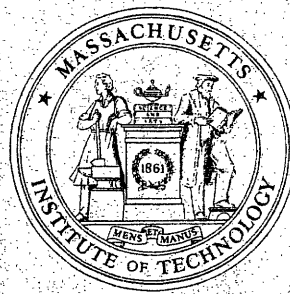# OPERATIONS RESEARCH CENTER
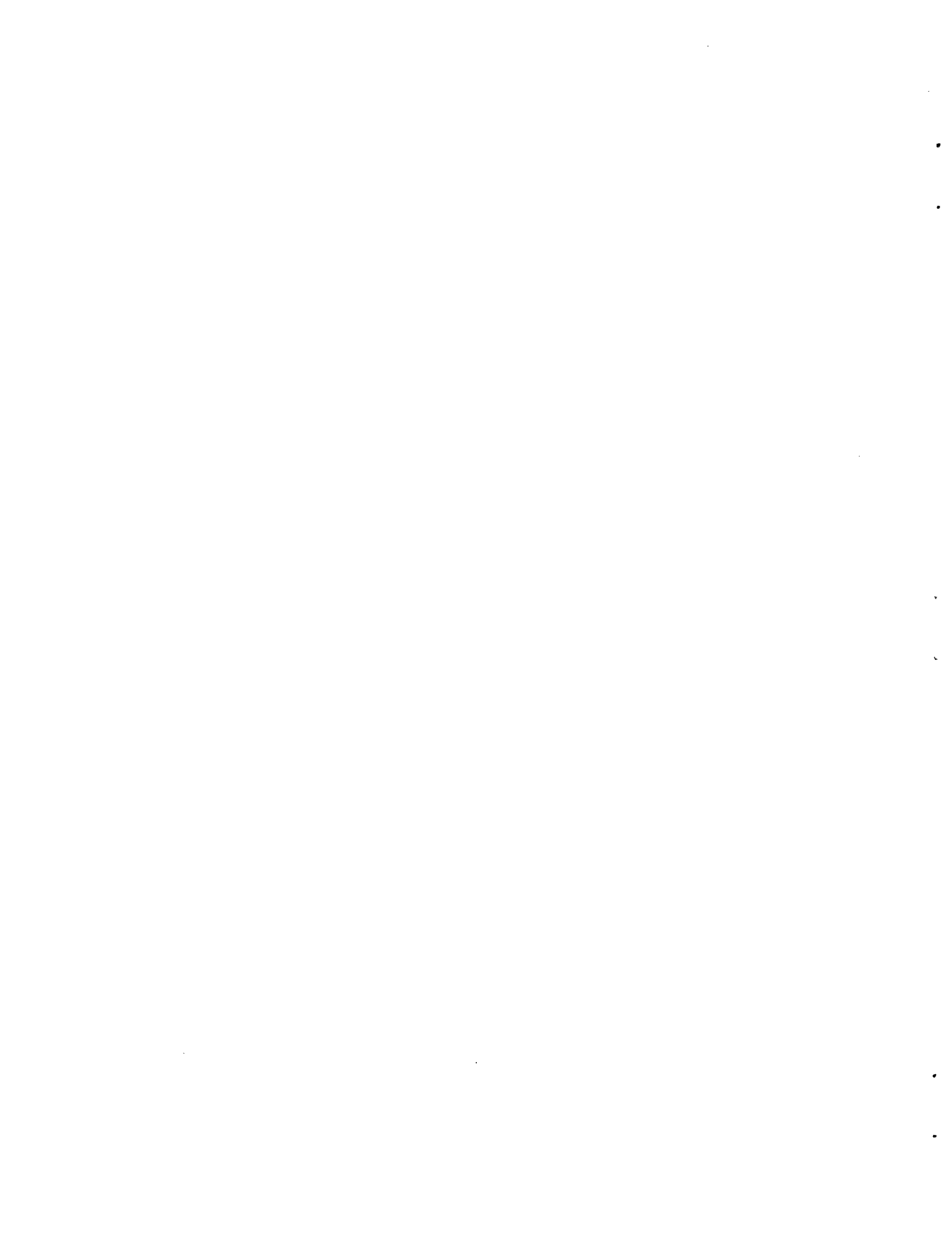
working paper

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

A POTENTIAL-FUNCTION REDUCTION ALGORITHM FOR SOLVING A LINEAR PROGRAM DIRECTLY FROM AN INFEASIBLE "WARM START"
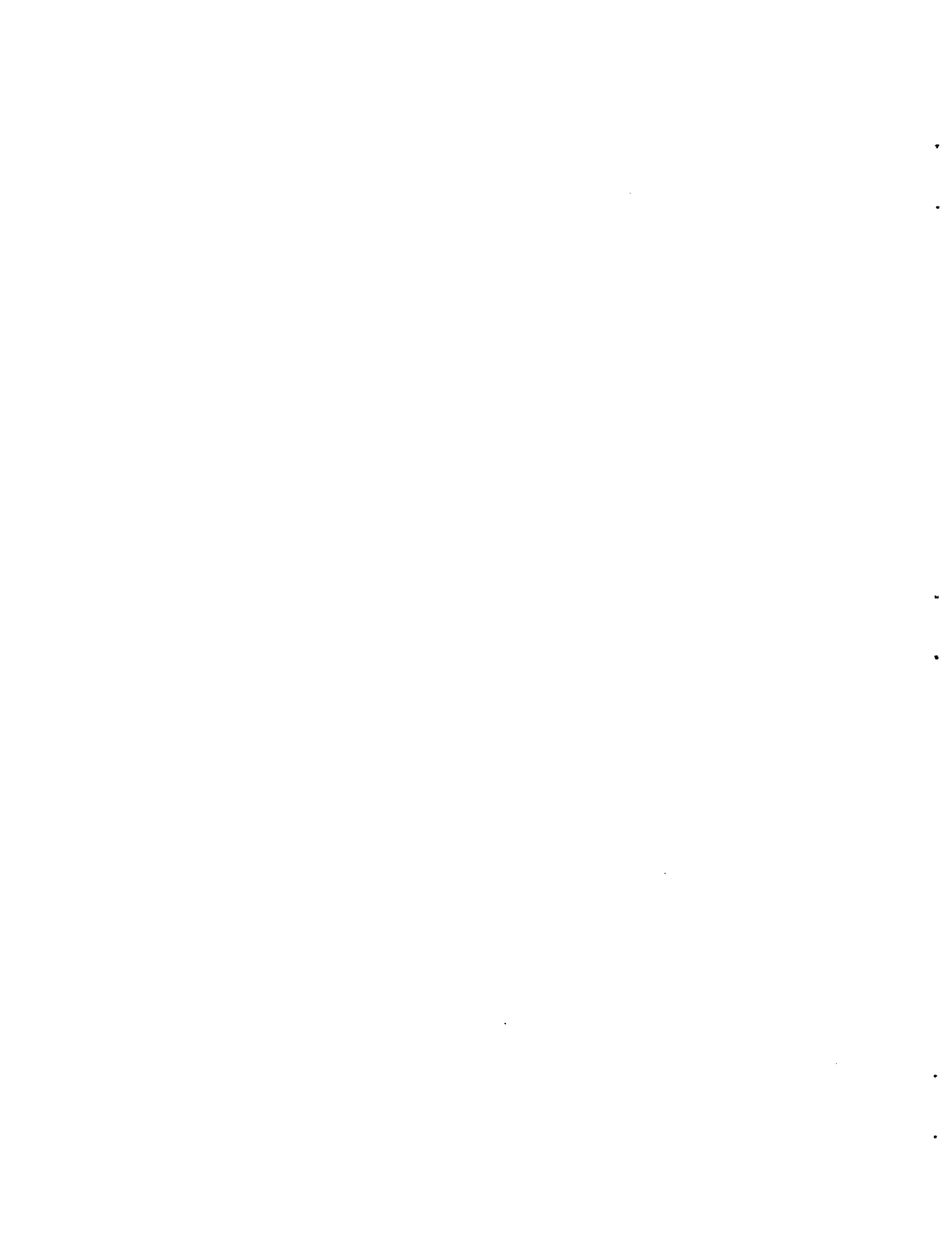
Robert M. Freund

OR-201-89 September, 1989

Abstract

This paper develops an algorithm for solving a standard-form linear program directly from an infeasible "warm start,"i.e., directly from a given infeasible solution $\hat{x}$ that satisfies $A\hat{x} = b$ but $\hat{x} \not\geq 0$. The algorithm is a potential function reduction algorithm, but the potential function is somewhat different than other interior-point method potential functions, and is given by

$$F(x, B) = q \ln (c^T x - B) - \sum_{j=1}^{n} \ln (x_j + h_j (c^T x - B))$$

where $q = n + \sqrt{n}$ is a given constant, h is a given strictly positive shift vector used to shift the nonnegativity constraints, and B is a lower bound on the optimal value of the linear program. The duality gap $c^T x - B$ is used both in the leading term as well as in the barrier term to help shift the nonnegativity constraints. The algorithm is shown under suitable conditions to achieve a constant decrease in the potential function and so achieves a constant decrease in the duality gap (and hence also in the infeasibility) in $O(n)$ iterations. Under more restrictive assumptions regarding the dual feasible region, this algorithm is modified by the addition of a dual barrier term, and will achieve a constant decrease in the duality gap (and in the infeasibility) in $O(\sqrt{n})$ iterations.

1. Introduction

This study is motivated by the problem of solving a linear program from an infeasible "warm start" solution, i.e., a solution that is not feasible for the linear program but is believed to be close to both feasibility and optimality for the linear program. The existence of such a "warm start" solution arises in many of the practical applications of linear programming. Quite often in the practice of using linear programming, it is necessary to make multiple runs of a given linear programming model, typically with relatively minor adjustments to the data of the given model. Over thirty years of experience with the simplex method has shown that the optimal basis (or equivalently the optimal solution) of one version of the model usually serves as an excellent starting basis for the next version of the model, whether or not the basis is even feasible for the next version of the model. When using the simplex method for solving linear programs, such a "warm start" infeasible solution can dramatically reduce the number of pivots and consequently the running time (both in Phase I and in Phase II) for solving multiple versions of a given base case linear programming model. In spite of the practical experience with using "warm start" solutions in the simplex method, there is no underlying complexity analysis that guarantees fast running times for such "warm start" solutions. This is due to the inevitable combinatorial aspects of the simplex algorithm itself.

In the case of interior-point algorithms for linear programming, much of the current complexity analysis of these algorithms is based on starting the algorithm from either an interior feasible solution (and only analyzing Phase II) or on starting the algorithm from a completely cold start, i.e., no known feasible solution. Anstreicher's combined Phase I-Phase II algorithm [2] is an exception to this trend, as is the shifted-barrier algorithm in [4]. (See Todd [10] for further analysis and extensions of Anstreicher's algorithm.) Both of these algorithms, as well as the algorithm presented in this paper, can be used to solve a linear program from an infeasible "warm start." Furthermore, all three algorithms have the following other desirable features: they simultaneously improve feasibility and optimality at each iteration, and so bypass the need for a Phase I-Phase II transition. Under suitable assumptions, these algorithms also have a worst-case computational complexity that is polynomial-time, and their theoretical performance is a function of how far the initial "warm start" is from being feasible and from being optimal (using a suitable measure of infeasibility and of optimality).

The algorithm developed in this paper is a potential function reduction algorithm, but the potential function is somewhat different than other interior-point method potential functions. The construction of the potential function is an extension of the shifted barrier function approach developed in [4]. Suppose we are interested in solving the linear program:

$$\text{LP:} \quad \underset{x}{\text{minimize}} \quad c^T x$$
$$\text{s.t.} \quad Ax = b, \quad x \geq 0 \,,$$

directly from a given infeasible "warm start" solution, i.e., a directly from a given solution $\hat{x}$ that is infeasible for LP in the sense that $A\hat{x} = b$ but $\hat{x} \not\geq 0$. Let $h \in R^n$ be a given strictly positive vector in R^n that is used to "shift" the nonnegativity constraints from $x \geq 0$ to $x + h\varepsilon \geq 0$ for some positive parameter ε. A shifted barrier function approach to solving LP is to solve the parameterized problem:

$$\text{Sh}(\varepsilon): \quad \underset{x}{\text{minimize}} \quad c^T x - \varepsilon \sum_{j=1}^{n} \ln (x_j + h_j \varepsilon)$$
$$\text{s.t.} \quad Ax = b,$$
$$x + h\varepsilon > 0 \,,$$

for a sequence of values of ε that converges to zero, see [4]. One can easily show that as ε goes to zero, optimal solutions to $\text{Sh}(\varepsilon)$ converge to a feasible and optimal solution to LP . (Problem $\text{Sh}(\varepsilon)$ above is a specific instance of a more general shifted barrier problem studied in Gill et. al. [5]). If B is a lower bound on the unknown optimal objective value of LP , denoted z^* , then the duality gap $c^T x - B$ can be used as a proxy for ε in problem $\text{Sh}(\varepsilon)$. This leads to the following potential function minimization problem:

2

$$\text{PF:} \quad \underset{x,\,B}{\text{minimize}} \quad F(x,B) = q\,ln\,(c^T x - B) - \sum_{j=1}^{n} ln\,(x_j + h_j\,(c^T x - B))$$

$$\text{s.t.} \quad Ax = b$$

$$x + h\,(c^T x - B) > 0\ ,$$

$$B \le \varepsilon^*\ ,$$

where $q > n$ is a given fixed scalar. Note that for a sufficiently small values of B, that $(\hat{x}, B)$ is feasible in PF .

An algorithm for solving PF is presented in Section 3, and this algorithm is denoted Algorithm 1. This algorithm is a direct extension of the potential function reduction algorithm of [3], which is a slightly altered version of Ye's algorithm [11] for linear programming. At each iteration, a primal step is taken if the norm of a certain vector is sufficiently large; otherwise an improved dual solution is produced. It is shown in Section 3 that under suitable assumptions the iterates of Algorithm 1 decrease the potential function $F(x, B)$ by at least 1/12 at each iteration, when $q = n + \sqrt{n}$. This leads to a complexity analysis of $O(n)$ iterations to achieve a constant decrease in the duality gap $c^T x - B$.

The assumptions that are needed to achieve the performance results for Algorithm 1 include very routine assumptions (i.e., A has full row rank, the sets of optimal primal and dual solutions are nonempty and bounded, and we know a lower bound $\hat{B}$ on z^*), plus one fairly restrictive assumption regarding the dual feasible region: it is assumed that the dual feasible region is bounded and that a bound on the size of the dual feasible is known in advance. The boundedness assumption is easy to coerce, but the known bound may not be very easy to satisfy in some circumstances, except by introducing large numbers (i.e., all dual solutions lie in a ball of radius 2^L , where L is the bit size representation of the linear program).

Section 4 of the paper examines a modification of the problem PF that includes a barrier term for dual variables:

HF: $\quad\underset{x,\pi,s,B}{\text{minimize}}$ $H(x,s,B) = q\,ln\,(c^T x - B) - \displaystyle\sum_{j=1}^{n} ln\,(x_j + h_j\,(c^T x - B)) - \sum_{j=1}^{n} ln\,s_j$

s.t. $\qquad Ax = b$

$\qquad x + h\,(c^T x - B) > 0$,

$\qquad A^T \pi + s = c$

$\qquad\qquad s > 0$

$\qquad\qquad B = b^T \pi$

Algorithm 1 is modified slightly to Algorithm 2 in this section. Under assumptions more restrictive than those of Algorithm 1, it is shown that the iterates of Algorithm 2 decrease the potential function $H(x, s, B)$ by at least 0.04 at each iteration, when $q = n + \sqrt{n}$. This leads to a complexity analysis of $O(\sqrt{n})$ iterations to achieve a constant decrease in the duality gap $c^T x - B$.

Section 2 of the paper presents notation, assumptions, and preliminary results. Section 3 contains the development and analysis of Algorithm 1, and Section 4 contains the analysis of Algorithm 2. Section 5 contains remarks concerning the role of dual feasible solutions in the algorithms and in the assumptions, and compares the strengths and weaknesses of Algorithm 1 and Algorithm 2. The Appendix contains inequalities concerning logarithms that are used in the analysis.

2. Notation, Assumptions, and Preliminaries

If s, y, t, or h is a vector in R^n, then S, Y, T, or H refers to the nxn diagonal matrix whose diagonal elements correspond to the components of s, y, t, or h, respectively. Let e be vector of ones, i.e., $e = (1, ..., 1)^T$. If $x \in R^n$, $||x||$ denotes the Euclidean norm of x, and $||x||_1$ denotes the L_1 norm of x, i.e.,

$$||x||_1 = \sum_{j=1}^{n} |x_j| \; .$$

Our concern is with solving a linear program of the form:

4

$$P: \quad \text{minimize} \quad c^T x$$
$$x$$
$$\text{s.t.} \quad Ax = b$$
$$x \geq 0 ,$$

whose dual is given by

$$D: \quad \text{maximize} \quad b^T \pi$$
$$(\pi, s)$$
$$\text{s.t.} \quad A^T \pi + s = c$$
$$s \geq 0 .$$

We make the following assumptions on P and D:

A1:   The rows of A have full rank.

A2:   The set of optimal solutions of P and D are nonempty and bounded.

Let z^* denote the optimal objective value of P. We also assume that we have the following initial information on P:

A3:   We have an initial vector $\hat{x}$ for which $A\hat{x} = b$ but $\hat{x} \not\geq 0$, and

A4:   We have an initial lower bound $\hat{B}$ on the unknown value z^*, i.e., we have a constant $\hat{B}$ for which $\hat{B} \leq z^*$.

Note that if $\hat{x}$ is the initial "warm start" for P, but $A\hat{x} \neq b$, then by performing a projection, we can modify $\hat{x}$ so that $A\hat{x} = b$. Furthermore, we can also assume with no loss of generality that the dual feasible region is bounded. (If the dual feasible region is not bounded, then by adding the constraint $b^T \pi \geq \hat{B}$ to the dual, the dual feasible region becomes bounded by assumption A2). We formally add this assumption as:

A5:   The dual feasible region is bounded.

Let $h \in R^n$ be a given positive vector, i.e., $h > 0$. Our interest lies in "shifting" the inequality constraints $x \geq 0$ to constraints of the form $x + \mu h \geq 0$,

for parameterized values of μ , so that the initial infeasible warm start solution $\hat{x}$ satisfies $\hat{x} + \mu h \geq 0$. Furthermore, our interest is in developing an algorithm for LP that will decrease the optimality gap $c^T x - z^*$ and will decrease the value of μ at each iteration. We refer to h as the given shift vector, and μ as the shift parameter. Our approach is as follows:

Suppose B is a lower bound on z^* . Consider the linear program:

$$\text{LP(B):} \qquad \underset{x}{\text{minimize}} \qquad c^T x$$

$$\text{s.t.} \qquad Ax = b$$

$$x + h(c^T x - B) \geq 0 .$$

Note that the constraints $x \geq 0$ in LP have been replaced in LP (B) by the constraints $x + \mu h \geq 0$, where the shift parameter μ is equal to $c^T x - B$, i.e., μ is the difference between the objective value of x and the bound B . It is straightforward to show the following:

<u>Proposition 2.1</u>: Let υ (B) denote that optimal objective value of LP (B) .

    (i)   If $B < z^*$, $B < \upsilon (B) < z^*$.

    (ii)   If $B = z^*$, $B = \upsilon (B) = z^*$. ∎

Based on the formulation of LP (B) , we consider the following potential function reduction problem:

$$\text{PF:} \qquad \underset{x, B}{\text{minimize}} \qquad F(x, B) = q \, ln \, (c^T x - B) - \sum_{j=1}^{n} ln \, (x_j + h_j \, (c^T x - B))$$

$$\text{s.t.} \qquad Ax = b$$

$$x + h \, (c^T x - B) > 0 ,$$

$$B \leq z^* ,$$

where $q > n$ is a given parameter. Note that the constraint $B \leq z^*$ is equivalent to the condition that $B \leq b^T \pi$ for some dual feasible solution (π, s) .

We make the following further assumptions on initial information about the dual feasible region:

<u>A6</u>:   A bound on the set of all dual feasible slack vectors s is known, and h has

been rescaled so that $h^T s \leq \dfrac{1}{k\sqrt{n}}$ for all feasible solutions (π, s) , where

$k = 9$.

Note that assumption A6 is satisfied if we know some information on the boundedness of the dual feasible region. For example, if we know that $\|s\| \leq R$ for

all dual feasible solutions (π, s) , then upon replacing $h \leftarrow h \dfrac{1}{k\sqrt{n}\|h\|R}$ where

$k = 9$, we have $h^T s \leq \|s\|\|h\| \leq \dfrac{1}{9\sqrt{n}}$. Of all of the assumptions, however, A6 appears to be the most restrictive.

Our final assumption is a technical consideration.

<u>A7</u>: $1 + h^T c \neq 0$.

This assumption can always be satisfied for a given h by slightly perturbing or rescaling h if necessary. It is a necessary assumption to ensure the invertability of an affine transformation defined in Section 3.

Assumptions A1 through A7 include the routine assumptions A1 – A4 (i.e., A has full row rank, the sets of optimal primal and dual solutions are nonempty and bounded, and we know a lower bound $\hat{B}$ on z^*), plus one fairly restrictive assumption regarding the dual feasible region: it is assumed that the dual feasible region is bounded (A5) and that a bound on the size of the dual feasible is known in advance (A6). The boundedness assumption is easy to coerce, but the known bound may not be very easy to satisfy in some circumstances, except by introducing large numbers (i.e., all dual solutions lie in a ball of radius 2^L , where L is the bit size representation of the linear program). Assumption A7 is a minor technical assumption.

Finally, we present the following technical remark.

<u>Remark 2.1</u>. Under assumptions A1 – A7, if $(\bar{\lambda}, \bar{\theta})$ satisfy $A^T \bar{\lambda} \leq \bar{\theta} c$, then $\bar{\theta} > 0$.

Proof: Because the dual is feasible, there exists $\tilde{\lambda}$ that solves $A^T\tilde{\lambda} \le c$. Suppose that $\bar{\theta} \le 0$. Then $A^T(-\bar{\theta}\,\bar{\lambda} + \tilde{\lambda}) \le -\bar{\theta}\,c + \bar{\theta}\,c = 0$, whereby $r = -\bar{\theta}\bar{\lambda} + \tilde{\lambda}$ is a ray of the dual feasible region. But because the dual feasible region is bounded, then $r = 0$. This in turn implies that $A^T\bar{\lambda} = c$, which implies that the objective value is constant on the primal feasible region. However, because the dual has a bounded region, the primal feasible region is unbounded, and so this now implies that the set of optimal solutions of P is unbounded, contradicting A2. ∎

3. Potential Function Reduction Algorithm 1

In this section we present an algorithm for LP that generates improving values of x and B in the potential function minimization problem PF:

$$
\text{PF:} \quad \underset{x,\,B}{\text{minimize}} \quad F(x, B) = q\,ln\,(c^T x - B) - \sum_{j=1}^{n} ln\,(x_j + h_j\,(c^T x - B))
$$

$$
\text{s.t.} \quad Ax = b
$$

$$
x + h\,(c^T x - B) > 0 ,
$$

$$
B \le z^* ,
$$

where $q > n$ is a given parameter. This algorithm is as follows:

Algorithm 1 $(A, b, c, h, \hat{x}, \hat{B}, \varepsilon^*, \theta, q, \bar{\lambda}, k)$

Step 0 (Initialization)

Define $\quad M = [I + hc^T]$, $M^{-1} = \left[I - \dfrac{h\,c^T}{1 + c^T h} \right]$. $\qquad$ (1)

$x^0 = \hat{x}$

Assign $\quad B^0 = \min \left\{ \hat{B} , \dfrac{\hat{x}_1 - 1}{h_1} + c^T\hat{x} , ..., \dfrac{\hat{x}_n - 1}{h_n} + c^T\hat{x} \right\}$

$\bar{x} = x^0$

$\bar{B} = B^0$

Step 1 (Test for Duality Gap Tolerance)
$\quad$ If $(c^T\bar{x} - \bar{B}) \le \varepsilon^*$, Stop.

8

Step 2 (Compute Direction)

$$\bar{y} = \bar{x} + h\,(c^T\bar{x} - \bar{B}) \tag{2a}$$

$$\bar{A} = AM^{-1}\bar{Y} \tag{2b}$$

$$\bar{c} = \bar{Y}c \tag{2c}$$

$$\bar{b} = b - \left(\frac{Ah\,\bar{B}}{1 + c^T h}\right) \tag{2d}$$

$$\bar{\Delta} = \left(\frac{\bar{c}^T e - \bar{B}}{1 + c^T h}\right) = c^T\bar{x} - \bar{B} \tag{3}$$

$$\bar{g} = \left(\frac{q}{\bar{\Delta}}\right)\left(\frac{\bar{c}}{1 + c^T h}\right) - e \tag{4}$$

$$\bar{d} = \left[I - \bar{A}^T\left(\bar{A}\,\bar{A}^T\right)^{-1}\bar{A}\right]\bar{g} \tag{5}$$

If $\|\bar{d}\| \geq \gamma$, go to Step 3. Otherwise go to Step 4.

Step 3 (Primal Step)

Set $\quad \bar{f} = M^{-1}d/\|\bar{d}\| \tag{6}$

Set $\quad \tilde{x} = \bar{x} - \alpha\bar{f}$

where $\quad \alpha = 1 - \dfrac{1}{\sqrt{1 + 2\gamma}}$, or α is determined by a line-search of the potential function $F(\bar{x} - \alpha\bar{f}, \bar{B})$.

Step 3a (Reset Primal Variables)

Reset $\quad \bar{x} = \tilde{x}$ and go to Step 1.

Step 4 (Dual Step)

Define $\quad \bar{t} = \left(\frac{\bar{\Delta}}{q}\right)\bar{Y}^{-1}(e + \bar{d}) \tag{7}$

$$\bar{\lambda} = \left(\frac{\bar{\Delta}}{q}\right)(\bar{A}\,\bar{A}^T)^{-1}\bar{A}\,\bar{g} \tag{8}$$

$$\tilde{s} = \frac{\bar{t}}{1 - h^T\bar{t}} \tag{9a}$$

$$\tilde{\pi} = \frac{\bar{\lambda}}{1 - h^T\bar{t}} \tag{9b}$$

9

$$\widetilde{B} = b^T \widetilde{\pi} \tag{10}$$

$$\beta = b^T \widetilde{\pi} - \overline{B} \tag{11}$$

Step 4a (Reset Dual Variables)

Reset $(\overline{\pi}, \overline{s}) = (\widetilde{\pi}, \widetilde{s})$

Reset $\overline{B} = \widetilde{B}$. Go to Step 1.

The data includes the original data for the LP , namely (A, b, c), the given shift vector $h > 0$, the initial "warm start" infeasible solution $\widetilde{x}$ and the initial lower bound $\widehat{B}$ on z^*. The scalar $\varepsilon*$ is a tolerance on the duality gap used to stop the algorithm (see Step 1). The constant q is the scalar used in the potential function $F(x, B)$ in problem PF. The constant γ is used in the algorithm and will be explained shortly. The constant k is the number $k = 9$ used in Assumption A6, i.e., $k = 9$. Each step of Algorithm 1 is summarized below.

Step 0:

In this step, the matrices M and M^{-1} are defined. Note that M^{-1} is well-defined due to Assumption A7. Next the initial values x^0 and B^0 are chosen. It is elementary to prove:

Proposition 3.1 (Initial Values). The values of x^0 and B^0 assigned in Step 0 are feasible for PF, and furthermore,

$$c^T x^0 - B^0 = \text{maximum } \{ c^T \widehat{x} - \widehat{B}, \frac{1 - \widehat{x}_1}{h_1}, ..., \frac{1 - \widehat{x}_n}{h_n} \} > 0 . \tag{12}$$

Expression (12) states that the initial gap $c^T x^0 - B^0$ is the maximum of the "warm start" gap $c^T \widehat{x} - \widehat{B}$ and the quantities $(1 - \widehat{x}_j)/h_j$, $j=1, ..., n$. Thus the initial gap is generally proportional to the extent of the initial gap and the infeasibility of $\widehat{x}$; the larger the negativity in $\widehat{x}_j$ or the larger the gap $c^T \widehat{x} - \widehat{B}$, the larger will be the initial gap $c^T x^0 - B^0$.

Step 1. This step tests whether or not the current gap value $c^T \overline{x} - \overline{B}$ is less than or equal to the initial tolerance $\varepsilon*$.

Step 2. The quantity $\overline{y}$ is the value of the slacks in the program PF at the current values of $(x, B) = (\overline{x}, \overline{B})$. Next the LP data is modified to $\overline{A}$, $\overline{c}$, and $\overline{b}$ in (2).

10

This modification will be explained below. The current value of the gap is set equal to $\overline{\Delta}$ in (3). The quantities $\overline{g}$ and $\overline{d}$ are defined next; $\overline{g}$ corresponds to a gradient and $\overline{d}$ corresponds to a projected gradient, as will be explained shortly. As in the algorithm of Ye [11] or [3], if $\overline{d}$ is "large", i.e., if $\|\overline{d}\| \geq \gamma$, the algorithm will take a primal step (Step 3). If, on the other hand, $\|\overline{d}\| \leq \gamma$, the algorithm updates the lower bound $\overline{B}$ by computing new dual variables (Step 4).

Step 3. In this step the algorithm computes the primal direction (6) and takes a step in the negative of this direction, where the length α is computed either analytically or by a line-search of the potential function.

Step 4. In this step the quantities $(\widetilde{\pi}, \widetilde{s})$ are defined in (9). Proposition 3.4 below demonstrates that the values $(\widetilde{\pi}, \widetilde{s})$ will be dual feasible if $\|\overline{d}\| \leq \gamma < 1$. The lower bound $\overline{B}$ on z^* is then updated to $\widetilde{B} = b^T \widetilde{\pi}$ in (10). It will be shown that if $q = n + \sqrt{n}$ then $\widetilde{B} - \overline{B} = \beta > 0$, where β is defined in (11).

Note that the major computational effort in this algorithm lies in the need to work with $(\overline{A}\,\overline{A}^T)^{-1}$, i.e., to solve a system of the form $(\overline{A}\,\overline{A}^T)v = r$ for v . However, because M^{-1} is a rank-1 matrix, $\overline{A}\,\overline{A}^T$ is a rank-3 modification of $A\,\overline{Y}^2 A^T$ (where $\overline{Y}^2$ is a diagonal matrix). Therefore methods that maintain sparsity in solving systems of the form $A\,\overline{Y}^2 A^T$ can be used to solve for $\overline{d}$ in Step 2 of the algorithm.

In the remainder of this section we will prove:

Lemma 3.1 (Primal Improvement)

If Algorithm 1 takes a primal step and $0 \leq \alpha < 1$, then $F(\overline{x} - \alpha \overline{f}, \overline{B}) -$
$F(\overline{x}, \overline{B}) \leq -\alpha \gamma + \dfrac{\alpha^2}{2(1-\alpha)}$. If $\gamma = 0.5$ and $\alpha = 1 - 1/\sqrt{1+2\gamma}$, then $F(\overline{x} - \alpha \overline{f}, \overline{B}) -$
$F(\overline{x}, \overline{B}) \leq -1/12$.

Lemma 3.2 (Dual Improvement)

If $q = n + \sqrt{n}$, $\gamma \in (0, 1)$, and $p = (1 + \gamma)/(k(1 - \gamma)) < 1$, then if Algorithm 1 takes a dual step, $F(\overline{x}, \widetilde{B}) - F(\overline{x}, \overline{B}) \leq -(1 - \gamma)\sqrt{n} + p + \dfrac{p^2}{2(1-p)}$. If $\gamma = 0.5$ and $k = 9$, then $p = 1/3$ and $F(\overline{x}, \widetilde{B}) - F(\overline{x}, \overline{B}) \leq -1/12$.

Lemmas 3.1 and 3.2 show that Algorithm 1 will reduce the potential function by at least 1/12 at each iteration. Lemmas 3.1 and 3.2 thus serve as the basis to analyze the complexity of Algorithm 1.

Let $S = \{x \in R^n \mid F(x, B) \le F(x^o, B^o)$ for some $B \in [B^o, z^*]\}$, let $\rho =$

$$\max_{x \in S} \sum_{j=1}^{n} ln\,(x_j + h_j\,(c^T x - B^o)), \text{ and let } \delta = \rho - \sum_{j=1}^{n} ln\,(x_j^o + h_j\,(c^Tx^o - B^o)) \text{ . It is}$$

straightforward to show that S is a bounded set, that ρ is finite, and so δ is finite.

<u>Theorem 3.1</u>. Suppose Algorithm 1 is initiated with $q = n + \sqrt{n}$, $k = 9$, and $\gamma = 0.5$.

Then after at most $K = \left\lceil 12\,(n + \sqrt{n})\,ln\left(\dfrac{c^Tx^o - B^o}{\varepsilon^*}\right) + 12\delta \right\rceil$ iterations, the algorithm will

stop with $c^T\bar{x} - \bar{B} \le \varepsilon^*$.

<u>Proof</u>: From Lemmas 3.1 and 3.2, we will have after K iterations,

$$F(\bar{x}, \bar{B}) \le F(x^o, B^o) - q\,ln\left(\frac{c^T x^o - B^o}{\varepsilon^*}\right) - \delta \text{ .}$$

Upon setting $y^o = x^o + h\,(c^T x^o - B^o)$, $\bar{y} = \bar{x} + h\,(c^T\bar{x} - \bar{B})$, we have

$$q\,ln\,(c^T\bar{x} - \bar{B}) - \sum_{j=1}^{n} ln\,\bar{y}_j \le q\,ln\,(c^T x^o - B^o) - \sum_{j=1}^{n} ln\,y_j^o - q\,ln\left(\frac{c^Tx^o - B^o}{\varepsilon^*}\right) - \delta,$$

i.e., $\quad q\,ln\,(c^T\bar{x} - \bar{B}) \le \sum_{j=1}^{n} ln\,\bar{y}_j - \sum_{j=1}^{n} ln\,y_j^o + q\,ln\,\varepsilon^* - \delta$.

However, $\sum_{j=1}^{n} ln\,\bar{y}_j \le \rho$, since $\bar{x} \in S$, and $\sum_{j=1}^{n} y_j^o = \rho - \delta$. Thus

$$q\,ln\,(c^T\bar{x} - \bar{B}) \le \rho - (\rho - \delta) + q\,ln\,\varepsilon^* - \delta = q\,ln\,\varepsilon^* \text{ ,}$$

whereby $c^T\bar{x} - \bar{B} \le \varepsilon^*$. ∎

We now proceed to prove Lemmas 3.1 and 3.2 by considering first the primal step and then the dual step.

Analysis of Primal Step

At the beginning of Step 2 of Algorithm 1, the values of $(x, B) = (\bar{x}, \bar{B})$ are feasible for PF, and the slacks are $\bar{y} = \bar{x} + h(c^T \bar{x} - \bar{B}) > 0$. Now consider the affine transformation:

$$y = T(x) = \bar{Y}^{-1}[x + h(c^T x - \bar{B})] = \bar{Y}^{-1} Mx - \bar{Y}^{-1} h \bar{B},$$

where M is defined in (1). The inverse of $T(x)$ is then:

$$x = T^{-1}(y) = \bar{Y}y - h\left(\frac{\bar{c}^T y - \bar{B}}{1 + c^T h}\right).$$

where $\bar{c} = \bar{Y}c$ is defined in (2).
Note that $T(\bar{x}) = e$. It is straightforward to verify that the affine transformation $y = T(x)$ transforms the problem PF to the following potential function reduction problem:

$$\text{PG:} \quad \underset{y}{\text{minimize}} \quad G(y, \bar{B}) = q \, ln\left(\frac{\bar{c}^T y - \bar{B}}{1 + c^T h}\right) - \sum_{j=1}^{n} ln\, y_j - \sum_{j=1}^{n} ln\, \bar{y}_j$$

$$\text{s.t.} \quad \bar{A}y = \bar{b}$$
$$y > 0 ,$$

where $\bar{y}, \bar{A}, \bar{b}, \bar{c}$ are defined in (2). Because $T(\bar{x}) = e$, $y = e$ is feasible in PG.

Proposition 3.2. If $(\bar{x}, \bar{B})$ are feasible for PF, then $T(x)$ and $T^{-1}(y)$ are well-defined, and for all $y = T(x)$, then
 (i) $Ax = b$ if and only if $\bar{A}y = \bar{b}$
 (ii) $x + h(c^T x - \bar{B}) > 0$ if and only if $y > 0$
 (iii) $F(x, \bar{B}) = G(y, \bar{B})$.

Proof: Follows from direct substitution. ∎

13

From (iii) above it follows that a decrease in $G(y, \overline{B})$ by a constant δ will correspond to an identical decrease in $F(x, \overline{B})$.

Much as in Gonzaga [6], Ye [11], and [3], we now show that the projected gradient of $G(y, \overline{B})$ in the y coordinates at the point $y = e$ is a good descent direction of $G(y, \overline{B})$. Note that $\overline{g}$ defined in (3) is the gradient of $G(y, \overline{B})$ in the y coordinate at $y = e$, and $\overline{d}$ is the projection of $\overline{g}$ onto the null space of the equality constraints of PG. Also note that $\overline{g}^T \overline{d} = \overline{d}^T \overline{d} = \|\overline{d}\|^2$. Finally, note that if a primal step is taken in Algorithm 1, then $\|\overline{d}\| \geq \gamma$.

$\underline{\text{Proposition 3.3.}}$ $G(e - \alpha \overline{d}/\|\overline{d}\|, \overline{B}) - G(e, \overline{B}) \leq -\alpha \gamma + \dfrac{\alpha^2}{2(1-\alpha)}$ for $\alpha \in [0, 1)$.

$\underline{\text{Proof:}}$ $G(e - \alpha \overline{d}/\|\overline{d}\|, \overline{B}) - G(e, \overline{B})$

$$= q \ln \left(\frac{\overline{c}^T (e - \alpha \overline{d}/\|\overline{d}\|) - \overline{B}}{\overline{c}^T e - \overline{B}} \right) - \sum_{j=1}^{n} \ln \left(1 - \left(\frac{\alpha \overline{d}_j}{\|\overline{d}\|} \right) \right)$$

$$\leq q \ln \left(1 - \left(\frac{\alpha \overline{c}^T \overline{d}/\|\overline{d}\|}{\overline{c}^T e - \overline{B}} \right) \right) + \alpha e^T \overline{d}/\|\overline{d}\| + \frac{\alpha^2}{2(1-\alpha)}$$

(from Proposition A.2 of the Appendix)

$$\leq - \left(\frac{q \alpha \overline{c}^T \overline{d}/\|\overline{d}\|}{\overline{c}^T e - \overline{B}} \right) + \alpha e^T \overline{d}/\|\overline{d}\| + \frac{\alpha^2}{2(1-\alpha)}$$

(from Proposition A.1 of the Appendix)

$$= \frac{-\alpha}{\|\overline{d}\|} \left(\left(\frac{q}{\overline{c}^T e - \overline{B}} \right) \overline{c} - e \right)^T \overline{d} + \frac{\alpha^2}{2(1-\alpha)}$$

$$= \frac{-\alpha}{\|\overline{d}\|} \overline{g}^T \overline{d} + \frac{\alpha^2}{2(1-\alpha)} = -\alpha \|\overline{d}\| + \frac{\alpha^2}{2(1-\alpha)} \leq -\alpha \gamma + \frac{\alpha^2}{2(1-\alpha)}. \quad \blacksquare$$

<u>Proof of Lemma 3.1</u>: Upon setting $\alpha = 1 - \dfrac{1}{\sqrt{1 + 2\gamma}}$ in Proposition 3.3,

$G\left(e - \alpha \overline{d} / \|\overline{d}\|, \overline{B}\right) - G\left(e, \overline{B}\right) \le -\left(1 + \gamma - \sqrt{1 + 2\gamma}\right) \le -0.085$ with $\gamma = 0.5$. Finally, from the definition of $T(x)$ and $T^{-1}(y)$ and Proposition 3.2, we obtain

$$F(\overline{x} - \alpha \overline{f}, \overline{B}) - F(\overline{x}, \overline{B}) \le -0.085 < -1/12 . \quad \blacksquare$$

<u>Analysis of Dual Step</u>

Algorithm 1 will take a dual step (Step 4) if $\|\overline{d}\| \le \gamma$. In Step 4, the quantities $\overline{t}$, $\overline{\lambda}$, $\tilde{s}$, and $\tilde{\pi}$ are defined. We first show:

<u>Proposition 3.4</u>. If $\|\overline{d}\| \le \gamma \le 1$ at Step 2, then $\left(\tilde{\pi}, \tilde{s}\right)$ is well-defined and $\left(\tilde{\pi}, \tilde{s}\right)$ is a dual feasible solution.

<u>Proof</u>: Because $\overline{\Delta} = c^T \overline{x} - \overline{B} > 0$, $q \ge 0$, $\overline{y} > 0$, and $\|\overline{d}\| \le \gamma \le 1$, then from (7) we have $\overline{t} \ge 0$. From (5) and (4) we have

$$\overline{d} = \left(\frac{q}{\overline{\Delta}}\right)\left(\frac{\overline{c}}{1 + c^T h}\right) - e - \overline{A}^T\left(\overline{A}\,\overline{A}^T\right)^{-1}\overline{A}\,\overline{g}$$

which after rearranging is

$$\frac{\overline{c}}{1 + c^T h} = \left(\frac{\overline{\Delta}}{q}\right)(e + \overline{d}) + \left(\frac{\overline{\Delta}}{q}\right)\overline{A}^T\left(\overline{A}\,\overline{A}^T\right)^{-1}\overline{A}\,\overline{g} = \left(\frac{\overline{\Delta}}{q}\right)(e + \overline{d}) + \overline{A}^T\overline{\lambda}$$

from (8). But from (2b) and (2c) this is

$$\left(\frac{1}{1 + c^T h}\right)\overline{Y}\,c = \left(\frac{\overline{\Delta}}{q}\right)(e + \overline{d}) + \overline{Y}\left(M^{-1}\right)^T A^T\overline{\lambda}$$

15

which from (7) is

$$\frac{c}{1 + c^T h} = \bar{t} + \left(M^{-1}\right)^T A^T \bar{\lambda} \ . \tag{13}$$

Premultiplying (13) by h^T yields

$$\frac{h^T c}{1 + c^T h} = h^T \bar{t} + h^T \left(M^{-1}\right)^T A^T \bar{\lambda} \ . \tag{14}$$

But $h^T \left(M^{-1}\right)^T = h^T \left(\dfrac{1}{1 + c^T h}\right)$ from (1), so (14) becomes

$$h^T \bar{t} = \frac{h^T c - h^T A^T \bar{\lambda}}{1 + c^T h}, \tag{15}$$

so that

$$1 - h^T \bar{t} = \frac{1 + h^T A^T \bar{\lambda}}{1 + c^T h} \ . \tag{16}$$

Expanding (13) using (1) gives

$$\frac{c}{1 + c^T h} = \bar{t} + A^T \bar{\lambda} - \frac{c\, h^T A^T \bar{\lambda}}{1 + c^T h} \ ,$$

i.e.,

$$A^T \bar{\lambda} + \bar{t} = c\left(\frac{1 + h^T A^T \bar{\lambda}}{1 + c^T h}\right) = c\left(1 - h^T \bar{t}\right) , \tag{17}$$

where the last equality is from (16).

16

But $\bar{t} \geq 0$, so from Remark 2.1, $1 - h^T \bar{t} > 0$. Therefore the definitions of $\tilde{s}$ and $\tilde{\pi}$ in (9) are well-defined, and from (17) we obtain $A^T \tilde{\pi} + \tilde{s} = c$. Finally, $\tilde{s} \geq 0$ because $\bar{t} \geq 0$. ∎

Our next task is to prove bounds on the quantity β defined in (11). Toward this end, we first prove two propositions.

<u>Proposition 3.5</u>. If $q = n + \sqrt{n}$ and $0 \leq \gamma < 1$ and Algorithm 1 is at Step 4, then

$$y^T \bar{t} \leq \left(\frac{\overline{\Delta}}{q}\right)(n + \sqrt{n}\,\gamma) .$$

<u>Proof</u>: From (7) we obtain

$$y^T \bar{t} = \left(\frac{\overline{\Delta}}{q}\right)(e^T e + e^T \overline{d}) \leq \left(\frac{\overline{\Delta}}{q}\right)(n + \sqrt{n}\,||\overline{d}||) \leq \left(\frac{\overline{\Delta}}{q}\right)(n + \sqrt{n}\,\gamma) . \quad ∎$$

<u>Proposition 3.6</u>. $y^T \bar{t} = c^T \overline{x} - \left(b^T \overline{\lambda} + h^T \bar{t}\,\overline{B}\right) .$

<u>Proof</u>: Premultiplying (13) by $\overline{y}^T$ gives

$$\frac{c^T \overline{y}}{1 + c^T h} = \overline{y}^T \bar{t} + e^T \overline{Y}(M^{-1})^T A^T \overline{\lambda} = \overline{y}^T \bar{t} + e^T \overline{A}^T \overline{\lambda} .$$

Thus $\quad \overline{y}^T \bar{t} = \dfrac{c^T \overline{y}}{1 + c^T h} - \overline{\lambda}^T \overline{A} e = \dfrac{c^T \overline{y}}{1 + c^T h} - \overline{\lambda}^T \overline{b} = \dfrac{c^T \overline{y}}{1 + c^T h} - \overline{\lambda}^T b + \dfrac{\overline{\lambda}^T A h \overline{B}}{1 + c^T h}$

(since $\overline{A} e = \overline{b}$ and using (2d))

$$= \frac{c^T \overline{y} - \overline{B}}{1 + c^T h} - b^T \overline{\lambda} + \frac{\overline{\lambda}^T A h \overline{B} + \overline{B}}{1 + c^T h}$$

$$= c^T \overline{x} - \overline{B} - b^T \overline{\lambda} + \overline{B}\left(\frac{\overline{\lambda}^T A h + 1}{1 + c^T h}\right) = c^T \overline{x} - \overline{B} - b^T \overline{\lambda} + \overline{B}\left(1 - h^T \bar{t}\right)$$

(from (16))

$$= \bar{c}^T \bar{x} - \left(b^T \bar{\lambda} + h^T \bar{t}\, \bar{B}\right) . \quad \blacksquare$$

<u>Proposition 3.7.</u> If $q = n + \sqrt{n}$ and $0 \le \gamma < 1$ and Algorithm 1 is set at Step 4, then

$$\frac{(1-\gamma)\sqrt{n}\,\bar{\Delta}}{q} \le \beta \le \frac{(1+\gamma)\sqrt{n}\,\bar{\Delta}}{q\left(1 - h^T \bar{t}\right)} , \text{ where } \beta \text{ is defined in (11)}.$$

<u>Proof:</u> From (9) and (11),

$$\beta = b^T \tilde{\pi} - \bar{B} \quad = \quad \frac{b^T \bar{\lambda}}{\left(1 - h^T \bar{t}\right)} - \bar{B} = \frac{b^T \bar{\lambda} + h^T \bar{t}\, \bar{B} - \bar{B}}{\left(1 - h^T \bar{t}\right)} = \frac{c^T \bar{x} - \bar{y}^T \bar{t} - \bar{B}}{1 - h^T \bar{t}}$$

(from Proposition 3.5)

$$= \left(\frac{1}{1 - h^T \bar{t}}\right)\left(\bar{\Delta} - \bar{y}^T \bar{t}\right) \tag{18}$$

(from (3))

$$\ge \left(\frac{1}{1 - h^T \bar{t}}\right)\left(\bar{\Delta} - \left(\frac{\bar{\Delta}}{q}\right)(n + \sqrt{n}\,\gamma)\right)$$

(from Proposition 3.5)

$$= \left(\frac{\bar{\Delta}}{1 - h^T \bar{t}}\right)\left(\frac{\sqrt{n}\,(1 - \gamma)}{q}\right)$$

$$\ge \frac{\bar{\Delta}\sqrt{n}\,(1 - \gamma)}{q} \text{ because } h^T \bar{t} > 0 .$$

18

This shows the first inequality. For the second inequality, note from (18) that

$$\beta = b^T \tilde{\pi} - \overline{B} \;=\; \left(\frac{1}{1-h^T\bar{t}}\right)\!\left(\overline{\Delta} - \overline{y}^T \bar{t}\right) \;=\; \left(\frac{1}{1-h^T\bar{t}}\right)\!\left(\overline{\Delta} - \left(\frac{\overline{\Delta}}{q}\right)\!\left(e^T e + e^T \overline{d}\right)\right)$$

$$\le \left(\frac{\overline{\Delta}}{1-h^T\bar{t}}\right)\!\left(1 - \left(\frac{n - \sqrt{n}\,\gamma}{q}\right)\right)$$

$$= \frac{\overline{\Delta}(1+\gamma)\sqrt{n}}{\left(1-h^T\bar{t}\right)q} \cdot \quad \blacksquare$$

Before proceeding with the proof of Lemma 3.2, we will need one more proposition.

<u>Proposition 3.8.</u> Suppose $0 \le \gamma < 1$ and that $\left(\dfrac{1+\gamma}{1-\gamma}\right) < k$, where k is the constant of Assumption A6. Suppose Algorithm 1 is at Step 4 and define β is as in (11) and define

$$p = \left(\frac{1+\gamma}{1-\gamma}\right)\left(\frac{1}{k}\right) . \tag{19}$$

Then   (i) $\beta\, h^T \overline{Y}^{-1} e \le p$

and   (ii) $\displaystyle\sum_{j=1}^{n} \frac{\left(\beta\, h_j / \overline{y}_j\right)^2}{2\left(1 - \beta\, h_j / \overline{y}_j\right)} \le \frac{p^2}{2(1-p)} .$

<u>Proof:</u>   (i) $\tilde{s} = \dfrac{\bar{t}}{1-h^T t} \;=\; \left(\dfrac{1}{1-h^T\bar{t}}\right)\!\left(\dfrac{\overline{\Delta}}{q}\right)\overline{Y}^{-1}\!\left(e + \overline{d}\right)$

(from (7) and (9))

$$\ge \left(\frac{1}{1-h^T t}\right)\left(\frac{\overline{\Delta}}{q}\right)\overline{Y}^{-1}(e-\gamma e) = \frac{(1-\gamma)\overline{\Delta}}{q(1-h^T t)}\overline{Y}^{-1}e \ .$$

Because $(\widetilde{\pi},\widetilde{s})$ is dual feasible, from Assumption A6 we have:

$$\frac{1}{k\sqrt{n}} \ge h^T\widetilde{s} \ge \frac{(1-\gamma)\overline{\Delta}}{q(1-h^T t)}h^T\overline{Y}^{-1}e \ . \tag{20}$$

From Proposition 3.7 we have $\beta \le \dfrac{(1+\gamma)\sqrt{n}\,\overline{\Delta}}{q(1-h^T t)}$, $\tag{21}$

where β is defined in (11). Combining (20) and (21) yields

$$\beta\, h^T\overline{Y}^{-1}e \le \frac{(1+\gamma)}{(1-\gamma)k} = p \ .$$

(ii)   For convenience, let $r = \beta\,\overline{Y}^{-1}h$. Then (i) states $e^T r \le p$, and

$$\sum_{j=1}^{n}\frac{r_j^2}{2(1-r_j)} \le \sum_{j=1}^{n}\frac{r_j^2}{2(1-p)} = \frac{\|r\|^2}{2(1-p)} \le \frac{\|r\|_1^2}{2(1-p)} = \frac{p^2}{2(1-p)} \ . \quad \blacksquare$$

Proof of Lemma 3.2:

$$F(\widetilde{x},\widetilde{B}) - F(\overline{x},\overline{B}) = q\,ln(c^T\widetilde{x}-\widetilde{B}) - \sum_{j=1}^{n} ln\,(\overline{y}_j-\beta\,h_j) - q\,ln(c^T\overline{x}-\overline{B}) + \sum_{j=1}^{n} ln\,(\overline{y}_j)$$

$$= q\,ln\left(1-\left(\frac{\beta}{\overline{\Delta}}\right)\right) - \sum_{j=1}^{n} ln(1-r_j)$$

where $r = \beta\,\overline{Y}^{-1}h$ and $\beta = \widetilde{B}-\overline{B} = b^T\widetilde{\pi}-\overline{B}$.

Thus from Proposition 3.8 and Propositions A.1 and A.2 of the Appendix

$$F\left(\overline{x}, \widetilde{B}\right) - F\left(\overline{x}, \overline{B}\right) \leq -\left(\frac{q}{\Delta}\right)\beta + e^T r + \sum_{j=1}^{n} \frac{r_j^2}{2\left(1 - r_j\right)}$$

$$\leq -\left(\frac{q}{\Delta}\right)\beta + p + \frac{p^2}{2\left(1 - p\right)}$$

$$\leq -\left(\frac{q}{\Delta}\right)\left(\frac{\left(1 - \gamma\right)\sqrt{n}\,\overline{\Delta}}{q}\right) + p + \frac{p^2}{2\left(1 - p\right)}$$

(from Proposition 3.7)

$$= -\left(1 - \gamma\right)\sqrt{n} + p + \frac{p^2}{2\left(1 - p\right)},$$

where $p = \left(\frac{1 + \gamma}{1 - \gamma}\right)\left(\frac{1}{k}\right)$. ∎

4.   Potential Function Reduction Algorithm 2

In this section we consider a modification of the potential problem PF
defined in Section 3 to the altered potential function problem HF presented below.
We then present Algorithm 2 which solves LP by seeking improving values of
primal and dual variables in HF . Algorithm 2 is in fact a slight modification of
Algorithm 1.

Consider the potential function minimization problem:

HF: $\underset{x, s, B, \pi}{\text{minimize}}$ $H(x, s, B) = q \, ln \, (c^T x - B) - \sum_{j=1}^{n} ln \, (x_j + h_j \, (c^T x - B)) - \sum_{j=1}^{n} ln \, s_j$

s.t. $\qquad Ax = b$

$\qquad\qquad x + h(c^T x - B) > 0 \, ,$

$\qquad\qquad A^T \pi + s = c$

$\qquad\qquad s > 0$

$\qquad\qquad B = b^T \pi \, ,$

where $q > n$ is a given parameter. Note that in this program that the potential function $H(x, s, B)$ is similar to $F(x, B)$ but also contains a barrier term for the dual slack variables s . Potential functions of this sort were first studied extensively by Todd and Ye [9] and by Ye [11], [3], and others.

Also note that because $Ax = b$ and $A^T \pi + s = b$, then

$$c^T x - B = c^T x - b^T \pi = x^T s \, , \qquad (22)$$

so that we could in fact rewrite $H(x, s, B)$ as

$$H(x, s) = H(x, s, B) = H\left(x, s, c^T x - x^T s\right) = q \, ln\left(x^T s\right) - \sum_{j=1}^{n} ln\left(x_j + h_j x^T s\right) - \sum_{j=1}^{n} ln \, s_j \, . \qquad (23)$$

The following algorithm (denoted Algorithm 2) is designed to generate improving values of x and (π, s) in the potential function minimization problem HF .

Algorithm 2 $(A, b, c, h, \hat{x}, \hat{\pi}, \hat{s}, \varepsilon^*, q, \lambda, k)$

Step 0 (Initialization)

Define $\quad M \qquad = [I + hc^T] \, , \, M^{-1} = \left[I - \dfrac{hc^T}{1 + c^T h}\right] . \qquad (1)$

$\qquad\qquad x^0 \qquad = \hat{x}$

$\qquad\qquad \left(\pi^0, s^0\right) \quad = \left(\hat{\pi}, \hat{s}\right)$

$\qquad\qquad B^0 \qquad = b^T \hat{\pi}$

$\qquad\qquad \bar{x} \qquad = x^0$

$\qquad\qquad \left(\bar{\pi}, \bar{s}\right) \quad = \left(\hat{\pi}, \hat{s}\right)$

$\qquad\qquad \bar{B} \qquad = b^T \hat{\pi}$

Step 1 (Test for Duality Gap Tolerance)

If $c^T \bar{x} - \bar{B} \leq \varepsilon_*$, Stop.

Step 2 (Compute Direction)

$$\bar{y} = \bar{x} + h(c^T\bar{x} - \bar{B}) \tag{2a}$$

$$\bar{A} = AM^{-1}\bar{Y} \tag{2b}$$

$$\bar{c} = \bar{Y}c \tag{2c}$$

$$\bar{b} = b - \frac{Ah\bar{B}}{1 + c^T h} \tag{2d}$$

$$\bar{\Delta} = \frac{\bar{c}^T e - \bar{B}}{1 + c^T h} = c^T\bar{x} - \bar{B} \tag{3}$$

$$\bar{g} = \left(\frac{q}{\bar{\Delta}}\right)\left(\frac{\bar{c}}{1 + c^T h}\right) - e \tag{4}$$

$$\bar{d} = \left[I - \bar{A}^T\left(\bar{A}\,\bar{A}^T\right)^{-1}\bar{A}\right]\bar{g} \tag{5}$$

If $\|\bar{d}\| \geq \gamma$, go to Step 3. Otherwise go to **Step 4**.

Step 3 (Primal Step)

Set $\quad \bar{f} = M^{-1}d/\|\bar{d}\| \tag{6}$

Set $\quad \tilde{x} = \bar{x} - \alpha\bar{f}$

where $\quad \alpha = 1 - \dfrac{1}{\sqrt{1 + 2\gamma}}$, or α is determined by a line-search of the potential function $H(\bar{x} - \alpha\bar{f}, \bar{s})$.

Step 3a (Reset Primal Variables)

Reset $\bar{x} = \tilde{x}$ and go to Step 1.

Step 4 (Dual Step)

Define $\quad \bar{t} = \left(\frac{\bar{\Delta}}{q}\right)\bar{Y}^{-1}(e + \bar{d}) \tag{7}$

$$\bar{\lambda} = \left(\frac{\bar{\Delta}}{q}\right)(\bar{A}\,\bar{A}^T)^{-1}\bar{A}\,\bar{g} \tag{8}$$

23

$$\tilde{s} \;=\; \frac{\bar{t}}{1 - h^T \bar{t}} \tag{9a}$$

$$\tilde{\pi} \;=\; \frac{\bar{\lambda}}{1 - h^T \bar{t}} \tag{9b}$$

$$\tilde{B} \;=\; b^T \tilde{\pi} \tag{10}$$

$$\beta \;=\; b^T \tilde{\pi} - \bar{B} \tag{11}$$

Step 4 (Reset Dual Variables)

Reset $(\bar{\pi}, \bar{s}) = (\tilde{\pi}, \tilde{s})$

Reset $\bar{B} = \tilde{B}$. Go to Step 1.

In Algorithm 2, the initial data for the problem is identical to the data for Algorithm 1, except that instead of having a lower bound $\tilde{B}$ on the optimal value z^* , we instead have an explicit dual feasible solution $(\tilde{\pi}, \tilde{s})$. Furthermore, we will need the following altered versions of Assumptions A4 and A6:

A4': We have an initial feasible solution $(\hat{\pi}, \hat{s})$ for which $\hat{\pi}$ and $\hat{s}$ are feasible in HF , i.e., $\hat{s} > 0$ and $\hat{x} + h\left(\hat{x}^T\hat{s}\right) > 0$.

A6': A bound on the set of all dual feasible slack vectors s is known, and h has been rescaled so that $h^T s \leq \dfrac{1}{k\sqrt{n}}$ for all dual feasible solutions (π, s) , where $k = 12\sqrt{n}$.

Assumption A4' assumes a known interior dual feasible solution $(\hat{\pi}, \hat{s})$. It also assume that $\hat{x}$ and $(\hat{\pi}, \hat{s})$ are feasible for HF . This assumption can be very restrictive. For instance, suppose that $\hat{x} \leq 0$. Then even if $(\hat{\pi}, \hat{s})$ is interior feasible, $\hat{x}^T\hat{s} \leq 0$, so that $\hat{x} + h\left(\hat{x}^T\hat{s}\right) \leq 0$, violating A4'. This point is discussed further in Section 5, which contains remarks. Assumption A6' is identical to A6, except that the constant k has been modified from k = 9 to $k = 12\sqrt{n}$. Note that other than the initialization step (Step 0), Algorithm 2 has an identical structure to Algorithm 1. Regarding the performance of Algorithm 2, we have

Lemma 4.1. (Primal Improvement). If Algorithm 2 takes a primal step and $0 \leq \alpha < 1$, then

$$H\left(\overline{x} - \alpha\overline{f}, \overline{s}\right) - H\left(\overline{x}, \overline{s}\right) \leq -\alpha\gamma + \frac{\alpha^2}{2(1-\alpha)} \ . \ \text{If } \gamma = 0.33 \text{ and } \alpha = 1 - 1/\sqrt{1+2\gamma} \ ,$$

then $H\left(\overline{x} - \alpha\overline{f}, \overline{s}\right) - H\left(\overline{x}, \overline{s}\right) \leq -0.04$.

<u>Proof</u>: If the algorithm takes a primal step, the additional potential function term

$\sum\limits_{j=1}^{n} \mathit{ln}\, s_j$ is unaffected. Therefore, the analysis is the same as in Algorithm 1, and

Lemma 3.1 applies. ∎

<u>Lemma 4.2. (Dual Improvement)</u>. If $q = n + \sqrt{n}$, $\gamma \in (0,1)$ and $p = \left(\dfrac{1+\gamma}{(1-\gamma)k}\right) < 1$,

then if Algorithm 2 takes a dual step,

$$H\left(\overline{x}, \tilde{s}\right) - H\left(\overline{x}, \overline{s}\right) \leq -(1-\gamma)/2 + \frac{\sqrt{n}}{k} + p + \frac{p^2}{2(1-p)} + \frac{\gamma^2}{2(1-\gamma)} \ . \ \text{If } \gamma = 0.33 \text{ and } k =$$

$12\sqrt{n}$, then $H\left(\overline{x}, \tilde{s}\right) - H\left(\overline{x}, \overline{s}\right) \leq -0.04$. ∎

Note in Lemma 4.2 that with $k = 12\sqrt{n}$ and $\gamma = 0.33$, that $p \leq 0.117$, because $n \geq 2$ (otherwise the dual feasible region would be unbounded, violating A5). Therefore $H\left(\overline{x}, \tilde{s}\right) - H\left(\overline{x}, \overline{s}\right) \leq -0.04$. Before we prove Lemma 4.2, we present a result on the complexity of Algorithm 2.

<u>Theorem 4.1</u>. Suppose Algorithm 2 is initiated with $q = n + \sqrt{n}$, $\gamma = 0.33$. Then

after at most $K = \left\lceil 25\sqrt{n} \ \ln\left(\dfrac{1}{\varepsilon^*}\right) + 25\, H\left(x^0, s^0\right)\right\rceil$ iterations, the algorithm will stop with

$c^T\overline{x} - \overline{B} = \overline{x}^T\overline{s} \leq \varepsilon^*$. ∎

This theorem will be proved at the end of this section.

Proof of Lemma 4.2. Suppose that Algorithm 2 is at Step 4. First note that

$$\bar{y}^T \bar{s} = \bar{s}^T\left(\bar{x} + \left(\bar{x}^T \bar{s}\right) h\right) = \bar{s}^T \bar{x}\left(1 + \bar{s}^T h\right) \leq \bar{s}^T \bar{x}\left(1 + \frac{1}{k\sqrt{n}}\right).$$

Therefore $n \, ln\left(\bar{y}^T \bar{s}\right) \leq n \, ln\left(\bar{x}^T \bar{s}\right) + n \, ln\left(1 + \frac{1}{k\sqrt{n}}\right)$.

However, from Proposition A.3 of the Appendix, $n \, ln\left(1 + \frac{1}{k\sqrt{n}}\right) \leq \frac{\sqrt{n}}{k}$. Thus

$$n \, ln\left(\bar{y}^T \bar{s}\right) \leq n \, ln\left(\bar{x}^T \bar{s}\right) + \frac{\sqrt{n}}{k}. \tag{23}$$

Next note that because $\bar{y} = \bar{x} + \left(\bar{x}^T \bar{s}\right) h \geq \bar{x}$ (because we must have $\bar{x}^T \bar{s} \geq 0$), then

$$n \, ln\left(\bar{y}^T \bar{s}\right) \geq n \, ln\left(\bar{x}^T \bar{s}\right). \tag{24}$$

From (7) and (9a)

$$\tilde{s} = \left(\frac{\bar{\Delta}}{q}\right)\left(\frac{1}{1 - h^T \bar{t}}\right) \bar{Y}^{-1}\left(e + \bar{d}\right)$$

and $\left\|\bar{d}\right\| \leq \gamma$. Therefore, from Proposition A.6 of the Appendix,

$$\sum_{j=1}^{n} ln \, \tilde{s}_j + \sum_{j=1}^{n} ln\left(\bar{y}_j\right) \geq n \, ln\left(\bar{y}^T \tilde{s}\right) - n \, ln \, n - \frac{\gamma^2}{2(1 - \gamma)} \tag{25}$$

$$\geq n \, ln\left(\bar{x}^T \bar{s}\right) - n \, ln \, n - \frac{\gamma^2}{2(1 - \gamma)} \tag{26}$$

(from (24)).

Also, from Proposition A.4 of the Appendix,

$$\sum_{j=1}^{n} \ln \bar{s}_j + \sum_{j=1}^{n} \ln \bar{y}_j \;\leq\; n \, \ln\left(\bar{y}^T \bar{s}\right) - n \, \ln n \;. \tag{27}$$

Let $\tilde{y} = \bar{x} + \left(\bar{x}^T \tilde{s}\right) h = \bar{x} + \left(c^T \bar{x} - \tilde{B}\right) h = \bar{x} + \left(c^T \bar{x} - \bar{B} - \beta\right) h$, then from Proposition 3.8 and the proof of Lemma 3.2 we obtain

$$\sum_{j=1}^{n} \ln \tilde{y}_j \;\geq\; \sum_{j=1}^{n} \ln \bar{y}_j - p - \frac{p^2}{2(1-p)} \;. \tag{28}$$

Finally, $H(\bar{x}, \tilde{s}) - H(\bar{x}, \bar{s}) = q \, \ln\left(\bar{x}^T \tilde{s}\right) - \sum_{j=1}^{n} \ln \tilde{y}_j - \sum_{j=1}^{n} \ln \tilde{s}_j - q \, \ln\left(\bar{x}^T \bar{s}\right) + \sum_{j=1}^{n} \ln \bar{y}_j + \sum_{j=1}^{n} \ln \bar{s}_j$

$$\leq \; q \, \ln\!\left(\frac{\bar{x}^T \tilde{s}}{\bar{x}^T \bar{s}}\right) + p + \frac{p^2}{2(1-p)} - \sum_{j=1}^{n} \ln \tilde{s}_j + \sum_{j=1}^{n} \ln \bar{s}_j$$

(from (28))

$$\leq \; q \, \ln\!\left(\frac{\bar{x}^T \tilde{s}}{\bar{x}^T \bar{s}}\right) + p + \frac{p^2}{2(1-p)} + \frac{\gamma^2}{2(1-\gamma)} - n \, \ln\left(\bar{x}^T \tilde{s}\right) + n \, \ln\left(\bar{y}^T \bar{s}\right)$$

(from (26) and (27))

$$\leq \; q \, \ln\!\left(\frac{\bar{x}^T \tilde{s}}{\bar{x}^T \bar{s}}\right) + p + \frac{p^2}{2(1-p)} + \frac{\gamma^2}{2(1-\gamma)} - n \, \ln\left(\bar{x}^T \tilde{s}\right) + n \, \ln\left(\bar{x}^T \bar{s}\right) + \frac{\sqrt{n}}{k}$$

(from (23))

$$= \; (q-n) \ln\!\left(\frac{\bar{x}^T \tilde{s}}{\bar{x}^T \bar{s}}\right) + p + \frac{p^2}{2(1-p)} + \frac{\gamma^2}{2(1-\gamma)} + \frac{\sqrt{n}}{k} \;. \tag{24}$$

However, $\left(\dfrac{\bar{x}^T\tilde{s}}{\bar{x}^T\bar{s}}\right) = \dfrac{\bar{\Delta}-\beta}{\bar{\Delta}} = 1 - \left(\dfrac{\beta}{\bar{\Delta}}\right) \leq 1 - \dfrac{(1-\gamma)\sqrt{n}}{q}$ from Proposition 3.7.

Thus, $(q-n)\,ln\left(\dfrac{\bar{x}^T\tilde{s}}{\bar{x}^T\bar{s}}\right) \leq (q-n)\,ln\left(1 - \dfrac{(1-\gamma)\sqrt{n}}{q}\right) \leq \dfrac{-(1-\gamma)\,n}{q} \leq \dfrac{-(1-\gamma)}{2}$. (25)

Inequalities (24) and (25) combine to yield the result. ∎

Proof of Theorem 4.1:

Let $\bar{x}$ and $(\bar{\pi},\bar{s})$ be the current primal and dual variables after K iterations of Algorithm 2.

Then

$$\sqrt{n}\,ln\left(\bar{x}^T\bar{s}\right) - \dfrac{\sqrt{n}}{k} \leq \sqrt{n}\,ln\left(\bar{x}^T\bar{s}\right) + n\,ln\left(\bar{x}^T\bar{s}\right) - n\,ln\left(\bar{y}^T\bar{s}\right)$$

(from (23))

$$= q\,ln\left(\bar{x}^T\bar{s}\right) - n\,ln\left(\bar{y}^T\bar{s}\right)$$

$$\leq q\,ln\left(\bar{x}^T\bar{s}\right) - \sum_{j=1}^{n} ln\,\bar{y}_j - \sum_{j=1}^{n} ln\,\bar{s}_j - n\,ln\,n$$

(from Proposition A.4)

$$= H(\bar{x},\bar{s}) - n\,ln\,n .$$

Thus $\sqrt{n}\,ln\left(\bar{x}^T\bar{s}\right) \leq H(\bar{x},\bar{s}) + \dfrac{\sqrt{n}}{k} - n\,ln\,n$

and so $\sqrt{n}\,ln\left(\bar{x}^T\bar{s}\right) \leq H(\bar{x},\bar{s})$, (26)

because from A6', $k = 12\sqrt{n}$, and so $\dfrac{\sqrt{n}}{k} \leq nln\ n$.

With K as given in the statement of Theorem 4.1,

$$\sqrt{n}ln\left(\overline{x}^T\overline{s}\right) \leq H(\overline{x},\overline{s}) \leq H(x^o,s^o) - 0.04K$$

$$\leq H(x^o,s^o) - \sqrt{n}ln\left(\frac{1}{\epsilon^*}\right) - H(x^o,s^o)$$

$$\leq \sqrt{n}ln(\epsilon^*) ,$$

and so $ln\left(\overline{x}^T\overline{s}\right) \leq \epsilon^*$. ∎

5.   Remarks

Relative Importance of Dual Feasible Solutions. The primary motivation behind the development of both Algorithm 1 and Algorithm 2 was to be able to solve a linear program from an initial infeasible solution, without having to perform a Phase I procedure or to induce feasibility by modifying the problem artificially. Both of these algorithms will generate a sequence of primal iterates that are becoming increasingly less infeasible and increasingly more optimal, and that converge to a feasible and optimal solution. Notice though that whenever either algorithm takes its first dual step, that the algorithm generates a dual feasible solution. Furthermore, Algorithm 2 actually presumes that an interior dual feasible solution is known in advance. In either algorithm, an interior dual feasible solution is produced or is known in advance, even though all primal iterates may be infeasible. This suggests perhaps that whenever a dual feasible solution is known or produced , that the LP instead be processed by a Phase II type polynomial-time algorithm working through the dual rather than the primal (for example, Todd and Burrell [8], Anstreicher [1], Ye [11], or [3]). Such an approach would render the results of this paper of little usefulness. However, there are at least two reasons why this strategy may not be wise. The dual feasible solution that is known or produced may have a very poor objective value, and so it may be a very poor candidate for a Phase II algorithm. Secondly, the initial infeasible primal solution may be very close to feasibility and to

optimality and so may be an excellent candidate for Algorithm 1 or Algorithm 2. In fact, this second condition may typically hold when making multiple runs of slightly altered versions of the same base case LP model, and it is this circumstance that has been the guiding motivation of this paper.

Comparison of Algorithm 1 and Algorithm 2. When implemented without a line-search, Algorithm 1 and Algorithm 2 differ in only three components: the choice of the control constant γ, the choice of the constant k that is used to rescale the shift vector h, and the initial presumed starting conditions of the two algorithms. However, as is evident from a comparison of Theorems 3.1 and 4.1, the complexity analysis of the two algorithms performed herein leads to different conclusions as to the efficiency of each algorithm. Herein we compare the two algorithms and discuss pros and cons of each algorithm.

A comparison of Theorems 3.1 and 4.1 suggests that Algorithm 2 is more attractive from the standpoint of efficiency, for at least two reasons. As a function of the duality gap tolerance ε^*, the iteration count constant K in Algorithm 2 is superior to that of Algorithm 1 by a factor of $O(\sqrt{n})$. The savings of $\sqrt{n}$ parallels the savings of $\sqrt{n}$ obtained for potential function reduction algorithms that use a symmetric primal-dual potential function versus a primal potential function, see Ye [11] or [3]. Secondly, the constant K in Theorem 4.1 for Algorithm 2 is readily computable as a function of the initial data for the problem. This contrasts with the constant K in Theorem 3.1 for Algorithm 1, which involves the unknown constants ρ and δ.

The attractiveness of Algorithm 2 over Algorithm 1 is diminished when considered from the standpoint of the initial assumptions. The additional initial assumptions needed for Algorithm 2 are assumptions A4' and A6'. Assumption A4' states that we know a dual feasible solution $(\hat{\pi}, \hat{s})$ and that $(\hat{\pi}, \hat{s})$ together with $\hat{x}$ are feasible for the potential function reduction problem HF. There are many instances where this assumption is not readily satisfiable. First, it assumes that an interior dual feasible solution is known, which is not usually the case in practice. Second, it assumes that this interior dual feasible solution results in a feasible solution for HF. However, in many cases this might be impossible. For example, suppose the initial value $\hat{x}$ has all components negative or zero, i.e., $\hat{x} \leq 0$. Then $\hat{x}^T \hat{s} \leq 0$ and so the initial feasibility condition for HF that $\hat{x} + h(\hat{x}^T \hat{s}) > 0$ cannot hold. In contrast, assumption A4 for Algorithm 1 only requires that a bound $\hat{B}$ on the optimal objective value z^* be known. This assumption is usually satisfied in practice.

30

One way to circumvent the restrictiveness of assumption A4' of Algorithm 2 is to first run Algorithm 1 (but with the constant $k = 12\sqrt{n}$) until the algorithm takes its first dual step. At that point the dual values $(\bar{\pi}, \bar{s})$ together with the current primal value $\bar{x}$ will satisfy assumption A4' and so now Algorithm 2 can be initiated. Note that this strategy will typically result in a larger initial duality gap (by a factor of $\sqrt{n}$) than if Algorithm 1 was run with the value of k set to $k = 9$. This is because the computation of B^0 in Step 0 of Algorithm 1 involves terms of the form $1/h_j$. Therefore with a constant of $k = 12\sqrt{n}$ versus $k = 9$ used to rescale the shift vector h, then the value of the gap $(c^T x^0 - B^0)$ could be larger by a factor of $(12/9)\sqrt{n} = 4\sqrt{n}/3$.

Appendix – Some Logarithmic Inequalities

In this appendix, we present a sequence of inequalities involving logarithms

Proposition A.1. If $a > -1$, then $ln\,(1 + a) \geq a$. ∎

Proposition A.2. If $|a| \leq b < 1$, $ln\,(1 + a) \leq a - \dfrac{a^2}{2(1 - b)}$. ∎

Proposition A.1 follows from the concavity of the function $ln\,(x)$, and a proof of Proposition A.2 can be found in Todd and Ye [9].

Proposition A.3. If $n > 0$ and $k > 0$, then $n \ln \left(1 + \dfrac{1}{k\sqrt{n}}\right) \leq \dfrac{\sqrt{n}}{k}$. ∎

Proposition A.3 follows from Proposition A.1 by setting $a = \dfrac{1}{k\sqrt{n}}$.

Proposition A.4. If $a \in R^n$, $b \in R^n$, $a > 0$, $b > 0$, then

$$n\,ln\,(a^T b) \geq \sum_{j=1}^{n} ln\,(a_j) + \sum_{j=1}^{n} ln\,(b_j) + n\,ln\,n.$$

Proof: This inequality is essentially the arithmetic-geometric mean inequality. Note that

$$\prod_{j=1}^{n} \left(\frac{a_j b_j}{a^T b}\right) \leq \left(\frac{1}{n}\right)^n$$, from which the stated result follows by taking logarithms . ∎

Proposition A.5. Let $s = \rho\, Y^{-1}\,(e + d)$, where $\rho > 0$, $y, e, d, s \in R^n$, and $y > 0$,

$\|d\| \leq \gamma < 1$, and $Y = \text{diag}\,(y)$. Then $\left\|Ys - \left(\dfrac{y^T s}{n}\right) e\right\| \leq \left(\dfrac{y^T s}{n}\right)\left(\dfrac{\gamma}{1 - \gamma}\right)$.

Proof: First note that $Ys - \left(\dfrac{y^T s}{n}\right) e = \rho\left[I - \dfrac{e\,e^T}{n}\right] d$, and because the matrix in brackets is a projection matrix, we have

32

$\left\| Ys - \left(\dfrac{y^T s}{n}\right) e \right\| \le \rho \|d\| \le \rho\,\gamma$. It thus remains to show that $\rho \le \dfrac{y^T s}{n\,(1-\gamma)}$.

To see this, note $y^T s = \rho\,(e^T e + e^T d) \ge \rho\,(n - \sqrt{n}\,\gamma) \ge \rho\,n\,(1-\gamma)$, from which it

follows that $\rho \le \dfrac{y^T s}{n\,(1-\gamma)}$. $\blacksquare$

<u>Proposition A.6</u>. Let $s = \rho\,Y^{-1}(e+d)$, where $\rho > 0$, $y, e, d \in R^n$, and $y > 0$,

$\|d\| \le \gamma < 1$. Then $\displaystyle\sum_{j=1}^{n} ln\, y_j + \sum_{j=1}^{n} ln\, s_j \ge n\, ln\, (y^T s) - n\, ln\, n - \dfrac{\gamma^2}{2(1-\gamma)}$.

<u>Proof</u>: For each j , $y_j s_j = \rho\,(1 + d_j)$, so that

$$ln\, y_j + ln\, s_j = \ln \rho + \ln (1 + d_j) \ge \ln \rho + d_j - \frac{d_j^2}{2(1-\gamma)} \; , \; \text{from Proposition A.2.}$$

Thus

$$\sum_{j=1}^{n} ln\, y_j + \sum_{j=1}^{n} ln\, s_j \ge n\, ln\, \rho + e^T d - \frac{\gamma^2}{2(1-\gamma)} \; . \tag{A1}$$

Also,

$$n\, ln\, (y^T s) = n\, ln\, (\rho\,(n + e^T d)) = n\, ln\, \rho + n\, ln\, (n + e^T d)$$

$$\le n\, ln\, \rho + n\, ln\, n + e^T d \tag{A2}$$

from Proposition A.1. Combining (A1) and (A2) gives the result. $\blacksquare$

References

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

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

[3]   Freund, R. M. (1988), "Polynomial-time algorithms for linear programming based only on primal scaling and projected gradients of a potential function," to appear in **Mathematical Programming**.

[4]   Freund, R. M. (1989), "Theoretical efficiency of a shifted barrier function algorithm for linear programming" Working paper OR 194-89, Operations Research Center, M.I.T., Cambridge, MA.

[5]   Gill, P., W. Murray, M. Saunders, J. Tomlin, and M. Wright (1989), "Shifted barrier methods for linar programming," forthcoming.

[6]   Gonzaga, C. C. (1988), "Polynomial affine algorithms for linear programming," Report ES-139/88, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

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

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

[9]   Todd, M. J. and Y. Ye (1987), "A centered projective algorithm for linear programming," Technical report No. 763, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY.

[10]   Todd, M. J. (1988), "On Anstreicher's Combined Phase I – Phase II projective algorithm for linear programming," Technical Report 776, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY.

[11]   Ye, Y. (1988), "A class of potential functions for linear programming," to appear in **Mathematical Programming**.