ANOTHER DERIVATION OF THE KARMARKAR DIRECTION FOR LINEAR PROGRAMMING

Michael J. Todd*, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York, USA

ABSTRACT
This paper provides another derivation of the Karmarkar direction for linear programming. It is strongly motivated by derivations of Gonzaga, but we show how the direction can be viewed as a steepest descent direction in the original feasible region corresponding to a metric different from the Euclidean one. We show that a fixed decrease in the potential function can be obtained by taking a step in this direction, as long as a certain assumption holds. We give an example showing that such a restriction is necessary, and discuss two ways to remove it.

Key words: Linear programming, interior-point methods, Karmarkar direction, steepest descent.

RESUMEN
Este artículo presenta una nueva derivación de la dirección Karmarkar para la programación lineal. Está motivada fuertemente por las derivaciones de Gonzaga. Nosotros mostramos cómo la dirección puede interpretarse como la dirección de descenso más empinada en la región factible original si se utiliza una métrica diferente de la Euclideana. Demostramos que se puede lograr un decremento fijo en la función de potencia tomando un paso en esta dirección, bajo cierta condición. Damos un ejemplo que demuestra que esta condición es necesaria, y exponemos dos maneras de eliminarla.

MSC: 90C05, 90C51.

1. INTRODUCTION

The last fifteen years has seen a revolution in algorithms for solving linear and certain convex programming problems. General discussions of these so-called interior-point methods, mostly in the context of linear programming, can be found in the texts by Roos, Terlaky, and Vial (1997) Vanderbei (1997), Wright (1997), and Ye (1997), while the monographs of Nesterov and Nemirovski (1994) and Renegar (2001) provide a deeper treatment of the application of these methods in a general setting. A survey of path-following methods can be found in Gonzaga (1992), while two surveys of potential-reduction algorithms appear in Anstreicher (1996) and Todd (1996).

The modern theory of interior-point methods originates with the work of Karmarkar (1985), whose algorithm was the first potential-reduction method. It also employed a projective scaling at each iteration, which made it less intuitive than the closely related (and much earlier) affine-scaling method of Dikin (1967) or the later affine-scaling potential-reduction algorithm of Gonzaga (1990). However, in contrast to Dikin's method, Karmarkar established polynomial convergence for his, by showing that the potential function he defines is reduced by a constant at each iteration. While Karmarkar's method has been supplanted by more efficient algorithms for practical computation, it remains of interest for historical reasons, and for the ideas it introduced. A nice elementary treatment has recently been given by Gonzaga (2002).

In this paper we give an alternate and more intuitive derivation of the search direction chosen in Karmarkar's algorithm for a standard-form linear programming problem, which makes clear its very close relationship to the simpler affine potential-reduction algorithms. Our derivation is strongly motivated by and closely related to several proposed by Gonzaga; see p. 162 in Gonzaga (1989), p. 162 in Gonzaga (1991a), and p. 222 in Gonzaga (1991b). However, while Gonzaga looks at the intersection of a spherical neighborhood and the cone generated by the feasible region, we keep the original feasible region and intersect it with a neighborhood that is a spherical cone. In this way, we can view Karmarkar's direction as a steepest descent direction with respect to a certain metric and we describe this precisely. The motivation here, of taking as a neighborhood of the current interior point not an ellipsoid but rather an ellipsoidal cone, was apparently first discussed by Megiddo (1985) in the context of a modification of the

*Research supported in part by NSF and ONR.

affine-scaling algorithm of Dikin (1967). See also p. 171 in Gonzaga (1991a) (note that "f" in (30) should be "r") and Padberg (1985).

We also prove directly that a fixed decrease in the potential function can be obtained by taking a step of an appropriate length in this direction. This proof is again very similar to those of Gonzaga (1989) and (1991a). We need, as he did, to make an additional assumption, which follows if the feasible region is compact. We provide an example showing that this restriction is necessary.

There are two ways to remove this unpleasant restriction. One is to use a monotonic variant of the direction, as discussed by Gonzaga (1989). The other is to convert the given standard-form linear programming problem into homogeneous form using an extra variable, as discussed by Anstreicher (1986), de Ghellinck and Vial (1986), Gay (1987), Jensen and Steger (1985), and Ye and Kojima (1987). As noted in Todd (1991), the latter approach is equivalent to adding a dummy variable set equal to one to the original standard-form problem and then following Gonzaga's approach.

2. THE KARMARKAR DIRECTION

We are concerned with the standard-form linear programming problem

$$\text{(P)} \qquad \text{minimize} \quad c^T x$$
$$\text{subject to} \quad Ax = b,\ x \geq 0,$$

where $A \in R^{mxn}$, $b \in R^m$ and $c \in R^n$. We assume that $F^0(P) := \{x \in R^n: Ax = b, x > 0\}$ is nonempty, that we have an initial point in this set, and that (P) has an optimal solution. Let us also assume that the objective function is not constant on the feasible region of (P); if so, this can easily be detected at the first iteration, and the algorithm halted. We further assume that b is nonzero; otherwise, the feasible region is a cone, and given that (P) has an optimal solution, the origin is optimal.

Interior-point methods for (P) generate a sequence of points in $F^0(P)$. At each iteration, given the current point $\hat{x} \in F^0(P)$, most methods scale the problem so that the current point is $e: = (1,1,...,1)^T$. To simplify the notation, we will assume that the current point is itself e.

Karmarkar's method tries to drive a potential function down to $-\infty$. Let v(P) denote the optimal value of (P), and suppose we have a lower bound z on v(P). By our assumptions, for any $x \in F^0(P)$, we have $c^T x > z$. For any $q \geq n$, let us define

$$\phi_q(x, z): = q\ln(c^T x - z) - \ln(x), \tag{1}$$

where

$$\ln(x): = \sum_{j=1}^{n} \ln(x_j). \tag{2}$$

It is helpful also to write ϕ_q in another form. We can easily find a vector g such that $g^T x = 1$ for all feasible x, for example by scaling a row of Ax = b corresponding to a nonzero component of b. In any case, g will be in the row space of A . Then we can alternatively write

$$\phi_q(x, z):= q\ln((c - zg)^T x) - \ln(x). \tag{3}$$

A natural way to choose the search direction is as the steepest descent direction for the function $\phi_q(.,z)$ for a certain q and z. This direction is the solution to the problem

$$\text{(SP)} \qquad \text{minimize} \quad \nabla_x \phi_q (e,z)^T d$$
$$\text{subject to} \quad Ad = 0$$
$$\|d\| \leq 1,$$

and is a positive multiple of $-P\nabla_x\phi_q(e, z)$. Here P denotes the orthogonal projection operator onto the null space of A; we also write v_P for the projection Pv of a vector v. (Note that, for any vectors u and v, we have $u_p^T v_p = u^t v_p = u_p^T v, u_p^T(v - v_p) = 0$, and $\| v_p \|^2 + \| v - v_p \|^2 = \| v \|^2$). We can alternatively define the steepest descent direction as the limit of the solutions (suitably scaled) to

$$
\begin{array}{lll}
(SP(\varepsilon)) & \text{minimize} & \phi_q(e + d) \\
 & \text{subject to} & Ad = 0, \\
 & & \| d \| \le \varepsilon,
\end{array}
$$

as $\varepsilon \downarrow 0$. The result is the same. This search direction is used in the affine potential-reduction algorithm of Gonzaga (1990), for q equal to $n + \sqrt{n}$. (Note that most of Gonzaga (1990) assumes that the optimal value is known to be zero, and uses z = 0, but that the final section shows how lower bounds on v(P) can be used and updated.)

In this derivation, the trust region constraint ($||d|| \le 1$ in (SP) or $||d|| \le \varepsilon$ in (SP(ε))) is added because the potential function ϕ_q is nonlinear, and so the linear approximation used in the objective of (SP) is only accurate close to e. However, if q = n, as in Karmarkar's original algorithm, then the potential function

$$\phi_n(x, z) = n \ln(c - zg)^T x) - \ln(x)$$

is homogeneous of degree zero in x. Moreover, the first-order Taylor approximation at the point x = e can easily be seen to be accurate for any point that is a positive multiple of e. Hence perhaps a more suitable trust region to use for this function is a neighborhood of e that is a spherical cone of small semi-angle. (An alternative trust region would seem to be a cylinder with spherical cross-section centered on the ray $\{\sigma e: \sigma \in R\}$ the same result as obtained below would hold in this case, but the neighborhood seems less suitable because the approximation is very inaccurate for points in the boundary of the cylinder near the origin.) Similar arguments can be made for (SP(ε)).

Such a spherical cone, with semi-angle $\arccos((1+\varepsilon)^{-1/2})$ can be written as the set of x satisfying

$$\frac{x^T e}{(x^T x)^{\frac{1}{2}}} \ge (1+\varepsilon))^{-1/2} \sqrt{n}\,.$$

Alternatively, this is the set of x with $x^T e \ge 0$ and

$$x^T \left(I - (1+\varepsilon) \frac{ee^T}{n} \right) x \le 0.$$

Now write x = e + d. Then x satisfies the linear constraints Ax = b and lies in this cone if and only if

$$
\begin{array}{c}
Ad = 0, \ e^T d \ge -n, \\
d^T \left(I - (1+e) \frac{ee^T}{n} \right) d - 2\varepsilon e^T d \le n\varepsilon.
\end{array}
\qquad (4)
$$

Observe that Ad = 0 implies that $e^T d = e_p^T d$, so we can replace "e" by "e_P" in the above constraints.

Next we use the following result.

Lemma 1. If b is nonzero, $I - \frac{e_p e_p^T}{n}$ is positive definite.

Proof: As a symmetric rank-one update of the identity, I - $\frac{e_p e_p^T}{n}$ is positive definite iff its determinant is positive. But the latter is 1 - $\frac{e_p^T e_p}{n} = \frac{n - e_p^T e_p}{n} = \frac{\|e\|^2 - \|e_p\|^2}{n} = \frac{\|e - e_p\|^2}{n}$. Since Ae = b ≠ 0 = Ae_p, this quantity is positive as desired.

From the lemma, it follows that for any sufficiently small ε, the matrix I - (1 + ε) $\frac{e_p e_p^T}{n}$ is also positive definite. Hence, for ε small, any d feasible in Gay (1987) is correspondingly small (also see the argument following (SP'(ε)) below). Thus for any sufficiently small ε, the constraint $e^T d \geq -n$ follows from the other constraints in Gay (1987). Hence, x = e + d satisfies Ax = b and lies in the conical neighborhood of e iff d lies in the null space of A and a small ellipsoid, for all sufficiently small ε. We can now proceed in either of two ways, corresponding to the two problems (SP) and (SP(ε)) above.

First, we see that as ε converges to 0, the ellipsoidal constraint on d tends to that corresponding to the metric defined by the positive definite matrix I – $\frac{e_p e_p^T}{n}$. Hence it is natural to choose our search direction d as the solution to

$$\begin{array}{lll} \text{(SP')} & \text{minimize} & \nabla_x \phi_n(e,z)^T d \\ & \text{subject to} & Ad = 0, \\ & & d^T(I - \left(I - \frac{e_p e_p^T}{n}\right) d \leq 1. \end{array}$$

Note that (SP') differs from (SP) only in that the Euclidean metric has been replaced by that determined by I - $\frac{e_p e_p^T}{n}$ (and q has been specialized to n).

Theorem 1. Let Δ: = $c^T e$ - z be the current duality gap. Then we have:

a) The solution to subproblem (SP) is a positive multiple of

$$\begin{aligned} d &= - P\nabla_x \phi_q(e, z) \\ &= - (q/\Delta)c_p + e_p. \end{aligned} \tag{5}$$

b) The solution to subproblem (SP') is a positive multiple of

$$\hat{d} = -(n/\Delta)c_p + \frac{n - (n/\Delta)c_p^T e_p}{n - e_p^T e_p} e_p. \tag{6}$$

Proof:

a) Since Ad = 0 for all feasible solutions, we may change the objective in (SP) to min $(P\nabla_x \phi_q(e, z))^T d$. Then the optimal solution to this problem, considering only the constraint ||d|| ≤ 1, is a positive multiple of d = - $P\nabla_x \phi_q(e,z)$. But this solution also satisfies the constraint Ad = 0, so is the optimal solution to the original problem (SP) also.

The second expression for d is obtained by using the definition of ϕ_q in Anstreicher (1986). Alternatively, if (3) is used, the first term becomes $(q/\Delta)(c_p - zg_p)$; but since g is in the row space of A, its projection vanishes and the same expression results.

b) Again we change the objective to min $(P\nabla_x \phi_n(e, z)^T d$. Also, let

$$M: = I - \frac{e_p e_p^T}{n}. \tag{7}$$

Then the optimal solution to the problem (SP'), considering only the constraint $d^T M d \le 1$, is a positive multiple of $\hat{d} := -M^{-1} P \nabla_x \phi_n(e,z)$. But from the Sherman-Morrison formula,

$$M^{-1} = I + \frac{e_p e_p^T}{n - e_p^T e_p}, \tag{8}$$

so we find

$$\begin{aligned}\hat{d} &= -M^{-1} P \nabla_x \phi_n(e,z) \\ &= -\left(I + \frac{e_p e_p^T}{n - e_p^T e_p}\right) P \nabla_x \phi_n(e,z) \\ &= -\left(I + \frac{e_p e_p^T}{n - e_p^T e_p}\right)((n/\Delta)\, c_p - e_p) \\ &= -(n/\Delta) c_p + \frac{n - (n/\Delta) c_p^T e_p}{n - e_p^T e_p} e_p. \end{aligned} \tag{9}$$

Moreover, we see that $\hat{d}$ automatically satisfies the omitted constraint $Ad = 0$, so that it is a positive multiple of the solution to (SP').

We wrote $\hat{d}$ in the form given in (b) above to show its similarity to the direction d in (a), which is familiar from the affine potential-reduction method. However, note that $\hat{d}$ is just a positive scalar multiple of

$$\hat{d} = -c_p + \frac{\Delta - c_p^T e_p}{n - e_p^T e_p} e_p, \tag{10}$$

which is the Karmarkar direction for the standard-form problem derived in Lemma 3.1 of Gonzaga (1991a).

For our second method, we proceed directly from the constraints (4), postponing taking limits. Let us choose $\hat{\varepsilon}$ sufficiently small that the smallest eigenvalue of $I - (1 + \varepsilon)\frac{e_p e_p^T}{n}$ is at least some fixed positive $\lambda = \lambda(\hat{\varepsilon})$ independent of ε whenever $0 < \varepsilon \le \hat{\varepsilon}$, cf Lemma 1. Then we consider, for such ε,

$$\begin{aligned} (SP'(\varepsilon))\ \text{minimize}\quad & \phi_n(e + d, z) \\ \text{subject to}\quad & Ad = 0,\ e_p^T d \ge -n, \\ & d^T\left(I - (1+\varepsilon)\frac{e_p e_p^T}{n}\right) d - 2\varepsilon e_p^T d \le n\varepsilon. \end{aligned}$$

Note that any d satisfying the last inequality above also satisfies

$$\lambda \|d\|^2 - 2\varepsilon\sqrt{n}\, \|d\| \le n\varepsilon,$$

so that

$$||d|| \le \frac{\varepsilon\sqrt{n} + \sqrt{n\varepsilon^2 + \lambda n\varepsilon}}{\lambda},$$

which tends to zero as $\varepsilon \downarrow 0$. Henceforth assume that ε is chosen sufficiently small that this bound on $||d||$ is less than 1. Then the constraint $e_p^T d \ge -n$ is automatically strictly satisfied, and e + d must be positive in each component, so that (SP'(ε)) is the minimization of a continuous function on a compact set and has an optimal solution, which we denote by $\bar{d}(\varepsilon)$. Moreover, since d = 0 satisfies the only nonlinear constraint of (SP'(ε)) strictly, and the constraints are convex, the Karush-Kuhn-Tucker conditions hold at $\bar{d}(\varepsilon)$. Hence for some $\mu = \mu(\varepsilon) \in R$ and $y = y(\varepsilon) \in R^m$ we have

$$\nabla_x \phi_n(e + \bar{d}(\varepsilon), z) + A^T y + \mu\left(2\left(I - (1+\varepsilon)\frac{e_p e_p^T}{n}\right)\bar{d}(\varepsilon) - 2\varepsilon e_p\right) = 0,$$

$$A\bar{d}(\varepsilon) = 0, \tag{11}$$

$$\mu \ge 0,\ \mu\left(\bar{d}(\varepsilon)^T\left(I - (1+\varepsilon)\frac{e_p e_p^T}{n}\right)\bar{d}(\varepsilon) - 2\varepsilon e_p^T \bar{d}(\varepsilon) - n\varepsilon\right) = 0.$$

Now multiply the first equation of (11) by $\bar{d}(\varepsilon)^T$ and use the other equations to deduce

$$\mu(\ 2n\varepsilon + 2\varepsilon e_p^T \bar{d}(\varepsilon)) + \bar{d}(\varepsilon)^T \nabla_x \phi_n(e + \bar{d}(\varepsilon), z = 0.$$

If we now take limits as $\varepsilon \downarrow 0$, and recall that $\bar{d}(\varepsilon) \to 0$ so that $\nabla_x\phi_n(e + \bar{d}(\varepsilon), z)$ remains bounded, we find that $\mu\varepsilon \to 0$ also. Now let

$$\hat{d}(\varepsilon) := 2\mu\bar{d}(\varepsilon),$$

and multiply the first equation of (11) by the projection matrix P to get

$$\left(I - (1+\varepsilon)\frac{e_p e_p^T}{n}\right)\hat{d}(\varepsilon) = -P\nabla_x \phi_n(e + \bar{d}(\varepsilon), z) + 2\mu\varepsilon e_p \tag{12}$$

Taking limits in (12) and using (9) and $\mu\varepsilon \to 0$, we arrive at

Theorem 2. As $\varepsilon \downarrow 0$, a positive multiple of the optimal solution to (SP'(ε)) converges to the direction $\hat{d}$ of (6).

Hence we have seen that the Karmarkar direction for standard-form problems can be derived by either minimizing the potential function over the intersection of the feasible region with an infinitesimally small spherical cone centered at the current solution e, or by taking the steepest descent direction for this function at e, where the metric is defined by the matrix $I - \frac{e_p e_p^T}{n}$.

3. CONVERGENCE ANALYSIS

In this section we show that the potential function can be reduced by a fixed constant at each iteration, as long as a certain assumption holds, which follows if the feasible region is compact. This then yields the standard convergence results. We give an example showing that such a potential reduction may not be achievable if the feasible region is unbounded, and provide a simple remedy. To show this potential reduction, we must have updated the lower bound z on v(P) suitably. Such bounds come from linear programming duality. The dual of (P) is

$$\text{(D)} \qquad \text{maximize} \quad b^T y$$

$$\text{subject to} \quad A^T y + s = c, \ \ s \geq 0,$$

and if (y,s) is feasible in (D), the corresponding duality gap with respect to x = e is $c^Te - b^Ty = e^T(c - A^Ty) = e^Ts$. Note that for any vector $v \in R^n$, Pv differs from v by a vector in the row space of A; we write y_v for a vector in R^m with $P_v = v - A^Ty_v$. We may assume that in calculating c_p and e_p, we also have y_c and y_e. Now observe that for any ν, $(y(\nu) := y_c - \nu y_e,\ s(\nu) := c_p + \nu(e - e_p))$ is feasible in the equality constraint of (D), and if s(ν) is nonnegative, the corresponding duality gap is $e^Ts(\nu) = e^Tc_p + \nu\|e - e_p\|^2$. Hence, if there is any ν such that s(ν) is nonnegative, we choose the smallest such ν and then update z if necessary so that

$$z \geq c^Te - e^Ts(\nu), \ \ \Delta \leq e^Ts(\nu). \tag{13}$$

Observe that, if an update is performed so that $\Delta = e^Ts(\nu)$, then $\nu = \dfrac{\Delta - c_p^Te_p}{n - e_p^Te_p}$, and this is the coefficient of e_p in (10).

From (9), we have

$$\hat{d} = -P\nabla\phi - \frac{e_p^TP\nabla\phi}{\|e - e_p\|^2} e_p = -P\nabla\phi - \rho e_p, \tag{14}$$

where $\nabla\phi$ is shorthand for $\nabla_x \phi_n(e, z)$ and

$$\rho := \frac{e_p^TP\nabla\phi}{\|e - e_p\|^2} = \frac{e^TP\nabla\phi}{\|e - e_p\|^2} = \frac{e_p^T\nabla\phi}{\|e - e_p\|^2} \tag{15}$$

In this section, ϕ_n is always defined using (3) rather than (1). The difference is significant since sometimes $\nabla\phi$ appears without the projection P, and then the result depends on which definition is used. We find it very convenient to consider also the direction

$$\begin{aligned} \bar{d} &:= \hat{d} + \rho e \\ &= -P\nabla\phi + \rho(e - e_p) \\ &= -\frac{n}{\Delta} c_p + e_p + \rho(e - e_p). \end{aligned} \tag{16}$$

One reason this direction is important is that we can show

Lemma 2. $\bar{d} \not\leq e$, so that $\|\bar{d}\| \geq 1$.

Proof. Suppose not, so that $e - \bar{d} > 0$. Note that

$$\frac{\Delta}{n}(e - \bar{d}) = c_p + \frac{\Delta}{n}(1 - \rho)(e - e_p) \tag{17}$$

which is of the form s(ν) above, and the corresponding duality gap is $e^Ts(\nu) = -\dfrac{\Delta}{n} e^T(e - \bar{d})$. Now

$$e^T\bar{d} = -e^T P\nabla\phi + e^T(e+e_p)\frac{e^T P\nabla\phi}{\|e-e_p\|^2} = 0\ . \tag{18}$$

Thus the duality gap is $\frac{\Delta}{n}e^T e = \Delta$. But since s(ν) is strictly positive, we can find a feasible dual slack with a smaller ν, and this would give a smaller duality gap than Δ, contradicting (13).

Note that Lemma 2 (with $e^T\bar{d} = 0$) also shows that $\hat{d}$ is nonzero.

Since $e^T\nabla\phi = e^T((n/\Delta)(c - zg) - e) = 0$, we can alternatively write using (15) and (16)

$$\hat{d} = -\left(P + \frac{(e-e_p)(e-e_p)^T}{\|e-e_p\|^2}\right)\nabla\phi =: -P_Q\nabla\phi, \tag{19}$$

where the last equation defines the matrix P_Q. It is easy to check that P_Q is a projection matrix; indeed, as shown in Lemma 2.2 of Gonzaga (1991a), P_Q is the matrix projecting onto $Q := \{x \in R^n: Ax = \sigma b$ for some $\sigma\}$.

Now we use the following standard lemma, due to Karmarkar; for a proof, see for example Ye (1991).

Lemma 3. If $e^T d = 0$ and $\alpha\|d\| < 1$, then

$$\phi_n(e + \alpha d, z) \le \phi_n(e, z) + \alpha\nabla\phi^T d + \frac{\alpha^2\|d\|^2}{2(1-\alpha\|d\|)}. \tag{20}$$

If we apply this lemma with $d = \bar{d}$ and $\alpha = \frac{1}{2\|d\|}$ we arrive at:

Proposition 1. If $\bar{\alpha} = \frac{1}{2\|d\|}$, then $e + \bar{\alpha}\bar{d} > 0$ and

$$\phi_n(e + \bar{\alpha}\bar{d}, z) \le \phi_n(e,z) - \frac{1}{4}. \tag{21}$$

Proof: Note that

$$\nabla\phi^T\bar{d} = -\nabla\phi^T P_Q\nabla\phi = -\|P_Q\nabla\phi\|^2 = -\|\bar{d}\|^2,$$

using (19) and the fact that P_Q is a projection matrix. Thus $\bar{\alpha}\nabla\phi^T\bar{d} = -\frac{\|\bar{d}\|}{2} \le -\frac{1}{2}$, using Lemma 2. Also, our choice of $\bar{\alpha}$ ensures that $\frac{\bar{\alpha}^2\|\bar{d}\|^2}{2(1-\bar{\alpha}\|d\|)} = \frac{1}{4}$, from which the result follows using (20).

How can we use Proposition 1, when the point $e + \bar{\alpha}\bar{d}$ is not feasible in (P)? We use the fact that the two directions $\bar{d}$ and $\hat{d}$ are *equivalent* in the sense of Gonzaga (1989), so that line searches in these directions in some sense give equivalent points. Indeed, using the homogeneity of ϕ_n we obtain:

$$\phi_n(e+\alpha\bar{d},z) = \phi_n(e+\alpha(\hat{d}+\rho e),z)$$
$$= \phi_n(1+\alpha\rho)(e+\frac{\alpha}{1+\alpha\rho}\hat{d}),z) \quad (22)$$
$$= \phi_n(e+\frac{\alpha}{1+\alpha\rho}\hat{d},z)$$

as long as $1 + \alpha\rho > 0$ (note that there is a slight mistake in Lemma 2.6 of Gonzaga (1991a) on the potential reduction achievable in two equivalent directions: the condition that $x + \alpha h > 0$ should be $x + \alpha h \in C$. This is related to the condition $1 + \alpha\rho > 0$ needed above.)

Theorem 3. Provided $\hat{d} \not> 0,$ we have

$$\sigma := \bar{\alpha}/(1+\bar{\alpha}\rho) > 0, \quad (23)$$

where $\bar{\alpha}$ is as in Proposition 1 and ρ is defined in (15), and

$$\phi_n(e+\sigma\hat{d},z) \le \phi_n(e,z) - \frac{1}{4}. \quad (24)$$

Moreover, the hypothesis holds if either the feasible region of (P) is compact or there is a nonnegative nonzero vector in the row space of A.

Proof: From the definition of $\bar{\alpha}$, we see that

$$e + \alpha\bar{d} = (1 + \alpha\rho)e + \alpha\hat{d}$$

is positive for all $0 < \alpha \le \hat{\alpha}$. Hence if $1 + \alpha\rho$ equalled zero for any such α, we could deduce $\hat{d} > 0$, contrary to hypothesis. It follows that $1 + \bar{\alpha}\rho > 0$, and hence σ is (well-defined and) positive as desired. Then (24) is implied by Proposition 1 and (22).

For the last part, note that $\hat{d}$ lies in the null space of A, and is therefore orthogonal to its row space. Thus we cannot have $\hat{d} > 0$ if either the feasible region of (P) is compact ($\hat{d}$ would be a direction of recession) or there is a nonnegative nonzero vector in this row space (since this cannot be orthogonal to a positive vector).€

We now provide an example to show that if the hypothesis fails, then a step in the direction $\hat{d}$ may not be able to provide a decrease of 1/4 in the potential function.

Example 1. Let A = [1,10,-10], b = 1, and $c = (0, 2, 1)^T$. Then x = e is feasible, and we can take $g = A^T$. We then find

$$c_p = \frac{1}{201}\begin{pmatrix}-10\\302\\301\end{pmatrix}, \quad e_p = \frac{1}{201}\begin{pmatrix}200\\191\\211\end{pmatrix}, \quad \text{and} \quad e-e_p = \frac{1}{201}\begin{pmatrix}1\\10\\-10\end{pmatrix}.$$

The optimal value of this problem is 0, corresponding to $x = (1,0,0)^T$, and the lower bound updating technique described at the beginning of this section will increase any negative lower bound to 0 (the smallest ν making $s(\nu) := c_p + \nu(e - e_p)$) nonnegative is 10, and then $e^Ts(\nu) = e^Tc_p + \nu\|e - e_p\|^2 = \frac{593}{201} + 10\frac{1}{201} = 3$, while the current objective value is $c^Te = 3$). Thus c - zg = c, and $\Delta = 3 - 0 = 3$, so $n/\Delta = 1$ and

$$\nabla\phi = c - e = \begin{pmatrix}-1\\1\\0\end{pmatrix}, P\nabla\phi = c_p - e_p = \frac{1}{201}\begin{pmatrix}-210\\111\\90\end{pmatrix}.$$

We then find $\rho = \dfrac{e_p^T P\nabla\phi}{\|e - e_p\|^2} = -9$, so $\hat{d} = -P\nabla\phi - \rho e_p = (10,8,9)^T$. The corresponding direction $\bar{d} = \hat{d} + \rho e = (1, -1, 0)^T$. We would like to move a stepsize $\bar{\alpha} = \dfrac{1}{2\|d\|} = \sqrt{\dfrac{2}{4}}$ in the direction $\bar{d}$. However, every point on the ray from e in the direction $\hat{d}$ is a positive multiple of a point of the form $e + \alpha\bar{d}$, with $\alpha < 1/9$. Indeed, an "infinite'' stepsize leads to the point at infinity in the direction of $\hat{d}$, which corresponds to (is a positive multiple of) the point $e + \frac{1}{9}\bar{d} = (10/9, 8/9, 1)^T$. The potential function $\phi_n(.,0)$ is decreasing all along this ray, converging to a value of $\phi_n(\hat{d}, 0) = 3.0774$ from its original value $\phi_n(e, 0) = 3.296$, a reduction of only about .22.

As we noted in the introduction, there are two ways to circumvent the problems illustrated by the example above. One is to require that the search direction be monotone with respect to the objective function, as suggested by Gonzaga (1989). In the example above, this leads to the direction $\frac{1}{593}(930,-31,62)^T$, and the potential function can be decreased by 2.81 by searching in this direction. However, the proof that the potential function can be reduced is more complicated than that above.

The other remedy is to use the last part of Theorem 3, and ensure that some nonnegative nonzero vector lies in the row space of A. The simplest way to do this is to introduce a dummy variable, so that (P) becomes

$$
\begin{array}{lll}
(P') & \text{minimize} & c^T x + 0\,\xi \\
 & \text{subject to} & Ax + 0\,\xi = b, \\
 & & 0^T x + 1\xi = 1, \\
 & & x \geq 0,\ \xi \geq 0,
\end{array}
$$

and the final equality constraint provides the desired nonnegative nonzero vector in the row space of the augmented A. This reformulation is equivalent to converting the given standard-form linear programming problem into homogeneous form using an extra variable, as discussed by Anstreicher (1986), de Ghellinck and Vial (1986), Gay (1987), Jensen and Steger (1985), and Ye and Kojima (1987); see Todd (1991).

If we modify our example thus, then c_p and e_p just have a zero component appended. The lower bounding technique can only guarantee a duality gap of at most 13 in this case, but for simplicity let us suppose that we have z = -1, so that $\Delta = 3 - (-1) = 4 = n$. Then $P\nabla\phi$ also just has a zero component appended. However, now it turns out that $\rho = -9/202$ and then $\hat{d}$ becomes $\frac{1}{202}(220, -103, -81)^T$. A line search in this direction can decrease the potential function by 3.16.

REFERENCES

ANSTREICHER, K.M. (1986): "A monotonic projective algorithm for fractional linear programming", **Algorithmica**, 1, 483-498.

_________________ (1996): "Potential reduction algorithm", in **Interior Point Methods in Mathematical Programming**, T. Terlaky, ed. Kluwer Academic Publishers, Dordrecht, The Netherlands.

DIKIN, I.I. (1967): "Iterative solution of problems of linear and quadratic programming", **Soviet Mathematics Doklady**, 8, 674-675.

GAY, D.M (1987): `"A variant of Karmarkar's linear programming algorithm for problems in standard form", **Mathematical Programming**, 37, 81-90.

GHELLINCK, G. de and J.P. VIAL (1986): "A polynomial Newton method for linear programming", **Algorithmica**, 1, 425-454.

GONZAGA, C.C. (1989): "Conical projection methods for linear programming", **Mathematical Programming**, 43, 151-173.

______________ (1990): "Polynomial affine algorithms for linear programming", **Mathematical Programming**, 49, 7-21.

______________ (1991a): "Search directions for interior linear-programming methods", **Algorithmica**, 6, 153-181.

______________ (1991b): "Interior point algorithms for linear programming with inequality constraints", **Mathematical Programming**, 52, 209-225.

______________ (1992): "Path following methods for linear programming", **SIAM Review**, 34, 167-227.

______________ (2002): "A simple presentation of Karmarkar's algorithm", Technical Report, Dept. of Mathematics, Federal University of Santa Catarina, Brazil.

KARMARKAR, N.K. (1985): "A new polynomial-time algorithm for linear programming", **Combinatorica**, 4, 373-395.

MEGIDDO, N. (1985): "A variation on Karmarkar's algorithm", Preliminary report, IBM Research Division, Almaden Research Center, San Jose, CA.

NESTEROV, YU. E. and A.S. NEMIROVSKI (1994): **Interior Point Polynomial Methods in Convex Programming: Theory and Algorithms,** SIAM Publications, SIAM, Philadelphia.

PADBERG, M.W. (1985): "Solution of a nonlinear programming problem arising in the projective method for linear programming", Technical Report, School of Business and Administration, New York University, New York, NY.

RENEGAR, J. (2001): **A Mathematical View of Interior-Point Methods in Convex Optimization**, SIAM Publications, SIAM, Philadelphia.

ROOS, C.; T. TERLAKY and J.P. VIAL (1997): **Theory and Algorithms for Linear Optimization: An Interior Point Approach**, John Wiley, Chichester.

STEGER, A.E. (1985): "An extension of Karmarkar's algorithm for bounded linear programming problems", Master's Thesis, State University of New York, Stonybrook, NY.

STURM, J.F. and S. ZHANG (1996): "An $O(\sqrt{n}L)$ iteration bound primal-dual cone affine scaling algorithm for linear programming", **Mathematical Programming**, 72, 177-194.

TODD, M.J. (1991): "The affine-scaling direction for linear programming is a limit of projective-scaling directions", **Linear Algebra and its Applications**, 152, 93-105.

__________ (1996): "Potential-reduction methods in mathematical programming", **Mathematical Programming**, 76, 3-45.

VANDERBEI, R.J. (1997): **Linear Programming: Foundations and Extensions**, Kluwer Academic Publishers, Boston.

WRIGHT, S.J. (1997): **Primal-Dual Interior-Point Methods**, SIAM, Philadelphia.

YE, Y. (1991): "An $O(n^3L)$ potential reduction algorithm for linear programming", **Mathematical Programming**, 50, 239-258.

_____ (1997): **Interior Point Algorithms: Theory and Analysis**, John Wiley, New York.

YE, Y. and M. KOJIMA (1987): "Recovering optimal dual solutions in Karmarkar's polynomial algorithm for linear programming", **Mathematical Programming**, 39, 305-317.