Barrier Functions and Interior-Point Algorithms for Linear Programming with Zero-, One-, or Two-Sided Bounds on the Variables

Robert M. Freund

and Michael J. Todd

WP# 3454-92-MSA July, 1992

Barrier Functions and Interior-Point Algorithms for Linear Programming with Zero-, One-, or Two-Sided Bounds on the Variables *

Robert M. Freund
Sloan School of Management, MIT
50 Memorial Drive
Cambridge, Mass. 02139

Michael J. Todd
School of Operations Research
and Industrial Engineering
Engineering and Theory Center
Cornell University
Ithaca, NY 14853

July, 1992

*Research supported in part by NSF, AFOSR, and ONR through NSF Grant DMS-8920550, and in part by the MIT-NTU Collaboration Research Fund. This work was initiated while the first author was visiting Cornell University, with support from the Center for Applied Mathematics and from DMS-8920550.

Abstract

This study examines two different barrier functions and their use in both path-following and potential-reduction interior-point algorithms for solving a linear program of the form: minimize $c^T x$ subject to $Ax = b$ and $\ell \leq x \leq u$, where components of ℓ and u can be nonfinite, so the variables x_j can have $0-$, $1-$,or 2-sided bounds, $j = 1, ..., n$. The barrier functions that we study include an extension of the standard logarithmic barrier function and an extension of a barrier function introduced by Nesterov. In the case when both ℓ and u have all of their components finite, these barrier functions are

$$\Psi(x) = \sum_j \{-\ln(u_j - x_j) - \ln(x_j - \ell_j)\}$$

and

$$\Psi(x) = \sum_j \{-\ln(\min\{u_j - x_j, x_j - \ell_j\}) + \min\{u_j - x_j, x_j - \ell_j\}/((u_j - \ell_j)/2)\}.$$

Each of these barrier functions gives rise to suitable primal and dual metrics that are used to develop both path-following and potential-reduction interior-point algorithms for solving such linear programming problems. The resulting complexity bounds on the algorithms depend only on the number of bounded variables, rather than on the number of finite inequalities in the system $\ell \leq x \leq u$, in contrast to the standard complexity bounds for interior-point algorithms. These enhanced complexity bounds stem directly from the choice of a "natural" metric induced by the barrier function. This study also demonstrates the interconnection between the notion of self-concordance (introduced by Nesterov and Nemirovsky) and properties of the two barrier functions that drive the results contained herein.

Key words: linear programming, barrier functions, interior-point algorithms, complexity theory, self-concordance.

Running Header: Barrier Functions for Linear Programming

1 Introduction

This study examines two different barrier functions and their use in both path-following and potential-reduction interior-point algorithms for solving a linear program of the form:

$$(P) \qquad \begin{array}{llcl} \min\limits_{x} & c^T x & & \\ \text{s.t.} & Ax & = & b, \\ \ell & \leq x & \leq & u, \end{array}$$

where the components of ℓ and u can be infinite, so that the variables x_j can have $0-, 1-$, or 2-sided bounds, $j = 1, ..., n$. Variable with no bounds are free variables, and the others are bounded variables.

This study is motivated in part by the desire to redress a weakness in much of the research in interior-point algorithms (both theoretical and computational) that assumes that all variables have one-sided bounds. To satisfy this assumption, a free variable x_j can be eliminated through row operations on A, but this is neither computationally convenient nor natural. We can deal with a 2-sided bounded variable x_j by adding an additional variable x'_j and a new equation $x_j + x'_j = u_j$. In this way both x_j and x'_j become one-sided bounded variables with finite lower bounds of ℓ_j and 0, respectively. This procedure seems also to be inconvenient and unnatural. Furthermore, by converting a two-sided bounded variable to two one-sided bounded variables, attention is drawn away from the inherent relation between the slacks on the two bounds, namely that they must sum to the positive constant $u_j - \ell_j$, and that the two bounds cannot therefore be simultaneously binding at any feasible solution.

The standard logarithmic barrier function for a two-sided bounded variable is

$$-\ln(u_j - x_j) - \ln(x_j - \ell_j). \tag{1.1}$$

This barrier function repels x_j from ℓ_j and u_j , but takes no advantage of the fact that x_j cannot simultaneously approach ℓ_j and u_j (unless $u_j = \ell_j$, which can easily be assumed away). It seems on intuitive grounds that a logical alternative barrier function for a two-sided bounded variable is $-\ln(\min\{u_j - x_j, x_j - \ell_j\})$, but this barrier function is not differentiable at $x_j = (u_j + \ell_j)/2$. Therefore we also consider herein a barrier function of the basic format

$$-\ln(\min\{u_j - x_j, x_j - \ell_j\}) + \min\{u_j - x_j, x_j - \ell_j\}/((u_j - \ell_j)/2), \tag{1.2}$$

which has an additional piecewise-linear term that causes the function to be twice differentiable. (The twice-differentiability of the barrier function is important for many reasons, including constructing good local approximations and using Newton's method, and is intimately related to the self-concordance notion of Nesterov and Nemirovsky [14], as is discussed in Section 2 of this paper. The above barrier function is a mild extension of a barrier function introduced in Nesterov [13].)

By providing a unified and consistent framework for studying barrier functions for the very general linear program (P) with possibly nonfinite values of ℓ and u, this study attempts to redress the weaknesses mentioned above: the awkwardness of assuming no free variables, and the unnaturalness of replacing a two-sided bounded variable by a pair of one-sided bounded variables. However, there is another added benefit from this study as well, related to the analysis of the computational complexity of interior-point algorithms for linear programming, that derives from the appropriate choice of metrics associated with the two barrier functions of the forms (1.1) and (1.2). By choosing suitable primal and dual metrics based on such barrier functions, we derive complexity bounds for path-following algorithms and potential-reduction algorithms that depend on the number of bounded variables rather than on the number of inequalities. (This runs counter to previous research in interior-point algorithms, in which any increase in the number of inequalities of the linear program necessitated an increase in the complexity of the algorithm.) The derivation of these complexity improvements stems from the choice of the metric used to measure displacements in both primal and dual space, and our results indicate that there is indeed a "natural" pair of primal-dual metrics derived from the choice of the barrier function. A similar connection between the barrier function and the choice of metric is also discussed in the context of the Riemannian geometry of linear programming problems in Karmarkar [8], and in particular in the work of Nesterov and Nemirovsky [14].

Another motivation of our study is to examine exactly which properties of an interior-point algorithm contribute to polynomial time bounds. Many methods use scaling by the components of the current iterate at each iteration; our development, for a slightly more general problem, shows that this scaling corresponds to using a primal metric derived from the Hessian of a barrier function. In addition, our problem seems to be about the simplest for which two different barrier functions can be used, while maintaining the same polynomial time complexity. Finally, our treatment is intended to illustrate in a fairly simple setting the ideas of the very general theory developed by Nesterov and Nemirovsky in [14].

The paper is organized as follows: In Section 2 we describe two barrier functions for the linear programming problem (P) with possibly nonfinite values of components of ℓ and u. The barrier functions are essentially the same as (1.1) and (1.2), with modifications to conveniently handle infinite values of ℓ_j and u_j. Basic differential properties of these barrier functions are developed as well. These barriers are used to define primal and dual metrics on displacements, and to define primal and dual projections. Central trajectories based on these barriers are also defined and duality gaps and measures of centrality are studied next. Finally, approximations to barrier functions and their gradients are developed. These approximations are derived using notions directly related to the self-concordance of the barrier functions, and the relation of self-concordance to the results is discussed.

Section 3 contains a description of a primal path-following algorithm (based on the use of Newton's method) for solving the linear programming problem (P). This algorithm obtains a fixed reduction in the duality gap in $O(\sqrt{p})$ iterations, where p is the number of bounded variables (as distinct from the the number of inequalities, which will be possibly larger).

Section 4 contains a description of a primal potential-reduction algorithm for solving the linear programming problem (P). This algorithm obtains a fixed reduction in the duality gap in $O(p)$ iterations.

Section 5 contains concluding remarks, open questions, and possible directions for future research.

Notation. For the most part, the notation is standard. If v is a vector in R^n, V denotes the diagonal matrix with diagonal entries corresponding to the components of v. The vector of ones, $(1, 1, ..., 1)^T$, is denoted by e, where the dimension is dictated by context. Let $\|\cdot\|$ and $\|\cdot\|_1$ denote the Euclidean and the ℓ_1- norm, respectively. In context, parentheses are used to denote indexing of components of a vector, e.g., $x = (\min\{v_j, 1 - v_j\})$ connotes that x is the vector whose jth component is $\min\{v_j, 1 - v_j\}, j = 1, ..., n$.

Finally, we point out that this study is intended to be self-contained, and does not rely on previous knowledge of results on interior-point methods in linear programming.

2 Barrier functions for interval bounds and their properties

In this section, we describe two barrier functions for the bounded linear programming problem. These barriers are a crucial ingredient in the algorithms we describe in the following sections. Among their roles are:

a) being part of primal potential functions;

b) defining metrics both on primal displacements and dual slack vectors;

c) determining central trajectories; and

d) permitting the derivation of dual slack vectors from primal solutions.

We will derive some key properties of our barrier functions that facilitate their use for the purposes above.

2.1 Problems, assumptions and duality

We are interested in the problem

$$(P) \qquad \begin{array}{rcl} \min_x & c^T x & \\ \text{s.t.} & Ax & = & b, \\ \ell \leq & x & \leq & u, \end{array}$$

where A is $m \times n$ and the vectors x, b, c, ℓ and u have appropriate dimensions. Here the components ℓ_j of ℓ can be $-\infty$ or finite, while the components of u_j of u can be $+\infty$ or finite. We insist that $\ell < u$. (If $\ell_j > u_j$ for some j, (P) is clearly infeasible; and if $\ell_j = u_j$, x_j is fixed and we can substitute for it and get a lower-dimensional problem.) We say x_j has 0-sided, 1-sided, or 2-sided bounds according as the number of finite elements of $\{\ell_j, u_j\}$ is 0, 1, or 2, respectively. Variables with 0-sided bounds are called free; all remaining variables are called bounded. Let p denote the number of bounded variables, and note $p \leq n$.

Let $F(P)$ denote the feasible region of (P) and $F^0(P)$ the set of strictly feasible points:

$$F^0(P) := \{x \in R^n : Ax = b, \ell < x < u\}.$$

We assume that $F^0(P)$ is nonempty, and that we know some $x^0 \in F^0(P)$. We suppose the set of optimal solutions of (P) is nonempty and bounded. We also assume without loss of generality that A has rank m.

Let $\tilde{A}$ consist of the columns of A corresponding to free variables. If for some nonzero $\tilde{d}$ we have $\tilde{A}\tilde{d} = 0$ and $\tilde{c}^T\tilde{d} \neq 0$, where $\tilde{c}$ is the corresponding subvector of c, then by moving the free variables in the direction $\pm\tilde{d}$ from any feasible solution we see that (P) is unbounded. If $\tilde{A}\tilde{d} = 0$ and $\tilde{c}^T\tilde{d} = 0$ for some nonzero $\tilde{d}$, then the set of optimal solutions cannot be bounded for the same reason. Hence our assumptions imply that $\tilde{A}$ has full column rank.

The standard dual of (P) is

$$\begin{array}{lccccccc} \max_{y,s',s''} & b^T y & + & \ell^T s' & - & u^T s'' & & \\ \text{s.t.} & A^T y & + & s' & - & s'' & = & c, \\ & s' & \geq & 0, & & s'' & \geq & 0. \end{array}$$

Since we are assuming that $\ell < u$, it is sufficient to consider solutions where for each j, s'_j or s''_j is zero. Writing $s = s' - s''$, we can alternatively express the dual in the form

$$(D) \qquad \begin{array}{lccccccc} \max_{y,s} & b^T y & + & \ell^T s^+ & - & u^T s^- & & \\ \text{s.t.} & A^T y & + & & & s & = & c, \end{array}$$

where $s^+ := (\max\{0, s_j\})$ and $s^- := (\max\{0, -s_j\})$. We have been implicitly assuming that ℓ and u are finite, but (D) is a valid dual problem even when some components of ℓ and/or u are infinite. If any $\ell_j = -\infty$, then $(s^+)_j$ must be zero so $s_j \leq 0$; if any $u_j = +\infty$, then $(s^-)_j$ must be zero so $s_j \geq 0$. We follow the usual understanding in extended-real-valued arithmetic that $0 \times (+\infty) = 0$ and $0 \times (-\infty) = 0$. In particular, $s_j = 0$ if x_j is free. With these conventions, (D) satisfies the usual duality relationships with (P).

It will be convenient to refer to the situation where $\ell = 0$ and each component of u is $+\infty$ as the standard case. Then $s \geq 0$ and $\ell^T s^+ - u^T s^- = 0$, so we recover the usual dual problem.

Let $F(D)$ denote the feasible region of the dual problem:

$$F(D) := \{(y, s) : A^T y + s = c, \quad \ell^T s^+ - u^T s^- > -\infty\}.$$

We also let

$$\begin{aligned} F^0(D) := \{(y, s) \in F(D) : s_j > 0 \text{ if } -\infty < \ell_j < u_j = +\infty, \\ s_j < 0 \text{ if } -\infty = \ell_j < u_j < +\infty\}. \end{aligned}$$

For any $x \in F(P), (y, s) \in F(D)$, the duality gap is

$$\begin{aligned} c^T x - b^T y - \ell^T s^+ + u^T s^- &= c^T x - (Ax)^T y - \ell^T s^+ + u^T s^- \\ &= x^T s - \ell^T s^+ + u^T s^- \\ &= (x - \ell)^T s^+ + (u - x)^T s^-. \end{aligned} \tag{2.1}$$

This shows that the duality gap is zero if and only if complementary slackness holds.

2.2 Two barrier functions

Let $X := \{x \in R^n : \ell \leq x \leq u\}$. By our assumptions, $\text{int}\, X = \{x \in R^n : \ell < x < u\}$ is nonempty.

The typical logarithmic barrier function associated with the linear program (P) associates a logarithmic penalty with each inequality slack, and it is of the form:

$$\bar{B}(x) := -\sum_{j=1}^{n} \ln(x_j - \ell_j) - \sum_{j=1}^{n} \ln(u_j - x_j).$$

When ℓ and u are both finite, then $\bar{B}(x)$ is well-defined. But if for some j, $\ell_j = -\infty$ or $u_j = +\infty$, then $\bar{B}(x)$ is not well-defined. In order to incorporate in a unifying framework the cases when some or all of the components of ℓ and/or u are not finite, we choose an arbitrary reference value of $x = \breve{x}$ with $\ell < \breve{x} < u$, and define

$$\bar{\Psi}(x) := \bar{\Psi}(x; \ell, u, \breve{x}) := -\sum_{j=1}^{n} \ln\left(\frac{x_j - \ell_j}{\breve{x}_j - \ell_j}\right) - \sum_{j=1}^{n} \ln\left(\frac{u_j - x_j}{u_j - \breve{x}_j}\right).$$

(In the standard case, we choose $\breve{x} = e$.) Then note that $\bar{\Psi}(x)$ is well-defined for all $x \in \operatorname{int} X$. We will rewrite $\bar{\Psi}(x)$ as

$$\bar{\Psi}(x) := \sum_j \bar{\psi}_j(x_j) := \sum_j \bar{\psi}(x_j; \ell_j, u_j, \breve{x}_j), \tag{2.2a}$$

where $\bar{\psi}$ is of the form:

$$\bar{\psi}(\xi) := \bar{\psi}(\xi; \lambda, v, \breve{\xi}) := -\ln(\frac{\xi - \lambda}{\breve{\xi} - \lambda}) - \ln(\frac{v - \xi}{v - \breve{\xi}}) \tag{2.3}$$

and $\lambda < \breve{\xi} < v$. When $\lambda = -\infty$ and/or $v = +\infty$, then $\bar{\psi}$ is defined to be the limit of the expression (2.3) as $\lambda \to -\infty$ and/or $v \to +\infty$. Note that when λ or v is infinite, the effect of the denominator in (2.3) is to erase the appropriate logarithmic barrier term. In particular, in the standard case, $\breve{\xi} = 1$ and we recover the usual logarithmic barrier function $-\ln(\xi)$. When λ and v are finite, the effect of the denominator is to add a scalar constant to the "normal" barrier function $-\ln(\xi - \lambda) - \ln(v - \xi)$.

In the case when both λ and v are finite, the purpose of the barrier function $\bar{\psi}(\xi)$ is to repel ξ from the boundary values of λ or v. Since $\lambda < v$, it is impossible for ξ to simultaneously approach λ and v. It therefore seems more efficient to consider a barrier of the form

$$-\ln(\min\{\xi - \lambda, v - \xi\}),$$

which only penalizes one of the slacks of the interval inequalities $\lambda < \xi < v$. This yields the barrier function

$$\tilde{B}(x) := -\sum_{j=1}^{n} \ln(\min\{x_j - \ell_j, u_j - x_j\}).$$

However, $\tilde{B}(x)$ is not differentiable at $x = (u - \ell)/2$ in the case when u and ℓ are finite. In order to remedy this, we augment $\tilde{B}(x)$ with an extra piecewise-linear term which will cause $\tilde{B}(x)$ to be twice continuously differentiable, as follows:

$$\tilde{B}(x) := \sum_{j=1}^{n} \left(\frac{\min\{x_j - \ell_j, u_j - x_j\}}{(u_j - \ell_j)/2} - \ln(\min\{x_j - \ell_j, u_j - x_j\}) \right).$$

(The differentiability of $\tilde{B}(x)$ will be proven in Lemma 2.1, to follow). However, just as in the function $\bar{B}(x)$, $\tilde{B}(x)$ is not well-defined if $\ell_j = -\infty$ or $u_j = +\infty$ for some j. Thus, as with $\bar{B}(x)$, we define:

$$\tilde{\Psi}(x) := \tilde{\Psi}(x; \ell, u, \check{x}) := \sum_{j=1}^{n} \left(\frac{\min\{x_j - \ell_j, u_j - x_j\}}{(u_j - \ell_j)/2} - \ln\left(\frac{\min\{x_j - \ell_j, u_j - x_j\}}{\min\{\check{x}_j - \ell_j, u_j - \check{x}_j\}} \right) \right).$$

We will write

$$\tilde{\Psi}(x) := \sum_j \tilde{\psi}_j(x_j) := \sum_j \tilde{\psi}(x_j; \ell_j, u_j, \check{x}_j). \tag{2.2b}$$

where $\tilde{\psi}$ is of the form

$$\tilde{\psi}(\xi) := \tilde{\psi}(\xi; \lambda, \upsilon, \check{\xi}) := \frac{\min\{\xi - \lambda, \upsilon - \xi\}}{(\upsilon - \lambda)/2} - \ln\left(\frac{\min\{\xi - \lambda, \upsilon - \xi\}}{\min\{\check{\xi} - \lambda, \upsilon - \check{\xi}\}} \right). \tag{2.4}$$

When $\lambda = -\infty$ and/or $\upsilon = +\infty$, then $\tilde{\psi}$ is defined to be the limit of the expression (2.4) as $\lambda = -\tau$ and/or $\upsilon = +\tau$ and $\tau \to +\infty$. (As in the case of the function $\bar{\psi}$, the role of $\check{\xi}$ in $\tilde{\psi}$ is to ensure that limits exist and different values of $\check{\xi}$ affect $\bar{\psi}$ and $\tilde{\psi}$ only through additive constants.)

Note then

$$\bar{\psi}(\xi; \lambda, +\infty, \check{\xi}) = \tilde{\psi}(\xi; \lambda, +\infty, \check{\xi}) = -\ln\left(\frac{\xi - \lambda}{\check{\xi} - \lambda} \right) \tag{2.5}$$

and

$$\bar{\psi}(\xi; -\infty, \upsilon, \check{\xi}) = \tilde{\psi}(\xi; -\infty, \upsilon, \check{\xi}) = -\ln\left(\frac{\upsilon - \xi}{\upsilon - \check{\xi}} \right). \tag{2.6}$$

(In the standard case, $\lambda = 0$, $\upsilon = +\infty$, and $\check{\xi} = 1$, and we obtain the standard logarithmic barrier function $-\ln \xi$ from (2.6).) If $\lambda = -\infty$ and $\upsilon = +\infty$, then

$$\bar{\psi}(\xi; -\infty, +\infty, \check{\xi}) = 0 \quad \text{and} \quad \tilde{\psi}(\xi; -\infty, +\infty, \check{\xi}) = 1. \tag{2.7}$$

For simplicity of notation we frequently omit the parameters λ, υ, and $\check{\xi}$. We will also write ψ for $\bar{\psi}$ or $\tilde{\psi}$ and Ψ for $\bar{\Psi}$ or $\tilde{\Psi}$, so that (2.2a) and (2.2b) become

$$\Psi(x) := \sum_j \psi_j(x_j) := \sum_j \psi(x_j; \ell_j, u_j, \check{x}_j), \tag{2.2c}$$

where ψ is either $\bar{\psi}$ or $\tilde{\psi}$. We will also use the notation

$$\rho = (\upsilon + \lambda)/2, \quad \nu = (\upsilon - \lambda)/2 \tag{2.8}$$

when υ and λ are finite. One can think of ρ as the midpoint and ν as the radius of the interval $[\lambda, \upsilon]$. Then $\min\{\xi - \lambda, \upsilon - \xi\}$ can also be written as $\nu - |\xi - \rho|$.

The barrier function $\bar{\psi}$ can be thought of as the sum of two standard logarithmic barriers, one for the slack variable $\xi - \lambda$ and the other for the slack variable $\upsilon - \xi$. We prefer to consider it as a single barrier associated with the variable ξ. The barrier function $\tilde{\psi}$ was introduced by Nesterov [13] for the case $\lambda = -1, \upsilon = +1, \check{\xi} = 0$, so that $\rho = 0$ and $\nu = 1$. In that case,

$$\tilde{\psi}(\xi) = 1 - |\xi| - \ln(1 - |\xi|).$$

We have

Lemma 2.1 *With obvious limits if $\lambda = -\infty$ or $\upsilon = +\infty$ or both,*

$$\bar{\psi}'(\xi) = -\frac{1}{\xi - \lambda} + \frac{1}{\upsilon - \xi}, \tag{2.9}$$

$$\bar{\psi}''(\xi) = \frac{1}{(\xi - \lambda)^2} + \frac{1}{(\upsilon - \xi)^2}, \tag{2.10}$$

$$\tilde{\psi}'(\xi) = \begin{cases} \frac{\xi - \rho}{\nu \min\{\xi - \lambda, \upsilon - \xi\}} & \text{if } \lambda \text{ and/or } \upsilon \text{ are finite} \\ 0 & \text{if } \lambda = -\infty, \upsilon = +\infty, \end{cases} \tag{2.11}$$

and

$$\tilde{\psi}''(\xi) = \frac{1}{(\min\{\xi - \lambda, \upsilon - \xi\})^2}. \tag{2.12}$$

Proof: If λ and υ are both finite, (2.9) and (2.10) follow immediately from the definition (2.3) of $\bar{\psi}$. If either or both are infinite, these results can be checked directly from (2.5)-(2.7).

Similarly, (2.11) and (2.12) are immediate from (2.5)-(2.7) if either or both of λ and υ are infinite. Suppose now both are finite. Then if $\check{\xi} < \rho$, then $\tilde{\psi}$

coincides with $(\xi-\lambda)/\nu - \ln(\xi-\lambda) + \ln(\min\{\breve{\xi}-\lambda, \upsilon-\breve{\xi}\})$ for ξ close to $\bar{\xi}$, and so

$$\begin{aligned}
\tilde{\psi}'(\bar{\xi}) &= 1/\nu - 1/(\bar{\xi}-\lambda) = (\bar{\xi}-\lambda-\nu)/(\nu(\bar{\xi}-\lambda)) \\
&= (\bar{\xi}-\rho)/(\nu\min\{\bar{\xi}-\lambda, \upsilon-\bar{\xi}\}) \qquad \text{and} \\
\tilde{\psi}''(\bar{\xi}) &= 1/(\bar{\xi}-\lambda)^2 = 1/(\min\{\bar{\xi}-\lambda, \upsilon-\bar{\xi}\})^2.
\end{aligned}$$

A similar argument holds for $\bar{\xi} > \rho$. Since these formulae are continuous at $\bar{\xi} = \rho$, it follows that they are valid for all $\bar{\xi}$. □

(Note in (2.9)-(2.12) that these derivatives are independent of the "reference" value $\breve{\xi}$. Also note interestingly that the third derivative of $\tilde{\psi}(\xi)$ is <u>not</u> well-defined at $\xi = \rho$.)

From their definitions and lemma 2.1, we see that $\bar{\Psi}$ and $\tilde{\Psi}$ are convex, and converge to $+\infty$ on a sequence of points in int X converging to a point of $X\backslash(\text{int } X)$. In particular, we have

$$\nabla^2\bar{\Psi}(x) = \bar{\Theta}^2, \quad \nabla^2\tilde{\Psi}(x) = \tilde{\Theta}^2, \tag{2.13}$$

where $\bar{\Theta}$ and $\tilde{\Theta}$ are the diagonal positive semi-definite matrices with diagonal entries

$$\bar{\theta}_j := \sqrt{(x_j-\ell_j)^{-2} + (u_j-x_j)^{-2}} \tag{2.14}$$

and

$$\tilde{\theta}_j := (\min\{x_j-\ell_j, u_j-x_j\})^{-1} \tag{2.15}$$

respectively. Note that $\bar{\theta}_j$ and $\tilde{\theta}_j$ are positive if x_j is a bounded variable, and that $\bar{\Theta}, \tilde{\Theta}, \bar{\theta}_j$ and $\tilde{\theta}_j$ depend on x. Finally, if there are no free variables, $\bar{\Theta}$ and $\tilde{\Theta}$ are positive definite and $\bar{\Psi}$ and $\tilde{\Psi}$ are strictly convex.

To conclude this subsection, we have the following:

Remark 2.1 *When λ and υ are finite, then*

$$\tilde{\psi}(\xi) = \max\left\{\frac{\xi-\lambda}{\nu} - \ln(\xi-\lambda), \frac{\upsilon-\xi}{\nu} - \ln(\upsilon-\xi)\right\} + \ln\min\{\breve{\xi}-\lambda, \upsilon-\breve{\xi}\}. \tag{2.16}$$

Thus $\tilde{\psi}$ is the maximum of two almost standard logarithmic barrier functions associated with the two bounds.

Proof: Let $R(\xi)$ be the difference between the expressions in the maximum in (2.16):

$$R(\xi) = \frac{v-\xi}{\nu} - \ln(v-\xi) - \frac{\xi-\lambda}{\nu} + \ln(\xi-\lambda).$$

Then $R(\rho) = 0$, and it suffices to show that $R(\xi) > 0$ for $\xi \in (\rho, v)$ and $R(\xi) < 0$ for $\xi \in (\lambda, \rho)$. Note that $R'(\xi) = -\frac{2}{\nu} + \frac{1}{v-\xi} + \frac{1}{\xi-\lambda}$, which after manipulation becomes

$$R'(\xi) = \frac{2(\xi-\rho)^2}{\nu(v-\xi)(\xi-\lambda)},$$

and so $R'(\xi) \geq 0$ and is zero only at $\xi = \rho$. Thus $R(\xi) > R(\rho) = 0$ for $\xi \in (\rho, v)$ and $R(\xi) < R(\rho) = 0$ for $\xi \in (\lambda, \rho)$. □

2.3 Metrics on primal and dual displacements

Let $\hat{x} \in \text{int}\, X$ be fixed. Let Ψ be either $\bar{\Psi}$ or $\tilde{\Psi}$, and let $\hat{\Theta}^2 := \nabla^2\Psi(\hat{x})$, so that $\hat{\Theta}$ is either $\bar{\Theta}$ or $\tilde{\Theta}$ with entries given by (2.14)-(2.15).

Definition 2.1 *For any $x', x'' \in F(P)$, the distance between them in the metric associated with $\hat{x}$ is*

$$\|x' - x''\|_{\hat{x}} := \|\hat{\Theta}(x' - x'')\|. \tag{2.17}$$

Recall that the norm on the right-hand side of (2.17) is the Euclidean norm. We call $\|\cdot\|_{\hat{x}}$ the (primal) barrier norm at $\hat{x}$. (The matrix $\bar{\Theta}$ is frequently used as a scaling matrix to define metrics in interior-point methods – see for instance McShane et al. [11] for a primal-dual version and Vanderbei [23] for a primal-only version. In fact, Vanderbei uses instead the scaling matrix $\tilde{\Theta}$, which he regards as an approximation to $\bar{\Theta}$. In our development it arises naturally in its own right.)

Proposition 2.1 $\|\cdot\|_{\hat{x}}$ *is a norm on the null space of A.*

Proof: It is clear that $\|\cdot\|_{\hat{x}}$ is positively homogeneous, nonnegative, and satisfies the triangle inequality. Now note from (2.14)-(2.15) that $\|d\|_{\hat{x}} = 0$ only if $d_j = 0$ for all bounded variables x_j. Since we are assuming that the columns of A corresponding to free variables are linearly independent, it follows that $\|d\|_{\hat{x}} = 0$ for d in the null space of A if and only if $d = 0$. □

We also write $\|d\|_{\hat{x}}$ for $\|\hat{\Theta}d\|$ for d not in the null space of A, although it may not be a norm.

Let S denote the set $\{s : \ell^T s^+ - u^T s^- > -\infty\}$. For $s \in S$, $s_j \geq 0$ if $-\infty < \ell_j < u_j = +\infty, s_j \leq 0$ if $-\infty = \ell_j < u_j < +\infty$, and $s_j = 0$ if $\ell_j = -\infty, u_j = +\infty$. Therefore $S - S = \{s : s_j = 0 \text{ if } \ell_j = -\infty \text{ and } u_j = +\infty\}$. We use rel int S to denote the relative interior of S. It is easy to see that $(y, s) \in F^0(D)$ if and only if $A^T y + s = c$ and $s \in$ rel int S. We can define a dual metric on displacements in S:

Definition 2.2 *For any s', s'' in S, the distance between them in the dual metric associated with $\hat{x}$ is*

$$\|s' - s''\|_{\hat{x}}^* := \|\hat{\Theta}^{-1}(s' - s'')\|. \tag{2.18}$$

Here a diagonal entry of $\hat{\Theta}^{-1}$ is taken to be $+\infty$ if the corresponding entry of $\hat{\Theta}$ is zero; but for such entries the corresponding component of $s' - s''$ is zero.

We call $\|\cdot\|_{\hat{x}}^*$ the dual barrier norm at $\hat{x}$. We obtain directly:

Proposition 2.2 $\|\cdot\|_{\hat{x}}^*$ *is a norm on the subspace* $S - S = \{s : s_j = 0$ *if* $\ell_j = -\infty$ *and* $u_j = +\infty\}$. □

(In the standard case, $\hat{\Theta} = \hat{X}^{-1}$ where $\hat{X}$ is the diagonal matrix containing the components of $\hat{x}$, and so

$$\|x' - x''\|_{\hat{x}} = \|\hat{X}^{-1}(x' - x'')\|$$

and

$$\|s' - s''\|_{\hat{x}}^* = \|\hat{X}(s' - s'')\|.)$$

Proposition 2.3 *Let $\hat{x} \in F^0(P)$. Then $\{\hat{x} + d : Ad = 0, \|d\|_{\hat{x}} < 1\} \subseteq F^0(P)$.*

Proof: We only need to show that $\hat{x} + d \in$ int X for such d. From (2.13)-(2.15) we see that $\|\hat{\Theta} d\| < 1$ implies $|d_j| / \min\{\hat{x}_j - \ell_j, u_j - \hat{x}_j\} < 1$ for each bounded variable. Hence if ℓ_j is finite, $\hat{x}_j + d_j - \ell_j = (\hat{x}_j - \ell_j)(1 + d_j/(\hat{x}_j - \ell_j)) > 0$, and similarly $u_j - \hat{x}_j - d_j > 0$ if u_j is finite. □

We use these metrics to perform projections. The Euclidean projection of a vector $v \in R^n$ onto the null space of A can easily be seen to be the unique solution to

$$\begin{array}{lrcl} \max\limits_{d} & v^T d & - & \frac{1}{2}\|d\|^2 \\ \text{s.t.} & Ad & = & 0. \end{array}$$

Correspondingly we have:

Theorem 2.1 *There is a unique solution to the problem*

$$\begin{array}{lrcl} \max\limits_{d} & v^T d & - & \frac{1}{2}\|d\|_\sharp^2 \\ \text{s.t.} & Ad & = & 0. \end{array} \tag{2.19}$$

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

$$\begin{pmatrix} \hat{\Theta}^2 & A^T \\ A & O \end{pmatrix} \begin{pmatrix} d \\ y \end{pmatrix} = \begin{pmatrix} v \\ 0 \end{pmatrix}, \tag{2.20}$$

and satisfies

$$\|\bar{d}\|_\sharp^2 = v^T \bar{d}; \tag{2.21}$$

$$\|\bar{d}\|_\sharp = \|v - A^T \bar{y}\|_\sharp^*; \tag{2.22}$$

and if $v \in S - S$

$$\|\bar{d}\|_\sharp \leq \|v\|_\sharp^*. \tag{2.23}$$

Proof: Since $v^T d - \frac{1}{2} d^T \hat{\Theta}^2 d$ is concave, the Karush-Kuhn Tucker conditions (2.20) are necessary and sufficient for d to solve (2.19). To see that (2.20) has a unique solution, we note that if (d, y) is a solution to the corresponding homogeneous solution, $0 = d^T \hat{\Theta}^2 d + d^T A^T y = d^T \hat{\Theta}^2 d$ and $Ad = 0$, so that by Proposition 2.1, $d = 0$. Then, since we are assuming that A has rank m, $A^T y = 0$ implies $y = 0$.

Premultiplying the first row of (2.20) by $\bar{d}$ yields $\bar{d}^T \hat{\Theta}^2 \bar{d} = \bar{d}^T v$, which is (2.21). If $v \in S - S$, then $\|\bar{d}\|_\sharp^2 = (\hat{\Theta}\bar{d})^T(\hat{\Theta}^{-1}v) \leq \|\hat{\Theta}\bar{d}\| \, \|\hat{\Theta}^{-1}v\| = \|\bar{d}\|_\sharp \, \|v\|_\sharp^*$, which is (2.23). Finally, $\hat{\Theta}^2 \bar{d} = v - A^T \bar{y}$ shows that $v - A^T \bar{y} \in S - S$, so that $\hat{\Theta}\bar{d} = \hat{\Theta}^{-1}(v - A^T \bar{y})$, whence taking norms gives (2.22). □

We write $P_\sharp(v)$ for the solution to (2.19). Dually, we have:

Theorem 2.2 *There is a unique solution to*

$$\begin{array}{ll} \min\limits_{y} & \frac{1}{2}(\|v - A^T y\|_\sharp^*)^2 \\ \text{s.t.} & v - A^T y \in S - S. \end{array} \tag{2.24}$$

Moreover, this solution $\bar{y}$ is part of the solution $(\bar{d}, \bar{y})$ to (2.20).

Proof: We can write (2.24) as

$$\begin{aligned}
\min_{y,s} \quad & \tfrac{1}{2} s^T \hat{\Theta}^{-2} s \\
\text{s.t.} \quad & A^T y + s = v, \\
& s \in S - S,
\end{aligned}$$

with $s = v - A^T y$, for which the Karush-Kuhn-Tucker conditions are again necessary and sufficient. But these conditions are again (2.20), where d is the Lagrange multiplier vector for the constraints $A^T y + s = v$. The results then follow from Theorem 2.1. (Note that (2.24) is feasible from our assumption that $\tilde{A}$ has full column rank.) □

Using (2.21) and (2.22) we see that (2.19) and (2.24) are dual problems with equal optimal values.

Proposition 2.4 *For any $\hat{x} \in$ int $X, \nabla\Psi(\hat{x}) \in S - S$ and $\|\nabla\Psi(\hat{x})\|_{\hat{x}}^* \leq \sqrt{p}$.*

Proof: It follows from Lemma 2.1 that $\nabla\Psi(\hat{x}) \in S - S$, so that $\|\nabla\Psi(\hat{x})\|_{\hat{x}}^*$ is well-defined, and $\|\nabla\Psi(\hat{x})\|_{\hat{x}}^* = \|\hat{\Theta}^{-1}\nabla\Psi(\hat{x})\|$. For the p nonzero components of $\hat{\Theta}^{-1}\nabla\Psi(\hat{x})$, from Lemma 2.1 and (2.13)-(2.15), we see that $|(\hat{\Theta}^{-1}\nabla\Psi(\hat{x}))_j| \leq 1$ for all non-free variable indices, and so $\|\hat{\Theta}^{-1}\nabla\Psi(\hat{x})\| \leq \sqrt{p}$. □
(Note that if all bounds are 1-sided, the inequality in the proposition becomes an equality.)

2.4 Central trajectories and conjugate barriers

Let $\mu > 0$. Here we consider the barrier problem

$$(BP) \qquad \begin{aligned}
\min_{x} \quad & c^T x \; + \; \mu\Psi(x) \\
\text{s.t.} \quad & Ax \; = \; b,
\end{aligned}$$

where Ψ is either $\bar{\Psi}$ or $\tilde{\Psi}$. Here we extend $\bar{\Psi}$ and $\tilde{\Psi}$ (and hence Ψ) to all of R^n, defining them to be $+\infty$ on $R^n \backslash(\text{int } X)$; then Ψ is a closed proper convex function on R^n.

By our assumption that $F^0(P)$ is nonempty, there is a feasible point for (BP) with finite objective value. Since we assume that (P) has a nonempty bounded set of optimal solutions, on any nonzero direction in the null space of A intersected with the recession cone of X, $c^T x$ increases linearly while $\Psi(x)$ decreases at most logarithmically. It follows that the convex function $c^T x + \mu\Psi(x)$ has no nonzero direction of recession in the null space of A, and hence by standard arguments that (BP) has an optimal solution. Moreover, since Ψ is strictly convex in the bounded variables, and $\tilde{A}$ has full column rank, we deduce that the optimal solution is unique. We denote the optimal solution by $x(\mu)$, and define the (primal) central trajectory to be $\{x(\mu) : \mu > 0\}$.

Again, the KKT conditions are necessary and sufficient for optimality in (BP) and we conclude that $x(\mu)$ with some $y(\mu), s(\mu)$ satisfies uniquely

$$Ax = b, \tag{2.25a}$$

$$A^T y + s = c, \tag{2.25b}$$

$$\mu \nabla \Psi(x) + s = 0. \tag{2.25c}$$

Condition (2.25c) implies that $s = -\mu \nabla \Psi(x) \in S$, and that $-s/\mu = \nabla \Psi(x) \in \partial \Psi(x)$, where $\partial \Psi$ denotes the subdifferential of Ψ. Let us define the convex conjugate of Ψ by

$$\Psi^*(s) := \sup_x \{-s^T x - \Psi(x)\}. \tag{2.26}$$

(Note that the usual convex conjugate replaces $-s^T x$ by $s^T x$; our modification turns out to be useful for the standard case.) Since Ψ is closed, proper, and convex, so is Ψ^*, and we have (2.25c) (i.e. $-s/\mu \in \partial \Psi(x)$) equivalent to

$$-x \in \partial \Psi^*(s/\mu) \tag{2.25c$'$}$$

(Rockafellar [16], §§12 and 23).

When we replace (2.25c) by (2.25c$'$), (2.25) becomes the necessary and sufficient conditions for (y, s) to solve

$$(BD) \qquad \begin{array}{llcllcl} \max\limits_{y,s} & b^T y & - & \mu \Psi^*(s/\mu) & & & \\ \text{s.t.} & A^T y & + & s & = & c. & \end{array}$$

(As we shall see below, $\Psi^*(s/\mu)$ is $+\infty$ unless $s \in$ rel int S.) We can also obtain (BD) as the ordinary (Lagrange) dual of the ordinary convex program (BP) (Rockafellar [16], §28). Note that, while (BP) is clearly closely related to (P), (BD) seems not so closely related to (D), since the terms $\ell^T s^+$ and $-u^T s^-$ appear to be missing. It turns out, however, that these two terms arise naturally as part of the barrier term $\mu \Psi^*(s/\mu)$ in (BD); see the Appendix for details.

(We remark that if there are no free variables, then Ψ is strictly convex as well as essentially smooth, and then Ψ^* is too, and can be viewed as the Legendre conjugate of Ψ (with appropriate sign changes) – see [16], §26.)

The detailed form of Ψ^* is not needed in the rest of the paper. For completeness, it is derived in the Appendix, and permits the proof of the theorem below.

We make two observations here: Ψ^* is probably important in the construction of primal-dual algorithms; and devising a suitable primal-dual potential function seems to be complicated by the presence of μ inside the function Ψ^*. (In the standard case, $\Psi^*(s) = -\sum_j \ln s_j - n$, so $\Psi^*(s/\mu) = -\sum_j \ln s_j + n(\ln \mu - 1)$; μ can be extracted from Ψ^* and causes no difficulty.) Nesterov and Nemirovsky [14] introduce a primal-dual potential function via the conjugate barrier function in a fairly general setting; but their development requires the problems to be in "conic" form and the barrier functions to be logarithmically homogeneous, neither of which holds in our framework. See also Section 5.

Summarizing the results of this subsection, and using the derivation in the Appendix, we have:

Theorem 2.3 *For any $\mu > 0$, there is a unique solution $x(\mu)$ to (BP) and a unique solution $(y(\mu), s(\mu))$ to (BD), and these solutions together solve the system (2.25). For any $x \in$ int X and any $s \in$ rel int S, $\mu\Psi(x) \to 0$ and $-\mu\Psi^*(s/\mu) \to \ell^T s^+ - u^T s^-$ as $\mu \downarrow 0$. Moreover, $\Psi^*(s) = \sum_j \psi^*(s_j)$, where the individual ψ^*'s are given by (A.1)-(A.3), (A.5) or (A.7) in the Appendix.* □

2.5 Duality gaps and measures of centrality

In the standard case (2.25c) becomes $x \circ s = \mu e$, where $x \circ s := (x_j s_j)$, and the duality gap is

$$(x-\ell)^T s^+ + (u-x)^T s^- = x^T s = e^T(x \circ s) = n\mu.$$

Thus if we follow the central trajectory as $\mu \downarrow 0$, any limit points of $x(\mu)$ and $(y(\mu), s(\mu))$ will necessarily be optimal primal and dual solutions to (P) and (D). We would like a similar result in the general case. In fact, we would like a generalization that allows x to be nearly, rather than exactly, central; this will have applications in the algorithms to follow.

We wish to permit s/μ to be close to $-\nabla\Psi(x)$. We will measure closeness in the norm $\|\cdot\|_x^*$. To simplify notation, we will write t for s/μ.

Theorem 2.4 *Suppose $t = -\nabla\Psi(x) + h$, where $x \in$ int X and $h \in S - S$ with $\|h\|_x^* \le \beta < 1$. Then $t \in$ rel int S and*

$$(x-\ell)^T t^+ + (u-x)^T t^- \le p + \beta\sqrt{p}. \tag{2.27}$$

(Recall that p is the number of bounded variables.)

This theorem shows that the duality gap corresponding to $x(\mu)$ and $(y(\mu), s(\mu))$ is at most $p\mu$, since we can set $t = s(\mu)/\mu$ and $\beta = 0$.

Proof: Let $\Theta^2 = \nabla^2\Psi(x)$, so that $\|\Theta^{-1}h\| \le \beta$. Since $h_j = 0$ if x_j is a free variable, $\|\Theta^{-1}h\|_1 \le \beta\sqrt{p}$. We prove (2.27) by showing that

$$(x_j - \ell_j)t_j^+ + (u_j - x_j)t_j^- \leq 1 + |\theta_j^{-1} h_j| \tag{2.28}$$

if x_j is a bounded variable. Since $(-\nabla\Psi(x))_j$ is zero if x_j is free, so is t_j and hence the left-hand side of (2.28) is zero for such j. Hence adding yields (2.27).

We use the notation of previous subsections, with τ denoting t_j and η denoting h_j. We also write θ for θ_j. Suppose first $-\infty < \lambda < \upsilon = +\infty$. Then $-\psi'(\xi) = \theta = (\xi - \lambda)^{-1}$. Since $|\theta^{-1}\eta| \leq \beta < 1, \tau = \theta + \eta > 0$ and

$$(\xi - \lambda)\tau^+ + (\upsilon - \xi)\tau^- = \theta^{-1}(\theta + \eta) \leq 1 + |\theta^{-1}\eta|$$

as desired. Similarly, if $-\infty = \lambda < \upsilon < +\infty$, $-\psi'(\xi) = -(\upsilon - \xi)^{-1}$ and $\theta = (\upsilon - \xi)^{-1}$. Thus $\tau = -\theta + \eta < 0$ and

$$\begin{aligned}(\xi - \lambda)\tau^+ + (\upsilon - \xi)\tau^- &= \theta^{-1}(-(-\theta + \eta)) \\ &\leq 1 + |\theta^{-1}\eta|.\end{aligned}$$

In addition to showing the appropriate part of (2.28), we have shown that each component of t has the correct sign if necessary, so that $t \in$ rel int S.

We still need to show (2.28) in the case $-\infty < \lambda < \upsilon < +\infty$. We first consider the case that $\psi = \bar{\psi}$. Then $-\psi'(\xi) = (\xi - \lambda)^{-1} - (\upsilon - \xi)^{-1}$ and $\theta = \sqrt{(\xi - \lambda)^{-2} + (\upsilon - \xi)^{-2}}$. If $\tau \geq 0$ we have to show that

$$\begin{aligned}(\xi - \lambda)\tau^+ &+ (\upsilon - \xi)\tau^- = (\xi - \lambda)\tau \\ &= (\xi - \lambda)((\xi - \lambda)^{-1} - (\upsilon - \xi)^{-1} + \eta) \leq 1 + |\theta^{-1}\eta|. \end{aligned} \tag{2.29}$$

This is trivial if $\eta \leq 0$, while if $\eta \geq 0$ it is equivalent to showing that

$$((\xi - \lambda)\theta - 1)(\theta^{-1}\eta) \leq (\xi - \lambda)(\upsilon - \xi)^{-1}.$$

But this holds since $(\xi - \lambda)^{-1} < \theta < (\xi - \lambda)^{-1} + (\upsilon - \xi)^{-1}$ and $0 \leq \theta^{-1}\eta < 1$. A similar argument holds in the case $\tau < 0$.

Finally, suppose $\psi = \tilde{\psi}$. Then $-\psi'(\xi) = (\rho - \xi)/(\nu \min\{\xi - \lambda, \upsilon - \xi\})$ and $\theta = (\min\{\xi - \lambda, \upsilon - \xi\})^{-1}$. Suppose $\tau \geq 0$. We want to show that

$$\begin{aligned}(\xi - \lambda)\tau^+ &+ (\upsilon - \xi)\tau^- = (\xi - \lambda)\tau \\ &= (\xi - \lambda)(-\psi'(\xi) + \eta) \leq 1 + |\theta^{-1}\eta|. \end{aligned} \tag{2.30}$$

If $\xi - \lambda \leq \upsilon - \xi$, this reduces to

$$(\rho - \xi)/\nu + \theta^{-1}\eta \leq 1 + |\theta^{-1}\eta|,$$

which is trivial. So assume $v - \xi < \xi - \lambda$, in which case (2.30) is equivalent to

$$\frac{\xi - \lambda}{v - \xi}\theta^{-1}\eta - |\theta^{-1}\eta| \leq 1 + \frac{(\xi - \lambda)(\xi - \rho)}{\nu(v - \xi)}. \tag{2.31}$$

If $\eta \leq 0$, this is again immediate, so suppose $\eta > 0$, in which case (2.31) holds if and only if (multiplying by $v - \xi$)

$$2(\xi - \rho)|\theta^{-1}\eta| \leq v - \xi + (\xi - \lambda)(\xi - \rho)/\nu.$$

But $|\theta^{-1}\eta| \leq \beta < 1, \xi - \rho > 0$ and $2 = \frac{v-\xi}{\nu} + \frac{\xi-\lambda}{\nu} \leq \frac{v-\xi}{\xi-\rho} + \frac{\xi-\lambda}{\nu}$, which together imply the desired inequality. A similar argument holds in the case $\tau < 0$. This completes the proof. □

We want t to be of the form s/μ, where $s = c - A^T y$ for some y. Then $\mu h = \mu t + \mu\nabla\Psi(x) = c + \mu\nabla\Psi(x) - A^T y$. Choosing y to make h small is then like problem (2.24). Then combining Theorems 2.1, 2.2 and 2.4 we obtain 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 2.5 *Suppose $\hat{x} \in F^0(P)$ is given. Let $\hat{\mu} > 0$ be chosen, and let*

$$v := c + \hat{\mu}\nabla\Psi(\hat{x}). \tag{2.32}$$

Let $(\hat{d}, \hat{y})$ be the solution to (2.20) for v given above, and hence define

$$\hat{s} := c - A^T\hat{y}. \tag{2.33}$$

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

$$\|\hat{d}/\hat{\mu}\|_{\hat{x}} = \|\hat{s}/\hat{\mu} + \nabla\Psi(\hat{x})\|^*_{\hat{x}} \leq \beta, \tag{2.34}$$

where $\beta < 1$, then

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

(ii) *the duality gap is* $(\hat{x} - \ell)^T\hat{s}^+ + (u - \hat{x})^T\hat{s}^- \leq \hat{\mu}(p + \beta\sqrt{p})$.

If (2.34) holds, we say $\hat{x}$ is β-close to $x(\hat{\mu})$.

Proof: The equality of the norms follows from (2.22). Now define $\hat{t} := \hat{s}/\hat{\mu}$ and $\hat{h} := \nabla\Psi(\hat{x}) + \hat{s}/\hat{\mu} = \nabla\Psi(\hat{x}) + \hat{t}$. From (2.20), (2.32) and (2.33) we have

$$\begin{aligned} A^T\hat{y} + \hat{s} &= c, \\ A^T\hat{y} + \hat{\Theta}^2\hat{d} &= c + \hat{\mu}\nabla\Psi(\hat{x}), \end{aligned}$$

and we readily derive

$$\hat{h} = \frac{1}{\hat{\mu}}\hat{\Theta}^2\hat{d},$$

and therefore $\hat{h} \in S - S$. Also, $\hat{t} = -\nabla\Psi(\hat{x}) + \hat{h}$, and so from (2.34) we can apply Theorem 2.4 to obtain

$$(\hat{x} - \ell)^T\hat{t}^+ + (u - \hat{x})^T\hat{t}^- \leq p + \beta\sqrt{p}.$$

Substituting $\hat{s} = \hat{\mu}\hat{t}$ in the above inequality gives (ii). Therefore $A^T\hat{y} + \hat{s} = c$, and $\ell^T\hat{s}^+ - u^T\hat{s}^- > -\infty$, so that $(\hat{y}, \hat{s}) \in F(D)$. Moreover, since $\hat{t} \in$ rel int S from Theorem 2.4, so is $\hat{s}$, and this shows (i). □

The last result of this subsection is a sufficient condition for $\hat{x}$ to be β-close to $x(\hat{\mu})$ for some $\hat{\mu} > 0$.

Proposition 2.5 *Suppose $\hat{x} \in F^0(P)$ and $\hat{\mu} > 0$ are given and suppose there exist $\tilde{y}$ and $\tilde{s}$ satisfying:*

(i) $A^T\tilde{y} + \tilde{s} = c$, $\tilde{s} \in S - S$,

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

Then $\hat{x}$ is β-close to $x(\hat{\mu})$.

Proof: According to Theorem 2.5 (equation (2.34)), $\hat{x}$ is β-close to $x(\hat{\mu})$ if $\|\hat{s}/\mu + \nabla\Psi(\hat{x})\|_{\hat{x}}^* \leq \beta$, where $\hat{s} = c - A^T\hat{y}$ and $(\hat{d}, \hat{y})$ is the solution (2.20) with $v = c + \hat{\mu}\nabla\Psi(\hat{x})$. From Theorem 2.2, $\hat{y}$ is the optimal solution to

$$\begin{aligned} \min_{y} \quad & \|c + \hat{\mu}\nabla\Psi(\hat{x}) - A^Ty\|_{\hat{x}}^*. \\ \text{s.t.} \quad & c + \hat{\mu}\nabla\Psi(\hat{x}) - A^Ty \in S - S. \end{aligned}$$

But from hypothesis (i) $\tilde{y}$ is feasible for this program. Therefore,

$$\begin{aligned} \|\hat{s} + \hat{\mu}\nabla\Psi(\hat{x})\|_{\hat{x}}^* = \|c + \hat{\mu}\nabla\Psi(\hat{x}) - A^T\hat{y}\|_{\hat{x}}^* &\leq \|c + \hat{\mu}\nabla\Psi(\hat{x}) - A^T\tilde{y}\|_{\hat{x}}^* \\ &= \|\tilde{s} + \hat{\mu}\nabla\Psi(\hat{x})\|_{\hat{x}}^* \leq \beta\hat{\mu} \end{aligned}$$

from (ii) above. Therefore $\|\hat{s}/\mu + \nabla\Psi(\hat{x})\|_{\hat{x}}^* \leq \beta$, proving the result. □

2.6 Approximations to barrier functions and their gradients

Here we provide bounds on the errors in first-order Taylor approximations to Ψ and $\nabla\Psi$, for Ψ equal to $\bar{\Psi}$ and $\tilde{\Psi}$. These bounds are crucial to the algorithms to follow. It turns out that such bounds follow naturally from the self-concordance of $\bar{\Psi}$ and $\tilde{\Psi}$ in the sense of Nesterov and Nemirovsky [14]. We will show the functions to be self-concordant, to stress the importance of this unifying concept. However, our proofs of the bounds are derived independently of the results in Nesterov and Nemirovsky [14]. In fact for the two barrier functions $\bar{\Psi}$ and $\tilde{\Psi}$, we are able to obtain slightly improved constants.

First we state the main results. Throughout, Ψ is either $\bar{\Psi}$ or $\tilde{\Psi}$.

Theorem 2.6 *Let $\hat{x}$, $x \in F^0(P)$ and let $\bar{d} = x - \hat{x}$. Let $\hat{\Theta}^2 = \nabla^2\Psi(\hat{x})$. Then*

$$\|\nabla\Psi(x) - \nabla\Psi(\hat{x}) - \hat{\Theta}^2\bar{d}\|_x^* \leq \|\bar{d}\|_{\hat{x}}^2. \tag{2.35}$$

(Note that the left-hand side is the error in the first-order approximation to $\nabla\Psi(x)$ based on $\hat{x}$.)

Theorem 2.7 *Let $\hat{x} \in F^0(P)$ and $\bar{d}$ be in the null space of A. Then, if $\gamma > 0$ is such that $\gamma\|\bar{d}\|_{\hat{x}} < 1$, $\hat{x} + \gamma\bar{d} \in F^0(P)$ and*

$$\begin{aligned}\Psi(\hat{x}) + \gamma\nabla\Psi(\hat{x})^T\bar{d} &\leq \Psi(\hat{x} + \gamma\bar{d}) \\ &\leq \Psi(\hat{x}) + \gamma\nabla\Psi(\hat{x})^T\bar{d} + \frac{\gamma^2\|\bar{d}\|_{\hat{x}}^2}{2(1-\gamma\|\bar{d}\|_{\hat{x}})}.\end{aligned} \tag{2.36}$$

(This result states that Ψ is well approximated by its first-order Taylor approximation. But note that the right-hand side of (2.36) can be written as

$$\Psi(\hat{x}) + \gamma\nabla\Psi(\hat{x})^T\bar{d} + \frac{\gamma^2}{2}\|\bar{d}\|_{\hat{x}}^2 + \frac{\gamma^3\|\bar{d}\|_{\hat{x}}^3}{2(1-\gamma\|\bar{d}\|_{\hat{x}})}.$$

which is close to its second-order Taylor approximation.)

Theorem 2.7 will basically follow from the fundamental theorem of calculus and Theorem 2.6. But in order to prove Theorem 2.7, we need to relate two norms of the form $\|\cdot\|_x$ and $\|\ \|_{\hat{x}}$. This is what the concept of self-concordance allows us to do. There are two possible definitions, depending on whether the function is twice or thrice continuously differentiable. The basic definition of self-concordance given in Nesterov and Nemirovsky [14] is for thrice continuously differentiable functions: a convex function Φ on an open subset Q of R^n is self-concordant (with parameter 1) if Φ is C^3 and for every $x \in Q$ and $d \in R^n$,

$$|D^3\Phi(x)[d,d,d]| \leq 2(d^T\nabla^2\Phi(x)d)^{3/2}. \tag{2.37}$$

Because $\tilde{\Psi}$ is not thrice continuously differentiable, we elect instead to use the following definition of self-concordance which is based on twice continuous differentiability: a convex function Φ on an open subset Q of R^n is self-concordant if Φ is C^2 and for every $x \in Q$, $d \in R^n$, and $h \in R^n$, if γ is a scalar satisfying $|\gamma|(d^T \nabla^2 \Phi(x) d)^{1/2} < 1$, then

$$\begin{aligned}(1 - |\gamma|(d^T \nabla^2 \Phi(x) d)^{1/2})^2 h^T \nabla^2 \Phi(x) h &\leq h^T \nabla^2 \Phi(x + \gamma d) h \qquad (2.38)\\ &\leq (1 - |\gamma|(d^T \nabla^2 \Phi(x) d)^{1/2})^{-2} h^T \nabla^2 \Phi(x) h.\end{aligned}$$

In [14], Nesterov and Nemirovsky prove that (2.37) implies (2.38), and it is (2.38) which is the basic property used to derive results. For our purposes, (2.38) will be the working definition of self-concordance.

Proposition 2.6 *Both $\bar{\Psi}$ and $\tilde{\Psi}$ are self-concordant on int X.*

The proofs of Proposition 2.6 and Theorem 2.6 are somewhat laborious, and are deferred to the Appendix. However, to show how useful self-concordance is, we now demonstrate how Theorem 2.7 follows from Theorem 2.6 and Proposition 2.6:

Proof of Theorem 2.7: The left-hand inequality is a consequence of the convexity of Ψ. For the right-hand inequality, first note that by replacing $\gamma \bar{d}$ by $\bar{d}$ we can assume that $\gamma = 1$ – the presence of γ is helpful for applications. Then

$$\begin{aligned}\Psi(\hat{x} + \bar{d}) &= \Psi(\hat{x}) + \int_0^1 \nabla \Psi(\hat{x} + \lambda \bar{d})^T \bar{d}\, d\lambda \\ &= \Psi(\hat{x}) + \int_0^1 (\nabla \Psi(\hat{x}) + \lambda \nabla^2 \Psi(\hat{x}) \bar{d})^T \bar{d}\, d\lambda \\ &+ \int_0^1 [\nabla \Psi(\hat{x} + \lambda \bar{d}) - \nabla \Psi(\hat{x}) - \lambda \nabla^2 \Psi(\hat{x}) \bar{d}]^T \bar{d}\, d\lambda. \qquad (2.39)\end{aligned}$$

The first two terms give

$$\Psi(\hat{x}) + \nabla \Psi(\hat{x})^T \bar{d} + \frac{1}{2} \|\bar{d}\|_{\hat{x}}^2,$$

so it suffices to show that

$$\int_0^1 |[\nabla \Psi(\hat{x} + \lambda \bar{d}) - \nabla \Psi(\hat{x}) - \lambda \hat{\Theta}^2 \bar{d}]^T \bar{d}| d\lambda \leq \frac{\|\bar{d}\|_{\hat{x}}^3}{2(1 - \|\bar{d}\|_{\hat{x}})}. \qquad (2.40)$$

Note that the integrand involves an expression we have bounded in Theorem 2.6. Let $x(\lambda) := \hat{x} + \lambda \bar{d}$, and observe that for $v \in S - S$,

$$|v^T \bar{d}| = |(\Theta^{-1}(\lambda)v)^T(\Theta(\lambda)\bar{d})| \leq \|v\|^*_{x(\lambda)} \|\bar{d}\|_{x(\lambda)},$$

where $\Theta^2(\lambda) := \nabla^2 \Psi(x(\lambda))$. We know from Theorem 2.6 that

$$\|\nabla \Psi(\hat{x} + \lambda \bar{d}) - \nabla \Psi(\hat{x}) - \lambda \hat{\Theta}^2 \bar{d}\|^*_{x(\lambda)} \leq \|\lambda \bar{d}\|^2_{\hat{x}},$$

so the left-hand side of (2.40) is at most

$$\int_0^1 \lambda^2 \|\bar{d}\|^2_{\hat{x}} \|\bar{d}\|_{x(\lambda)} d\lambda. \tag{2.41}$$

We note here two different norms of $\bar{d}$, but we can use the self-concordance of Ψ (Proposition 2.6) to obtain (see (2.38))

$$\begin{aligned} \|\bar{d}\|_{x(\lambda)} &\leq (1 - \lambda \|\bar{d}\|_{\hat{x}})^{-1} \|\bar{d}\|_{\hat{x}} \\ &\leq (1 - \|\bar{d}\|_{\hat{x}})^{-1} \|\bar{d}\|_{\hat{x}}. \end{aligned} \tag{2.42}$$

Hence (2.41) is at most

$$(1 - \|\bar{d}\|_{\hat{x}})^{-1} \|\bar{d}\|^3_{\hat{x}} \int_0^1 \lambda^2 d\lambda;$$

since $\int_0^1 \lambda^2 d\lambda = \frac{1}{3} < \frac{1}{2}$, this establishes (2.40). □

3 Tracing the Central Trajectory Using Newton's Method

3.1 The Trajectory-Following Algorithm

In this section we study the use of Newton's method for tracing a sequence of points near the optimal solution of the barrier problem

$$(BP) \qquad \begin{array}{rcccl} \min\limits_x & c^T x & + & \mu \Psi(x) \\ \text{s.t.} & Ax & = & b, \\ \ell & \leq & x & \leq & u, \end{array}$$

for a decreasing sequence of positive values of the parameter μ converging to zero. Here Ψ is either $\bar{\Psi}$ or $\tilde{\Psi}$.

The idea of the algorithm is the same as in many of the path-following algorithms for linear programming, see, e.g., Renegar [15], Gonzaga [3], Kojima et al. [9], and Monteiro and Adler [12]. At the start of an iteration, we have on hand a value $\hat{\mu} > 0$ of the parameter μ, and a vector $\hat{x}$ that is β-close to $x(\hat{\mu})$ for some fixed value of $\beta < 1$. We then want to generate a new value of x by Newton's method, and to shrink $\hat{\mu}$ to $\mu := \alpha\hat{\mu}$ for some fraction $\alpha < 1$. Finally, we want to show that x is β-close to $x(\mu)$.

Let $\hat{x}, \hat{\mu}$, and β be given as above. Then, just as in the "approximately-centered" Theorem 2.5, define

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

let $(\hat{d}, \hat{y})$ be the solution to (2.20) with this v, and define

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

From Theorem 2.5, $\|\hat{d}/\hat{\mu}\|_{\hat{x}} = \|\hat{s}/\hat{\mu} + \nabla\Psi(\hat{x})\|_{\hat{x}}^{*} \leq \beta$. Now note that v is the gradient of the objective function of (BP) at $\hat{x}$, while its Hessian is $\hat{\mu}\nabla^2\Psi(\hat{x}) = \hat{\mu}\hat{\Theta}^2$. Therefore, since $\hat{d}$ is the solution to (2.19) for this v, $-\hat{d}/\hat{\mu}$ 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 $\hat{x}$ is at most β in the barrier norm at $\hat{x}$. Now let x be the Newton iterate, i.e.,

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

Proposition 3.1 $x \in F^0(P)$.

Proof: From (2.20), $\hat{d}$ and hence $\bar{d}$ satisfy $A\bar{d} = 0$. Also $\|\bar{d}\|_{\hat{x}} \leq \beta < 1$, and so the result follows from Proposition 2.3. □

The next result shows that Newton's method exhibits quadratic convergence if $\hat{x}$ is β-close to $x(\hat{\mu})$ with $\beta < 1$.

Theorem 3.1 *Suppose $\hat{\mu} > 0$ is given and $\hat{x} \in F^0(P)$ is β-close to $x(\hat{\mu})$, where $\beta < 1$. Let x be defined by (3.4) via (3.1), (2.20), (3.2), and (3.3). Then x is β^2-close to $x(\hat{\mu})$.*

Proof: From Proposition 2.5, it suffices to exhibit (y, s) that satisfy

$$A^T y + s = c, \quad s \in S - S, \quad \text{and} \quad \|s/\hat{\mu} + \nabla\Psi(x)\|_x^* \le \beta^2. \tag{3.5}$$

We will show that (3.5) holds for $(y, s) = (\hat{y}, \hat{s})$.

Let us recall some relationships. Let $\hat{\Theta}^2 := \nabla^2\Psi(\hat{x})$. Then from (2.20) we have

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

so that

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

Hence

$$\|\hat{s}/\hat{\mu} + \nabla\Psi(x)\|_x^* = \|\nabla\Psi(x) - \nabla\Psi(\hat{x}) - \hat{\Theta}^2\bar{d}\|_x^*,$$

and by Theorem 2.6 this is at most $\|\bar{d}\|_{\hat{x}}^2 \le \beta^2$. This proves (3.5) and hence x is β^2-close to $x(\hat{\mu})$ from Proposition 2.5. □

As a consequence of this result, we obtain:

Theorem 3.2 *Let $\hat{x} \in F^0(P)$ be β-close to $x(\hat{\mu})$ for $\beta < 1$, and define x as above. Let*

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

Then $x \in F^0(P)$ and x is β-close to $x(\mu)$. (Recall that p is the number of bounded variables.)

Proof: That x lies in $F^0(P)$ follows from Proposition 3.1. To show that x is β-close to $x(\mu)$, note that

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

where the last inequality follows from (3.5) for $s = \hat{s}$ and Proposition 2.4 and the final equation from the definition of α. This establishes that x is β-close to $x(\mu)$ as desired. □

This theorem states that the Newton step allows us to reduce the barrier parameter $\hat{\mu}$ by the factor $\alpha < 1$ while remaining β-close to the central trajectory. Note that with $\beta = \frac{1}{2}$, then $\alpha = 1 - \frac{1}{2+4\sqrt{p}} \leq 1 - \frac{1}{6\sqrt{p}}$. Thus repeating the Newton procedure k times, we can shrink the barrier parameter to at most $\left(1 - \frac{1}{6\sqrt{p}}\right)^k$ of its original value. This is summarized in:

Theorem 3.3 *Suppose x^0 is β-close to $x(\mu^0)$ for some $\mu^0 > 0$ and $\beta = \frac{1}{2}$. Let $\alpha := 1 - \frac{1}{2+4\sqrt{p}}$, and define the iterates (x^k, y^k, s^k) recursively 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 2.5. Let $(y^k, s^k) := (\hat{y}, \hat{s})$, define $x^{k+1} := x$ from (3.3)-(3.4), and set $\mu^{k+1} := \alpha\mu^k$. Then*

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

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

(iii) the duality gap is $(x^k - \ell)^T(s^k)^+ + (u - x^k)^T(s^k)^- \leq (\alpha)^k\mu^0(p + \beta\sqrt{p})$. □

Therefore, given an initial vector x^0 that is β-close to $x(\mu^0)$ for $\beta = \frac{1}{2}$, (a bound on) the duality gap can be shrunk by a constant amount in at most $6\sqrt{p}$ iterations. The next subsection describes how to find such an initial vector.

3.2 Initializing the Algorithm

Here we show how to transform a given linear program into an equivalent linear programming problem (P) so that the new problem satisfies the condition that we have available some $\mu^0 > 0$ and a point $x^0 \in F^0(P)$ that is β-close to $x(\mu^0)$. Our construction generalizes the derivation in Monteiro and Adler [12]. To simplify the notation, we frequently omit the transposes in concatenating several vectors in this subsection (and similarly in Section 4.2) – no confusion should result.

Suppose our given linear program is

$$(\tilde{P}) \qquad \begin{array}{rcl} \min_{\tilde{x}} & \tilde{c}^T\tilde{x} & \\ \text{s.t.} \quad \tilde{A}\tilde{x} & = & \tilde{b}, \\ \tilde{\ell} \leq \tilde{x} & \leq & \tilde{u}, \end{array} \qquad (3.7)$$

which satisfies the assumptions in Section 2.1. Then, because the columns of $\tilde{A}$ corresponding to the free variables are linearly independent, there exists $\tilde{y}$

for which the components of $\tilde{c} - \tilde{A}^T\tilde{y}$ corresponding to free variables are all zero. We can then replace $\tilde{c}$ by $\tilde{c} - \tilde{A}^T\tilde{y}$, and thus can assume without loss of generality that $\tilde{c} \in S - S$. Also, we can assume (by reversing signs if necessary) that all one-sided bounds on variables $\tilde{x}_j$ are of the form $\tilde{\ell}_j \le \tilde{x}_j \le \tilde{u}_j = +\infty$. Therefore we can partition the variables and their indices into three groups and rewrite $(\tilde{P})$ as

$$(\tilde{P}) \qquad \begin{array}{rcccccl} \min & & & c_2^T x_2 & + & c_3^T x_3 & \\ \text{s.t.} & A_1x_1 & + & A_2x_2 & + & A_3x_3 & = \tilde{b} \\ (\ell_1,\ell_2,\ell_3) & \le & & (x_1,x_2,x_3) & & \le & (u_1,u_2,u_3), \end{array} \tag{3.8}$$

where $\ell_1 = (-\infty)e$, $u_1 = (+\infty)e$ (x_1 is free), $\ell_2 > (-\infty)e$, $u_2 < (+\infty)e$ (x_2 has two-sided bounds), and $\ell_3 > (-\infty)e$, $u_3 = (+\infty)e$ (x_3 has one-sided bounds). Next let $\lambda > 0$ be chosen large enough so that $x_3 \le \ell_3 + \lambda e$ for any basic feasible solution (x_1, x_2, x_3) to $(\tilde{P})$. Then $(\tilde{P})$ is equivalent to

$$\begin{array}{rcccccccl} \min & & & c_2x_2 & + & c_3x_3 & & & \\ \text{s.t.} & A_1x_1 & + & A_2x_2 & + & A_3x_3 & & & = \tilde{b} \\ & & & & & e^Tx_3 & + & x_4 & = \lambda(n_3+1) + e^T\ell_3 \\ (\ell_1,\ell_2,\ell_3,\ell_4) & \le & & (x_1,x_2,x_3,x_4) & & \le & & & (u_1,u_2,u_3,u_4), \end{array} \tag{3.9}$$

where n_3 is the number of one-sided bounded variables, and where $\ell_4 := 0$ and $u_4 := +\infty$, so that x_4 is an ordinary slack variable. Finally, we want to choose a convenient solution $\hat{x}$ that will be feasible for the new problem. Let $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4) := (0, (l_2 + u_2)/2, \ell_3 + \lambda e, \lambda)$. In order to make this solution feasible, we introduce an artificial variable x_5 with a large cost M to obtain:

$$(P) \qquad \begin{array}{rcccccccccl} \min & & & c_2^Tx_2 & + & c_3^Tx_3 & & & + & Mx_5 & \\ \text{s.t.} & A_1x_1 & + & A_2x_2 & + & A_3x_3 & & & + & a_5x_5 & = \tilde{b} \\ & & & & & e^Tx_3 & + & x_4 & + & x_5 & = \lambda(n_3+2) + e^T\ell_3 \\ (\ell_1,\ell_2,\ell_3,\ell_4,\ell_5) & \le & & (x_1,x_2,x_3,x_4,x_5) & & \le & & & & & (u_1,u_2,u_3,u_4,u_5), \end{array} \tag{3.10}$$

where $a_5 := \lambda^{-1}(\tilde{b} - A_1\hat{x}_1 - A_2\hat{x}_2 - A_3\hat{x}_3)$, $\ell_5 := 0$ and $u_5 := +\infty$. Let $x := (x_1, x_2, x_3, x_4, x_5)$ and define b, c, A, ℓ, and u so that (3.10) is (P) of Section 2. Then note that for M sufficiently large, any optimal solution x to (3.10) will have $x_5 = 0$, and (x_1, x_2, x_3) will solve $(\tilde{P})$. Also note that $\hat{x} := (\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4, \hat{x}_5) := (0, (\ell_2 + u_2)/2, \ell_3 + \lambda e, \lambda, \lambda)$ is feasible for (3.10). Furthermore, for $\Psi(x) = \tilde{\Psi}(x)$

or $\Psi(x) = \tilde{\Psi}(x)$, $\nabla\Psi(\hat{x}) = (0, 0, \lambda^{-1}e, \lambda^{-1}, \lambda^{-1})$. We can partition the rows of A so that

$$A = \begin{bmatrix} B \\ f^T \end{bmatrix}$$

We now define $\hat{\mu}$ and $(\hat{y}, \hat{s})$ and show that $\hat{x}$ is β-close to $x(\hat{\mu})$. Let

$$\hat{\mu} := (\|c\|_{\hat{x}}^*)/\beta,$$

where $\beta < 1$ is a given positive constant (e.g., one can use $\beta = \frac{1}{2}$), and let $\hat{y} := (0, ..., 0, \hat{\mu}\lambda^{-1})^T$. Then define $\hat{s} := c - A^T\hat{y}$. Now note that

$$\begin{aligned} \nabla\Psi(\hat{x}) + \hat{s}/\hat{\mu} &= \lambda^{-1}f + \hat{\mu}^{-1}(c - A^T\hat{y}) \\ &= \lambda^{-1}f + \hat{\mu}^{-1}c - \lambda^{-1}f \\ &= \hat{\mu}^{-1}c. \end{aligned}$$

Therefore $\|\nabla\Psi(\hat{x}) + \hat{s}/\hat{\mu}\|_{\hat{x}}^* = (\|c\|_{\hat{x}}^*)/\hat{\mu} = \beta$, and from Proposition 2.5, $\hat{x}$ is β-close to $x(\hat{\mu})$. Therefore we can initialize the algorithm of this section with $\mu^0 := \hat{\mu}$ and $x^0 := \hat{x}$. It should be pointed out that, in the integer model, the input size of the M and λ and μ^0 can be bounded by $2^{O(L)}$, where L is the input length of the bit representation of the original data, and that as a consequence the input size of the data for problem (P) of (3.10) is bounded by a constant times that of the original data in (3.7) or (3.8).

4 A Potential Reduction Algorithm for Solving (P)

4.1 The Algorithm

In this section, we present a potential function reduction algorithm for solving the linear programming problem (P) that is an extension of the algorithm of Gonzaga [4], see also Ye [24], or [1]. Given a finite lower bound z on z^*, the unknown optimal value of (P), we consider the following potential function minimization problem:

$$\min_{x,z} \phi(x,z) := q\ln(c^T x - z) + \Psi(x) \tag{4.1a}$$

$$\text{s.t.} \quad Ax = b, \tag{4.1b}$$

$$\ell \leq x \leq u, \tag{4.1c}$$

$$z \leq z^*,$$

where q is a given parameter (we will want to set $q = p + \sqrt{p}$), and Ψ is either $\bar{\Psi}$ or $\tilde{\Psi}$.

At the start of an iteration, $\hat{x} \in F^0(P)$ and $\hat{z} \leq z^*$ are given. Let $\tilde{v}$ denote the gradient of $\phi(x,z)$ at $(x,z) = (\hat{x},\hat{z})$ with respect to x, and note that

$$\tilde{v} = \left(\frac{q}{c^T\hat{x} - \hat{z}}\right) c + \nabla\Psi(\hat{x}). \tag{4.2}$$

Let $\bar{d}$ be the projection of $\tilde{v}$ onto the null space of A as given by the solution to (2.19) with $v = \tilde{v}$. This solution $\bar{d}$ is part of the solution $(\bar{d}, \bar{y})$ to the system (2.20) with $v = \tilde{v}$, and is unique (see Theorem 2.1). Just as in [1], the potential function can be reduced by at least a constant amount of $\delta = \frac{1}{6}$ by taking a step in the direction $-\bar{d}$ from the current point $x = \hat{x}$, as long as $\|\bar{d}\|_{\hat{x}}$ is sufficiently large.

Proposition 4.1 *Suppose $\hat{x} \in F^0(P), \hat{z} \leq z^*$, $v = \tilde{v}$ is given by (4.2), and $(\bar{d}, \bar{y})$ is given by (2.20). Then if $\|\bar{d}\|_{\hat{x}} \geq \frac{4}{5}$ and $\alpha \in [0,1)$,*

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

and $x(\alpha)$ satisfies

$$\phi(x(\alpha), \hat{z}) \leq \phi(\hat{x}, \hat{z}) - \frac{4}{5}\alpha + \frac{\alpha^2}{2(1-\alpha)}. \tag{4.3a}$$

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

$$\phi(x(\frac{2}{5}), \hat{z}) \leq \phi(\hat{x}, \hat{z}) - \frac{1}{6}. \tag{4.3b}$$

Proof: That $x(\alpha) \in F^0(P)$ follows as in Proposition 3.1. Also we have

$$
\begin{aligned}
&\phi(x(\alpha), \hat{z}) - \phi(\hat{x}, \hat{z}) \\
&\quad = q \ln\left(1 - \frac{\alpha c^T \bar{d}}{\|\bar{d}\|_{\hat{x}}(c^T\hat{x} - \hat{z})}\right) + \Psi(x(\alpha)) - \Psi(\hat{x}) \\
&\quad \leq -\frac{\alpha q c^T \bar{d}}{\|\bar{d}\|_{\hat{x}}(c^T\hat{x} - \hat{z})} + \Psi(\hat{x} - \alpha\bar{d}/\|\bar{d}\|_{\hat{x}}) - \Psi(\hat{x}) \\
&\quad \text{(from the concavity of the logarithm function)} \\
&\quad \leq -\frac{\alpha q c^T \bar{d}}{\|\bar{d}\|_{\hat{x}}(c^T\hat{x} - \hat{z})} - \frac{\alpha \nabla\Psi(\hat{x})^T \bar{d}}{\|\bar{d}\|_{\hat{x}}} + \frac{\alpha^2}{2(1-\alpha)} \\
&\quad \text{(from Theorem 2.7)} \\
&\quad = \frac{-\alpha}{\|\bar{d}\|_{\hat{x}}}\left(\frac{q}{c^T\hat{x} - \hat{z}}c + \nabla\Psi(\hat{x})\right)^T \bar{d} + \frac{\alpha^2}{2(1-\alpha)} \\
&\quad = \frac{-\alpha}{\|\bar{d}\|_{\hat{x}}}\bar{v}^T\bar{d} + \frac{\alpha^2}{2(1-\alpha)} = -\alpha\|\bar{d}\|_{\hat{x}} + \frac{\alpha^2}{2(1-\alpha)} \\
&\quad \text{(from (2.21) of Theorem 2.1)} \\
&\quad \leq -\frac{4}{5}\alpha + \frac{\alpha^2}{2(1-\alpha)}.
\end{aligned}
$$

This shows (4.3a). Then (4.3b) follows by substituting $\alpha = \frac{2}{5}$. □

Thus from Proposition 4.1, $\phi(\hat{x}, \hat{z})$ can be reduced by at least $\frac{1}{6}$ whenever $\|\bar{d}\|_{\hat{x}} \geq \frac{4}{5}$, by taking a step in the direction $-\bar{d}$.

Now suppose instead that $(\bar{d}, \bar{y})$ is computed and $\|\bar{d}\|_{\hat{x}} < \frac{4}{5}$. Then we can update the lower bound $\hat{z}$ on z^* as follows. Let

$$\hat{\mu} := (c^T\hat{x} - \hat{z})/q, \tag{4.4a}$$

$$\hat{v} := \hat{\mu}\bar{v} = c + \hat{\mu}\nabla\Psi(\hat{x}), \tag{4.4b}$$

$$\hat{y} := \hat{\mu}\bar{y}, \tag{4.4c}$$

and

$$\hat{s} := c - A^T\hat{y}. \tag{4.4d}$$

Then $\hat{d} := P_{\hat{x}}(\hat{v}) = \hat{\mu} P_{\hat{x}}(\bar{v}) = \hat{\mu}\bar{d}$, so that $\|\hat{d}/\hat{\mu}\|_{\hat{x}} = \|\bar{d}\|_{\hat{x}} < \frac{4}{5}$. We can therefore apply the "approximately-centered" Theorem 2.5 again to conclude that

$$(\hat{x} - \ell)^T \hat{s}^+ + (u - \hat{x})^T \hat{s}^- \leq \hat{\mu}(p + \frac{4}{5}\sqrt{p}). \tag{4.5a}$$

Let

$$\begin{aligned} z &:= c^T\hat{x} - (\hat{x} - \ell)^T \hat{s}^+ - (u - \hat{x})^T \hat{s}^- \\ &= b^T \hat{y} + \ell^T \hat{s}^+ - u^T \hat{s}^-. \end{aligned} \tag{4.5b}$$

We have:

Proposition 4.2 *Suppose $q \geq p + \sqrt{p}$, and suppose $\hat{x} \in F^0(P)$, $\hat{z} \leq z^*$, $\bar{v}$ is given by (4.2), and $(\bar{d}, \bar{y})$ is given by (2.20) with $v = \bar{v}$. Let $(\hat{y}, \hat{s})$ and z be given by (4.4) and (4.5). If $\|\bar{d}\|_{\hat{x}} < \frac{4}{5}$, then $(\hat{y}, \hat{s}) \in F^0(D)$, z is a finite lower bound on z^*, and*

$$\phi(\hat{x}, z) \leq \phi(\hat{x}, \hat{z}) - \frac{1}{5\sqrt{p}} \leq \phi(\hat{x}, \hat{z}) - \frac{1}{6}.$$

Proof: That $(\hat{y}, \hat{s}) \in F^0(D)$ follows from Theorem 2.5, and then z is a finite lower bound on z^* by (4.5b). Now

$$\begin{aligned} c^T\hat{x} - z &= (\hat{x} - \ell)^T \hat{s}^+ + (u - \hat{x})^T \hat{s}^- \\ &\leq \hat{\mu}(p + \frac{4}{5}\sqrt{p}) \quad \text{(by (4.5a))} \\ &\leq (c^T\hat{x} - \hat{z})\frac{p + \frac{4}{5}\sqrt{p}}{q} \quad \text{(by (4.4a))}. \end{aligned}$$

This gives

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

where the second to last inequality follows again from the concavity of the logarithm function, and the final inequality derives from $p \geq 1$. □
We summarize Proposition 4.1 and 4.2 as follows:

Remark 4.1 *Assume $q \geq p + \sqrt{p}$. Let $\hat{x} \in F^0(P)$ be given, along with $\hat{z} \leq z^*$. Compute $v = \bar{v}$ from (4.2), and compute $\bar{d}, \bar{y}, \hat{y}, \hat{s}, z$ from (2.20), (4.4) and (4.5). Then:*

(i) If $\|\hat{d}\|_{\hat{x}} \geq \frac{4}{5}$, ϕ can be decreased by $\delta \geq \frac{1}{6}$ by taking a step in x in the direction $-\bar{d}$.
(ii) If $\|\bar{d}\|_{\hat{x}} < \frac{4}{5}$, ϕ can be decreased by $\delta \geq \frac{1}{6}$ by replacing $\hat{z}$ with the new lower bound z.

This leads to the following algorithm for solving (P). Let (A, b, c, ℓ, u) be the data for (P), and let $x^0 \in F^0(P)$ and $z^0 \leq z^*$ be initial values.

Step 0 Set $q = p + \sqrt{p}$,
$k = 0$.
Step 1 Set $\hat{x} = x^k, \hat{z} = z^k$.
Compute $\bar{v}$ from (4.2).
Compute $(\bar{d}, \bar{y})$ from (2.20).
Step 2 (Step in x)
If $\|\bar{d}\|_{\hat{x}} \geq \frac{4}{5}$, let
$\tilde{x} = \hat{x} - \frac{2}{5}\bar{d}/\|\bar{d}\|_{\hat{x}}$.
Set $x^{k+1} = \tilde{x}, z^{k+1} = \hat{z}$, and $k = k + 1$, and go to Step 1.
Step 3 (Update Bound)
If $\|\bar{d}\|_{\hat{x}} < \frac{4}{5}$, let $(\hat{y}, \hat{s}, z)$ be given by (4.4)-(4.5).
Set $x^{k+1} = \hat{x}, z^{k+1} = z$, and $k = k + 1$, and go to Step 1.

Of course, it is possible to make a line-search in Step 2 of the algorithm as long as a potential reduction of at least $1/6$ is attained.

Theorem 4.1 *For $k = 0, 1, \ldots$, the above algorithm satisfies $x^k \in F^0(P), z^k \leq z^*$, and $\phi(x^{k+1}, z^{k+1}) \leq \phi(x^k, z^k) - \frac{1}{6}$.*

Proof: Follows from Proposition 4.1, 4.2, and Remark 4.1. □

4.2 Initialization and Complexity of the Potential Reduction Algorithm

In this subsection, we discuss a way to initialize the potential reduction algorithm and state complexity results. We proceed in a manner almost identical to Section 3.2. Just as in Section 3.2, we suppose our given linear program is $(\tilde{P})$ of (3.7) and that this program satisfies the assumptions of Section 2.1. Then exactly as in Section 3.2, (3.7) can be transformed into the equivalent programs (3.8), (3.9), and finally (P) of (3.10), where

$$x = (x_1, x_2, x_3, x_4, x_5) = \hat{x} = (0, (\ell_2 + u_2)/2, \ell_3 + \lambda e, \lambda, \lambda) \qquad (4.6)$$

is feasible for (P). Furthermore, upon partitioning the rows of A into

$$A = \begin{bmatrix} B \\ f^T \end{bmatrix}$$

we note that $\nabla\Psi(\hat{x}) = (0, 0, \lambda^{-1}e, \lambda^{-1}, \lambda^{-1}) = \lambda^{-1}f = A^T\hat{y}$, where $\hat{y} = (0, ..., 0, \lambda^{-1})^T$. Thus $x = \hat{x}$ solves the optimality conditions for the barrier problem

$$(B) \qquad \begin{array}{rrcl} \min\limits_{x} & \Psi(x) & & \\ \text{s.t.} & Ax & = & b, \\ \ell & \leq \; x & \leq & u, \end{array} \qquad (4.7)$$

and so $\Psi(x) \leq \Psi(\hat{x})$ for any $x \in F^0(P)$. In this way, we see that if the potential reduction algorithm is applied to (P) of (3.10) starting with $x^0 = \hat{x}$ and z^0 is a suitably chosen lower bound on z^*, then from Theorem 4.1,

$$\begin{aligned} q\ln(c^Tx^k - z^k) &= \phi(x^k, z^k) + \Psi(x^k) \\ &\leq -k/6 + \phi(x^0, z^0) + \Psi(x^k) \\ &= -k/6 + q\ln(c^Tx^0 - z^0) + \Psi(x^k) - \Psi(x^0) \\ &\leq -k/6 + q\ln(c^Tx^0 - z^0), \end{aligned}$$

and so

$$c^Tx^k - z^k \leq \exp(-k/6q)(c^Tx^0 - z^0). \qquad (4.8)$$

With $q = p + \sqrt{p}$, (4.8) states that a fixed reduction in the duality gap $(c^Tx - z)$ can be achieved in $O(p)$ iterations. We have thus shown:

Theorem 4.2 *Suppose that the linear programming problem $(\tilde{P})$ of (3.7) is transformed into the linear programming problem (P) of (3.10), and that the potential reduction algorithm is applied to (P) starting at $x^0 = \hat{x}$ of (4.6) and $z^0 \leq z^*$, with $q = p + \sqrt{p}$. Then each (x^k, z^k) satisfies: $x^k \in F^0(P), z^k \leq z^*$, and $c^Tx^k - z^* \leq c^Tx^k - z^k \leq \exp(-k/6q)(c^Tx^0 - z^0)$.* □

One can think of the results of Theorem 4.2 as being driven by three factors: the potential reduction theorem (Theorem 4.1), the construction of (P) in (3.10) so that we have a known feasible solution $\hat{x}$, and the property of (P) in (3.10) that guarantees that $\hat{x}$ is "centered," i.e., that $\hat{x}$ solves the barrier problem (B) of (4.7). In the integer model of computation, where L is the input length of the bit representation of the original linear program $\tilde{P}$, then one can choose *a priori* $M = 2^{O(L)}$, $\lambda = 2^{O(L)}$ and $z^0 = -2^{O(L)}$ in the construction of (P) via (3.10).

It should be noted that in some instances, we may be given a linear programming problem (P) directly, with a point $x^0 \in F^0(P)$ and a lower bound $z^0 \le z^*$. In this case, it is possible to prove the following result:

Theorem 4.3 *Suppose that the linear programming problem (P) of Section 2.1 is given directly, with a point $x^0 \in F^0(P)$ and a lower bound $z^0 \le z^*$. Let (x^k, z^k) be the iterates generated by the potential reduction algorithm applied to (P) starting at (x^0, z^0), with $q = p + \sqrt{p}$. Then $x^k \in F^0(P)$ and $z^k \le z^*$. Also there exist two finite positive constants C_1 and C_2 with the property that for all $k \ge C_1$,*

$$(c^T x^k - z^k) \le C_2 \exp(-k/6q).$$

Finally, in the integer model of computation, if L is the input length of the bit representation of the problem (P), then $C_1 = O(L)$ and $C_2 = 2^{O(L)}$ as long as $\ln(c^T x^0 - z^0) = O(L)$ and $\Psi(x^0) = O(L)$. □

The proof of this result follows from similar arguments appearing in [2] and in [19].

5 Concluding Remarks

5.1 Metrics, Complexity, and Self-Concordance

The choice of the metric in which primal and dual displacements are measured is perhaps the main point of differentiation between the analysis in this paper as compared to the bulk of the literature on interior-point algorithms. This metric is defined for each $\hat{x} \in F^0(P)$ by first constructing the scaling matrix $\hat{\Theta}$ (see (2.13)) and then defining the metric $\|\cdot\|_{\hat{x}}$ by $\|x' - x''\|_{\hat{x}} = \|\hat{\Theta}(x' - x'')\|$, see (2.17).

The essential difference between this metric and the traditional scaling metric lies in the case of two-sided bounded variables. Traditionally the slacks $u_j - \hat{x}_j$ and $\ell_j - \hat{x}_j$ are treated as separate and unrelated slack variables, and hence the displacement in each slack becomes a separate component of the vector of primal displacements; see, for instance, McShane et al. [11]. The net result is that the two slacks each contribute separately to the accounting of inequalities in the complexity analysis of any algorithm, and so this complexity is a function of the total number of (finite) inequalities, which we denote by k. In the analysis in this paper, however, the scaling metric combines the slacks $u_j - \hat{x}_j$ and $\ell_j - \hat{x}_j$ in the definition of the scaling matrix $\hat{\Theta}$ (see (2.13)). The net result is that the two bounds ℓ_j and u_j do not contribute separately to the accounting of inequalities in the complexity analysis, and the complexity is a function of p, the number of variables with finite bounds, which will always be less than or equal to k.

The definition of the scaling matrix $\hat{\Theta}$ (2.13) used to define the displacement metrics is very much related to the theory of self-concordance developed by Nesterov and Nemirovsky [14]. In [14], as in this paper, the use of the square-root of the Hessian of the barrier function to define a metric is a central concept.

5.2 Towards Symmetry and a Primal-Dual Approach

The duality results of linear programming demonstrate that the primal and dual linear programming problems are fundamentally symmetric, in that each problem can be cast in the format of the other. This symmetry has been manifest as well in a number of ways in interior-point algorithms. We call an algorithm symmetric if its iterates are invariant under a reversal of the roles of the primal and dual variables.

Many path-following interior-point algorithms are symmetric algorithms, e.g., those of Kojima et al. [9] and Monteiro and Adler [12]. Many of these algorithms attain the best known complexity bounds in that they achieve fixed improvements in the duality gap in $O(\sqrt{n})$ iterations (where n is the number of inequalities in the primal (or the dual). However, there are also nonsymmetric path-following algorithms that achieve this complexity–see, e.g., Roos and Vial [17] or Tseng [22]. The path-following algorithm presented in Section 3 of this paper is a nonsymmetric algorithm: the Newton step is derived for the primal barrier problem (BP) by optimizing the quadratic approximation to its objective function. The dual iterates are computed at each iteration as part of the projection problem (2.19) and (2.20), but the scaling used in (2.19) and (2.20) is induced only by the primal iterate x. Nevertheless, this algorithm also exhibits the "low complexity" worst case behavior of $O(\sqrt{p})$ iterations for a fixed improvement in the duality gap. A natural question to ask is whether or not there exists a symmetric algorithm for following the central trajectory $(x(\mu), y(\mu), s(\mu))$ of (P). The answer to this question will probably rely on the construction of an appropriate combined primal-dual scaling for use as a metric for measuring primal and dual displacements, that generalizes the metrics used in traditional primal-dual path-following algorithms such as those of Kojima et al. [9] and Monteiro and Adler [12]. We will discuss this further below.

In the arena of potential-reduction algorithms, symmetry has been a more elusive target. Karmarkar's original algorithm [7] made no mention of the dual problem at all (except that it was possibly applied to a combined primal-dual formulation). Todd and Burrell [20] showed how Karmarkar's algorithm produces dual variables for which the duality gap converges to zero. Tanabe [18] and Todd and Ye [21] introduced a symmetric primal-dual potential function that was used by Todd and Ye to produce the first potential-reduction algorithm for linear programming that achieved constant duality gap improvement in $O(\sqrt{n})$ iterations. Their algorithm was symmetric, but required fixed step sizes, not permitting a line-search. Then Ye [24] introduced an algorithm based on the Tanabe-Todd-Ye potential function that allowed line-searches at each iteration. However, this was not a symmetric algorithm, even though it used a symmetric potential function. Kojima et al. [10] constructed the first symmetric "low complexity" (i.e., $O(\sqrt{n})$ iterations for constant improvement) potential reduction algorithm with line-searches for linear programming (see also Gonzaga and Todd [6]). Also, Gonzaga [5] showed that a symmetric potential function was not necessary to achieve low complexity, but the arguments in his analysis are somewhat cumbersome. Nesterov and Nemirovsky [14] extended Ye's algorithm to a very general "conic" setting, using a logarithmically homogeneous barrier function and its conjugate to construct a symmetric primal-dual potential function.

The potential reduction algorithm presented in Section 4 of this paper is a nonsymmetric algorithm, and is modeled after one of the algorithms in [1]. It uses the nonsymmetric potential function $\phi(x, z)$, and it does not achieve as low a complexity bound (requiring $O(p)$ iterations for constant improvement in the duality gap) as the path-following algorithm of Section 3. We suspect that in order to achieve $O(\sqrt{p})$-iteration constant improvement, it will be necessary to use a potential function that incorporates the dual barrier function $\Psi^*(s)$ in some way and that employs the dual variables in a more symmetric manner. As in the case of symmetric path-following algorithms, this will also probably entail the examination of primal-dual metrics for displacements in the primal and dual variables.

We anticipate that symmetric algorithms that extend the algorithms of Section 3 and Section 4 herein will use a scaling metric that is a combination of a primal scaling metric and a dual scaling metric. If $\hat{x} \in F^0(P)$ and $(\hat{y}, \hat{s}) \in F^0(D)$, then $\hat{\Theta}$ of (2.13) is the scaling matrix used to define the primal displacement metric

$$\|x' - x''\|_{\hat{x}} = \|\hat{\Theta}(x' - x'')\|$$

of (2.17), and is based on the primal iterate $\hat{x}$. Borrowing ideas from traditional primal-dual symmetric algorithms such as those of Kojima et al. [9] and Monteiro and Adler [12], we might construct a combined primal-dual scaling matrix of the form $\hat{D}$ where

$$\hat{D} = (\hat{\Theta}\hat{\Sigma}^{-1})^{1/2} \tag{5.1}$$

and $\hat{\Sigma}$ is a dual scaling matrix, and primal displacements $x' - x''$ and dual displacements $(s' - s'')$ would then be measured with the metrics

$$||\hat{D}(x' - x'')|| \text{ and } ||\hat{D}^{-1}(s' - s'')||, \tag{5.2}$$

respectively. We end this remark by suggesting how the dual scaling $\hat{\Sigma}$ might be constructed based on the dual iterates $(\hat{y}, \hat{s})$ and the conjugate of the barrier function $\Psi^*(s)$.

The conjugate function $\Psi^*(s)$ is finite for all $s \in$ rel int S, and we can define $\nabla\Psi^*(s)$ and $\nabla^2\Psi^*(s)$ as follows:

$$[\nabla\Psi^*(s)]_j = \begin{cases} (\psi_j^*)'(s_j) & \text{if} \quad \ell_j > -\infty \text{ or } u_j < +\infty, \\ +\infty & \text{else,} \end{cases} \tag{5.3}$$

$$[\nabla^2\Psi^*(s)]_{ij} = \begin{cases} 0 & \text{if} \quad i \neq j, \\ (\psi_j^*)''(s_j) & \text{if} \quad i = j, \text{ and } \ell_j > -\infty \text{ or } u_j < +\infty, \\ +\infty & \text{else,} \end{cases} \tag{5.4}$$

for $i, j = 1, ..., n$. Then one can verify via the analysis of $\Psi^*(s)$ in the Appendix that $\nabla\Psi^*(s)$ and $\nabla^2\Psi^*(s)$ are well-defined and correspond to the regular gradient and Hessian functions of $\Psi^*(s)$ in the cases when their values are finite. Next consider a value $\mu = \hat{\mu}$ of the barrier parameter, and a feasible solution $(\hat{y}, \hat{s})$ of (BD) with finite objective value, i.e., $b^T\hat{y} - \hat{\mu}\Psi^*(\hat{s}/\hat{\mu}) > -\infty$. Define $\hat{\Sigma}$ by:

$$\hat{\Sigma}^2 = (\hat{\mu})^{-2}\nabla^2\Psi^*(\hat{s}/\hat{\mu}). \tag{5.5}$$

Then it can be shown that $\hat{\Sigma}$ is a diagonal positive definite matrix, and $\hat{\Sigma}^{-1}$ is a positive semi-definite matrix. Let $\hat{\Theta}$ be defined for $\hat{x} \in F^0(P)$ as in (2.13). Then $\hat{\Sigma}$ and $\hat{\Theta}$ can be combined as in (5.1) to give the metrics (5.2). One clear difficulty in this approach is that $\hat{\Sigma}$ depends on the chosen barrier parameter $\hat{\mu}$ when there are two-sided bounded variables. Note that in the standard case $(s \geq 0)$, (5.5) gives $\hat{\Sigma} = \hat{S}^{-1}$ to go with $\hat{\Theta} = \hat{X}^{-1}$. Further development of these ideas is left to future work.

References

[1] 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.

[2] R. M. Freund. A potential reduction algorithm with user-specified Phase I - Phase II balance, for solving a linear program from an infeasible warm start. Technical Report 981, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853, USA, October 1991.

[3] 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.

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

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

[6] C. C. Gonzaga and M. J. Todd. An $O(\sqrt{n}L)$–iteration large–step primal-dual affine algorithm for linear programming. To appear in *SIAM Journal on Optimization 2(3), 1992.*

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

[8] N. K. Karmarkar. Riemannian geometry underlying interior–point methods for linear programming. In J. C. Lagarias and M. J. Todd, editors, *Mathematical Developments Arising from Linear Programming : Proceedings of a Joint Summer Research Conference held at Bowdoin College, Brunswick, Maine, USA, June/July 1988*, volume 114 of *Contemporary Mathematics*, pages 51–75. American Mathematical Society, Providence, Rhode Island, USA, 1990.

[9] M. Kojima, S. Mizuno, and A. Yoshise. A polynomial–time algorithm for a class of linear complementarity problems. *Mathematical Programming*, 44:1–26, 1989.

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

[11] K. A. McShane, C. L. Monma, and D. F. Shanno. An implementation of a primal–dual interior point method for linear programming. *ORSA Journal on Computing*, 1:70–83, 1989.

[12] R. D. C. Monteiro and I. Adler. Interior path following primal–dual algorithms : Part I : Linear programming. *Mathematical Programming*, 44:27–41, 1989.

[13] Y. E. Nesterov. Polynomial–time dual algorithms for linear programming. *Kibernetika*, 25(1):34–40, 1989. Translated in : *Cybernetics*, 25(1):40–49, 1989.

[14] Y. E. Nesterov and A. S. Nemirovsky. *Interior point polynomial methods in convex programming: Theory and Applications.* To appear in Lecture Notes in Mathematics, Springer-Verlag, 1992.

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

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

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

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

[19] M. J. Todd. Combining phase I and phase II in a potential reduction algorithm for linear programming. Technical Report 907, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853, USA, July 1990. To appear in *Mathematical Programming.*

[20] M. J. Todd and B. P. Burrell. An extension of Karmarkar's algorithm for linear programming using dual variables. *Algorithmica*, 1(4):409–424, 1986.

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

[22] P. Tseng. A simple complexity proof for a polynomial–time linear programming algorithm. *Operations Research Letters*, 8:155–159, 1989.

[23] R. J. Vanderbei. Affine–scaling for linear programs with free variables. *Mathematical Programming*, 43:31–44, 1989.

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

A Appendix

A.1 Derivation of the Conjugate Barrier

Here we compute Ψ^* as defined in (2.26) in Section 2.4. Since $\Psi(x) = \sum_j \psi_j(x_j)$, we find that

$$\Psi^*(s) = \sum_j \psi_j^*(s_j), \text{ with } \psi_j^*(\sigma) = \psi^*(\sigma; \ell_j, u_j, \check{x}_j).$$

It is convenient to treat separately the cases of 0-, 1- and 2-sided bounded variables.

If $\lambda = -\infty$, and $\upsilon = +\infty$, then $\bar{\psi}(\xi) \equiv 0$ and $\tilde{\psi}(\xi) \equiv 1$, so that

$$\bar{\psi}^*(\sigma) = \begin{cases} 0 & \text{for} \quad \sigma = 0 \\ +\infty & \text{for} \quad \sigma \neq 0, \end{cases} \tag{A.1a}$$

$$\tilde{\psi}^*(\sigma) = \begin{cases} -1 & \text{for} \quad \sigma = 0 \\ +\infty & \text{for} \quad \sigma \neq 0. \end{cases} \tag{A.1b}$$

If $-\infty < \lambda < \upsilon = +\infty$, then $\bar{\psi}(\xi) = \tilde{\psi}(\xi) = -\ln(\xi - \lambda) + \ln(\check{\xi} - \lambda)$ and an easy computation gives

$$\bar{\psi}^*(\sigma) = \tilde{\psi}^*(\sigma) = \begin{cases} -1 - \lambda\sigma - \ln\sigma - \ln(\check{\xi} - \lambda) & \text{if} \quad \sigma > 0 \\ +\infty & \text{if} \quad \sigma \leq 0. \end{cases} \tag{A.2}$$

In particular, in the standard case we get $\psi^*(\sigma) = -1 - \ln\sigma$ for $\sigma > 0$.

Similarly, if $-\infty = \lambda < \upsilon < +\infty$, we find

$$\bar{\psi}^*(\sigma) = \tilde{\psi}^*(\sigma) = \begin{cases} -1 - \upsilon\sigma - \ln(-\sigma) - \ln(\upsilon - \check{\xi}) & \text{if} \quad \sigma < 0 \\ +\infty & \text{if} \quad \sigma \geq 0. \end{cases} \tag{A.3}$$

Finally, we consider the general 2-sided case. We treat the two cases $\psi = \bar{\psi}$ and $\tilde{\psi}$ separately. In the first case we have by definition

$$\bar{\psi}^*(\sigma) = \sup_\xi\{-\sigma\xi + \ln(\xi - \lambda) + \ln(\upsilon - \xi)\} - \ln(\check{\xi} - \lambda) - \ln(\upsilon - \check{\xi}).$$

The expression inside the braces is strictly convex and maximized by ξ with $-\sigma + (\xi - \lambda)^{-1} - (\upsilon - \xi)^{-1} = 0$. After some algebraic manipulation, this yields

$$\xi = \begin{cases} \rho & \text{if} \quad \sigma = 0 \\ \frac{1+\rho\sigma-\sqrt{1+\nu^2\sigma^2}}{\sigma} & \text{if} \quad \sigma \neq 0 \end{cases} \tag{A.4}$$

with ρ and ν given by (2.8). Substituting this value in our formula for $\bar{\psi}^*$ yields

$$\bar{\psi}^*(\sigma) = -1 - \rho\sigma + \sqrt{1+\nu^2\sigma^2} + \ln\left(\frac{2\nu^2}{1+\sqrt{1+\nu^2\sigma^2}}\right) - \ln(\bar{\xi} - \lambda) - \ln(\upsilon - \bar{\xi}). \tag{A.5}$$

If $\psi = \tilde{\psi}$, we find

$$\begin{aligned}\tilde{\psi}^*(\sigma) &= \sup_{\xi}\{-\sigma\xi - \frac{\min\{\xi-\lambda, \upsilon-\xi\}}{\upsilon} \\ &\quad + \ln(\min\{\xi-\lambda, \upsilon-\xi\})\} - \ln(\min\{\bar{\xi}-\lambda, \upsilon-\bar{\xi}\}).\end{aligned}$$

Again the supremum is attained where the derivative is zero, and this gives

$$\xi = \rho - \frac{\nu^2\sigma}{1+\nu|\sigma|}. \tag{A.6}$$

Substituting this in our expression for $\tilde{\psi}^*$ gives

$$\tilde{\psi}^*(\sigma) = -1 - \rho\sigma + \nu|\sigma| + \ln\left(\frac{\nu}{1+\nu|\sigma|}\right) - \ln(\min\{\bar{\xi}-\lambda, \upsilon-\bar{\xi}\}). \tag{A.7}$$

Note the similarity between (A.5) and (A.7): apart from an additive constant, (A.7) differs from (A.5) only in that $\nu|\sigma|$ replaces $\sqrt{1+\nu^2\sigma^2}$, to which it is asymptotic as $|\sigma| \to \infty$.

Next consider $\mu\Psi^*(s/\mu)$ as $\mu \downarrow 0$. Looking at (A.1)-(A.7), we see that all constant and logarithmic terms vanish in the limit, and we find

$$\mu\psi^*(\sigma/\mu) \to \begin{cases} 0 & \text{if} \quad \lambda = -\infty,\ \upsilon = +\infty \quad \text{and} \quad \sigma = 0 \\ -\lambda\sigma & \text{if} \quad \lambda > -\infty,\ \upsilon = +\infty \quad \text{and} \quad \sigma > 0 \\ -\upsilon\sigma & \text{if} \quad \lambda = -\infty,\ \upsilon < +\infty \quad \text{and} \quad \sigma < 0 \\ -\rho\sigma + \nu|\sigma| & \text{if} \quad \lambda > -\infty,\ \upsilon < +\infty, \end{cases}$$

and $+\infty$ otherwise. Note that $-\rho\sigma + \nu|\sigma| = -\lambda\sigma^+ + \upsilon\sigma^-$. Thus $\mu\Psi^*(s/\mu) \to -\ell^T s^+ + u^T s^-$, and we see that the barrier term in (BD) contains terms like $\ell^T s^+ - u^T s^-$, and converges to it as $\mu \downarrow 0$.

(An expression similar to (A.5) can be obtained from a different perspective. If we had modelled the dual problem with two slack vectors s' and s'', and imposed standard logarithmic barriers on each of them, we would be led to the subproblem

$$\min\{-\lambda\sigma' + \upsilon\sigma'' - \ln\sigma' - \ln\sigma'' : \sigma' - \sigma'' = \sigma\}$$

for finding the "best" s' and s'' corresponding to a given y and $s = c - A^T y$. The solution to this is given by $\sigma', \sigma'' = (1 + \sqrt{1 + \nu^2\sigma^2})/2\nu \pm \sigma/2$, and the resulting objective function is

$$1 - \rho\sigma + \sqrt{1 + \nu^2\sigma^2} + \ln\left(\frac{2\nu^2}{1 + \sqrt{1 + \nu^2\sigma^2}}\right),$$

which differs by a constant from (A.5).)

A.2 Proofs of Theorem 2.6 and Proposition 2.6

Proof of Proposition 2.6: Note that since $\Psi = \bar{\Psi}$ or $\Psi = \tilde{\bar{\Psi}}$ is separable, then (2.38) holds as long as it holds when h and d are unit vectors. Hence we need to show that for each j

$$(1 - |\gamma|(\psi_j''(x_j))^{1/2})^2 \psi_j''(x_j) \le \psi_j''(x_j + \gamma) \le (1 - |\gamma|(\psi_j''(x_j))^{1/2})^{-2} \psi_j''(x_j) \quad \text{(A.8)}$$

if $|\gamma|(\psi_j''(x_j))^{1/2} < 1$. Let us drop the subscripts and use the notation of section 2, and so we need to prove that

$$(1 - |\gamma|(\psi''(\xi))^{1/2})^2 \psi''(\xi) \le \psi''(\xi + \gamma) \le (1 - |\gamma|(\psi''(\xi))^{1/2})^{-2} \psi''(\xi) \quad \text{(A.9)}$$

if $|\gamma|(\psi''(\xi))^{1/2} < 1$.

We first prove (A.9) when $\Psi = \bar{\Psi}$, so that $\psi = \bar{\psi}$. Then if $\lambda = -\infty$ and $v = +\infty, \psi''(\xi) = 0$ for all ξ, so (A.9) holds trivially. Next suppose $\lambda > -\infty$ and $v = +\infty$. If $1 > |\gamma|(\psi''(\xi))^{1/2} = \frac{|\gamma|}{\xi - \lambda}$, then

$$\frac{1}{\xi + \gamma - \lambda} \le \frac{1}{\xi - |\gamma| - \lambda} = \left(1 - \frac{|\gamma|}{\xi - \lambda}\right)^{-1} \left(\frac{1}{\xi - \lambda}\right), \quad \text{(A.10)}$$

and squaring both sides gives the right-most inequality of (A.9). Also,

$$\frac{1}{\xi + \gamma - \lambda} \ge \frac{1}{\xi + |\gamma| - \lambda} = \left(1 + \frac{|\gamma|}{\xi - \lambda}\right)^{-1} \left(\frac{1}{\xi - \lambda}\right) \ge \left(1 - \frac{|\gamma|}{\xi - \lambda}\right)\left(\frac{1}{\xi - \lambda}\right), \quad \text{(A.11)}$$

and squaring both sides gives the left-most inequality of (A.9). Next, suppose that $\lambda = -\infty$ and $v < +\infty$. This case is symmetric to the previous case, and parallel arguments apply.

Finally, suppose that $\lambda > -\infty$ and $v < +\infty$. Then ψ is the sum of two one-sided barriers, each of which we have shown is self-concordant. However, it is straightforward to show that the sum of two self-concordant functions is also self-concordant (this follows directly from the definition (2.38)), and hence ψ is self-concordant.

Now, suppose $\Psi = \bar{\Psi}$, so that $\psi = \bar{\psi}$. Unless $-\infty < \lambda$ and $\upsilon < +\infty$, $\bar{\psi}$ and $\tilde{\psi}$ have the same derivatives, and so the self-concordance of $\bar{\psi}$ follows just as in the case when $\psi = \bar{\psi}$. It only remains to show (A.9) holds when $-\infty < \lambda < \upsilon < +\infty$.

Without loss of generality, we assume that $\xi \leq \rho$, so $\psi''(\xi) = (\xi - \lambda)^{-2}$. If also $\xi + \gamma \leq \rho$, then (A.9) reduces to (A.10) and (A.11), which we have already established. So suppose $\xi + \gamma > \rho$ so that $\psi''(\xi + \gamma) = (\upsilon - \xi - \gamma)^{-2}$. Then the left-hand inequality of (A.9) holds, since $(\upsilon - \xi - \gamma)^{-2} \geq (\xi + \gamma - \lambda)^{-2}$, for which we already have the required bound. The right-hand inequality becomes

$$\frac{1}{(\upsilon - \xi - \gamma)^2} \leq \left(1 - \frac{|\gamma|}{\xi - \lambda}\right)^{-2} \frac{1}{(\xi - \lambda)^2} = \frac{1}{(\xi - \lambda - |\gamma|)^2},$$

and this follows because $\xi - \lambda - |\gamma| > 0$ ($|\gamma|(\psi''(\xi))^{1/2} < 1$), and $\xi - \lambda - |\gamma| \leq \upsilon - \xi - \gamma$ ($\xi \leq \rho$). This completes the proof. □

As a stepping stone toward the proof of Theorem 2.6, we first prove:

Lemma A.1 *Suppose $x \in F^0(P)$ and $\hat{x} \in F^0(P)$ and let $\bar{d} = x - \hat{x}$. Let θ_j and $\hat{\theta}_j$ denote the j^{th} diagonal entries of Θ (with $\Theta^2 = \nabla^2\Psi(x)$) and $\hat{\Theta}$ (with $\hat{\Theta}^2 = \nabla^2\Psi(\hat{x})$), respectively. Then for each $j = 1, ..., n$,*

$$|\theta_j^{-1}(\psi_j'(x_j) - \psi_j'(\hat{x}_j) - \hat{\theta}_j^2\bar{d}_j)| \leq \hat{\theta}_j^2\bar{d}_j^2.$$

Proof: We write ξ for x_j, $\hat{\xi}$ for $\hat{x}_j$, λ for ℓ_j, υ for u_j, $\bar{\delta}$ for $\bar{d}_j$, and suppress the subscripts. Hence we wish to show

$$|\theta^{-1}(\psi'(\xi) - \psi'(\hat{\xi}) - \hat{\theta}^2\bar{\delta})| \leq \hat{\theta}^2\bar{\delta}^2. \tag{A.12}$$

We consider the various possibilities for the bounds separately.

Case 1. x_j is free. Then $\hat{\theta}$ is zero, as are $\psi'(\hat{\xi})$ and $\psi'(\xi)$, so both sides of (A.12) are zero.

Case 2. $-\infty < \lambda < \upsilon = +\infty$. Then $\theta = (\xi - \lambda)^{-1} = -\psi'(\xi)$, $\hat{\theta} = (\hat{\xi} - \lambda)^{-1} = -\psi'(\hat{\xi})$, so the left-hand side of (A.12) is

$$\begin{aligned} &|(\xi - \lambda)(-(\hat{\xi} - \lambda)^{-2}\bar{\delta} + (\hat{\xi} - \lambda)^{-1} - (\xi - \lambda)^{-1})| \\ = \ &|(\xi - \lambda)(-(\hat{\xi} - \lambda)^{-2}\bar{\delta} + (\hat{\xi} - \lambda)^{-1}(\xi - \lambda)^{-1}\bar{\delta})| \\ &\text{(using } (\xi - \lambda) - (\hat{\xi} - \lambda) = \xi - \hat{\xi} = \bar{\delta}) \\ = \ &|(\hat{\xi} - \lambda)^{-2}\bar{\delta}(-(\xi - \lambda) + (\hat{\xi} - \lambda))| \\ = \ &|-(\hat{\xi} - \lambda)^{-2}\bar{\delta}^2|, \end{aligned} \tag{A.13}$$

which is $\hat{\theta}^2\bar{\delta}^2$.

Case 3. $-\infty = \lambda < v < +\infty$. This case is completely symmetric to case 2.

Case 4. $-\infty < \lambda < v < +\infty$. In this case we have to distinguish whether Ψ is $\bar{\Psi}$ or $\tilde{\Psi}$. In the former case,

$$\begin{aligned}
\theta^{-1} &= ((\xi-\lambda)^{-2}+(v-\xi)^{-2})^{-1/2} < \min\{\xi-\lambda, v-\xi\} \\
\hat{\theta}^2 &= (\hat{\xi}-\lambda)^{-2}+(v-\hat{\xi})^{-2}, \\
\psi'(\hat{\xi}) &= -(\hat{\xi}-\lambda)^{-1}+(v-\hat{\xi})^{-1}, \quad \text{and} \\
\psi'(\xi) &= -(\xi-\lambda)^{-1}+(v-\xi)^{-1}.
\end{aligned}$$

Hence the left-hand side of (A.12) is

$$\begin{aligned}
&|\theta^{-1}(-((\hat{\xi}-\lambda)^{-2}+(v-\hat{\xi})^{-2})\bar{\delta}+(\hat{\xi}-\lambda)^{-1}-(v-\hat{\xi})^{-1} \\
&- \quad (\xi-\lambda)^{-1}+(v-\xi)^{-1})| \\
&\leq \quad |(\xi-\lambda)(-(\hat{\xi}-\lambda)^{-2}\bar{\delta}+(\hat{\xi}-\lambda)^{-1}-(\xi-\lambda)^{-1})| \\
&+ \quad |(v-\xi)(-(v-\hat{\xi})^{-2}\bar{\delta}-(v-\hat{\xi})^{-1}+(v-\xi)^{-1}|,
\end{aligned}$$

where we have used the over-estimate $\xi-\lambda$ for θ^{-1} in the first term and the over-estimate $v-\xi$ in the second. Now by (A.13) in case 2, the first term is $|-(\hat{\xi}-\lambda)^{-2}\bar{\delta}^2|$, and symmetrically the second term is $|-(v-\hat{\xi})^{-2}\bar{\delta}^2|$, so the left-hand side of (A.12) is bounded by

$$((\hat{\xi}-\lambda)^{-2}+(v-\hat{\xi})^{-2})\bar{\delta}^2 = \hat{\theta}^2\bar{\delta}^2$$

as desired.

Finally we suppose Ψ is $\tilde{\Psi}$. Without loss of generality, $\hat{\xi}$ is closer to λ than to v (i.e. $\hat{\xi} \leq \rho$). If also $\xi \leq \rho$, then both θ and $\hat{\theta}$ are as in case 2, and both $\psi'(\xi)$ and $\psi'(\hat{\xi})$ differ by the same constant $1/\nu$ from their values in case 2 (see Remark 2.1), so the proof of that case applies.

It only remains to assume $\hat{\xi} \leq \rho < \xi$. Then

$$\begin{aligned}
\theta &= (v-\xi)^{-1}, \\
\hat{\theta} &= (\hat{\xi}-\lambda)^{-1}, \\
\psi'(\hat{\xi}) &= (\hat{\xi}-\rho)\nu^{-1}(\hat{\xi}-\lambda)^{-1}, \quad \text{and} \\
\psi'(\xi) &= (\xi-\rho)\nu^{-1}(v-\xi)^{-1} \\
&= (\xi-\rho)\nu^{-1}((\xi-\lambda)^{-1}+((v-\xi)^{-1}-(\xi-\lambda)^{-1})).
\end{aligned}$$

Thus the left-hand side of (A.12) is

$$
\begin{aligned}
& |(v-\xi)[-(\hat{\xi}-\lambda)^{-2}\bar{\delta} \\
& + \{-(\hat{\xi}-\rho)\nu^{-1}(\hat{\xi}-\lambda)^{-1}+(\xi-\rho)\nu^{-1}(\xi-\lambda)^{-1}\} \\
& + (\xi-\rho)\nu^{-1}((v-\xi)^{-1}-(\xi-\lambda)^{-1})]| \\
& = |(v-\xi)[-(\hat{\xi}-\lambda)^{-2}\bar{\delta}+\{(\hat{\xi}-\lambda)^{-1}-(\xi-\lambda)^{-1}\} \\
& + 2(v-\xi)^{-1}(\xi-\lambda)^{-1}\nu^{-1}(\xi-\rho)^{2}]| \\
& = |-(v-\xi)(\xi-\lambda)^{-1}(\hat{\xi}-\lambda)^{-2}\bar{\delta}^{2})+2(\xi-\lambda)^{-1}\nu^{-1}(\xi-\rho)^{2}|,
\end{aligned}
$$

where we have used (A.13) to get the first term. Now the expression within the absolute value signs is either negative or nonnegative. If negative, its absolute value is at most

$$
\begin{aligned}
(v-\xi)(\xi-\lambda)^{-1}(\hat{\xi}-\lambda)^{-2}\bar{\delta}^{2} &\leq (\hat{\xi}-\lambda)^{-2}\bar{\delta}^{2} \\
&= \hat{\theta}^{2}\bar{\delta}^{2}.
\end{aligned}
$$

Finally, suppose it is nonnegative. Then its absolute value is

$$
2(\xi-\lambda)^{-1}\nu\nu^{-2}(\xi-\rho)^{2}-(v-\xi)(\xi-\lambda)^{-1}(\hat{\xi}-\lambda)^{-2}\bar{\delta}^{2}. \tag{A.14}
$$

But $\hat{\xi} \leq \rho$ implies $\hat{\xi}-\lambda \leq \nu$, so $\nu^{-2} \leq (\hat{\xi}-\lambda)^{-2}$, and $\xi-\rho \leq \xi-\hat{\xi} = \bar{\delta}$. Hence (A.14) is bounded by

$$
(\hat{\xi}-\lambda)^{-2}\bar{\delta}^{2}(\xi-\lambda)^{-1}(2\nu-(v-\xi)) = (\hat{\xi}-\lambda)^{-2}\bar{\delta}^{2}(\xi-\lambda)^{-1}(\xi-\lambda) = \hat{\theta}^{2}\bar{\delta}^{2}.
$$

This completes the proof. □

Proof of Theorem 2.6: The proof is a consequence of Lemma A.1. To prove (2.35) note that

$$
\begin{aligned}
\|\nabla\Psi(x)-\nabla\Psi(\hat{x})-\hat{\Theta}^{2}\bar{d}\|_{x}^{*} &= \|\Theta^{-1}(\nabla\Psi(x)-\nabla\Psi(\hat{x})-\hat{\Theta}^{2}\bar{d})\| \\
&\leq \|\Theta^{-1}(\nabla\Psi(x)-\nabla\Psi(\hat{x})-\hat{\Theta}^{2}\bar{d})\|_{1} \\
&\leq \sum_{j}\hat{\theta}_{j}^{2}\bar{d}_{j}^{2} = \|\hat{\Theta}\bar{d}\|^{2} = \|\bar{d}\|_{\hat{x}}^{2},
\end{aligned}
$$

where the second inequality follows from Lemma A.1. □