Potential-Reduction Methods in Mathematical Programming

Michael J. Todd*

January 1995, revised October 1995

*School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853. E-mail: miketodd@cs.cornell.edu. URL: http://www.orie.cornell.edu/~miketodd/todd.html. Research supported in part by NSF, AFOSR and ONR through NSF Grant DMS-8920550. Some of this work was done while the author was on a sabbatical leave from Cornell University visiting the Department of Mathematics at the University of Washington. Partial support from the department and from NSF Grants CCR-9103804 and DMS-9303772 is gratefully acknowledged.

Abstract

We provide a survey of interior-point methods for linear programming and its extensions that are based on reducing a suitable potential function at each iteration. We give a fairly complete overview of potential-reduction methods for linear programming, focusing on the possibility of taking long steps and the properties of the barrier function that are necessary for the analysis. We then describe briefly how the methods and results can be extended to certain convex programming problems, following the approach of Nesterov and Todd. We conclude with some open problems.

Key words: linear programming, potential functions, interior-point methods, self-concordant barriers, self-scaled barriers

Running Header: Potential-Reduction Methods

1 Introduction

The aim of this paper is to provide a self-contained overview of interior-point methods for mathematical programming problems that are based on potential reduction.

Interior-point methods for optimization problems defined with equality and inequality constraints generate a sequence of points which satisfy all the equality constraints and satisfy the inequality constraints strictly. Such points are called *strictly feasible.* There are also so-called infeasible-interior-point methods, whose iterates need not satisfy the equality constraints, but we discuss these only briefly. There has been enormous interest in interior-point methods in the last decade, since their introduction in the seminal paper of Karmarkar [20], in particular for linear programming; they combine very attractive theoretical properties with highly competitive practical performance on large-scale problems. It is not our concern here to discuss computational issues; see the excellent survey of Lustig, Marsten, and Shanno [24] and the commentaries following it. For a comprehensive (up to mid 1993) bibliography on interior-point methods, see Kranich [23].

Most recent research in the area has concentrated on path-following methods. (The more computationally-oriented papers discuss numerical methods inspired by such algorithms, but with much longer steps.) These methods are motivated by a desire to follow, more or less closely, the so-called central path to an optimal solution. In contrast, we will study the class of potential-reduction methods, which have received less attention in the past few years. Our reasons for this focus are several.

The first is historical. Karmarkar's original algorithm was driven by a potential function, and much of the ensuing research followed this lead. We feel that this provides strong motivation for studying this class of methods. Indeed, our purpose here is partly pedagogical, and we believe these earlier methods provide useful perspective on later path-following approaches. However, a historical justification alone would not be too compelling.

As we mentioned above, practical methods inspired by path-following approaches eliminate most of the path-following restrictions, and this leads to an unfortunate gap between theory and practice which is only now being slowly bridged. On the other hand, potential-reduction methods bring theory and practice much closer, since their choices of search direction and step size as well as their analysis are all driven by a certain potential or merit function.

Two natural questions motivate many interior-point methods: how "good" is a given strictly feasible point, and how do we "improve" such a point. Thus we seek an appropriate merit function and ways to decrease it. For simplex methods, and other related feasible-point methods, the objective function alone is used to measure how good a point is, and the constraints only come in when we try to improve it. In interior-point methods, the view is taken that it may be worthwhile to sacrifice the objective function value in order to gain "centrality"; so a measure of distance from the constraint boundaries is combined with the objective function in a certain way.

Potential-reduction methods are in many respects the purest expression of this perspective. A single function, the potential function, is used to measure the quality of a point and to decide how to improve it, and the sole goal of each iteration is to decrease this function as much as possible. A suitable decrease leads to a corresponding *complexity bound* – a bound on the number of iterations necessary to attain a certain accuracy for the optimization problem. We will describe particular algorithms that ensure that an appropriate decrease can be obtained, but it is important to note that *any method* of updating the iterates that guarantees such a decrease will lead to corresponding results. This gives great flexibility to methods based on potential reduction.

Another important justification for our focus on such methods is that they can be extended nicely to certain nonlinear convex programming problems, including for instance optimizing over the cone of positive semidefinite matrices. We will discuss this further below.

A last, practical reason for concentrating on potential-reduction methods is that Gonzaga has provided a state-of-the-art review of path-following algorithms up to 1992 in [16]. (The reader may also like to consult the earlier elementary treatment in Goldfarb and Todd [13].)

As we mentioned above, most of our discussion focuses on algorithms that generate sequences of strictly feasible solutions, and these methods clearly need such a solution to start; on the other hand, practical problems rarely provide a convenient initial point. Infeasible-interior-point methods give one solution to this difficulty, but these are usually path-following algorithms. There are primal-only potential-reduction algorithms that deal with this issue (see Anstreicher [2, 3], Freund [11] and [38, 39]), but these methods seem not as efficient as primal-dual algorithms. Recently, however, Ye, Todd, and Mizuno [49] have shown how to construct a homogeneous self-dual problem with a known strictly feasible initial point, whose solution yields solutions to the original linear programming problem and its dual, and (a slight modification of) the primal-dual potential-reduction algorithm of Kojima, Mizuno, and Yoshise [22] can be applied to this problem.

The paper is organized as follows. In Section 2, we formally state the linear programming problems with which we are most concerned and establish some preliminary results. Then Section 3 introduces the primal potential function and describes the important properties that underlie its use in an interior-point method, without specifying any particular algorithm. The following section describes two such methods, the primal-only potential-reduction methods of Karmarkar [20] and Gonzaga [14]. Our development stresses the properties of the barrier that allow the analysis to go through and the possibility of taking long steps. In Section 5, we provide a corresponding definition and justification for the primal-dual potential function, and this leads to the discussion of the potential-reduction methods of Todd and Ye [41], Ye [46], and Kojima, Mizuno, and Yoshise [22] in Section 6.

Section 7 considers two extensions of the latter algorithm, to monotone linear complementarity problems and to an infeasible-interior-point method. Section 8 generalizes the previous sections to certain convex programming problems, in particular those in conic form. We describe the work of Nesterov and Nemirovskii [33] on potential-reduction methods for such problems, but concentrate on the approach of Nesterov and Todd [34], which requires the cone involved to have a special kind of self-concordant barrier function, a so-called self-scaled barrier. This special class of problems still contains many of practical significance, and also allows the analysis of long-step algorithms and those that use symmetric primal-dual scaling. While we do not go into details, our development in the previous sections was designed to make it easy to see how the analysis extends to this more general setting; thus the material in this section provides an introduction to [34] which demonstrates the importance of certain key properties of the barrier function. We also note that the analysis of the symmetric primal-dual potential-reduction method for the monotone linear complementarity problem extends to such problems defined on self-scaled cones.

The final section gives a brief description of some open problems related to potential-reduction methods.

As mentioned above, part of the motivation for this paper is pedagogical, so let me briefly describe how this material can be incorporated in a course in interior-point methods for graduate students. I like to start with a description of the affine-scaling algorithm, and then thinking of the step as along a steepest descent direction with respect to some norm leads to the logarithmic barrier, whose Hessian provides just the right norm. After stating some convergence results for affine-scaling, I come back to the barrier function and establish the basic results of Section 2 as well as existence, smoothness, and convergence of the central path. A key insight is that being close to the primal central path yields a dual feasible solution. Then I discuss primal potential-reduction methods, first with the assumption that the optimal value is known and zero. This leads to a discussion of what the methods would do if the optimal value is not known, but a lower bound on it is given, and how never updating the lower bound would lead to a point near the central path, and hence a way to update the bound; then the general form of the methods can be presented. I usually then turn to path-following methods, either primal and primal-dual or just the latter, and the desire to remove path-following restrictions leads back to primal-dual potential-reduction methods. The course concludes with discussions of infeasible-interior-point methods and extensions to convex programming using self-concordance.

We end this introduction with a brief mention of some topics in potential-function methods which we could not cover. Bayer and Lagarias [6] and de Ghellinck and Vial [12] show connections between Karmarkar's method and Newton's method. Primal potential-function methods that handle the case where no initial strictly feasible solution is known include those of Anstreicher [2, 3], Freund [11], and [38, 39]. Further properties of the potential function are given by Imai [18], and reduced-complexity versions of methods using partial updating of the projection matrix are described by Bosch and Anstreicher [8] and Mizuno [26]. Trajectories associated with Karmarkar's algorithm are studied in Monteiro [29]. Extensions of the methods described here are provided by Kojima, Megiddo, and Ye [21] for some non-monotone linear complementarity problems, and by Ye [47] for large-scale problems requiring column generation. Obtaining lower bounds on the required number of iterations is the subject of Anstreicher [4], Ji and Ye [19], Powell [36], and Bertsimas and Luo [7]. Finally, Mizuno and Nagasawa [28] and Tunçel [43] give interesting algorithms combining affine-scaling and potential functions. This is of course only a very partial list; again we refer to the bibliography [23].

Our notation is fairly standard, except that we use subscripts and not superscripts to indicate sequences, to avoid confusion with powers. For a vector v, v^α denotes the vector whose components are the α-th powers of those of v. We try to avoid reference to specific components of a vector as much as possible, but it is necessary at times: $v^{(j)}$ denotes the jth component of v. We use an upper case roman letter (e.g., V) to denote the diagonal matrix whose diagonal entries contain the components of the corresponding lower case letter (e.g., v), and e to denote the vector of ones of appropriate dimension, so that $Ve = v$. Finally, we often study how to improve a current or base point, which we denote by x_- or (x_-, s_-), to be read x-bar and s-bar; when we decide on the step to be taken, the next point is denoted x_+ or (x_+, s_+), and then it is natural to think of x_- and s_- as x-minus and s-minus.

Much of the development here was inspired by the analysis of long steps due to Yurii Nesterov that appears in [34], and I am very grateful for his insights.

2 Preliminaries

For most of this paper we are concerned with the linear programming problem, which we write in standard form as

$$
\begin{array}{llrcl}
(P) & \min_x & c^T x & & \\
 & & Ax & = & b \\
 & & x & \geq & 0,
\end{array}
$$

where A is a real $m \times n$ matrix, $b \in I\!R^m$, $c \in I\!R^n$, and $x \in I\!R^n$. The dual problem can be written as

$$
\begin{array}{llrcrcl}
(D) & \max_{y,s} & b^T y & & & & \\
 & & A^T y & + & s & = & c \\
 & & & & s & \geq & 0,
\end{array}
$$

where $y \in I\!R^m$ and we have included an explicit slack vector $s \in I\!R^n$.

We may view either (P) or (D) as the primal and the other as the dual. If one is interested in a problem with inequality constraints in unrestricted variables, one can put primary emphasis on (D). Below we concentrate on algorithms for (P) alone or for (P) and (D) together; however, we also indicate how to convert an algorithm for a problem in form (P) to one for (D).

We will denote by $\mathcal{F}(P)$ and $\mathcal{F}^0(P)$ the feasible and strictly feasible region for (P), and by $v(P)$ its optimal value:

$$\begin{aligned} \mathcal{F}(P) &:= \{x : Ax = b,\ x \geq 0\}, \\ \mathcal{F}^0(P) &:= \{x : Ax = b,\ x > 0\}, \\ v(P) &:= \inf\{c^T x : x \in \mathcal{F}(P)\}. \end{aligned}$$

We use similar notation for (D):

$$\begin{aligned} \mathcal{F}(D) &:= \{(y,s) : A^T y + s = c,\ s \geq 0\}, \\ \mathcal{F}^0(D) &:= \{(y,s) : A^T y + s = c,\ s > 0\}, \\ v(D) &:= \sup\{b^T y : (y,s) \in \mathcal{F}(D)\}. \end{aligned}$$

Here $u > 0$ means each component of u is positive.

While (D) is usually thought of as a problem involving the dual variables y, and it seems that s has been introduced purely for convenience, it will be useful for us to think of (D) as primarily a problem in the dual slack s. For this, it is helpful to make the blanket assumption that A has full row rank m. This is without real loss of generality. Consider first the primal problem (P). If the rows of A are linearly dependent, then either (P) is infeasible or redundant linear equations can be removed sequentially until the resulting A does have full row rank. Moreover this process can be carried out easily. Next consider (D). If the columns of A^T exhibit a linear dependence, say $A^T z = 0$, and $b^T z \neq 0$, then it is easy to see that (D), if it is feasible, must be unbounded. If $b^T z = 0$ for every such z, then we may eliminate any one of the columns corresponding to a nonzero component of z without affecting (D). Continuing in this way yields an equivalent problem where A has full row rank.

With this assumption we proceed to write (D) as a problem in s alone. Let G be a real $(n-m) \times n$ matrix whose rows form a basis for the nullspace of A. Then if $A^T y + s = c$, $Gs = GA^T y + Gs = Gc$, and if $Gs = Gc$, then $c - s$ lies in the nullspace of G, hence the row space of A, and thus $A^T y + s = c$

for some y; since A has full row rank, y is in fact unique. Hence there is a one-to-one correspondence between $\mathcal{F}(D)$ and $\widetilde{\mathcal{F}}(D)$ and between $\mathcal{F}^0(D)$ and $\widetilde{\mathcal{F}}^0(D)$, where

$$\begin{aligned}
\widetilde{\mathcal{F}}(D) &:= \{s : \exists y \text{ such that } (y, s) \in \mathcal{F}(D)\}, \\
\widetilde{\mathcal{F}}^0(D) &:= \{s : \exists y \text{ such that } (y, s) \in \mathcal{F}^0(D)\}.
\end{aligned}$$

Next, if $\tilde{x}$ satisfies $A\tilde{x} = b$ (such an $\tilde{x}$ exists by our assumption on A), then $b^T y = (A\tilde{x})^T y = \tilde{x}^T c - \tilde{x}^T s$ for any $(y, s) \in \mathcal{F}(D)$. It follows that (D) is equivalent to

$$(\widetilde{D}) \qquad \begin{array}{lrcl} \min & \tilde{x}^T s & & \\ & Gs & = & Gc \\ & s & \geq & 0, \end{array}$$

in the sense that there is a one-to-one correspondence between their feasible and strictly feasible regions which preserves optimality. Moreover, the argument above shows that $\mathcal{F}(\widetilde{D}) = \widetilde{\mathcal{F}}(D)$, $\mathcal{F}^0(\widetilde{D}) = \widetilde{\mathcal{F}}^0(D)$.

Hence any problem in the form (D) can be re-expressed as an equivalent problem $(\widetilde{D})$ in standard form. We rarely need to carry out this process explicitly. However, it provides a prescription for converting an algorithm for a standard-form problem (P) into one for (D): consider the steps of the algorithm as applied to the standard-form problem $(\widetilde{D})$, and then perform the requisite algebraic manipulation to re-express these steps in terms of the original problem (D). Because of this general recipe, we will consider only algorithms for (P) alone, or for (P) and (D) together, called primal-only and primal-dual methods, respectively. A primal-dual method is called *symmetric* iff it yields the same iterates when applied to (P) and (D) as when applied to $(\widetilde{D})$ and its dual.

We are concerned with interior-point methods for linear programming and some of its extensions, and almost all of the time, with feasible-interior-point methods, which deal with strictly feasible iterates. As is conventional, we call such methods just interior-point methods, and use the term infeasible-interior-point algorithm for one that generates possibly infeasible iterates with x, s, or x and s in the interior of the nonnegative orthant. We distinguish primal-only methods (generating a sequence $\{x_k\}$ in $\mathcal{F}^0(P)$), dual-only methods (generating $\{(y_k, s_k)\}$ in $\mathcal{F}^0(D)$), and primal-dual methods (generating a sequence $\{(x_k, y_k, s_k)\}$ in $\mathcal{F}^0(P) \times \mathcal{F}^0(D)$). For the reasons given above, we confine ourselves to primal-only and primal-dual algorithms. We would like the sequences to satisfy $c^T x_k \to v(P)$ for a primal-only method and $c^T x_k \to v(P) = v(D) \leftarrow b^T y_k$ for a primal-dual method.

Let x and (y, s) be feasible for (P) and (D) respectively. Then

$$c^T x - b^T y = (A^T y + s)^T x - (Ax)^T y = x^T s \geq 0,$$

which is the weak duality property that all feasible objective values for (P) are at least as great as those for (D). Moreover, strong duality implies that if both (P) and (D) have feasible solutions, then both have optimal solutions, say x_* and (y_*, s_*), such that $c^T x_* = v(P) = v(D) = b^T y_*$ and $x_*^T s_* = 0$, so there is no duality gap. Moreover, $x_* \geq 0$, $s_* \geq 0$, and $x_*^T s_* = 0$ imply that $X_* S_* e = 0$.

The equation $X_* S_* e = 0$ is equivalent to the assertion that, of every pair of corresponding components of x_* and s_*, at least one is zero. This is the condition of complementary slackness. A theorem of A. W. Tucker states that, if (P) and (D) have optimal solutions, these can be chosen so that "at least one" can be replaced by "exactly one" above; such solutions are called strictly complementary.

An easy consequence of complementary slackness is

Proposition 2.1 *If c does not lie in the row space of A, then $c^T x > v(P)$ for every $x \in \mathcal{F}^0(P)$.*

Proof. Suppose that $c^T x = v(P)$ for some $x \in \mathcal{F}^0(P)$. Then (P), and hence (D), has an optimal solution. Complementary slackness implies that $s_* = 0$ for any (y_*, s_*) optimal in (D), whence $c = A^T y_*$. □

Our final definition is the set of optimal solutions of a problem:

$$\mathcal{S}(P) := \{x \in \mathcal{F}(P) : c^T x = v(P)\}.$$

Theorem 2.1 *Suppose that $\mathcal{F}(P)$ is nonempty. Then the following are equivalent:*

a) $\mathcal{F}^0(D)$ *is nonempty;*
b) For all $\gamma \in I\!R$, $\{x \in \mathcal{F}(P) : c^T x \leq \gamma\}$ *is bounded;*
c) $\mathcal{S}(P)$ *is nonempty and bounded; and*
d) For some $\gamma \in I\!R$, $\{x \in \mathcal{F}(P) : c^T x \leq \gamma\}$ *is nonempty and bounded.*

Proof.

a) $\Rightarrow$ b): Suppose $(y_0, s_0) \in \mathcal{F}^0(D)$. Then, using duality, $x \in \mathcal{F}(P)$ and $c^T x \leq \gamma$ imply that $s_0^T x = c^T x - b^T y_0 \leq \gamma - b^T y_0$. Hence such x's must lie in the bounded set $\{x : s_0^T x \leq \gamma - b^T y_0,\ x \geq 0\} \subseteq \{x : 0 \leq x \leq (\gamma - b^T y_0)s_0^{-1}\}$. (Recall that s_0^{-1} stands for the positive vector whose components are the reciprocals of those of s_0.)

b) $\Rightarrow$ c): Choosing $\gamma = c^T x_0$ for some $x_0 \in \mathcal{F}(P)$ shows that (P) is equivalent to minimizing $c^T x$ over the nonempty compact set $\{x \in \mathcal{F}(P) : c^T x \leq \gamma\}$, so that it has an optimal solution; moreover, all optimal solutions lie in this compact set.

c) $\Rightarrow$ d): This is immediate by choosing γ equal to $v(P)$.

d) $\Rightarrow$ a): Suppose $\{x \in \mathcal{F}(P) : c^T x \leq \gamma\}$ is nonempty and bounded. Choose some $s_0 > 0$. Then the problem $\min\{c^T x : Ax = 0,\ s_0^T x = 1,\ x \geq 0\}$ is either infeasible or has positive optimal value. Thus its dual, $\max\{\zeta : A^T y + s_0 \zeta + s = c,\ s \geq 0\}$, which clearly has a feasible solution (take $y = 0$ and ζ sufficiently negative), has a feasible solution with ζ positive. This yields a point in $\mathcal{F}^0(D)$. $\square$

3 The Primal-Only Potential Function

In this section we describe a potential function involving only the primal variables and discuss its properties. As we mentioned in the introduction, a potential function is a merit function for strictly feasible points that combines a measure of the objective function, or closeness to optimality, with one of the "centrality," or distance to the boundary of the feasible region, for a strictly feasible point.

In the primal setting, any monotonic function of $c^T x$ provides a measure for optimality. The measure of strict feasibility is given by the logarithmic barrier function

$$F(x) := -\Sigma_j \ln x^{(j)}. \tag{3.1}$$

This is a convex function defined on the interior of the nonnegative orthant which approaches $+\infty$ if x tends to a point in the boundary of the orthant. The reason for choosing this particular function, as opposed to say $\Sigma_j\, 1/x^{(j)}$, will emerge as we discover the very attractive properties satisfied by F. Roughly, we can bound the errors in Taylor approximations to F and ∇F very nicely, because its Hessian satisfies a relative Lipschitz condition that limits how fast it can vary – in the terminology of Nesterov and Nemirovskii [33], it is a self-concordant barrier function for $\mathbb{R}^n_+$. As far as possible, we will analyze our algorithms in terms of these basic properties of F rather than its particular form, so that the same analysis will apply in other settings.

Notice that F only takes into account the nonnegativity constraints, since the equality constraints $Ax = b$ are easy to handle by restricting primal directions to the nullspace of A. An appropriate barrier function for (D) is $F(s)$, or, in terms of y alone,

$$-\,\Sigma_j \ln(c - A^T y)^{(j)}. \tag{3.2}$$

One immediately obvious advantage of the logarithmic barrier function, as opposed to alternatives like the inverse barrier mentioned above, is its insensitivity to scalings. Indeed, if the primal variables are measured in different units, so that x undergoes a diagonal scaling, i.e., premultiplication by a positive definite diagonal matrix, $F(x)$ only changes by an additive constant. This is also the result of a scaling of the columns of c^T and A on the function in (3.2).

Perhaps the most natural combination of the objective function with the barrier is the *penalized* function

$$\psi_\mu(x) := \frac{1}{\mu} c^T x + F(x), \tag{3.3}$$

whose minimizers in $\mathcal{F}^0(P)$, as μ ranges over positive scalars, define the primal central path. To approach optimality, μ must tend to 0. In order to construct a "single" merit function, we devise instead a combination that automatically increases the importance of the objective function (like μ tending to 0) as optimality is approached. The function depends on a parameter $\rho \geq n$ and also a lower bound $\zeta \leq \zeta_* := v(P)$. Thus, for $x \in \mathcal{F}^0(P)$, define

$$\phi(x;\zeta) := \phi_\rho(x;\zeta) := \rho \ln(c^T x - \zeta) + F(x), \tag{3.4}$$

called the *primal potential* function. This function was first introduced by Karmarkar [20] with $\rho = n$ and $\zeta = \zeta_*$ for a specially formulated problem, but the more general form above is necessary for general formulations of the problem or if the optimal value is unknown. When we use this function in an algorithm, ρ is fixed (hence often omitted from the notation), the lower bound ζ is updated at some but possibly not all iterations, and the iterate x is usually updated at every iteration.

Note that ϕ is not well-defined at $x \in \mathcal{F}^0(P)$ iff $c^T x - \zeta = 0$, i.e., $c^T x = \zeta = \zeta_*$. According to Proposition 2.1, this can only happen if c lies in the row space of A, which implies that all feasible solutions to (P) are optimal. This case can be detected at the beginning of any primal-only interior-point method and the method then stops with an optimal solution. Henceforth, in this section and the next, we assume that $c^T x > \zeta_*$ and thus ϕ is well-defined at all $x \in \mathcal{F}^0(P)$.

Now we give some properties of ϕ_ρ. Algorithms using it will be described in the next section. First, we note that

$$\nabla F(x) = -x^{-1}, \quad \nabla^2 F(x) = X^{-2}, \tag{3.5}$$

while

$$\nabla \psi_\mu(x) = c/\mu - x^{-1}, \quad \nabla^2 \psi_\mu(x) = X^{-2}, \tag{3.6}$$

so that F and ψ_μ are strictly convex. In contrast,

$$\nabla_x \phi_\rho(x;\zeta) = \frac{\rho}{c^T x - \zeta} c + \nabla F(x) = \frac{\rho}{c^T x - \zeta} c - x^{-1}, \tag{3.7}$$

$$\nabla^2_{xx} \phi_\rho(x;\zeta) = -\frac{\rho}{(c^T x - \zeta)^2} cc^T + \nabla^2 F(x) = -\frac{\rho}{(c^T x - \zeta)^2} cc^T + X^{-2}. \tag{3.8}$$

Thus ϕ is in general not convex in x, but a positive definite part of its Hessian coincides with that of F and of ψ_μ. Also, the gradients of ϕ and ψ_μ coincide for certain values of the parameters, so we can expect that steepest descent steps or pseudo-Newton steps (where we use the positive definite part of the Hessian of ϕ) will correspond for appropriate values of these parameters. Hence there will be a close relation between potential-reduction algorithms and central path-following methods, although their motivations are in many respects very different.

Let us give three related examples in $I\!R^2$ that suggest the relevant properties of ϕ and their limitations.

Examples.

a) Consider $\min\{x^{(1)} : x^{(1)} - x^{(2)} = -1,\ x \geq 0\}$. The optimal solution is $x_* = (0,1)^T$ with value 0. Let $\zeta = \zeta_* = 0$, and consider any strictly feasible solution $x(\tau) := (\tau, 1+\tau)^T$, $\tau > 0$. We then have (for $\rho \geq 2$)

$$\phi_\rho(x(\tau);\zeta) = \rho \ln \tau - \ln \tau - \ln(1+\tau).$$

Then:

(i) $\phi_\rho(x(\tau);\zeta) \to -\infty$ as $\tau \to 0$. It is possible to drive ϕ to $-\infty$.

(ii) If $\phi_\rho(x(\tau);\zeta) \to -\infty$, $\tau \to 0$. Driving ϕ to $-\infty$ forces x to approach optimality.

(iii) $\{x \in \mathcal{F}^0(P) : \phi_\rho(x;\zeta) \leq \delta\}$ is bounded for all δ if $\rho > n = 2$. For $\rho = 2$, this level set is unbounded for $\delta \geq 0$, but bounded for negative δ.

b) Now consider $\min\{x^{(1)} : x^{(1)} - x^{(2)} = 0,\ x \geq 0\}$, and again let $\zeta = \zeta_* = 0$ and $x(\tau)$ denote an arbitrary strictly feasible solution, now $(\tau,\tau)^T$. Then properties (i)–(iii) still hold as long as $\rho > n = 2$, but for $\rho = 2$, $\phi_\rho(x(\tau);\zeta) \equiv 0$, so even (i) fails.

c) Finally, consider $\min\{x^{(1)} : x \geq 0\}$. For any $\rho \geq 2$, and any $\zeta \leq \zeta_* = 0$, let $x(\kappa) := (\kappa, \kappa^\rho)^T$ for $\kappa > 0$, a strictly feasible point, and note that

$$\begin{aligned}
\phi_\rho(x(\kappa);\zeta) &= \rho \ln(\kappa - \zeta) - \ln \kappa - \ln \kappa^\rho \\
&= \rho \ln\left(\frac{\kappa - \zeta}{\kappa}\right) - \ln \kappa,
\end{aligned}$$

which tends to $-\infty$ as $\kappa \to \infty$. Thus here the potential function approaches $-\infty$ for a sequence of strictly feasible solutions which diverges and diverges from optimality. So properties (ii) and (iii) fail.

These examples show that we may need to choose $\rho > n$ and/or restrict our problems so that they have bounded sets of optimal solutions. Indeed, we have

Theorem 3.1 *Suppose $\mathcal{F}^0(P)$ and $\mathcal{S}(P)$ are nonempty. Then, as long as $\rho > n$ or $b \neq 0$, $\phi_\rho(x;\zeta)$ is unbounded below as a function of x and ζ.*

Proof. Let us choose $x_0 \in \mathcal{F}^0(P)$ and $x_* \in \mathcal{S}(P)$, set $\zeta = \zeta_* = v(P) = c^T x_*$, and consider $\phi_\rho(x;\zeta)$ on the line segment joining x_* and x_0. Let $x(\lambda) := \lambda x_0 + (1-\lambda)x_*$ for $\lambda \in (0,1]$, so that

$$\begin{aligned}
\phi_\rho(x(\lambda);\zeta) &= \rho \ln(\lambda c^T x_0 + (1-\lambda)c^T x_* - c^T x_*) + F(x(\lambda)) \\
&= \rho \ln \lambda + \rho \ln(c^T x_0 - c^T x_*) + F(x(\lambda)) \\
&= (\rho - n)\ln \lambda + \rho \ln(c^T x_0 - c^T x_*) + n \ln \lambda \\
&\qquad + F(\lambda x_0 + (1-\lambda)x_*) \\
&= (\rho - n)\ln \lambda + \rho \ln(c^T x_0 - c^T x_*) + F(x_0 + \frac{1-\lambda}{\lambda} x_*),
\end{aligned}$$

where we have used the fact, easily derived from its definition, that

$$F(\lambda x) = F(x) - n \ln \lambda. \tag{3.9}$$

Now let $\lambda \downarrow 0$. Then if $\rho > n$, $(\rho - n)\ln\lambda \to -\infty$ while $F(x_0 + (1-\lambda)\lambda^{-1}x_*) \leq F(x_0)$, since $x_* \geq 0$, so that $\phi_\rho \to -\infty$. Alternatively, if $\rho = n$ but $b \neq 0$, then $x_* \in \mathcal{F}(P)$ implies $x_* \neq 0$, so $F(x_0 + (1-\lambda)\lambda^{-1}x_*) \to -\infty$ as $\lambda \downarrow 0$. Thus again $\phi_\rho \to -\infty$. $\square$

Theorem 3.2 *Suppose $\mathcal{F}^0(P)$ is nonempty and $\mathcal{S}(P)$ is nonempty and bounded. Then, if $\phi_\rho(x;\zeta) \to -\infty$, for $x \in \mathcal{F}^0(P)$ and $\zeta \leq \zeta_*$, we have $c^T x \to \zeta_*$ and $\zeta \to \zeta_*$. Indeed, there is a constant κ depending only on the data of (P) such that, for $\phi_\rho(x;\zeta)$ sufficiently small, we have*

$$c^T x - \zeta \leq \kappa \exp\{\phi_\rho(x;\zeta)/\rho\}. \tag{3.10}$$

Moreover, for sufficiently small Δ, $\{x \in \mathcal{F}^0(P) : \phi_\rho(x;\zeta) \leq \Delta$ for some $\zeta \leq \zeta_*\}$ is bounded.

Proof. The hypotheses of the theorem imply that $\mathcal{F}^0(D)$ is nonempty by Theorem 2.1. Let us choose $(y_0, s_0) \in \mathcal{F}^0(D)$. We now use an important inequality, which we record for future use as

Proposition 3.1 *For positive vectors x and s in $\mathbb{R}^n$,*

$$n \ln x^T s + F(x) + F(s) \geq n \ln n, \tag{3.11}$$

with equality iff $XSe = \mu e$ for some μ.

Proof of proposition. Note that the concavity of the logarithm implies that $\ln(\tau) \leq \tau - 1$ for any positive τ, with equality iff $\tau = 1$. Hence we have

$$\begin{aligned} n \ln x^T s + F(x) + F(s) - n \ln n &= -\sum_j \ln \frac{x^{(j)} s^{(j)}}{x^T s/n} \\ &\geq -\sum_j \left(\frac{x^{(j)} s^{(j)}}{x^T s/n} - 1 \right) \\ &= -(n-n) = 0. \end{aligned}$$

Moreover, equality holds iff each $\frac{x^{(j)} s^{(j)}}{x^T s/n}$ equals one, or equivalently iff $XSe = \mu e$ for some μ. □

We use this proposition to replace the term $F(x)$ in ϕ_ρ by one involving $x^T s_0$. We have

$$\begin{aligned} \phi_\rho(x;\zeta) &= \rho \ln(c^T x - \zeta) + F(x) \\ &\geq \rho \ln(c^T x - \zeta) - n \ln x^T s_0 - F(s_0) + n \ln n \\ &= \rho \ln(c^T x - \zeta) - n \ln(c^T x - b^T y_0) - F(s_0) + n \ln n \\ &= (\rho - n) \ln(c^T x - \zeta) + n \ln \tfrac{c^T x - \zeta}{c^T x - b^T y_0} - F(s_0) + n \ln n. \end{aligned} \tag{3.12}$$

Now suppose $c^T x \geq 2\zeta_* - b^T y_0 + 1$. Then

$$c^T x - \zeta \geq c^T x - \zeta_* \geq \zeta_* - b^T y_0 + 1$$

and

$$\begin{aligned} c^T x - \zeta &\geq c^T x - \zeta_* = \frac{1}{2}(c^T x - \zeta_*) + \frac{1}{2}(c^T x - \zeta_*) \\ &> \frac{1}{2}(c^T x - \zeta_*) + \frac{1}{2}(\zeta_* - b^T y_0) = \frac{1}{2}(c^T x - b^T y_0). \end{aligned}$$

It follows that

$$\phi_\rho(x;\zeta) > (\rho - n) \ln(\zeta_* - b^T y_0 + 1) + n \ln \frac{1}{2} - F(s_0) + n \ln n =: \Delta_0.$$

Hence if $\phi_\rho(x;\zeta) \leq \Delta_0$, $c^T x < 2\zeta_* - b^T y_0 + 1$ and thus x lies in the bounded set $\{x \in \mathcal{F}^0(P) : c^T x < 2\zeta_* - b^T y_0 + 1\}$ (see Theorem 2.1). This proves the last part of the theorem. Moreover, for $\phi_\rho(x;\zeta) \leq \Delta_0$, (3.12) gives

$$\phi_\rho(x;\zeta) \geq \rho \ln(c^T x - \zeta) - n \ln(2(\zeta_* - b^T y_0) + 1) - F(s_0) + n \ln n,$$

from which by rearranging the terms we obtain (3.10) with

$$\kappa := \exp\left\{\left[n \ln \frac{2(\zeta_* - b^T y_0) + 1}{n} + F(s_0)\right] / \rho\right\} \tag{3.13}$$

as desired. ∎

Theorem 3.2 is the main tool in establishing the complexity of primal potential-reduction algorithms. Indeed, we have

Corollary 3.1 *Suppose we have available some $x_0 \in \mathcal{F}^0(P)$ and $\zeta_0 \leq \zeta_*$ and an algorithm which at each iteration replaces x_k and ζ_k with x_{k+1} and ζ_{k+1} such that*

$$\phi_\rho(x_{k+1}; \zeta_{k+1}) \leq \phi_\rho(x_k; \zeta_k) - \delta \tag{3.14}$$

for some positive absolute constant δ. Then, for sufficiently small $\epsilon > 0$,

$$c^T x_k - \zeta_k \leq \epsilon \tag{3.15}$$

within

$$K = \left\lceil \frac{\rho}{\delta}\left(\ln \kappa + \frac{\phi_\rho(x_0; \zeta_0)}{\rho} + \ln \frac{1}{\epsilon}\right)\right\rceil \tag{3.16}$$

iterations, where κ is given by (3.13).

Proof. Note that $\phi_\rho(x_k; \zeta_k) \leq \phi_\rho(x_0; \zeta_0) - k\delta$ by hypothesis. Substituting this into (3.10) with K as given above yields $c^T x_k - \zeta_k \leq \epsilon$. If ϵ is sufficiently small, K in (3.16) is large enough that $\phi_\rho(x_0; \zeta_0) - K\delta$ is sufficiently small that (3.10) holds. ∎

In (3.16), the first two terms in the brackets depend on the problem (P) and the initialization. We will assume that ϵ is sufficiently small that these two terms are dominated by a constant multiple of the third. Then (3.15), i.e., ϵ-optimality, is achieved within $O(\rho \ln \frac{1}{\epsilon})$ iterations. We call such an algorithm an $O(\rho \ln \frac{1}{\epsilon})$-iteration algorithm loosely while recognizing that other quantities enter into the complexity bound. For example, for a problem in Karmarkar's form (see Section 4.1) with $g = e$ and $x_0 = e/n \in \mathcal{F}^0(P)$, one can show from the above that

$$K = \left\lceil \frac{n}{\delta} \ln \frac{2(c^T x_0 - \zeta_0)}{\epsilon} \right\rceil$$

iterations suffice for Karmarkar's algorithm to achieve ϵ-optimality, while the usual special analysis for this case only eliminates the factor 2.

4 Primal Potential-Reduction Methods

Here we describe two methods based on reducing the primal-only potential function studied in the last section. These are due to Karmarkar [20] and Gonzaga [14]; we stress here the derivation of the search directions, the ability to take long steps, and the properties of the barrier function F necessary for the analysis.

Both algorithms rely on achieving a constant decrease in the potential function at each iteration, and both use $\rho = O(n)$ and are thus $O(n \ln \frac{1}{\epsilon})$-iteration methods.

We assume here (as is natural from the examples and theorems of the previous section) that $\mathcal{F}^0(P)$ is nonempty and $\mathcal{S}(P)$ nonempty and bounded. The algorithms require that we know some initial point $x_0 \in \mathcal{F}^0(P)$ and some initial lower bound $\zeta_0 \leq \zeta_* = v(P)$.

4.1 Karmarkar's Algorithm

Here we assume that the constraints $Ax = b$ take the special form $Bx = 0$, $g^T x = 1$, where $g \geq 0$. Thus the problem and its dual can be written as

$$
\begin{array}{llrcl}
(P) & \min & c^T x & & \\
& & Bx & = & 0 \\
& & g^T x & = & 1 \\
& & x & \geq & 0,
\end{array}
\qquad
\begin{array}{llrrrcl}
(D) & \max & & \zeta & & & \\
& & B^T y & +g\zeta & +s & = & c \\
& & & & s & \geq & 0.
\end{array}
$$

Note that this form is without loss of generality. Indeed, the standard-form problem (P) of Section 2 can be written as

$$
\begin{array}{lrcrcl}
\min & c^T x & & & & \\
& Ax & - & b\tau & = & 0 \\
& & & \tau & = & 1 \\
& x, & & \tau & \geq & 0,
\end{array}
$$

which is of the form of (P) above. For this subsection, (P) and (D) refer to the problems stated above.

Since b $(= (0^T, 1)^T)$ is nonzero, we can choose $\rho = n$ (see Theorem 3.1) in the potential function. Also, for feasible x, $c^T x - \zeta = (c - \zeta g)^T x$. We therefore write our potential function as

$$\phi(x; \zeta) = \phi_n(x; \zeta) = n \ln(c - \zeta g)^T x + F(x), \tag{4.1}$$

which, if we use (3.9) again, can easily be seen to be homogeneous of degree 0 in x. We can therefore define ϕ in this way for all $\zeta \leq \zeta_*$ and all x in $\widehat{\mathcal{F}}^0(P) := \{x : Bx = 0, x > 0\}$, since (with $g \geq 0$) all points in this set are positive scalar multiples of points in $\mathcal{F}^0(P)$.

Suppose we have a current point $x_- \in \mathcal{F}^0(P)$ and a current lower bound $\zeta_- \leq \zeta_*$. We wish to update x_- to x_+ and ζ_- to ζ_+ (possibly equal to ζ_-) to achieve a constant reduction in ϕ. Moreover, it suffices to have x_+ in $\widehat{\mathcal{F}}^0(P)$, for then $\tilde{x}_+ := x_+/g^T x_+$ lies in $\mathcal{F}^0(P)$ (note $g^T x_+ > 0$) and yields the same reduction.

We obtain such a reduction by taking a steepest descent step for ϕ in the nullspace of B with respect to the norm defined by the positive definite matrix $\nabla^2 F(x_-)$. Equivalently, we take a pseudo-Newton step for ϕ where its Hessian is approximated by its positive definite part $\nabla^2 F(x_-)$. Thus our step will be a positive multiple of the solution to the direction-finding problem

$$(DFP) \qquad \begin{array}{rcl} \min\limits_p \quad u^T p & + & \frac{1}{2} p^T \nabla^2 F(x_-) p \\ Bp & = & 0, \end{array}$$

for $u = \nabla_x \phi(x_-; \zeta)$, where ζ is a possibly updated lower bound. Note that, even for the steepest descent motivation, we put the second-order term in the objective rather than as a constraint with an arbitrary right-hand side, since we want to avoid any suggestion that our step will be limited by $(p^T \nabla^2 F(x_-) p)^{\frac{1}{2}} \leq \alpha$ for some α of order 1; this would be considered a short step and we are interested in long steps.

An interesting alternative to (DFP) was proposed by de Ghellinck and Vial [12]. We have seen that the normalization $g^T x = 1$ can be removed in our search for x_+. Alternatively, the normalization can be changed to another. De Ghellinck and Vial use the normalization $(c - \zeta g)^T x = (c - \zeta g)^T x_-$. Note that, with this restriction, ϕ_n of (4.1) differs by a constant from the convex function F. Thus they take a step in the Newton direction for F in the nullspace of B intersected with that of $(c - \zeta g)^T$. It turns out that this method yields the same iterates as we will derive, given corresponding step sizes. See also Bayer and Lagarias [6].

Before we continue, we remark that our derivation of (DFP) and our subsequent analysis seem very different from the usual motivation and analysis using projective transformations. Here we briefly digress to mention the relationship. A full description of this geometric view of Karmarkar's algorithm can be found in Anstreicher [1] for instance; see also [13].

This geometric picture corresponds to yet another normalization, defined by $-\nabla F(x_-)^T x = (x_-^{-1})^T x = n$, so that $\mathcal{F}^0(P)$ is replaced by $\widetilde{\mathcal{F}}^0(P) := \{x : Bx = 0, -\nabla F(x_-)^T x = n, x > 0\}$. It is clear that this set contains our current iterate x_-, and that in fact x_- minimizes F over $\widetilde{\mathcal{F}}^0(P)$. Indeed, x_- minimizes F over the larger simplex $\widetilde{\Delta} := \{x : -\nabla F(x_-)^T x = n, x > 0\}$. So x_- can be viewed as a sort of center of these sets; in fact, it is the barycenter or centroid of $\widetilde{\Delta}$.

Now $\widetilde{\mathcal{F}}^0(P)$ is a projective transform of $\mathcal{F}^0(P)$, where points are related if they lie on a single ray from the origin. We can make a further affine transformation, taking x to $\check{x} := X_-^{-1}x$. This takes $\widetilde{\mathcal{F}}^0(P)$ to the set $\check{\mathcal{F}}^0(P) := \{\check{x} : (BX_-)\check{x} = 0, e^T\check{x} = n, \check{x} > 0\}$, the simplex $\widetilde{\Delta}$ to the standard simplex $\check{\Delta} := \{\check{x} : e^T\check{x} = n, \check{x} > 0\}$, and x_- to its barycenter e. Now the ball around e of radius 1 (in the hyperplane $e^T\check{x} = n$) is contained in $\check{\Delta}$, while the ball of radius n (in the same hyperplane) contains it. The same is true when we intersect the simplex and the balls with the subspace $\{\check{x} : (BX_-)\check{x} = 0\}$. Thus, after the projective transformation, the feasible region is "well-rounded" and the current iterate "well-centered." It therefore seems promising to move in a direction of steepest descent (with respect to the Euclidean norm) in $\check{x}$-space. This corresponds to moving in a direction of steepest descent with respect to the norm defined by the positive definite matrix $\nabla^2 F(x_-)$ in x-space, within $\widetilde{\mathcal{F}}^0(P)$. But it turns out that the steepest descent direction within $\widehat{\mathcal{F}}^0(P)$ automatically remains within $\widetilde{\mathcal{F}}^0(P)$ (see (4.12) below), and this direction is exactly the solution of (DFP) above. Hence this problem with its simple motivation corresponds to using a (Euclidean) steepest descent direction in a projective transformation of the original problem.

The solution to (DFP) will be denoted $p(u)$. Writing the Karush-Kuhn-Tucker (KKT) conditions for the problem, we obtain

$$\begin{array}{rcl} B^T y(u) - \nabla^2 F(x_-)p(u) & = & u \\ Bp(u) & = & 0, \end{array} \tag{4.2}$$

for some (unique, since B has full row rank) vector $y(u)$ of Lagrange multipliers. Since $\nabla^2 F(x_-) = X_-^{-2}$, we can explicitly calculate that

$$p(u) = -X_- P_{BX_-} X_- u, \tag{4.3}$$

where P_M denotes the orthogonal projection onto the nullspace of M. In any case, (4.2) uniquely determines $p(u)$ and shows that it depends linearly on u.

In our case, we can write u as

$$\begin{array}{rcl} \nabla_x \phi(x_-;\zeta) & = & \frac{n}{c^T x_- - \zeta}(c - \zeta g) + \nabla F(x_-) \\ & = & \frac{n}{(c-\zeta g)^T x_-}\left[c - \zeta g + \frac{(c-\zeta g)^T x_-}{n}\nabla F(x_-)\right] \\ & = & \frac{n}{(c-\zeta g)^T x_-}[\tilde{c} - \zeta \tilde{g}], \end{array} \tag{4.4}$$

where

$$\tilde{c} := c + \frac{c^T x_-}{n}\nabla F(x_-), \quad \tilde{g} := g + \frac{g^T x_-}{n}\nabla F(x_-). \tag{4.5}$$

Hence, for any $\zeta \le \zeta_*$, the solution to (DFP) is a positive scalar multiple of

$$p(\tilde{c}) - \zeta p(\tilde{g}). \tag{4.6}$$

We therefore compute separately $p(\tilde{c})$ and $p(\tilde{g})$, so that after possibly updating ζ_- we can compute our search direction.

The update of the lower bound is based on Todd and Burrell [40]. Let us write (4.2) for $u = \tilde{c}$ and $u = \tilde{g}$:

$$\begin{array}{rcl} B^T y(\tilde{c}) - \nabla^2 F(x_-)p(\tilde{c}) & = & c - \dfrac{c^T x_-}{n}\nabla^2 F(x_-)x_-, \\ B^T y(\tilde{g}) - \nabla^2 F(x_-)p(\tilde{g}) & = & g - \dfrac{g^T x_-}{n}\nabla^2 F(x_-)x_-, \end{array}$$

where we have used

$$\nabla^2 F(x_-)x_- = X_-^{-2}x_- = x_-^{-1} = -\nabla F(x_-), \tag{4.7}$$

so that

$$B^T(y(\tilde{c}) - \zeta y(\tilde{g})) + g\zeta + \nabla^2 F(x_-)\left[\frac{(c-\zeta g)^T x_-}{n}x_- - (p(\tilde{c}) - \zeta p(\tilde{g}))\right] = c. \tag{4.8}$$

Note that this equation expresses precisely that we have a dual feasible solution of value ζ, as long as

$$x(\zeta) := \frac{(c-\zeta g)^T x_-}{n}x_- - (p(\tilde{c}) - \zeta p(\tilde{g})) \ge 0, \tag{4.9}$$

which implies that $s(\zeta) := \nabla^2 F(x_-)x(\zeta) \geq 0$. Thus we choose $\check{\zeta}$ as the maximum ζ so that (4.9) holds (or $-\infty$ if there is no such ζ), and update ζ_- to

$$\zeta_+ := \max\{\zeta_-, \check{\zeta}\}, \quad \check{\zeta} := \max\{\zeta : x(\zeta) \geq 0\}. \tag{4.10}$$

Note that, as a result of our choice, $x(\zeta_+) \not> 0$, i.e., it has some nonpositive component.

A few remarks should be made. First,

$$\nabla F(x_-)^T x_- = -(x_-^{-1})^T x_- = -n, \tag{4.11}$$

so that $\tilde{c}^T x_- = \tilde{g}^T x_- = 0$. Using (4.2), we see that $\nabla F(x_-)^T p(\tilde{c}) = \nabla F(x_-)^T p(\tilde{g}) = 0$ (using (4.7)), so that

$$\nabla F(x_-)^T (p(\tilde{c}) - \zeta p(\tilde{g})) = 0 \tag{4.12}$$

and, with $-\nabla F(x_-) \geq 0$, (4.9) implies

$$(c - \zeta g)^T x_- = -\nabla F(x_-)^T x(\zeta) \geq 0$$

(using (4.11) again), or $\zeta \leq c^T x_-$. This shows that $\check{\zeta}$ is not $+\infty$. Next, since $x(\zeta)$ is affine in ζ, the computation of $\check{\zeta}$ is trivial (like a minimum ratio test in the simplex method). Lastly, since ζ_+ is either $\zeta_- \leq \zeta_*$ or $\check{\zeta}$, the value of a feasible dual solution, we have

Proposition 4.1 *If ζ_- is updated to ζ_+ as in (4.10), then $\zeta_+ \leq \zeta_*$.* □

Since $\zeta_+ \geq \zeta_-$ by definition, and $g^T x_- = 1 > 0$, we immediately have

$$\phi(x_-; \zeta_+) \leq \phi(x_-; \zeta_-). \tag{4.13}$$

It therefore suffices to take a step in x to reduce $\phi(\cdot; \zeta_+)$ by a constant δ. We take a step in the direction (see (4.6))

$$p := p(\tilde{c}) - \zeta_+ p(\tilde{g}) = p(\tilde{c} - \zeta_+ \tilde{g}). \tag{4.14}$$

Note that, as a result of (4.9) and the remark below (4.10), we have

$$\frac{(c - \zeta_+ g)^T x_-}{n} x_- - p \not> 0. \tag{4.15}$$

This allows us to prove an important result on the quality of the updated lower bound ζ_+.

Proposition 4.2 *Let* $\max(v)$ *denote the largest component of a vector* v*. Then we have*

$$c^T x_- - \zeta_+ \leq n \max(X_-^{-1}p) \leq n\|X_-^{-1}p\|_\infty \leq n\|X_-^{-1}p\|_2. \tag{4.16}$$

Proof. From (4.15), $\frac{c^T x_- - \zeta_+}{n} e - X_-^{-1}p$ has a nonpositive component, which gives the first inequality; the others are immediate from the definition of the norms. ∎

Let us now introduce some useful notation for the quantities appearing in (4.16). Given vectors $x_- > 0$ and v in $I\!R^n$, we set

$$\begin{aligned} \|v\|_{x_-} &:= \|X_-^{-1}v\|_2 = (v^T \nabla^2 F(x_-) v)^{1/2}; \\ \sigma_{x_-}(v) &:= \max\{0, \max(X_-^{-1}v)\} \\ &= \min\{\sigma \geq 0 : \sigma x_- - v \geq 0\} \\ &= \min\{\sigma \geq 0 : \sigma e - X_-^{-1}v \geq 0\}; \\ |v|_{x_-} &:= \max\{\sigma_{x_-}(v), \sigma_{x_-}(-v)\} = \|X_-^{-1}v\|_\infty. \end{aligned} \tag{4.17}$$

The norm dual to $\|\cdot\|_{x_-}$ is

$$\|u\|_{x_-}^* := \|X_- u\|_2 = (u^T [\nabla^2 F(x_-)]^{-1} u)^{1/2}. \tag{4.18}$$

Note that $\|\cdot\|_{x_-}$ is the norm we have used in defining steepest descent steps; it is appropriate for measuring displacements at x_-, while $\|\cdot\|_{x_-}^*$ is appropriate for measuring gradients at x_-. Indeed, (4.7) and (4.11) show that $\|x_-\|_{x_-} = \|\nabla F(x_-)\|_{x_-}^* = \sqrt{n}$.

We can now show that p is a good search direction at least initially, by computing the directional derivative of ϕ in the direction p. Indeed,

$$\begin{aligned} \nabla_x \phi(x_-; \zeta_+)^T p &= \tfrac{n}{c^T x_- - \zeta_+}(\tilde{c} - \zeta_+ \tilde{g})^T p \qquad \text{(using (4.4))} \\ &= -\tfrac{n}{c^T x_- - \zeta_+} p^T \nabla^2 F(x_-) p \\ &\quad \text{(using (4.2) with } u = \tilde{c} - \zeta_+ \tilde{g}) \\ &\leq -\tfrac{n}{n|p|_{x_-}} \|p\|_{x_-}^2 \\ &\quad \text{(using (4.16))} \\ &\leq -\tfrac{n}{n\|p\|_{x_-}} \|p\|_{x_-}^2 = -\|p\|_{x_-}. \end{aligned} \tag{4.19}$$

Thus the directional derivative in the direction $p/\|p\|_{x_-}$ is at most -1, and we can take a unit step in this direction, suggesting that a constant decrease is possible. In fact, since the ∞-norm is usually much less than the 2-norm, the directional derivative is typically much smaller and we can typically take a much longer step, resulting in a considerably larger decrease. For this, we need to control the higher-order terms in ϕ. We use

Theorem 4.1 *For any x_-, $p \in \mathbb{R}^n$ and $\alpha > 0$ such that $x_- > 0$ and $x_- + \alpha p > 0$, we have*

$$F(x_- + \alpha p) \le F(x_-) + \alpha \nabla F(x_-)^T p + \frac{\alpha^2 p^T \nabla^2 F(x_-) p}{2(1 - \alpha \sigma_{x_-}(-p))}. \tag{4.20}$$

Proof. This result is usually proved using a series expansion of the logarithm. We give a simple proof that is easy to generalize (see Section 8). We have

$$\begin{aligned} F(x_- + \alpha p) &= F(x_-) + \alpha \nabla F(x_-)^T p + \int_0^1 \int_0^\xi \alpha^2 p^T \nabla^2 F(x_- + \tau \alpha p) p \, d\tau \, d\xi \\ &= F(x_-) + \alpha \nabla F(x_-)^T p + \int_0^1 (1-\tau) \alpha^2 p^T \nabla^2 F(x_- + \tau \alpha p) p \, d\tau. \end{aligned}$$

Now, with $\sigma := \sigma_{x_-}(-p)$, we have $\sigma x_- + p \ge 0$, so $x_- + \tau \alpha p \ge (1 - \tau\alpha\sigma) x_-$ and hence

$$\nabla^2 F(x_- + \tau \alpha p) \le \nabla^2 F((1 - \tau\alpha\sigma) x_-) = (1 - \tau\alpha\sigma)^{-2} \nabla^2 F(x_-),$$

in the sense that the right-hand side minus the left is positive semidefinite. Using this inequality in the equation above we get

$$F(x_- + \alpha p) \le F(x_-) + \alpha \nabla F(x_-)^T p + p^T \nabla^2 F(x_-) p \int_0^1 (1-\tau)(1 - \tau\alpha\sigma)^{-2} d\tau.$$

Now $(1 - \tau\alpha\sigma)^{-2}$ is a nondecreasing function of τ, so the integral above is at most $\frac{1}{2} \int_0^1 (1 - \tau\alpha\sigma)^{-2} d\tau$, and evaluating this gives (4.20). $\square$

Theorem 4.1 can be interpreted in two ways. First, since $\sigma_{x_-}(-p) \le |p|_{x_-} \le \|p\|_{x_-}$, it shows that F can be well approximated by its first-order Taylor approximation in a neighborhood $\{x_- + w : \|w\|_{x_-} \le \beta\}$ with small β, since the error is at most $\|w\|_{x_-}^2 / [2(1 - \|w\|_{x_-})] \le \beta^2 / [2(1-\beta)]$. Second, it shows that F can be well approximated by its second-order Taylor approximation in a one-sided way in a much larger neighborhood $\{x_- + w : |w|_{x_-} \le \beta\}$ or even $\{x_- + w : \sigma_{x_-}(-w) \le \beta\}$ for small β, since the second-order term

$w^T\nabla^2 F(x_-)w/2$ can be replaced by $w^T\nabla^2 F(x_-)w/[2(1-\beta)]$ to give a valid bound.

We think of algorithms whose steps yield points within the first kind of neighborhood as *short-step* methods, while those that take steps to points near the boundary of the latter two kinds of neighborhood are *long-step* methods.

We are now ready for the main result of this subsection.

Theorem 4.2 *Suppose $x_- \in \mathcal{F}^0(P)$ and $\zeta_- \le \zeta_*$. Let ζ_- be updated to ζ_+ as in (4.10), and suppose p is given by (4.14). Then, for $\alpha := [\sigma_{x_-}(p) + \sigma_{x_-}(-p)]^{-1}$, we have $x_+ := x_- + \alpha p \in \widehat{\mathcal{F}}^0(P)$ and*

$$\begin{aligned} \phi(x_+;\zeta_+) &\le \phi(x_-;\zeta_-) - \tfrac{1}{4}\tfrac{\|p\|_{x_-}^2}{|p|_{x_-}^2} \\ &\le \phi(x_-;\zeta_-) - \tfrac{1}{4}. \end{aligned} \tag{4.21}$$

Proof. Using (4.13) it suffices to prove the inequality with ζ_- replaced by ζ_+.

From Proposition 4.2, we have (unless x_- is optimal, which we have excluded) that $\sigma_{x_-}(p)$ is positive. From (4.12), we have $\nabla F(x_-)^T p = 0$, or $e^T X_-^{-1} p = 0$, so $X_-^{-1}p$ has both positive and negative components and $\sigma_{x_-}(-p)$ is also positive. Thus α is well-defined and $0 < \alpha < 1/\sigma_{x_-}(-p)$. This shows that x_+ is positive, and since $Bp = 0$, x_+ lies in $\widehat{\mathcal{F}}^0(P)$. Now

$$\begin{aligned} \Delta\phi(\alpha) &:= \phi(x_+;\zeta_+) - \phi(x_-;\zeta_+) \\ &= n\ln\frac{(c-\zeta_+ g)^T x_+}{(c-\zeta_+ g)^T x_-} + F(x_+) - F(x_-) \\ &\le n\ln\left(1 + \alpha\frac{(c-\zeta_+ g)^T p}{(c-\zeta_+ g)^T x_-}\right) + \alpha\nabla F(x_-)^T p + \alpha^2\frac{p^T\nabla^2 F(x_-)p}{2[1-\alpha\sigma_{x_-}(-p)]} \\ &\qquad \text{(using Theorem 4.1)} \\ &\le \alpha\left(\frac{n}{(c-\zeta_+ g)^T x_-}(c-\zeta_+ g) + \nabla F(x_-)\right)^T p + \alpha^2\frac{\|p\|_{x_-}^2}{2[1-\alpha\sigma_{x_-}(-p)]} \\ &\qquad \text{(using the concavity of the logarithm)} \\ &= -\alpha\frac{n}{(c-\zeta_+ g)^T x_-}p^T\nabla^2 F(x_-)p + \alpha^2\frac{\|p\|_{x_-}^2}{2[1-\alpha\sigma_{x_-}(-p)]} \\ &\qquad \text{(using (4.19))} \end{aligned}$$

$$
\leq \left[-\frac{\alpha}{\sigma_{x_-}(p)} + \frac{\alpha^2}{2[1-\alpha\sigma_{x_-}(-p)]}\right] \|p\|_{x_-}^2
$$
$$
\text{(using (4.16)).}
$$

Now we substitute $\alpha = [\sigma_{x_-}(p) + \sigma_{x_-}(-p)]^{-1}$ to get

$$
\begin{aligned}
\Delta\phi(\alpha) &\leq -\frac{1}{2\sigma_{x_-}(p)[\sigma_{x_-}(p) + \sigma_{x_-}(-p)]} \|p\|_{x_-}^2 \\
&\leq -\frac{1}{4}\frac{\|p\|_{x_-}^2}{|p|_{x_-}^2},
\end{aligned}
$$

since $\sigma_{x_-}(p)$ and $\sigma_{x_-}(-p)$ are both at most $|p|_{x_-}$. □

This theorem demonstrates that a constant reduction in ϕ (in fact, a large reduction if $\|p\|_{x_-}$ is much larger than $|p|_{x_-}$) can be achieved at each iteration of Karmarkar's algorithm and completes the analysis.

Note that the step size α chosen is between $(2|p|_{x_-})^{-1}$ and $|p|_{x_-}^{-1}$, whereas the largest feasible step size is $\sigma_{x_-}(-p)^{-1}$, which is of comparable size if $\sigma_{x_-}(p)$ and $\sigma_{x_-}(-p)$ are. Thus this is a long-step algorithm. Taking such a step assures a reduction in the potential function that is usually much larger than $1/4$; indeed, we might hope that it is typically $\Omega(n/\ln n)$. The long step and large reduction are byproducts of the analysis, not of a hopeful line search with a constant reduction assured by a small step size. It can be seen from the proofs that the key ingredients in the improved analysis are the uses of $\sigma_{x_-}(p)$ in Proposition 4.2 and $\sigma_{x_-}(-p)$ in Theorem 4.1; looser bounds using $\|p\|_{x_-}$ would only permit short steps and a constant decrease in ϕ in the analysis. Finally, we note that tighter bounds on F and hence an improvement in Theorems 4.1 and 4.2 (with $\frac{1}{4}$ replaced by $1 - \ln 2$) are possible by specializing to $\mathbb{R}_+^n$ the arguments of Nesterov and Todd [34], but the present development is simpler and achieves qualitatively similar results.

4.2 Gonzaga's Algorithm

The analysis of Section 4.1 relied heavily on the homogeneity of ϕ in (4.1) and the special form of the constraints. Indeed, the iterates could be thought of as rays in $\widehat{\mathcal{F}}^0(P)$ rather than points in $\mathcal{F}^0(P)$; after each iteration, or once at the end of the process, a point in the latter could be obtained by replacing a ray represented by $x \in \widehat{\mathcal{F}}^0(P)$ by $x/g^T x \in \mathcal{F}^0(P)$. In Karmarkar's original presentation of the method, these ideas were encapsulated in a projective transformation which was made at every iteration to convert the current point into e/n and the normalizing constraint into $e^T x = 1$ (or equivalently $-\nabla F(e/n)^T x = -\nabla F(e/n)^T e/n = n$). In our analysis, the corresponding relation $\nabla F(x_-)^T p = 0$ came for free (see (4.12)) from our derivation of the search direction, in particular from our choice of norm.

Gonzaga [14] showed that neither the special form of the constraints nor the homogeneity of the potential function was required to ensure a constant decrease. He considered the standard-form problem (P) of Section 2 in its general form, and used the potential function ϕ_ρ with $\rho \geq n + \sqrt{n}$. Here we will describe briefly his method, again stressing the derivation of the direction and the possibility of long steps, and laying emphasis on the choice of ρ. (Gonzaga showed that $\rho \geq n + \sqrt{n}$ suffices for his analysis, but his later paper [15] recommended larger values for ρ of at least $2n$, based on his analysis of the amount by which the lower bound is updated in a path-following version of his method. We will provide arguments based on the length of steps and the amount of decrease in the potential function for such a choice.)

Once again we assume we have a current iterate $x_- \in \mathcal{F}^0(P)$ and a lower bound $\zeta_- \leq \zeta_*$. Our direction will be the steepest descent direction for ϕ_ρ in the nullspace of A with respect to the norm $\|\cdot\|_{x_-}$ defined by $\nabla^2 F(x_-)$, i.e., the solution to

$$(DFP') \qquad \begin{array}{rrcl} \min & u^T p & + & \frac{1}{2} p^T \nabla^2 F(x_-) p \\ & Ap & = & 0, \end{array}$$

for $u = \nabla \phi_\rho(x_-; \zeta)$, where ζ is a possibly updated lower bound. Since

$$\nabla \phi_\rho(x_-; \zeta) = \lambda^{-1}(c + \lambda d), \tag{4.22}$$

where

$$\lambda = \lambda(\zeta) := \frac{c^T x_- - \zeta}{\rho} \tag{4.23}$$

and

$$d := \nabla F(x_-), \tag{4.24}$$

we compute the solution $p(u)$ of (DFP') for $u = c$ and $u = d$ separately. Note that we have

$$\begin{array}{rcl} A^T y(u) - \nabla^2 F(x_-)p(u) & = & u \\ Ap(u) & = & 0, \end{array} \tag{4.25}$$

for all u. We therefore have

$$A^T[y(c) + \lambda y(d)] + [-\lambda d - \nabla^2 F(x_-)(p(c) + \lambda p(d))] = c. \tag{4.26}$$

Thus, let

$$x(\lambda) := \lambda x_- - p(c) - \lambda p(d) \tag{4.27}$$

and

$$s(\lambda) := \nabla^2 F(x_-)x(\lambda) = -\lambda d - \nabla^2 F(x_-)(p(c) + \lambda p(d)) \tag{4.28}$$

(using (4.7)), so that, if $s(\lambda) \geq 0$ (or equivalently $x(\lambda) \geq 0$), $s(\lambda) \in \widetilde{\mathcal{F}}(D)$. The corresponding duality gap is

$$\begin{array}{rcl} x_-^T s(\lambda) & = & -d^T x(\lambda) \qquad \text{(from (4.28) and (4.7))} \\ & = & d^T p(c) - \lambda[d^T x_- - d^T p(d)] \qquad \text{(from (4.27))} \\ & = & d^T p(c) + \lambda \left[d^T[\nabla^2 F(x_-)]^{-1} d + d^T p(d)\right] \\ & & \text{(using (4.7) again).} \end{array} \tag{4.29}$$

Now, from (4.25), we get

$$\begin{array}{rcl} \|p(d)\|_{x_-}^2 & = & p(d)^T \nabla^2 F(x_-)p(d) \\ & = & -d^T p(d) \leq \left(d^T[\nabla^2 F(x_-)]^{-1} d\right)^{\frac{1}{2}} \left(p(d)^T \nabla^2 F(x_-)p(d)\right)^{\frac{1}{2}} \\ & = & \|d\|_{x_-}^* \|p(d)\|_{x_-} \end{array} \tag{4.30}$$

from the Cauchy-Schwartz inequality; hence $\|p(d)\|_{x_-} \leq \|d\|_{x_-}^*$ and thus $x_-^T s(\lambda)$ is a nondecreasing function of λ. In fact, if we have equality in (4.30), then $p(d)$ is a multiple of $-\nabla^2 F(x_-)d = x_-$, so $b = Ax_-$ is zero. In this case, $x_* = 0$ is optimal in (P). If we exclude this case, the duality gap $x_-^T s(\lambda)$ is an increasing function of λ.

Thus we choose $\hat{\lambda}$ as the smallest λ so that $x(\lambda)$ or equivalently $s(\lambda)$ is nonnegative, or $+\infty$ if there is no such λ, $\hat{\zeta} = c^T x_- - x_-^T s(\hat{\lambda})$ from (4.29) ($-\infty$ if $\hat{\lambda} = +\infty$), and choose

$$\zeta_+ = \max\{\zeta_-, \hat{\zeta}\} \tag{4.31}$$

as our possibly updated lower bound. Since $x_-^T s(\lambda)$ is nonnegative if $s(\lambda)$ is, and (excluding the case $b = 0$) $x_-^T s(\lambda)$ is a strictly increasing function of λ, $\hat{\lambda}$ is not $-\infty$, so $\hat{\zeta}$ is either $-\infty$ or finite. Also, ζ_+ is either $\zeta_- \leq \zeta_*$ or the value of a feasible dual solution, so we have

Proposition 4.3 *If ζ_- is updated to ζ_+ as in (4.31), then $\zeta_+ \leq \zeta_*$.* □

We also obtain a result on the quality of the updated bound similar to Proposition 4.2. Let

$$\lambda := \lambda(\zeta_+), \quad p := p(c) + \lambda p(d), \tag{4.32}$$

so that, from (4.22), p is λ times the steepest descent direction for $\phi_\rho(\cdot;\zeta_+)$ at x_-.

Proposition 4.4 *If $\rho \geq n + \sqrt{n}$, then*

$$c^T x_- - \zeta_+ \leq \rho \|p\|_{x_-}, \quad \lambda \leq \|p\|_{x_-}, \tag{4.33}$$

while if $\rho \geq 2n$,

$$c^T x_- - \zeta_+ \leq \rho |p|_{x_-}, \quad \lambda \leq |p|_{x_-}. \tag{4.34}$$

Proof. The two parts of (4.33) (or of (4.34)) are equivalent by the definition of λ in (4.32) and (4.23). Suppose the result is false. Then in either case we have $\lambda > |p|_{x_-}$, so that $x(\lambda) = \lambda x_- - p > 0$ and so $\hat{\lambda} < \lambda$. It follows that

$$\begin{aligned} c^T x_- - \zeta_+ < x_-^T s(\lambda) &= x_-^T \nabla^2 F(x_-)(\lambda x_- - p) \\ &= \lambda n - e^T X_-^{-1} p. \end{aligned}$$

Suppose first that $\rho \geq n + \sqrt{n}$ and $\lambda > \|p\|_{x_-}$. Then

$$\begin{aligned} c^T x_- - \zeta_+ &< \lambda n + \|e\|_2 \|p\|_{x_-} \\ &< \lambda n + \sqrt{n}\lambda = \lambda(n + \sqrt{n}) \leq \lambda\rho, \end{aligned}$$

contradicting the definition of λ. On the other hand, if $\rho \geq 2n$ and $\lambda > |p|_{x_-}$, then

$$\begin{aligned} c^T x_- - \zeta_+ &< \lambda n + \|e\|_1 |p|_{x_-} \\ &< \lambda n + n\lambda = \lambda(2n) \leq \lambda\rho, \end{aligned}$$

again yielding a contradiction. □

We are now ready to prove the main theorem of this section.

Theorem 4.3 *Suppose $x_- \in \mathcal{F}^0(P)$ and $\zeta_- \le \zeta_*$. Let ζ_- be updated to ζ_+ as in (4.31) and suppose p is given by (4.32). Then*

a) if $\rho \ge n + \sqrt{n}$, for $\alpha = [\|p\|_{x_-} + \sigma_{x_-}(-p)]^{-1}$, we have $x_+ := x_- + \alpha p \in \mathcal{F}^0(P)$ and

$$\phi_\rho(x_+;\zeta_+) \le \phi_\rho(x_-;\zeta_-) - \frac{1}{4};$$

b) if $\rho \ge 2n$, for $\alpha = [|p|_{x_-} + \sigma_{x_-}(-p)]^{-1}$, we have $x_+ := x_- + \alpha p \in \mathcal{F}^0(P)$ and

$$\begin{aligned}\phi_\rho(x_+;\zeta_+) &\le \phi_\rho(x_-;\zeta_-) - \tfrac{1}{4}\tfrac{\|p\|_{x_-}^2}{|p|_{x_-}^2} \\ &\le \phi_\rho(x_-;\zeta_-) - \tfrac{1}{4}.\end{aligned}$$

Proof. Proceeding exactly as in the proof of Theorem 4.2, we find

$$\begin{aligned}\Delta\phi(\alpha) &:= \phi_\rho(x_+;\zeta_+) - \phi_\rho(x_-;\zeta_+) \\ &\le \left[-\tfrac{\alpha}{\lambda} + \tfrac{\alpha^2}{2[1-\alpha\sigma_{x_-}(-p)]}\right]\|p\|_{x_-}^2.\end{aligned}$$

For $\rho \ge n + \sqrt{n}$, we have $\lambda \le \|p\|_{x_-}$, so

$$\Delta\phi(\alpha) \le \left[-\frac{\alpha}{\|p\|_{x_-}} + \frac{\alpha^2}{2[1-\alpha\sigma_{x_-}(-p)]}\right]\|p\|_{x_-}^2,$$

and substituting $\alpha = [\|p\|_{x_-} + \sigma_{x_-}(-p)]^{-1}$, we obtain

$$\Delta\phi(\alpha) \le -\frac{1}{2\|p\|_{x_-}[\|p\|_{x_-} + \sigma_{x_-}(-p)]}\|p\|_{x_-}^2 \le -\frac{1}{4},$$

since $\sigma_{x_-}(-p) \le \|p\|_{x_-}$.

For $\rho \ge 2n$, we have $\lambda \le |p|_{x_-}$, so

$$\Delta\phi(\alpha) \le \left[-\frac{\alpha}{|p|_{x_-}} + \frac{\alpha^2}{2[1-\alpha\sigma_{x_-}(-p)]}\right]\|p\|_{x_-}^2,$$

and substituting $\alpha = [|p|_{x_-} + \sigma_{x_-}(-p)]^{-1}$, we obtain

$$\begin{aligned}\Delta\phi(\alpha) &\le -\tfrac{1}{2|p|_{x_-}[|p|_{x_-} + \sigma_{x_-}(-p)]}\|p\|_{x_-}^2 \\ &\le -\tfrac{1}{4}\tfrac{\|p\|_{x_-}^2}{|p|_{x_-}^2} \le -\tfrac{1}{4},\end{aligned}$$

since $\sigma_{x_-}(-p) \leq |p|_{x_-}$ and $|p|_{x_-} \leq \|p\|_{x_-}$. This yields the result. $\square$

For $n + \sqrt{n} \leq \rho < 2n$, the analysis guarantees a constant reduction of $\frac{1}{4}$ by choosing a short step: α is between $(2\|p\|_{x_-})^{-1}$ and $\|p\|_{x_-}^{-1}$. But a larger value of ρ, $\rho \geq 2n$, allows a longer step, between $(2|p|_{x_-})^{-1}$ and $|p|_{x_-}^{-1}$, that guarantees at least a reduction of $\frac{1}{4}$ but typically a much larger decrease. This provides another justification for Gonzaga's recommendation of a value for ρ of at least $2n$. Again, the bounds given here can be improved ($\frac{1}{4}$ becomes $(1 - \ln 2)$) at the expense of a slightly more complicated analysis [34].

5 The Primal-Dual Potential Function

In order to avoid problems with updating lower bounds and to obtain more symmetric primal-dual methods, we here investigate a potential function that measures the merit of a pair (x, s) in $\mathcal{F}^0(P) \times \tilde{\mathcal{F}}^0(D)$.

It is natural to combine a monotonic function of the duality gap $x^T s$ with barrier terms for both x and s. Thus, for $\rho \geq n$ and $x \in \mathcal{F}^0(P)$, $s \in \tilde{\mathcal{F}}^0(D)$, we define the primal-dual potential function by

$$\Phi_\rho(x, s) := \rho \ln x^T s + F(x) + F(s). \tag{5.1}$$

A very important observation comes from Proposition 3.1, which gives a bound on $\Phi_n(x, s)$. Hence we have

$$\begin{array}{rcl} \Phi_\rho(x, s) & = & (\rho - n) \ln x^T s + \Phi_n(x, s) \\ & \geq & (\rho - n) \ln x^T s + n \ln n, \end{array} \tag{5.2}$$

with equality iff $XSe = \mu e$ for some $\mu > 0$. This demonstrates the very close relationship between this potential function and the duality gap, and permits the easy proof of versions of the primal-only theorems of Section 3. The function Φ_ρ was independently introduced in 1987 by Tanabe [37] (in multiplicative form) and Todd and Ye [41].

Once again, our potential function is related to a penalized function, here

$$\Psi_\mu(x, s) := \frac{x^T s}{\mu} + F(x) + F(s). \tag{5.3}$$

This function is convex on $\mathcal{F}^0 := \mathcal{F}^0(P) \times \tilde{\mathcal{F}}^0(D)$, since we can rewrite $x^T s$ as a linear function of x and s: choose any $(x_0, s_0) \in \mathcal{F}^0$ and note that $x - x_0$ lies in the nullspace and $s-s_0$ in the row space of A, so that $(x-x_0)^T(s-s_0)=0$ or $x^Ts = x_0^Ts + s_0^Tx - x_0^Ts_0$.

We have

$$\begin{aligned}
\nabla_x \Psi_\mu(x,s) &= \frac{s}{\mu} + \nabla F(x) = \frac{s}{\mu} - x^{-1}, \\
\nabla^2_{xx} \Psi_\mu(x,s) &= X^{-2},
\end{aligned}$$

and

$$\begin{aligned}
\nabla_x \Phi_\rho(x,s) &= \frac{\rho}{x^Ts}s + \nabla F(x) = \frac{\rho}{x^Ts}s - x^{-1}, && (5.4)\\
\nabla^2_{xx} \Phi_\rho(x,s) &= -\frac{\rho}{(x^Ts)^2}ss^T + \nabla^2 F(x) = -\frac{\rho}{(x^Ts)^2}ss^T + X^{-2}, && (5.5)
\end{aligned}$$

and similar results for derivatives with respect to s. The remarks we made below (3.8) are also valid in this setting. The gradients of Ψ_μ and Φ_ρ with respect to x and s coincide for suitable values of μ and ρ, and hence steepest descent or pseudo-Newton steps (where we use the positive definite part $\nabla^2 F(x)$ of the Hessian of Φ_ρ) for path-following and potential-reduction methods will bear a strong resemblance.

Now we establish the results for Φ_ρ analogous to those we proved for the primal-only potential function ϕ_ρ.

Theorem 5.1 *If $\mathcal{F}^0$ is nonempty and $\rho > n$, then Φ_ρ is unbounded below.*

Proof. We could use a proof like that of Theorem 3.1, considering Φ_ρ on a line segment joining $(x_0, s_0) \in \mathcal{F}^0$ and (x_*, s_*), where x_* and (y_*, s_*) are optimal solutions to (P) and (D). It turns out to be necessary to assume that x_* and (y_*, s_*) are strictly complementary. Since there may not exist a strictly complementary solution in the more general situation of a monotone linear complementarity problem, to be considered in Section 7, we provide an alternative proof.

It is known that, for every $\mu > 0$, there is a corresponding point on the primal-dual central path, i.e., a pair $(x(\mu), s(\mu)) \in \mathcal{F}^0$ with $X(\mu)S(\mu)e = \mu e$. Using (5.2) we find $\Phi_\rho(x(\mu), s(\mu)) = (\rho - n)\ln n\mu + n \ln n$, which tends to $-\infty$ as μ decreases to zero. □

The second result is even easier:

Theorem 5.2 *Suppose $\mathcal{F}^0 \neq \phi$ and $\Phi_\rho(x,s) \to -\infty$ for $(x,y,s) \in \mathcal{F}^0(P) \times \mathcal{F}^0(D)$ and $\rho > n$. Then $c^T x \to v(P) = v(D) \leftarrow b^T y$. Indeed,*

$$c^T x - b^T y = x^T s \leq \exp\{\Phi_\rho(x,s)/(\rho - n)\}. \tag{5.6}$$

Proof. Since $n \ln n \geq 0$, this follows directly from (5.2). □

Corollary 5.1 *Suppose we have available some $(x_0, s_0) \in \mathcal{F}^0$ and an algorithm that at each iteration replaces $(x_k, s_k) \in \mathcal{F}^0$ with $(x_{k+1}, s_{k+1}) \in \mathcal{F}^0$ such that*

$$\Phi_\rho(x_{k+1}, s_{k+1}) \leq \Phi_\rho(x_k, s_k) - \delta \tag{5.7}$$

for some positive absolute constant δ. Then for any positive ϵ,

$$x_k^T s_k \leq \epsilon \tag{5.8}$$

within

$$K = \left\lceil \frac{\rho - n}{\delta} \left(\frac{\Phi_\rho(x_0, s_0)}{\rho - n} + \ln \frac{1}{\epsilon} \right) \right\rceil \tag{5.9}$$

iterations. □

Hence, if (5.7) is achievable for $\rho = n + O(\sqrt{n})$, we obtain an $O(\sqrt{n} \ln \frac{1}{\epsilon})$-iteration algorithm.

6 Primal-Dual Potential-Reduction Methods

We assume here that $\mathcal{F}^0$ is nonempty and that we know some initial point $(x_0, y_0, s_0) \in \mathcal{F}^0(P) \times \mathcal{F}^0(D)$.

The first algorithm to use the primal-dual potential function in its analysis was that of Todd and Ye [41]. This method was symmetric between the primal and dual, but not a pure potential-reduction algorithm. Indeed, all iterates lay in a neighborhood of the central path, and this restriction was necessary for the analysis of the decrease in the potential function. (Actually, the neighborhood restriction was not imposed explicitly; the initial iterates had to be near the central path, and short steps were employed, so that the method was path-following *de facto*. However, line searches to decrease the potential function further were forbidden.) In addition, the algorithm used projective transformations, making it complicated; we will not discuss it further here.

The first pure potential-reduction algorithm requiring only $O(\sqrt{n}\ln\frac{1}{\epsilon})$ iterations was that of Ye [46] (see also Freund [10]). This algorithm achieved a constant reduction in the symmetric primal-dual potential function from any pair $(x_-, s_-) \in \mathcal{F}^0$ and permitted line searches. However, it was not a symmetric primal-dual method. Instead, some steps were identical to those of Gonzaga's primal potential-reduction algorithm, motivated by the fact that, from (3.4) and (5.1),

$$\Phi_\rho(x, s) = \phi_\rho(x; \zeta) + F(s)$$

for $\zeta := b^T y$, where $(y, s) \in \mathcal{F}^0(D)$, so that a constant decrease in ϕ_ρ keeping ζ constant yields a constant decrease in Φ_ρ holding s constant. Similarly, we see from (3.7) and (5.4) that, for this value of ζ, the gradients of ϕ_ρ and Φ_ρ with respect to x differ by an element of the row space of A, so that their steepest descent directions in the nullspace of A coincide. Ye's algorithm either takes a step in x in this direction, holding s constant, or updates s (and hence ζ) by a method similar to that in (4.28), but without moving λ to $\hat{\lambda}$ which would cause $F(s)$ in Φ_ρ to explode. Details may be found in [46].

We will devote the rest of this section to a discussion of the symmetric primal-dual potential-reduction algorithm of Kojima, Mizuno, and Yoshise [22]. To derive this method, let us consider first separately steepest descent directions for x and for s at a point $(x_-, y_-, s_-) \in \mathcal{F}^0(P) \times \mathcal{F}^0(D)$. For x, we restrict ourselves to the nullspace of A and use the norm defined by $\nabla^2 F(z)$, where we think of z as in some sense close to x_-. Thus the primal direction solves

$$\begin{array}{rrcl} \min & \nabla_x \Phi_\rho(x_-, s_-)^T p & + & \frac{1}{2} p^T \nabla^2 F(z) p \\ & Ap & = & 0, \end{array}$$

and hence, together with some Lagrange multiplier vector w, the system

$$\begin{array}{rcrcl} Ap & & & = & 0 \\ \nabla^2 F(z) p & - & A^T w & = & -\nabla_x \Phi_\rho(x_-, s_-), \end{array}$$

or, writing v for $-A^T w$ and substituting the formula (5.4) for the gradient,

$$\begin{array}{lcrcrcl} & & Ap & & & = & 0 \\ A^T w & & & + & v & = & 0 \\ & & \nabla^2 F(z) p & + & v & = & -\frac{\rho}{x_-^T s_-} s_- + x_-^{-1}. \end{array} \tag{6.1}$$

Similarly, the dual steepest descent direction (r, q) for (y, s) will be restricted to the nullspace of (A^T, I) and based on the norm defined by $\nabla^2 F(t)$ in the s-variables, where again t is an approximation to s_-. Thus (r, q) solves

$$\begin{array}{rrcl} \min & \nabla_s \Phi_\rho(x_-, s_-)^T q & + & \frac{1}{2} q^T \nabla^2 F(t) q \\ & A^T r & + & q \quad = \quad 0, \end{array}$$

and hence, together with some Lagrange multiplier vector u, the system

$$\begin{array}{rcccl} & Au & & = & 0 \\ A^T r & & + \quad q & = & 0 \\ & u & + \quad \nabla^2 F(t) q & = & -\frac{\rho}{x_-^T s_-} x_- + s_-^{-1}, \end{array} \tag{6.2}$$

where we have substituted for $\nabla_s \Phi_\rho(x_-, s_-)$ the expression $\frac{\rho}{x_-^T s_-} x_- + \nabla F(s_-)$.

Note the strong similarity between (6.1) and (6.2). Indeed, the two systems coincide if we choose z and t so that

$$\nabla^2 F(z) = [\nabla^2 F(t)]^{-1}, \quad \nabla^2 F(z) x_- = s_-, \quad \nabla^2 F(z) s_-^{-1} = x_-^{-1} \tag{6.3}$$

all hold. The first condition is in fact quite natural; it is that the norms $\|v\|_z := (v^T \nabla^2 F(z) v)^{\frac{1}{2}}$ and $\|u\|_t := (u^T \nabla^2 F(t) u)^{\frac{1}{2}}$ are dual to each other. All these conditions hold if we choose

$$z = X_-^{\frac{1}{2}} S_-^{-\frac{1}{2}} e, \quad t = X_-^{-\frac{1}{2}} S_-^{\frac{1}{2}} e, \tag{6.4}$$

from the expression (3.5) for $\nabla^2 F$. Thus we choose z and t as in (6.4) and our search directions (p, q, r) from

$$\begin{array}{rcccl} & Ap & & = & 0 \\ A^T r & & + \quad q & = & 0 \\ & \nabla^2 F(z) p & + \quad q & = & -\frac{\rho}{x_-^T s_-} s_- + x_-^{-1}, \end{array} \tag{6.5}$$

where the last equation can equivalently be written as

$$p + \nabla^2 F(t) q = -\frac{\rho}{x_-^T s_-} x_- + s_-^{-1}, \tag{6.6}$$

showing that the method is symmetric, or in the symmetric form

$$S_- p + X_- q = -\frac{\rho}{x_-^T s_-} X_- S_- e + e. \tag{6.7}$$

A similar motivation is used by Nazareth [31] for the primal-dual affine-scaling directions.

There is a closely related choice for z and t that should also be mentioned. Let $\mu_- = x_-^T s_-/n$ and

$$z = \mu_-^{\frac{1}{2}} X_-^{\frac{1}{2}} S_-^{-\frac{1}{2}} e, \quad t = \mu_-^{\frac{1}{2}} X_-^{-\frac{1}{2}} S_-^{\frac{1}{2}} e.$$

Then, if (x_-, s_-) lies on the central path so that $X_- S_- e = \mu_- e$, it is easy to see that $z = x_-$ and $t = s_-$; if (x_-, s_-) is close to the central path, z and t are close to x_- and s_-. If we use these values in (6.1) and (6.2), then the systems, viewed as equations in $(p, \mu_- v, \mu_- w)$ and in $(\mu_- u, q, r)$, again become identical and can be written in terms of p, q and r as

$$\begin{array}{rcrcrcl} & & Ap & & & = & 0 \\ A^T r & & & + & q & = & 0 \\ & & S_- p & + & X_- q & = & \mu_- \left(e - \frac{\rho}{x_-^T s_-} X_- S_- e \right) \\ & & & & & = & \frac{\rho}{n} (\gamma \mu_- e - X_- S_- e), \end{array}$$

where $\gamma := n/\rho$. The resulting directions are just scalar multiples (by μ_-) of those coming from (6.5), and the system above is almost that (since $\rho/n \approx 1$ for $\rho = n + \sqrt{n}$) arising in a (short-step, since γ is close to 1) path-following method. However, the iterates need not be close to the central path and a line search can be made along the search directions. We use z and t in (6.4) to ensure the properties (6.3), which can be extended beyond the case of linear programming (see Section 8).

Let us evaluate the directional derivative of Φ_ρ in the directions given by (6.5). We have

$$\begin{array}{rcl} \nabla_x \Phi_\rho(x_-, s_-)^T p & + & \nabla_s \Phi_\rho(x_-, s_-)^T q \\ & = & (\frac{\rho}{x_-^T s_-} x_- + \nabla F(s_-))^T (\nabla^2 F(z) p + q) \\ & = & -(\frac{\rho}{x_-^T s_-} x_- + \nabla F(s_-))^T (\frac{\rho}{x_-^T s_-} s_- + \nabla F(x_-)) \qquad (6.8) \\ & & \text{(from (6.5))} \\ & = & -[\nabla F(x_-)^T \nabla F(s_-) + \frac{\rho^2}{x_-^T s_-} \\ & & + \frac{\rho}{x_-^T s_-} (\nabla F(x_-)^T x_- + \nabla F(s_-)^T s_-)] \end{array}$$

$$= \quad -[\nabla F(x_-)^T \nabla F(s_-) - \frac{\rho(2n-\rho)}{x_-^T s_-}],$$

where the last line uses $\nabla F(x_-)^T x_- = \nabla F(s_-)^T s_- = -n$ from (4.11). We bound this using

Proposition 6.1 *[22] For any $x_- > 0$, $s_- > 0$, $\rho \geq n + \sqrt{n}$,*

$$\nabla F(x_-)^T \nabla F(s_-) - \frac{\rho(2n-\rho)}{x_-^T s_-} \geq \frac{3}{4}\sigma^2 + \frac{\rho^2 - 2\rho n + n^2 - n}{x_-^T s_-} \geq \frac{3}{4}\sigma^2, \quad (6.9)$$

where

$$\sigma := \sigma_{x_-}(z) = \sigma_{s_-}(t) = \max(X_-^{-1/2} S_-^{-1/2} e). \quad (6.10)$$

Proof. Note that the first inequality in (6.9) can be written as

$$\nabla F(x_-)^T \nabla F(s_-) - \frac{n^2 - n}{x_-^T s_-} \geq \frac{3}{4}\sigma^2, \quad (6.11)$$

or, in terms of

$$v := X_-^{1/2} S_-^{1/2} e, \quad v_{\min} := \min\{v^{(j)}\} = \sigma^{-1}, \quad (6.12)$$

$$(v^{-1})^T v^{-1} - \frac{n^2 - n}{v^T v} \geq \frac{3}{4} v_{\min}^{-2}.$$

The second inequality of (6.9) is clear, since

$$\rho^2 - 2\rho n + n^2 - n = (\rho - n - \sqrt{n})^2 + 2\sqrt{n}(\rho - n - \sqrt{n}).$$

But we have, with $\mu := v^T v / n$,

$$\begin{aligned}
(v^{-1})^T v^{-1} - \frac{n^2 - n}{v^T v} &= \|v^{-1} - \frac{n + \sqrt{n}}{v^T v} v\|^2 = \|v^{-1} - \frac{v}{\mu} - \frac{v}{\sqrt{n}\mu}\|^2 \\
&= \|v^{-1} - \frac{v}{\mu}\|^2 + \|\frac{v}{\sqrt{n}\mu}\|^2 \\
&\qquad \text{(since } v \text{ and } v^{-1} - \tfrac{v}{\mu} \text{ are orthogonal)} \\
&\geq (v_{\min}^{-1} - \frac{v_{\min}}{\mu})^2 + \frac{1}{\mu} = \frac{1}{\mu^2}[v_{\min}^{-2}\mu^2 - \mu + v_{\min}^2] \\
&= \frac{1}{\mu^2}[(\frac{1}{2} v_{\min}^{-1}\mu - v_{\min})^2 + \frac{3}{4} v_{\min}^{-2}\mu^2] \geq \frac{3}{4} v_{\min}^{-2}.
\end{aligned}$$

∎

We can now prove the main result of this section.

Theorem 6.1 *Let $(x_-, y_-, s_-) \in \mathcal{F}^0(P) \times \mathcal{F}^0(D)$ and $\rho \geq n + \sqrt{n}$, and suppose p, q, and r are obtained as above. Then there is some $\alpha > 0$ such that $x_+ := x_- + \alpha p$, $y_+ := y_- + \alpha r$, and $s_+ := s_- + \alpha q$ satisfy $(x_+, y_+, s_+) \in \mathcal{F}^0(P) \times \mathcal{F}^0(D)$ and*

$$\Phi_\rho(x_+, s_+) \leq \Phi_\rho(x_-, s_-) - 1/8.$$

By Corollary 5.1, this yields an $O(\sqrt{n} \ln \frac{1}{\epsilon})$-iteration algorithm for $\rho = n + \sqrt{n}$.

Proof. It is clear that, if $\alpha \sigma_{x_-}(-p) < 1$ and $\alpha \sigma_{s_-}(-q) < 1$, then $(x_+, y_+, s_+) \in \mathcal{F}^0(P) \times \mathcal{F}^0(D)$. Also, proceeding exactly as in Theorems 4.2 and 4.3 and using (6.8), we obtain

$$\begin{array}{rcl} \Delta\Phi_\rho(\alpha) & := & \Phi_\rho(x_+, s_+) - \Phi_\rho(x_-, s_-) \\ & \leq & -\alpha\pi^2 + \frac{\alpha^2 p^T \nabla^2 F(x_-) p}{2[1 - \alpha\sigma_{x_-}(-p)]} + \frac{\alpha^2 q^T \nabla^2 F(s_-) q}{2[1 - \alpha\sigma_{s_-}(-q)]}, \end{array} \tag{6.13}$$

where

$$\pi^2 := \nabla F(x_-)^T \nabla F(s_-) - \frac{\rho(2n - \rho)}{x_-^T s_-}. \tag{6.14}$$

We now use the key fact that, since $\sigma x_- \geq z$ (where σ is given by (6.10)),

$$\nabla^2 F(x_-) \leq \sigma^2 \nabla^2 F(z), \tag{6.15}$$

in the sense that the right-hand side minus the left is positive semidefinite, and similarly

$$\nabla^2 F(s_-) \leq \sigma^2 \nabla^2 F(t). \tag{6.16}$$

Thus $\|p\|_{x_-} \leq \sigma \|p\|_z$ and $\|q\|_{s_-} \leq \sigma \|q\|_t = \sigma \|q\|_z^*$. Hence, if we define

$$\bar{\sigma} := \max\{\sigma_{x_-}(-p), \sigma_{s_-}(-q)\}, \tag{6.17}$$

the second and third terms on the right-hand side of (6.13) are at most

$$\frac{\alpha^2 \sigma^2 (\|p\|_z^2 + \|q\|_t^2)}{2(1 - \alpha\bar{\sigma})}. \tag{6.18}$$

Now, by taking the scalar products of the last equation in (6.5) and (6.6), we obtain

$$\|p\|_z^2 + \|q\|_t^2 + 2p^T q = (\frac{\rho}{x_-^T s_-} s_- + \nabla F(x_-))^T (\frac{\rho}{x_-^T s_-} x_- + \nabla F(s_-)) = \pi^2. \tag{6.19}$$

Since $p^T q = 0$, we obtain

$$\begin{aligned}\Delta\Phi_\rho(\alpha) &\leq -\alpha\pi^2 + \frac{\alpha^2\sigma^2\pi^2}{2(1-\alpha\bar{\sigma})}\\ &\leq -\alpha\pi^2 + \frac{\alpha^2\sigma^2\pi^2}{1-\alpha\bar{\sigma}}\end{aligned} \tag{6.20}$$

(the weakening here will be useful in Section 7).

Let us choose

$$\alpha := [2\sigma^2 + \bar{\sigma}]^{-1},$$

so that $\alpha\bar{\sigma} < 1$ implying that the next iterates are feasible and

$$\Delta\Phi_\rho(\alpha) \leq -\frac{\pi^2}{4\sigma^2 + 2\bar{\sigma}}. \tag{6.21}$$

Now (6.15) and (6.16) imply that

$$\sigma_{x_-}(-p) \leq |p|_{x_-} \leq \|p\|_{x_-} \leq \sigma\|p\|_z \leq \sigma\pi, \tag{6.22}$$

and similarly $\sigma_{s_-}(-q) \leq \sigma\pi$, so that $\bar{\sigma} \leq \sigma\pi$. We then have

$$\Delta\Phi_\rho(\alpha) \leq -\frac{\pi^2}{4\sigma^2 + 2\sigma\pi} = -\frac{(\pi/\sigma)^2}{4 + 2(\pi/\sigma)}.$$

The expression on the right-hand side is decreasing with π/σ, and Proposition 6.1 shows that this is at least $\sqrt{3}/2$, so

$$\Delta\Phi_\rho(\alpha) \leq -\frac{3/4}{4+\sqrt{3}} < -\frac{1}{8},$$

as required. $\square$

Note that, in contrast to Theorems 4.2 and 4.3, Theorem 6.1 guarantees only a constant decrease in the primal-dual potential function, rather than one that is always at least a constant but likely to be much larger. Yet the proof relies on a "long step"; α depends on "σ"-quantities, related to longest feasible steps, rather than on 2-norms. Can the analysis be improved to yield a larger decrease? The answer is yes, but the result is hard to express succinctly, which is why Theorem 6.1 is stated as it is.

Let us suppose $\rho = 2n$. Then the theorem still guarantees a constant decrease in Φ_ρ, leading to a $O(n \ln \frac{1}{\epsilon})$-iteration algorithm. However, we can expect a larger decrease typically. From Proposition 6.1, we have

$$\begin{aligned}
\pi^2 &\geq \frac{3}{4}\sigma^2 + \frac{n^2 - n}{x_-^T s_-} \\
&= [\frac{3}{4} + (n-1)\frac{v_{\min}^2}{v^T v/n}]\sigma^2,
\end{aligned}$$

where π, σ, v, and $v_{\min}$ are as in (6.14), (6.10), and (6.12). Let us suppose that the current point lies in a wide neighborhood of the central path, so that $\frac{v_{\min}^2}{v^T v/n} \geq .01$. Then we have

$$\pi/\sigma \geq .1\sqrt{n}. \tag{6.23}$$

Next, (6.22) gives

$$\sigma_{x_-}(-p) \leq |p|_{x_-} \leq \|p\|_{x_-} \leq \sigma\|p\|_z \leq \sigma\pi,$$

and similarly for $\sigma_{s_-}(-q)$. Since these chains of inequalities include one relating an ∞-norm to a 2-norm, we can hope that $\bar{\sigma} \leq \sigma\pi$ can be sharpened to

$$\bar{\sigma} \leq 2\sigma\pi\sqrt{\ln n}/\sqrt{n} \tag{6.24}$$

(this is based on an event that happens with probability close to 1 if the components of $X_-^{-1}p$ and $S_-^{-1}q$ have independent identical normal distributions). If both (6.23) and (6.24) hold, then (6.21) gives

$$\begin{aligned}
\Delta\Phi_\rho(\alpha) &\leq -\frac{(\pi/\sigma)^2}{4 + 2(\pi/\sigma)(\bar{\sigma}/[\sigma\pi])} \\
&\leq -\frac{(\pi/\sigma)^2}{4 + (4\sqrt{\ln n}/\sqrt{n})(\pi/\sigma)} \\
&\leq -\frac{(.1\sqrt{n})^2}{4 + (4\sqrt{\ln n}/\sqrt{n})(.1\sqrt{n})} \\
&\leq -\frac{n}{440\sqrt{\ln n}}.
\end{aligned}$$

Thus we have a decrease of $\Omega(n/\sqrt{\ln n})$.

If we choose $\rho = n + (\kappa+1)\sqrt{n}$, where κ is a large constant, a similar analysis shows that Φ_ρ always decreases by at least 1/8, but typically decreases by $\kappa^2/[4 + 2\kappa\sqrt{\ln n}/\sqrt{n}]$.

7 The Linear Complementarity Problem and Infeasible-Interior-Point Methods

In this section we describe two extensions of the Kojima-Mizuno-Yoshise algorithm of Section 6.

The first is to the monotone linear complementarity problem (indeed this was the main focus of [22]): given an $n \times n$ positive semidefinite matrix M and an n-vector a, find $x, s \in I\!R^n$ with

$$s = Mx + a, \quad x \geq 0, \quad s \geq 0, \quad x^T s = 0. \tag{7.1}$$

This fundamental problem includes convex quadratic programming, certain equilibrium problems from economics and engineering, etc.; see Cottle, Pang and Stone [9].

Let $\mathcal{F}^0 := \{(x, s) : s = Mx + a, \quad x > 0, \quad s > 0\}$, and define Φ_ρ on $\mathcal{F}^0$ exactly as in (5.1). The results of Section 5 (eliminating all references to y and the objective functions) remain true, so again we wish to devise an algorithm that decreases Φ_ρ by a constant at each iteration.

In this case, the primal and dual directions are linked by the matrix M, and the appropriate direction-finding subproblem turns out to be

$$\begin{array}{rrcl} \min & \nabla_x \Phi_\rho(x_-, s_-)^T p & + & \nabla_s \Phi_\rho(x_-, s_-)^T q \\ & & + & \frac{1}{2}(\nabla^2 F(z)p + q)^T (p + \nabla^2 F(t) q) \\ & Mp - q & = & 0, \end{array} \tag{7.2}$$

where z and t are again given by (6.4) so that they satisfy (6.3). The KKT conditions for this problem are (with multipliers w)

$$\begin{array}{rcrcrcrcl} & & Mp & - & & q & = & 0 \\ M^T w & + & \nabla^2 F(z) p & + & & q & = & -\nabla_x \Phi_\rho(x_-, s_-) \\ -w & + & p & + & \nabla^2 F(t) q & & = & -\nabla_s \Phi_\rho(x_-, s_-). \end{array}$$

As in Section 6, $\nabla^2 F(z) \nabla_s \Phi_\rho(x_-, s_-) = \nabla_x \Phi_\rho(x_-, s_-)$, so the last two equations imply $M^T w = -\nabla^2 F(z) w$, which gives

$$0 \leq w^T M^T w = -w^T \nabla^2 F(z) w = -\|w\|_z^2 \leq 0.$$

Hence $w = 0$, and the equations defining p and q are $Mp - q = 0$ and the last equation of (6.6) or alternatively (6.7).

Let us show that Theorem 6.1 still remains true in this more general setting (with references to y eliminated), as long as $n + \sqrt{n} \leq \rho \leq 2n$. The main difference is that $p^T q$ may be nonzero. Indeed, $p^T q = p^T M p \geq 0$ since M is positive semidefinite. This yields another second-order term in (6.13) that comes from $x_+^T s_+$:

$$\frac{\alpha^2 \rho p^T q}{{x_-}^T s_-}.$$

Now since ${x_-}^T s_- = \|x_-\|_z^2 \geq \|x_-\|_{x_-}^2/\sigma^2 = n/\sigma^2$ (using (6.3), (6.15), (4.7), and (4.11)), this is at most

$$\frac{\rho}{n}\alpha^2\sigma^2 p^T q \leq \frac{\rho}{n}\frac{\alpha^2\sigma^2}{1-\alpha\overline{\sigma}} p^T q \leq \frac{\alpha^2\sigma^2}{1-\alpha\overline{\sigma}} \cdot 2p^T q.$$

Now, considering (6.19), we see that this extra term can be combined with the bound (6.18) on the other second-order terms, so that (6.20) is still valid. The rest of the proof is the same as before.

The second extension is to an infeasible-interior-point potential-reduction method. Let us return to the setting of linear programming (although the approach also works for the monotone linear complementarity problem). We no longer assume we have a point $(x_-, y_-, s_-) \in \mathcal{F}^0(P) \times \mathcal{F}^0(D)$. Instead, we have $x_- > 0$ and $s_- > 0$ in $I\!R^n$, and $y_- \in I\!R^m$, but the equality constraints may be violated.

Mizuno, Kojima, and Todd [27] propose two algorithms that iterate points of this type while decreasing either Φ_ρ or

$$\Psi_\rho(x,y,s) := \Phi_\rho(x,s) - \ln(x^T s - \sigma\|(Ax-b; A^T y + s - c)\|_2).$$

(In the first case, an extra constraint must be imposed on the step; the second is a pure potential-reduction method.) In either case, the search direction at a point (x_-, y_-, s_-) is (p, r, q), the solution to

$$\begin{array}{rcrcrcl} & & Ap & & & = & b - Ax_- \\ A^T r & & & + & q & = & c - A^T y_- - s_- \\ & & S_- p & + & X_- q & = & \gamma\mu_- e - X_- S_- e \end{array} \qquad (7.3)$$

(compare with (6.5)–(6.7)), where $\mu_- := {x_-}^T s_-/n$ and $\gamma = n/\rho$. Here we only provide an interpretation of these directions as steepest descent directions for Φ_ρ. For the analysis of these methods, see [27].

We must consider the primal and the dual together. The direction p for x is chosen so that $x_- + p$ satisfies the constraints $Ax = b$. Thus $Ap = b - Ax_-$. Similarly $A^T r + q = c - A^T y_- - s_-$. Thus we will not necessarily have $p^T q = 0$. We consider

$$\begin{array}{rcl} \min\limits_{p,q,r} \quad \lambda \nabla_x \Phi_\rho(x_-, s_-)^T p & + & \lambda \nabla_s \Phi_\rho(x_-, s_-)^T q \\ & + & \frac{1}{2}(\nabla^2 F(z) p + q)^T (p + \nabla^2 F(t) q) \\ Ap & = & b - Ax_- \\ A^T r + q & = & c - Ay_- - s_-. \end{array} \tag{7.4}$$

If the KKT conditions are written, it is not hard to see that the multipliers on the constraints are again zero, by reasoning like that for the linear complementarity problem. Moreover, if λ is chosen to be $x_-{}^T s_- / \rho$, it can be seen that these conditions give exactly (7.3). Hence the search directions in [27] can be viewed in a sense as steepest descent directions for Φ_ρ in a combined primal-dual space.

The norm here, as in the problem (7.2) for the linear complementarity problem is given by $\|(p,q)\| := \left[(\nabla^2 F(z) p + q)^T (p + \nabla^2 F(t) q)\right]^{\frac{1}{2}} = \|Z^{-1} p + T^{-1} q\|_2$. This norm was introduced by Kojima-Mizuno-Yoshise in their original paper [22]. Note that, if $(x_-, y_-, s_-) \in \mathcal{F}^0(P) \times \mathcal{F}^0(D)$, then the constraints of (7.4) imply $p^T q = 0$, so $\|(p,q)\|^2 = \|p\|_z^2 + \|q\|_t^2$, and (7.4) separates into the two direction-finding problems of Section 6, with a factor λ which only affects the lengths of the directions. Unfortunately, I can see no justification for the particular value of λ that is necessary for (7.4) to yield the directions in (7.3).

8 Extensions to Certain Nonlinear Problems

Here we outline extensions of the algorithms we have presented to a class of nonlinear optimization problems. Since we are concerned with potential-reduction algorithms, we assume that the problem is given in conic form as

$$(P) \qquad \begin{array}{rcl} \min\limits_x & \langle c, x \rangle & \\ & Ax & = & b \\ & x & \in & K, \end{array}$$

following Nesterov and Nemirovskii [33]. Here x lies in a finite-dimensional real vector space E, and c lies in the dual space E^*. A is a linear transformation from E to another such space Y^*, and $b \in Y^*$. Finally, K is a closed convex cone in E that has nonempty interior and is pointed, i.e., $K \cap (-K) = \{0\}$. While this problem looks special, it is in fact a general form for a constrained convex optimization problem [33].

Let K^* denote the dual cone

$$\{s \in E^* : \langle s, x\rangle \geq 0 \text{ for all } x \in K\}.$$

Then the dual problem can be written

$$(D) \qquad \begin{array}{rrcrcl} \max_{y,s} & \langle b, y\rangle & & & & \\ & A^*y & + & s & = & c \\ & & & s & \in & K^*, \end{array}$$

where $y \in Y$, the dual of Y^*, and $A^* \in L(Y, E^*)$ is the adjoint of A. Conditions guaranteeing strong duality and the attainment of optimal values are given in [33], for example. For our purposes, it is enough to record that sufficient conditions for strong duality are the nonemptiness of at least one of

$$\begin{array}{rcl} \mathcal{F}^0(P) & := & \{x \in \text{int } K : Ax = b\}, \\ \mathcal{F}^0(D) & := & \{(y,s) \in Y \times \text{ int } K^* : A^*y + s = c\}; \end{array}$$

optimal solutions exist if both are nonempty. Note that weak duality is trivial:

$$\begin{array}{rcl} \langle c, x\rangle - \langle b, y\rangle & = & \langle A^*y + s, x\rangle - \langle Ax, y\rangle \\ & = & \langle s, x\rangle \geq 0 \end{array}$$

for all feasible solutions x and (y, s) to (P) and (D).

Two helpful examples to keep in mind are the following. First, if

$$K = \mathbb{R}^n_+ \subseteq E = \mathbb{R}^n, \tag{8.1}$$

then (P) is the standard linear programming problem, E^* can be identified with $\mathbb{R}^n$, and then K^* is again the nonnegative orthant, so that (D) is the usual dual problem. Second, let E be the space $S\mathbb{R}^{n\times n}$ of symmetric matrices of order n and K be the cone $P^{n\times n}$ of symmetric positive semidefinite matrices:

$$K = P^{n\times n} \subseteq E = S\mathbb{R}^{n\times n}. \tag{8.2}$$

Again, E^* can be identified with $S\mathbb{R}^{n\times n}$ with the scalar product $\langle u, v\rangle = \operatorname{tr} uv$, and then the dual cone K^* coincides with $P^{n\times n}$. In this case, (P) and (D) are called semidefinite programming problems, and these include many important applications – see Vandenberghe and Boyd [45].

Nesterov and Nemirovskii [33] give potential-reduction algorithms extending those of Karmarkar and Ye (Sections 4.1 and 6 above) to certain problems of the form (P). They require that K have a suitable barrier function, which we will again denote by F. First, F must be a thrice-differentiable convex barrier function for K: finite on int K, and tending to $+\infty$ if x converges to a point in the boundary of K. Next, $D^2F(x)$ (a mapping, in contrast to the matrix $\nabla^2 F(x)$ we have considered before) should be positive definite:

$$D^2F(x)[v,v] \geq 0 \text{ for all } x \in \text{ int } K, v \in E.$$

This allows us to define a norm (compare (4.17))

$$\|v\|_x := (D^2F(x)[v,v])^{\frac{1}{2}}.$$

The next property is that F should be a *1-self-concordant function.* This requires that

$$|D^3F(x)[v,v,v]| \leq 2 \quad \text{ for all } x \in \text{ int } K, \quad v \in E \text{ with } \|v\|_x = 1. \tag{8.3}$$

An equivalent way of stating this condition is as follows: for every $x \in \text{int } K$ and $v \in E$ with $\|v\|_x = 1$, the function $f(\alpha) := F(x + \alpha v)$ satisfies

$$|f'''(0)| \leq 2. \tag{8.4}$$

By bounding the third derivative, Nesterov and Nemirovskii ensure that the second derivative doesn't change too rapidly, which seems very natural since it is used to define search directions in the algorithm (see also the proof of Theorem 4.1 and (6.15)-(6.16)). To get a feel for the naturalness of (8.3) and (8.4), the reader can consider the logarithmic barrier function $-\ln x$ for the case $K = \mathbb{R}_+ \subseteq E = \mathbb{R}$, and note that (8.3) and (8.4) hold with equality.

Moreover, alternative barriers such as $x^{-\tau}$, $\tau > 0$, fail to satisfy (8.3) and (8.4).

It is not hard to see that

$$F(x) := -\Sigma_j \ln x^{(j)}, \tag{8.5}$$

the logarithmic barrier function of (3.1), satisfies (8.3) and (8.4) for the case (8.1). For the case of positive semidefinite matrices, the function

$$F(x) := -\ln \det x \tag{8.6}$$

(recall that here x is a symmetric matrix of order n) is 1-self-concordant. This is easier to check using (8.4), since $f(\alpha) := F(x + \alpha v) = -\ln \det x - \ln \det(i + \alpha x^{-\frac{1}{2}} v x^{-\frac{1}{2}})$, with i the identity matrix.

The final property used by Nesterov and Nemirovskii is *logarithmic homogeneity*: for some parameter $\nu > 0$, F should satisfy

$$F(\tau x) = F(x) - \nu \ln \tau \tag{8.7}$$

for all $\tau > 0$. The parameter ν determines the complexity of the resulting algorithms. Note that (8.5) and (8.6) both satisfy (8.7), in both cases with $\nu = n$. This is the dimension of the space in case (8.1), and indeed Nesterov and Nemirovskii show that every cone K satisfying the conditions we have stated admits a ν-logarithmically homogeneous self-concordant barrier (they call it a ν-normal barrier) with ν of order the dimension of K. In case (8.2), the dimension of K is $n(n+1)/2$, so here $\nu = n$ is considerably smaller, leading to (at least theoretically) efficient algorithms for semidefinite programming.

The simple property (8.7) leads to some important consequences, which we list here. Let $F'(x) \in E^*$ and $F''(x) \in L(E, E^*)$ satisfy

$$\langle F'(x), v \rangle = DF(x)[v], \quad \langle F''(x)v, w \rangle = D^2F(x)[v, w].$$

Then differentiating (8.7) once and twice with respect to x yields

$$F'(\tau x) = F'(x)/\tau, \quad F''(\tau x) = F''(x)/\tau^2; \tag{8.8}$$

differentiating with respect to τ and then setting $\tau = 1$ gives

$$\langle F'(x), x \rangle = -\nu; \tag{8.9}$$

and differentiating this last with respect to x yields

$$F''(x)x = -F'(x). \tag{8.10}$$

The latter two relations correspond to (4.7) and (4.11).

Just as $F''(x)$ defines the norm

$$\|v\|_x := \langle F''(x)v, v\rangle^{\frac{1}{2}} \tag{8.11}$$

on E, its inverse (in $L(E^*, E)$) defines one on E^*:

$$\|u\|_x^* := \langle u, [F''(x)]^{-1}u\rangle^{\frac{1}{2}}. \tag{8.12}$$

Substituting for either x or $F'(x)$ in (8.9) using (8.10), we find

$$\|x\|_x = \|F'(x)\|_x^* = \sqrt{\nu}. \tag{8.13}$$

(Compare with the paragraph containing (4.17).) This bound on a suitable norm of $F'(x)$ contrasts nicely with the bound (8.3) on $D^3F(x)$; together they imply that F is a ν-self-concordant barrier for K [33].

We should mention the intriguing connections between self-concordant barriers for a cone and its characteristic function, introduced almost forty years ago by Köcher, and since used extensively in several areas of pure mathematics, e.g., to classify homogeneous cones. This connection and its consequences were found by Güler [17], to which we refer for further details.

Given the convex function F defined on int K, we can construct its dual F_* on E^* by

$$F_*(s) := \sup\{-\langle s, x\rangle - F(x) : x \in \text{int } K\}. \tag{8.14}$$

This differs by just a change in sign in s from the usual convex conjugate function. For the cases (8.5) and (8.6), we find respectively

$$\begin{aligned} F_*(s) &= -\Sigma_j \ln s^{(j)} - n, & (8.15)\\ F_*(s) &= -\ln\det s - n, & (8.16) \end{aligned}$$

so that F_* differs by a constant from F. In general, Nesterov and Nemirovskii show that whenever F is a ν-logarithmically homogeneous self-concordant barrier for K, F_* is one for K^*.

Standard results in convex analysis provide the following relations. For any $x \in \text{int } K$, $s \in \text{int } K^*$,

$$-F'(x) \in \text{int } K^*, \quad -F'_*(s) \in \text{int } K; \tag{8.17}$$

we always have (from the definition)

$$F_*(s) + F(x) + \langle s, x\rangle \geq 0, \tag{8.18}$$

and equality holds iff $s = -F'(x)$. Thus

$$F_*(-F'(x)) = \langle F'(x), x\rangle - F(x) = -F(x) - \nu \tag{8.19}$$

(using (8.9)), and similarly

$$F(-F'_*(s)) = -F_*(s) - \nu. \tag{8.20}$$

Also,

$$\begin{array}{ll} F'_*(-F'(x)) = -x, & F'(-F'_*(s)) = -s, \\ F''_*(-F'(x)) = [F''(x)]^{-1}, & F''(-F'_*(s)) = [F''_*(s)]^{-1}. \end{array} \tag{8.21}$$

An important inequality results from (8.18) using (8.7). We have, writing τx for x,

$$\tau\langle s, x\rangle + F(x) + F_*(s) - \nu \ln \tau \geq 0$$

for all $\tau > 0$. Minimizing the left-hand side as a function of τ leads to $\tau = \nu/\langle s, x\rangle$, and substituting this value gives

$$\nu \ln\langle s, x\rangle + F(x) + F_*(s) \geq \nu \ln \nu - \nu. \tag{8.22}$$

This result (with its very trivial proof) extends Proposition 3.1 (the extra "$-\nu$" above comes from the "$-n$" in (8.15)), which was used above in both primal-only and primal-dual potential reduction methods for linear programming. Moreover, it is easy to see that equality holds in (8.22) iff $s = -\mu F'(x)$ for some $\mu > 0$ (or equivalently $x = -\mu F'_*(s)$ for some $\mu > 0$).

We therefore see that several properties of the standard logarithmic barrier function for $\mathbb{R}^n_+$ extend to a general logarithmically homogeneous self-concordant barrier for a possibly non-polyhedral cone K. This opens the way for potential-reduction algorithms for the general problem (8.1). Indeed, as we have noted, Nesterov and Nemirovskii provide extensions of certain algorithms to this setting. However, their analysis does not extend to all the algorithms we have studied above, nor does it cover the long-step variants we have stressed. There seem to be two difficulties. First, the general theory for such barriers provides good approximations for F, F', and F'' (like our Theorem 4.1, for example), but *only* within a small neighborhood of the form $\{x_- + v : \|v\|_{x_-} \leq \beta\}$ for small β. Second, the symmetric primal-dual method requires the existence of $z \in \text{int } K$ satisfying (6.3) for every $x_- \in \text{int } K$, $s_- \in \text{int } K^*$, and it is not clear that such a z exists in the general case.

It appears that further conditions must be placed on the cone K and the barrier F. Nesterov and Todd [34] require that they be *self-scaled*: for each $x, w \in \text{int } K$,

$$\begin{aligned} F''(w)x &\in \text{int } K^*, && (8.23)\\ F_*(F''(w)x) &= F(x) - 2F(w) - \nu. && (8.24) \end{aligned}$$

It is easy to check that these conditions hold for case (8.1), with F and F_* given by (8.5) and (8.15). They also hold in case (8.2), with F and F_* given by (8.6) and (8.16). Indeed, in this case we find $F''(w)x = w^{-1}xw^{-1} \in P^{n\times n}$, with $-\ln\det(w^{-1}xw^{-1}) - n = -\ln\det\, x + 2\ln\det\, w - n$. Finally, the conditions hold for the so-called second-order cone, and for combinations of these examples. Thus, while the conditions (8.23) and (8.24) appear quite special, they hold in important examples.

We do not intend to provide details of how the algorithms and analysis of Sections 2 through 6 (and some of 7) can be extended to the case of self-scaled cones. However, we will provide some insight into the key properties of self-scaled barriers that allow such extensions. The reader should have no difficulty in seeing how the algorithms extend, and should be able to get a feel for how the analysis works. Details can be found in [34].

First, we note that (almost) all expressions can be translated into the present general setting. Here is an incomplete list of appropriate substitutions:

$$\begin{array}{cccccc} n, & x^T s & \longrightarrow & \nu, & \langle s, x\rangle; \\ x^{-1}, & s^{-1} & \longrightarrow & -F'(x), & -F'_*(s); \\ c^T x, & b^T y & \longrightarrow & \langle c, x\rangle, & \langle b, y\rangle; \\ F(s), & F(t) & \longrightarrow & F_*(s)+\nu, & F_*(t)+\nu; \\ \nabla F(x)^T p, & \nabla F(s)^T q & \longrightarrow & \langle F'(x), p\rangle, & \langle q, F'_*(s)\rangle; \\ v^T \nabla^2 F(x) v, & u^T \nabla^2 F(s) u & \longrightarrow & \langle F''(x)v, v\rangle, & \langle u, F''_*(s)u\rangle; \\ x \geq 0, & s \geq 0 & \longrightarrow & x \in K, & s \in K^*; \\ x > 0, & s > 0 & \longrightarrow & x \in \text{int } K, & s \in \text{int } K^*. \end{array}$$

Notable absences here are any expressions involving X, X^{-1}, S, and S^{-1}. However, we can usually avoid primal or dual scaling (as we have done above) by using different norms at the current point. The condition for lying on the central path, $XSe = \mu e$ for some $\mu > 0$, can be translated into $x - \mu s^{-1} = 0$ or $s - \mu x^{-1} = 0$, which becomes $x + \mu F'_*(s) = 0$ or equivalently $s + \mu F'(x) = 0$. Note that we established this as the necessary and sufficient condition for equality in (8.22).

Let us now examine the algorithms and analysis in the paper, to see how they extend to the general case of self-scaled cones.

The results of Section 2 remain true. For Theorem 2.1 we note that $\{x \in K : \langle s_0, x\rangle \leq \gamma\}$ is bounded for $s_0 \in \text{int } K^*$, and that $\max\{\zeta : A^* y + s_0\zeta + s = c, s \in K^*\}$ always has a strictly feasible solution, so that strong duality holds. For Proposition 2.1, the proof is almost the same, but it is necessary to use Theorem 2.1; $\mathcal{F}^0(P)$ nonempty implies that (D), which has finite value, has an optimal solution.

Theorem 3.1 mainly used (3.9), which is replaced by (8.7). We also used $F(x_0 + w) \leq F(x_0)$ for $x_0 \in \text{int } K$ and $w \in K$, which follows from convexity and the fact that $F(x_0 + \lambda w)$, itself a self-concordant function of λ, is unbounded below as $\lambda \to \infty$ if $w \neq 0$, which comes from Proposition 2.3.2 in Nesterov and Nemirovskii [33]. Theorem 3.2 and Corollary 3.1 only required (3.11) of Proposition 3.1, which is replaced by (8.22).

The homogeneity of ϕ in (4.1) follows from (8.7). We can again generate search directions from (DFP), and again they are characterized using (4.2), although (4.3) no longer makes sense. (4.7) and (4.11) are replaced by (8.10) and (8.9) respectively. The fact that $x(\zeta) \in K$ implies $s(\zeta) \in K^*$, below (4.9), follows from (8.23). The norms in (4.17) and (4.18) are defined as in

(8.11) and (8.12), and we define

$$\begin{aligned}\sigma_{x_-}(v) &:= \min\{\sigma \geq 0 : \sigma x_- - v \in K\},\\ |v|_{x_-} &:= \max\{\sigma_{x_-}(v), \sigma_{x_-}(-v)\}\end{aligned}$$

as in one of the expressions in (4.17). Then Proposition 4.2, in the form

$$\langle c, x_-\rangle - \zeta_+ \leq \nu\sigma_{x_-}(p) \leq \nu|p|_{x_-} \leq \nu\|p\|_{x_-}$$

remains true, with the same proof; the last inequality requires $\{x_- + v : \|v\|_{x_-} \leq 1\} \subseteq K$, which again comes from Proposition 2.3.2 of [33]. Hence (4.19) remains valid. Theorem 4.1 remains true, with the same proof. This is a key property of self-scaled barriers. The inequality used in its proof follows form another key property for such barriers: if $\sigma_{x_-}(-v) < 1$,

$$\frac{1}{(1+\sigma_{x_-}(v))^2}F''(x_-) \leq F''(x_- + v) \leq \frac{1}{(1-\sigma_{x_-}(-v))^2}F''(x_-) \qquad (8.25)$$

(see Theorem 4.1 of [34]). (A similar but much weaker result holds for general self-concordant functions, roughly with the σ's replaced by $\|v\|_{x_-}$; see Theorem 2.1.1 of [33].) Then Theorem 4.2 follows as before. The two parts of Proposition 4.4 require slightly modified proofs. For the first, we have

$$\langle F'(x_-), p\rangle \leq \|F'(x_-)\|^*_{x_-}\|p\|_{x_-} \leq \sqrt{\nu}\|p\|_{x_-},$$

using (8.13), while if $\sigma = \sigma_{x_-}(-p)$, $\langle -F'(x_-), \sigma x_- + p\rangle \geq 0$ (since $-F'(x_-) \in K^*$, $\sigma x_- + p \in K$) gives

$$\langle F'(x_-), p\rangle \leq \nu\sigma \leq \nu|p|_{x_-}$$

using (8.9). Then Theorem 4.3 follows as before.

The existence of the central path, used in the proof of Theorem 5.1, is established by Nesterov [32]. For such points, $\Phi_n(x,s) = 0$, as we saw in the conditions for equality in (8.22). Theorem 5.2 only needs (8.22).

For the symmetric primal-dual algorithm of Section 6, a key requirement is the existence of $z \in \operatorname{int} K$ and $t \in \operatorname{int} K^*$ satisfying (6.3). While (6.4) no longer makes sense, z and t exist by virtue of Theorem 3.2 of [34]. The other ingredient is Proposition 6.1, whose proof above used v of (6.12) and its individual components, and thus does not extend to arbitrary self-scaled cones. The general result is given as Theorem 5.2 of [34], with a very intricate proof; a simpler derivation appears as a remark following Theorem 4.2 of [35]. Finally, (6.15) and (6.16) follow from (8.25). Since $\sigma x_- - z \in K$, $\sigma_z(-(\sigma x_- - z)) = 0$, so $F''(\sigma x_-) \leq F''(z)$, which with (8.9) gives (6.15). Thus the proof of Theorem 6.1 goes through as before. We remark that the observations following the proof still hold qualitatively if not quantitatively; a larger value of ρ is likely to lead to a larger decrease in the potential function.

Finally, we note that the argument of the first part of Section 7 carries over directly, so that we have an $O\left(\sqrt{\nu}\ln\frac{1}{\epsilon}\right)$-iteration algorithm for a monotone linear complementarity problem defined with ν-self-scaled cones. This seems to be a new result. The derivation of the directions for the infeasible-interior-point method described in the second half of Section 7 also carries over, but the analysis is likely to be much more complicated, and it is unknown whether the same complexity bound as in [27] holds.

9 Open Problems

Here we briefly describe three open problems related to potential-reduction methods. The first two are concerned with feasible-interior-point algorithms, while the last addresses the important subject of infeasible-interior-point methods.

There has been great interest in the last few years in devising interior-point methods that both are globally (and preferably polynomially) convergent and possess fast asymptotic convergence. See, for example, Monteiro and Wright [30] and the references therein. Most of these methods are based on path-following ideas, and many use a tight neighborhood of the central path for their iterates. Unfortunately, such restrictions add to the divergence of theory and practice, since they are very undesirable in practical computation. Because potential-reduction methods allow long steps provided the potential function decreases, one could hope that these methods would also allow fast asymptotic convergence. Unfortunately, results of this type are rare.

McShane [25] used the primal-dual potential function (with a variable parameter ρ) to motivate the search directions of his method, but the parameter and step size were chosen to stay close to the central path. Ye et al. [48] showed that a wide range of parameters ρ could be used in choosing the search directions, while still ensuring that a constant reduction of the potential function for a fixed value of ρ could be attained. Their motivation was to allow freedom in directions and step sizes to permit fast local convergence, but no superlinear results were obtained. Monteiro and Wright [30] have recently given a potential-reduction method with superlinear convergence. The search directions are the primal-dual affine-scaling directions, while the step size comes from an Armijo-type rule on the primal-dual potential function, using a parameter ρ very close to n. However, the first step size tested is defined by a rule that has nothing to do with the value of the potential function, so the method is not a pure potential-reduction method. Monteiro and Wright show that asymptotically, a constant reduction in the potential function is achieved by this step size; indeed, the reduction obtained tends to infinity. Also, the step size converges to the full step to the boundary. This gives strong evidence that fast convergence is possible for a pure potential-reduction algorithm, but we suspect that it may only be R- rather than Q-superlinear.

The second open problem concerns other ways to combine barrier terms and functions of the optimality or duality gap to form potential functions. Clearly the functions we have studied are the most natural; the logarithm of the duality gap will lead to global linear convergence if a constant decrease can be assured, while the logarithmic barrier function possesses all the nice properties of self-scaled barriers in general and hence leads to polynomial convergence. However, the method of Monteiro and Wright mentioned above as well as methods using affine-scaling directions together with step sizes defined by the contours of a potential function as in Mizuno and Nagasawa [28] and Tunçel [43] use different strategies to achieve global convergence. Nesterov [32] also uses neighborhoods defined by a potential function in his long-step path-following method. In particular, Tunçel shows a relationship between neighborhoods and potential functions that might be useful in developing new potential functions for which earlier "constrained" potential-reduction methods become "pure."

Finally, we feel infeasible-interior-point primal-dual potential-reduction methods need much more development. (This is also true for primal-only methods. Even though these are quite well-developed (see [2, 3, 11, 38, 39]), the results obtained are rather weak if either the primal or the dual problem is infeasible. Moreover, these methods do not seem to perform very well in practice, mainly because of difficulties in updating lower bounds. Applying such methods to an artificial combined primal-dual problem as in Anstreicher [5] seems much more attractive, but there are computational costs to this approach as well as the lack of flexibility mentioned below caused by tying the primal and dual problems together.) In the feasible case, a primal-dual potential-reduction method can use any technique to decrease the potential function; different step sizes can be taken in the primal and dual spaces, and these directions do not have to come from a particular direction-finding subproblem. The same is true for the infeasible-interior-point methods mentioned in Section 7, although some extra restrictions on the step sizes may be necessary. But the convergence results of Mizuno et al. [27] are only powerful if a bound is known on the size of optimal solutions. However, if no strictly feasible initial solutions are available, we would prefer our algorithms to approximate optimal solutions if they exist, and otherwise to give some indication of the infeasibility of the primal or dual problem; see Todd and Ye [42]. With this strengthened termination criterion, it appears that very few interior-point methods can guarantee convergence. One method that appears attractive theoretically is the path-following algorithm using a homogeneous self-dual formulation developed by Ye et al. [49]. We can also apply a primal-dual potential-reduction method to this formulation, with a mild restriction on the step size to guarantee that strictly complementary solutions will be found. But this formulation links together the primal and dual problems, and hence the freedom usually available in the feasible case seems not to be present any more; this is a considerable practical disadvantage. One interesting paper on this subject is Tütüncü [44], but more work needs to be done, both for linear programming and for its extensions.

Acknowledgement:
The author would like to thank the anonymous referees for their very constructive comments and suggestions.

References

[1] K. M. Anstreicher, "A monotonic projective algorithm for fractional linear programming," *Algorithmica* 1 (1986) 483–498.

[2] K. M. Anstreicher, "A combined phase I – phase II projective algorithm for linear programming," *Mathematical Programming* 43 (1989) 209–223.

[3] K. M. Anstreicher, "A combined phase I – phase II scaled potential algorithm for linear programming," *Mathematical Programming* 52 (1991) 429–439.

[4] K. M. Anstreicher, "On the performance of Karmarkar's algorithm over a sequence of iterations," *SIAM Journal on Optimization* 1 (1991) 22–29.

[5] K. M. Anstreicher, "On interior algorithms for linear programming with no regularity assumptions," *Operations Research Letters* 11 (1992) 209–212.

[6] D. A. Bayer and J. C. Lagarias, "Karmarkar's linear programming algorithm and Newton's method," *Mathematical Programming* 50 (1991) 291–330.

[7] D. Bertsimas and X. Luo, "On the worst case complexity of potential reduction algorithms for linear programming," Working Paper 3558-93, Sloan School of Management, MIT, Cambridge, MA 02139, USA (1993).

[8] R. A. Bosch and K. M. Anstreicher, "On partial updating in a potential reduction linear programming algorithm of Kojima, Mizuno and Yoshise," *Algorithmica* 9 (1993) 184–197.

[9] R.W. Cottle, J.-S. Pang, and R.E. Stone, *The Linear Complementarity Problem* (Academic Press, Boston, 1992).

[10] R. M. Freund, "Polynomial–time algorithms for linear programming based only on primal scaling and projected gradients of a potential function," *Mathematical Programming* 51 (1991) 203–222.

[11] R. M. Freund, "A potential–function reduction algorithm for solving a linear program directly from an infeasible 'warm start'," *Mathematical Programming* 52 (1991) 441–466.

[12] G. de Ghellinck and J. P. Vial, "A polynomial Newton method for linear programming," *Algorithmica* 1 (1996) 425–453.

[13] D. Goldfarb and M. J. Todd, "Linear Programming," in: G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, eds., *Optimization*, volume 1 of *Handbooks in Operations Research and Management Science* (North Holland, Amsterdam, The Netherlands, 1989) pp. 141–170.

[14] C. C. Gonzaga, "Polynomial affine algorithms for linear programming," *Mathematical Programming* 49 (1990) 7–21.

[15] C. C. Gonzaga, "Large steps path–following methods for linear programming, Part II : Potential reduction method," *SIAM Journal on Optimization* 1 (1991) 280–292.

[16] C. C. Gonzaga, "Path following methods for linear programming," *SIAM Review* 34 (1992) 167–227.

[17] O. Güler, "Barrier functions in interior-point methods," Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD (1994).

[18] H. Imai, "On the convexity of the multiplicative version of Karmarkar's potential function," *Mathematical Programming* 40 (1988) 29–32.

[19] J. Ji and Y. Ye, "A complexity analysis for interior–point algorithms based on Karmarkar's potential function," *SIAM Journal on Optimization* 4 (1994) 512–520.

[20] N. K. Karmarkar, "A new polynomial–time algorithm for linear programming," *Combinatorica* 4 (1984) 373–395.

[21] M. Kojima, N. Megiddo, and Y. Ye, "An interior point potential reduction algorithm for the linear complementarity problem," *Mathematical Programming* 54 (1992) 267–279.

[22] M. Kojima, S. Mizuno, and A. Yoshise, "An $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems," *Mathematical Programming* 50 (1991) 331–342.

[23] E. Kranich, "Interior point methods for mathematical programming : A bibliography," available through NETLIB: send e-mail to netlib@research.att.com containing the message "send intbib.bib from bib" and/or "send intbib.bbl from bib."

[24] I. J. Lustig, R. E. Marsten, and D. F. Shanno, "Interior point methods : Computational state of the art," *ORSA Journal on Computing* 6 (1994) 1–14.

[25] K. A. McShane, "Superlinearly convergent $O(\sqrt{n}L)$-iteration interior-point algorithms for linear programming and the monotone linear complementarity problem," *SIAM Journal on Optimization* 4 (1994) 247–261.

[26] S. Mizuno, "$O(n^{\rho}L)$ iteration $O(n^3L)$ potential reduction algorithms for linear programming," *Linear Algebra and Its Applications* 152 (1991) 155–168.

[27] S. Mizuno, M. Kojima, and M. J. Todd, "Infeasible–interior–point primal–dual potential–reduction algorithms for linear programming," *SIAM Journal on Optimization* 5 (1995) 52–67.

[28] S. Mizuno and A. Nagasawa, "A primal–dual affine scaling potential reduction algorithm for linear programming," *Mathematical Programming* 62 (1993) 119–131.

[29] R. D. C. Monteiro, "Convergence and boundary behavior of the projective scaling trajectories for linear programming," in: J. C. Lagarias and M. J. Todd, eds., *Mathematical Developments Arising from Linear Programming : Proceedings of a Joint Summer Research Conference held at Bowdoin College, Brunswick, Maine, USA, June/July 1988*, volume 114 of *Contemporary Mathematics* (American Mathematical Society, Providence, Rhode Island, USA, 1990) pp. 213–229.

[30] R. D. C. Monteiro and S. J. Wright, "Superlinear primal–dual affine scaling algorithms for LCP," *Mathematical Programming* 69 (1995) 311–333.

[31] J. L. Nazareth, "Quadratic and conic approximating models in linear programming," *Mathematical Programming Society COAL Bulletin* 23 (1994).

[32] Yu. E. Nesterov, "Long-step strategies in interior point potential-reduction methods," Department SES COMIN, Geneva University, Switzerland (1993).

[33] Yu. E. Nesterov and A. S. Nemirovskii, *Interior Point Polynomial Methods in Convex Programming : Theory and Algorithms* (SIAM, Philadelphia, USA, 1993).

[34] Yu. E. Nesterov and M. J. Todd, "Self-scaled barriers and interior-point methods for convex programming," Technical Report No. 1091, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853-3801, USA (1994).

[35] Yu. E. Nesterov and M. J. Todd, "Primal-dual interior-point methods for self-scaled cones," Technical Report No. 1125, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853-3801, USA (1995).

[36] M. J. D. Powell, "On the number of iterations of Karmarkar's algorithm for linear programming," *Mathematical Programming* 62 (1993) 153–197.

[37] K. Tanabe, "Centered Newton method for mathematical programming," in: M. Iri and K. Yajima, eds., *System Modelling and Optimization : Proceedings of the 13th IFIP Conference, Tokyo, Japan, Aug./Sept. 1987*, volume 113 of *Lecture Notes in Control and Information Sciences* (Springer Verlag, Berlin, Germany, 1988) pp. 197–206.

[38] M. J. Todd, "On Anstreicher's combined phase I – phase II projective algorithm for linear programming," *Mathematical Programming* 55 (1992) 1–15.

[39] M. J. Todd, "Combining phase I and phase II in a potential reduction algorithm for linear programming," *Mathematical Programming* 59 (1993) 133–150.

[40] M. J. Todd and B. P. Burrell, "An extension of Karmarkar's algorithm for linear programming using dual variables," *Algorithmica* 1 (1986) 409–424.

[41] M. J. Todd and Y. Ye, "A centered projective algorithm for linear programming," *Mathematics of Operations Research* 15 (1990) 508–529.

[42] M. J. Todd and Y. Ye, "Approximate Farkas lemmas and stopping rules for iterative infeasible-point algorithms for linear programming," Technical Report No. 1109, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853-3801, USA (1994).

[43] L. Tunçel, "Constant potential primal–dual algorithms : A framework," *Mathematical Programming* 66 (1994) 145–159.

[44] R. H. Tütüncü, "A Practical Infeasible-Interior-Point Potential-Reduction Algorithm for Linear Programming," Technical Report No. 1136, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853-3801, USA (1995).

[45] L. Vandenberghe and S. Boyd, "Semidefinite programming," Technical Report, Information Systems Laboratory, Dept. of Electrical Engineering, Stanford University, Stanford, CA 94305, USA (1994).

[46] Y. Ye, "An $O(n^3L)$ potential reduction algorithm for linear programming," *Mathematical Programming* 50 (1991) 239–258.

[47] Y. Ye, "A potential reduction algorithm allowing column generation," *SIAM Journal on Optimization* 2 (1992) 7–20.

[48] Y. Ye, K. O. Kortanek, J. A. Kaliski, and S. Huang, "Near–boundary behavior of primal–dual potential reduction algorithms for linear programming," *Mathematical Programming* 58 (1993) 243–255.

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