DIMACS Technical Report 98-37
June 1998

SCALING DUALITIES AND SELF-CONCORDANT HOMOGENEOUS PROGRAMMING IN FINITE DIMENSIONAL SPACES [1]

by

Bahman Kalantari [2]

[1] A preliminary version of this paper was presented at SIAM Optimization Conference, Victoria, British Columbia, May 1996

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

DIMACS is a partnership of Rutgers University, Princeton University, AT&T Labs-Research, Bell Labs, Bellcore and NEC Research Institute.

DIMACS is an NSF Science and Technology Center, funded under contract STC–91–19999; and also receives support from the New Jersey Commission on Science and Technology.

ABSTRACT

In this paper first we prove four fundamental theorems of the alternative, called *scaling dualities*, characterizing exact and approximate solvability of four significant conic problems in finite dimensional spaces, defined as: *homogeneous programming* (HP), *scaling problem* (SP), *homogeneous scaling problem* (HSP), and *algebraic scaling problem* (ASP). Let ϕ be a homogeneous function of degree $p > 0$, K a pointed closed convex cone, W a subspace, and F a θ-logarithmically homogeneous barrier for K°. HP tests the existence of a nontrivial zero of ϕ over $W \cap K$. SP, and HSP test the existence of finite minimum of $\psi = \phi + F$, and $X = \phi / exp(-pF/\theta)$ over $W \cap K^\circ$, respectively. ASP tests the solvability of the *scaling equation* (SE), a fundamental equation inherited from properties of homogeneity and those of the *operator-cone*, $T(K) = \{D \equiv \nabla^2 F(d)^{-1/2} : d \in K^\circ\}$. Each D induces a *scaling* of $\nabla\phi(d)$ (or $\nabla^2\phi(d)$), and SE is solvable if and only if there exists a fixed-point under this scaling. In case K is a symmetric cone, the fixed-point is its *center*, e. These four problems together with the scaling dualities offer a new point of view into the theory and practice of convex and nonconvex programming. Nontrivial special cases over the nonnegative orthant include: testing if the convex-hull of a set of points contains the origin (equivalently, testing the solvability of Karmarkar's canonical LP), computing the minimum of the arithmetic-geometric mean ratio over a subspace, testing the solvability of the diagonal matrix scaling equation ($DADe = e$), as well as NP-complete problems. The scaling dualities closely relate these seemingly unrelated problems. Via known conic LP dualities, convex programs can be formulated as HP. Scaling dualities go one step further, allowing us to view HP as a problem *dual* to the corresponding SP, HSP, or ASP. This duality is in the sense that HP is solvable if and only if the other three are not. Using the scaling dualities we describe algorithms that attempt to solve SP, HSP, or ASP. If any of these problems is unsolvable, as a by-product, we solve HP. Our scaling dualities give nontrivial generalization of the arithmetic-geometric mean, trace-determinant, and Hadamard inequalities; matrix scaling theorems; and the classic duality of Gordan. We describe potential-reduction and path-following algorithms for these four problems, resulting in novel and conceptually simple polynomial-time algorithms for linear, quadratic, semidefinite, and self-concordant programming. Furthermore, the algorithms are more powerful than their existing counterparts, since they also establish the polynomial-time solvability of the corresponding SP, HSP, as well as many cases of ASP. The scaling problems either have not been addressed in the literature, or have been treated in very special cases. The algorithms are based on the scaling dualities, significant bounds obtained in this paper, properties of homogeneity, as well as Nesterov and Nemirovskii's machinery of self-concordance. We prove that if ϕ is β-compatible with F, then ϵ-approximate versions of HP, SP, and HSP are all solvable in polynomial-time. Additionally, if the ratio $||D||/||d||$ is uniformly bounded, ASP is also solvable in polynomial-time. The latter result extends the polynomial-time solvability of matrix scaling equation (even in the presence of W) to general cases of SE over the nonnegative cone, or the semidefinite cone, or the second-order cone.

Key words. convex programming, homogeneous functions, Euler's equation, homogeneous programming, Karmarkar's canonical LP, interior-point algorithms, diagonal matrix scaling, quadratic programming, semidefinite programming, self-concordance, Newton's method, convex cones, logarithmically homogeneous barriers, complexity, duality, homogeneous self-dual cones, chain rule, Lagrange multipliers, Hilbert spaces, arithmetic-geometric mean inequality, Hadamard inequality

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

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 function over a pointed closed convex cone and its intersection with a subspace, or proving the nonexistence of such zero. In particular, HP is a canonical formulation of convex programming problems (possible via known conic linear programming dualities), as well as nonconvex, and NP-complete problems. 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 via a simple inductive argument this minimum is well known to be 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 totally 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 matrix. The matrix scaling problem has been a problem of interest since at least the thirties.

There are fundamental theorems of the alternatives relating exact and approximate solutions of these four problems. We will prove four distinct such theorems, and will refer to these theorems as *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 mathematical programming problems gives new insight into the theory of convex and nonconvex programming 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.

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 fundamental inequalities that generalize the classic arithmetic-geometric mean, trace-determinant, and Hadamard inequalities. Also, solvability theorems for scaling equations that generalize the known diagonal matrix scaling theorems, as well as theorems of the alternative that generalize 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 nonnegative 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 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 amount of literature on interior-point algorithms for linear, quadratic, linear complementarity, semidefinite programming, or self-concordant programming that has emerged since Karmarkar's LP algorithm, few papers have dealt with HP, and even fewer with the other three scaling problems, or the fundamental dualities relating them. Even in the monumental work of Nesterov and Nemirovskii, despite their development of a new conic LP duality which makes it possible to state the HP formulation of convex programs, there is no mention of the 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, [34], on the complexity of the path-following method of centers for linear programming, and Nesterov and Nemirovskii's fundamental theory on the class of self-concordant convex programming, [31]. 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, [10], more than a century ago, is equivalent to several HP formulation of linear programming (over the rational or algebraic numbers), as well as Karmarkar's canonical LP.

In this paper we employ basic but important properties of homogeneous functions to prove four fundamental scaling dualities, derive several significant bounds, and by building on Nesterov and Nemirovskii's machinery of self-concordance, we obtain novel polynomial-time potential-reduction and path-following algorithms for many cases of 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. Furthermore, both algorithms are more powerful than their existing counterparts, since they also establish the polynomial-time solvability of the corresponding SP, HSP, and many cases of ASP. Our result extend the polynomial-time solvability of matrix scaling equation (even in the presence of a subspace) to very general cases of SE over the nonnegative cone, or the semidefinite cone, or the second-order cone. The polynomial-time solvable cases of HSP, a nonconvex program, even if the homogeneous function is convex, includes the deceptively simple looking, but nontrivial problem of computing the minimum of the arithmetic-geometric mean ratio of nonnegative numbers, over an arbitrary subspace of the Euclidean space. It also includes 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 significant theory of self-concordance, many of the results proved in this paper, including the scaling dualities and bounds, are indispensable ingredient in establishing the polynomial-time solvability of the scaling problems.

Before describing the precise and general definitions of HP, SP, HSP, and ASP, in § 2 we review history, examples, and significance of these four problems when defined over the nonnegative orthant. In doing so, we will familiarize the reader with the more formal definitions to be described in § 3. Once the preliminary mathematical results are presented in § 3, at the end of that section we will describe a precise summary of the results of subsequent sections. The intricate relationship between the results are also described in a schematic form. Although this is a long paper, the results are quite related and are presented in a logical order. Once the reader is convinced of this, the preferred order for detailed examination can be decided by his/her choice. Since no proofs will be given in § 2, in principle the reader may begin with § 3. However, § 2 will provide an introductory overview that gives the flavor of some of the main results.

2. History, significance, and examples of HP, SP, HSP, ASP, and scaling dualities over the nonnegative orthant. Homogeneous programming problems over the nonnegative orthant, $K = \{x \in \Re^n : x \geq 0\}$, form a fundamental subclass of the class of problems to be considered in this paper. Let $\phi(x)$ be a real-valued function, twice continuously differentiable, and homogeneous of degree $p > 0$. Let

$W = \{x \in \Re^n : Ax = 0\}$, where A is an $m \times n$ matrix of rank m. Assume $x^0 \in W \cap K^\circ$ is available, where K° is the interior of K. Let $F(x) = -\sum_{i=1}^n \ln x_i$. Let $S = \{x : \|x\| = 1\}$. HP is the problem of testing if $\mu \leq 0$, where

$$\mu = \min\{\phi(x) : x \in W \cap K \cap S\}. \tag{2.1}$$

Without loss of generality we may assume that $\mu \leq 1$. If $\phi(x^0) \leq 0$, then the problem is trivially solvable. Thus, we may assume that $\phi(x^0) > 0$. Then, from the intermediate value theorem, and the convexity of the feasible domain, HP reduces to the problem of computing a nontrivial root of ϕ over $W \cap K$, or proving that such root does not exist. Define the *logarithmic potential*, and the *homogeneous potential* as

$$\psi(x) = \phi(x) + F(x), \quad X(x) = \frac{\phi(x)}{(\Pi_{i=1}^n x_i)^{p/n}}, \tag{2.2}$$

respectively. Let

$$\psi^* = \inf\{\psi(x) : x \in W \cap K^\circ\}, \quad X^* = \inf\{X(x) : x \in W \cap K^\circ\}. \tag{2.3}$$

Scaling problem SP is to determine if $\psi^* = -\infty$, and if ψ^* is finite, to compute its value, together with a corresponding minimizer d^*. Homogeneous scaling problem HSP is to determine if $X^* \leq 0$, and if $X^* > 0$ is finite, to compute its value, together with a corresponding minimizer $\bar{d}^*$. If $\phi(x)$ is convex, then $\psi(x)$ is strictly convex, thus SP is a convex programming problem. However, HSP is a nonconvex optimization, even if ϕ is convex. Note that X is constant along rays, i.e., $X(\alpha x) = X(x)$, for all $\alpha > 0$. This implies the equivalence of the optimization of X over $W \cap K^\circ$, and its optimization over $W \cap K^\circ \cap S$. Thus, we may assume that if $\bar{d}^*$ exists, it is a vector of unit norm. The algebraic scaling problem (ASP) is to test the solvability of the scaling equation (SE)

$$P_d \nabla \phi_d(e) = e, \tag{2.4}$$

where

$$d \in W \cap K^\circ, \quad e = (1, \ldots, 1)^T, \quad D = \mathrm{diag}(d), \quad \phi_d(x) = \phi(Dx), \tag{2.5}$$

and $P_d = I - DA^T(AD^2A^T)^{-1}AD$, the orthogonal projection operator onto the subspace $W_d = D^{-1}W$. If $W = \Re^n$, then $P_d = I$.

Given $\epsilon \in (0, 1]$, ϵ-HP, ϵ-SP, ϵ-HSP, and ϵ-ASP can also be defined. These four problems and their approximation versions, defined below, are intimately related via scaling dualities. ϵ-HP is to compute $d \in W \cap K \cap S$ such that $\phi(d) \leq \epsilon$, or proving that such point does not exist. ϵ-ASP is to test the solvability of $\|P_d \nabla \phi_d(e) - e\| \leq \epsilon$. ϵ-SP is to compute, if it exists, a point $d \in W \cap K^\circ$ such that $\psi(d) - \psi^* \leq \epsilon$. ϵ-HSP is to compute, if it exists, a point $d \in W \cap K^\circ$ such that $X(d)/X^* \leq exp(\epsilon)$.

2.1. Examples. Perhaps the most familiar of the above four conic problems is the special case of HSP, where $\phi(x) = e^T x/n$, and $W = \Re^n$, i.e., the problem of computing the minimum of $X(x) = e^T x/n(\Pi_{i=1}^n x_i)^{1/n}$ over K°. From the classic arithmetic-geometric mean inequality, it follows that $X^* = X(e) = 1$. But how do we generalize this inequality when W is a subspace that does not contain e? In other words, what is the new value of X^*, or the corresponding $\bar{d}^*$? More generally, given an arbitrary vector $c \in \Re^n$, and a subspace W, how do we compute the corresponding X^*, or $\bar{d}^*$? These are nontrivial questions. We shall provide some partial answer in this section, and complete answer in the remaining of the paper. Indeed one can view Karmarkar's algorithm as an algorithm that tests the solvability of HSP for the case where $\phi(x) = c^T x$, but in the weaker sense that it only provides a yes or no answer. This point will be made more precise.

Next, we consider some special cases of HP. Let Q be an $n \times n$ symmetric matrix. We shall refer to the HP where $\phi(x) = \frac{1}{2}x^T Qx$, and $W = \Re^n$, as *Gordan's HP*, since it is very much related to Gordan's theorem. According to this theorem, one of the first known dualities on linear inequalities (see Dantzig [3], and Schrijver [37]): given any $m \times n$ real matrix H, either there exists $x \geq 0$, $x \neq 0$ such that $Hx = 0$, or there exists y such that $H^T y > 0$. Geometrically, either the convex-hull of the columns of H contains the origin, or there exists a hyperplane (induced by y) separating these points from the origin. On the one hand, when Q is indefinite, Gordan's HP is NP-complete, see [13]. On the other hand, linear programming over the rational or algebraic numbers can be formulated as Gordan's HP with a positive semidefinite Q (see [20]). This can be done as follows. Firstly, through the use of linear programming dualities, equivalently, the application of Farkas' lemma, it follows that such linear programming problem is equivalent to the geometric problem of determining if the convex-hull of a set of n points in $\Re^m$ contains the origin. Letting H denote the $m \times n$ matrix of the points, this geometric problem is to decide if $V = \{x : Hx = 0, x \geq 0, x \neq 0\}$ is nonempty. Let $Q = H^T H$. Since Q is positive semidefinite, the convex-hull problem is trivially equivalent to the corresponding Gordan's HP.

Karmarkar's canonical linear programming, [22], is the special case of HP where $\phi(x) = c^T x$. 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 it ignores the fundamental dualities that relate HP to SP, HSP, or ASP. Why should one assume that the convex-hull of a given set of points contains the origin, when this is precisely the problem being tested? Perhaps this assumption stems from the fact that in Karmarkar's first version of the projective algorithm, [21], the minimum objective value is assumed to be zero. However, in the version appearing in [22], the assumption is removed. Karmarkar's HP can be formulated as Gordan's HP, simply by taking $\phi(x) = (c^T x)^2 + x^T A^T A x$. Conversely, the convex-hull problem can be formulated as Karmarkar's HP by letting $\phi(\alpha, x) = \alpha$, $W = \{(x, \alpha) : Hx - \alpha He = 0\}$, $e = (1, \ldots, 1)^T \in \Re^n$, and K the nonnegative orthant in $\Re^{n+1}$. In [16], we described a variant of Karmarkar's algorithm with simple analysis for solving the Karmarkar canonical LP. This algorithm either computes $d \in W \cap K^\circ$ such that $P_d Dc > 0$, or it computes a nontrivial zero of $c^T x$ over $W \cap K$. When this algorithm is applied to the HP formulation of the convex-hull problem, the condition $P_d Dc > 0$, in fact gives rise to a separating hyperplane, and the algorithm can be viewed as an algorithmic proof of Gordan's theorem.

In our next example we consider SP and ASP corresponding to Gordan's HP, i.e., $\phi(x) = \frac{1}{2} x^T Q x$, $W = \Re^n$. If the corresponding ψ^* is finite, attained at d^*, optimality condition implies that $\nabla \psi(d^*) = 0$. This equation is trivially seen to be equivalent to the diagonal scaling equation $D^* Q D^* e = e$. The solvability of this equation was of interest prior to the interior-point research which began with Karmarkar's algorithm. However, the problem was primarily studied in the context of nonnegative matrices, see e.g. Sinkhorn [38], Brualdi, Prater and Schneider [2], Marshall and Olkin [29], and Kalantari [17]. It is interesting and surprising that the earlier application and existence theorems on the diagonal matrix scaling equation preceded the discovery of the fact that the vector $d^* = D^* e$ is a stationary point of the corresponding logarithmic potential function. The polynomial-time solvability of ASP for positive semidefinite matrices was established in [23], and for nonnegative matrices in [19]. For positive semidefinite matrix scaling over the algebraic numbers, see [20], and for nonnegative matrix scaling, also see [18]. When Q is an arbitrary symmetric matrix, the corresponding ASP is NP-hard (see [24]), and the corresponding HP is NP-complete (see [13]).

Aside from linear programming, many other linearly constrained convex programming problems can be formulated as an HP over the nonnegative cone. For instance, using ordinary dualities, see e.g. Mangasarian [28], linearly constrained convex quadratic programming can be reduced to the problem of testing if a convex quadratic function has a zero over a polytope. This problem in turn can be reduced to a convex HP over $K = \Re^n_+$ with $\phi(x) = \frac{1}{2} x^T Q x / e^T x + c^T x$, where Q is a symmetric positive semidefinite matrix, and $c \in \Re^n$. We shall discuss the advantages of this formulation of linearly constrained convex quadratic programming in § 11. To understand this reduction to an HP, consider the more general problem of determining if a real-valued convex function $g(x)$ has a zero over a polytope $P = \{x \in \Re^n : Bx \leq b\}$, where it is assumed that we are given $x^0 \in P$, satisfying $g(x^0) \geq 0$. Given a bound on the norm of potential zeros (as in the case of linear or convex quadratic programming), without loss of generality we may assume $P = \{x \in \Re^n : x \in W \cap K : e^T x = 1\}$, where $W = \{x : Ax = 0\}$, A is an $m \times n$ matrix of full-rank, and K the nonnegative orthant. Finally, the problem of testing if $g(x)$ has a zero over P, is equivalent to an HP with $\phi(x) = (e^T x) g(x / e^T x)$, a function that is easily seen to be convex and homogeneous of degree $p = 1$. If x^0 is not strictly positive, a strictly positive point can be produced as follows. Let I (possibly empty) be the index set of zero components of x^0. If a_i denotes the i-th column of A, for each $i \in I$, we introduce a new variable ϵ_i, and the corresponding constraint column $-a_i$. With the new homogeneous function $\phi(x)+\sum_{i\in I}\epsilon_i$, the corresponding HP will have a strictly positive point.

2.2. Scaling dualities. Scaling dualities are theorems of the alternative that relate exact and ϵ-approximate solvability of the four problems HP, SP, HSP, and ASP. We shall summarize these scaling dualities over the nonnegative orthant. These dualities have been established in the papers Kalantari [12], [13], [14], and [17].

The connection between linear programming and matrix scaling was established in Kalantari [13], via a scaling duality. In a subsequent technical report, Kalantari [14], whose more general version appeared as Kalantari [17], it was proved that the stationary points of ψ and X can be obtained from each other by scalar multiplication. In particular, under the assumption of convexity of ϕ, the minimizer d^* of ψ, if it exists, is also the minimizer $\bar{d}^*$ of X. Hence, the minimization of the nonconvex Karmarkar potential function, i.e., $\log X(x)$, is equivalent to the minimization of the corresponding convex logarithmic potential, $\psi(x)$. To the best of our knowledge the results in [14], and [17] established for the first time the precise relationship between SP and HSP, and for general ϕ, although for linear ϕ a resemblance of Karmarkar potential function minimization and the logarithmic potential minimization was claimed in the paper of Gill, Murray, Saunders, Tomlin, and Wright, [7].

We now list four distinct scaling dualities over the nonnegative cone, called the *Weak Scaling Duality*, the *Scaling Duality Theorem*, the *Scaling Separation Duality*, and the *Uniform Scaling Duality*.

The Weak Scaling Duality:

$$\mu \leq 0, \quad \text{or} \quad P_d \nabla \phi_d(e) = e, \quad \text{for some} \quad d \in W \cap K^\circ. \tag{2.6}$$

We shall refer to the equation $P_d \nabla \phi_d(e) = e$, as the *scaling equation* (SE). If $p \neq 1$, then

$$P_d \nabla \phi_d(e) = e \quad \Longleftrightarrow \quad P_d D^T \nabla^2 \phi(d) De = \frac{1}{p-1} e. \tag{2.7}$$

The Scaling Duality Theorem: assume that ϕ is convex. Then,

$$\mu > 0 \iff P_d \nabla \phi_d(e) = e, \quad \text{for some} \quad d \in W \cap K^\circ. \tag{2.8}$$

Moreover, d is unique and it coincides with d^*.

To see the significance of these two dualities alone, let us return to the arithmetic-geometric mean inequality. We can prove this classic inequality as follows: consider an HSP where $c = e/n$, and $W = \Re^n$. From the Scaling Duality Theorem, d^* exists. Since d^* satisfies the corresponding scaling equation, it follows that $d^* = ne$. The strength of the Scaling Duality Theorem is not that it gives yet another proof of this classic inequality, rather its generality. In the case of minimization of the arithmetic-geometric mean ratio over a proper subspace, W, not containing e, again d^* must satisfy the corresponding scaling equation. However, in this more general case the approximation of the new minimizer, d^*, or the new minimum ratio, $X^* = X(d^*)$, are nontrivial problems whose solution requires other scaling dualities, results that in particular include new bounds, as well as the self-concordance machinery of Nesterov and Nemirovskii. We will show that if ϕ is convex, given any $x \in W \cap K^\circ$, we have

$$\ln \left(\frac{X(x)}{X^*} \right)^{\frac{n}{p}} \leq \psi(x) - \psi^*. \tag{2.9}$$

In particular, exact and approximate minimization of ψ and X are identical. More detailed algorithmic significance of this relationship will be considered in the next section. Two other scaling dualities are:

The Scaling Separation Duality: assume that ϕ is convex. Then,

$$\mu > 0 \iff P_d \nabla \phi_d(e) > 0, \quad \text{for some} \quad d \in W \cap K^\circ. \tag{2.10}$$

The Uniform Scaling Duality: assume that ϕ is convex. Then,

$$\mu \leq 0 \iff \|P_d \nabla \phi_d(e) - e\| \geq 1, \quad \forall d \in W \cap K^\circ. \tag{2.11}$$

The Uniform Scaling Duality is a consequence of the Scaling Separation Duality, and the Scaling Duality Theorem. The fact that $\mu \leq 0$ implies the uniform lower bound of one on $||P_d \nabla \phi_d(e) - e||$ follows from the fact that if $z \in \Re^n$ is such that $||z - e|| < 1$, then in particular, $|z_i - 1| < 1$. But this implies $z > 0$.

In particular, if $\phi(x) = \frac{1}{2}x^T Qx$, Q a symmetric positive definite matrix, and $W = \Re^n$, then the Scaling Duality Theorem (which reduces to a matrix scaling duality), and the Scaling Separation Duality (which reduces to Gordan's theorem), and the Uniform Scaling Duality are, respectively

$$\exists x \geq 0,\ x \neq 0,\ Qx = 0,\ \text{ or } \exists d > 0,\ DQDe = e,\ \text{ not both.} \tag{2.12}$$

$$\exists x \geq 0,\ x \neq 0,\ Qx = 0,\ \text{ or } \exists d > 0, DQDe > 0,\ \text{ not both.} \tag{2.13}$$

$$\exists x \geq 0,\ x \neq 0,\ Qx = 0,\ \text{ or } \exists d > 0, ||DQDe - e|| < 1,\ \text{ not both.} \tag{2.14}$$

The Uniform Scaling Duality is a significant duality with respect to a path-following algorithm, applicable to HP, SP, HSP, or ASP, assuming that ϕ is convex, and $p > 1$. This algorithm will be sketched in the next section. The Scaling Separation Duality in (2.13) is in fact equivalent to Gordan's separation theorem.

In the case of Gordan's HP where Q is positive semidefinite, Gordan's theorem implies that if $\mu = 0$, then $||DQDe - e|| \geq 1$, i.e., a weaker version of the Uniform Scaling Duality, given as (2.14). In the final section of this paper (§ 11.3) we will prove this implication of Gordan's theorem and will show that even this weaker version of the Uniform Scaling Duality is a main ingredient behind a very simple polynomial-time path-following algorithm for solving Gordan's HP (hence linear programming), or the corresponding ASP (positive semidefinite matrix scaling). In the process we will derive the linear programming/matrix scaling algorithm described in [23], but from a fresh point of view, even simplifying the previous version, while also exhibiting the dependence of the algorithm on Gordan's theorem, a dependence that is transparent in [23]. As a byproduct of this algorithm we will prove the Scaling Duality Theorem for the positive semidefinite matrix, i.e., (2.12). Ironically, in the presence of a proper subspace $W = \{x : Ax = 0\}$, the very simple analysis of § 11.3 does appear to be extendible, and the proof of the polynomial-time solvability of the corresponding HP and scaling problems, although is based on a simple algorithm, essentially requires all the main results, including scaling dualities, bounds, as well as theory of self-concordance.

The extension of these four scaling dualities to more general cones will demand many new results. In particular, this generalization relies on another important and new duality, to be referred to as the *Conic Convex Programming Duality.* For the class of HP's considered in this section the statement of this duality is as follows: let $g(x)$ be any convex function defined over $W \cap K$, and $F(x) = -\sum_{i=1}^{n} \ln x_i$. Then, either $g(x)$ has a recession direction in $W \cap K$, or the infimum of $g(x) + F(x)$ is attained. This duality in fact implies the Scaling Duality Theorem. The relevance of this duality theorem is in proving the unproven cases of the Scaling Duality Theorem with respect to more general cones, as well as implication of other important results.

2.3. Algorithms based on the scaling dualities. Here we will describe three algorithms for solving the four conic problems HP, SP, HSP, or ASP, over the nonnegative orthant. The first algorithm is a generalization of Karmarkar's projective algorithm, and assumes only the convexity of ϕ. The next two algorithms consist of a new potential-reduction algorithm, and a new path-following algorithm, applicable when ϕ is β-compatible with F (see Definition 3.3). All three algorithms make use of the scaling dualities, and the latter two, to be analyzed in this paper, make use of some results from Nesterov and Nemirovskii's theory of self-concordance (summarized as Theorem 3.23).

2.3.1. A projective algorithm. We first describe a projective algorithm for solving any convex HP. This algorithm, given in [15], is a generalization of Karmarkar's algorithm. Its iterative step is as follows: given any $d \in W \cap K^\circ$, where $e^T d = 1$, we either determine that $\mu > 0$, or compute $d' \in W \cap K^\circ$, satisfying $e^T d' = 1$, and $X(d') \leq \delta X(d)$, where $\delta \in (0, 1)$ is a fixed constant depending only on n. In particular, the above implies $\mu > 0$ if and only if $X^* > 0$. The iterative step of the projective algorithm consists of two minimizations. First we compute the minimum of $\phi_d(x)$ over the intersection of $W_d \cap K$, the simplex $\{x : e^T x = 1, x \geq 0\}$, and the smallest circumscribing ball containing the simplex, centered at e. If the minimum value is positive, the algorithm halts and $\mu > 0$. Otherwise, an analogous minimization is carried out, replacing the smallest circumscribing ball with the largest inscribed ball. The latter optimization leads to the point d' stated above. In view of the scaling dualities the algorithm is also capable of solving a *decision* version of SP, HSP, or ASP. For a description of this algorithm, as applied to Karmarkar's canonical LP, and its general case of arbitrary convex HP over the nonnegative orthant, see Kalantari [12] and [15], respectively.

2.3.2. A potential-reduction algorithm. The projective algorithm, stated above, tests the solvability of HP, but only a decision version of the remaining three problems, SP, HSP, or ASP. This projective algorithm can be viewed as a *homogeneous potential-reduction* algorithm. When ϕ is β-compatible with F, another potential-reduction algorithm can be described whose iterative step replaces the minimization of $\phi_d(x)$ with a Newton step applied directly to ψ. More specifically, given an interior-point d, the algorithm replaces d with $d' = d + \sigma(d)y$, where y satisfies $P\nabla^2\psi(d)y = -P\nabla\psi(d)$, P is the orthogonal projection operator onto W, and $\sigma(d)$ is an appropriately selected step size (see Theorem 3.23). Using our scaling dualities, other ingredients that include bounds, the relationship between ψ and X, and results from Nesterov and Nemirovskii's self-concordance theory (see Theorem 3.23), we will show that all four problems are solvable in polynomial-time. More specifically, set $\sigma = exp[\frac{p}{n}\psi(x^0) - 1 + \ln\frac{n}{p}]$, and $q = \sup\{||\nabla^2\phi(d)|| : d \in W \cap K^\circ, ||d|| = 1\}$. We will prove that if $\mu \leq 0$, then the number of iterations of this potential-reduction algorithm to solve ϵ-HP is $O(n \ln \frac{\sigma}{\epsilon})$. If $\mu > 0$, then the number of iterations to solve ϵ-SP or ϵ-HSP is $O(n \ln \frac{\sigma}{\mu} + \ln\ln\frac{1}{\epsilon})$. Also, the number of iterations to solve ϵ-ASP is $O(n \ln \frac{\sigma}{\mu} + \ln\ln\frac{q}{\epsilon})$. In particular, the minimum of the arithmetic-geometric mean ratio over an arbitrary subspace can be computed in polynomial-time. More generally, we can solve linear or convex quadratic programming and their corresponding SP, HSP, or ASP in polynomial-time. The fact that quadratic programming can be solved via this simple potential-reduction algorithm is in particular the consequence of a new result proved in this paper: the function $\phi(x) = \frac{1}{2}x^T Qx/e^T x + c^T x$, although is not self-concordant, it is $\sqrt{3}$-compatible with $F(x)$ (see § 11.2). Assuming that ϕ is β-compatible, this potential-reduction algorithm is simpler and more powerful than Karmarkar's algorithm, or its generalization for arbitrary convex ϕ (described in § 2.3.1). We will also see that the analogue of this algorithm for solving the HP formulation of conic LP is simpler than Nesterov and Nemirovskii's corresponding algorithm (see Chapter 4 of [31]). Moreover, the algorithm is more powerful than Nesterov and Nemirovskii's algorithm, since additionally it can also solve any of the other three problems, SP, HSP, and ASP, in polynomial-time. In particular, for linearly constrained quadratic programming, in view of our β-compatibility proof, and the scaling dualities, this potential-reduction algorithm is also simpler and more powerful than Ye and Tse's generalization of Karmarkar's algorithm, [39].

2.3.3. A path-following algorithm. Suppose that ϕ is convex and $p > 1$. Let $u = -P\nabla\psi(x^0)$, where x^0 is a given point in $W \cap K^\circ$. For each $t \in (0, \infty)$, define $f^{(t)}(x) = t\phi(x) + tu^T x + F(x)$. Using our Conic Convex Programming Duality (Theorem 5.3), we will prove that for any t, $d_t^* = \text{argmin}\{f^{(t)}(x) : x \in W \cap K^\circ\}$ exists. Then, using the Uniform Scaling Duality together with properties of homogeneous functions, we will prove that if $\mu \leq 0$, then $\phi(d_t^*/||d_t^*||) = O(t^{p-1})$, and if $\mu > 0$, then as t approaches 0, $\hat{d}_t^* = t^{1/p}d_t^*$ converges to d^*. Moreover, instead of computing d_t^* we can compute any approximate minimizer of $f^{(t)}$ such that $||P_d\nabla f_d^{(t)}(e)||$ is sufficiently small. In particular, if ϕ is β-compatible with F, this gives a conceptually simple path-following algorithm for computing ϵ-approximate versions of HP, SP, HSP, and ASP (see § 3.6). The path-following algorithm consists of two phases. In Phase I, having selected an appropriate $t_* \in (0, 1)$, to be determined via an important theorem to be proved in this paper, called the Path-Following Theorem, starting with $t = 1$, we repeatedly apply a path-following step, until we have obtained a point within the region of quadratic convergence of Newton decrement of $f^{(t_*)}$ (see Theorem 3.23). In Phase II, starting with the terminal interior point of Phase I, we repeatedly apply Newton's method to $f^{(t_*)}$, until an interior point d is computed such that $||P_d\nabla f_d^{(t_*)}(e)||$ is sufficiently small. For a precise description of the algorithm, see § 3.6. Given t_*, the number of steps of Phase I can be estimated via Nesterov and Nemirovskii's theory (Theorem 3.23). However, in order to estimate the number of Newton iterations of Phase II, we need to relate the Newton decrement to the norm of the scaled projected gradient. The derivation of such relationship is nontrivial and requires the development of new bounds. The path-following algorithm has the following property. Let $q = \sup\{||\nabla^2\phi(d)|| : d \in W \cap K^\circ, ||d|| = 1\}$. The number of Newton iterations to solve ϵ-HP, the number of Newton iterations to solve ϵ-ASP, and those to solve ϵ-SP (or ϵ-HSP) are, respectively

$$O(\sqrt{n}\ln\frac{n||u||}{\epsilon} + \ln\ln q), \quad O(\sqrt{n}\ln\frac{n||u||}{\mu\epsilon} + \ln\ln q), \quad O(\sqrt{n}\ln\frac{n||u||}{\mu} + \ln\ln\frac{q}{\epsilon}).$$

In order to apply the path-following algorithm to linear programming, we formulate linear programming as Gordan's HP. Equivalently, we can apply the path-following algorithm to Karmarkar's canonical LP, where we simply replace the linear objective c^Tx with $\phi(x) = (c^Tx)^2$. Using this algorithm, we can also compute the minimum ratio of the arithmetic-geometric mean over an arbitrary subspace. More generally, given any linear ϕ, we can solve the corresponding SP, HSP, or ASP. Moreover, in polynomial-time, we can solve the diagonal matrix scaling equation $P_dDQDP_de = e$, for any given positive semidefinite matrix Q, assuming that $\mu > 0$. This problem can be viewed as diagonal matrix scaling in the presence of a proper subspace W. In the special case where $\phi(x) = \frac{1}{2}x^TQx$, and $W = \Re^n$, the corresponding path-following algorithm is in fact quite trivial, requiring no results from the theory of self-concordance, and is based on a weaker version the Uniform Scaling Duality, implied by Gordan's theorem (see § 11.3).

We mention that path-following algorithms for Karmarkar's canonical LP have been given before, see e.g. Goldfarb and Shaw [8]. However, the proposed path-following algorithm of this paper, although using many results, including the theory of self-concordance, is conceptually simpler than these algorithms. 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 $p > 1$. More importantly, the results are generalizable to the case of the four conic problems over other cones. We also would like to mention that although the path-following algorithm considered in this paper 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 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 (see Fiacco and McCormick [5], also Gill et all [7]). Thus, even for LP the proposed path-following algorithm of this paper is for instance different than that of Gonzaga's polynomial-time path-following algorithm for LP over bounded domains, or other path-following methods, see [9], including those based on primal-dual formulations, see also Nesterov and Nemirovskii [31]. 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 3.23). However, in general the proof of 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 paper, as opposed to the mere application of general results from convex programming, or those implied by the theory of self-concordance.

2.4. Generalizations. Given a finite dimensional space E, we define HP to be the problem of testing if a given smooth homogeneous function $\phi(x)$ of degree $p > 0$, has a nontrivial root over a pointed closed convex cone K (having nonempty interior), and its intersection with a subspace W. Such a cone admits a θ-logarithmically homogeneous barrier $F(x)$ (see Definition 3.2) which plays the role of the logarithmic barrier of $K = \Re^n_+$. Any convex programming problem over a closed convex set can be formulated as a conic linear programming: minimization of a linear functional $\langle c, x\rangle$ over a pointed closed convex cone, and its intersection with an affine subspace. By applying a duality for conic LP, developed by Nesterov and Nemirovskii (see [31], Chapter 4), the conic LP formulation of convex programs in turn can be reformulated as a special case of HP with a linear objective.

For instance, semidefinite programming can be formulated as the following HP with a linear homogeneous objective: given a symmetric matrix c, test the existence of a nontrivial zero of $\phi(x) = \langle c, x\rangle = tr(cx)$ over S^+_n, the cone of positive semidefinite symmetric matrices, and its intersection with a subspace W of $E = S_n$, the Hilbert space of $n \times n$ real symmetric matrices. Semidefinite programming finds applications in combinatorial optimization, as well as eigenvalue problems (see e.g. [26], [33], [1], [40]), and is a special case of the general class of self-concordant programming, defined and studied in [31].

It is natural to ask if the scaling dualities proved for $K = \Re^n_+$ extend to the HP formulation of semidefinite programming, or more generally to the HP formulation of self-concordant programming, or even to a general HP. For instance, is there an analogue of the diagonal matrix scaling equation, $DQDe = e$, or Gordan's separation theorem over S^+_n? If so, what are they? Can we test their solvability efficiently? Can we compute the minimum of the trace-determinant ratio, $tr(x)/det(x)^{1/n}$, over the intersection of S^+_n and a given subspace W of S_n? If so, can this be done in polynomial-time? We will prove that the answer to these questions is affirmative. Moreover, SP, HSP, and ASP, as well as several corresponding scaling dualities can be stated, not only with respect to the HP formulation of semidefinite programming, but for any HP with a smooth homogeneous ϕ defined over a general pointed closed convex cone K, and its possible intersection with a subspace W. In particular, if in a given HP, ϕ does not possess a nontrivial zero over $W \cap K$, then the corresponding scaling equation, $P_d \nabla \phi_d(e_d) = e_d$ (equivalently, $P_d \nabla^2 \phi_d(e_d) P_d e_d = \frac{e_d}{p-1}$, $p \neq 1$) holds, where $D = \nabla^2 F(d)^{-1/2}$, $d \in W \cap K^\circ$, $e_d = D^{-1}d$, and $\phi_d(x) = \phi(Dx)$. In case K is a symmetric cone (also known as homogeneous self-dual cone), $e_d = e$, the *center* of K. Symmetric cones include the nonnegative cone, as well as the semidefinite cone, and the second-order cone. The solvability of ASP over general cones is the solvability of the scaling equation (SE), a fundamental equation derivable from the solution of SP, while invoking properties inherited from homogeneity and those of the *operator-cone*, $T(K) = \{D \equiv \nabla^2 F(d)^{-1/2} : d \in K^\circ\}$. Each operator D induces a *scaling* of the corresponding Hessian or gradient of ϕ, and SE is solvable if and only if under this scaling there exists a fixed-point, e_d.

We prove that when ϕ is β-compatible with a θ-normal barrier F for the underlying cone K (see Definition 3.3), then ϵ-approximate versions of HP, SP, and HSP are all solvable in polynomial-time. Additionally, if the ratio $||D||/||d||$ is uniformly bounded, then ASP is also solvable in polynomial-time. In particular, the latter result extends the polynomial-time solvability of matrix scaling equation (even in the presence of a subspace W) to very general cases of SE over the nonnegative cone, or the semidefinite cone, or the second-order cone. The corresponding complexities for these problems essentially amount to replacing n with θ in the stated complexity formulas for the potential-reduction and path-following algorithms, sketched in the previous two sections.

3. Mathematical preliminaries. In this section we first formally define homogeneous programming (HP), scaling problem (SP), homogeneous scaling problem (HSP), 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 fundamental properties of homogeneous functions, and introduce the notion of operator-cone needed to define the scaling equation (SE). We give the precise statement of four distinct scaling dualities that are to be proved in subsequent sections. We summarize some fundamental results from Nesterov and Nemirovskii's theory of self-concordance in the form of a single theorem (Theorem 3.23), suited for our conic problems. This theorem will be repeatedly utilized throughout the paper. The assertion of the theorem is for convenience and in order to make it easier for the reader to follow the new results. 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 detailed summary of the precise results that will follow in the subsequent sections of the paper. This will also be done in a schematic form in order to help the reader to view and recognize the connection between different components, needed to prove the main results. This can also guide the reader in deciding the desired order for his/her detailed examination of the building blocks.

3.1. 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 $W = \{x \in E : \langle a_i, x\rangle = 0,\ i = 1, \ldots, m\}$, 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 $\{e_1, \ldots, e_n\}$, for each $i = 1, \ldots, m$, 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

$$v \geq u \iff v - u \in K \cap W, \qquad v > u \iff v - u \in K^\circ \cap W. \tag{3.1}$$

For properties of this partial order as well as background on functional analysis utilized in this paper, see the excellent books of Zeidler [42], and [43]. From (3.1), we can write

$$x \geq 0 \iff x \in K \cap W, \qquad x > 0 \iff x \in K^\circ \cap W. \tag{3.2}$$

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 (3.2).

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 $x > 0$,

$$\phi(\alpha x) = \alpha^p \phi(x). \tag{3.3}$$

For simplicity, we will assume that ϕ has a continuous extension to the boundary of K. We do not assume convexity of ϕ. However, in this paper we are primarily interested in convex cases.

DEFINITION 3.1. *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 3.2. *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

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

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

DEFINITION 3.3. *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 $\langle x, y\rangle \equiv x^TQy$. This inner product turns E into a Hilbert space. Thus we can identify $D^kF(x)$ (or $D^k\phi(x)$) with the multidimensional symmetric matrix $\nabla^kF(x) = (\partial^kF(x)/\partial x_{i_1}\cdots\partial x_{i_k})$ (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, $\nabla^kF(x)$ is defined with respect to the original inner product.

DEFINITION 3.4. *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 3.5. *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 3.6. *The homogeneous scaling problem (HSP) is to determine if $X^* \leq 0$, and if $X^* > 0$, 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 $X^* > 0$, 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 if ϕ is convex, then SP is a convex programming problem. However, HSP is a nonconvex program, even if ϕ is convex. An important property of HSP is that $X(x)$ is homogeneous of degree zero. If ϕ is convex and 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 3.7. *Assume that ϕ is convex, and F is θ-normal. The "parameterized family of logarithmic potentials" is defined as*

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

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

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

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 [31]. 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, $G_t = \{x \in G : f(x) \le t\}$ is closed; and for all $x \in G$, and $h \in E$, we have

$$D^3 f(x)[h,h,h] \le \frac{2}{\sqrt{a}}(D^2 f(x)[h,h])^{3/2}. \tag{3.4}$$

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 3.23 (Nesterov and Nemirovskii's Theorem). This theorem is an adaptation of several significant result of Nesterov and Nemirovskii. It is a fundamental theorem and is instrumental in solving HP/SP/HSP, and ASP, which is the problem of testing the solvability of the scaling equation (SE), a fundamental algebraic equation induced from homogeneity and differentiability. The scaling equation, already defined in the previous section for the case where the underlying cone is the nonnegative orthant, will be defined over more general cones. Theorem 3.23 on self-concordance does not by itself solve these four problems. Additionally, we need the development of several fundamental results. In particular, the scaling dualities, and significant bounds.

Over the nonnegative orthant, since $F(x) = -\sum_{i=1}^n \ln x_i$ is available, the fact that SP and HSP are well-defined problems is immediate. However, over more general cones, their definition requires justification. The existence of θ-logarithmically homogeneous barriers and θ-normal barriers for the cone K of HP is a consequence of some old and new results. Nesterov and Nemirovskii proved the existence of a universal strongly self-concordant barrier for any convex set with nonempty interior (see Nesterov and Nemirovskii [31], 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. On the other hand, the notion of characteristic function for cones is an old notion that was introduced by Koecher [25]:

$$C(x) = \int_{K^*} e^{-\langle x,y\rangle} dy, \quad K^* = \{x \in E : \langle x,y\rangle \ge 0, \forall y \in K\}. \tag{3.5}$$

It is straightforward to show that

$$F(x) = \log C(x), \tag{3.6}$$

is a θ-logarithmically homogeneous barrier. In fact, using Nesterov and Nemirovskii's universal barrier, Güler [11] shows that $F(x)$ is within a constant factor of the universal barrier. Thus, $F(x)$ is strongly self-concordant. The cone K is said to be *homogeneous* if given $x, y \in K^\circ$, there exists an invertible linear transformation A, such that

$$A(K) = K, \quad A(x) = y. \tag{3.7}$$

For homogeneous cones, $F(x)$ can easily be shown to be strongly self-concordant, see e.g. Güler [11] (Theorem 4.3). Thus, for homogeneous cones, one need not resort to Nesterov and Nemirovskii's theorem on the existence of universal barriers.

3.2. 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 d^* must satisfy, if it exists. We shall also obtain an analogous equation for d_t^*.

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

PROPOSITION 3.8. (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) $\nabla\phi(x)[x] \ = p\phi(x)$ (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) $\nabla F(x)[x] \ = -\theta$.
(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 [31]. However, these are actually a consequence of the first two well-known properties, as applied to the function $\Gamma(x) = exp(-F(x))$.

DEFINITION 3.9. *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:

$$x >_d 0 \quad \Longleftrightarrow \quad x \in W_d \cap K_d^\circ. \tag{3.8}$$

DEFINITION 3.10. *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., $P_d = I - D^T A^T (ADD^T A^T)^{-1}AD$. If $W = E$, then $P_d = I$.

PROPOSITION 3.11. *If SP is solvable, then SE is solvable. Moreover, if $p \neq 1$, then SE is equivalent to the equation*

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

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

$$\nabla g_d(x) = D^T\nabla g(Dx), \tag{3.9}$$

$$\nabla^2 g_d(x) = D^T\nabla^2 g(Dx)D. \tag{3.10}$$

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,

$$\nabla\psi(d) = \nabla\phi(d) + \nabla F(d) = A^T v, \tag{3.11}$$

where v is the vector of Lagrange multipliers. Let $e_d = D^{-1}d$. From property (P8) of Proposition 3.8, $\nabla F(d) = -\nabla^2 F(d)d$. Using this and by applying the operator D^T to (3.11), together with the chain rule, see (3.9), we get

$$\nabla\psi_d(e_d) = \nabla\phi_d(e_d) - e_d = D^T A^T v, \tag{3.12}$$

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

$$v = (ADD^TA^T)^{-1}AD\nabla\psi_d(e_d). \tag{3.13}$$

Substituting the above in (3.12), 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 3.8. □

DEFINITION 3.12. *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

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

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 $D^T\nabla^2F(d)D = I$. However, these problems can be defined in more generality as will be seen next.

3.3. 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. These properties will be used heavily in the subsequent sections.

DEFINITION 3.13. *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 $d \in K^\circ$, $||e_d||^2 \le N < \infty$, 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*

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

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

DEFINITION 3.14. *An operator-cone $T(K)$ is said to be bounded if there exists a fixed constant ρ such that*

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

Remark. The operator-cone 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 3.15. *Let $T_F = \{D = \nabla^2F(d)^{-1/2} : d \in K^\circ\}$. Then, T_F is an operator-cone with $M = 1$, $N = \theta$.*

Proof. From property (P7) of Proposition 3.8, we have $\nabla^2F(\alpha d) = \alpha^{-2}\nabla^2F(d)$. Thus, $\nabla^2F(\alpha d)^{1/2} = \alpha^{-1}\nabla^2F(d)^{1/2}$. But this implies the validity of condition (1) of Definition 3.13. We have $||e_d||^2 = d^TD^{-2}d =$

$d^T\nabla^2F(d)d$. But from properties (P8) and (P5) of Proposition 3.8, we have $d^T\nabla^2F(d)d = -d^T\nabla F(d) = \theta$. Hence, the validity of condition (2) of Definition 3.13. Clearly, condition (3) is valid. The validity of (4) is a consequence of the chain rule, see (3.10), 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 $T(K) = T_F$. 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 ϵ-ASP in polynomial time. Three important cones that will admit bounded operator-cones are 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) $K = SO^{n+1} = \{x = (\tau, z) \in \Re^{n+1} : \tau \geq \|z\|\}$. If $K = S_n^+$, K° is the set of positive definite matrices. This cone can be viewed as a cone in S_n, the Hilbert space of the $n \times n$ real symmetric matrices, or as a cone in $E = L(\Re^n, \Re^n)$ the set of $n \times n$ real matrices. The inner product in either case is $\langle x, y\rangle = tr(x^Ty)$. The following proposition reveals some properties of these cones. In particular, the fact that $T_F = \{D = \nabla^2F(d)^{-1/2} : d \in K^\circ\}$ is a bounded operator-cone.

PROPOSITION 3.16. *Let the pair (K, E) represent either $(\Re^n_+, \Re^n)$, or (S_n^+, S_n), or $(S_n^+, L(\Re^n, \Re^n))$, or $(SO^{n+1}, \Re^{n+1})$, respectively. Let F be respectively, $F(x) = -\ln det(x) = -\sum_{i=1}^n \ln x_i$, $F(x) = -\ln det(x)$, $F(x) = -\ln det(x)$, and $F(x) = -\frac{1}{2}\ln(\tau^2 - \|z\|^2)$. Then, $T_F = \{D = \nabla^2F(d)^{-1/2} : d \in K^\circ\}$ is a bounded operator-cone, satisfying,*

$$Dx = \begin{cases} \mathrm{diag}(d)x, \\ d^{1/2}xd^{1/2}, \\ dx, \\ \begin{pmatrix} \tau & z^T \\ z & \tau I\end{pmatrix} x; \end{cases} \qquad e = D^{-1}d = \begin{cases} (1,\ldots,1)^T, \\ I, \\ I, \\ (1,0,\ldots,0)^T; \end{cases} \qquad \rho = \begin{cases} 1, \\ 1, \\ 1, \\ \sqrt{2}; \end{cases}$$

respectively. Moreover, $K_d = K$ in the first two cases.

Proof. For $K = \Re^n_+$, one can easily show that given $d > 0$, $\nabla^2F(d) = D^{-2}$. Clearly, $\|D\| \leq \|d\|$, i.e., $\rho = 1$. Also, $D^{-1}d = e = (1,\ldots,1)^T$, and $K_d = K$. Since for $K = S_n^+$, $F_d(x) = -\ln det(d^{1/2}xd^{1/2}) = F(x) + F(d)$, we have $\nabla^2F_d(e) = \nabla^2F(e)$. But from the definition of determinant, it follows that $\nabla^2F(e) = I$. From this and the chain rule, $D\nabla^2F(d)D = \nabla^2F_d(e) = I$, i.e., $\nabla^2F(d) = D^{-2}$, the operator that maps z to $d^{-1}zd^{-1}$. To prove that $\rho = 1$, we have

$$\|Dx\| = \|d^{1/2}xd^{1/2}\| = \sqrt{tr((d^{1/2}xd^{1/2})^2} \leq tr(d^{1/2}xd^{1/2}) = tr(dx) \leq \|d\|\,\|x\|.$$

The first inequality follows from the fact that if Q is symmetric positive semidefinite, $tr(Q^2) \leq tr(Q)^2$. Thus, $\|D\| \leq \|d\|$. Also, $D^{-1}d = e = I$, and $K_d = K$, for all $d > 0$. When S_n^+ is viewed as a cone in $E = L(\Re^n, \Re^n)$, the fact that $\rho = 1$ is trivially provable. In this case, it is easy to verify that $\nabla^2F_d(e)$ is the operator that maps $x \in S_n^+$ to $d^{-1}xd$. Although in this case too $D^{-1}d = e = I$, K_d is not invariant. For the second-order cone it can easily be verified that with D as claimed, we have $D^{-1}d = e$, and $\nabla^2F(d) = D^{-2}$. We prove that $\rho = \sqrt{2}$. Since D is symmetric, to compute $\|D\|$, it suffices to compute its eigenvalues. Letting $z = (\tau', x')$, the equation $Dz = \alpha z$ gives the equations $\tau\tau' + x^Tx' = \alpha\tau'$, and $\tau'x + \tau x' = \alpha x'$. Solving for α gives $\alpha = \tau \pm \|x\|$. Thus, $\|D\| = \tau + \|x\|$, and is easily seen to be bounded above by $\sqrt{2}\,\|d\| = \sqrt{2(\tau^2 + \|x\|^2)}$. □

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

PROPOSITION 3.17. *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 $\nabla^2F(d)$ 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

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

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

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

But from property (P7) of Proposition 3.8, replacing d by αd in the above does not change the ratio. Thus, we may assume that $\|d\| = 1$. The converse trivially follows. □

In fact it can be shown that given any *symmetric* cone, T_F is a bounded operator-cone. Symmetric cones are cones that are homogeneous (see (3.7)), and additionally *self-dual*, i.e., there exists an invertible linear transformation A such that $A(K) = K^*$. It is well-known that any symmetric cone can be written as the direct sum of five distinct irreducible cones. Three distinct types are the cone of positive semidefinite symmetric, Hermitian, and Hermitian quaternion matrices, respectively. The fourth type is the second-order cone (also known as Lorentz cone), and the fifth type is a 27-dimensional cone. For results regarding cones, see Koecher [25], Rothaus [36], Vinberg [41], and Faraut and A. Koranyi [4]. Symmetric cones have been recently studied by Güler [11], and Nesterov and Todd [32], who have renamed them as *self-scaled* cones, in order to emphasize other property of these cones.

As we shall see in this paper, the boundedness property of the underlying operator-cone will play a significant algorithmic role in all the four problems HP, SP, HSP, and ASP. The study of cones with bounded operator-cones although is important and interesting, is not the goal of this paper, rather the subject of future investigation.

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

DEFINITION 3.18. *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 $\phi = \frac{1}{2}x^TQx$, Q a symmetric matrix, $W = \Re^n$, and $K = \Re^n_+$, the above definition coincides with the usual definition of copositive plusness as defined in textbooks (see e.g. Murty [30]). The above definition, 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 3.19. *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 (3.8), it follows that $x^TP\nabla\phi(x) = p\phi(x)$. Since $p > 0$, $\phi(x) = 0$. □

The following result characterizes the solvability of HP when ϕ is convex and $p > 1$.

PROPOSITION 3.20. *Suppose in a given HP the homogeneous function ϕ is defined over W, it is convex, and $p > 1$. 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. □

Note that Proposition 3.20 (originally proved in Kalantari [14] for HP over the nonnegative orthant) is not true if $p = 1$, e.g. if ϕ is linear. From this proposition and additional results we obtain theorems that generalize Gordan's theorem (see (2.13) and Theorem 7.2), and the matrix scaling duality for positive semidefinite symmetric matrices (see (2.12) and Theorem 5.7). According to our 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.

3.5. Definition of scaling dualities. The following definition classifies some important dualities that we shall prove in this paper.

DEFINITION 3.21. (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* $P_d \nabla \phi_d(e) \in K^\circ$.

3.6. 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 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 3.22. *Assume* ϕ *is* β*-compatible. Given* $d > 0$, *let* $P_2(y) = \psi(d) + \nabla\psi(d)^T y + \frac{1}{2} y^T \nabla^2 \psi(d) y$. *Then, Newton direction is defined as*

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

Newton decrement is defined as

$$\lambda(d) = (1+\beta)\sqrt{\Delta(d)}, \quad \Delta(d) = y(d)^T \nabla^2 \psi(d) y(d).$$

Let $\bar{\lambda}$ *be a number in* $[\lambda_*, 1)$, *where* $\lambda_* = 2 - \sqrt{3}$. *The Newton iterate is defined as*

$$d' = NEW(\psi, d) = d + \sigma(\lambda(d)) y(d),$$

where

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

Similarly, 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 [31], combining them into a format suitable for our conic problem. It is a theorem that will be used repeatedly in the paper. This theorem combines a theorem characterizing basic properties of Newton's method under self-concordance ([31], Theorem 2.2.2), a theorem on the main property of self-concordant families ([31], Theorem 3.1.1), and its consequence under β-compatibility ([31], Proposition 3.2.2).

THEOREM 3.23. (Nesterov and Nemirovskii's Theorem) *Let* ϕ *be* β*-compatible. Then,* $d' = NEW(\psi, d) > 0$, $d'_t = NEW(f^{(t)}, d) > 0$, *and*

$$\begin{cases} \psi(d') - \psi(d) \le -\frac{\lambda(d) - \ln(1+\lambda(d))}{(1+\beta)^2}, & \text{if } \lambda(d) > \bar{\lambda}; \\ \lambda(d') \le \frac{1}{4}(6\lambda(d) - \lambda^2(d) - 1), & \text{if } \lambda(d) \in [\lambda_*, \bar{\lambda}]; \\ \lambda(d') \le \frac{\lambda^2(d)}{(1-\lambda(d))^2}, & \text{if } \lambda(d) < \lambda_*. \end{cases} \tag{3.14}$$

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

$$\psi(d) - \psi^* \le \frac{\omega^2(\lambda(d))(1+\omega(\lambda(d)))}{2(1+\beta)^2(1-\omega(\lambda(d)))}, \quad \omega(\lambda(d)) = 1 - (1 - 3\lambda(d))^{1/3}. \tag{3.15}$$

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

$$\lambda^{(t)}(d) \le \kappa < \lambda_*. \tag{3.16}$$

Then, for any t' *satisfying*

$$|\ln \frac{t'}{t}| \left(1 + (1+\beta)\frac{\sqrt{\theta}}{\kappa}\right) \le 1 - \frac{\kappa}{(1-\kappa)^2}, \tag{3.17}$$

we have

$$\lambda^{(t')}(d) \leq \kappa. \tag{3.18}$$

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

$$t' = r_* t, \quad r_* = exp\left(\frac{-c}{(1+\beta)\sqrt{\theta}}\right), \quad c = \frac{5}{9}\left(\frac{(1+\beta)\sqrt{\theta}}{1+4(1+\beta)\sqrt{\theta}}\right) \geq \frac{1}{9}. \tag{3.19}$$

Thus, given any $t \in (0,1]$, since $\lambda_1(x^0) = 0$, the number of iterations, k_t, to obtain $d > 0$ such that $\lambda_t(d) < \lambda_*$, satisfies the inequality $r_*^{k_t} \leq t$. Equivalently, it suffices to choose

$$k_t = \lceil 9(1+\beta)\sqrt{\theta}\ln(\frac{1}{t})\rceil. \ \square \tag{3.20}$$

Remark. In particular, the central-path $\Pi = \{d_t^* : t \in (0, \infty)\}$ is well-defined.

Assuming β-compatibility, we will solve HP, SP, HSP, and ASP using the following Potential-Reduction algorithm, as well as the Path-Following algorithm, for appropriately chosen $t_* \in (0,1)$.

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

Path-Following:
Initialization. Let $t = 1$, $d = x^0$.
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.

3.7. Summary of the results. Here we summarize the main results of the paper, and also give a figure describing their intricate relationships.

In § 4, we prove the first scaling duality (see Definition 3.21), the Weak Scaling Duality (Theorem 4.2). This theorem also gives an inequality on homogeneous potentials, more general than the arithmetic-geometric mean, trace-determinant, and Hadamard inequalities. This is the only section in which ϕ may be nonconvex. Under the assumption of convexity of ϕ, in subsequent sections, the Weak Scaling Duality will be strengthened. In § 5, we prove a new duality on convex programming, the Conic Convex Programming Duality (Theorem 5.3). Using this duality, we prove the Scaling Duality Theorem (Theorem 6.3), given that F is θ-normal. 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 § 6, we study Newton's method and its convergence analysis in the context of SP. In this section we also prove the Uniform Scaling Duality (Theorem 6.3), given any operator-cone $T(K)$, or the specific operator-cone $T_F = \{D = \nabla^2 F(d)^{-1/2} : d \in K^\circ\}$, assuming that ϕ is β-compatible. In § 7, we prove the Scaling Separation Duality (Theorem 7.1) over three symmetric cones, the nonnegative cone, the semidefinite cone, and the second-order cone. This theorem will also imply that over these symmetric cones the Uniform Scaling Duality Theorem is valid, assuming only convexity of ϕ, and with a better parameter γ^* than that derived for β-compatible ϕ over general cones. In § 8, we derive a bound on the norm of the scaling vector, d^*, and the norm of parameterized scaling vectors, d_t^*. Also, given $d \in S(\bar{\lambda}_*) = \{x > 0 : \lambda(x) < \bar{\lambda}_*\}$, $\bar{\lambda}_* = \lambda_*(1+\beta)/(1+\beta+\rho)$, we bound the norm of d, and the scaled gradient projection $P_d \nabla \psi_d(e_d)$. Likewise, given $d \in S_t(\bar{\lambda}_*) = \{x > 0 : \lambda_t(x) < \bar{\lambda}_*\}$, we bound the norm of d, and $P_d \nabla f_d^{(t)}(e_d)$. In § 9, we prove a complexity theorem, the Potential-Reduction Complexity Theorem (Theorem 9.3). This theorem analyzes the application of the Potential-Reduction algorithm for solving HP/SP/HSP, when ϕ is β-compatible. Additionally, given a bounded operator-cone, the theorem gives the complexity analysis for solving ASP. The theorem employs scaling dualities, bounds, and Theorem 3.23 of Nesterov and Nemirovskii, referred in the figure as "NN's Theorem". In § 10, we prove the Path-Following Theorem (Theorem 10.3), for solving HP/SP/HSP/ASP, given a bounded operator-cone. Also, we prove a complexity theorem, the Path-Following

Complexity Theorem (Theorem 10.4), for solving β-compatible cases of HP/SP/HSP/ASP, applicable when the homogeneous degree p is larger than one. In § 11, we consider the applications of the two algorithms in linear programming, quadratic programming, and semidefinite programming, as well as in deriving algorithmic proof of some important inequalities over the three symmetric cones. In this final section, we also prove a theorem on β-compatibility. Finally, we rederive, while further simplifying the matrix scaling/linear programming algorithm of [23], using elementary and self-contained analysis. The following figure describes the intricate relationships and implications between various main results.

[Relationship between homogeneous potential X, and logarithmic potential ψ] $\Longrightarrow$
[Weak Scaling Duality] $\Longrightarrow$
[generalization of matrix scaling theorems and classic homogeneous inequalities].

[Conic Convex Programming Duality] $\Longrightarrow$
[Scaling Duality Theorem for θ-normal F] **and**
[generalization of matrix scaling/duality theorems] **and**
[existence theorem on the central-path Π of $f^{(t)}$ for θ-normal F].

[Homogeneous properties] $\Longrightarrow$ [$O((\frac{1}{\mu})^{1/p})$-bound on $||d^*||$]

[Derived bound on $||d^*||$] **and** [given a bounded operator-cone] **with**
[NN's Theorem] $\Longrightarrow$ [$O((\frac{1}{\mu})^{1/p})$-bound on $||d||$, $d \in S(\bar{\lambda}_*)$]

[A derived bound on $||P\nabla\psi(d)||$] **and** [given a bounded operator-cone] **and**
[derived bound on $||d||$, $d \in S(\bar{\lambda}_*)$] $\Longrightarrow$ [$O(\frac{\lambda(d)}{\mu})$-bound on $||P_d\nabla\psi_d(e_d)||$, $d \in S(\bar{\lambda}_*)$]

[Scaling Duality Theorem] **and** [derived bound on $||d^*||$] **and**
[assuming β-compatibility] **and** [relationship between ψ and X] **with**
[NN's Theorem] $\Longrightarrow$ [Potential-Reduction Complexity Theorem for HP/SP/HSP]

[Scaling Duality Theorem] **and** [assuming β-compatibility] **with** [NN's Theorem] $\Longrightarrow$
[Uniform Scaling Duality]

[Uniform Scaling Duality] **and** [derived bound on $||d^*||$] **and**
[assuming β-compatibility] **and** [given a bounded operator-cone] **and**
[derived bound on $||d||$, $d \in S(\bar{\lambda}_*)$] **with** [NN's Theorem] $\Longrightarrow$
[Potential-Reduction Complexity Theorem for HP/SP/HSP/ASP]

[Homogeneous properties] $\Longrightarrow$ [Scaling Separation Duality over symmetric cones] $\Longrightarrow$
[Uniform Scaling Duality over symmetric cones (with or without β-compatibility)]

[Given a bounded operator-cone] **with** [NN's Theorem] **and** [bounding results]$\Longrightarrow$
[$O\left((\frac{1}{t})^{9\rho\sqrt{\theta}}\right)$-bound on $||d_t^*||$] $\Longrightarrow$ [$O\left((\frac{1}{t})^{9\rho\sqrt{\theta}}\right)$-bound on $||d||$, $d \in S_t(\bar{\lambda}_*)$]

[A derived bound on $||P\nabla f^{(t)}(d)||$] **and** [given a bounded operator-cone] **and**
[derived bound on $||d||$, $d \in S_t(\bar{\lambda}_*)$] $\Longrightarrow$
[$O(\lambda_t(d)\frac{1}{t}^{9p\rho\sqrt{\theta}})$-bound on $||P_d\nabla f_d^{(t)}(e_d)||$, $d \in S_t(\bar{\lambda}_*)$]

[Uniform Scaling Duality] **and** [given a bounded operator-cone] **and**
[existence theorem on the central-path Π of $f^{(t)}$] $\Longrightarrow$
[Path-Following Theorem with or without β-compatibility]

[Uniform Scaling Duality] **and** [assuming β-compatibility] **and**
[given a bounded operator-cone] **and** [derived bounds on $||d^*||$, $||d_t^*||$, $||\hat{d}_*||$] **and**
[derived bounds on norm of d in $S(\bar{\lambda}_*)$, or in $S_t(\bar{\lambda}_*)$] **and**
[derived bounds on $||P_d\nabla\psi_d(e_d)||$, and $||P_d\nabla f^{(t)}(e_d)||$] **and**
[Path-Following Theorem] **with** [NN's Theorem] $\Longrightarrow$
[Path-Following Complexity Theorem for HP/SP/HSP/ASP when $p > 1$]

[Potential-Reduction and Path-Following Complexity Theorems] $\Longrightarrow$
[new algorithms for linear/quadratic/semidefinite/self-concordant programming] **and**
[proof of polynomial-time solvability of many new scaling equations (SE)] **and**

[proof of polynomial-time solvability of arithmetic-geometric, or trace-determinant ratio minimization over arbitrary subspace] **and** [many new HSP, and SP's.]

FIG 1. Schematric representation of main results, and their implications

4. Proof of the Weak Scaling Duality and an inequality. In this section we consider cases of the four conic problems, where ϕ is possibly nonconvex. In the following lemma we establish a relationship between the stationary points of ψ and those of X.

LEMMA 4.1. *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

$$\nabla X(x) = \Big(\nabla\phi(x) + \frac{p}{\theta}\phi(x)\nabla F(x)\Big) exp(\frac{p}{\theta}F(x)). \tag{4.1}$$

Since d is a stationary point of ψ, we have

$$P\nabla\psi(d) = P(\nabla\phi(d) + \nabla F(d)) = 0. \tag{4.2}$$

Multiplying the above by d^T, and applying properties (P1) and (P5) of Proposition 3.8 we get

$$\phi(d) = \frac{\theta}{p}. \tag{4.3}$$

Substituting (4.3) in (4.1), 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 3.8, 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. ☐

The above result was previously established for the case of $K = \Re^n_+$ in Kalantari [14] (Lemma 3.1 in [17]). The following theorem for ϕ a quadratic form over $K = \Re^n_+$, was established previously in Kalantari [13], and for general ϕ over this cone in [17]. We shall refer to this theorem as the Weak Scaling Duality/Homogeneous Potential Inequality Theorem. This theorem gives a nonconvex version of the Scaling Duality Theorem to be proved in the next section. The theorem also characterizes an optimality condition on the homogeneous potential function X, which is nonconvex even if ϕ is convex. In particular, the theorem gives a generalization of the arithmetic-geometric mean, trace-determinant, and Hadamard inequalities.

THEOREM 4.2. (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 $D = \nabla^2 F(d)^{-1/2}$. If $p > 1$, either $\mu \leq 0$, or $P_d\nabla^2\phi_d(e_d)P_d e_d = e_d/(p-1)$, 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 4.1, 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 4.1, 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. □

Theorem 4.2 has some important consequences. Note that in particular it implies the diagonal matrix scaling result, $DQDe = e$, for symmetric matrices of positive entries (positive matrices), or more generally, $P_d DQDP_d e = e$. We now list some others.

COROLLARY 4.3. (Generalization of the Arithmetic-Geometric Mean Inequality) *Consider the case of HP where* $K = \Re^n_+$, $\phi(x) = c^T x$, $c \in \Re^n$, *and* $F(x) = -\sum_{i=1}^n \ln x_i$. *If* $\mu > 0$, *there exists a unique* $d = (d_1, \ldots, d_n) \in W \cap K^\circ$ *such that for all* $x \in W \cap K^\circ$,

$$\frac{c^T x}{(\prod_{i=1}^n x_i)^{1/n}} \geq \frac{c^T d}{(\prod_{i=1}^n d_i)^{1/n}},$$

and

$$c^T x - \sum_{i=1}^n \ln x_i \geq c^T d - \sum_{i=1}^n \ln d_i.$$

Moreover,

$$P_d Dc = (I - DA^T (AD^2 A^T)^{-1} AD) Dc = e, \quad D = \mathrm{diag}(d), \quad e = (1, \ldots, 1)^T.$$

Proof. Apply Theorem 4.2, and note that since $\psi(x)$ is necessarily strictly convex, its stationary point is unique. □

Remark. In fact from the Scaling Duality Theorem (Theorem 5.5) it follows that $\mu > 0$ if and only if the corresponding HP is unsolvable. This also applies to the following corollary of the Weak Scaling Duality.

COROLLARY 4.4. (Generalization of the Trace-Determinant Inequality) *Consider the case of HP where* $K = S^+_n$, $\phi(x) = tr(cx)$, $c \in S^+_n$, *and* $F(x) = -\ln det(x)$. *If* $\mu > 0$, *there exists a unique* $d \in W \cap K^\circ$ *such that for all* $x \in W \cap K^\circ$,

$$\frac{tr(cx)}{(det(x))^{1/n}} \geq \frac{tr(cd)}{(det(d))^{1/n}},$$

and

$$tr(cx) - \ln det(x) \geq tr(cd) - \ln det(d).$$

Moreover,

$$P_d Dc = P_d d^{1/2} c d^{1/2} = e, \quad e = I.$$

Proof. Apply Theorem 4.2, and note that since $\psi(x)$ is necessarily strictly convex, its stationary point is unique. □

COROLLARY 4.5. (Hadamard Inequality) *For all* $x = (x_{ij}) \in S^+_n$,

$$\phi(x) = \prod_{i=1}^n \sum_{j=1}^n x_{ij}^2 \geq det(x)^2.$$

Proof. Let $\epsilon > 0$ be given. Define

$$\phi_\epsilon(x) = \phi(x) + \epsilon tr(x)^{2n}. \tag{4.4}$$

Note that $\phi_\epsilon(x)$ is homogeneous of degree $2n$. Denote the corresponding μ by μ_ϵ. Since given $x \in S_n^+$, $tr(x) = 0$ if and only if $x = 0$, it follows that $\mu_\epsilon > 0$. Let $\psi_\epsilon(x) = \phi_\epsilon(x) - \ln det(x)$. Since $\mu_\epsilon > 0$, from Theorem 4.2, there exists $d_\epsilon > 0$ such that

$$\nabla \psi_\epsilon(d_\epsilon) = 0. \tag{4.5}$$

We will prove that $d_\epsilon = \alpha I$, for some unique positive real α. From this and Theorem 4.2, it follows that

$$\frac{\phi_\epsilon(x)}{det(x)^2} \geq \frac{\phi_\epsilon(x_\epsilon)}{det(x_\epsilon)^2} = \frac{\alpha^{2n} + \epsilon(n\alpha)^{2n}}{\alpha^{2n}} = 1 + \epsilon n^{2n}. \tag{4.6}$$

Taking the limit in (4.6), as ϵ approaches zero, we get the desired inequality. To prove the desired result on d_ϵ, from the definition of ψ_ϵ, it follows that (4.5) is equivalent to the equation

$$\Delta d_\epsilon + 2n\epsilon\ tr(d_\epsilon)^{2n-1} I - d_\epsilon^{-1} = 0, \quad \Delta = \mathrm{diag}\left(\frac{\phi(d_\epsilon)}{||d_\epsilon^1||}, \cdots, \frac{\phi(d_\epsilon)}{||d_\epsilon^n||}\right), \tag{4.7}$$

where d_ϵ^i is the i-th column of d_ϵ. Multiplying (4.7) by d_ϵ, from left and right, gives two equations from which we get

$$\Delta d_\epsilon^2 = d_\epsilon^2 \Delta. \tag{4.8}$$

Since d_ϵ commutes with Δ, either Δ is a scalar multiple of I, or d_ϵ is a diagonal matrix. In either case, from this together with equation (4.7), we get the matrix equation $ad_\epsilon^2 + bd_\epsilon - cI = 0$, where a, b, and c are positive reals. Any eigenvalue of d_ϵ must satisfy the equation $az^2 + bz - c = 0$. Since this equation has a unique positive root, it follows that $d_\epsilon = \alpha I$, $\alpha > 0$. □

5. The Conic Convex Programming Duality and its applications. In this section we prove a duality for conic convex programming (Theorem 5.3). The theorem will be applied to get three different results. Firstly, we apply this theorem to prove the Scaling Duality Theorem (Theorem 5.5). Secondly, we use it to prove the existence of Π (Theorem 5.8), the central-path of the family $f^{(t)}(x)$, $t \in (0, \infty)$, independent of strong self-concordance. The latter theorem will be used within an important theorem called the Path-Following Theorem (Theorem 10.3). 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 5.9). The latter theorem will be invoked several times in subsequent sections (in particular, in § 8).

DEFINITION 5.1. *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 5.2. *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 $v \in \bar{W} \cap K$, $v \neq 0$.*

Proof. The set $K_\alpha(f)$ is closed and convex. Thus, if unbounded, it has a recession direction, see Rockafellar [35] (Theorem 8.4). 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 5.3. (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 5.2, 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

$$p(\alpha) \geq p(\alpha_0) + (\alpha - \alpha_0)p'(\alpha_0) \geq c_0 + \epsilon\alpha, \tag{5.1}$$

for some constant c_0. Thus,

$$f(x + \alpha v) \geq L(\alpha) = c_0 + \epsilon\alpha + \bar{F}(x + \alpha v). \tag{5.2}$$

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 (see (2.12), Marshall and Olkin [29], and Kalantari [13], [17]).

COROLLARY 5.4. *Let Q be an $n \times n$ positive semidefinite symmetric matrix such that $x^T Q x$ has no nontrivial zero in $\Re^n_+$. Then, given any natural number k, there exists a positive diagonal matrix D such that*

$$D^k Q D e = e.$$

Proof. Let $g(x) = \frac{1}{2} x^T Q x$, 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 5.3. □

5.1. Proof of the Scaling Duality Theorem. The goal of this section is to prove the Scaling Duality Theorem, i.e.,

THEOREM 5.5. (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 5.3, the Weak Scaling Duality (Theorem 4.2), and the following result.

LEMMA 5.6. *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 5.1. 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 [31] (page 39), we have

$$-\nabla^2 F(x)[h,h]^{1/2} \geq \nabla F(x)[h]. \tag{5.3}$$

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 [31] (page 34), for any $x, y \in K^\circ$, we have

$$\nabla F(x)[y-x] \leq 1. \tag{5.4}$$

In particular,

$$q'(\alpha) = \nabla F(x+\alpha v)[v] = \frac{1}{\alpha}\nabla F(x+\alpha v)[\alpha v] \geq \frac{-1}{\alpha}. \tag{5.5}$$

Thus, $-1/\alpha \leq q'(\alpha) < 0$. Taking the limit as α approaches infinity, we see that F satisfies the third condition of Definition 5.1. □

We now state a theorem that is a consequence of the Scaling Duality Theorem and Proposition 3.20 on copositive plusness. This theorem is the generalization of a matrix scaling duality for positive semidefinite symmetric matrices (see (2.12)).

THEOREM 5.7. (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.* □

5.2. Existence of central-path. Let $\bar{F}(x)$, and $g(x)$ be as in Theorem 5.3. 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 5.8. *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 5.3, $g(x) + \langle \bar{u}, x\rangle$ has no recession direction over K. This implies $t(g(x) + \langle \bar{u}, x\rangle)$ has no recession direction. Again by Theorem 5.3, $\bar{f}^{(t)}(x)$ must have a stationary point. □

Theorem 5.8 will be used to prove the Path-Following Theorem (Theorem 10.3). Theorem 5.8 can also be used to prove that if $g(x)$ has no recession direction, then

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

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.

5.3. Convergence of Newton iterates.

THEOREM 5.9. *Assume ϕ is β-compatible. For a given $\alpha \in [0, \frac{1}{3}]$, the sets $S(\alpha) = \{d > 0 : \lambda(d) < \alpha\}$ and $S_t(\alpha) = \{d : d > 0, \lambda_t(d) < \alpha\}$, $t \in (0, 1]$, 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

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

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

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

Proof. From Proposition 5.2, the minimizer d^* of ψ exists if and only if for each $\alpha \in \Re$, the level set $K_\alpha(\psi) = \{x \in W \cap K^\circ : \psi(x) \le \alpha\}$ is compact. From this and Theorem 3.23, see (3.15), 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 5.8, d_t^* exists. In particular, from Proposition 5.2, $K_\alpha(f^{(t)}) = \{x \in W \cap K^\circ : \psi(x) \le \alpha\}$ is compact. Now using an analogous argument as those used for ψ, the convergence of d_t^k to d_t^* can be established. ☐

6. 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 3.22) $\lambda(d)$ and $\lambda_t(d)$ (Lemma 6.2). The first upper bound will be used to prove the Uniform Scaling Duality under the assumption of β-compatibility (Theorem 6.3). The lower bounds will be used to derive bounds on the norm of points within the quadratic regions of convergence (Lemma 8.3).

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

$$\Phi_d(x) = \psi(d) + \nabla\psi(d)[x - d] + \frac{1}{2}\nabla^2\psi(d)[x - d, x - d]. \tag{6.1}$$

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

$$\frac{\Delta(d)}{2} = \psi(d) - \Phi_d(\bar{x}), \quad y = \bar{x} - d, \quad \bar{x} = d + y. \tag{6.2}$$

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

$$P\nabla^2\psi(d)y = -P\nabla\psi(d), \quad \Delta(d) = y^T\nabla^2\psi(d)y, \tag{6.3}$$

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 $c = \nabla\psi(d)$. Then, as derived in textbooks (see e.g. [27], [6]), it is easy to show that y satisfies

$$PQy = PQPy = -Pc, \quad y = -Q^{-1}c + Q^{-1}A^T(AQ^{-1}A^T)^{-1}AQ^{-1}c. \tag{6.4}$$

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

$$y = -Q^{-1/2}P_{AQ^{-1/2}}Q^{-1/2}c = -L^{-T}P_{AL^{-T}}L^{-1}c. \tag{6.5}$$

In particular, the Newton decrement satisfies

$$\Delta(d) = y^TQy = \|P_{AQ^{-1/2}}Q^{-1/2}c\|^2 = \|P_{AL^{-T}}L^{-1}c\|^2 \leq \|Q^{-1/2}c\|^2. \tag{6.6}$$

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

PROPOSITION 6.1. *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 6.2. *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

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

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 $\nabla\psi(d) + \nabla^2\psi(d)y = A^Tv$, for some vector of Lagrange multipliers, v. From this, chain rule, and the definition of z it is easy to show that

$$P_d\nabla^2\psi_d(e_d)z = -P_d\nabla\psi_d(e_d). \tag{6.7}$$

Thus,

$$\|P_d\nabla\psi_d(e_d)\| = \|P_d\nabla^2\psi_d(e_d)P_dz\| = \|Hz\|, \tag{6.8}$$

where

$$H = P_d\nabla^2\psi_d(e_d)P_d = P_d\nabla^2\phi_d(e_d)P_d + P_d\nabla^2F_d(e_d)P_d. \tag{6.9}$$

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

$$z^THz \geq z^T\nabla^2F_d(e_d)z. \tag{6.10}$$

Now from (6.10) and property (4) of operator-cones (see Definition 3.13), together with Proposition 6.1, we get

$$\|z\|^2 \leq z^THz \leq \|Hz\|^2. \tag{6.11}$$

We have

$$\Delta(d) = y^T\nabla\psi(d)y = z^T\nabla\psi_d(e_d)z = z^THz. \tag{6.12}$$

The proof now follows from (6.11), (6.12), and since $\lambda(d) = (1+\beta)\sqrt{\Delta(d)}$ (see Definition 3.22). □

THEOREM 6.3. (Uniform Scaling Duality Theorem) *Assume ϕ is β-compatible, and $T(K)$ a given operator-cone. Then, $\mu \le 0$ if and only if*

$$\forall\ d > 0, \quad \|P_d \nabla \psi_d(e_d)\| \ge \gamma^* = \frac{1}{(1+\beta)}. \tag{6.13}$$

Proof. From the Scaling Duality Theorem (Theorem 5.5), $\mu > 0$ if and only if the scaling equation is solvable. Thus, if $\mu > 0$, (6.13) is violated for some $d > 0$. Conversely, suppose that (6.13) is violated for some $d > 0$. Then, from Lemma 6.2, $\lambda(d) < 1$. From Theorem 3.23, it follows that $S(\lambda_*) = \{x > 0 : \lambda(x) < \lambda_*\}$ is nonempty. Then, from Theorem 5.9, starting from any point in this set, Newton's iterates converge to a point d satisfying the scaling equation. Hence, $\mu > 0$. □

7. The Scaling Separation Duality over three symmetric cones. In this section we consider HP's over the three special cones $\Re^n_+$, S^+_n, and SO^{n+1}, with their corresponding θ-logarithmic barriers. The aim of this section is to prove the Scaling Separation Duality. This is an important duality that is an analogue of Gordan's theorem, viewed in a diagonal scaling format, i.e., given a positive semidefinite symmetric matrix Q, either $Qx = 0$, $x \ge 0$, $x \ne 0$, or $DQDe > 0$ for some $D = \mathrm{diag}(d)$, $d > 0$, but not both. Aside from being a characterization theorem, the Scaling Separation Duality is the theorem that gives rise to the Uniform Scaling Duality for these special cones. The Uniform Scaling Duality Theorem in turn gives rise to the Path-Following Theorem (Theorem 10.3).

We assume the convexity of ϕ, but no β-compatibility is required here. The Scaling Separation Duality also has direct algorithmic applications. For instance, when solving a convex HP over $\Re^n_+$ or S^+_n, once we have obtained a point $d > 0$ such that $P_d \nabla \phi_d(e) > 0$, where e is the center of the cone, we can terminate the Potential-Reduction algorithm, or the Path-Following algorithm (see § 3.6), with the conclusion that $\mu > 0$.

THEOREM 7.1. (Scaling Separation Duality Theorem) *Assume that ϕ is convex, $K = \Re^n_+$, S^+_n, or SO^{n+1}, F the corresponding θ-logarithmically homogeneous barrier, and $T(K) = T_F = \{D = \nabla^2 F(d)^{-1/2} : d \in K^\circ\}$. Then,*

$$\mu > 0 \quad \iff \quad \exists d > 0 \ \text{ s.t. } \ P_d \nabla \phi_d(e) >_d 0.$$

If $K_d = K$ for all $d > 0$, then

$$\mu > 0 \quad \iff \quad \exists d > 0 \ \text{ s.t. } \ P_d \nabla \phi_d(e) \in K^\circ.$$

Equivalently,

$$\mu \le 0 \quad \iff \quad \forall\ d > 0, \quad \|P_d \nabla \phi_d(e) - e\| \ge \gamma^*,$$

where γ^* is the largest $\gamma > 0$ such that the ball $B_\gamma(e) = \{x \in E : \|x - e\| \le \gamma\}$ is contained in K, i.e., the Uniform Scaling Duality holds with parameter γ^*.

We now state a theorem that is a consequence of the Scaling Separation Duality Theorem and Proposition 3.20 on copositive plusness:

THEOREM 7.2. (Generalization of Gordan's Theorem) *Assume that ϕ is defined over W, it is convex, $p > 1$, $K = \Re^n_+$, S^+_n, or SO^{n+1}, F the corresponding θ-logarithmically homogeneous barrier, and $T(K) = T_F = \{D = \nabla^2 F(d)^{-1/2} : d \in K^\circ\}$. Then, either $P\nabla\phi(x) = 0$ for some $x \ge 0$, $x \ne 0$, or $P_d \nabla \phi_d(e_d) >_d 0$ for some $d > 0$. Moreover, precisely one of the two conditions hold.* □

Remark. It is easy to see that $\gamma^* = 1$ for $K = \Re^n_+$. We claim that this is also the case for $K = S^+_n$, $T(K) = \{D \in L(S_n, S_n) : Dx = d^{1/2} x d^{1/2}, d \in K^\circ\}$. To see this assume that Q is a symmetric matrix satisfying $tr((I-Q)^2) < 1$. Then, for any nonzero $x \in \Re^n$, $x^T(I-Q)x = \|x\|_2^2 - x^TQx \le \|x\|_2^2 \|Q - I\|_2 \le \|x\|_2^2 (tr((I-Q)^2))^{1/2} < \|x\|_2^2$. Thus, $x^TQx > 0$.

In order to prove Theorem 7.1 we will first give a definition, and prove several lemmas.

DEFINITION 7.3. *A convex cone C in a Hilbert space is said to be acute if whenever $x, y \in C$, we have $\langle x, y\rangle > 0$.*

Clearly, an acute cone is necessarily pointed. The following result gives a condition for hyperplanes whose cross section with an acute cone is a bounded set, i.e., hyperplanes that completely *intersect* the cone.

LEMMA 7.4. *Assume that the cone K of HP is acute. Let c be a given positive real number, and $\lambda \in K^\circ$. Consider the hyperplane*

$$H = \{x \in E : \langle \lambda, x \rangle = c\}.$$

Then, given any $z \geq 0$, $z \neq 0$, we have $\langle \lambda, \alpha z \rangle = c$, for some $\alpha > 0$.

Proof. The proof is obvious if $z > 0$. Assume $z \geq 0$, $z \neq 0$. Let $\{z_k > 0\}_{k=0}^{\infty}$ be a sequence of points converging to z, satisfying $||z_k|| = ||z||$. We have $\langle \lambda, \alpha_k z_k \rangle = c$, for some $\alpha_k > 0$. In particular,

$$0 < \frac{c}{\alpha_k} = \langle \lambda, z_k \rangle \leq ||\lambda|| \, ||z_k|| < \infty. \tag{7.1}$$

Thus, there exists $\delta > 0$ such that $\alpha_k > \delta$, for all k. Using the continuity of inner product, α_k converges to a number $\alpha > 0$. □

The proof of Theorem 7.1 will be broken up into several lemmas. Define

$$\mu_d = \min\{\phi_d(x) : x \in W_d \cap K_d^\circ, \ ||x|| = 1\}. \tag{7.2}$$

LEMMA 7.5. *If $\mu > 0$, then there exists $d > 0$ such that $P_d \nabla \phi_d(e) \in K_d^\circ$.*

Proof. From the Weak Scaling Duality (Theorem 4.2), $\mu > 0$ implies there exists $d > 0$ such that $P_d \nabla \phi_d(e) = e$. Since $e \in K_d^\circ$ for all d, the proof is immediate. □

LEMMA 7.6. *$\mu_d \leq 0$ if and only if $\mu \leq 0$.*

Proof. The proof follows from the continuity property of the operator-cone and homogeneity of ϕ. □

LEMMA 7.7. *For each $d > 0$, K_d is acute.*

Proof. We first claim the following holds (in fact over any symmetric cone):

$$\nabla^2 F(d) x \in K^\circ, \quad \forall \ x, d \in K^\circ. \tag{7.3}$$

It can trivially be proved for the case of $K = \Re_+^n$, and S_n^+, and is also easily verifiable for SO^{n+1} (see e.g. Nesterov and Todd [32]). Let $x, y, d \in K^\circ$ be given. Let $D = \nabla^2 F(d)^{-1/2}$. From the chain rule, we have $\nabla^2 F(d) = D^{-T} \nabla^2 F_d(e) D^{-1}$. From this and since $\nabla^2 F_d(e)$ is the identity operator (see Proposition 3.15), we get $y^T \nabla^2 F(d) x = y^T D^{-T} D^{-1} x = (D^{-1} y)^T (D^{-1} x)$. The proof is now immediate since K is acute. □

LEMMA 7.8. *Suppose there exists $d > 0$ such that $\lambda_d = P_d \nabla \phi_d(e) \in K_d^\circ$, then $\mu > 0$.*

Proof. Since $e \in K_d$ and K_d is acute, the hyperplane $H = \{x : \langle \lambda_d, x \rangle = \langle \lambda_d, e \rangle\}$ intersects K_d in the sense of Lemma 7.4. Define

$$\hat{\mu}_d = \min\{\phi_d(x) : x \in W_d \cap K_d^\circ, \quad \langle \lambda_d, x \rangle = \langle \lambda_d, e \rangle\}. \tag{7.4}$$

We claim that the center e is the minimizer of the above optimization problem. Since $\phi_d(x)$ is convex, this optimization problem is a convex program. Thus, to prove the claim it suffices to show that the Lagrange multiplier condition holds at e. Since $P_d^2 = P_d$, $P_d \lambda_d = \lambda_d$. Thus, $\lambda_d \in W_d$. It can be verified that this together with the fact that $P_d \nabla \phi_d(e) = \lambda_d$ is precisely the optimality condition at e. Thus,

$$\hat{\mu}_d = \phi_d(e) = \phi(d). \tag{7.5}$$

From the acuteness of K_d, the fact that P_d is self-adjoint, and property (P1) of Proposition 3.8, we have:

$$0 < \langle e, \lambda_d \rangle = \langle e, P_d \nabla \phi_d(e) \rangle = \langle P_d e, \nabla \phi_d(e) \rangle = \langle e, \nabla \phi_d(e) \rangle = p\phi(d). \tag{7.6}$$

Thus, $\hat{\mu}_d > 0$. It is easy to show that

$$\mu_d > 0 \quad \Longleftrightarrow \hat{\mu}_d > 0 \tag{7.7}$$

Hence, $\mu_d > 0$. Finally, from Lemma 7.6, we must have $\mu > 0$. □

LEMMA 7.9. *Suppose $K_d = K$ for all $d > 0$. Then,*

$$\mu \leq 0 \quad \Longleftrightarrow \quad \forall\ d > 0, \quad ||P_d \nabla \phi_d(e) - e|| \geq \gamma^*.$$

Proof. Suppose that $\mu > 0$. From Theorem 4.2, there exists $d > 0$ satisfying the scaling equation. But such d violates the inequality in the lemma. Conversely, suppose there exists $d > 0$ such that $||P_d \nabla \phi_d(e) - e|| < \gamma^*$. In particular, from the definition of γ^* it follows that $P_d \nabla \phi_d(e) \in K^\circ$. Then, Lemma 7.8 implies $\mu > 0$. □

This complete the proof of Theorem 7.1.

8. Bounds on the norm of scaling vectors, central-paths, regions of quadratic convergence, and scaled gradient projections. 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 $t \in (0,1]$, $d \in S_t(\bar{\lambda}_*)$ is obtained via Phase I of the Path-Following algorithm (see § 3.6), 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.

8.1. Bounds on the norm of scaling vector.

THEOREM 8.1. *Suppose $\mu > 0$. Then,*

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

Proof. The point d^* satisfies

$$P\nabla\psi(d^*) = P\nabla\phi(d^*) + P\nabla F(d^*) = 0. \tag{8.1}$$

Taking the inner product of (8.1) with d^*, and using properties (P1) and (P5) of Proposition 3.8, we get $p\phi(d^*) = \theta$. Dividing the latter by $||d^*||^p$, using homogeneity of ϕ, and the definition of μ, it follows that

$$p\mu \leq p\phi\left(\frac{d^*}{||d^*||}\right) = \frac{\theta}{||d^*||^p}. \square \tag{8.2}$$

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

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

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

Proof. We have

$$P\nabla f^{(t)}(d_t^*) = tP\nabla\phi(d_t^*) + tPu + P\nabla F(d_t^*) = 0. \tag{8.3}$$

Taking the inner product of the above with d_t^*, and using properties (P1) and (P5) of Proposition 3.8, we get

$$pt\phi(d_t^*) + tu^T d_t^* - \theta = 0. \tag{8.4}$$

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

$$\phi\left(\frac{d_t^*}{||d_t^*||}\right) = \frac{1}{||d_t^*||^p}\phi(d_t^*) \geq \mu, \quad tu^T d_t^* \geq -t||u||\ ||d_t^*||. \tag{8.5}$$

Using the inequalities of (8.5) in (8.4), we get

$$tp\mu||d_t^*||^p - t||u||\ ||d_t^*|| - \theta \leq 0. \tag{8.6}$$

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

$$p\mu||t^{1/p}d_t^*||^p - ||u||\ ||t^{1/p}d_t^*|| - \theta \leq 0. \tag{8.7}$$

The function

$$w(r) = p\mu r^p - ||u||r - \theta \tag{8.8}$$

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

Remark. In § 10 (Corollary 10.2), we show that if $\mu \leq 0$, then both $||d_t^*||$ and $||\hat{d}_t^*||$ approach infinity as t goes to zero. Simple examples can be constructed for which $||d_t^*||$ diverges, even if d^* exists.

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

LEMMA 8.3. *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

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

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

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

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 6.2, Cauchy-Schwarz inequality, and the fact that $||D|| \leq \rho||d||$, we get

$$||d|| - ||d'|| \leq ||d' - d|| \leq ||Dz|| \leq ||D||\ ||z|| \leq \rho||d||\frac{\lambda(d)}{(1+\beta)}. \tag{8.9}$$

Rearranging the above, we get

$$||d||[1 - \frac{\rho}{1+\beta}\lambda(d)] \leq ||d'||. \tag{8.10}$$

We have

$$(1 - \frac{\rho}{1+\beta}\lambda(d)) = (1 - \bar{\alpha}\lambda(d)) \geq (1 - \lambda_*) > 0. \tag{8.11}$$

The desired inequality now follows from (8.10) and (8.11). □

THEOREM 8.4. *Assume that ϕ is β-compatible, and $T(K)$ is a bounded operator-cone. Let*

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

If $d \in S(\bar{\lambda}_*)$, then $||d|| \leq h(\beta, \rho)||d^*||$.

If $d \in S_t(\bar{\lambda}_*)$, then $||d|| \leq h(\beta, \rho)||d_t^*||$. In particular, $||t^{1/p}d|| \leq h(\beta, \rho)||\hat{d}_t^*||$.

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 8.3 as applied to d^k gives

$$||d|| \leq \frac{||d^{k+1}||}{\prod_{j=0}^{k}(1 - \bar{\alpha}\lambda_j)}. \tag{8.12}$$

From Nesterov and Nemirovskii's Theorem (Theorem 3.23), we have $\lambda_{k+1} < \lambda_k^2/(1 - \lambda_*)^2$. This gives, for each $j = 1, \ldots, k$,

$$\lambda_j \leq \bar{\lambda}_* r^{2^j - 1}, \quad r = \frac{\bar{\lambda}_*}{(1 - \lambda_*)^2}. \tag{8.13}$$

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

$$\lambda_j \leq r^{2^j}. \tag{8.14}$$

Also, since $\lambda_*/(1 - \lambda_*)^2 < 1$, we have $r < (1 + \beta)/(1 + \beta + \rho)$. Thus, $1 - r^2(1 + \bar{\alpha}) > 0$. We have

$$\prod_{j=0}^{k}(1 - \bar{\alpha}\lambda_j) \geq (1 - \bar{\alpha}\lambda_*)\prod_{j=1}^{k}(1 - \bar{\alpha}\lambda_j) \geq (1 - \bar{\alpha}\lambda_*)(1 - \bar{\alpha}r^2\sum_{i=0}^{\infty} r^{2i}) \tag{8.15}$$

$$= (1 - \bar{\alpha}\lambda_*)(1 - \frac{\bar{\alpha}r^2}{1 - r^2}) = (1 - \bar{\alpha}\lambda_*)\frac{1 - r^2(1 + \bar{\alpha})}{1 - r^2} > 0.$$

Substituting (8.15) in (8.12), it follows that $||d|| \leq h(\beta, \rho)||d^k||$, for all k. Since from Theorem 5.9, d^k converges to d^*, the proof of the desired inequality follows. □

THEOREM 8.5. *Assume that ϕ is β-compatible, and $T(K)$ is 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 within Phase I of the Path-Following algorithm (see § 3.6) starting at x^0. 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 3.23), 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

$$d^{k+1} = d^k + y_{t_k}(d^k) = d^k + D^k z_{t_k}, \tag{8.16}$$

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 6.2, and (8.16), we get

$$||d^{k+1}|| \le ||d^k|| + ||D^k z_{t_k}|| \le ||d^k|| + \rho ||d^k||\ ||z_{t_k}|| \le ||d^k||(1 + \frac{\rho}{1+\beta}). \tag{8.17}$$

From the repeated application of (8.17), and since $||x^0|| = 1$, we get

$$||d^k|| \le (1 + \frac{\rho}{1+\beta})^k. \tag{8.18}$$

By setting $k = k_t \le [9(1+\beta)\sqrt{\theta}\ln(1/t) + 1]$ in (8.18), and $G = (1 + \rho/(1+\beta))^k$, we have

$$\ln G \le \left(\ln(\frac{1}{t})^{9(1+\beta)\sqrt{\theta}} + 1 \right) \ln(1 + \frac{\rho}{1+\beta}). \tag{8.19}$$

From the above we get

$$G \le (1 + \frac{\rho}{1+\beta})(\frac{1}{t})^{9(1+\beta)\sqrt{\theta}\ln(1+\frac{\rho}{1+\beta})}. \tag{8.20}$$

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

$$(1+\beta)\sqrt{\theta}\ln(1 + \frac{\rho}{1+\beta}) = \rho\sqrt{\theta}\frac{(1+\beta)}{\rho}\ln(1 + \frac{\rho}{1+\beta}) \le \rho\sqrt{\theta}. \tag{8.21}$$

From (8.21), (8.18), the bound on G in (8.20), 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.

8.3. Bound on the norm of scaled gradient projections. The following can be proved (see [20], Lemma 3):

PROPOSITION 8.6. *Let H be a self-adjoint (symmetric) and positive operator on a finite dimensional Hilbert space E. For any $x \in E$*

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

The following is an important implication of the above proposition.

LEMMA 8.7. *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,*

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

Proof. Let $H = P\nabla^2\psi(d)P$. Since $P\nabla^2\psi(d)Py = -P\nabla\psi(d)$, and $||H|| \le ||P||\ ||\nabla^2\psi(d)||\ ||P|| \le ||\nabla^2\psi(d)||$, Proposition 8.6 implies the desired result. □

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

LEMMA 8.8. *Assume that ϕ is β-compatible, and $T(K)$ is a bounded operator-cone. Let $q = \sup\{||\nabla^2\phi(d)|| : d > 0, ||d|| = 1\}$. Then, $d \in S(\lambda_*)$ implies*

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

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 $H = P_d \nabla^2 \psi_d(e_d) P_d = P_d \nabla^2 \phi_d(e_d) P_d + P_d \nabla^2 F_d(e_d) P_d$. Let z be the solution to $Hz = -P_d \nabla \psi_d(e_d)$ (the scaled Newton direction). From Proposition 8.6, $\|Hz\| \leq \|H\| \sqrt{z^T Hz}$. On the one hand, $\lambda(d) = (1+\beta)\sqrt{z^T Hz}$. On the other hand, we can bound $\|H\|$. Firstly, recalling the definition of an operator-cone (Definition 3.13) we have $\|P_d \nabla^2 F_d(e_d) P_d\| \leq \|P_d\| \, \|\nabla^2 F_d(e_d)\| \, \|P_d\| \leq \|\nabla^2 F_d(e_d)\| \leq M$. Also

$$\|P_d \nabla^2 \phi_d(e_d) P_d\| = \|P_d D^T \nabla^2 \phi(d) D P_d\| \leq \tag{8.22}$$

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

But, from property (P3) of Proposition 3.8 and definition of q, we have

$$\|\nabla^2 \phi(d)\| = \|\nabla^2 \phi(\|d\| \frac{d}{\|d\|})\| = \|d\|^{p-2} \|\nabla^2 \phi(\frac{d}{\|d\|})\| \leq q\|d\|^{p-2}. \tag{8.23}$$

Substituting (8.23) in (8.22), we get $\|H\| \leq (q\rho^2 \|d\|^p + M)$. Hence the proof. □

We now state the following fundamental theorem that is a strengthened version of Lemma 6.2. This theorem will be used to prove complexity theorems on the Potential-Reduction and Path-Following algorithms.

THEOREM 8.9. *Assume ϕ is β-compatible, and $T(K)$ is a bounded operator-cone. Let $\bar{\alpha} = \rho/(1+\beta)$, $\bar{\lambda}_* = \lambda_*/(1+\bar{\alpha})$, and $q = \sup\{\|\nabla^2 \phi(d)\| : d > 0, \|d\| = 1\}$. If $\mu > 0$, then given $d \in S(\bar{\lambda}_*) = \{x > 0 : \lambda(x) < \bar{\lambda}_*\}$, we have*

$$\|z\| = \|D^{-1} y(d)\| \leq \frac{\lambda(d)}{(1+\beta)} \leq \|P_d \nabla \psi_d(e_d)\| \leq \tag{8.24}$$

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

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 § 3.6). Then,

$$\|z_t\| = \|D^{-1} y_t(d)\| \leq \frac{\lambda_t(d)}{(1+\beta)} \leq \|P_d \nabla f_d^{(t)}(e_d)\| \leq \tag{8.25}$$

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

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

$$\|P_d \nabla f_d^{(t)}(e_d)\| \leq \tag{8.26}$$

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

Proof. The first two inequalities in (8.24) have already been proved in Lemma 6.2. The next inequality follows by bounding $\|d\|$, $d \in S(\bar{\lambda}_*)$ in Lemma 8.8, using Theorem 8.1, and Theorem 8.4. Again the first two inequalities in (8.25) have already been proved in Lemma 6.2. The next inequality follows by bounding $d \in S_t(\bar{\lambda}_*)$ in Lemma 8.8, using Theorem 8.5. The bound in (8.26) follows from bounding $t^{1/p} d$, in Lemma 8.8, using Theorem 8.2 and Theorem 8.4. □

9. The Potential-Reduction Complexity Theorem. In this section we will study the application of the Potential-Reduction algorithm, described in § 3.6 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 3.23), 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 [31]). Given $d > 0$, the algorithm simply replaces d with $d' = NEW(\psi, d)$ and repeats. By making use of the Scaling Duality Theorem (Theorem 5.5), we will see that this simple algorithm solves the desired problems. We need some preliminary results.

LEMMA 9.1. *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

$$\psi(x) \geq \min\{\psi(tx) : t \in (0, \infty)\} = \tag{9.1}$$

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

Let

$$R = \sup\{exp(-\frac{p}{\theta}F(x)) : x \in W \cap K^\circ, ||x|| = 1\}. \tag{9.2}$$

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 9.2. *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 9.1, definitions of R, μ, and that $X(x)$ is homogeneous of degree zero. To prove the second result, from Theorem 4.2 if d^* (the minimizer of ψ) exists, then it is also the minimizer of $X(x)$. From this, and the inequality in Lemma 9.1, we have

$$\psi^* = \frac{\theta}{p}(1 - \ln\frac{\theta}{p} + \ln X^*). \tag{9.3}$$

Subtracting the above from the inequality in Lemma 9.1, the desired results follow. $\square$

THEOREM 9.3. (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

$$O\left(\theta \ln \frac{R\sigma}{\epsilon}\right). \tag{9.4}$$

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

$$O\left(\theta \ln \frac{R\sigma}{\mu} + \ln\ln \frac{1}{\epsilon}\right). \tag{9.5}$$

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

$$O\left(\theta \ln \frac{R\sigma}{\mu} + \ln\ln(1 + \frac{\rho}{1+\beta}) + \ln\ln \frac{q}{\epsilon}\right). \tag{9.6}$$

Proof. Let the k-th iterate of the algorithm be denoted by x^k. Let $\lambda_k = \lambda(x^k)$. We claim that if $\lambda_k \geq \bar{\lambda}$ (see Definition 3.22), then

$$\phi\left(\frac{x^k}{||x^k||}\right) \leq R\sigma \ exp(-\frac{k\delta p}{\theta}), \quad \delta = \frac{(\bar{\lambda} - \ln(1+\bar{\lambda}))}{(1+\beta)^2}. \tag{9.7}$$

From Theorem 3.23, if λ_j is less than $\bar{\lambda}$, then so is λ_{j+1}. Thus, if $\lambda_k \geq \bar{\lambda}$, then $\lambda_j \geq \bar{\lambda}$ 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 9.1, and the definition of σ it easily follows that

$$X(x^k) \leq \sigma \ exp(-\frac{k\delta p}{\theta}). \tag{9.8}$$

From (9.8), the fact that $X(tx) = X(x)$, and the definition of R (see (9.2)), the proof of (9.7) is immediate.

Suppose that $\mu \leq 0$. Then, we claim that the algorithm will never obtain a point $d > 0$ such that $\lambda(d) < \bar{\lambda}$. Otherwise, from Theorem 3.23 the set $S(\lambda_*) = \{d > 0 : \lambda(d) < \lambda_*\}$ is nonempty. Then, from Theorem 5.9, the minimizer d^* of ψ exists. But, by the Scaling Duality Theorem (Theorem 5.5), either $\mu \leq 0$, or d^* exists. This proves the claim. Now the inequality in (9.7) proves the claimed bound on the number of iterations to solve ϵ-HP.

Suppose that $\mu > 0$. From Corollary 9.2, 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

$$\phi\left(\frac{x^k}{||x^k||}\right) \geq \mu, \tag{9.9}$$

from (9.7), the Scaling Duality Theorem (Theorem 5.5), and the second part of (3.14) in Theorem 3.23, 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 3.23, starting with $d^0 = d$, after k subsequent iterations, we obtain $d^k > 0$ such that $\lambda_k = \lambda(d^k)$ satisfies

$$\lambda_k \leq {r^*}^{2^k}, \quad r^* = \frac{\lambda_*}{(1-\lambda_*)^2}. \tag{9.10}$$

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

$$\lambda_k < \min\{\lambda_*, \frac{\epsilon}{3}(1+\beta)^2(1-\omega_*)\}, \tag{9.11}$$

where $\omega_* = \omega(\lambda_*)$, and $\omega(\lambda) = 1 - (1-3\lambda)^{1/3}$. We claim that $\psi(d^k) - \psi^* \leq \epsilon$. Clearly, from (9.10), 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 3.23, see (3.15), it follows that

$$\epsilon \geq \frac{\omega(\lambda_k)}{(1+\beta)^2(1-w_*)} \geq \frac{\omega^2(\lambda_k)}{(1+\beta)^2(1-w_*)} \geq \frac{\omega^2(\lambda_k)}{(1+\beta)^2(1-w(\lambda_k))} \tag{9.12}$$

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

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 (9.10), 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 8.9, 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 $d^k > 0$ such that $\lambda_k = \lambda(d^k)$ satisfies

$$\lambda_k \leq r^{2^k}, \quad r = \frac{\bar{\lambda}_*}{(1-\lambda_*)^2}. \tag{9.13}$$

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. $\square$

Remark. Even for $\phi(x)$ linear, the above algorithm is simpler than a potential-reduction algorithm described by Nesterov and Nemirovskii [31] (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 (when a bounded operator-cone is available).

Moreover, since for any β-compatible ϕ, the Uniform Scaling Duality is valid with parameter $\gamma^* = \frac{1}{(1+\beta)}$, within each iteration a duality check can be implemented. More precisely, when solving ϵ-HP, if we encounter an iterate d, such that $||P_d \nabla \psi_d(e_d)|| < \gamma^*$, then we terminate the algorithm with the conclusion that $\mu > 0$. 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 $p = 1$. In § 11 it is shown that quadratic programming can be formulated as an HP with a nonlinear ϕ, where $p = 1$. This ϕ will however be shown to be β-compatible.

10. The Path-Following Complexity Theorem. In this section we first prove a lemma, called the Path-Following Lemma (Lemma 10.1). 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 10.3), a theorem that establishes the significance of computing approximate minimizers of $f^{(t)}(x) = t\phi(x) + tu^T x + F(x)$, 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 8.9, gives appropriate values for $t_* \in (0,1)$ so that the Path-Following algorithm (see § 3.6) can solve in polynomial-time any of the four problems: ϵ-HP, ϵ-SP, ϵ-HSP, or ϵ-ASP.

LEMMA 10.1. (Path-Following Lemma) *Assume that $p > 1$, and $T(K)$ is a given bounded operator-cone. Given positive numbers $\gamma_1 < \gamma_2$, define*

$$C(p, \theta, \rho, N, ||u||, \gamma_1, \gamma_2) = \frac{[(\theta + \gamma_1\sqrt{N})\rho + (\gamma_2 - \gamma_1)]\rho^{p-1}||u||^p}{p(\gamma_2-\gamma_1)^p}. \tag{10.1}$$

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

$$||P_d \nabla f_d^{(t)}(e_d)|| \leq \gamma_1. \tag{10.2}$$

Let

$$\hat{d} = t^{1/p} d. \tag{10.3}$$

If

$$\|P_{\hat{d}} \nabla \psi_{\hat{d}}(e_{\hat{d}})\| \geq \gamma_2, \tag{10.4}$$

then

$$\phi\left(\frac{d}{\|d\|}\right) \leq C(p, \theta, \rho, N, \|u\|, \gamma_1, \gamma_2) t^{p-1}. \tag{10.5}$$

Proof. Let $d \in W \cap K^\circ$ be a point satisfying (10.2). We have

$$P_d \nabla f_d^{(t)}(e_d) = t P_d \nabla \phi_d(e_d) + t P_d D^T u + P_d \nabla F_d(e_d), \tag{10.6}$$

where $D = T(d)$. Let $\alpha = t^{1/p}$. Thus, $\hat{d} = \alpha d$. From the chain rule and property (P2) of Proposition 3.8, we have

$$\nabla \phi_{\hat{d}}(e_{\hat{d}}) = \hat{D}^T \nabla \phi(\hat{d}) = \alpha D^T \nabla \phi(\alpha d) = \alpha \alpha^{p-1} D^T \nabla \phi(d) = \alpha^p \nabla \phi_d(e_d). \tag{10.7}$$

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

$$P_{\hat{d}} \hat{D}^T u = \alpha P_d D^T u. \tag{10.8}$$

From the chain rule and property (P6) of Proposition 3.8, we have

$$\nabla F_{\hat{d}}(e_{\hat{d}}) = \hat{D}^T \nabla F(\hat{d}) = \alpha D^T \nabla F(\alpha d) = \alpha \alpha^{-1} D^T \nabla F(d) = \nabla F_d(e_d). \tag{10.9}$$

Substituting (10.7) and (10.9) into (10.6), we get

$$P_d \nabla f_d^{(t)}(e_d) = P_{\hat{d}} \nabla \phi_{\hat{d}}(e_{\hat{d}}) + t^{(p-1)/p} P_{\hat{d}} \hat{D}^T u + P_{\hat{d}} \nabla F_{\hat{d}}(e_{\hat{d}}). \tag{10.10}$$

Now assume $\hat{d}$ satisfies (10.4) and let

$$v = P_{\hat{d}} \nabla \psi_{\hat{d}}(e_{\hat{d}}) = P_{\hat{d}} \nabla \phi_{\hat{d}}(e_{\hat{d}}) + P_{\hat{d}} \nabla F_{\hat{d}}(e_{\hat{d}}), \qquad w = t^{(p-1)/p} P_{\hat{d}} \hat{D}^T u. \tag{10.11}$$

Thus, from (10.2) and (10.4), $\|v + w\| \leq \gamma_1$ and $\|v\| \geq \gamma_2$, respectively. Since $\|v\| - \|w\| \leq \|v + w\|$, we have $\|w\| \geq \|v\| - \|v + w\|$. Thus,

$$t^{(p-1)/p} \|P_{\hat{d}} \hat{D}^T u\| \geq (\gamma_2 - \gamma_1). \tag{10.12}$$

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

$$\|P_{\hat{d}} \hat{D} u\| \leq \rho \|P_{\hat{d}}\| \, \|\hat{d}\| \, \|u\| \leq \rho \|\hat{d}\| \, \|u\|. \tag{10.13}$$

Thus, from (10.12) and (10.13), we get

$$\frac{1}{\|\hat{d}\|} \leq \frac{\rho t^{(p-1)/p} \|u\|}{(\gamma_2 - \gamma_1)}. \tag{10.14}$$

From the chain rule and properties (P1) and (P8) of Proposition 3.8, we have

$$e_{\hat{d}}^T \nabla \phi_{\hat{d}}(e_{\hat{d}}) = p \phi_{\hat{d}}(e_{\hat{d}}) = p \phi(\hat{d}), \quad e_{\hat{d}}^T \nabla F_{\hat{d}}(e_{\hat{d}}) = \hat{d}^T \nabla F(\hat{d}) = -\theta. \tag{10.15}$$

Taking the inner product with $e_{\hat{d}}$ in (10.10), using the fact that $P_{\hat{d}} e_{\hat{d}} = e_{\hat{d}}$, together with (10.15), Cauchy-Schwarz inequality, and condition (2) of operator-cone (see 3.13), we get

$$e_{\hat{d}}^T P_d \nabla f_d^{(t)}(e_d) = p\phi(\hat{d}) + t^{(p-1)/p} u^T \hat{d} - \theta \le \|e_{\hat{d}}\| \; \|P_d \nabla f_d^{(t)}(e_d)\| \le \sqrt{N}\gamma_1. \tag{10.16}$$

This implies,

$$p\phi(\hat{d}) \le \theta + \sqrt{N}\gamma_1 + t^{(p-1)/p}\|\hat{d}\| \; \|u\|. \tag{10.17}$$

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 (10.14) we have

$$\frac{\theta + \sqrt{N}\gamma_1}{\|\hat{d}\|^p} \le \frac{\rho^p t^{(p-1)}\|u\|^p}{(\gamma_2 - \gamma_1)^p}. \tag{10.18}$$

Also, from (10.14) we have

$$\frac{t^{(p-1)/p}\|\hat{d}\| \; \|u\|}{\|\hat{d}\|^p} \le \frac{\rho^{p-1} t^{(p-1)}\|u\|^p}{(\gamma_2 - \gamma_1)^{p-1}}. \tag{10.19}$$

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

COROLLARY 10.2. *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 6.3) 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 (10.14) is valid for all $\gamma_1 < \gamma^*$. Now letting γ_1 converge to zero, we get the desired result. □

THEOREM 10.3. (Path-Following Theorem) *Assume that ϕ is convex, $p > 1$, $T(K)$ is a given bounded operator-cone with respect to which the Uniform Scaling Duality holds with parameter γ^*, and F is a θ-normal barrier.*

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

$$C(p, \theta, \rho, N, \|u\|, \frac{1}{2}\epsilon, \epsilon)t^{p-1} < \mu. \tag{10.20}$$

There exists $d > 0$ satisfying

$$\|P_d \nabla f_d^{(t)}(e_d)\| \le \frac{\epsilon}{2}. \tag{10.21}$$

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

$$\|P_{\hat{d}} \nabla \psi_{\hat{d}}(e_{\hat{d}})\| < \epsilon. \tag{10.22}$$

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

$$C(p, \theta, N, \|u\|, \frac{1}{2}\gamma^*, \gamma^*)t^{p-1} \le \epsilon. \tag{10.23}$$

There exists $d > 0$ satisfying

$$\|P_d \nabla f_d^{(t)}(e_d)\| \le \frac{1}{2}\gamma^*. \tag{10.24}$$

Moreover,

$$\phi\left(\frac{d}{||d||}\right) \leq \epsilon. \tag{10.25}$$

Proof. From Theorem 5.8, for each t, $d_t^* = \operatorname{argmin}\{f^{(t)}(x) : x > 0\}$ 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 10.1). □

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

THEOREM 10.4. (Path-Following Complexity Theorem) *Assume ϕ is β-compatible, $p > 1$, and $T(K)$ is 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 § 3.6). More precisely:*

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

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

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

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

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

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

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

Proof. The stated right-hand-side bounds assume that the parameter ρ, γ^* are $O(1)$ constants, and $N = O(\theta)$. 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 10.3), 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

$$\frac{1}{t_*} = \left(C(p,\theta,\rho,N,||u||,\frac{1}{2}\gamma^*,\gamma^*)\frac{1}{\epsilon}\right)^{\frac{1}{p-1}}. \tag{10.26}$$

From the above, and Theorem 3.23, 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 $k_{t_*} = \lceil 9(1+\beta)\sqrt{\theta}\ln(\frac{1}{t_*})\rceil$. 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 3.23, it follows that $\lambda_k \leq r^{2^k}$ (see (8.14)), where $r = \bar{\lambda}_*/(1-\lambda_*)^2$, $\bar{\lambda}_* = \lambda_*(1+\beta)/(1+\beta+\rho)$. From Theorem 8.9, if $\lambda_k < \bar{\lambda}_*$, then we have

$$||P_{d^k} \nabla f_{d^k}^{(t_*)}(e_{d^k})|| \leq \left((1+\frac{\rho}{1+\beta})^p q\rho^2\left(\frac{1}{t_*}\right)^{9p\rho\sqrt{\theta}} + M\right)\frac{r^{2^k}}{(1+\beta)}. \tag{10.27}$$

Substituting in the right-hand-side of (10.27) for $1/t_*$ from the equation (10.26), and bounding the resulting expression by $\gamma^*/2$, with the assumption that $M = O(1)$, we get

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

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

$$\frac{1}{t_*} = \left(C(p, \theta, \rho, N, \|u\|, \frac{1}{2}\epsilon, \epsilon) \frac{1}{2\mu} \right)^{\frac{1}{p-1}}. \tag{10.28}$$

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 (10.27), but using the new value of $1/t_*$ given in (10.28). Note that in this case, since $\mu > 0$, we can use the alternative bound from Theorem 8.9. 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 8.9, 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 9.3). Thus, the proof of the theorem is complete. ∎

11. Applications of the Potential-Reduction and the Path-Following Complexity Theorems.

In this section we will consider several applications of the Potential-Reduction and the Path-Following Complexity Theorems in deriving polynomial-time algorithms for linear, quadratic, and semidefinite programming, as well as solution of the corresponding scaling equations. Also, we will use these theorems to derive polynomial-time algorithms for computing the minimum of classic homogeneous ratios, such as the arithmetic-geometric mean, or trace-determinant, but over an arbitrary subspace, intersecting the corresponding cones.

11.1. Applications of the Path-Following Complexity Theorem.

THEOREM 11.1. *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 β-compatible with the corresponding θ-normal barrier F, $p > 1$, 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. In particular, these apply to linear programming, convex quadratic programming over linear or convex quadratic constraints, semidefinite programming, and their corresponding scaling problems. In particular, 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.*

Proof. As shown in Proposition 3.16, T_F is a bounded operator-cone. Thus, both the Potential-Reduction Complexity Theorem and the Path-Following Complexity Theorem (Theorem 9.3, and Theorem 10.4) apply to any β-compatible ϕ.

To apply the Path-Following algorithm to linear programming, we consider Karmarkar's canonical LP and simply replace the linear objective $c^T x$ with $\phi(x) = (c^T x)^2$. Also, to apply the Path-Following algorithm to semidefinite programming, we 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 known fact that it can be formulated as a semidefinite programming problem.

To compute the minimum of $c^T x / (\prod_{i=1}^n x_i)^{1/n}$ over the intersection of the nonnegative orthant and an arbitrary subspace W of $\Re^n$, we minimize $(c^T x)^2 / (\prod_{i=1}^n x_i)^{2/n}$, i.e., solve ϵ-HSP with $\phi(x) = (c^T x)^2$. 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, 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, we can incorporate a simple duality test in either the Potential-Reduction, or the Path-Following algorithm. Since the Uniform Scaling Duality holds, within each iteration we check if the current iterate, d, satisfies $\|P_d \nabla \phi_d(e) - e\| < \gamma^*$. If so HP is unsolvable. Given that the Scaling Separation Duality holds, we can strengthen this duality test. For instance, for the nonnegative and semidefinite cone we can check if the current iterate satisfies $P_d \nabla \phi_d(e) \in K^\circ$. The set $\{d > 0 : \|P_d \nabla \phi_d(e) - e\| < \gamma^*\}$, if nonempty, is a proper subset of the set $\{d > 0 : P_d \nabla \phi_d(e) \in K^\circ\}$.

11.2. A theorem on compatibility and its application to quadratic programming. Convex quadratic programming can be formulated as an HP over the semidefinite cone. However, as discussed in § 2.1 linearly constrained convex quadratic programming in $\Re^n$ can be formulated as an HP over $K = \Re^n_+$, where the corresponding $\phi(x)$ is homogeneous of degree 1. The aim of this section is to show that the corresponding ϕ is β-compatible with $F(x) = -\sum_{i=1}^n \ln x_i$. This in particular implies the applicability of the Potential-Reduction Complexity Theorem (Theorem 9.3).

THEOREM 11.2. *Assume Q is an $n \times n$ symmetric positive semidefinite matrix, $c \in \Re^n$, and $e = (1, \ldots, 1)^T \in \Re^n$. Consider an HP over $K \cap W$, where $K = \Re^n_+$, W a subspace of $\Re^n$,*

$$\phi(x) = \frac{1}{2} \frac{x^T Q x}{e^T x} + c^T x,$$

and $F(x) = -\sum_{i=1}^n \ln x_i$. Then, $\phi(x)$ is $\sqrt{3}$-compatible with F (i.e., ψ is strongly self-concordant with parameter $a = 1/(1+\sqrt{3})^2$).

Proof. Since in the presence of a subspace, strong self-concordance is preserved (see Proposition 2.1.1 in [31]), to prove the theorem we may assume $W = \Re^n$. Also, without loss of generality we may assume that $c = 0$. For any $x \in K^\circ$, $h \in \Re^n$, from the Taylor's expansion of $\phi(x+h)$ it is easy to show

$$\nabla^2 \phi(x)[h,h] = \frac{(e^T h)^2}{e^T x} \left(\frac{x}{e^T x} - \frac{h}{e^T h}\right)^T Q \left(\frac{x}{e^T x} - \frac{h}{e^T h}\right). \tag{11.1}$$

Since Q is positive semidefinite, from (11.1) it follows that $\nabla^2 \phi(x)$ is positive semidefinite. Hence, the proof of convexity of ϕ over K°. From the Taylor's expansion, it also follows that

$$\nabla^3 \phi(x)[h,h,h] = -3 \frac{(e^T h)^3}{(e^T x)^2} \left(\frac{x}{e^T x} - \frac{h}{e^T h}\right)^T Q \left(\frac{x}{e^T x} - \frac{h}{e^T h}\right). \tag{11.2}$$

Next we prove

$$|\nabla^3 \phi(x)[h,h,h]| \le \beta \left(3 \nabla^2 \phi(x)[h,h]\right) \left(3 \nabla^2 F(x)[h,h]\right)^{1/2}, \tag{11.3}$$

where $\beta = \sqrt{3}$. Clearly, to prove (11.3) it suffices to show

$$\frac{\sum_{i=1}^n h_i}{\sum_{i=1}^n x_i} \le \left(\sum_{i=1}^n \left(\frac{h_i}{x_i}\right)^2\right)^{1/2}. \tag{11.4}$$

To prove (11.4), assume without loss of generality $\sum_{i=1}^n h_i = \sum_{i=1}^n x_i = 1$. This assumption implies the inequality $|h_i| < x_i$ is not valid for all $i = 1, \ldots, n$. But, this implies $\sum_{i=1}^n (h_i/x_i)^2 \ge 1$, hence the proof. □

Remark. Since the corresponding μ, if positive, must be greater than or equal to $O(2^{-L})$, where L is the size of the quadratic problem, the number of Newton iterations of the Potential-Reduction algorithm to determine if $\mu = 0$ is at most $O(nL)$ (see Theorem 9.3). Since for the above HP $p = 1$, the Path-Following algorithm does not directly apply. However the Path-Following Theorem (Theorem 10.3) applies to $\phi(x)^2$, since under the operation of squaring convexity remains intact. But, to prove the applicability of the Path-Following Complexity Theorem, β-compatibility of $\phi(x)^2$ remains to be checked.

11.3. Matrix scaling and linear programming revisited. In this final part of the paper we consider the problem of computing a nonnegative nontrivial zero of Q, a given positive semidefinite symmetric $n \times n$ matrix, i.e., testing if $\mu = \min\{\phi(x) = \frac{1}{2}x^TQx : x \geq 0, ||x|| = 1\}$ is zero; and the problem of computing a positive diagonal matrix D such that $DQDe = e$, where e is the vector of ones.

The results presented here are completely independent of the rest of the paper, do not rely on the of theory self-concordance, nor do they rely on the bounds derived in this paper. It is completely self-contained, and very simple. We will rederive the polynomial-time linear programming/matrix scaling algorithm described in [23], but from a fresh point of view, and even with simpler analysis. In particular, the new derivation emphasizes the dependence of the algorithm in the fundamental separation theorem of Gordan, rather than the matrix scaling duality theorem according to which precisely one of the two problems is solvable. In fact we make use of the algorithm to prove this duality.

We will consider ϵ-approximate version of these problems. For detailed analysis of how over the rational or algebraic numbers, an approximate zero can be rounded into an exact zero, and for deriving a lower bound on μ, see [20].

Define ϵ-HP as the problem of computing $d > 0$, $||d|| = 1$, such that $\phi(d) \leq \epsilon$, or proving its unsolvability. Define ϵ-ASP as the problem of computing a diagonal matrix D with positive diagonal entries, such that $||DQDe - e|| \leq \epsilon$, or proving its unsolvability. Let $F(x) = -\sum_{i=1}^n \ln x_i$, $\psi(x) = \phi(x) + F(x)$. Let $u \in \Re^n$ be arbitrary, and for each $t \in (0,1]$, define $f^{(t)}(x) = t\phi(x) + tu^Tx + F(x)$. For a given $d > 0$, let $D = \mathrm{diag}(d)$, $\phi_d(x) = \phi(Dx)$, $\psi_d(x) = \psi(Dx)$, and $f_d^{(t)}(x) = f^{(t)}(Dx)$. Let $y_t(d)$ be the solution to $\nabla^2 f^{(t)}(d)y_t(d) = -\nabla f^{(t)}(d)$. Define Newton iterate and Newton decrement as $d_t' = NEW(f^{(t)}, d) \equiv d + y_t(d)$ and $\lambda_t(d) = [y_t(d)^T\nabla^2 f^{(t)}(d)y_t(d)]^{1/2}$, respectively. Set $z_t = D^{-1}y_t(d)$, $Z_t = \mathrm{diag}(z_t)$, and $z_t' = D_t'^{-1}y_t(d_t')$. It is easy to see that z_t is the solution to $\nabla^2 f_d^{(t)}(e)z_t = -\nabla f_d^{(t)}(e)$, and $\lambda_t(d) = [z_t^T\nabla^2 f_d^{(t)}(e)z_t]^{1/2}$.

LEMMA 11.3. (Corollary of Gordan's theorem) *If $\mu = 0$, then for all $d > 0$, $||DQDe - e|| \geq 1$.*

Proof. Suppose there exists $d > 0$ such that $||DQDe - e|| < 1$. This implies that $DQDe > 0$. Thus, $Qd > 0$. From Gordan's theorem the system $Qx = 0$ has no nontrivial nonnegative solution. Thus, $\mu > 0$, a contradiction. □

LEMMA 11.4. (Path-Following Lemma) *Let γ be a number in $(0,1]$. Given $t \in (0,1]$, suppose there exists $d > 0$ such that $||\nabla f_d^{(t)}(e)|| = ||tDQDe + tDu - e|| \leq \frac{1}{2}\gamma$. Let $\hat{d} = \sqrt{t}d$. If $||\nabla\psi_{\hat{d}}(e)|| = ||\hat{D}Q\hat{D}e - e|| \geq \gamma$, then $\phi(\frac{d}{||d||}) \leq C(\gamma)t$, where $C(\gamma) = \frac{1}{\gamma^2}[2n + \gamma(\sqrt{n}+1)]||u||^2$.*

Proof. Clearly, $\nabla f_d^{(t)}(e) = (\hat{D}Q\hat{D}e - e) + \sqrt{t}\hat{D}u$. From Cauchy-Schwarz inequality, and the assumed upper bound on $||\nabla f_d^{(t)}(e)||$, we get

$$e^T\nabla f_d^{(t)}(e) = 2\phi(\hat{d}) + \sqrt{t}u^T\hat{d} - n \leq ||e||\;||\nabla f_d^{(t)}(e)|| \leq \frac{\sqrt{n}}{2}\gamma.$$

This implies,

$$2\phi(\hat{d}) \leq n + \frac{\sqrt{n}}{2}\gamma + \sqrt{t}||\hat{d}||\;||u||. \tag{11.5}$$

Since

$$||\sqrt{t}\hat{D}u|| \geq ||\hat{D}Q\hat{D}e - e|| - ||\nabla f_d^{(t)}(e)|| \geq \gamma - \frac{1}{2}\gamma = \frac{1}{2}\gamma,$$

and $||\hat{d}||\;||u|| \geq ||\hat{D}||\;||u|| \geq ||\hat{D}u||$, we get the bound $1/||\hat{d}|| \leq 2\sqrt{t}||u||/\gamma$. Dividing (11.5) by $||\hat{d}||^2$, and using the bound on $1/||\hat{d}||$, we get the desired result. □

LEMMA 11.5. *Let $t \in (0,1]$, $d > 0$, and $d_t' = NEW(f^{(t)}, d)$. We have*

$$||z_t'|| \leq \lambda_t(d_t') \leq ||\nabla f_{d_t'}^{(t)}(e)|| = ||Z_tz_t|| \leq ||z_t||^2 \leq \lambda_t(d)^2 \leq ||\nabla f_d^{(t)}(e)||^2.$$

Proof. The verification of the first two inequalities, or the last two inequalities follows from the fact that if H is an $n \times n$ symmetric positive definite matrix all of whose eigenvalues are bounded below by one, then for any $w \in \Re^n$, $\|w\|^2 \leq w^T H w \leq \|Hw\|^2$. For instance, to prove the last two inequalities we take $H = \nabla^2 f_d^{(t)}(e) = t\nabla^2 \phi_d(e) + I$. The linking inequality is the identity

$$\nabla f_{d'_t}^{(t)}(e) = -Z_t z_t. \tag{11.6}$$

To prove this identity, consider the equation $\nabla^2 f_d^{(t)}(e) z_t = -\nabla f_d^{(t)}(e)$. Equivalently,

$$(t\nabla^2 \phi_d(e) + I)z_t = e - t\nabla\phi_d(e) - tDu = e - t\nabla^2\phi_d(e)e - tDu.$$

Regrouping terms in the above gives

$$t\nabla^2\phi_d(e)(e + z_t) = e - z_t - tDu.$$

Since $\nabla^2\phi_d(e) = D\nabla^2\phi(d)D$, $D(e + z_t) = d'_t$, $\nabla^2\phi(d) = \nabla^2\phi(d'_t)$, and $\nabla^2\phi(d'_t)d'_t = \nabla\phi(d'_t)$, from the above we get

$$tD\nabla\phi(d'_t) = e - z_t - tDu.$$

Multiplying the above by $I + Z_t = D'_t D^{-1}$, and using that $D'_t\nabla\phi(d'_t) = \nabla\phi_{d'_t}(e)$, (11.6) follows. □

LEMMA 11.6. *Fix $\gamma_0 \in (0,1)$. Given $t \in (0,1]$, suppose $d > 0$ satisfies $\|\nabla f_d^{(t)}(e)\| \leq \gamma_0$. Then, $d'_t = NEW(f^{(t)}, d) > 0$, and if $t' = tr_*$, where $r_* = (\frac{\sqrt{n}-\gamma_0}{\sqrt{n}-\gamma_0^2})$, then $\|\nabla f_{d'_t}^{(t')}(e)\| \leq \gamma_0$.*

Proof. Since $\|\nabla f_d^{(t)}(e)\| \leq \gamma_0 < 1$, from Lemma 11.5 we get $\|z_t\| < 1$. This implies that $d'_t = D(e + z_t) > 0$. Also, from this lemma it follows that $\|\nabla f_{d'_t}^{(t)}(e)\| \leq \gamma_0^2$. Let $a = D'_t Q D'_t e + D'_t u$. For any $\tau \in (0, t]$, we have

$$\|\nabla f_{d'_t}^{(\tau)}(e)\| = \|\tau a - e\| = \|\frac{\tau}{t}ta - \frac{\tau}{t}e + \frac{\tau}{t}e - e\| \leq \frac{\tau}{t}\|ta - e\| + (1 - \frac{\tau}{t})\|e\| = \sqrt{n} - \frac{\tau}{t}(\sqrt{n} - \gamma_0^2).$$

Setting the right-hand-side of the above equal to γ_0, and solving for $\tau = t'$ gives the desired result. □

Let $u = e - Qe$ in the definition of $f^{(t)}(x)$. Given $t_* \in (0,1)$, consider:

Path-Following:

Initialization. Let $t = 1$, $d = e$.

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 11.7. *If $\mu = 0$, the algorithm solves ϵ-HP in $O(\sqrt{n}\ln\frac{n\|u\|}{\epsilon})$ iterations of Phase I. If $\mu > 0$, the algorithm solves ϵ-ASP in $O(\sqrt{n}\ln\frac{n\|u\|}{\mu\epsilon})$ combined iterations of Phase I and Phase II.*

Proof. For a given t_*, let the k-th iterate of Phase I of the Path-Following algorithm be denoted by (d^k, t_k). Since $(d^0, t_0) = (e, 1)$, $\|\nabla f_{d^0}^{(t_0)}(e)\| = 0$. Thus, Lemma 11.6 implies that for all k, $d^k > 0$, $\|\nabla f_{d^k}^{(t_k)}(e)\| \leq \gamma_0$, and $t_k = r_*^k \leq exp(k(r_* - 1))$. Thus, if the number of iterations of Phase I is k_1, then $t_* \leq exp(k_1(r_* - 1))$. This implies $k_1 = O(\sqrt{n}\ln\frac{1}{t_*})$.

If $\mu = 0$, from Lemma 11.3, and the Path-Following Lemma (Lemma 11.4) it follows that $\phi(d^k/\|d^k\|) \leq C(\gamma)t_k$. Thus, to solve ϵ-HP it suffices to choose t_* satisfying $C(\gamma_0)t_* = \epsilon$. From the definition of $C(\gamma_0)$, we have $\frac{1}{t_*} = O(\frac{n\|u\|^2}{\epsilon})$. The claimed complexity for ϵ-HP follows.

If $\mu > 0$, since $\phi(d^k/\|d^k\|) \geq \mu$, from the Path-Following Lemma (Lemma 11.4), it suffices to implement the Path-Following algorithm with t_* satisfying $C(\epsilon)t_* = \frac{1}{2}\mu$. The algorithm will give (d,t) such that $d > 0$, $t \leq t_*$, and if $\hat{D} = \sqrt{t}D$, then $\|\hat{D}Q\hat{D}e - e\| \leq \epsilon$. In this case $\frac{1}{t_*} = O(\frac{n\|u\|^2}{\mu\epsilon^2})$. Hence, $k_1 = O(\sqrt{n}\ln\frac{n\|u\|}{\mu\epsilon})$. At the completion of Phase I we have a point d^{k_1} satisfying $\|\nabla f_{d^{k_1}}^{(t_{k_1})}(e)\| \leq \gamma_0$. Denote the iterates of Phase II by $\bar{x}^j$ where $\bar{x}^0 = d^{k_1}$. From Lemma 11.5, we have

$$\|\nabla f_{\bar{x}^j}^{(t_{k_1})}(e)\| \leq \|\nabla f_{\bar{x}^{(j-1)}}^{(t_{k_1})}(e)\|^2 \leq \cdots \leq \|\nabla f_{\bar{x}^0}^{(t_{k_1})}(e)\|^{2^j} \leq \gamma_0^{2^j}.$$

Thus, the number of iterations of Phase II is $O(\ln\ln\frac{1}{\epsilon})$. This complexity is dominated by that of Phase I. $\square$

THEOREM 11.8. *$\mu > 0$ if and only if there exists $d > 0$ such that $DQDe = e$.*

Proof. From Theorem 11.7, since $\mu > 0$, there exists a point $d > 0$ such that $\|DQDe - e\| \leq \gamma_0 < 1$. Set $d^0 = d$, and for each $k \geq 0$ define $d^{k+1} = d^k + y^k = d^k + D^k z^k$ where y^k is the solution to $\nabla^2\psi(d^k)y^k = -\nabla\psi(d^k)$. Note that $d^{k+1} = NEW(f^{(t)}, d^k)$, where $f^{(t)}(x) = t\phi(x) + tu^T x + F(x)$, $t = 1$, and $u = 0$. Thus, from Lemma 11.6 it follows that $\|D^k Q D^k e - e\|$ converges to zero. To prove the theorem we only need to show that d^k's form a bounded sequence. We have

$$\|d^{k+1}\| \leq \|d^k\| + \|D^k\|\ \|z^k\| \leq \|d^k\|(1 + \|z^k\|).$$

From Lemma 11.5, $\|z^k\| \leq \|z^{k-1}\|^2$. Thus, $\|d^k\| \leq \|d^0\| \sum_{i=1}^k \|z^0\|^i < \frac{\gamma_0}{1-\gamma_0}\|d_0\|$. $\square$

REFERENCES

[1] F. Alizadeh, *Interior point methods in semidefinite programming with applications to combinatorial optimization*, SIAM J. Optim., 5 (1995), pp. 13-51.

[2] R.A. Brualdi, S.V. Prater and H. Schneider, *The diagonal equivalence of a nonnegative matrix to a stochastic matrix*, J. Math. Ann. Appl, 16 (1966), pp. 31-50.

[3] G.B. Dantzig, *Linear Programming and Extensions*, Princeton University Press, Princeton, New Jersey, 1963.

[4] J. Faraut and A. Koranyi, *Analysis on symmetric cones*, Oxford University Press, New York, 1994.

[5] A.V. Fiacco and G.P. McCormick, *Nonlinear Programming*, John Wiley, New York, 1968.

[6] P. Gill, W. Murray, and M. Wright, *Practical Optimization*, Academic Press, London and New York, 1986.

[7] P. Gill, W. Murray, M. Saunders, J. Tomlin, and M. Wright, *On projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method*, Math. Programming, 36 (1986), pp. 183-209.

[8] D. Goldfarb and D.X. Shaw, *On the complexity of a class of projective interior point algorithms*, Math. of Oper. Res., 20 (1995), pp. 116-134.

[9] C.C. Gonzaga, *Path-following methods for linear programming*, SIAM Review, 34 (1992), pp. 167-224.

[10] P. Gordan, *Ueber die Auflosung linearer Gleichungen mit reellen Coefficienten*, Mathematische Annalen (1873), pp. 23-28.

[11] O. Güler, *Barrier functions in interior point methods*, Math. Oper. Res., 21 (1996), pp. 860-885.

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

[13] 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.

[14] B. Kalantari, *Derivation of a generalized and strengthened Gordan theorem from generalized Karmarkar potential and logarithmic barrier functions*. Technical Report LCSR-TR-121, Department of Computer Science, Rutgers University, New Brunswick, NJ, 1989.

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

[16] 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.

[17] 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.

[18] B. Kalantari and L. Khachiyan, *On the rate of convergence of deterministic and randomized RAS matrix scaling method*, Oper. Res. Lett., 14 (1993), pp. 237-244.

[19] B. Kalantari and L. Khachiyan, *On the complexity of nonnegative-matrix scaling*, Linear Algebra and its Applications, 240 (1996), pp. 87-103.

[20] 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.

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

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

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

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

[25] M. Koecher, *Positivitäbereiche im* $\Re^n$, Amer. Jour. Math., 79 (1957), pp. 575-596.

[26] L. Lovász and A. Schrijver, *Cones of matrices and set-functions, and 0-1 optimization*, SIAM J. of Optim., 1 (1991), pp. 166-190.

[27] D.G. Luenberger, *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, Menlo Park, California, second edition, 1984.

[28] O.L. Mangasarian, *Nonlinear Programming, Classics in Applied Mathematics*, 10, SIAM, Philadelphia, PA, 1994.

[29] A.W. Marshall and I. Olkin, *Scaling of matrices to achieve specified row and column sums*, Numer. Math, 12 (1968), pp. 83-90.

[30] K.G. Murty, *Linear Complementarity, Linear and Nonlinear Programming*, Heldermann-Verlag, West Berlin, 1988.

[31] Y. Nesterov and A.S. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, PA, 1994.
[32] Y.E. Nesterov and M.J. Todd, *Self-scaled barriers and interior-point methods for convex programming*, Math. Oper. Res., 22 (1997), pp. 1-42.
[33] M.L. Overton and R.S. Womersely, *Optimality conditions and duality for minimizing sums of the largest eigenvalues of symmetric matrices*, Math. Programming, 62 (1993), pp. 321-357.
[34] J. Renegar, *A polynomial-time algorithm based on Newton's method for linear programming*, Math. Programming, 40 (1988), pp. 59-93.
[35] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ., 1969.
[36] O. Rothaus, *Domain of positivity*, Abh. Math. Sem. Univ. Hamburg, 24 (1960), pp. 189-235.
[37] A. Schrijver, *Theory of Linear and Integer Programming*, John Wiley and Sons, New York, 1986.
[38] R. Sinkhorn, *A relationship between arbitrary positive matrices and doubly stochastic matrices*, Ann. Math. Statist., 35 (1964), pp. 876-879.
[39] Y. Ye and E. Tse, *An extension of Karmarkar's projective algorithm for convex quadratic programming*, Math. Programming, 44 (1989), pp. 159-179.
[40] L. Vandenberghe and S. Boyd, *Semidefinite programming*, *SIAM Review*, 38 (1996), pp. 49-95.
[41] E.B. Vinberg, *The theory of convex homogeneous cones*, Trans. Moscow Math. Soc., 12 (1963), pp. 340-403.
[42] E. Zeidler, *Applied Functional Analysis: Applications to Mathematical Physics Applied Mathematical Sciences*, Vol. 108, Springer-Verlag, New York, 1995.
[43] E. Zeidler, *Applied Functional Analysis: Main Principles and Their Applications Applied Mathematical Sciences*, Vol. 109, Springer-Verlag, New York, 1995.