = \frac{(\bar{\lambda} - \ln(1+\bar{\lambda}))}{(1+\beta)^2}. \tag{7.7}$$

From Theorem 2.24, if $\lambda_j < \bar{\lambda}$, then so is λ_{j+1}. Thus, if $\lambda_k \geq \bar{\lambda}$, then $\lambda_j \geq 1$ for all $j = 0, \ldots, k$. From the same theorem, $\psi(x^k) - \psi(x^0) \leq -k\delta$. From this and the inequality relating $X(x)$ and $\psi(x)$ in Lemma 7.1, and the definition of σ it easily follows that

$$X(x^k) \leq \sigma \; exp(-\frac{k\delta p}{\theta}). \tag{7.8}$$

From (7.8), the fact that $X(tx) = X(x)$, and the definition of R (see (7.2)), the proof of (7.7) is immediate.

Suppose that $\mu \leq 0$. Then, we claim that the algorithm will never obtain a point $d > 0$ such that $\lambda(d) < 1$. Otherwise, from Theorem 2.24 the set $S(\lambda_*) = \{d > 0 : \lambda(d) < \lambda_*\}$ is nonempty. Then, from Theorem 4.9, the minimizer d^* of ψ exists. But, by the Scaling Duality Theorem (Theorem 4.5), either $\mu \leq 0$, or d^* exists. This proves the claim. Now the inequality in (7.7) proves the claimed bound on the number of iterations to solve ϵ-HP.

Suppose that $\mu > 0$. From Corollary 7.2, in order to find the number of iterations to solve ϵ-HSP, it suffices to bound the number of iterations to solve ϵ-SP. We next bound the number of steps to solve ϵ-SP. Since we must have

$$\phi\left(\frac{x^k}{\|x^k\|}\right) \geq \mu, \tag{7.9}$$

from (7.7) and the Scaling Duality Theorem (Theorem 4.5), the number of iterations to obtain a point $d \in S(\lambda_*)$ satisfies $O(\theta \ln \frac{R\sigma}{\mu})$.

We claim that once we have a point $d \in S(\lambda_*)$, the number of subsequent steps to get a point d' such that $\psi(d') - \psi^* \leq \epsilon$, is $O(\ln \ln \frac{1}{\epsilon})$.

To prove this we first note that from Theorem 2.24, starting with $d^0 = d$, after k subsequent iterations, we obtain $d^k > 0$ such that $\lambda_k = \lambda(d^k)$ satisfies $\lambda_{k+1} < \lambda_k^2/(1-\lambda_*)^2$. This and a simple induction gives $\lambda_k \leq \lambda_* \bar{r}^{2^k-1}, \quad j = 0, \ldots, k$, where $\bar{r} = \lambda_*/(1-\lambda_*)^2$. Thus we get

$$\lambda_k \leq \bar{r}^{2^k}. \tag{7.10}$$

Now suppose we have obtained a point d^k such that

$$\lambda_k < \min\{\lambda_*, \frac{\epsilon}{3}(1+\beta)^2(1-\omega_*)\}, \tag{7.11}$$

where $\omega_* = \omega(\lambda_*)$, and $\omega(\lambda) = 1-(1-3\lambda)^{1/3}$. We claim that $\psi(d^k)-\psi^* \leq \epsilon$. Clearly, from (7.10), the number of iterations to obtain d^k is $O(\ln \ln \frac{1}{\epsilon})$. Since $\lambda_k < \lambda_*$, and $\omega(\lambda_k) \leq 3\lambda_k < 1$, together with Theorem 2.24, see (2.12), it follows that

$$\epsilon \geq \frac{\omega(\lambda_k)}{(1+\beta)^2(1-w_*)} \geq \frac{\omega^2(\lambda_k)}{(1+\beta)^2(1-w_*)} \geq \frac{\omega^2(\lambda_k)}{(1+\beta)^2(1-w(\lambda_k))} \tag{7.12}$$

$$\geq \frac{\omega^2(\lambda_k)(1+w(\lambda_k))}{2(1+\beta)^2(1-w(\lambda_k))} \geq \psi(d') - \psi^*.$$

Now suppose that $\mu > 0$, and $T(K)$ is a bounded operator-cone. We will bound the number of iterations to solve ϵ-ASP. From (7.10), once we have obtained a point $d \in S(\lambda_*)$, the number of iterations to obtain a point $\bar{d} \in S(\bar{\lambda}_*) = \{x > 0 :$

$\lambda(x) < \bar{\lambda}_*\}$, where $\bar{\lambda}_* = \lambda_*(1+\beta)/(1+\beta+\rho)$, satisfies $O(\ln\ln 1/\bar{\lambda}_*)$. In order to estimate the number of subsequent iterations to solve ϵ-ASP, from Theorem 6.9, we have $\|P_{\bar{d}}\nabla\psi_{\bar{d}}(e_{\bar{d}})\| = O(\theta q\lambda(\bar{d})/\mu)$. Starting with $d^0 = \bar{d}$, after k subsequent iterations, we obtain from (7.10) a point $d^k > 0$ such that $\lambda_k = \lambda(d^k) \leq \bar{r}^{2^k}$. Thus, to have $\theta q\lambda_k/\mu = O(\epsilon)$, it suffices to have $k = O(\ln\ln\frac{\theta q}{\mu\epsilon})$. This completes the proof. □

Remark. Even for $\phi(x)$ linear, the above algorithm is simpler than a potential-reduction algorithm described by Nesterov and Nemirovskii [17] (Chapter 4) which is essentially capable of solving ϵ-HP, where it is assumed that $\mu = 0$. We have shown that we need not make any assumptions on μ, and that the algorithm is also capable of solving ϵ-SP, ϵ-HSP, and ϵ-ASP with a bounded operator-cone.

For the case where ϕ is linear, ϵ-ASP is trivial since in this case there is a constant a such that for all d, $\|P_d\nabla\psi_d(e_d)\| = a\lambda(d)$. This is no longer the case when ϕ is nonlinear, even if $p = 1$. For instance in Kalantari [9] it is shown that convex quadratic programming over linear constraints can be formulated as an HP with a nonlinear but convex ϕ, where $p = 1$, nevertheless shown to be β-compatible.

8. The Path-Following Complexity Theorem. In this section we first prove a lemma, called the Path-Following Lemma (Lemma 8.1). It reveals an important property under the satisfiability of the Uniform Scaling Duality, and the boundedness of the corresponding operator-cone, $T(K)$. We then prove a theorem, called the Path-Following Theorem (Theorem 8.3) establishing the significance of computing approximate minimizers of $f^t(x) = t\phi(x) + tu^Tx + F(x)$, given that $F(x)$ is θ-normal. These results do not require β-compatibly of ϕ. However, given β-compatibility, the Path-Following Theorem together with the previously derived results, in particular the bounds of Theorem 6.9, give appropriate values for $t_* \in (0,1)$ so that the Path-Following algorithm (see § 2.7) can solve, in polynomial-time, any of the four problems: ϵ-HP, ϵ-SP, ϵ-HSP, or ϵ-ASP.

LEMMA 8.1. (Path-Following Lemma) *Assume $p > 1$ and $T(K)$ a given bounded operator-cone. Given positive numbers $\gamma_1 < \gamma_2$, define*

$$C(p,\theta,\rho,N,\|u\|,\gamma_1,\gamma_2) = \frac{[(\theta+\gamma_1\sqrt{N})\rho + (\gamma_2-\gamma_1)]\rho^{p-1}\|u\|^p}{p(\gamma_2-\gamma_1)^p}. \tag{8.1}$$

Given $t \in (0,1]$, suppose there exists $d \in W \cap K^\circ$ such that

$$\|P_d\nabla f_d^t(e_d)\| \leq \gamma_1. \tag{8.2}$$

Let

$$\hat{d} = t^{1/p}d. \tag{8.3}$$

If

$$\|P_{\hat{d}}\nabla\psi_{\hat{d}}(e_{\hat{d}})\| \geq \gamma_2, \tag{8.4}$$

then

$$\phi\left(\frac{d}{\|d\|}\right) \leq C(p,\theta,\rho,N,\|u\|,\gamma_1,\gamma_2)t^{p-1}. \tag{8.5}$$

Proof. Let $d \in W \cap K^\circ$ be a point satisfying (8.2). We have

$$P_d\nabla f_d^t(e_d) = tP_d\nabla\phi_d(e_d) + tP_dD^Tu + P_d\nabla F_d(e_d), \tag{8.6}$$

where $D = T(d)$. Let $\alpha = t^{1/p}$. Thus, $\hat{d} = \alpha d$. From the chain rule and property (P2) of Proposition 2.8, we have

$$\nabla \phi_{\hat{d}}(e_{\hat{d}}) = \hat{D}^T \nabla \phi(\hat{d}) = \alpha D^T \nabla \phi(\alpha d) = \alpha \alpha^{p-1} D^T \nabla \phi(d) = \alpha^p \nabla \phi_d(e_d). \tag{8.7}$$

Let $\hat{D} = T(\hat{d})$. Since $P_{\hat{d}} = P_d$,

$$P_{\hat{d}} \hat{D}^T u = \alpha P_d D^T u. \tag{8.8}$$

From the chain rule and property (P6) of Proposition 2.8, we have

$$\nabla F_{\hat{d}}(e_{\hat{d}}) = \hat{D}^T \nabla F(\hat{d}) = \alpha D^T \nabla F(\alpha d) = \alpha \alpha^{-1} D^T \nabla F(d) = \nabla F_d(e_d). \tag{8.9}$$

Substituting (8.7) and (8.9) into (8.6), we get

$$P_d \nabla f_d^t(e_d) = P_{\hat{d}} \nabla \phi_{\hat{d}}(e_{\hat{d}}) + t^{(p-1)/p} P_{\hat{d}} \hat{D}^T u + P_{\hat{d}} \nabla F_{\hat{d}}(e_{\hat{d}}). \tag{8.10}$$

Now assume $\hat{d}$ satisfies (8.4) and let

$$v = P_{\hat{d}} \nabla \psi_{\hat{d}}(e_{\hat{d}}) = P_{\hat{d}} \nabla \phi_{\hat{d}}(e_{\hat{d}}) + P_{\hat{d}} \nabla F_{\hat{d}}(e_{\hat{d}}), \qquad w = t^{(p-1)/p} P_{\hat{d}} \hat{D}^T u. \tag{8.11}$$

Thus, from (8.2) and (8.4), $\|v+w\| \le \gamma_1$ and $\|v\| \ge \gamma_2$, respectively. Since $\|v\| - \|w\| \le \|v + w\|$, we have $\|w\| \ge \|v\| - \|v + w\|$. Thus,

$$t^{(p-1)/p} \|P_{\hat{d}} \hat{D}^T u\| \ge (\gamma_2 - \gamma_1). \tag{8.12}$$

Since $\|\hat{D}^T\| = \|\hat{D}\| \le \rho \|\hat{d}\|$, and $\|P_{\hat{d}}\| = 1$, we get

$$\|P_{\hat{d}} \hat{D} u\| \le \rho \|P_{\hat{d}}\| \ \|\hat{d}\| \ \|u\| \le \rho \|\hat{d}\| \ \|u\|. \tag{8.13}$$

Thus, from (8.12) and (8.13), we get

$$\frac{1}{\|\hat{d}\|} \le \frac{\rho t^{(p-1)/p} \|u\|}{(\gamma_2 - \gamma_1)}. \tag{8.14}$$

From the chain rule and properties (P1) and (P8) of Proposition 2.8, we have

$$e_{\hat{d}}^T \nabla \phi_{\hat{d}}(e_{\hat{d}}) = p \phi_{\hat{d}}(e_{\hat{d}}) = p \phi(\hat{d}), \quad e_{\hat{d}}^T \nabla F_{\hat{d}}(e_{\hat{d}}) = \hat{d}^T \nabla F(\hat{d}) = -\theta. \tag{8.15}$$

Taking the inner product with $e_{\hat{d}}$ in (8.10), using the fact that $P_{\hat{d}} e_{\hat{d}} = e_{\hat{d}}$, together with (8.15), Cauchy-Schwarz inequality, and condition (2) of operator-cone (see 2.13), we get

$$e_{\hat{d}}^T P_d \nabla f_d^t(e_d) = p \phi(\hat{d}) + t^{(p-1)/p} u^T \hat{d} - \theta \le \|e_{\hat{d}}\| \ \|P_d \nabla f_d^t(e_d)\| \le \sqrt{N} \gamma_1. \tag{8.16}$$

This implies,

$$p \phi(\hat{d}) \le \theta + \sqrt{N} \gamma_1 + t^{(p-1)/p} \|\hat{d}\| \ \|u\|. \tag{8.17}$$

Dividing the above inequality by $\|\hat{d}\|^p$, from homogeneity of ϕ the resulting left-hand-side reduces to $p\phi(d/\|d\|)$. To bound the resulting right-hand-side, from (8.14) we have

$$\frac{\theta + \sqrt{N} \gamma_1}{\|\hat{d}\|^p} \le \frac{\rho^p t^{(p-1)} \|u\|^p}{(\gamma_2 - \gamma_1)^p}. \tag{8.18}$$

Also, from (8.14) we have

$$\frac{t^{(p-1)/p}\|\hat{d}\| \, \|u\|}{\|\hat{d}\|^p} \le \frac{\rho^{p-1} t^{(p-1)} \|u\|^p}{(\gamma_2 - \gamma_1)^{p-1}}. \tag{8.19}$$

From these two bounds the desired bound on $\phi(d/\|d\|)$ follows. □

COROLLARY 8.2. *Suppose that ϕ is β-compatible. If $\mu \le 0$, then*

$$\|\hat{d}_t^*\| \ge \frac{\gamma^*}{\rho \|u\| t^{(p-1)/p}}.$$

In particular, $\lim_{t\to 0} \|\hat{d}_t^*\| = \infty$, *and* $\lim_{t\to 0} \|\hat{d}_t^*\| = \infty$.

Proof. Since ϕ is β-compatible, the Uniform Scaling Duality (Theorem 5.3) is valid . Thus, $\mu \le 0$ implies that $\|P_d \nabla \psi_d(e_d)\| \ge \gamma^*$, for all $d > 0$. Since d_t^* exists, choosing $\gamma_2 = \gamma^*$, the inequality (8.14) is valid for all $\gamma_1 < \gamma^*$. Now letting γ_1 converge to zero, we get the desired result. □

THEOREM 8.3. (Path-Following Theorem) *Assume that ϕ is convex, $p > 1$, $T(K)$ a given bounded operator-cone with respect to which the Uniform Scaling Duality holds with parameter γ^*, and let F be a θ-normal barrier for the cone K.*

(I): Assume that $\mu > 0$. Given $\epsilon \in (0, \gamma^*]$, let $t \in (0, 1]$ satisfy

$$C(p, \theta, \rho, N, \|u\|, \frac{1}{2}\epsilon, \epsilon) t^{p-1} < \mu. \tag{8.20}$$

There exists $d > 0$ satisfying

$$\|P_d \nabla f_d^t(e_d)\| \le \frac{\epsilon}{2}. \tag{8.21}$$

Moreover, $\hat{d} = t^{1/p} d$ satisfies

$$\|P_{\hat{d}} \nabla \psi_{\hat{d}}(e_{\hat{d}})\| < \epsilon. \tag{8.22}$$

(II): Assume $\mu \le 0$. Given $\epsilon \in (0, \gamma^*]$, let $t \in (0, 1]$ satisfy

$$C(p, \theta, N, \|u\|, \frac{1}{2}\gamma^*, \gamma^*) t^{p-1} \le \epsilon. \tag{8.23}$$

There exists $d > 0$ satisfying

$$\|P_d \nabla f_d^t(e_d)\| \le \frac{1}{2}\gamma^*. \tag{8.24}$$

Moreover,

$$\phi\left(\frac{d}{\|d\|}\right) \le \epsilon. \tag{8.25}$$

Proof. From Theorem 4.8, for each t, $d_t^* = \text{argmin}\{f^t(x) : x > 0\}$ exists. It follows that if $D_t^* = T(d_t^*)$, then $P_{d_t^*} \nabla f_{d_t^*}^t(e_{d_t^*}) = 0$. The rest of the proof is the immediate consequence of the Path-Following Lemma (Lemma 8.1). □

The following significant theorem reveals the algorithmic implication of the Path-Following Theorem.

THEOREM 8.4. (Path-Following Complexity Theorem) *Assume ϕ is β-compatible, $p > 1$, and $T(K)$ a bounded operator-cone. Let $\epsilon \in (0, \gamma^*]$ be given. Let $q = \sup\{\|\nabla^2\phi(d)\| : d > 0, \|d\| = 1\}$. Then, the solvability of ϵ-HP ϵ-SP, ϵ-HSP, or ϵ-ASP can all be tested in polynomial-time via the Path-Following algorithm (see § 2.7). More precisely:*

if $\mu \leq 0$, the number of Newton iterations to solve ϵ-HP is

$$O\biggl(\frac{1+\beta}{p-1}\sqrt{\theta}\ln\Bigl[C(p,\theta,\rho,N,\|u\|,\frac{1}{2}\gamma^*,\gamma^*)\frac{1}{\epsilon}\Bigr]+\ln\ln q\biggr)=O\biggl(\sqrt{\theta}\ln\frac{\theta\|u\|}{\epsilon}+\ln\ln q\biggr),$$

if $\mu > 0$, the number of Newton iterations to solve ϵ-ASP is

$$O\biggl(\frac{1+\beta}{p-1}\sqrt{\theta}\ln\Bigl[C(p,\theta,\rho,N,\|u\|,\frac{1}{2}\epsilon,\epsilon)\frac{1}{\mu}\Bigr]+\ln\ln q\biggr)=O\biggl(\sqrt{\theta}\ln\frac{\theta\|u\|}{\mu\epsilon}+\ln\ln q\biggr),$$

and the number of Newton iterations to solve ϵ-SP or ϵ-HSP is

$$O\biggl(\frac{1+\beta}{p-1}\sqrt{\theta}\ln\Bigl[C(p,\theta,\rho,N,\|u\|,\frac{\lambda_*}{2(1+\beta)},\frac{\lambda_*}{1+\beta})\frac{1}{\mu}\Bigr]+\ln\ln\frac{q}{\epsilon}\biggr)=$$

$$O\biggl(\sqrt{\theta}\ln\frac{\theta\|u\|}{\mu}+\ln\ln\frac{q}{\epsilon}\biggr).$$

Proof. The stated right-hand-side bounds assume that the parameter ρ, γ^* are $O(1)$ constants, and $N = O(\theta)$ (note from Proposition 2.15 for T_F, $N = \theta$). Suppose that $\mu \leq 0$. To prove the bound on the number of Newton iterations to solve ϵ-HP, from the Path-Following Theorem (Theorem 8.3), it suffices to compute $d > 0$ such that $\|P_d \nabla f_d^{t_*}(e_d)\| \leq \gamma^*/2$, where t_* is selected so that the following equation is satisfied

$$\frac{1}{t_*}=\Bigl(C(p,\theta,\rho,N,\|u\|,\frac{1}{2}\gamma^*,\gamma^*)\frac{1}{\epsilon}\Bigr)^{\frac{1}{p-1}}. \tag{8.26}$$

From the above, and Theorem 2.24, we can bound the number of steps, k_1, of Phase I of the Path-Following algorithm. This is the number of iterations needed to compute $d' > 0$ such that $\lambda_{t_*}(d') < \lambda_*$ and is equal to $k_{t_*} = \lceil \frac{9}{4}(1+\beta)\sqrt{\theta}\ln(\frac{1}{t_*})\rceil$. Now we need to bound the number of steps, k_2 of Phase II, i.e., the number of Newton iterations applied to f^{t_*}, in order to get from $d^0 = d'$ to a point $d^{k_2} = d > 0$, satisfying $\|P_d \nabla f_d^{t_*}(e_d)\| \leq \gamma^*/2$. Let $\lambda_k = \lambda_{t_*}(d^k)$, $k = 0, \ldots, k_2$. From Theorem 2.24, it follows that $\lambda_k \leq r^{2^k}$ (see (6.14)), where $r = \bar{\lambda}_*/(1-\lambda_*)^2$, $\bar{\lambda}_* = \lambda_*(1+\beta)/(1+\beta+\rho)$. From Theorem 6.9, if $\lambda_k < \bar{\lambda}_*$, then we have

$$\|P_{d^k}\nabla f_{d^k}^{t_*}(e_{d^k})\| \leq \Bigl((1+\frac{\rho}{1+\beta})^p q\rho^2\Bigl(\frac{1}{t_*}\Bigr)^{9p\rho\sqrt{\theta}}+M\Bigr)\frac{r^{2^k}}{(1+\beta)}. \tag{8.27}$$

Substituting in the right-hand-side of (8.27) for $1/t_*$ from the equation (8.26), and bounding the resulting expression by $\gamma^*/2$, with the assumption that $M = O(1)$, we get

$$k_2=O\biggl(\ln\ln(1+\frac{\rho}{1+\beta})+\ln\ln q+\ln\ln\Bigl(C(p,\theta,\rho,N,\|u\|,\frac{1}{2}\gamma^*,\gamma^*)\frac{1}{\epsilon}\Bigr)^{\frac{9p\rho\sqrt{\theta}}{p-1}}\biggr).$$

Combining the number of iterations of Phase I and Phase II, we get the desired result on the complexity of ϵ-HP.

Suppose that $\mu > 0$. To solve ϵ-ASP, from the Path-Following Theorem, it suffices to compute $d > 0$ such that $\|P_d \nabla f_d^{t_*}(e_d)\| \leq \epsilon/2$, where t_* satisfies

$$\frac{1}{t_*} = \left(C(p, \theta, \rho, N, \|u\|, \frac{1}{2}\epsilon, \epsilon) \frac{1}{2\mu} \right)^{\frac{1}{p-1}}. \tag{8.28}$$

As in the case of ϵ-HP, from this we can bound the number of steps of Phase I of the Path-Following algorithm. The bound on the number of steps of Phase II follows in similar fashion as in the case of ϵ-HP, again by bounding the right-hand-side of (8.27), but using the new value of $1/t_*$ given in (8.28). Note that in this case, since $\mu > 0$, we can use the alternative bound from Theorem 6.9. This implies that once we have a point $d \in S_{t_*}(\lambda_*)$, the number of steps of Phase II is $O(\ln \ln \frac{1}{\epsilon\mu})$, independent of t_*. However, using either bound, the total complexity remains unchanged.

To bound the number of steps of ϵ-SP, or ϵ-HSP, we first solve $\frac{\lambda_*}{(1+\beta)}$-ASP, i.e., compute a point $d > 0$ such that $\|P_d \nabla \psi_d(e_d)\| \leq \frac{\lambda_*}{(1+\beta)}$. The number of necessary iterations can thus be bounded from the previous bound of the theorem. From Theorem 6.9, it follows that $\lambda(d) < \lambda_*$. The number of subsequent iterations to get a point d' such that $\psi(d') - \psi^* \leq \epsilon$ is $O(\ln \ln \frac{1}{\epsilon})$. The proof of the latter result was already established within the proof of the Potential-Reduction Complexity Theorem (Theorem 7.3). Thus, the proof of the Path-Following theorem is complete. □

9. Applications. The following theorem shows the wide range of applications of the two algorithms.

THEOREM 9.1. *Consider any HP/SP/HSP/ASP over a pointed convex cone K where ϕ is of homogeneous degree p and β-compatible with the corresponding θ-normal barrier F. Assume that $T(K) = T_F = \{T(d) = D = \nabla^2 F(d)^{-1/2} : d \in K^\circ\}$. Given $\epsilon \in (0,1)$, the solvability of ϵ-HP ϵ-SP, ϵ-HSP, or ϵ-ASP can all be tested in polynomial-time via the Potential-Reduction algorithm and if $p > 1$ via the Path-Following algorithm. In particular, these algorithms apply to the case where ϕ is linear, or quadratic. To apply the Path-Following algorithm to a linear ϕ we simply need to square it.*

Proof. From Theorem 2.17, T_F is a bounded operator-cone. Thus, given any β-compatible ϕ the Potential-Reduction Complexity Theorem (Theorem 7.3) holds, and when $p > 1$, the Path-Following Complexity Theorem (Theorem 8.4). □

COROLLARY 9.2. *Consider any HP/SP/HSP/ASP over the nonnegative cone $K = \Re^n_+$, or the semidefinite cone $K = S_n^+$, or the second-order cone $K = SO^{n+1}$, where ϕ is of homogeneous degree p and β-compatible with the corresponding θ-normal barrier F, and $T(K) = T_F = \{T(d) = D = \nabla^2 F(d)^{-1/2} : d \in K^\circ\}$. Given $\epsilon \in (0,1)$, the solvability of ϵ-HP ϵ-SP, ϵ-HSP, or ϵ-ASP can all be tested in polynomial-time via the Potential-Reduction algorithm, and when $p > 1$ the Path-Following algorithm. In particular, both algorithms apply to linear programming, convex quadratic programming over linear or convex quadratic constraints, semidefinite programming, as well as their corresponding scaling problems. Moreover, the problem of computing the minimum ratio of the arithmetic-geometric means over an arbitrary subspace of $\Re^n$, or the minimum ratio of trace-determinant over an arbitrary subspace of S_n, can be established in polynomial-time via both algorithms.*

Proof. Clearly the problems are special cases of those stated in the previous theorem. To apply the Path-Following algorithm to linear programming, we consider Karmarkar's canonical LP and simply replace the linear objective c^Tx with $\phi(x) = (c^Tx)^2$. We may also formulate linear programming as that of testing if a positive semidefinite quadratic matrix has a nontrivial nonnegative zero. To apply the Path-Following algorithm to semidefinite programming, we may use Nesterov and Nemirovskii's HP formulation of semidefinite programming with a linear objective over S_n^+, again replacing the linear objective $tr(cx)$ with the quadratic objective $\phi(x) = tr(cx)^2$. To apply the Path-Following algorithm to quadratic programming, we use the fact that it can be formulated as a semidefinite programming problem.

To compute the minimum of $c^Tx/(\prod_{i=1}^n x_i)^{1/n}$ over the intersection of the nonnegative orthant and an arbitrary subspace W of $\Re^n$ via the Path-Following algorithm, we minimize $(c^Tx)^2/(\prod_{i=1}^n x_i)^{2/n}$, i.e., solve ϵ-HSP with $\phi(x) = (c^Tx)^2$. To compute the minimum of $tr(cx)/det(x)^{1/n}$ over the intersection of the semidefinite cone and an arbitrary subspace W of S_n via the Path-Following algorithm, we solve ϵ-HSP with $\phi(x) = tr(cx)^2$. □

Remark. When trying to determine the solvability of ϵ-HP corresponding to any β-compatible ϕ, regardless of the underlying cone K, we can incorporate a simple duality test in either the Potential-Reduction algorithm, or the Path-Following algorithm. To see this recall the definition of scaling dualities (Definition 2.22). Since the Uniform Scaling Duality is valid with parameter $\gamma^* = \frac{1}{(1+\beta)}$, within each iteration we check if the current iterate, d, satisfies $\|P_d\nabla\psi_d(e_d)\| < \gamma^*$. If so, then $\mu > 0$. Hence HP is unsolvable. Given that the Scaling Separation Duality holds , we can strengthen this duality test. For instance, for the nonnegative orthant and semidefinite cone this duality hods, see Kalantari [9]. Moreover, for these cones $\gamma^* = 1$. Thus within each iteration we check if the current iterate satisfies $P_d\nabla\phi_d(e) > 0$. The set $\{d > 0 : \|P_d\nabla\psi_d(e)\| = \|P_d\nabla\phi_d(e) - e\| < \gamma^*\}$, if nonempty, is a proper subset of the set $\{d > 0 : P_d\nabla\phi_d(e) > 0\}$.

Remark. As with the proposed Potential-Reduction algorithm, the proposed Path-Following algorithm of this paper, although making use of the theory of self-concordance, is conceptually simpler than the path-following algorithm suggested by Nesterov and Nemirovskii. Moreover, our path-following algorithm is also capable of solving - in polynomial-time - any of the other three problems, SP, HSP, and ASP, and is applicable to any β-compatible ϕ with $p > 1$. We also mention that although our Path-Following algorithm has similarities to the class of so-called barrier-generated path-following algorithms, it is fundamentally different than those algorithms, and other existing path-following algorithms. Firstly, and ironically, the case of HP with $p = 1$, to which linear programming (or more generally conic linear programming) belong, is in fact a singular case for our Path-Following algorithm. However, to apply our Path-Following algorithm to HP with ϕ linear, all is needed is to replace ϕ with ϕ^2. Secondly, the domain of the optimization of f^t is unbounded since it is the intersection of a cone and a subspace. It is even non-trivial to show that the corresponding central-path exists. This is contrary to barrier-generated methods for the minimization of a convex function, say, $g(x)$, over a bounded convex domain, say, G. Assuming that G has an interior point, and given a smooth convex barrier $B(x)$, in view of the fact that G is bounded, it is easy to show that the central-path $\{x_t^* = \text{argmin}\{tg(x) + B(x) : x \in G\}, t \in (0,\infty)\}$ exists. Also, under mild assumptions it can be shown that x_t^* converges to the minimizer of g. Thus, even for LP our proposed Path-Following algorithm is different than the existing polynomial-time path-following algorithm for LP over bounded domains, or other path-following methods, including those based on primal-dual formulations, see e.g. Nesterov and

Nemirovskii [17]. In the case of f^t, under the assumption of β-compatibility of ϕ, the existence of the central-path can be established from a theorem on self-concordance (see Theorem 2.24). However, in general the existence of the central-path employs our new Convex Conic Programming Duality. Thirdly, unlike the existing barrier-generated path-following methods, in our path-following method the central-path by itself is of no direct significance. Rather, its projection onto the unit sphere. Fourthly, issues regarding the approximation of the projected central-path, and the significance of this approximation, as well as their application in terms of ϵ-approximate version of the four problems are issues whose answers rely on the scaling dualities, bounds, and sensitivity analysis, specifically developed in this article, as opposed to the mere application of general results from convex programming, or those implied by the theory of self-concordance.

REFERENCES

[1] V. Chvátal, *Notes on the Khachiyan-Kalantari algorithm*, Rutgers University, New Brunswick, NJ (2002). `http://www.cs.rutgers.edu/~chvatal/521/khakal.pdf`

[2] Y. Jin and B. Kalantari, *Chvátal's Lemma and matrix scaling.* Technical Report DCS-TR-581, Department of Computer Science, Rutgers University, New Bruswick, NJ, 2004.

[3] B. Kalantari, *Karmarkar's algorithm with improved steps*, Math. Programming, 46 (1990), pp. 73-78.

[4] B. Kalantari, *Canonical problems for quadratic programming and projective methods for their solution*, Proceedings of AMS conference "Mathematical problems arising from linear programming", 1988. In: Contemporary Mathematics, J.C. Lagarias and M. Todd (eds.), (1990) Volume 114, pp. 243-263.

[5] B. Kalantari, *Generalization of Karmarkar's algorithm to convex homogeneous functions*, Oper. Res. Lett., 11 (1992), pp. 93-98.

[6] B. Kalantari, *A simple polynomial time algorithm for a convex hull problem equivalent to linear programming,* Combinatorics Advances, C.J. Colbourn and E.S. Mahmoodian (eds.), Kluwer Academic Publishers (1995), pp. 207-216.

[7] B. Kalantari, *A Theorem of the alternative for multihomogeneous functions and its relationship to diagonal scaling of matrices*, Linear Algebra and its Applications, 236 (1996), pp. 1-24.

[8] B. Kalantari and M.R. Emamy-K, *On linear programming and matrix scaling over the algebraic numbers*, Linear Algebra and its Applications, 262 (1997), pp. 283-306.

[9] B. Kalantari, *Scaling dualities and self-concordant homogeneous programming in finite dimensional spaces.* Technical Report DCS-TR-359, Department of Computer Science, Rutgers University, New Brunswick, NJ, 1998.

[10] B. Kalantari, *A note on a boundedness property of normal barriers on convex cones.* Technical Report LCSR-TR-398, Department of Computer Science, Rutgers University, New Brunswick, NJ, 1999.

[11] B. Kalantari, *Semidefinite programming and matrix scaling over the semidefinite cone*, Linear Algebra and its Applications, 375 (2003), pp. 221-243.

[12] B. Kalantari, *My memories of Leonid Khachiyan and a personal tribute for his contributions in linear programming*, Rutgers University, New Brunswick, NJ (2005). `http://www.cs.rutgers.edu/~kalantar/khachiyan2005.pdf`

[13] N. Karmarkar, *A new polynomial time algorithm for linear programming*, Proceedings of Symposium on Theory of Computing, Washington D.C. (1984), pp. 302-311.

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

[15] L. Khachiyan and B. Kalantari, *Diagonal matrix scaling and linear programming*, SIAM J. Optim., 4 (1992), pp. 668-672.

[16] L. Khachiyan, *Diagonal matrix scaling is NP-hard*, Linear Algebra and its Applications, 234 (1996), pp. 173-179.

[17] Y. Nesterov and A.S. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, PA, 1994.