SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NY 14853-3801

TECHNICAL REPORT NO. 1044

February 1993

ANALYSIS OF INTERIOR-POINT METHODS FOR LINEAR PROGRAMMING PROBLEMS WITH VARIABLE UPPER BOUNDS

by

Michael J. Todd[1]

[1] Research supported in part by NSF, AFOSR and ONR through NSF Grant DMS-8920550.

Abstract

We describe path-following and potential-reduction algorithms for linear programming problems with variable upper bounds. Both methods depend on a barrier function for the cone of solutions to the variable upper bounds, and we establish the key properties of this barrier that allow the complexity of the algorithms to be analyzed. These properties mostly follow from the self-concordance of the function, a notion introduced by Nesterov and Nemirovsky. Our analysis follows that of Freund and Todd for problems with (possibly two-sided) simple bounds.

Key words: linear programming, variable upper bounds, barrier functions, interior-point algorithms, self-concordance.

Running Header: Interior-Point Methods for Variable Upper Bounds

1 Introduction

The aim of this paper is to describe two polynomial-time interior-point algorithms for linear programming problems with variable upper bounds, one based on path-following and the other on potential reduction. In this endeavor, we are continuing the work of Freund and Todd [5] for the case of simple upper and lower bounds; the idea is to highlight the role of the barrier function and elucidate in a simple setting the notion of self-concordance introduced by Nesterov and Nemirovsky [12]. At the same time we show how variable upper bounds can be handled implicitly, without increasing the size of the linear system to be solved at each iteration.

Exploiting the structure of variable upper bounds has been discussed by Schrage [16] and Todd [17] in the context of the simplex method, by Todd [18] for Karmarkar's projective algorithm, and by Choi and Goldfarb [2] for a short-step primal-dual path-following method. In addition, since variable upper bounds define a cone, the general primal-dual potention-reduction method of Nesterov and Nemirovsky (see [11] or Chapter 3 of [12]) can be adapted to this problem. However, as we argue in Section 2.5, their algorithm is also restricted to relatively short step sizes.

Let $N = \{1, 2, \ldots, n\}$ index the variables of a linear programming problem, and suppose $J \cup K \cup L \cup F = N$ is a partition of N. A variable upper bound is a constraint of the form $x_j \le x_{k(j)}$ where $j \in J$ and $k(j) \in K$. We call j (or x_j) the *child* of its *parent* $k(j)$ (or $x_{k(j)}$); each child has only one parent, but a parent may have several (but at least one) children. Variables in L (as well as those in $J \cup K$) are required to be nonnegative, while those in F are free. We assume there are no simple upper bounds. Let $J(k) = \{j \in J : k(j) = k\}$. Then our problem can be written

$$(P) \qquad \begin{aligned} \min_x \quad & c^\top x \\ Ax &= b \\ x &\in C, \end{aligned}$$

where

$$C := \{x \in I\!\!R^n : \begin{array}{lll} 0 \le x_j \le x_k, & j \in J(k), & k \in K, \\ 0 \le x_\ell, & \ell \in L\} & \end{array} \tag{1.1}$$

and A is an $m \times n$ matrix. If we added the constraints $x_j \leq x_k$ with slack variables explicitly to the constraint matrix, m and n would increase by $|J|$, a substantial increase if $m \ll n$ and $|J|$ is of the same order as n.

Writing the dual problem directly, we obtain

$$
(D) \qquad
\begin{array}{llcll}
\max_{y,s} & b^\top y & & & \\
& a_j^\top y & - \quad t_j & \leq c_j, & j \in J, \\
& a_k^\top y & + \sum\limits_{j \in J(k)} t_j & = c_k, & k \in K, \\
& a_\ell^\top y & & \leq c_\ell, & \ell \in L, \\
& a_f^\top y & & = c_f, & f \in F, \\
& & t_j \geq 0, & & j \in J.
\end{array}
$$

It is easy to see that y is feasible in (D) with some t iff $s = c - A^\top y$ lies in

$$
\begin{array}{rlll}
C^* := \{ s \in I\!R^n : & s_k + \sum\limits_{i \in I(k)} s_i \geq 0, & I(k) \subseteq J(k), & k \in K, \\
& s_\ell \geq 0, & \ell \in L, & \\
& s_f = 0, & f \in F \},&
\end{array}
\tag{1.2}
$$

the dual or polar cone of C: $C^* := \{ s \in I\!R^n : x^\top s \geq 0 \text{ for all } x \in C \}$. Note that C can also be described in terms of its generators,

$$
\begin{array}{rlll}
C = \text{cone}\{ & e^k + \sum\limits_{i \in I(k)} e^i, & I(k) \subseteq J(k), & k \in K, \\
& e^\ell, & \ell \in L, & \\
& e^f, -e^f, & f \in F \},&
\end{array}
\tag{1.3}
$$

where e^j denotes the jth unit vector in $I\!R^n$, and similarly

$$
\begin{array}{rll}
C^* = \text{cone}\{ & e^j, e^{k(j)} - e^j, & j \in J, \\
& e^\ell, & \ell \in L \}.
\end{array}
\tag{1.4}
$$

We can then describe (D) compactly as

$$(D) \qquad \begin{array}{rl} \max_{y,s} & b^{\mathsf{T}} y \\ & A^{\mathsf{T}} y + s = c \\ & s \in C^*, \end{array}$$

and this is (easily seen to be equivalent to) the dual problem of Nesterov and Nemirovsky ([11] or Chapter 3 of [12]).

For any x feasible in (P) and (y, s) feasible in (D), the duality gap is seen to be

$$c^{\mathsf{T}} x - b^{\mathsf{T}} y = (A^{\mathsf{T}} y + s)^{\mathsf{T}} x - (Ax)^{\mathsf{T}} y = x^{\mathsf{T}} s \geq 0, \tag{1.5}$$

so $x^{\mathsf{T}} s$ is the gap as in the standard case of nonnegativities only in C.

We make the following assumptions throughout the paper:

(A1) $F^0(P) := \{x \in \text{int } C : Ax = b\} \neq \emptyset$;

(A2) the set of optimal solutions of (P) is nonempty and bounded; and

(A3) A has rank m.

The assumption (A3) is for convenience only; it can easily be relaxed and only minor modifications are necessary. Of course, given feasibility of (P), linearly dependent rows of A give redundant constraints, which can be deleted.

If there were a linear dependence among the columns of A corresponding to the free variables, say $A_F x_F = 0$, $x_F \neq 0$, then necessarily $c_F^{\mathsf{T}} x_F = 0$ (else there would be no optimal solutions to (P)) and hence the set of optimal solutions of (P) would be unbounded. Hence (A2) implies that

$$A_F \quad \text{has full column rank.} \tag{1.6}$$

Of course, this assumption in A_F is fairly harmless; as long as (P) has an optimal solution, $A_F x_F = 0$ implies $c_F^{\mathsf{T}} x_F = 0$, and so linearly dependent columns of A_F could be deleted without changing the problem. (This is dual to removing dependent rows.) (A2) is stronger than this requirement; as in the standard case, it is not hard to show that it implies

$$F^0(D) := \{(y,s) \in I\!R^m \times \text{ri } C^* : A^\top y + s = c\} \neq \emptyset, \tag{1.7}$$

where ri C^* is the relative interior of C^*, obtained by making all the inequalities in (1.2) strict. Conversely, (1.6) and (1.7) imply (A2), and similarly (A1) and (A3) imply that the set of optimal solutions of (D) is nonempty and bounded.

In Section 2 we describe and study a barrier function for C and show how it determines primal and dual norms and projections as well as central trajectories. These form the basis for the algorithms given in Section 3.

2 Analysis

In this section we define a barrier function for int C, and develop its key properties for our algorithms. Section 2.1 defines the barrier and computes its gradient and Hessian matrix. In Section 2.2 we use these to define primal and dual metrics. Then we establish key self-concordance properties and Taylor approximation results in Section 2.3. (The reader may prefer to skip the proofs here on a first reading.) Section 2.4 shows how the metrics are used to define projections. The central path is defined in Section 2.5, and Section 2.6 uses the projections defined in Section 2.4 to analyze near-central points.

2.1 A barrier function for int C

Here we describe a barrier function for int C, a convex function that tends to $+\infty$ if the argument converges to a point of $C \backslash (\text{int } C)$, and compute its first and second derivatives. Our barrier is a simple logarithmic function, following the standard techniques for constructing barriers, with one term for each inequality defining C. We group these as follows. For each x in

$$\text{int } C = \{x \in I\!R^n : \begin{array}{ll} 0 < x_j < x_k, & j \in J(k), k \in K, \\ 0 < x_\ell, & \ell \in L\}, \end{array} \tag{2.1}$$

define

$$\Psi^j(x) := -\ln x_j - \ln(x_{k(j)} - x_j), \quad j \in J, \tag{2.2}$$

and

$$\Psi^{\ell}(x) := -\ln x_{\ell}, \quad \ell \in L, \tag{2.3}$$

and then set

$$\Psi(x) := \sum_{j\in J} \Psi^j(x) + \sum_{\ell \in L} \Psi^{\ell}(x). \tag{2.4}$$

From (2.2) we have, for $j \in J$,

$$\begin{aligned} \nabla \Psi^j(x) &= ((x_k - x_j)^{-1} - x_j^{-1})e^j - (x_k - x_j)^{-1} e^k \\ &= -x_j^{-1} e^j - (x_k - x_j)^{-1}(e^k - e^j) \end{aligned} \tag{2.5}$$

and

$$\begin{aligned} \nabla^2 \Psi^j(x) &= ((x_k - x_j)^{-2} + x_j^{-2}) e^j (e^j)^{\top} \\ &\quad + (-(x_k - x_j)^{-2})(e^j (e^k)^{\top} + e^k (e^j)^{\top}) \\ &\quad + (x_k - x_j)^{-2} e^k (e^k)^{\top}, \\ &= x_j^{-2} e^j (e^j)^{\top} + (x_k - x_j)^{-2} (e^k - e^j)(e^k - e^j)^{\top}, \end{aligned} \tag{2.6}$$

where $k := k(j)$, and from (2.3) we have

$$\nabla \Psi^{\ell}(x) = -x_{\ell}^{-1} e^{\ell} \tag{2.7}$$

and

$$\nabla^2 \Psi^{\ell}(x) = x_{\ell}^{-2} e^{\ell} (e^{\ell})^{\top}. \tag{2.8}$$

From these elements we can assemble the gradient and Hessian of Ψ. (Note that we use superscripts for both powers and indices; no confusion should result.) For example, if $n = 7$, $J = \{1, 2, 4\}$, $K = \{3, 5\}$, $k(1) = k(2) = 3$, $k(4) = 5$, and $L = \{6\}$, we find

$$\nabla^2\Psi(x) = \begin{bmatrix} \alpha_1^{-2}+\beta_1^{-2} & & -\beta_1^{-2} & & & & \\ & \alpha_2^{-2}+\beta_2^{-2} & -\beta_2^{-2} & & & & \\ -\beta_1^{-2} & -\beta_2^{-2} & \beta_1^{-2}+\beta_2^{-2} & & & & \\ & & & \alpha_4^{-2}+\beta_4^{-2} & -\beta_4^{-2} & & \\ & & & -\beta_4^{-2} & \beta_4^{-2} & & \\ & & & & & \alpha_6^{-2} & \\ & & & & & & 0 \end{bmatrix},$$

where $\alpha_i = x_i$, $i = 1, 2, 4, 6$; $\beta_1 = x_3 - x_1$, $\beta_2 = x_3 - x_2$; and $\beta_4 = x_5 - x_4$. All unmarked entries are zero. We see that $\nabla^2\Psi$ has a block diagonal structure where each block has an arrowhead pattern if "families" are grouped together, with the parent appearing last. (The same structure appears in the matrix S in Choi and Goldfarb [2]; see their (4.2).)

2.2 Primal and dual metrics

Let us fix some $\hat{x} \in \operatorname{int} C$ and denote $\nabla^2\Psi(\hat{x})$ by $\hat{\Theta}^2$. We note from (2.6) and (2.8) that $\nabla^2\Psi(\hat{x})$ is symmetric positive semi-definite; thus this notation is appropriate if we use $\hat{\Theta}$ for the symmetric positive semi-definite square root of $\nabla^2\Psi(\hat{x})$. We further note that $\hat{\Theta}^2$ is positive definite on

$$I\!\!R^{J\cup K\cup L} := \{x \in I\!\!R^n : x_f = 0 \quad \text{for all} \quad f \in F\}. \tag{2.9}$$

Let us write $v^{J\cup K\cup L}$ for the projection of a vector $v \in I\!\!R^n$ into $I\!\!R^{J\cup K\cup L}$, obtained by setting its components indexed by F to zero. Let V be any subspace of $I\!\!R^n$ satisfying

$$v \in V \quad \text{and} \quad v^{J\cup K\cup L} = 0 \quad \text{imply} \quad v = 0. \tag{2.10}$$

By the discussion above,

Proposition 2.1 $\|\cdot\|_{\hat{x}}$, *defined by*

$$\|v\|_{\hat{x}} := \|\hat{\Theta}v\|_2, \tag{2.11}$$

is a norm on V. □

From (1.6), we see that the null space of A, $\mathcal{N}(\mathcal{A})$, satisfies (2.10), so $\|\cdot\|_{\hat{x}}$ provides a norm on this space, and hence a metric on $F^0(P)$.

Note that the lineality space of C (the largest subspace contained in it) is

$$\mathbb{R}^F := \{x \in \mathbb{R}^n : x^{J \cup K \cup L} = 0\},$$

while its linear span is $C - C = \mathbb{R}^n$. Also, the lineality space of C^* is $\{0\}$ and its span is $C^* - C^* = \mathbb{R}^{J \cup K \cup L}$. The difference of s's for any two feasible solutions in $F^0(D)$ lies in $C^* - C^* = \mathbb{R}^{J \cup K \cup L}$, and this subspace also contains the ranges of $\nabla\Psi$ and of $\hat{\Theta}^2$. We now define a dual norm on this space.

Since $\hat{\Theta}^2$ is nonsingular (and positive definite) as an operator on $\mathbb{R}^{J \cup K \cup L}$, it has an inverse, also nonsingular and positive definite, which we denote $\hat{\Theta}^{-2}$. We also use $\hat{\Theta}^{-2}$ to denote the matrix that is block diagonal, representing the operator $\hat{\Theta}^{-2}$ on $\mathbb{R}^{J \cup K \cup L}$, and diagonal with diagonal entries equal to $+\infty$ on $\mathbb{R}^F$. Our convention is that $(+\infty)0 = 0$, so $\hat{\Theta}^{-2}v$ is well-defined for $v \in \mathbb{R}^{J \cup K \cup L}$, whether $\hat{\Theta}^{-2}$ is viewed as an operator or a matrix. We define $\hat{\Theta}^{-1}$ similarly.

Proposition 2.2 $\|\cdot\|_{\hat{x}}^*$, *defined by*

$$\|v\|_{\hat{x}}^* = \|\hat{\Theta}^{-1}v\|_2, \tag{2.12}$$

is a norm on $C^* - C^* = \mathbb{R}^{J \cup K \cup L}$. □

2.3 Self-concordance and Taylor approximations to Ψ

Here we show that Ψ is self-concordant in the sense of Nesterov and Nemirovsky [12] and also establish some key bounds on the errors in Taylor approximations to Ψ and $\nabla\Psi$.

A convex function Φ on an open subset $\mathcal{Q}$ of $\mathbb{R}^n$ is said to be *self-concordant* (with parameter 1) if Φ is C^3 and for every $x \in \mathcal{Q}$ and $d \in \mathbb{R}^n$,

$$|D^3\Phi(x)[d,d,d]| \le 2(d^\top \nabla^2 \Phi(x) d)^{3/2}. \tag{2.13}$$

Here $D^3\Phi(x)$ denotes the third derivative of Φ. We easily find

$$D^3\Psi(x)[d,d,d] = \sum_{k\in K}\sum_{j\in J(k)} \left[-2\frac{d_j^3}{x_j^3} - 2\frac{(d_k-d_j)^3}{(x_k-x_j)^3}\right] + \sum_{\ell\in L}\left(-2\frac{d_\ell^3}{x_\ell^3}\right) \tag{2.14}$$

while

$$d^\top \nabla^2\Psi(x) d = \sum_{k\in K}\sum_{j\in J(k)} \left[\frac{d_j^2}{x_j^2} + \frac{(d_k-d_j)^2}{(x_k-x_j)^2}\right] + \sum_{\ell\in L}\frac{d_\ell^2}{x_\ell^2}, \tag{2.15}$$

so (2.13) holds for Ψ by the inequality $\sum \mu_i^3 \le (\sum \mu_i^2)^{3/2}$ for nonnegative μ_i's.

Note that Ψ also satisfies

$$\Psi(\lambda x) = \Psi(x) - (2|J| + |L|)\ln\lambda, \tag{2.16}$$

for $\lambda > 0$ and $x \in \operatorname{int} C$, directly from the definition. For ease of notation, we define

$$p := |J| + \frac{1}{2}|L|, \tag{2.17}$$

and note that $p \le n - |F|$, the number of non-free variables. Equation (2.16) is then the defining relationship for Ψ to be logarithmically homogeneous with homogeneity parameter $2p$ (Nesterov and Nemirovsky [11]). Hence we have:

Theorem 2.1 *Ψ is a $(2p)$-logarithmically homogeneous self-concordant barrier for* int C. □

(The terminology $(2p)$-normal barrier is used in ([12], Chapter 3).)

We can now follow the development of [11, 12]: (2.16) implies (differentiating with respect to λ at $\lambda = 1$)

$$\nabla\Psi(x)^\top x = -2p \tag{2.18}$$

and (differentiating with respect to x)

$$\nabla\Psi(\lambda x) = \lambda^{-1}\nabla\Psi(x), \tag{2.19}$$

and hence (differentiating (2.19) with respect to λ at $\lambda = 1$)

$$\nabla^2\Psi(x)\cdot x = -\nabla\Psi(x), \tag{2.20}$$

and

$$\nabla^2\Psi(x)\cdot x^{J\cup K\cup L} = -\nabla\Psi(x). \tag{2.21}$$

From this we obtain (similar to [5]):

Proposition 2.3 *For any $x \in \text{int } C$, $\nabla\Psi(x) \in -\text{ri } C^* \subseteq \mathbb{R}^{J\cup K\cup L}$ and*

$$\|\nabla\Psi(x)\|_x^* = (2p)^{1/2} \leq (2n)^{1/2}. \tag{2.22}$$

Proof. (2.5) and (2.7) show that $-\nabla\Psi(x)$ is a positive combination of the generators of C^* (see (1.4)), establishing the first part. For the second,

$$\begin{aligned} \nabla\Psi(x)^\top(\nabla^2\Psi(x))^{-1}\nabla\Psi(x) &= -\nabla\Psi(x)^\top x^{J\cup K\cup L} \\ &= -\nabla\Psi(x)^\top x \\ &= 2p, \end{aligned}$$

where the first equation comes from (2.21) and the last from (2.18). This proves (2.22). □

The following result is essentially a general consequence of self-concordance (see Theorem 1.1 of Nesterov and Nemirovsky [12]) but we provide a direct proof because it is so simple.

Proposition 2.4 *If $\hat{x} \in \text{int } C$ and $d \in \mathbb{R}^n$ with $\|d\|_{\hat{x}} < 1$, then $\hat{x}+d \in \text{int } C$.*

(Note that we can use $\|\cdot\|_{\hat{x}}$, defined in (2.11), even for vectors not lying in a subspace V satisfying (2.10).)

Proof. Using (2.6) and (2.8) we find

$$
\begin{aligned}
1 > \|d\|_{\hat{x}}^2 &= d^\top \nabla^2 \Psi(\hat{x}) d \\
&= \sum_J \left[\frac{d_j^2}{\hat{x}_j^2} + \frac{(d_{k(j)} - d_j)^2}{(\hat{x}_{k(j)} - \hat{x}_j)^2} \right] + \sum_L \frac{d_\ell^2}{\hat{x}_\ell^2},
\end{aligned}
$$

whence $|d_j| < \hat{x}_j$ and $|d_{k(j)} - d_j| < \hat{x}_{k(j)} - \hat{x}_j$ for each $j \in J$ and $|d_\ell| < \hat{x}_\ell$ for each $\ell \in L$. This implies $\hat{x} + d$ lies in int C. □

We can use this property of self-concordance to eliminate another natural candidate for a self-concordant barrier. Indeed, by analogy with [5], we might suppose that Ψ^j in (2.2) could be replaced by

$$
\tilde{\Psi}^j(x) := \frac{\min\{x_j, x_k - x_j\}}{x_k/2} - \ln(\min\{x_j, x_k - x_j\})
$$

for $k = k(j)$. This $\tilde{\Psi}^j$ can be shown to be twice continuously differentiable (but not thrice), and its Hessian at a point x with $x_j < x_k/2$ is

$$
\nabla^2 \tilde{\Psi}^j(x) = x_j^{-2} e^j (e^j)^\top - 2x_k^{-2}(e^j(e^k)^\top + e^k(e^j)^\top) + 4x_j x_k^{-3} e^k (e^k)^\top .
$$

Suppose $k \in K$ and $\hat{x} \in$ int C has $\hat{x}_j < \hat{x}_k/2$ for all $j \in J(k)$. Let $d = d_k e^k$, and note that

$$
d^\top \nabla^2 \tilde{\Psi}(\hat{x}) d = 4(\sum_{J(k)} \hat{x}_j / \hat{x}_k)(d_k/\hat{x}_k)^2,
$$

where $\tilde{\Psi} := \sum_J \tilde{\Psi}^j + \sum_L \Psi^\ell$ is the new barrier function. It follows that we may have $d^\top \nabla^2 \tilde{\Psi}(\hat{x}) d$ less than 1 while $d_k < -\hat{x}_k$ (so $\hat{x}_k + d_k < 0$) as long as all $\hat{x}_j$'s, $j \in J(k)$, are sufficiently small. Hence Proposition 2.4 fails for $\tilde{\Psi}$, and we conclude that $\tilde{\Psi}$ is not self-concordant. Indeed, for any positive constant ϵ, $d^\top \nabla^2 \tilde{\Psi}(\hat{x}) d < \epsilon$ does not imply $d_k < -\hat{x}_k$ for all $\hat{x}$, so no positive multiple of $\tilde{\Psi}$ is self-concordant (or in other words, $\tilde{\Psi}$ is not self-concordant with any positive parameter [12]).

We now establish a key result on the first-order Taylor approximation of $\nabla \Psi$.

Theorem 2.2 *Let $\hat{x}, x \in \operatorname{int} C$ and let $\bar{d} := x - \hat{x}$ and $\hat{\Theta}^2 := \nabla^2\Psi(\hat{x})$. Then*

$$\|\nabla\Psi(x) - \nabla\Psi(\hat{x}) - \hat{\Theta}^2\bar{d}\|_x^* \leq \|\bar{d}\|_{\hat{x}}^2. \tag{2.23}$$

(Note that the vector appearing on the left-hand side of (2.23) is the error in the first-order Taylor approximation to $\nabla\Psi(x)$ based on the point $\hat{x}$. Since all vectors in this expression lie in $\mathbb{R}^{J\cup K\cup L}$, its dual norm is defined, and indeed appropriate—gradients of Ψ lie in dual space. On the right-hand side we have the square of the primal norm of the primal displacement $\bar{d}$. Note also that the dual norm is with respect to the "new point" x, while the primal norm corresponds to the base point $\hat{x}$.)

Proof. We obtain a bound on the norm of each constituent of the vector appearing on the left in (2.23). For ease of notation, let

$$\begin{array}{l}
\alpha_j = x_j, \quad \hat{\alpha}_j = \hat{x}_j, \quad \beta_j = x_{k(j)} - x_j, \quad \hat{\beta}_j = \hat{x}_{k(j)} - \hat{x}_j, \\
\alpha_\ell = x_\ell, \quad \hat{\alpha}_\ell = \hat{x}_\ell, \\
\delta_j = x_j - \hat{x}_j = \alpha_j - \hat{\alpha}_j, \quad \epsilon_j = x_{k(j)} - x_j - (\hat{x}_{k(j)} - \hat{x}_j) = \beta_j - \hat{\beta}_j, \\
\delta_\ell = x_\ell - \hat{x}_\ell = \alpha_\ell - \hat{\alpha}_\ell.
\end{array}$$

We also omit subscripts when they are clear from the context. Then

$$\begin{array}{rl}
\multicolumn{2}{l}{\nabla\Psi^j(x) - \nabla\Psi^j(\hat{x}) - \nabla^2\Psi^j(\hat{x})(x - \hat{x})} \\
= & -\alpha_j^{-1}e^j - \beta_j^{-1}(e^k - e^j) + \hat{\alpha}_j^{-1}e^j + \hat{\beta}_j^{-1}(e^k - e^j) \\
 & -\hat{\alpha}_j^{-2}\delta_j e^j - \hat{\beta}_j^{-2}\epsilon_j(e^k - e^j) \\
= & \alpha^{-1}\hat{\alpha}^{-2}(-\hat{\alpha}^2 + \alpha\hat{\alpha} - \alpha\delta)e^j + \beta^{-1}\hat{\beta}^{-2}(-\hat{\beta}^2 + \beta\hat{\beta} - \beta\epsilon)(e^k - e^j) \\
= & -\alpha^{-1}\hat{\alpha}^{-2}\delta^2 e^j - \beta^{-1}\hat{\beta}^{-2}\epsilon^2(e^k - e^j),
\end{array} \tag{2.24}$$

where $k = k(j)$.

Now if P and Q are symmetric positive semi-definite matrices with $P \geq Q$ ($P - Q$ positive semi-definite), and $v \in \operatorname{Im}(Q) \subseteq \operatorname{Im}(P)$, then we can show that

$$v^\top P^{-1} v \leq v^\top Q^{-1} v,$$

where $P^{-1}v$ denotes the vector in the range of P with $P(P^{-1}v) = v$, and similarly for $Q^{-1}v$. Indeed, this is a standard result when P and Q are positive definite, and the general result follows by a limiting argument, using $P_k := P + \frac{1}{k}I$ and similarly for Q_k. (I am grateful to C. Van Loan for this line of reasoning, which simplifies my earlier argument using Schur complements.)

Applying this inequality to the vector in (2.24), with $P = \nabla^2\Psi(x)$ and $Q = \nabla^2\Psi^j(x)$, we obtain

$$
\begin{aligned}
&(\|\nabla\Psi^j(x) - \nabla\Psi^j(\hat{x}) - \nabla^2\Psi^j(\hat{x})(x-\hat{x})\|_x^*)^2 \\
&\leq \begin{pmatrix} -\alpha^{-1}\hat{\alpha}^{-2}\delta^2 + \beta^{-1}\hat{\beta}^{-2}\epsilon^2 \\ -\beta^{-1}\hat{\beta}^{-2}\epsilon^2 \end{pmatrix}^\top \begin{pmatrix} \alpha^{-2}+\beta^{-2} & -\beta^{-2} \\ -\beta^{-2} & \beta^{-2} \end{pmatrix}^{-1} \\
&\quad \begin{pmatrix} -\alpha^{-1}\hat{\alpha}^{-2}\delta^2 + \beta^{-1}\hat{\beta}^{-2}\epsilon^2 \\ -\beta^{-1}\hat{\beta}^{-2}\epsilon^2 \end{pmatrix} \\
&= \begin{pmatrix} -\alpha^{-1}\hat{\alpha}^{-2}\delta^2 + \beta^{-1}\hat{\beta}^{-2}\epsilon^2 \\ -\beta^{-1}\hat{\beta}^{-2}\epsilon^2 \end{pmatrix}^\top \begin{pmatrix} \alpha^2 & \alpha^2 \\ \alpha^2 & \alpha^2+\beta^2 \end{pmatrix} \\
&\quad \begin{pmatrix} -\alpha^{-1}\hat{\alpha}^{-2}\delta^2 + \beta^{-1}\hat{\beta}^{-2}\epsilon^2 \\ -\beta^{-1}\hat{\beta}^{-2}\epsilon^2 \end{pmatrix} \\
&= \hat{\alpha}^{-4}\delta^4 + \hat{\beta}^{-4}\epsilon^4 \leq (\hat{\alpha}^{-2}\delta^2 + \hat{\beta}^{-2}\epsilon^2)^2 \\
&= (\bar{d}^\top\nabla^2\Psi^j(\hat{x})\bar{d})^2.
\end{aligned}
$$

In the same way, using $\nabla^2\Psi(x) \geq \nabla^2\Psi^\ell(x)$, we get

$$
(\|\nabla\Psi^\ell(x) - \nabla\Psi^\ell(\hat{x}) - \nabla^2\Psi^\ell(\hat{x})(x-\hat{x})\|_x^*)^2 \leq \hat{\alpha}_\ell^{-4}\delta_\ell^4 = (\bar{d}^\top\nabla^2\Psi^\ell(\hat{x})\bar{d})^2.
$$

Adding the square roots of all these inequalities gives the desired inequality (2.23). □

From this result we can prove, exactly as in [5],

Theorem 2.3 *Let $\hat{x} \in \operatorname{int} C$. If $\bar{d} \in \mathbb{R}^n$ and $\gamma > 0$ are such that $\gamma\|\bar{d}\|_{\hat{x}} < 1$, then $\hat{x} + \gamma\bar{d} \in \operatorname{int} C$ and*

$$
\begin{aligned}
&\Psi(\hat{x}) + \gamma\nabla\Psi(\hat{x})^\top\bar{d} \leq \Psi(\hat{x}+\gamma\bar{d}) \\
&\quad \leq \Psi(\hat{x}) + \gamma\nabla\Psi(\hat{x})^\top\bar{d} + \tfrac{\gamma^2\|\bar{d}\|_{\hat{x}}^2}{2(1-\gamma\|\bar{d}\|_{\hat{x}})}.
\end{aligned} \tag{2.25}
$$

Proof. The first part follows from Proposition 2.4, while (2.25) is derived from (2.23) and the fundamental theorem of calculus exactly as in [5], using the following consequence of the self-concordance property established in Theorem 2.1: For every $\in$ int $C, d \in I\!R^n$, and $h \in I\!R^n$, if $\gamma \in I\!R$ satisfies $|\gamma|\|d\|_x < 1$, then

$$(1-|\gamma|\|d\|_x)^2 h^\top \nabla^2\Psi(x)h \leq h^\top \nabla^2\Psi(x+\gamma d)h \leq (1-|\gamma|\|d\|_x)^{-2} h^\top \nabla^2\Psi(x)h.$$

(This implication is Theorem 1.1 of Nesterov and Nemirovsky [12].) □

2.4 Projections

We can use the metrics in Section 2.2 to perform projections. Note that the Euclidean projection of a vector $v \in I\!R^n$ onto $\mathcal{N}(\mathcal{A})$ can easily be seen to be the unique solution to

$$\max_d \{v^\top d - \frac{1}{2}\|d\|^2 : Ad = 0\}.$$

Correspondingly, we have (as in [5]):

Theorem 2.4 *Fix $\hat{x} \in$ int C. For each $v \in I\!R^n$, there is a unique solution to the problem*

$$\begin{array}{ll} \max\limits_d & v^\top d - \frac{1}{2}\|d\|_{\hat{x}}^2 \\ & Ad = 0. \end{array} \tag{2.26}$$

Moreover, this solution $\bar{d}$ is part of the solution $(\bar{d}, \bar{y})$ to the system

$$\begin{pmatrix} \hat{\Theta} & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} d \\ y \end{pmatrix} = \begin{pmatrix} v \\ 0 \end{pmatrix}, \tag{2.27}$$

where $\hat{\Theta}^2 := \nabla^2\Psi(\hat{x})$, and satisfies

$$\|\bar{d}\|_{\hat{x}}^2 = v^\top \bar{d}; \tag{2.28}$$

$$\|\bar{d}\|_{\hat{x}} = \|v - A^\top \bar{y}\|_{\hat{x}}^*; \tag{2.29}$$

and, if $v \in I\!R^{J\cup K\cup L}$,

$$\|\bar{d}\|_{\hat{x}} \leq \|v\|_{\hat{x}}^*. \tag{2.30}$$

Proof. Exactly as in [5]. The system (2.27) forms the Karush-Kuhn-Tucker conditions for the concave maximization problem (2.26), and these conditions are necessary and sufficient for optimality. The homogeneous system corresponding to (2.27) has only the trivial solution (using Proposition 2.1 and the remark following it, and assumption (A3)), so (2.27) has a unique solution. Then (2.28) – (2.30) follow from the first set of equations in (2.27). □

We write $P_{\hat{x}}(v)$ for the solution to (2.26). Simple algebra shows that (2.27) also forms the KKT conditions for

$$\begin{aligned} \min_{y,s} \quad & \tfrac{1}{2} s^\top \hat{\Theta}^{-2} s \\ & A^\top y + s = v \\ & s \in I\!R^{J \cup K \cup L}, \end{aligned}$$

which is feasible since A has rank m. Hence we obtain the dual result:

Theorem 2.5 *There is a unique solution to*

$$\begin{aligned} \min_{y} \quad & \tfrac{1}{2}(\|v - A^\top y\|_{\hat{x}}^*)^2 \\ & v - A^\top y \in I\!R^{J \cup K \cup L}. \end{aligned} \tag{2.31}$$

Moreover, this solution $\bar{y}$ is part of the solution $(\bar{d}, \bar{y})$ to (2.27). □

Hence (2.26) and (2.31) can be viewed as dual problems with equal optimal values.

Since computing projections is the basic task at each iteration, we conclude this section by describing how (2.27) can be solved efficiently. Of course, (2.27) is a sparse symmetric indefinite system of size $m + n$, and it can be solved by general techniques as in the standard case—see, for instance, Fourer and Mehrotra [3] or Vanderbei and Carpenter [19]. However, here we show how it can be reduced to a positive definite system of order m (assuming for simplicity there are no free variables).

Note that solving the system (2.27) reduces to the solution of a smaller system with coefficient matrix $A\hat{\Theta}^{-2}A^\top$. Since we are assuming $F = \emptyset$, $\hat{\Theta}^2$ is positive definite and hence its inverse is finite and also positive definite.

However, $\hat{\Theta}^{-2}$ is not diagonal when variable upper bounds are present, so our matrix is more complicated than $AD^2A^\top = \sum_j d_{jj}^2 a_j a_j^\top$ as in the standard case.

As noted in Section 2.1, $\hat{\Theta}^2$ is block diagonal. If we set

$$\begin{array}{lll} \hat{\alpha}_j = \hat{x}_j, & \hat{\beta}_j = \hat{x}_k - \hat{x}_j, & \hat{\gamma}_j = (\hat{\alpha}_j^{-2} + \hat{\beta}_j^{-2})^{1/2}, \\ \hat{\eta}_j = -\hat{\beta}_j^{-2}\hat{\gamma}_j^{-1}, & \text{and} & \hat{\kappa} = (\sum_{J(k)} \hat{\alpha}_j^{-2}\hat{\beta}_j^{-2}\hat{\gamma}_j^{-2})^{1/2}, \end{array} \tag{2.32}$$

for $j \in J(k)$, the kth block of $\hat{\Theta}^2$ is

$$\begin{pmatrix} \ddots & & & \vdots \\ & \hat{\alpha}_j^{-2} + \hat{\beta}_j^{-2} & & -\hat{\beta}_j^{-2} \\ & & \ddots & \vdots \\ \cdots & -\hat{\beta}_j^{-2} & \cdots & \sum_{J(k)} \hat{\beta}_j^{-2} \end{pmatrix}, \tag{2.33}$$

which has Cholesky factorization $WW^\top$ with

$$W = \begin{pmatrix} \ddots & & & \\ & \hat{\gamma}_j & & \\ & & \ddots & \\ \cdots & \hat{\eta}_j & \cdots & \hat{\kappa} \end{pmatrix}, \tag{2.34}$$

as can easily be checked. Thus the corresponding part of $A\hat{\Theta}^{-2}A^\top$ is

$$(\cdots a_j \cdots a_k)W^{-\top}W^{-1} \begin{pmatrix} \vdots \\ a_j^\top \\ \vdots \\ a_k \end{pmatrix}. \tag{2.35}$$

From (2.34) we find that

$$(\cdots a_j \cdots a_k)W^{-\top} = (\cdots \hat{\gamma}_j^{-1} a_j \cdots \hat{\kappa}^{-1}(a_k - \sum_{J(k)} \hat{\eta}_j \hat{\gamma}_j^{-1} a_j)). \tag{2.36}$$

Assembling all these pieces, and substituting back from (2.32), we find

$$
\begin{aligned}
A\hat{\Theta}^{-2}A^{\top} \;=\; & \textstyle\sum_{k\in K}\left[\sum_{j\in J(k)} \frac{a_j a_j^{\top}}{\hat{x}_j^{-2}+(\hat{x}_k-\hat{x}_j)^{-2}} + \frac{\tilde{a}_k\tilde{a}_k^{\top}}{\sum_{J(k)}\frac{1}{\hat{x}_j^2+(\hat{x}_k-\hat{x}_j)^2}}\right] \\
+ \; & \textstyle\sum_{\ell\in L}\hat{x}_\ell^2 a_\ell a_\ell^{\top},
\end{aligned}
\tag{2.37}
$$

where

$$
\tilde{a}_k := a_k + \sum_{j\in J(k)} \frac{(\hat{x}_k-\hat{x}_j)^{-2}}{\hat{x}_j^{-2}+(\hat{x}_k-\hat{x}_j)^{-2}} a_j. \tag{2.38}
$$

Notice that $A\hat{\Theta}^{-2}A^{\top}$ in (2.37) is expressed as $\tilde{A}D^2\tilde{A}^{\top}$ with D diagonal, where $\tilde{A}$ differs from A only in that its parent columns a_k are augmented by a linear combination of their children columns. (If each $\hat{x}_j$ converges to 0 or to $\hat{x}_k$, then $\tilde{a}_k$ converges to $a_k + \sum\{a_j : j \in J(k),\quad \hat{x}_j \text{ converges to } \hat{x}_k\}$, and this is exactly the modification of parent columns in the working basis scheme of [17].)

A similar analysis of the simplifications resulting from the special structure of $\hat{\Theta}^2$ appears in Choi and Goldfarb [2], but without noting the relationship to perturbing the parent columns as above.

2.5 Central paths

Consider the barrier problem

$$
(BP) \qquad \begin{aligned}
\min_x \quad c^{\top}x \;&+\; \mu\Psi(x) \\
Ax \;&=\; b, \\
x \;&\in\; \operatorname{int} C,
\end{aligned}
$$

for $\mu > 0$. By our assumption that $F^0(P)$ is nonempty, (BP) has a feasible solution. Since we also assume that (P) has a nonempty bounded set of optimal solutions, along any direction in $\mathcal{N}(A) \cap C$ $c^{\top}x$ increases linearly while Ψ decreases at most logarithmically. Hence standard arguments from convex analysis (e.g., see Rockafellar [14]) imply that (BP) has an optimal solution. Moreover, Ψ is strictly convex in $\mathbb{R}^{J\cup K\cup L}$, while the columns of A_F

are linearly independent, so the optimal solution is unique. We denote it by $x(\mu)$, and define the primal central trajectory to be $\{x(\mu) : \mu > 0\}$.

The KKT conditions are necessary and sufficient for (BP), so $x(\mu)$ together with some $y(\mu), s(\mu)$ satisfies uniquely

$$Ax = b, \tag{2.39}$$

$$A^\top y + s = c, \tag{2.40}$$

$$\mu \nabla \Psi(x) + s = 0. \tag{2.41}$$

Condition (2.41) implies that $s = -\mu \nabla \Psi(x) \in \text{ri } C^*$ by Proposition 2.3 so that $(y, s) \in F^0(D)$, and that $-s/\mu = \nabla \Psi(x) \in \partial \Psi(x)$ where $\partial \Psi$ denotes the subdifferential of the convex function Ψ. We define the convex conjugate of Ψ by

$$\Psi^*(s) := \sup_x \{-s^\top x - \Psi(x)\}.$$

(Note that we use $-s^\top x$ instead of the more usual $s^\top x$ here, in order to get the usual formula when there are only nonnegativity constraints.) Then (2.41) is equivalent [14] to

$$-x \in \partial \Psi^*(s/\mu).$$

This can be made even more symmetric with (2.41). Indeed, since Ψ is a $2p$-logarithmically homogeneous barrier function (recall $2p := 2|J| + |L|$), so is Ψ^* (Nesterov and Nemirovsky [11, 12]), and hence (or directly) $\partial \Psi^*(s/\mu) = \mu \partial \Psi^*(s)$. Thus (2.41) is equivalent to

$$\mu \partial \Psi^*(s) + x \ni 0. \tag{2.42}$$

Thus conditions (2.39) – (2.41) also form the optimality conditions for

$$(BD) \quad \begin{aligned} \max_{y,s} \quad & b^\top y - \mu \Psi^*(s), \\ & A^\top y + s = c, \\ & s \in \text{ri } C^*. \end{aligned}$$

(As in [11, 12], Ψ^* is finite exactly on ri C^*.) Hence $(y(\mu), s(\mu))$ lies on the central trajectory for (D), defined as the set of solutions to (BD) for $\mu > 0$.

We could use Ψ and Ψ^* to construct a primal-dual potential-reduction algorithm for (P) and (D) following the general scheme of Nesterov and Nemirovsky [11, 12]. However, in our case it turns out to be impossible to obtain Ψ^* in closed form, and this precludes the possibility of line searches in the dual space. (This is in interesting contrast to [5], where Ψ^* could be obtained explicitly, but because Ψ and Ψ^* were not logarithmically homogeneous (the problems treated were not "conical"), we could not use the simplification (2.42) and (BD) involved $\mu\Psi^*(s/\mu)$.)

Hence in Section 4, we will confine ourselves to primal algorithms.

2.6 Duality gaps and near-central points

For every $\mu > 0$, we have $x(\mu) \in F^0(P)$ and $(y(\mu), s(\mu)) \in F^0(D)$. Our first result concerns the corresponding duality gap.

Proposition 2.5 *We have $x(\mu)^\top s(\mu) = 2p\mu$. (Recall that $p = |J| + \frac{1}{2}|L|$.)*

Proof. Since $s \in \mathbb{R}^{J\cup K\cup L}$ (we omit the argument μ for ease of notation),

$$x^\top s = (x^{J\cup K\cup L})^\top s = [-(\nabla^2\Psi(x))^{-1}\nabla\Psi(x)]^\top[-\mu\nabla\Psi(x)]$$

(using (2.21) and (2.41)), so

$$x^\top s = \mu\nabla\Psi(x)(\nabla^2\Psi(x))^{-1}\nabla\Psi(x) = \mu(\|\nabla\Psi(x)\|_x^*)^2.$$

The result now follows from Proposition 2.3. □

Hence, if we could follow the path $\{x(\mu)\}$, we could get arbitrarily close to optimal. Unfortunately, we cannot follow the path exactly. Thus we will be interested in pairs $x \in F^0(P)$ and $(y, s) \in F^0(D)$ satisfying (2.39) and (2.40) exactly but (2.41) only approximately. The following result allows us to bound the duality gap of such a pair. (Think of t as s/μ.)

Theorem 2.6 *Suppose $t = -\nabla\Psi(x) + h$, where $x \in$ int C and $h \in \mathbb{R}^{J\cup K\cup L}$ with $\|h\|_x^* \leq \beta < 1$. Then $t \in$ ri C^* and*

$$x^\top t \leq 2p + \beta\sqrt{2p}. \tag{2.43}$$

Proof. From Proposition 2.3 it is immediate that $t \in \mathbb{R}^{J\cup K\cup L}$. To show $t \in \mathrm{ri}\ C^*$ we show that $v^\top t > 0$ for each generator (other than e^f or $-e^f$) of C (see (1.3)), or in other words

$$-v^\top h < -v^\top \nabla\Psi(x)$$

for such v. Since $h \in \mathbb{R}^{J\cup K\cup L}$, this holds if

$$\|v\|_x \|h\|_x^* < -v^\top \nabla\Psi(x),$$

hence if

$$\beta\|v\|_x < -v^\top \nabla\Psi(x). \tag{2.44}$$

For $v = e^\ell$, the left-hand side is βx_ℓ^{-1}, which is less than x_ℓ^{-1}, the right-hand side. Now suppose $v = e^k + \sum_{i\in I(k)} e^i$, where $I(k) \subseteq J(k)$, $k \in K$. Then

$$\begin{aligned}
\|v\|_x^2 &= v^\top \nabla^2\Psi(x) v \\
&= \sum_{j\in J}\left[\frac{v_j^2}{x_j^2} + \frac{(v_{k(j)} - v_j)^2}{(x_{k(j)} - x_j)^2}\right] + \sum_{\ell\in L}\frac{v_\ell^2}{x_\ell^2} \\
&= \sum_{k\in K}\left[\sum_{i\in I(k)} x_i^{-2} + \sum_{j\in J(k)\setminus I(k)} (x_k - x_j)^{-2}\right].
\end{aligned}$$

On the other hand, (2.5) and (2.7) show that

$$-v^\top \nabla\Psi(x) = \sum_{k\in K}\left[\sum_{i\in I(k)} x_i^{-1} + \sum_{j\in J(k)\setminus I(k)} (x_k - x_j)^{-1}\right].$$

Thus (2.44) follows from $\beta < 1$ and the inequality between the 2-norm and the 1-norm of a vector.

To prove (2.43), we have as in Proposition 2.5

$$\begin{aligned}
x^\top t &= \nabla\Psi(x)(\nabla^2\Psi(x))^{-1}(\nabla\Psi(x) - h) \\
&= (\|\nabla\Psi(x)\|_x^*)^2 - \nabla\Psi(x)(\nabla^2\Psi(x))^{-1} h \\
&\le (\|\nabla\Psi(x)\|_x^*)^2 + \|\nabla\Psi(x)\|_x^* \|h\|_x^* \\
&\le 2p + \sqrt{2p}\beta,
\end{aligned}$$

as required. □

Note that the first conclusion of the theorem can be viewed as a dual version of Proposition 2.4.

As mentioned above, t should be thought of as s/μ, where $s = c - A^\top y$ for some y. Then $\mu h = \mu t + \mu \nabla\Psi(x) = c + \mu\nabla\Psi(x) - A^\top y$, and choosing y to make h small is an instance of problem (2.31). Combining Theorems 2.4, 2.5 and 2.6 gives us the following important result, which we call the *approximately-centered theorem.* It allows us to obtain a feasible dual solution from a sufficiently central primal solution.

Theorem 2.7 *Suppose $\hat{x} \in F^0(P)$ is given. Choose $\hat{\mu} > 0$, and let*

$$v := c + \hat{\mu}\nabla\Psi(\hat{x}). \tag{2.45}$$

Let $(\hat{d}, \hat{y})$ be the solution to (2.27) for this v, and hence define

$$\hat{s} := c - A^\top \hat{y}. \tag{2.46}$$

Then $\|\hat{d}\|_{\hat{x}} = \|\hat{s} + \hat{\mu}\nabla\Psi(\hat{x})\|_{\hat{x}}^*$. If

$$\|\hat{d}/\hat{\mu}\|_{\hat{x}} = \|\hat{s}/\hat{\mu} + \nabla\Psi(\hat{x})\|_{\hat{x}}^* \leq \beta, \tag{2.47}$$

where $\beta < 1$, then

(i) $(\hat{y}, \hat{s}) \in F^0(D)$;

(ii) *the duality gap is $\hat{x}^\top \hat{s} \leq \hat{\mu}(2p + \beta\sqrt{2p})$.*

If (2.47) holds, we say $\hat{x}$ is *β-close* to $x(\hat{\mu})$.

Proof. The equality of the norms follows from (2.29). Now define $\hat{t} := \hat{s}/\hat{\mu}$ and $\hat{h} := \nabla\Psi(\hat{x}) + \hat{t}$. Then we find $\hat{h} \in I\!R^{J\cup K\cup L}$ and $\|\hat{h}\|_{\hat{x}}^* \leq \beta < 1$. From Theorem 2.6, $\hat{t} \in \text{ri } C^*$, so $\hat{s} \in \text{ri } C^*$ and $(\hat{y}, \hat{s}) \in F^0(D)$; and $\hat{x}^\top \hat{t} \leq 2p + \beta\sqrt{2p}$, whence (ii) follows. □

To conclude this section, we give as in [5] a sufficient condition for $\hat{x} \in F^0(D)$ to be β-close to $x(\hat{\mu})$. This follows from Theorem 2.5.

Proposition 2.6 *Suppose* $\hat{x} \in F^0(P)$ *and* $\hat{\mu} > 0$ *are given. If there exist* (y, s) *satisfying*

(i) $A^{\mathsf{T}} y + s = c, \quad s \in \mathbb{R}^{J \cup K \cup L}$,

(ii) $\|s/\hat{\mu} + \nabla\Psi(\hat{x})\|^*_{\hat{x}} \leq \beta$,

then $\hat{x}$ *is* β*-close to* $x(\hat{\mu})$. □

3 Algorithms

Here we describe two algorithms for problem (P) based on the barrier function Ψ. The first is a path-following method, using the measure of closeness given in the approximately-centered theorem. Progress sufficient for polynomiality is assured by Proposition 2.3 and Theorem 2.2. The second algorithm is a potential-reduction method, using the barrier function Ψ as part of the potential function. Constant decrease of the latter is guaranteed by Theorems 2.3 and 2.7. We do not deal with initialization of the methods; techniques similar to those in [5] can be employed.

3.1 A path-following method

Here we generate a sequence of points approximating $x(\mu)$ for a geometrically decreasing sequence of values of μ. The idea is similar to that in the algorithms of Renegar [13], Gonzaga [6], and Roos and Vial [15], for instance; see also Gonzaga [8]. Our argument follows [5].

Suppose we have some $\hat{\mu} > 0$ and $\hat{x} \in F^0(P)$ that is β-close to $x(\hat{\mu})$ for some $\beta < 1$. We generate a new value of x by applying Newton's method to (BP), and then shrink $\hat{\mu}$ to $\mu := \alpha\hat{\mu}$ for some $\alpha < 1$. We then want to show that x is again β-close to $x(\mu)$.

With $\hat{x}$, $\hat{\mu}$ and β as above, define

$$v := c + \hat{\mu}\nabla\Psi(\hat{x}) \tag{3.1}$$

as in the approximately-centered Theorem 2.7. Let $(\hat{d}, \hat{y})$ be the solution to (2.27) for this v, so that $\hat{d} = P_{\hat{x}}(v)$, and let

$$\hat{s} := c - A^{\mathsf{T}}\hat{y}. \tag{3.2}$$

From our assumption and the theorem, $\|\hat{d}/\hat{\mu}\|_{\hat{x}} = \|\hat{s}/\hat{\mu} + \nabla\Psi(\hat{x})\|^*_{\hat{x}} \leq \beta$. Now note that v is the gradient of the objective function of (BP) at $\hat{x}$, while its Hessian is $\hat{\mu}\nabla^2\Psi(\hat{x}) = \hat{\mu}\hat{\Theta}^2$. It follows that $-\hat{d}/\hat{\mu}$ is the Newton step for (BP) at $\hat{x}$. We write

$$\bar{d} := -\hat{d}/\hat{\mu}. \tag{3.3}$$

Thus, being β-close to $x(\hat{\mu})$ means precisely that the length of the Newton step for (BP) at $\hat{x}$, measured in the primal norm associated with $\hat{x}$, is at most β. Now let x be the Newton iterate

$$x := \hat{x} + \bar{d}. \tag{3.4}$$

Proposition 3.1 *With the notation above, $x \in F^0(P)$ is β^2-close to $x(\hat{\mu})$.*

Proof. From (2.27), $\bar{d}$ lies in the null space of A, so that $Ax = A\hat{x} = b$. Also, since $\|\bar{d}\|_{\hat{x}} \leq \beta < 1$, Proposition 2.4 guarantees that $x \in \text{int } C$; hence $x \in F^0(P)$ as desired.

Now to show that x is β^2-close to $x(\hat{\mu})$, it is enough by Proposition 2.6 to find (y, s) with

$$A^\top y + s = c, \quad s \in I\!R^{J \cup K \cup L}, \quad \|s/\hat{\mu} + \nabla\Psi(x)\|_x^* \leq \beta^2. \tag{3.5}$$

We prove that (3.5) holds for $(y, s) = (\hat{y}, \hat{s})$. This vector certainly satisfies the first two conditions, and we only need the norm inequality.

From (2.27),

$$\hat{\Theta}^2\hat{d} + A^\top\hat{y} = c + \hat{\mu}\nabla\Psi(\hat{x}),$$

where $\hat{\Theta}^2 := \nabla^2\Psi(\hat{x})$, so we find

$$\begin{aligned} \hat{s}/\hat{\mu} = (c - A^\top\hat{y})/\hat{\mu} &= -\nabla\Psi(\hat{x}) - \hat{\Theta}^2(-\hat{d}/\hat{\mu}) \\ &= -\nabla\Psi(\hat{x}) - \hat{\Theta}^2\bar{d}. \end{aligned}$$

Hence the norm in (3.5) is exactly that on the left-hand side of (2.23) in Theorem 2.2, and thus at most $\|\bar{d}\|_{\hat{x}}^2$. But by hypothesis, this is at most β^2 and the proof is complete. □

Now we show how $\hat{\mu}$ can be decreased:

Proposition 3.2 *Let $\hat{x}, \hat{\mu}, \hat{\beta}$, and x be as above. Let*

$$\alpha := 1 - \frac{\beta - \beta^2}{\beta + \sqrt{2p}} = \frac{\beta^2 + \sqrt{2p}}{\beta + \sqrt{2p}}, \quad \mu = \alpha\hat{\mu}. \tag{3.6}$$

Then x is β-close to $x(\mu)$. (Recall, $p \leq n$ is defined in (2.17).)

Proof. Again, we use Proposition 2.6 with $(y, s) = (\hat{y}, \hat{s})$. We have

$$\begin{aligned} \|\hat{s}/\mu + \nabla\Psi(x)\|_x^* &= \|\hat{s}/(\alpha\hat{\mu}) + \nabla\Psi(x)\|_x^* \\ &= \|\tfrac{1}{\alpha}\left(\tfrac{\hat{s}}{\hat{\mu}} + \nabla\Psi(x)\right) - \left(\tfrac{1}{\alpha} - 1\right)\nabla\Psi(x)\|_x^* \\ &\leq \tfrac{1}{\alpha}\|\tfrac{\hat{s}}{\hat{\mu}} + \nabla\Psi(x)\|_x^* + \left(\tfrac{1}{\alpha} - 1\right)\|\nabla\Psi(x)\|_x^* \\ &\leq \tfrac{1}{\alpha}\beta^2 + \left(\tfrac{1}{\alpha} - 1\right)\sqrt{2p} \\ &= \beta, \end{aligned} \tag{3.7}$$

where the second inequality follows from (3.5) with $s = \hat{s}$ and Proposition 2.3. □

Thus μ can be reduced by a constant factor at each iteration, and Theorem 2.7 translates this into a geometrically decreasing bound on the duality gap at each iteration. Let us use $\beta = \frac{1}{2}$, so $\alpha = 1 - \frac{1}{2+4\sqrt{2p}} < 1 - \frac{1}{8\sqrt{p}}$. Thus repeating the Newton procedure k times, we reduce the bound by the factor at most $(1 - 1/8\sqrt{p})^k$, and hence $O(\sqrt{p})$ iterations reduce it by a constant factor. From this discussion and the propositions above, we have

Theorem 3.1 *Suppose x^0 is β-close to some $x(\mu^0)$, where $\mu^0 > 0$ and $\beta = \frac{1}{2}$. Let $\alpha := 1 - \frac{1}{2+4\sqrt{2p}}$, and define the iterates (x^k, y^k, s^k) as follows. For each $k = 0, 1, \ldots$, let $\hat{\mu} := \mu^k$ and $\hat{x} := x^k$ and define $\hat{d}, \hat{y}$, and $\hat{s}$ as in the approximately-centered Theorem 2.7. Let $(y^k, s^k) := (\hat{y}, \hat{s})$, define $x^{k+1} := x$ from (3.3) and (3.4), and set $\mu^{k+1} := \alpha\mu^k$. Then, for each k,*

(i) $x^k \in F^0(P), (y^k, s^k) \in F^0(D)$;

(ii) $\|s^k/\mu^k + \nabla\Psi(x^k)\|_{x^k}^* \leq \beta$; *and*

(iii) the duality gap is bounded by $(x^k)^\top s^k \leq (\alpha)^k \mu^0 (2p + \beta\sqrt{2p})$.

Moreover, $(x^k)^\top s^k \leq \epsilon$ within $O(p \ln(p\mu^0/\epsilon))$ iterations. □

To conclude the section, we remark that from the proofs of Theorems 2.6 and 2.7 we can deduce that $(x^k)^\mathsf{T} s^k \geq (\alpha)^k \mu^0 (2p - \beta\sqrt{2p})$. Thus the only way to accelerate the algorithm is to decrease μ faster. However, note that, as in (3.7),

$$\begin{aligned} \|\hat{s}/\mu + \nabla\Psi(x)\|_x^* &\geq \left(\frac{1}{\alpha} - 1\right) \|\nabla\Psi(x)\|_x^* - \frac{1}{\alpha}\|\frac{\hat{s}}{\hat{\mu}} + \nabla\Psi(x)\|_x^* \\ &\geq \left(\frac{1}{\alpha} - 1\right)\sqrt{2p} - \frac{1}{\alpha}\beta^2, \end{aligned}$$

so that we can only prove that x is β-close to $x(\mu)$ using $\hat{s}$ if $\alpha \geq 1 - \frac{\beta+\beta^2}{\beta+\sqrt{2p}}$. Therefore, without solving another linear system, the rate of decrease of μ to guarantee path-following is severely restricted.

3.2 A potential-reduction method

Let us suppose that $c^\mathsf{T} x$ is not constant on $F(P)$. (If it were, c would be in the row space of A; then solving (2.27) for any $\hat{x} \in F^0(P)$ with $v = c$ would give $\bar{d} = 0$, confirming that $\hat{x}$ is optimal.) In this case, $c^\mathsf{T} x$ is greater than the optimal value z^* of (P) at any $x \in F^0(P)$, and we can thus define

$$\phi(x, z) := q \ln(c^\mathsf{T} x - z) + \Psi(x), \tag{3.8}$$

where $z \leq z^*$ and q is a positive parameter. Our algorithm is based on reducing this potential function (closely related to that of Karmarkar [10]) as in Gonzaga [7], Ye [20], or Freund [4].

Suppose at the start of an iteration we have $\hat{x} \in F^0(P)$ and $\hat{z} \leq z^*$. Then the gradient of ϕ with respect to x at $(\hat{x}, \hat{z})$ is

$$\nabla_x \phi(\hat{x}, \hat{z}) =: \tilde{v} = \frac{q}{c^\mathsf{T}\hat{x} - \hat{z}} c + \nabla\Psi(\hat{x}). \tag{3.9}$$

(Note the similarity to v in (3.1).) Let $(\tilde{d}, \tilde{y})$ solve (2.27) for $v = \tilde{v}$, so that $\tilde{d} = P_{\hat{x}}(\tilde{v})$. Then we show, as in [5], that the potential function can be reduced by a constant by taking a step in the direction $-\tilde{d}$, as long as $\|\tilde{d}\|_{\hat{x}}$ is sufficiently large.

Proposition 3.3 *Suppose $\hat{x}, \hat{z}, \tilde{v}, \tilde{d}$ and $\tilde{y}$ are as above. Then, if $\|\tilde{d}\|_{\hat{x}} \geq \frac{4}{5}$ and $\gamma \in (0,1)$,*

$$x(\gamma) := \hat{x} - \gamma \tilde{d}/\|\tilde{d}\|_{\hat{x}} \in F^0(P)$$

and

$$\phi(x(\gamma), \hat{z}) \leq \phi(\hat{x}, \hat{z}) - \frac{4}{5}\gamma + \frac{\gamma^2}{2(1-\gamma)}. \tag{3.10}$$

In particular, $x(\frac{2}{5}) \in F^0(P)$ and

$$\phi(x(\frac{2}{5}, \hat{z}) \leq \phi(\hat{x}, \hat{z}) - \frac{1}{6}. \tag{3.11}$$

Proof. As in the proof of Proposition 3.1, $\tilde{d}$ lies in the null space of A, so $Ax(\gamma) = A\hat{x} = b$, and $\|x(\gamma) - \hat{x}\|_{\hat{x}} = \gamma < 1$, so $x(\gamma) \in F^0(P)$ by Proposition 2.4. Also, we have

$$\begin{aligned}
\phi(x(\gamma), \hat{z}) - \phi(\hat{x}, \hat{z}) &= q \ln\left(1 - \frac{\gamma c^\top \tilde{d}}{\|\tilde{d}\|_{\hat{x}}(c^\top \hat{x} - \hat{z})}\right) + \Psi(x(\gamma)) - \Psi(\hat{x}) \\
&\leq -\frac{\gamma q c^\top \tilde{d}}{\|\tilde{d}\|_{\hat{x}}(c^\top \hat{x} - z)} + \Psi(\hat{x} - \gamma \tilde{d}/\|\tilde{d}\|_{\hat{x}}) - \Psi(\hat{x}) \\
&\quad \text{(from the concavity of the logarithm function)} \\
&\leq -\frac{\gamma q c^\top \tilde{d}}{\|\tilde{d}\|_{\hat{x}}(c^\top \hat{x} - z)} - \frac{\gamma \nabla\Psi(\hat{x})^\top \tilde{d}}{\|\tilde{d}\|_{\hat{x}}} + \frac{\gamma^2}{2(1-\gamma)} \\
&\quad \text{(from Theorem 2.3)} \\
&= -\frac{\gamma}{\|\tilde{d}\|_{\hat{x}}}\left(\frac{q}{c^\top \hat{x} - \hat{z}} c + \nabla\Psi(\hat{x})\right)^\top \tilde{d} + \frac{\gamma^2}{2(1-\gamma)} \\
&= -\frac{\gamma}{\|\tilde{d}\|_{\hat{x}}} \tilde{v}^\top \tilde{d} + \frac{\gamma^2}{2(1-\gamma)} = -\gamma\|\tilde{d}\|_{\hat{x}} + \frac{\gamma^2}{2(1-\gamma)} \\
&\quad \text{(from (2.28) in Theorem 2.4)} \\
&\leq -\frac{4}{5}\gamma + \frac{\gamma^2}{2(1-\gamma)}.
\end{aligned}$$

This proves (3.10), and (3.11) follows by substituting $\gamma = \frac{2}{5}$. □

Suppose now $\|\tilde{d}\|_{\hat{x}} < \frac{4}{5}$. Then let

$$\begin{aligned} \hat{\mu} &:= (c^\top \hat{x} - \hat{z})/q, \\ \hat{v} &:= \hat{\mu}\tilde{v} = c + \hat{\mu}\nabla\Psi(\hat{x}), \\ \hat{y} &:= \hat{\mu}\tilde{y}, \quad \text{and} \\ \hat{s} &:= c - A^\top \hat{y}. \end{aligned} \tag{3.12}$$

Let $\hat{d} := P_{\hat{x}}(\hat{v}) = \hat{\mu} P_{\hat{x}}(\tilde{v}) = \hat{\mu}\tilde{d}$, and note that $\|\hat{d}/\hat{\mu}\|_{\hat{x}} = \|\tilde{d}\|_{\hat{x}} < \frac{4}{5}$, so $\hat{x}$ is β-close to $x(\hat{\mu})$ for $\beta = \frac{4}{5}$. We can therefore apply the approximately-centered Theorem 2.7 to obtain

$$(\hat{y}, \hat{s}) \in F^0(D) \quad \text{and} \tag{3.13}$$

$$\hat{x}^\top \hat{s} \leq \hat{\mu}(2p + \frac{4}{5}\sqrt{2p}). \tag{3.14}$$

Hence

$$z := b^\top \hat{y} = c^\top \hat{x} - \hat{x}^\top \hat{s} \tag{3.15}$$

is a valid lower bound on the optimal value z^* of (P) and (D). We show as in [5] that this update provides a sufficient decrease in ϕ as long as q is sufficiently large.

Proposition 3.4 *Suppose $q \geq 2p + \sqrt{2p}$. If $\hat{x} \in F^0(P)$, and, with the notation above, $\|\tilde{d}\|_{\hat{x}} < \frac{4}{5}$, then $(\hat{y}, \hat{s}) \in F^0(D)$, $z \leq z^*$ and*

$$\phi(\hat{x}, z) \leq \phi(\hat{x}, \hat{z}) - \frac{1}{6}. \tag{3.16}$$

Proof. From the discussion above, we only need to establish (3.16). From (3.12) - (3.15),

$$c^\top \hat{x} - z = \hat{x}^\top \hat{s} \leq \hat{\mu}(2p + \frac{4}{5}\sqrt{2p}) = \frac{2p + \frac{4}{5}\sqrt{2p}}{q}(c^\top \hat{x} - \hat{z}).$$

Hence,

$$
\begin{aligned}
\phi(\hat{x}, z) - \phi(\hat{x}, \hat{z}) &= q \ln \left(\frac{c^{\mathsf{T}} \hat{x} - z}{c^{\mathsf{T}} \hat{x} - \hat{z}} \right) \\
&\leq q \ln \left(\frac{2p + \frac{4}{5}\sqrt{2p}}{q} \right) \\
&\leq q \ln \left(1 - \frac{\frac{1}{5}\sqrt{2p}}{q} \right) \\
&\quad (\text{since} \quad q \geq 2p + \sqrt{2p}) \\
&\leq -\frac{1}{5}\sqrt{2p} \leq -\frac{1}{6}. \quad \Box
\end{aligned}
$$

These two results naturally suggest an algorithm for which we obtain the following convergence result.

Theorem 3.2 *Let $x^0 \in F^0(P)$ and $z^0 \leq z^*$ be given, and choose $q \geq 2p + \sqrt{2p}$. Suppose $\{x^k\} \subseteq I\!R^n$ and $\{z^k\} \subseteq I\!R$ are obtained as follows. For each k, let $\hat{x} := x^k$ and $\hat{z} := z^k$ and define $\tilde{v}$ by (3.9). Compute $\tilde{d} := P_{\hat{x}}(\tilde{v})$. If $\|\tilde{d}\|_{\hat{x}} \geq \frac{4}{5}$, set $x^{k+1} := x(\frac{2}{5})$ as in Proposition 3.3, and let $z^{k+1} := z^k$; else, set $x^{k+1} := x^k$ and update $z^{k+1} := z$ as in Proposition 3.4. Then, for each k,*

(i) $x^k \in F^0(P)$ and $z^k \leq z^*$;

(ii) $\phi(x^k, z^k) \leq \phi(x^0, z^0) - k/6$; and

(iii) for some constant C, $c^{\mathsf{T}} x^k - z^k \leq C \exp(-k/6q)$. $\Box$

The last part follows from the same analysis used in the proofs of Theorem 4.1 and Lemma 4.2 of [1].

If we choose $q = O(p)$, this yields a bound of $O(p)$ iterations to reduce the bound on the optimality gap by a constant factor. This is worse than the bound for the path-following method; but the present algorithm allows considerably more flexibility which might improve its practical performance. For instance, as in Freund [4] and Gonzaga [7], we can try to improve the lower bound as in Proposition 3.4 at every iteration, performing a line search on $\mu = \hat{\mu}$ (of which $\hat{y}$ and $\hat{s}$ are linear functions) to obtain the best bound, and we can similarly perform a line search on γ to approximately minimize $\phi(x(\gamma), \hat{z})$, even allowing $\gamma > 1$ as long as $x(\gamma)$ remains feasible.

References

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

[2] I. C. Choi and D. Goldfarb. Exploiting special structure in a primal–dual path–following algorithm. Technical Report, Dept. of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, USA, 1991. To appear in *Mathematical Programming.*

[3] R. Fourer and S. Mehrotra. Performance of an augmented system approach for solving least–squares problems in an interior–point method for linear programming. *COAL Newsletter*, 19:26–31, August 1991.

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

[5] R. M. Freund and M. J. Todd. Barrier functions and interior–point algorithms for linear programming with zero–, one–, or two–sided bounds on the variables. Technical Report 1016, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853–3801, USA, 1992.

[6] C. C. Gonzaga. An algorithm for solving linear programming problems in $O(n^3L)$ operations. In N. Megiddo, editor, *Progress in Mathematical Programming : Interior Point and Related Methods*, pages 1–28. Springer Verlag, New York, 1989.

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

[8] C. C. Gonzaga. Large steps path–following methods for linear programming, Part I : Barrier function method. *SIAM Journal on Optimization*, 1:268–279, 1991.

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

[10] N. K. Karmarkar. A new polynomial–time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[11] Y. E. Nesterov and A. S. Nemirovsky. 'Conic' formulation of a convex programming problem and duality. Manuscript, Central Economic and Mathematical Institute, USSR Academy of Science, Moscow, USSR, 1991.

[12] Y. E. Nesterov and A. S. Nemirovsky. *Interior Point Polynomial Methods in Convex Programming : Theory and Algorithms.* To appear in SIAM Publications. SIAM, Philadelphia, USA, 1993.

[13] J. Renegar. A polynomial–time algorithm, based on Newton's method, for linear programming. *Mathematical Programming*, 40:59–93, 1988.

[14] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, 1970.

[15] C. Roos and J. P. Vial. A polynomial method of approximate centers for linear programming. *Mathematical Programming*, 54:295–305, 1992.

[16] L. Schrage. Implicit representation of variable upper bounds in linear programming. *Mathematical Programming Study*, 4:118–132, 1975.

[17] M. J. Todd. An implementation of the simplex method for linear programming problems with variable upper bounds. *Mathematical Programming*, 23:34–49, 1982.

[18] M. J. Todd. Exploiting special structure in Karmarkar's algorithm for linear programming. *Mathematical Programming*, 41:97–113, 1988.

[19] R. J. Vanderbei and T. J. Carpenter. Symmetric indefinite systems for interior point methods. Technical Report SOR 91–7, Dept. of Civil Engineering and Operations Research, Princeton University, Princeton, NJ 08544, USA, 1991. Revised June 1992.

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