Quadratic Convergence of Potential-Reduction Methods for Degenerate Problems

Reha H. Tütüncü *

December 2, 1998

Abstract

Global and local convergence properties of a primal-dual interior-point *pure* potential-reduction algorithm for linear programming problems is analyzed. This algorithm is a primal-dual variant of the Iri-Imai method and uses modified Newton search directions to minimize the Tanabe-Todd-Ye (TTY) potential function. A polynomial time complexity for the method is demonstrated. Furthermore, this method is shown to have a unique accumulation point even for degenerate problems and to have Q-quadratic convergence to this point by an appropriate choice of the step-sizes. This is, to the best of our knowledge, the first superlinear convergence result on degenerate linear programs for primal-dual interior-point algorithms that do not follow the central path.

*Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA (`reha+@andrew.cmu.edu`). Research supported in part by NSF through grant DMS-9706950 and a Faculty Development Grant from Carnegie Mellon University.

1 Introduction

In this article, we study the solution of linear programming problems using interior-point potential-reduction algorithms. We consider linear programs in the standard form:

$$(LP) \qquad \begin{array}{rcl} \min_x & c^T x & \\ & Ax & = & b \\ & x & \geq & 0, \end{array} \tag{1}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ are given, and $x \in \mathbb{R}^n$. The matrix A is assumed to have full row rank without loss of generality.

One of the most desirable properties of an optimization algorithm is its ability to converge fast to a solution once the iterates are sufficiently close to such a point. The quadratic convergence of Newton's method is a well-known example of this behavior. Path-following variants of interior-point methods for linear programming enjoy this feature. These variants generate iterates that are restricted to remain in some neighborhood of the so-called *central path* to reach the optimal solution. While the proofs of superlinear convergence results for path-following algorithms invariably and explicitly use these neighborhood restrictions, such restrictions are routinely ignored in practical implementations of these algorithms without destroying the superlinear convergence property. This confusing fact raises the following fundamental question addressed in this article: Can one have a superlinearly convergent interior-point algorithm without restricting the iterates to any wide or narrow neighborhood of the central path?

We emphasize that our focus is on degenerate problems. For these problems the Jacobian of the nonlinear system we would like to solve is singular at the solution set. Therefore, although the algorithm we will present can be viewed as a variant of Newton's method for the solution of this system, Kantarovich type of convergence analysis is not applicable. Path-following methods circumvent this difficulty by restricting their iterates to the central path; see Chapter 7 in [13] for a detailed discussion of this phenomenon. In this paper, for the first time, we develop an algorithm and a convergence theory for the solution of degenerate linear programs that establishes quadratic convergence without any path-following restrictions. We achieve this result using a *potential-reduction* algorithm, which we describe below.

One of the most useful theoretical and algorithmic tools in the development and study of interior-point methods for optimization problems is the concept of *potential functions*, i.e., functions that measure the quality of a trial solution of the problem. Such functions often balance measures of proximity to the set of optimal solutions, proximity to the feasible set in the case of infeasible-interior-points, and a measure of centrality within the feasible region. Potential functions are chosen such that one approaches an optimal solution of the problem by reducing the potential function. Methods using this strategy are called *potential-reduction* algorithms. The algorithm presented in this article is a potential-reduction algorithm that uses the so-called Tanabe-Todd-Ye potential function [7, 9]. Elaborate discussions of potential-reduction algorithms can be found in two recent surveys by Anstreicher and Todd [1, 8].

The present work simultaneously improves and generalizes the results of three earlier papers. The first one of these papers is by the present author [12], where a potential-reduction algorithm with polynomial and quadratic convergence for nondegenerate problems is developed. We use a similar algorithm here but the convergence analysis is significantly more involved in the degenerate case. The algorithm in [12] can be viewed as a primal-dual variant of the method of Iri and Imai [3], which is a damped Newton's method to minimize the multiplicative analogue of the potential function introduced by Karmarkar [5]. We also generalize the results of a paper by Iri [4], where he proved an $\mathcal{O}(nL)$ iteration complexity for the Iri-Imai algorithm. Iri-Imai algorithm assumes that an exact line search is employed along the search directions. Further, it is assumed that the optimal value of the linear programming problem in question is known and equals zero. We remove this assumption in our primal-dual set-up and also do not assume that exact line searches are used. Yet, we prove that our algorithm has the same worst-case complexity as the Iri-Imai method. The last paper that the current article generalizes is by Tsuchiya [11], where he proves that the Iri-Imai method enjoys quadratic convergence even for degenerate problems, provided that one uses exact line searches to minimize the potential function along the generated directions. We prove a Q-quadratic convergence result for our algorithm without using Tsuchiya's assumptions.

The paper is organized as follows: After this introduction, in Section 2, we review potential-reduction algorithms and summarize some of the results from [12] that will be useful in the remainder of the paper. These results include the convexity of the multiplicative primal-dual potential function with a large enough parameter and an explicit formula for the search directions to be used in the algorithm. Global convergence property of the algorithm is analyzed in Section 3 where a polynomial iteration complexity is demonstrated. Section 4 is devoted to the analysis of the asymptotic behavior of our search directions and iterates as well as the demonstration of the uniqueness of accumulation points of the algorithm.

After proving our quadratic convergence results in Section 5, we conclude in Section 6. Section 7 contains the lengthy proof of a key lemma.

The notation used here is mostly standard. $\|\cdot\|$ (without a subscript) denotes the 2-norm of real vectors; we will use subscripts with 1-norms and ∞-norms. We use subscripts to denote components of vectors or matrices, and superscripts to denote iteration indices. We also use standard order notation. For sequences $\{x^k\}$ and $\{y^k\}$ of real numbers with $y^k > 0$, $x^k = \mathcal{O}(y^k)$ means that the sequence $\{\frac{x^k}{y^k}\}$ is bounded above by a number independent of k. If $x^k > 0$ also, then $x^k = \Theta(y^k)$ means that $x^k = \mathcal{O}(y^k)$ and $y^k = \mathcal{O}(x^k)$. If $\{x^k\}$ is a sequence of matrices, $x^k = \mathcal{O}(y^k)$ and $x^k = \Theta(y^k)$ mean that $x^k_{ij} = \mathcal{O}(y^k)$ and $x^k_{ij} = \Theta(y^k)$, respectively, for each i and j. We will denote primal-only and primal-dual potential functions using lower-case and upper-case letters, respectively. For an n-dimensional vector x, the corresponding capital letter X denotes the $n \times n$ diagonal matrix with $X_{ii} \equiv x_i$. Section 4.1 introduces the notation employed in Sections 4 and 5.

2 A Primal-Dual Variant of the Iri-Imai Method

Modern developments in the interior-point methods theory started with Karmarkar's 1984 paper [5], where he introduced a potential function to measure the quality of different feasible points of a linear programming problem. He showed that for linear programs given in a particular form, this function tends to $-\infty$ only along the sequences that approach an optimal solution of the problem. Therefore, a search for an optimal solution can be performed by minimizing this function. The potential function introduced by Karmarkar is given below:

$$\phi_\rho(x) := \rho \ln(c^T x) - \sum_{i=1}^{n} \ln x_i \tag{2}$$

with $\rho = n + 1$. Karmarkar assumes that the problem (1) is given in a non-standard special form, and the optimal objective value is *a priori* known to be equal to zero. With these assumptions, he shows that $\phi_\rho(x)$ can be reduced by at least a constant amount from any feasible interior-point and thus proves that his algorithm converges to an optimal solution in polynomial time.

Karmarkar's function is not convex. Thus, an algorithm using Newton type search directions may not converge to a global minimizer of the function. Instead, Iri and Imai [3] consider the multiplicative analogue of the function ϕ:

$$f_\rho(x) := \frac{(c^T x)^\rho}{\prod_{i=1}^n x_i}. \tag{3}$$

As it turns out, this function is strictly convex in the relative interior of the primal feasible region if the parameter $\rho \geq n+1$. Iri and Imai apply Newton's method for the minimization of this strictly convex function and obtain quadratic convergence for nondegenerate problems when an exact line search is employed [3]. We mimic this approach in a primal-dual setting.

Primal-dual interior-point methods generate solutions for both the problem (1), called the *primal* problem, and a related problem called its *dual*. These methods emerged as the most successful variants of interior-point methods and also have an elegant complexity theory [13]. The dual of the linear programming problem given in (1) is:

$$\begin{array}{llrcrcl} (LD) & \max_{y,s} & b^T y & & & & \\ & & A^T y & + & s & = & c \\ & & & & s & \geq & 0, \end{array} \tag{4}$$

where $y \in I\!R^m$ and $s \in I\!R^n$.

Let G^T be a null-space basis matrix for A, that is, G is an $(n-m) \times n$ matrix with rank $n-m$ and it satisfies $AG^T = 0$, $GA^T = 0$. Using this matrix and a vector $d \in I\!R^n$ satisfying $Ad = b$, one can rewrite (4) in a form that is identical to (1):

$$\begin{array}{llrcl} (LD') & \min_s & d^T s & & \\ & & Gs & = & Gc \\ & & s & \geq & 0. \end{array} \tag{5}$$

This form of the dual was discussed, among others, by Gonzaga [2]. We prefer this form of the dual since it allows us to reproduce the results we develop for primal variables in the context of dual variables with no additional effort.

Let $\mathcal{F}$ and $\mathcal{F}^0$ denote the primal-dual feasible region and its relative interior:

$$\begin{array}{rcl} \mathcal{F} & := & \{(x,s) : Ax = b, Gs = Gc, (x,s) \geq 0\} \\ \mathcal{F}^0 & := & \{(x,s) : Ax = b, Gs = Gc, (x,s) > 0\} \end{array}$$

We will assume that $\mathcal{F}^0$ is non-empty and a point $(x^0, s^0) \in \mathcal{F}^0$ is available. This assumption is not restrictive; any LP can be embedded in an artificial problem with a known point in the relative interior of its feasible region; see, e.g., Ye, Todd, and Mizuno [14]. See also the related remarks in Section 6. Furthermore, certain solutions to this artificial problem will either give the optimal solution to the original LP, or reveal that it is either infeasible or unbounded. A key consequence of our assumption on the existence of a strictly feasible primal-dual pair of solutions is that the optimal solution set Ω defined below is nonempty and bounded; see, e.g., [13].

$$\Omega \;\;:=\;\; \{(x,s) \in \mathcal{F} : x^T s = 0\} \tag{6}$$

A primal-dual variant of Karmarkar's potential function was introduced by Tanabe [7], and Todd and Ye [9] independently and has been a useful tool in the construction and analysis of efficient interior-point algorithms for linear programming and linear complementarity problems. For the primal-dual pair of problems (1) and (5), this function is defined as

$$\Phi_\rho(x,s) \;\;:=\;\; \rho \ln(x^T s) - \sum_{i=1}^{n} \ln(x_i s_i), \text{ for every } (x,s) > 0. \tag{7}$$

Using a primal-dual update, Kojima, Mizuno, and Yoshise showed that when $\rho \geq n + \sqrt{n}$, Tanabe-Todd-Ye (TTY) potential function can be reduced by at least 0.2 from any feasible point (x,y,s) with $(x,s) > 0$ [6]. This guaranteed constant reduction in the potential function leads to an algorithm with $\mathcal{O}((\rho - n)\ln\frac{1}{\varepsilon})$ complexity. However, the Kojima-Mizuno-Yoshise algorithm cannot have better than linear convergence rate, since the search directions they use have a positive centering component that does not diminish as iterates approach a solution. We evaluate the gradient and the Hessian of this function for future reference:

$$\nabla \Phi_\rho(x,s) \;=\; \begin{bmatrix} \frac{\rho}{x^T s} s - x^{-1} \\ \frac{\rho}{x^T s} x - s^{-1} \end{bmatrix}, \tag{8}$$

$$\nabla^2 \Phi_\rho(x,s) \;=\; \begin{bmatrix} X^{-2} & \\ & S^{-2} \end{bmatrix} + \frac{\rho}{x^T s}\begin{bmatrix} & I \\ I & \end{bmatrix} - \frac{\rho}{(x^T s)^2}\begin{bmatrix} s \\ x \end{bmatrix}\begin{bmatrix} s \\ x \end{bmatrix}^T. \tag{9}$$

In the vicinity of an optimal solution to the linear programming problem under consideration, the potential function Φ_ρ tends to $-\infty$, and for faster convergence one needs to proceed along directions of negative curvature. In our previous work [12], we established that such directions can be obtained for nondegenerate problems using a primal-dual variant of Iri and Imai's strategy. Next, we outline this strategy and also state some of the results from [12] which will be useful later in our analysis.

Consider the multiplicative analogue of the Tanabe-Todd-Ye primal-dual potential function:

$$F_\rho(x,s) \quad := \quad \frac{(x^T s)^\rho}{\prod_{i=1}^n x_i s_i} = \exp\{\Phi_\rho(x,s)\}, \text{ for every } (x,s) > 0. \tag{10}$$

Since $F_\rho(x,s)$ is a monotone transformation of $\Phi_\rho(x,s)$, minimizing one of these functions is equivalent to minimizing the other. Along a sequence of points (x^k, s^k) where $\Phi_\rho(x^k, s^k)$ tends to $-\infty$, $F_\rho(x^k, s^k)$ tends to zero. Note the following identities:

$$\nabla F_\rho(x,s) \quad = \quad F_\rho(x,s)\nabla\Phi_\rho(x,s) \tag{11}$$

$$\nabla^2 F_\rho(x,s) \quad = \quad F_\rho(x,s)\left(\nabla^2\Phi_\rho(x,s) + \nabla\Phi_\rho(x,s)\nabla\Phi_\rho(x,s)^T\right). \tag{12}$$

Our motivation for considering $F_\rho(x,s)$ is the following result:

Theorem 2.1 (Theorem 3.1, [12]) *The function $F_\rho(x,s)$ is strictly convex on the set $\mathcal{F}^0$ if $\rho \geq 2n+1$.* □

In view of Theorem 2.1, we will assume that $\rho \geq 2n+1$ in the rest of this paper. Let d_{RN} be the reduced Newton direction for $F_\rho(x,s)$ at some strictly feasible point (x,s). Note that for feasibility, a search direction must satisfy

$$\begin{bmatrix} A & \\ & G \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta s \end{bmatrix} = 0. \tag{13}$$

Therefore, d_{RN} can be written as the unique solution of the following system:

$$\left(Z^T\nabla^2 F_\rho(x,s)Z\right) d_{RN} = -Z^T\nabla F_\rho(x,s) \tag{14}$$

where Z be a null space basis for the constraint matrix of (13). In this paper, we will use the following convenient form for Z:

$$Z \quad = \quad \begin{bmatrix} G^T & \\ & A^T \end{bmatrix} \tag{15}$$

In view of the identities (11) and (12), (14) is equivalent to

$$\left(Z^T\left[\nabla^2\Phi_\rho(x,s) + \nabla\Phi_\rho(x,s)\nabla^T\Phi_\rho(x,s)\right] Z\right) d_{RN} = -Z^T\nabla\Phi_\rho(x,s). \tag{16}$$

Since $\nabla^2 F_\rho(x,s)$ is positive definite on $\mathcal{F}^0$, the direction Zd_{RN} is guaranteed to be a descent direction for F_ρ and, consequently, for Φ_ρ. Also note that the direction d_{RN} that solves (16) is a scalar multiple of the reduced Newton direction to minimize the TTY potential function Φ_ρ. More precisely, if $\widetilde{d_{RN}}$ denotes the solution of the system

$$\left(Z^T \nabla^2 \Phi_\rho(x,s) Z\right) \widetilde{d_{RN}} = -Z^T \nabla \Phi_\rho(x,s) \tag{17}$$

then the following relationship holds:

$$\widetilde{d_{RN}} = \frac{1}{1 + d_{RN}^T Z^T \nabla^2 \Phi_\rho(x,s) Z d_{RN}} d_{RN}. \tag{18}$$

Since the Hessian matrix $\nabla^2 \Phi_\rho(x,s)$ may be indefinite, it may happen that the scalar in front of d_{RN} in (18) is negative so that d_{RN} and $\widetilde{d_{RN}}$ point in opposite directions. Since d_{RN} leads to a guaranteed descent direction, we will focus on it in the remainder of this paper.

The left hand side matrix of the equation (16) is the sum of a block diagonal matrix and a rank-two matrix. Therefore, this system can be solved efficiently using the Sherman-Morrison-Woodbury formula. Although the derivation is somewhat tedious, the resulting directions are relatively simple. Below, these directions will be presented formally, but first we introduce some notation. We will make use of two orthogonal projection matrices:

$$\Xi = \Xi(x) := X^{-1} G^T (G X^{-2} G^T)^{-1} G X^{-1}, \tag{19}$$

and

$$\Sigma = \Sigma(s) := S^{-1} A^T (A S^{-2} A^T)^{-1} A S^{-1}. \tag{20}$$

We will suppress the dependence of Ξ and Σ to (x,s) in our notation. We note that Ξ and Σ are orthogonal projection matrices into the range spaces of $X^{-1}G^T$ and $S^{-1}A^T$, respectively. These range spaces are the same as the null spaces of AX and GS, respectively. Consequently,

$$\Xi = I - X A^T (A X^2 A^T)^{-1} A X, \tag{21}$$

and

$$\Sigma = I - S G^T (G S^2 G^T)^{-1} G S. \tag{22}$$

We will denote the normalized complementarity vector with ν:

$$\nu \quad := \quad \frac{XSe}{x^T s}. \tag{23}$$

The analysis of the search direction defined in (16) relies heavily on an accurate estimation of the matrices Ξ and Σ, and of the orthogonal projections of the vectors e and ν using these matrices. To this end, we define

$$\begin{aligned}
\beta_1 &:= \nu^T(\Xi+\Sigma)\,\nu, && (24)\\
\beta_2 &:= \nu^T(\Xi+\Sigma)\,e, && (25)\\
\beta_3 &:= e^T(\Xi+\Sigma)\,e, && (26)\\
\Delta &:= (\rho\beta_1-1)(\rho-\beta_3-1)+\rho(1-\beta_2)^2. && (27)
\end{aligned}$$

Above, Δ is the determinant of the 2×2 matrix encountered when using the Sherman-Morrison-Woodbury formula to invert the left-hand-side matrix in (16). Note that the scalars defined above depend on the current iterate (x,s), but we will suppress this dependence in our notation.

Proposition 2.1 (Prop. 3.1, [12]) *Let d_{RN} be the unique solution to (16). Then,*

$$d_{RN} \quad = \quad -\frac{1}{\Delta}\begin{bmatrix}(GX^{-2}G^T)^{-1}GX^{-1}\\ (AS^{-2}A^T)^{-1}AS^{-1}\end{bmatrix}\left(\rho(1-\beta_2)\nu+(\rho\beta_1-1)e\right). \tag{28}$$

Now let $(\Delta x^T,\Delta s^T)^T = Zd_{RN}$. Then, Δx and Δs satisfy the following equations:

$$X^{-1}\Delta x = -\frac{\rho(1-\beta_2)}{\Delta}\Xi\nu - \frac{\rho\beta_1-1}{\Delta}\Xi e, \tag{29}$$

$$S^{-1}\Delta s = -\frac{\rho(1-\beta_2)}{\Delta}\Sigma\nu - \frac{\rho\beta_1-1}{\Delta}\Sigma e. \tag{30}$$

□

In relation to the remarks following equation (16), we note that $\widetilde{d_{RN}} = \frac{\Delta}{\rho\beta_1-1}d_{RN}$, so that if $(\widetilde{\Delta x}^T,\widetilde{\Delta s}^T)^T = Z\widetilde{d_{RN}}$ we then have:

$$X^{-1}\widetilde{\Delta x} = -\frac{\rho(1-\beta_2)}{\rho\beta_1-1}\Xi\nu - \Xi e, \tag{31}$$

$$S^{-1}\widetilde{\Delta s} = -\frac{\rho(1-\beta_2)}{\rho\beta_1-1}\Sigma\nu - \Sigma e. \tag{32}$$

Note that the bulk of the work in computing Δx and Δs from (29) and (30) is in the factorization of the matrices AX^2A^T and $AS^{-2}A^T$. This is roughly twice the work required for an iteration of most interior-point methods. However, most path-following algorithms with quadratic convergence on degenerate problems require that each predictor step is followed by a corrector step to maintain the proximity to the central path. The combined progress of the predictor and corrector steps of a path-following algorithm is comparable to a single step of ours, where we measure the progress in terms of the total complementarity, x^Ts. Therefore, the ratio of effort to progress for our algorithm is similar to the same ratio for predictor-corrector path-following algorithms.

Proposition 2.2 (Prop. 3.2, [12]) *Let Δx and Δs be defined by (29) and (30), and let $(x^+, s^+) = (x, s) + \alpha(\Delta x, \Delta s)$. Then,*

$$(x^+)^T s^+ \quad = \quad \left(1 - \frac{\alpha}{\Delta}(\rho\beta_1 - \beta_2)\right) x^T s. \qquad (33)$$

□

Proposition 2.3 (Prop. 4.1, [12]) *The search direction given by (29) and (30) is primal-dual symmetric and scale invariant.* □

Proposition 2.4 (Prop. 4.2, [12]) *Let (x, s) be a point on the central path $\mathcal{C}$. Then, the search direction given by (29) and (30) is a scalar multiple of the primal-dual affine scaling direction.* □

We will end this section by describing our algorithm formally. The reasons for stepsize selection rules below will be apparent after the analysis of the next three sections.

Algorithm 1 *Let ρ be an $\mathcal{O}(n)$ parameter greater than $2n+1$ and let $\varepsilon > 0$. Finally, let (x^0, s^0) be a strictly feasible solution for this LP such that $\Phi_\rho\left((x^0)^T s^0\right) = \mathcal{O}(n \ln \frac{1}{\varepsilon})$. Let k=0.*

1. *If $(x^k)^T s^k < \varepsilon$, stop.*
2. *Compute the search direction $(\Delta x, \Delta s)$ from (29) and (30). Choose a step size $\frac{1}{5} \leq \alpha^k = \tau^k \alpha^k_{\max}$, with*

$$\alpha^k_{\max} = \max\{\alpha : x^k + \alpha\Delta x^k \geq 0, s^k + \alpha\Delta s^k \geq 0\},$$
$$\tau_k = 1 - \Theta\left((x^k)^T s^k\right),$$

such that

$$\Phi_\rho(x^+, s^+) - \Phi_\rho(x, s) \leq -0.04.$$

3. let $(x^{k+1}, s^{k+1}) := (x^k, s^k) + \alpha^k(\Delta x, \Delta s)$, $k = k + 1$. *Go to step 1.*

3 Polynomiality of the Algorithm

In this section, we will demonstrate that it is always possible to reduce the potential function Φ_ρ by at least a constant amount as prescribed in the algorithm of the previous section. This will lead to the conclusion that the algorithm finds an ε-complementary solution in polynomial time. We achieve this result by first deriving a quadratic underestimate of the reduction in the potential function and then showing that the linear part of this estimate is sufficiently negative.

Let $(\Delta x, \Delta s)$ be a feasible direction for the primal-dual pair of problems and let (x, s) be the current iterate. For the feasibility of the next iterate $(x + \alpha\Delta x, s + \alpha\Delta s)$, a sufficient condition on the stepsize α is:

$$\alpha \max\left(\|X^{-1}\Delta x\|_\infty, \|S^{-1}\Delta s\|_\infty\right) \leq 1.$$

By requiring the stepsize to satisfy a more stringent condition we can estimate the change in the value of the potential function as a quadratic function of the stepsize:

Lemma 3.1 *Let* $(\Delta x, \Delta s)$ *be a feasible direction for the primal-dual pair of problems and let* (x, s) *be the current feasible iterate. If the stepsize* α *satisfies*

$$\alpha \max\left(\|X^{-1}\Delta x\|_\infty, \|S^{-1}\Delta s\|_\infty\right) \leq 1/2, \tag{34}$$

then we have

$$\Phi_\rho(x + \alpha\Delta x, s + \alpha\Delta s) - \Phi_\rho \leq \alpha u_1 + \alpha^2 u_2, \tag{35}$$

where

$$u_1 = \rho\frac{x^T\Delta s + s^T\Delta x}{x^T s} - e^T\left(X^{-1}\Delta x + S^{-1}\Delta s\right) = [\nabla\Phi_\rho]^T \begin{bmatrix} \Delta x \\ \Delta s \end{bmatrix}, \tag{36}$$

$$u_2 = \|X^{-1}\Delta x\|_2^2 + \|S^{-1}\Delta s\|_2^2. \tag{37}$$

Proof:
The inequality (35) is well known, see, e.g., [13, page 71] and substitute $\tau = 0.5$. $\square$

To show that a reduction is always possible, we will demonstrate that u_1 is sufficiently negative using techniques similar to those of [4]:

Lemma 3.2 *If $(\Delta x, \Delta s)$ is given by (29) and (30) then, the scalar u_1 defined in (36) is less than $-\frac{1}{2}$.*

Proof:
Let us denote the reduced gradient vector $Z^T \nabla \Phi_\rho$ and the reduced Hessian matrix $Z^T(\nabla^2 \Phi_\rho + \nabla \Phi_\rho \nabla \Phi_\rho^T)Z$ with g and H to simplify the notation. By definition, $Hd_{RN} = -g$. We start by noting that the reduced Newton direction d_{RN} given by (16) maximizes the following function:

$$h(d) \quad := \quad -\frac{g^T d}{\sqrt{d^T H d}}. \tag{38}$$

Hence,

$$\left(\max_d h(d)\right)^2 = \frac{(g^T d_{RN})^2}{d_{RN}^T H d_{RN}} = -g^T d_{RN}. \tag{39}$$

Next, we will show that there is a vector $\hat{d}$ satisfying $h(\hat{d}) > \frac{1}{\sqrt{2}}$ which will lead to the conclusion that $u_1 = g^T d_{RN} < -\frac{1}{2}$ using (39).

Let (x^*, s^*) be an optimal pair of primal-dual solutions to problems (1) and (4), and let (x, s) be the current iterate. Then, the vector $\eta := \begin{bmatrix} x^* - x \\ s^* - s \end{bmatrix}$ is in the range space of the matrix Z. Let $\hat{d}$ be a vector such that $Z\hat{d} = \eta$. Note that $\begin{bmatrix} s \\ x \end{bmatrix}^T \eta = x^T s^* + s^T x^* - 2x^T s = -x^T s$, since $(x^* - x)^T (s^* - s) = 0$.

Using (8) we observe that

$$\begin{aligned} g^T \hat{d} &= \nabla \Phi_\rho^T Z \hat{d} = \nabla \Phi_\rho^T \eta \\ &= \frac{\rho}{x^T s} \left(\begin{bmatrix} s \\ x \end{bmatrix}^T \eta \right) - \left(e^T X^{-1}(x^* - x) + e^T S^{-1}(s^* - s) \right) \\ &= -(\rho - 2n) - \left(e^T X^{-1} x^* + e^T S^{-1} s^* \right). \end{aligned} \tag{40}$$

Next, using (8) and (9) we obtain

$$\begin{aligned}
\hat{d}^T H \hat{d} &= \eta^T(\nabla^2\Phi_\rho + \nabla\Phi_\rho\nabla\Phi_\rho^T)\eta = \eta^T\nabla^2\Phi_\rho\eta + (g^T\hat{d})^2 \\
&= (x^* - x)^T X^{-2}(x^* - x) + (s^* - s)^T S^{-2}(s^* - s) + 2(x^* - x)^T(s^* - s) \\
&\quad -\frac{\rho}{(x^T s)^2}\left((x^* - x)^T s + (s^* - s)^T x\right)^2 + (g^T\hat{d})^2 \\
&= \left((x^*)^T X^{-2} x^* + (s^*)^T S^{-2} s^*\right) - 2\left(e^T X^{-1} x^* + e^T S^{-1} s^*\right) + 2n \\
&\quad -\rho + (g^T\hat{d})^2 \\
&< 2(g^T\hat{d})^2. \qquad (41)
\end{aligned}$$

Above, we used the identity $\begin{bmatrix} s \\ x \end{bmatrix}^T \eta = -x^T s$. The final inequality can be verified by squaring both sides of (40) and using the inequality $\sum w_i^2 \le \left(\sum w_i\right)^2$ for nonnegative w_i.

Since (41) implies that $h(\hat{d}) > \frac{1}{\sqrt{2}}$, using (39) we conclude

$$u_1 = [\nabla\Phi_\rho]^T \begin{bmatrix} \Delta x \\ \Delta s \end{bmatrix} = g^T d_{RN} < -\frac{1}{2}. \qquad (42)$$

□

Next, we will demonstrate that the term u_2 is bounded above. This analysis is also similar to that of [4]. For this proof we will need a larger value for the potential function parameter ρ. Fortunately, this requirement does not affect the overall complexity.

Lemma 3.3 *Let $(\Delta x, \Delta s)$ be given by (29) and (30). Further, assume that $\rho \ge 4n + 2$. Then, the scalar u_2 defined in (37) is at most $\frac{25}{18}$.*

Proof:
Let $\hat{H} = d_{RN}^T H d_{RN}$. Since $H d_{RN} = -g$, we have that $\hat{H} = d_{RN}^T H d_{RN} = -g^T d_{RN} > \frac{1}{2}$ and that

$$\hat{H} = d_{RN}^T H d_{RN} = -g^T d_{RN} = -\rho\gamma_1 + \gamma_2 \qquad (43)$$

where

$$\gamma_1 = \frac{s^T\Delta x + x^T\Delta s}{x^T s} \quad \text{and} \quad \gamma_2 = e^T(X^{-1}\Delta x + S^{-1}\Delta s).$$

On the other hand, using the definition of H, (9), and (12) we obtain

$$\hat{H} = u_2 - \rho\gamma_1^2 + \hat{H}^2. \tag{44}$$

Since $u_2 = (\|X^{-1}\Delta x\|_2^2 + \|S^{-1}\Delta s\|_2^2) = \sigma^2 + \frac{\gamma_2^2}{2n}$, where

$$\sigma^2 = \left\| \begin{array}{c} X^{-1}\Delta x - \frac{e^T(X^{-1}\Delta x + S^{-1}\Delta s)}{2n} e \\ S^{-1}\Delta s - \frac{e^T(X^{-1}\Delta x + S^{-1}\Delta s)}{2n} e \end{array} \right\|_2^2$$

we can rewrite (44) as

$$\rho\gamma_1^2 - \frac{\gamma_2^2}{2n} = \hat{H}^2 - \hat{H} + \sigma^2. \tag{45}$$

Now we can solve for γ_1 and γ_2 in terms of $\hat{H}$ and σ using (43) and (45). Since (45) is quadratic we get two roots for each γ_i:

$$\gamma_1 = -\frac{1}{\rho - 2n}\hat{H} \pm \sqrt{\frac{2n}{\rho(\rho - 2n)}\left(\hat{H} - \frac{\rho - 2n - 1}{\rho - 2n}\hat{H}^2 - \sigma^2\right)} \tag{46}$$

$$\gamma_2 = -\frac{2n}{\rho - 2n}\hat{H} \pm \sqrt{\frac{2n\rho}{\rho - 2n}\left(\hat{H} - \frac{\rho - 2n - 1}{\rho - 2n}\hat{H}^2 - \sigma^2\right)}. \tag{47}$$

Since γ_i's are real numbers, the term inside the square-root must be nonnegative. Since $\sigma^2 \geq 0$ and $\rho > 2n + 1$, we conclude that

$$\hat{H} \leq \frac{\rho - 2n}{\rho - 2n - 1}. \tag{48}$$

Also note that

$$\hat{H} - \frac{\rho - 2n - 1}{\rho - 2n}\hat{H}^2 = \hat{H}(1 - \frac{\rho - 2n - 1}{\rho - 2n}\hat{H}) \leq \frac{\rho - 2n}{4(\rho - 2n - 1)}. \tag{49}$$

Now,

$$\begin{aligned} u_2 &= \sigma^2 + \frac{\gamma_2^2}{2n} \\ &\leq \sigma^2 + 2n\left(\frac{1}{\rho - 2n}\hat{H} + \sqrt{\frac{\rho}{2n(\rho - 2n)}\left(\hat{H} - \frac{\rho - 2n - 1}{\rho - 2n}\hat{H}^2 - \sigma^2\right)}\right)^2 && (50) \\ &\leq 2n\left(\frac{1}{\rho - 2n}\hat{H} + \sqrt{\frac{\rho}{2n(\rho - 2n)}\left(\hat{H} - \frac{\rho - 2n - 1}{\rho - 2n}\hat{H}^2\right)}\right)^2 && (51) \\ &\leq \left(\frac{\sqrt{2n}}{\rho - 2n - 1} + \frac{1}{2}\sqrt{\frac{\rho}{\rho - 2n - 1}}\right)^2. && (52) \end{aligned}$$

Above, (50) uses (47) and the fact that $\hat{H}$ is positive. (51) holds since the right-hand-side of (50) is a decreasing function of σ^2. Finally, (52) follows from (48) and (49).

Now, using the assumption $\rho \geq 4n+2$ we get

$$\begin{aligned} u_2 &\leq \left(\frac{\sqrt{2n}}{\rho - 2n - 1} + \frac{1}{2}\sqrt{\frac{\rho}{\rho - 2n - 1}} \right)^2 \\ &\leq \left(\frac{\sqrt{2n}}{2n+1} + \frac{1}{2}\sqrt{\frac{4n+2}{2n+1}} \right)^2 \leq (\frac{\sqrt{2}}{3} + \frac{1}{2}\sqrt{2})^2 = \frac{25}{18}. \end{aligned}$$

□

Now we are ready to demonstrate that Φ_ρ can be reduced by at least a constant using a step in the direction given by (29) and (30):

Lemma 3.4 *Let $(\Delta x, \Delta s)$ be given by (29) and (30). Further, assume that $\rho \geq 4n+2$. Then there exists a stepsize α no less than $\frac{1}{5}$ that yields at least a reduction of 0.04 in Φ_ρ.*

Proof:
We will show that $\alpha = \frac{1}{5}$ satisfies the statement of the lemma. First, note that $\alpha = \frac{1}{5}$ satisfies (34). Indeed,

$$\max\left(\|X^{-1}\Delta x\|_\infty, \|S^{-1}\Delta s\|_\infty\right) = \left\| \begin{array}{c} X^{-1}\Delta x \\ S^{-1}\Delta s \end{array} \right\|_\infty \leq \left\| \begin{array}{c} X^{-1}\Delta x \\ S^{-1}\Delta s \end{array} \right\|_2 = \sqrt{u_2} \leq \sqrt{\frac{25}{18}}. \quad (53)$$

Accordingly, it follows from Lemma 3.1 that

$$\Phi_\rho(x + \frac{1}{5}\Delta x, s + \frac{1}{5}\Delta s) - \Phi_\rho \leq \frac{1}{5}u_1 + \frac{1}{5^2}u_2 < -\frac{1}{10} + \frac{1}{18} < -0.04. \quad (54)$$

□

The lemma above leads to the following polynomial complexity result:

Theorem 3.1 *Given $(x^0, s^0) \in \mathcal{F}^0$ and $\rho \geq 4n+2$, Algorithm 1 given in Section 3 generates iterates satisfying $\Phi_\rho(x^+, s^+) - \Phi_\rho(x, s) \leq -0.04$ using steps of size at least $\frac{1}{5}$ and there is an index K given by*

$$K = \frac{1}{0.04}(\Phi_\rho(x^0, s^0) + (\rho - n)\log(1/\varepsilon))$$

such that $\frac{(x^k)^T s^k}{n} \leq \varepsilon$ for $k \geq K$. □

We omit the simple proof of the theorem above, see, e.g., [13] for a similar result.

4 Asymptotic Analysis of the Search Directions

4.1 Basics and Notation

The local convergence analysis of the interior-point algorithm described in Section 2 requires an accurate estimation of the different vectors and scalars appearing in equations (29) and (30). This, in turn, requires a careful analysis of the the projection matrices $\Xi(x^k)$ and $\Sigma(s^k)$ in the neighborhood of the optimal solution set. In [12], such an analysis was performed with the assumption that the LP and its dual have nondegenerate solutions.

In the nondegenerate case, the optimal solution set Ω is a single vertex and therefore, convergence to Ω implies the convergence of the iterates to this unique point. In the degenerate case, however, even when Ω is bounded, a proof of the convergence of the iterates is more involved since there may be several accumulation points in Ω. Our strategy to prove convergence is as follows: First, through a careful asymptotic analysis of the Newton search directions we will demonstrate that all accumulation points lie in the relative interior of a particular face $\mathcal{F}$ of Ω (Theorem 4.2)[1]. Then, we will show that either $\mathcal{F}$ is a vertex, or there is a unique accumulation point which is the relative analytic center of $\mathcal{F}$ (Theorem 4.3).

We start by identifying the face $\mathcal{F}$ of Ω mentioned in the previous paragraph. To prove convergence of the iterates of a potential-reduction algorithm, one normally has to assume that the optimal solution set is bounded. Otherwise, one may be able to keep reducing the potential function and keep approaching the optimal face with divergent iterates. With our assumption of the existence of a strictly feasible primal-dual pair, we have that Ω is bounded. Since the sequence of iterates generated by our algorithm will converge to the set Ω, as illustrated in Theorem 3.1, this sequence will have all its accumulation points on Ω.

Let $(x^*, s^*) \in \Omega$ be an accumulation point of the sequence of iterates generated by Algorithm 1 that has the maximum number of zeros, in x and s variables combined. We define $\mathcal{F}$ to be the face of Ω that contains (x^*, s^*) in its relative interior. If (x^*, s^*) is a vertex, then $\mathcal{F}$ is chosen to be this vertex. This definition of $\mathcal{F}$ is analogous to Tsuchiya's choice in the proof of Theorem 4.2 in [11]. Note that, given $(x^*, s^*) \in \Omega$, the face $\mathcal{F}$ is uniquely determined. Potentially, there may be several accumulation points with the same maximum number of zeros. We will show in Theorem 4.2 that they must all lie in the relative interior of $\mathcal{F}$. We also

[1] The relative interior of a vertex is taken to be the vertex itself.

note for future reference that no proper face of $\mathcal{F}$ can contain an accumulation point since such points would have more zeros than the points in the relative interior of $\mathcal{F}$, contradicting our choice of $\mathcal{F}$.

Next, we introduce the notation we will employ in the remainder of our analysis. Let $\mathcal{F}$ be the face of the primal-dual optimal solution set Ω defined in the previous paragraph. Let $\mathcal{F}_x$ and $\mathcal{F}_s$ denote the restrictions of the set $\mathcal{F}$ to the x and s components, respectively. In other words,

$$\begin{aligned}\mathcal{F}_x &:= \{x : \exists s \in \Re^n \text{ s.t. } (x, s) \in \mathcal{F}\},\\ \mathcal{F}_s &:= \{s : \exists x \in \Re^n \text{ s.t. } (x, s) \in \mathcal{F}\}.\end{aligned}$$

Define F_x^0 as the set of indices i for which $x_i = 0$ whenever $x \in \mathcal{F}_x$, and let $F_x^+ = \{1, \cdots, n\} \setminus F_x^0$. Therefore,

$$F_x^+ = \{i : \exists x \in \mathcal{F}_x \text{ s.t. } x_i > 0\}.$$

Likewise, define F_s^0 as the set of indices i for which $s_i = 0$ whenever $s \in \mathcal{F}_s$ and let $F_s^+ = \{1, \cdots, n\} \setminus F_s^+$. From complementary slackness conditions it follows that $F_x^0 \cup F_s^0 = \{1, \cdots, n\}$, and that $F_s^+ \subset F_x^0$ as well as $F_x^+ \subset F_s^0$. The inclusions may be strict if the LP is degenerate.

In what follows, given a matrix H and an index set I, let H_I denote the submatrix of H consisting of columns indexed by I, unless specified otherwise. Choose a subset B_x of F_x^+ such that columns of A_{B_x} form a basis for the range space of $A_{F_x^+}$. Let $N_x = F_x^+ \setminus B_x$. Note that there exists a matrix $\Gamma_{N_x B_x}$ that satisfies

$$A_{N_x} = A_{B_x} \Gamma_{N_x B_x}. \tag{55}$$

Since the matrix A is assumed to have full row rank, we can find an index set $\overline{B}_x \subset F_x^0$ such that the matrix $A_{B_x \cup \overline{B}_x}$ is square and nonsingular. Define $\overline{N}_x := F_x^0 \setminus \overline{B}_x$. Therefore, the index sets B_x, N_x, $\overline{B}_x$, and $\overline{N}_x$ form a partition of $\{1, \cdots, n\}$. We will use the following notation for the inverse of the matrix $A_{B_x \cup \overline{B}_x}$:

$$A_{B_x \cup \overline{B}_x}^{-1} = \begin{bmatrix} A_{B_x} & A_{\overline{B}_x} \end{bmatrix}^{-1} = \begin{bmatrix} \overline{A}_{B_x} \\ \overline{A}_{\overline{B}_x} \end{bmatrix}. \tag{56}$$

Above, $\overline{A}_{B_x}$ and $\overline{A}_{\overline{B}_x}$ denote the submatrices of $A_{B_x \cup \overline{B}_x}^{-1}$ consisting of *rows* indexed by B_x and $\overline{B}_x$ respectively. From (56) it follows that

$$A_{B_x} \overline{A}_{B_x} + A_{\overline{B}_x} \overline{A}_{\overline{B}_x} = I_m, \tag{57}$$

$$\overline{A}_{B_x} A_{B_x} = I_{|B_x|}, \text{ and } \overline{A}_{\overline{B}_x} A_{\overline{B}_x} = I_{|\overline{B}_x|}, \tag{58}$$

$$\overline{A}_{B_x} A_{\overline{B}_x} = 0, \text{ and } \overline{A}_{\overline{B}_x} A_{B_x} = 0. \tag{59}$$

Using the partition described above, we can write the matrix AX as follows:

$$\begin{aligned} AX &= \left[\begin{array}{cc} A_{F_x^+} X_{F_x^+} & A_{F_x^0} X_{F_x^0} \end{array} \right] = \left[\begin{array}{cccc} A_{B_x} X_{B_x} & A_{N_x} X_{N_x} & A_{\overline{B}_x} X_{\overline{B}_x} & A_{\overline{N}_x} X_{\overline{N}_x} \end{array} \right] \\ &= \left[\begin{array}{cc} A_{B_x} X_{B_x} & A_{\overline{B}_x} X_{\overline{B}_x} \end{array} \right] \left[\begin{array}{cccc} I & X_{B_x}^{-1} \overline{A}_{B_x} A_{N_x} X_{N_x} & 0 & X_{B_x}^{-1} \overline{A}_{B_x} A_{\overline{N}_x} X_{\overline{N}_x} \\ 0 & X_{\overline{B}_x}^{-1} \overline{A}_{\overline{B}_x} A_{N_x} X_{N_x} & I & X_{\overline{B}_x}^{-1} \overline{A}_{\overline{B}_x} A_{\overline{N}_x} X_{\overline{N}_x} \end{array} \right]. \end{aligned}$$

Let $R_{B_x N_x} = X_{B_x}^{-1} \overline{A}_{B_x} A_{N_x} X_{N_x}$, and define $R_{B_x \overline{N}_x}$, etc., similarly. From (55) and (59), we have that $R_{\overline{B}_x N_x} = 0$. Next, we define

$$R_{B_x F_x^+} = \left[\begin{array}{cc} R_{B_x B_x} & R_{B_x N_x} \end{array} \right] = \left[\begin{array}{cc} I_{|B_x|} & R_{B_x N_x} \end{array} \right], \tag{60}$$

$$R_{B_x F_x^0} = \left[\begin{array}{cc} R_{B_x \overline{B}_x} & R_{B_x \overline{N}_x} \end{array} \right] = \left[\begin{array}{cc} 0 & R_{B_x \overline{N}_x} \end{array} \right], \tag{61}$$

$$R_{\overline{B}_x F_x^0} = \left[\begin{array}{cc} R_{\overline{B}_x \overline{B}_x} & R_{\overline{B}_x \overline{N}_x} \end{array} \right] = \left[\begin{array}{cc} I_{|\overline{B}_x|} & R_{\overline{B}_x \overline{N}_x} \end{array} \right], \tag{62}$$

where the zero matrix in (61) is $|B_x| \times |\overline{B}_x|$. Above, $R_{B_x F_x^+} \in I\!R^{|B_x| \times |F_x^+|}$, $R_{B_x F_x^0} \in I\!R^{|B_x| \times |F_x^0|}$, and $R_{\overline{B}_x F_x^0} \in I\!R^{|\overline{B}_x| \times |F_x^0|}$. Using this notation, we can simplify the expression for AX:

$$\begin{aligned} AX &= \left[\begin{array}{cc} A_{B_x} X_{B_x} & A_{\overline{B}_x} X_{\overline{B}_x} \end{array} \right] \left[\begin{array}{cccc} I & R_{B_x N_x} & 0 & R_{B_x \overline{N}_x} \\ 0 & 0 & I & R_{\overline{B}_x \overline{N}_x} \end{array} \right] \\ &= A_{B_x \cup \overline{B}_x} X_{B_x \cup \overline{B}_x} \left[\begin{array}{cc} R_{B_x F_x^+} & R_{B_x F_x^0} \\ 0 & R_{\overline{B}_x F_x^0} \end{array} \right]. \end{aligned} \tag{63}$$

Using the definitions above we introduce two orthogonal projection matrices that will be useful in the decomposition of the matrix Ξ:

$$\overline{\Xi}_+ := R_{B_x F_x^+}^T \left(R_{B_x F_x^+} R_{B_x F_x^+}^T \right)^{-1} R_{B_x F_x^+}, \tag{64}$$

$$\overline{\Xi}_0 := R_{\overline{B}_x F_x^0}^T \left(R_{\overline{B}_x F_x^0} R_{\overline{B}_x F_x^0}^T \right)^{-1} R_{\overline{B}_x F_x^0}. \tag{65}$$

An analogous development with the dual variables leads to the two orthogonal projection matrices given below:

$$\overline{\Sigma}_+ := R_{B_s F_s^+}^T \left(R_{B_s F_s^+} R_{B_s F_s^+}^T \right)^{-1} R_{B_s F_s^+}, \tag{66}$$

$$\overline{\Sigma}_0 := R_{\overline{B}_s F_s^0}^T \left(R_{\overline{B}_s F_s^0} R_{\overline{B}_s F_s^0}^T \right)^{-1} R_{\overline{B}_s F_s^0}, \tag{67}$$

where B_s, $\overline{B}_s$, $R_{B_s F_s^+}$, etc., are defined analogously to the corresponding definitions involving x variables. Finally, let $\Xi_+ = I - \overline{\Xi}_+$, $\Sigma_0 = I - \overline{\Sigma}_0$, etc.

4.2 Estimation of the Projection Matrices

In this section we will provide asymptotic estimates of the projection matrices Ξ and Σ defined in (19)-(20) using the partitions described in the previous subsection. These estimates will involve the matrices Ξ_+, Ξ_0, Σ_+, and Σ_0 and an error term.

The accuracy of our estimates will depend on the magnitude of the error term, which is bounded by a constant multiple of $\|X_{F_x^+}^{-1}\| \cdot \|X_{F_x^0}\| + \|S_{F_s^+}^{-1}\| \cdot \|S_{F_s^0}\|$. We will later show that all accumulation points of our algorithm are in the relative interior of the face $\mathcal{F}$. Therefore, the sequence $\{(x_{F_x^+}, s_{F_s^+})\}$ remains bounded away from zero and $\{(x_{F_x^0}, s_{F_s^0})\}$ approaches zero, which indicates that the error term approaches zero in magnitude.

Lemma 4.1 *Let $\mathcal{F}$, Ξ_+, Ξ_0, Σ_+, Σ_0, and B_x, B_s, etc., be as defined above. Let $\overline{\Xi}_+ = I - \Xi_+$, etc. Further assume that $\|X_{F_x^+}^{-1}\| \cdot \|X_{F_x^0}\| \to 0$ and $\|S_{F_s^+}^{-1}\| \cdot \|S_{F_s^0}\| \to 0$. Then, the matrices Ξ and Σ defined in (19) and (20) can be partitioned as follows (after possible row/column permutations):*

$$\Xi = \begin{bmatrix} \Xi_+ & \\ & \Xi_0 \end{bmatrix} + \mathcal{O}\Big(\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\|\Big), \tag{68}$$

$$\Sigma = \begin{bmatrix} \Sigma_0 & \\ & \Sigma_+ \end{bmatrix} + \mathcal{O}\Big(\|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|\Big). \tag{69}$$

Proof:
We give a proof of this lemma in Section 7. □

4.3 Estimation of the Projected Vectors

In order to estimate the asymptotic values of the search directions Δx and Δs as iterates approach the optimal face, we need to evaluate the asymptotic values of the matrix-vector products $(\Xi + \Sigma)\nu$ and $(\Xi + \Sigma)e$ and the parameters β_i, for $i = 1, 2, 3$, and Δ. We will rely on Lemma 4.1 and the following two lemmas:

Lemma 4.2 *Let e be a column vector of ones of appropriate dimension in the following statements. Then, $e \in \mathcal{N}(R_{\overline{B}_x F_x^0})$ and $e \in \mathcal{N}(R_{\overline{B}_s F_s^0})$. Therefore,*

$$\Xi_0 e = e, \tag{70}$$

$$\Sigma_0 e = e. \tag{71}$$

Proof:
Since x is a feasible vector for (1), $Ax = AXe = b$. Therefore, from (63) we have

$$A_{B_x \cup \overline{B}_x} X_{B_x \cup \overline{B}_x} \begin{bmatrix} R_{B_x F_x^+} e + R_{B_x F_x^0} e \\ R_{\overline{B}_x F_x^0} e \end{bmatrix} = b. \tag{72}$$

Since $\hat{x}_i = 0$ for $i \in F_x^0$ whenever $\hat{x} \in \mathcal{F}_x$, and since $\mathcal{F}$ is assumed to be nonempty, the right-hand-side vector b is in the range space of $A_{F_x^+}$, which is the same as the range space of A_{B_x}. Therefore, there exists a vector $z \in \Re^{|B_x|}$ such that $A_{B_x} z = b$. So, $\begin{bmatrix} A_{B_x} & A_{\overline{B}_x} \end{bmatrix} \begin{bmatrix} z \\ 0 \end{bmatrix} = b$, or $\begin{bmatrix} \overline{A}_{B_x} b \\ \overline{A}_{\overline{B}_x} b \end{bmatrix} = \begin{bmatrix} z \\ 0 \end{bmatrix}$. Now, premultiplying both sides of (72) by $\overline{A}_{\overline{B}_x}$, using (58), (59), and the identity $\overline{A}_{\overline{B}_x} b = 0$, we obtain

$$X_{\overline{B}_x} R_{\overline{B}_x F_x^0} e = 0.$$

Since $X_{\overline{B}_x}$ is a positive diagonal matrix, this last result indicates that $R_{\overline{B}_x F_x^0} e = 0$, i.e., $e \in \mathcal{N}(R_{\overline{B}_x F_x^0})$. Furthermore, since Ξ_0 is the orthogonal projection matrix onto the null-space of $R_{\overline{B}_x F_x^0}$, it maps e onto itself. The corresponding result for the dual variables is proved identically. □

We have a similar result involving the matrices $R_{B_x F_x^+}$, $R_{B_s F_s^+}$, and the vector $\nu := \frac{XSe}{x^T s}$:

Lemma 4.3 *Let ν be as defined in (23). Then, $\nu_{F_x^+} \in \mathcal{R}(R^T_{B_x F_x^+})$ and $\nu_{F_s^+} \in \mathcal{R}(R_{B_s F_s^+})$. Therefore,*

$$\Xi_+ \nu_{F_x^+} = 0, \tag{73}$$
$$\Sigma_+ \nu_{F_s^+} = 0. \tag{74}$$

Proof:
We will prove the statement of the lemma only for the x variables; the analogous result for the dual variables can be proved identically. First, recall from (63) that

$$XA^T = \begin{bmatrix} R^T_{B_x F_x^+} & 0 \\ R^T_{B_x F_x^0} & R^T_{\overline{B}_x F_x^0} \end{bmatrix} X_{B_x \cup \overline{B}_x} A^T_{B_x \cup \overline{B}_x}.$$

Since $X_{B_x \cup \overline{B}_x}$ and $A^T_{B_x \cup \overline{B}_x}$ are both nonsingular square matrices, the range space of $(XA^T)_{F_x^+} = X_{F_x^+} A^T_{F_x^+} = \begin{bmatrix} X_{B_x} A^T_{B_x} \\ X_{N_x} A^T_{N_x} \end{bmatrix}$ coincides with the range space of $R^T_{B_x F_x^+}$.

From complementary slackness conditions, $\hat{s}_{F_x^+} = 0$ whenever $\hat{s} \in \mathcal{F}_s$. Since $\mathcal{F}_s$ is nonempty and since $(c - \hat{s}) \in \mathcal{R}(A^T)$ (recall that $G\hat{s} = Gc$ and $\mathcal{R}(A^T) = \mathcal{N}(G)$) we must have $c_{F_x^+} \in \mathcal{R}(A^T_{F_x^+})$. Further, since $(c - s)_{F_x^+} \in \mathcal{R}(A^T_{F_x^+})$ whenever $Gs = Gc$, we conclude that $s_{F_x^+} \in \mathcal{R}(A^T_{F_x^+})$ whenever s is feasible for the dual problem.

Now combining the fact that $\mathcal{R}(R^T_{B_x F_x^+}) = \mathcal{R}(X_{F_x^+} A^T_{F_x^+})$, and that $s_{F_x^+} \in \mathcal{R}(A^T_{F_x^+})$, we conclude that $\nu_{F_x^+} = \frac{X_{F_x^+} s_{F_x^+}}{x^T s} \in \mathcal{R}(R^T_{B_x F_x^+})$. This also indicates that $\nu_{F_x^+}$ is perpendicular to the null space of $R_{B_x F_x^+}$. Therefore, $\Xi_+ \nu_{F_x^+}$, the projection of $\nu_{F_x^+}$ into this null space, is zero. $\square$

Using Lemmas 4.2 and 4.3 we can now evaluate the asymptotic values of the scalars β_i and Δ:

Lemma 4.4 *Let $\mathcal{F}$, F_x^+, F_x^0, F_s^+, and F_s^0 be as defined in Section 4.1. If $\|X^{-1}_{F_x^+}\|\|X_{F_x^0}\| \to 0$ and $\|S^{-1}_{F_s^+}\|\|S_{F_s^0}\| \to 0$, the scalars β_i, $i = 1, 2, 3$, and Δ defined in (24)-(27) satisfy the following relations:*

$$\frac{1}{2n} + \mathcal{O}\Big(\|X^{-1}_{F_x^+}\|\|X_{F_x^0}\| + \|S^{-1}_{F_s^+}\|\|S_{F_s^0}\|\Big) \le \beta_1 \le 2, \tag{75}$$

$$\beta_2 = 1 + \mathcal{O}\Big(\|X^{-1}_{F_x^+}\|\|X_{F_x^0}\| + \|S^{-1}_{F_s^+}\|\|S_{F_s^0}\|\Big), \tag{76}$$

$$0 \le \beta_3 \le 2n, \tag{77}$$

$$\Delta = \Theta(1). \tag{78}$$

Proof:
We start by noting that the union of the index sets F_x^0 and F_s^0 is $\{1, \ldots, n\}$ from complementary slackness conditions, and that these sets may have a non-empty intersection if $\mathcal{F}$ does not contain a strictly complementary solution.

Since $\nu = \frac{XSe}{x^T s} = \Theta(1)$, using Lemmas 4.1 and 4.3 we have:

$$\Xi\nu = \begin{bmatrix} \Xi_+ \nu_{F_x^+} \\ \Xi_0 \nu_{F_x^0} \end{bmatrix} + \mathcal{O}\Big(\|X^{-1}_{F_x^+}\|\|X_{F_x^0}\|\Big) = \begin{bmatrix} 0 \\ \Xi_0 \nu_{F_x^0} \end{bmatrix} + \mathcal{O}\Big(\|X^{-1}_{F_x^+}\|\|X_{F_x^0}\|\Big), \tag{79}$$

$$\Sigma\nu = \begin{bmatrix} \Sigma_+ \nu_{F_s^+} \\ \Sigma_0 \nu_{F_s^0} \end{bmatrix} + \mathcal{O}\Big(\|S^{-1}_{F_s^+}\|\|S_{F_s^0}\|\Big) = \begin{bmatrix} 0 \\ \Sigma_0 \nu_{F_s^0} \end{bmatrix} + \mathcal{O}\Big(\|S^{-1}_{F_s^+}\|\|S_{F_s^0}\|\Big). \tag{80}$$

Therefore, using Lemma 4.2, $\beta_2 = e^T(\Xi + \Sigma)\nu$ can be written as

$$\beta_2 = e^T \Xi_0 \nu_{F_x^0} + e^T \Sigma_0 \nu_{F_s^0} + \mathcal{O}\Big(\|X^{-1}_{F_x^+}\|\|X_{F_x^0}\| + \|S^{-1}_{F_s^+}\|\|S_{F_s^0}\|\Big)$$

$$
\begin{aligned}
&= e^T \nu_{F_x^0} + e^T \nu_{F_s^0} + \mathcal{O}\Big(\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\| + \|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|\Big) \\
&= 1 + \mathcal{O}\Big(\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\| + \|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|\Big),
\end{aligned}
$$

which proves (76). The last equality above follows from the fact that

$$
e^T \nu_{(F_x^0 \cap F_s^0)} = \mathcal{O}\Big(\|X_{F_x^0}\|\|S_{F_s^0}\|\Big) = \mathcal{O}\Big(\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\| + \|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|\Big),
$$

since both $x_{F_x^+}$ and $s_{F_s^+}$ are bounded.

Next, note that $\beta_1 = \nu^T \Xi \nu + \nu^T \Sigma \nu = \nu^T \Xi^2 \nu + \nu^T \Sigma^2 \nu$, since Ξ and Σ are orthogonal projection matrices. Therefore, $\beta_1 = \|\Xi\nu\|_2^2 + \|\Sigma\nu\|_2^2 \le 2\|\nu\|_2^2 \le 2\|\nu\|_1^2 = 2$, establishing the second inequality in (75). A similar argument shows that $\beta_3 = e^T(\Xi + \Sigma)e \le 2\|e\|_2^2 = 2n$. Since Ξ and Σ are also positive semidefinite, $\beta_3 \ge 0$. Therefore, (77) holds.

From (79) and (80) it also follows that $\beta_1 = \nu_{F_x^0}^T \Xi_0 \nu_{F_x^0} + \nu_{F_s^0}^T \Sigma_0 \nu_{F_s^0} + \mathcal{O}\Big(\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\| + \|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|\Big)$. On the other hand, from the Cauchy-Schwartz inequality we have that

$$
\begin{aligned}
(e^T \Xi_0 \nu_{F_x^0})^2 = (e^T \nu_{F_x^0})^2 &\le |F_x^0| (\nu_{F_x^0}^T \Xi_0 \nu_{F_x^0}) \le n(\nu_{F_x^0}^T \Xi_0 \nu_{F_x^0}) \\
(e^T \Sigma_0 \nu_{F_s^0})^2 = (e^T \nu_{F_s^0})^2 &\le |F_s^0| (\nu_{F_s^0}^T \Sigma_0 \nu_{F_s^0}) \le n(\nu_{F_s^0}^T \Xi_0 \nu_{F_s^0})
\end{aligned}
$$

Since $e^T \nu_{F_x^0} + e^T \nu_{F_s^0} \ge 1$, the sum of the left hand side terms of the two inequalities above is at least $\frac{1}{2}$. Therefore,

$$
\beta_1 \;\ge\; \frac{1}{2n} + \mathcal{O}\Big(\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\| + \|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|\Big).
$$

Finally, recall that

$$
\Delta \;=\; (\rho\beta_1 - 1)(\rho - \beta_3 - 1) + \rho(1 - \beta_2)^2.
$$

Since $\rho > 2n + 1$, (77) indicates that $(\rho - \beta_3 - 1)$ is always positive and (75) indicates that $\rho\beta_1 - 1$ is positive when $\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\|$ and $\|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|$ are sufficiently small. In any case, $\Delta = \Theta(1)$ and positive for sufficiently small $\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\|$ and $\|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|$. This concludes the proof. □

We now are ready to give an estimation of the search directions obtained when iterates approach the face $\mathcal{F}$ of the optimal set. This is a less precise analogue of Lemma 4.3 in our previous paper [12] that does not require a nondegeneracy assumption.

Theorem 4.1 *Let $\mathcal{F}$, F_x^+, F_x^0, F_s^+, and F_s^0 be as defined in Section 4.1. If $\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\| \to 0$ and $\|S_{F_s^+}^{-1}\|\|S_{F_s^0}\| \to 0$, the search direction defined by (29) and (30) satisfies the following relations:*

$$X^{-1}\Delta x = -\frac{1}{\rho-\beta_3-1}\begin{bmatrix}\Xi_+ e\\ e\end{bmatrix} + \mathcal{O}\Big(\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\| + \|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|\Big), \quad (81)$$

$$S^{-1}\Delta s = -\frac{1}{\rho-\beta_3-1}\begin{bmatrix}\Sigma_+ e\\ e\end{bmatrix} + \mathcal{O}\Big(\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\| + \|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|\Big). \quad (82)$$

Proof:
First, note that Ξ and Σ are both orthogonal projection matrices, and $\nu = \Theta(1)$. Hence, the vectors $\Xi\nu$ and $\Sigma\nu$ are both $\mathcal{O}(1)$ vectors. Recall from Lemma 4.4 that $(1-\beta_2) = \mathcal{O}\Big(\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\| + \|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|\Big)$, and that $\Delta = \Theta(1)$. Therefore, both of the terms $-\frac{\rho(1-\beta_2)}{\Delta}\Xi\nu$ and $-\frac{\rho(1-\beta_2)}{\Delta}\Sigma\nu$ are $\mathcal{O}\Big(\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\| + \|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|\Big)$.

Using Lemmas 4.1 and 4.2 we also have:

$$\Xi e = \begin{bmatrix}\Xi_+ e\\ \Xi_0 e\end{bmatrix} + \mathcal{O}\Big(\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\|\Big) = \begin{bmatrix}\Xi_+ e\\ e\end{bmatrix} + \mathcal{O}\Big(\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\|\Big), \quad (83)$$

$$\Sigma e = \begin{bmatrix}\Sigma_+ e\\ \Sigma_0 e\end{bmatrix} + \mathcal{O}\Big(\|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|\Big) = \begin{bmatrix}\Sigma_+ e\\ e\end{bmatrix} + \mathcal{O}\Big(\|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|\Big). \quad (84)$$

Recalling that $\Delta = (\rho\beta_1 - 1)(\rho - \beta_3 - 1) + \rho(1-\beta_2)^2$ and using Lemma 4.4, we observe that

$$\frac{\rho\beta_1 - 1}{\Delta} = \frac{1}{\rho-\beta_3-1} + \mathcal{O}\Big(\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\| + \|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|\Big).$$

Now the assertion of the theorem follows by noting that $\|(\Xi_+ e, \Sigma_+ e)\| = \Theta(1)$. □

Theorem 4.1 forms, in a sense, the backbone of our analysis. We will later see that $\Xi_+ e$ and $\Sigma_+ e$ approach zero as iterates approach $\mathcal{F}$. This, in combination with Theorem 4.1, indicates that search directions of our algorithm approach the affine-scaling direction. This is a promising feature of our search directions since it is necessary to phase out centering from search directions to ensure eventual superlinear convergence.

Also, observe that, if $\mathcal{F}$ is a vertex, elementary theory of linear programming indicates that the columns of the matrix $A_{F_x^+}$ are linearly independent, i.e., $N_x = \emptyset$. Therefore, $R_{B_x F_x^+}$ and $\overline{\Xi}_+$ are identity matrices, and $\Xi_+ e = 0$. An identical result holds for $\Sigma_+ e$.

4.4 Accumulation Points

In this subsection, we will prove that the iterates of Algorithm 1 has a unique accumulation point. To this end, we first prove that all the accumulation points of our algorithm lie in the interior of the face $\mathcal{F}$ defined above. This result is an extension of a theorem by Tsuchiya [11], who proves a similar result for the Iri-Imai algorithm. Next, we will demonstrate that even when $\mathcal{F}$ is not a vertex, the accumulation point is unique: the relative analytical center of $\mathcal{F}$.

Theorem 4.2 (Theorem 4.2, [11]) *Let $\mathcal{F}$ be as defined above. Then, all the accumulation points of the algorithm described in Section 2 lie in the relative interior of $\mathcal{F}$.*

Proof:
A detailed proof of the statement of the theorem in the primal-only setting is given in [11]. His proof can be generalized into our primal-dual setting in a straightforward manner. Therefore, we only give a sketch of the proof and skip some details. Recall that, by definition, $\mathcal{F}$ contains in its relative interior an accumulation point of our iteration sequence with the maximum number of zeros. Since (x^*, s^*) is in the relative interior of $\mathcal{F}$ we must have that $x^*_{F_x^+} > 0$ and that $s^*_{F_s^+} > 0$. Also, by definition, $x^*_{F_x^0} = 0$ and $s^*_{F_s^0} = 0$.

Note that, no other acceleration point (x, s) can have $x_{F_x^0} = 0$ and $s_{F_s^0} = 0$, as well as $x_i = 0$ for some $i \in F_x^+$ or $s_i = 0$ for some $i \in F_s^+$. Therefore, for large enough k, having small $\|X^k_{F_x^0}\|$ and $\|S^k_{F_s^0}\|$ will imply that $\|(X^k_{F_x^+})^{-1}\|$ and $\|(S^k_{F_s^+})^{-1}\|$ are both bounded. Consequently, we can choose a large enough index k, for which $\|(X^k_{F_x^+})^{-1}\|\|X^k_{F_x^0}\| + \|(S^k_{F_s^+})^{-1}\|\|S^k_{F_s^0}\|$ is sufficiently small.

Since Algorithm 1 uses step sizes that are at least 1/5, using Theorem 4.1 it is observed that the components of the vectors $x^{k+1}_{F_x^0}$ and $s^{k+1}_{F_s^+}$ will be smaller than those of $x^k_{F_x^0}$ and $s^k_{F_s^+}$. Furthermore, k can be chosen large enough so that $x^{k+1}_{F_x^0} \leq \gamma x^k_{F_x^0}$, for some constant $\gamma \in (0, 1)$. Then, repeating this argument with the new iterate one observes that $x^k_{F_x^0}$ and $s^k_{F_s^0}$ both converge to zero and that they do so at least linearly. So, $x_{F_x^0}$ and $s_{F_s^0}$ are both zero for any accumulation point (x, s) of our algorithm. Now, using the maximality of the number of zeros of (x^*, s^*), we conclude that $x_{F_x^+}$ and $s_{F_s^+}$ must be strictly positive for any accumulation point (x, s) of the algorithm. This completes the proof of the theorem. □

Since the accumulation points of our algorithm are in the relative interior of the face $\mathcal{F}$, the vectors $x_{F_x^+}$ and $s_{F_s^+}$ remain bounded away from zero. Therefore, the error terms $\mathcal{O}\Big(\|X_{F_x^+}^{-1}\|\|X_{F_x^0}\|\Big)$ and $\mathcal{O}\Big(\|S_{F_s^+}^{-1}\|\|S_{F_s^0}\|\Big)$ in Lemmas 4.1, 4.4, and Theorem 4.1 can be replaced by $\mathcal{O}\Big(\|X_{F_x^0}\|\Big)$ and $\mathcal{O}\Big(\|S_{F_s^0}\|\Big)$.

At the end of Section 4.3 we observed that $\Xi_+ e$ and $\Sigma_+ e$ are both zero vectors if $\mathcal{F}$ is a vertex. Next, we prove that these vectors converge to vectors of zeros when $\mathcal{F}$ is not a vertex. This result is also based on a similar result by Tsuchiya given for the primal-only version of our algorithm.

Lemma 4.5 *Let $\mathcal{F}$, Ξ_+, and Σ_+ be as defined in Section 4.1. If $\mathcal{F}$ is not a vertex, $\Xi_+ e$ and $\Sigma_+ e$ converge to vectors of zeros.*

Proof:
Note that $\|\Xi_+ e\|_2^2 = e^T \Xi_+ e$ and $\|\Sigma_+ e\|_2^2 = e^T \Sigma_+ e$. We will show that $\sum_k (e^T \Xi_+^k e + e^T \Sigma_+^k e)$ is a convergent series, which immediately implies the lemma. The superscripts in Ξ_+^k and Σ_+^k signify the dependence on the iterate (x^k, s^k).

Consider the function defined below:

$$\Lambda_{\mathcal{F}}(x,s) \quad := \quad \sum_{i \in F_x^+} \ln x_i + \sum_{i \in F_s^+} \ln s_i. \tag{85}$$

Since Λ is a concave function, the change in its value from one iteration of our algorithm to the next can be bounded above by the change in its linear approximation at the current iterate. Thus,

$$\begin{aligned}
\Lambda_{\mathcal{F}}(x+\alpha\Delta x, s+\alpha\Delta s) - \Lambda_{\mathcal{F}}(x,s) &\le \alpha e^T \nabla \Lambda_{\mathcal{F}}(x,s) \begin{bmatrix} \Delta x \\ \Delta s \end{bmatrix} \\
&= -\alpha \tfrac{1}{\rho-\beta_3-1}(e^T \Xi_+ e + e^T \Sigma_+ e) + \mathcal{O}\Big(\|X_{F_x^0}\| + \|S_{F_s^0}\|\Big) \\
&\le -\alpha \tfrac{1}{\rho-2n-1}(e^T \Xi_+ e + e^T \Sigma_+ e) + \mathcal{O}\Big(\|X_{F_x^0}\| + \|S_{F_s^0}\|\Big).
\end{aligned}$$

The equality above follows from the fact that $\nabla \Lambda_{\mathcal{F}}(x,s) = \mathrm{diag}(X_{F_x^+}^{-1} e, S_{F_s^+}^{-1} e)$ and Theorem 4.1. The second inequality uses 77.

Since, by Theorem 4.2, all accumulation points of our algorithm are in the relative interior of $\mathcal{F}$, x_i^k and s_j^k remain bounded away from zero for $i \in F_x^+$ and $j \in F_s^+$. Therefore, if we add the above inequality for successive iterates, the telescopic series on the left will remain bounded below. Furthermore, from the proof of Theorem 4.2 we have that $x_{F_x^0}^k$ and $s_{F_s^0}^k$ both converge to zero, at least linearly. So, the sum of the $\mathcal{O}\Big(\|X_{F_x^0}\| + \|S_{F_s^0}\|\Big)$ terms on the right-hand-side of the inequality is bounded above by a constant independent from the number of terms being added. Recalling that the stepsize α is at least $\frac{1}{5}$ in every iteration of Algorithm 1, this last conclusion indicates that the sum $\sum_k (e^T \Xi_+^k e + e^T \Sigma_+^k e)$ is also bounded above. Since, this is a monotone increasing series, it is convergent. This completes our proof. □

Recall from Lemma 4.3 that $\Xi_+ \nu_{F_x^+} = 0$ and $\Sigma_+ \nu_{F_s^+} = 0$, while from Lemma 4.5 $\Xi_+ e \to 0$ and $\Sigma_+ e \to 0$ when $\mathcal{F}$ is not a vertex. This happens because $\nu_{F_x^+ \cup F_s^+}$ and e are asymptotically collinear. Indeed, the iterates converge to the relative analytical center of the face $\mathcal{F}$ when it is not a vertex.

Theorem 4.3 *The iterates converge to the relative analytical center of the face $\mathcal{F}$.*

Proof:
The relative analytical center of the face $\mathcal{F}$ is the unique point that minimizes $-\Lambda_{\mathcal{F}}(x, s) = -\sum_{i \in F_x^+} \ln x_i - \sum_{i \in F_s^+} \ln s_i$ among all the points on $\mathcal{F}$. KKT conditions for this optimization problem indicate that there exist vectors y and w such that:

$$\begin{aligned} X_{F_x^+}^{-1} e &= -A_{F_x^+}^T y, \\ S_{F_s^+}^{-1} e &= -G_{F_s^+}^T w, \\ A_{F_x^+} x_{F_x^+} &= b, \\ G_{F_s^+} s_{F_s^+} &= Gc, \end{aligned}$$

as well as $(x_{F_x^0}, s_{F_s^0}) = 0$, and $(x_{F_x^+}, s_{F_s^+}) > 0$. All but the first two conditions are satisfied by all points in the relative interior of $\mathcal{F}$. Therefore, it suffices to prove that any accumulation point also satisfies the first two conditions. Since all the accumulation points are in the relative interior of the face $\mathcal{F}$, x_{F_x+} is bounded away from zero for all iterates, and the matrix $R_{B_x F_x^+} = \left[\begin{array}{cc} I_{|B_x|} & X_{B_x}^{-1} \overline{A}_{B_x} A_{N_x} X_{N_x} \end{array}\right]$ is well-defined and varies continuously with x. Therefore, Ξ_+, the orthogonal projection matrix into the null space of $R_{B_x F_x^+}$, is also well-defined and a continuous function of x. Since $\Xi_+ e \to 0$, as iterates approach $\mathcal{F}$, $\Xi_+ e = 0$ for an accumulation point (x, s) on $\mathcal{F}$ of our iteration sequence. Therefore, e is in the range space of $R_{B_x F_x^+}^T$, which, as we have seen in the proof of Lemma 4.3, is the same as the range space of $X_{F_x^+} A^T_{F_x^+}$. This indicates that $X^{-1}_{F_x^+} e$ is in the range space of $A^T_{F_x^+}$, i.e., there exists a vector y such that

$$X^{-1}_{F_x^+} e \quad = \quad -A^T_{F_x^+} y.$$

Similarly, for an accumulation point (x, s) on $\mathcal{F}$ of our iteration sequence we have that $\Sigma_+ e = 0$, which leads to an analogous conclusion: There exists a vector w such that

$$S^{-1}_{F_s^+} e \quad = \quad -G^T_{F_s^+} w.$$

Therefore, our accumulation points in the relative interior of the face $\mathcal{F}$ all satisfy the first two equations defining the relative analytical center of this face. Since this center is unique, there is a single accumulation point of our iteration sequence and it is the relative analytical center of the face $\mathcal{F}$. $\square$

5 Quadratic Convergence

We will prove a Q-quadratic convergence result for our algorithm in this section. This objective requires a delicate balance in the selection of the step sizes in final iterations of Algorithm 1. Larger steps necessary for quadratic convergence may not guarantee reduction in the potential function while such reductions are necessary for polynomial convergence.

Recall that $\alpha^k_{\max}$ is defined to be the largest feasible step size in iteration k. We will choose stepsizes of the form $\alpha^k = \tau^k \alpha^k_{\max} \geq 1/5$ where $\tau^k = 1 - \Theta\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right) < 1$ as given in Algorithm 1. This strategy ensures that more aggressive steps can be taken as we approach the optimal set while it also keeps us a healthy distance away from the boundary so that the barrier terms in the potential function do not explode and the function can still be reduced. Our first result in this section indicates that this stepsize selection strategy is consistent with potential reduction:

Lemma 5.1 *Let Δx and Δs be given by (29) and (30). Let $\alpha^k = \tau^k \alpha^k_{\max}$ with $\tau^k = 1 - \Theta\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right) < 1$, and $(x^{k+1}, s^{k+1}) = (x^k, s^k) + \alpha^k(\Delta x, \Delta s)$. Then,*

$$\Phi_\rho(x^{k+1}, s^{k+1}) - \Phi_\rho(x^k, s^k) \quad \leq \quad (\rho - n) \ln\left(\mathcal{O}\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right)\right).$$

Proof:
To simplify the notation we let $(x^+, s^+) = (x^{k+1}, s^{k+1})$, $(x, s) = (x^k, s^k)$, and $\alpha = \alpha^k$. From Proposition 2.2,

$$\begin{aligned}
&\Phi_\rho(x^+, s^+) - \Phi_\rho(x, s) \\
&= \rho \ln\left(\frac{(x^+)^T s^+}{x^T s}\right) - \sum_{i=1}^n \ln\left(1 + \alpha\frac{\Delta x_i}{x_i}\right) - \sum_{i=1}^n \ln\left(1 + \alpha\frac{\Delta s_i}{s_i}\right) \\
&= \rho \ln\left(1 - \frac{\alpha}{\Delta}(\rho\beta_1 - \beta_2)\right) - \sum_{i=1}^n \ln\left(1 + \alpha\frac{\Delta x_i}{x_i}\right) - \sum_{i=1}^n \ln\left(1 + \alpha\frac{\Delta s_i}{s_i}\right).
\end{aligned}$$

Next, note that from Theorem 4.1 and Lemma 4.5 we have $\alpha_{\max} = (\rho - \beta_3 - 1) + \mathcal{O}\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right)$. Recalling the definition of Δ and using Lemma 4.4, we conclude that

$$\frac{\alpha}{\Delta}(\rho\beta_1 - \beta_2) \;=\; 1 - \Theta\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right).$$

From Theorem 4.1 and Lemma 4.5 we also have that

$$\begin{aligned}
1 + \alpha\frac{\Delta x_i}{x_i} &= \Theta\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right) \; i \in F_x^0, \qquad 1 + \alpha\frac{\Delta x_i}{x_i} = \Theta(1) \; i \in F_x^+ \;, \\
1 + \alpha\frac{\Delta s_i}{s_i} &= \Theta\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right) \; i \in F_s^0, \qquad 1 + \alpha\frac{\Delta s_i}{s_i} = \Theta(1) \; i \in F_s^+.
\end{aligned}$$

From the definition of the order notation, there exist positive real numbers r and R independent from the current iterate (x, s) such that, any number $z = \Theta\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right)$ satisfies

$$r\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right) \le z \le R\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right)$$

and any number $z = \mathcal{O}\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right)$ satisfies $z \le R\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right)$. Using the asymptotic relations above, we obtain:

$$\begin{aligned}
&\Phi_\rho(x^+, s^+) - \Phi_\rho(x, s) \\
&\le \rho \ln(R\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right)) - n\ln(r\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right)) - \mathcal{O}(1) \\
&= (\rho - n)\ln(\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right)) + \mathcal{O}(1) \\
&= (\rho - n)\ln\left(\mathcal{O}\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right)\right).
\end{aligned}$$

Therefore, as $\|X_{F_x^0}\| + \|S_{F_s^0}\|$ tends to zero, the reduction in the potential function tends to $-\infty$. ∎

Next, we observe that with the given stepsize selection strategy the variables that converge to zero do so quadratically:

Lemma 5.2 *Let Δx and Δs be given by (29) and (30). Let $\alpha^k = \tau^k \alpha^k_{\max}$ with $\tau^k = 1 - \Theta\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right) < 1$, and $(x^{k+1}, s^{k+1}) = (x^k, s^k) + \alpha^k(\Delta x, \Delta s)$. Then,*

$$\|(x^{k+1}_{F_x^0}, s^{k+1}_{F_s^0})\| \quad \leq \quad C_1 \|(x^k_{F_x^0}, s^k_{F_s^0})\|^2,$$

where C_1 is a constant independent from k.

Proof:
As in the proof of Lemma 5.1,

$$x_i^{k+1} \quad = \quad x_i^k \left(1 + \alpha \frac{\Delta x_i^k}{x_i^k}\right) = x_i^k \Theta\left(\|X^k_{F_x^0}\| + \|S^k_{F_s^0}\|\right),$$

for $i \in F_x^0$, and similarly for s_i^{k+1} with $i \in F_s^0$. ∎

Corollary 5.1 *Let Δx and Δs be given by (29) and (30). Let $(x^{k+1}, s^{k+1}) = (x^k, s^k) + \alpha^k(\Delta x, \Delta s)$ with $\alpha^k = \tau^k \alpha^k_{\max}$ and $\tau^k = 1 - \Theta\left(\|X_{F_x^0}\| + \|S_{F_s^0}\|\right) < 1$. If $\mathcal{F}$ is a vertex, then the sequence (x^k, s^k) converges to $\mathcal{F}$ Q-quadratically.*

Proof:
Let $\mathcal{F} = (x^*, s^*)$. It suffices to prove that $x^k_{F_x^+} = x^k_{B_x}$ converges to $x^*_{B_x}$ quadratically and likewise for s variables. Note that $x^*_{B_x} = \overline{A}_{B_x} b$ and since our iterates are feasible, $x^k_{B_x} = \overline{A}_{B_x} b - \overline{A}_{B_x} A_{F_x^0} x_{F_x^0}$. Now, the result follows from Lemma 5.2. The proof with the dual variables is identical. ∎

The task that remains is to prove Q-quadratic convergence for the case when $\mathcal{F}$ is not a vertex. As in the corollary above, we will reach this conclusion by showing that $\|x^k_{F_x^+} - x^*_{F_x^+}\|$ is bounded by a scalar multiple of $\|X^k_{F_x^0}\| + \|S^k_{F_s^0}\|$, and similarly for the dual variables. To prove this result we first observe that the F_x^+ components of the search directions generated by Algorithm 1 converge to a *negative* scalar multiple of the Newton direction for a system of nonlinear equations with a nonsingular Jacobian matrix at the solution of the system. More precisely, we will demonstrate that the directions

$$\widehat{\Delta x_+} = X_{F_x^+}\Xi_+ e, \qquad \widehat{\Delta s_+} = S_{F_s^+}\Sigma_+ e \tag{86}$$

are Newton directions for a nonlinear system whose solution is the convergence point of our algorithm.

Let (x^*, s^*) be the relative analytical center of the face $\mathcal{F}$. Recall the definitions of the index sets B_x, N_x, etc. from Section 4.1. Also recall that $\Gamma_{N_x B_x}$ was defined as the matrix satisfying $A_{N_x} = A_{B_x}\Gamma_{N_x B_x}$. Below, we will refer to this matrix as Γ (without the subscripts) for simplicity. Consider the following system:

$$x_{B_x} + \Gamma x_{N_x} = \overline{A}_{B_x} b, \tag{87}$$

$$\Gamma^T X_{B_x}^{-1} e - X_{N_x}^{-1} e = 0. \tag{88}$$

We first show that $x^*_{F_x^+}$ solves the system above. Recall the KKT conditions given in the proof of Theorem 4.3 that define x^*. Premultiplying the equality $A_{F_x^+} x^*_{F_x^+} = b$ by $\overline{A}_{B_x}$ we observe that $x^*_{F_x^+}$ satisfies (87). Next, by partitioning the equality $X_{F_x^+}^{-1} e = -A_{F_x^T}^T y$ we get $(X^*_{B_x})^{-1} e = -A_{B_x}^T y$, and $(X^* N_x)^{-1} e = -A_{N_x}^T y = -\Gamma^T A_{B_x}^T y$, from which it follows that $x^*_{F_x^+}$ also satisfies (88).

Next, we observe that the Jacobian of the system (87)–(88) is nonsingular at $x^*_{F_x^+}$. Indeed, the Jacobian can be partitioned into components corresponding to B_x and N_x:

$$J(x) = \begin{bmatrix} I & \Gamma \\ -\Gamma^T X_{B_x}^{-2} & X_{N_x}^{-2} \end{bmatrix}$$

Since $x^*_{F_x^+} > 0$, both diagonal matrices $X_{B_x}^{-2}$ and $X_{N_x}^{-2}$ are strictly positive at $x = x^*_{F_x^+}$. If $J(x^*)[u^T, \;\; v^T]^T = 0$, the first block indicates that $u = -\Gamma v$. Substituting it in the second block we get $(\Gamma^T X_{B_x}^{-2}\Gamma + X_{N_x}^{-2})v = 0$, which implies that $v = 0$ since the left-hand-side matrix of this equation is positive definite. Now, $u = 0$ also follows and therefore $J(x^*)$ is nonsingular.

Next, we will observe that the Newton direction for finding a zero of the system (87)–(88) coincides with $\widehat{\Delta x_+}$ given in (86). Let Δx_N be the Newton direction for the system (87)–(88) and let $\overline{\Delta x_N} = X_{F_x^+}^{-1}\Delta x_N$. Then, $\overline{\Delta x_N} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$ satisfies

$$\begin{bmatrix} X_{B_x} & \Gamma X_{N_x} \\ -\Gamma^T X_{B_x}^{-1} & X_{N_x}^{-1} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = - \begin{bmatrix} 0 \\ \Gamma^T X_{B_x}^{-1} e - X_{N_x}^{-1} e \end{bmatrix}.$$

On the other hand, $\Xi_+ e$ is the orthogonal projection of the vector e into the null space of $R_{B_x F_x^+} = \begin{bmatrix} I & X_{B_x}^{-1}\overline{A}_{B_x} A_{N_x} X_{N_x} \end{bmatrix} = \begin{bmatrix} I & X_{B_x}^{-1}\Gamma X_{N_x} \end{bmatrix}$ which is the same as the null space of $R := \begin{bmatrix} X_{B_x} & \Gamma X_{N_x} \end{bmatrix}$. Now note that $\overline{\Delta x_N}$ is in the null space of R because of the first block equation in the system above. Let $y = X_{B_x}^{-1}e - X_{B_x}^{-1}d_1$ and observe that $X_{N_x}\Gamma^T y = e - d_2$. In other words, $R^T y = e - \overline{\Delta x_N}$, i.e., the vectors $\overline{\Delta x_N}$ and $e - \overline{\Delta x_N}$ are in orthogonal spaces (null space of R and range space of R^T) and their sum is e. Therefore, $\overline{\Delta x_N}$ is the projection of the vector e into $\mathcal{N}(R)$, which is exactly $\Xi_+ e$. So, $\Delta x_N = X_{F_x^+}\overline{\Delta x_N} = X_{F_x^+}\Xi_+ e = \widehat{\Delta x_+}$. Identical arguments can be carried out with s variables also. Thus, we proved:

Lemma 5.3 *Let $\mathcal{F}$ be a face of dimension at least one of the optimal solution set containing the convergence point (x^*, s^*) of Algorithm 1 in its relative interior. Let $\widehat{\Delta x}$ and $\widehat{\Delta s}$ be given by (86). These directions are Newton search directions for particular nonlinear systems which have their solutions at x^* and s^*, respectively, and have nonsingular Jacobians at these points.* □

Lemma 5.4 *Let (x^*, s^*) be the convergence point of Algorithm 1. If stepsizes in Algorithm 1 are chosen as in Lemma 5.1, then there exists a positive constant C_3 such that the inequalities*

$$\|x^k_{F_x^+} - x^*_{F_x^+}\| \leq C_3\Big(\|x^k_{F_x^0}\| + \|s^k_{F_s^0}\|\Big), \tag{89}$$

$$\|s^k_{F_s^+} - s^*_{F_s^+}\| \leq C_3\Big(\|x^k_{F_x^0}\| + \|s^k_{F_s^0}\|\Big) \tag{90}$$

are satisfied in all but a finite number of iterations.

Proof:
We will show that if the inequalities above are not satisfied for a certain C_3 then the term on the left-hand-side of each inequality grows significantly from one iteration to the other. From Lemma 5.3, using the standard theory of Newton updates, we have that there exist a neighborhood of x^* and a scalar $0 < \gamma < 1/2$ such that

$$\|x^k_{F_x^+} + X^k_{F_x^+}\Xi_+ e - x^*_{F_x^+}\| \leq \gamma \|x^k_{F_x^+} - x^*_{F_x^+}\|. \tag{91}$$

Also observe that,

$$\begin{aligned} x^{k+1}_{F_x^+} &= x^k_{F_x^+} - \alpha^k \left[-\frac{1}{\rho - \beta_3 - 1} X^k_{F_x^+}\Xi_+ e + \mathcal{O}\left(\|X^k_{F_x^0}\| + \|S^k_{F_s^0}\|\right)\right] \\ &= x^k_{F_x^+} - X^k_{F_x^+}\Xi_+ e + \Theta\left(\|X^k_{F_x^0}\| + \|S^k_{F_s^0}\|\right). \end{aligned}$$

Above, the first equality follows from Theorem 4.1 and the second equality follows from the fact that $\alpha^k = (\rho - \beta_3 - 1) + \Theta\left(\|X^k_{F^0_x}\| + \|S^k_{F^0_s}\|\right)$. Therefore,

$$\begin{aligned}
\|x^{k+1}_{F^+_x} - x^*_{F^+_x}\| &= \|x^k_{F^+_x} - X^k_{F^+_x}\Xi_+ e - x^*_{F^+_x}\| + \Theta\left(\|X^k_{F^0_x}\| + \|S^k_{F^0_s}\|\right) \\
&= \|2(x^k_{F^+_x} - x^*_{F^+_x}) - (x^k_{F^+_x} + X^k_{F^+_x}\Xi_+ e - x^*_{F^+_x})\| + \Theta\left(\|X^k_{F^0_x}\| + \|S^k_{F^0_s}\|\right) \\
&\geq (2-\gamma)\|x^k_{F^+_x} - x^*_{F^+_x}\| - C_2\left(\|X^k_{F^0_x}\| + \|S^k_{F^0_s}\|\right)
\end{aligned}$$

where C_2 is a positive constant and the last inequality uses (91). Let $C_3 = \frac{C_2}{\gamma}$. Then, if (89) is violated, we have that $\|x^{k+1}_{F^+_x} - x^*_{F^+_x}\| \geq (2-2\gamma)\|x^k_{F^+_x} - x^*_{F^+_x}\|$, where $2-2\gamma$ is a constant greater than 1. Furthermore, as long as the current iterate remains in the neighborhood of x^* where (91) holds, all the arguments remain true (note that $\|X^k_{F^0_x}\| + \|S^k_{F^0_s}\|$ decreases to zero quadratically) and the norm of the residual $x^k_{F^+_x} - x^*_{F^+_x}$ is multiplied by at least $2-2\gamma$ in each iteration. Since (x^*, s^*) is the only accumulation point of Algorithm 1, the iterates cannot leave a neighborhood of (x^*, s^*) infinitely many times. Therefore, (89) can be violated only finitely many times. An identical proof works for the second inequality of the Lemma. □

Now, we are ready to prove:

Theorem 5.1 *Iterates of Algorithm 1 converge Q-quadratically to* (x^*, s^*).

Proof:
In view of Lemmas 5.4 and 5.2, there exist positive constants C_4 and C_5 such that

$$\begin{aligned}
\|(x^{k+1}, s^{k+1}) - (x^*, s^*)\| &\leq C_4\|(x^{k+1}_{F^0_x}, s^{k+1}_{F^0_s})\| \leq C_5\|(x^k_{F^0_x}, s^k_{F^0_s})\|^2 \\
&\leq C_5\|(x^k, s^k) - (x^*, s^*)\|^2. \qquad (92)
\end{aligned}$$

□

6 Conclusion

We developed a potential-reduction algorithm that converges to optimal solutions of linear programming problems quadratically. This fast convergence is obtained even with degenerate problems and *without* making any path-following restrictions on the iterates. To the best of our knowledge, this is the first such result for a primal-dual interior-point algorithm. Finally, what we have is a Q-order quadratic convergence result rather than the inferior R-order convergence as Todd anticipated for potential-reduction algorithms [8].

The only significant assumption we made to achieve the quadratic convergence result is the existence (and the availability) of a strictly feasible solution to the primal-dual pair of problems. Under this assumption, the optimal solution set is bounded, which is often a necessary assumption for the convergence of potential-reduction methods. For any linear programming problem, there is a corresponding artificial problem, a so-called homogeneous self-dual LP, which has a known strictly feasible solution [14]. Furthermore, certain solutions to this artificial problem provide either the solution to the original primal-dual pair of problems or demonstrate that one (or both) of these problems must be infeasible. For our potential-reduction algorithm, however, it is not clear that the solutions generated will produce such solutions, i.e., to use the notation of [14], it is not clear that we will have $\kappa^* + \tau^* > 0$. Indeed, the behavior of potential-reduction methods on the self-dual homogeneous formulations is not well understood and remains to be investigated.

One of the main difficulties in developing a complete quadratic convergence theory for potential-reduction methods is the possibility of having a convergence point that is not a vertex. While this possibility exists and cannot be excluded, our preliminary computational experiments demonstrated that such situations are rare. In fact, in all but a few specially constructed examples, we observed convergence to vertices of the optimal solution set. This phenomenon is contrary to the behavior of most interior-point algorithms that converge to the relative analytical center of the full optimal face. It may be possible to get more insight into this behavior by analyzing the existence and convergence of certain trajectories that have the property that the tangent to the trajectory at a given point is equal to the search direction used by our algorithm. This, also, is a topic for future research.

7 Proof of Lemma 4.1

Below, positive definite and positive semidefinite will be abbreviated as pd and psd. Using the partition of the matrix AX given in (63), we evaluate the matrix

AX^2A^T:

$$\begin{aligned} AX^2A^T &= BX_B \begin{bmatrix} R_{B_xF_x^+}R^T_{B_xF_x^+} + R_{B_xF_x^0}R^T_{B_xF_x^0} & R_{B_xF_x^0}R^T_{\overline{B}_xF_x^0} \\ R_{\overline{B}_xF_x^0}R^T_{B_xF_x^0} & R_{\overline{B}_xF_x^0}R^T_{\overline{B}_xF_x^0} \end{bmatrix} X_B B^T \\ &= BX_B \begin{bmatrix} R_{B_xF_x^+}R^T_{B_xF_x^+} + R_{B_x\overline{N}_x}R^T_{B_x\overline{N}_x} & R_{B_x\overline{N}_x}R^T_{\overline{B}_x\overline{N}_x} \\ R_{\overline{B}_x\overline{N}_x}R^T_{B_x\overline{N}_x} & I + R_{\overline{B}_x\overline{N}_x}R^T_{\overline{B}_x\overline{N}_x} \end{bmatrix} X_B B^T \end{aligned}$$

where $B = A_{B_x \cup \overline{B}_x}$ and $X_B = X_{B \cup \overline{B}_x}$. Let Z denote the 2×2 block matrix in the middle of the right-hand-side expression above. Z can be written as

$$Z = \begin{bmatrix} R_{B_xF_x^+}R^T_{B_xF_x^+} & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} R_{B_x\overline{N}_x} \\ R_{\overline{B}_x\overline{N}_x} \end{bmatrix} \begin{bmatrix} R_{B_x\overline{N}_x} \\ R_{\overline{B}_x\overline{N}_x} \end{bmatrix}^T.$$

To determine the projection matrix into the null space of AX, we will need to evaluate the inverse of AX^2A^T, and consequently, of Z. Recall the Sherman-Morrison-Woodbury formula:

$$(E + UV^T)^{-1} = E^{-1} - E^{-1}U\left(I + V^TE^{-1}U\right)^{-1}V^TE^{-1}. \quad (93)$$

Let $W = (R_{B_xF_x^+}R^T_{B_xF_x^+})^{-1} = (I + R_{B_xN_x}R^T_{B_xN_x})^{-1}$. The inverse exists since the term inside the parentheses is the sum of the identity matrix and a psd matrix, and therefore is pd. Note also that W itself is pd. Using (93) and letting $U = \left(I + R^T_{B_x\overline{N}_x}WR_{B_x\overline{N}_x} + R^T_{\overline{B}_x\overline{N}_x}R_{\overline{B}_x\overline{N}_x}\right)^{-1}$, we obtain the following expression for the inverse of the matrix Z:

$$\begin{aligned} Z^{-1} &= \begin{bmatrix} W & 0 \\ 0 & I \end{bmatrix} - \begin{bmatrix} WR_{B_x\overline{N}_x} \\ R_{\overline{B}_x\overline{N}_x} \end{bmatrix} U \begin{bmatrix} R^T_{B_x\overline{N}_x}W & R^T_{\overline{B}_x\overline{N}_x} \end{bmatrix} \\ &= \begin{bmatrix} W - WR_{B_x\overline{N}_x}UR^T_{B_x\overline{N}_x}W & -WR_{B_x\overline{N}_x}UR^T_{\overline{B}_x\overline{N}_x} \\ -R_{\overline{B}_x\overline{N}_x}UR^T_{B_x\overline{N}_x}W & I - R_{\overline{B}_x\overline{N}_x}UR^T_{\overline{B}_x\overline{N}_x} \end{bmatrix} =: \begin{bmatrix} \overline{Z}_{11} & \overline{Z}_{12} \\ \overline{Z}^T_{12} & \overline{Z}_{22} \end{bmatrix}. \end{aligned}$$

Let us analyze U first. Since W is pd, $R^T_{B_x\overline{N}_x}WR_{B_x\overline{N}_x}$ is psd. Further, $R^T_{\overline{B}_x\overline{N}_x}R_{\overline{B}_x\overline{N}_x}$ is also psd and therefore U^{-1} and U are pd. Since all eigenvalues of U^{-1} are at least one, $\|U\|$, the operator norm of U, is at most 1.

Now, we are ready to evaluate the orthogonal projection matrix $\Xi = I - XA^T(AX^2A^T)^{-1}AX$:

$$\Xi = I - XA^T(AX^2A^T)^{-1}AX$$

$$
\begin{aligned}
&= I - \begin{bmatrix} R^T_{B_x F_x^+} & 0 \\ R^T_{B_x F_x^0} & R^T_{\overline{B}_x F_x^0} \end{bmatrix} Z^{-1} \begin{bmatrix} R_{B_x F_x^+} & R_{B_x F_x^0} \\ 0 & R_{\overline{B}_x F_x^0} \end{bmatrix} \\
&= \begin{bmatrix} \Xi_{F_x^+ F_x^+} & \Xi_{F_x^+ F_x^0} \\ \Xi_{F_x^0 F_x^+} & \Xi_{F_x^0 F_x^0} \end{bmatrix},
\end{aligned}
$$

with

$$
\begin{aligned}
\Xi_{F_x^+ F_x^+} &= I - R^T_{B_x F_x^+} \overline{Z}_{11} R_{B_x F_x^+} \\
\Xi_{F_x^+ F_x^0} &= -(R^T_{B_x F_x^+} \overline{Z}_{11} R_{B_x F_x^0} + R^T_{B_x F_x^+} \overline{Z}_{12} R_{\overline{B}_x F_x^0}) \quad = \quad \Xi^T_{F_x^0 F_x^+} \\
\Xi_{F_x^0 F_x^0} &= I - R^T_{B_x F_x^0} \overline{Z}_{11} R_{B_x F_x^0} - R^T_{B_x F_x^0} \overline{Z}_{12} R_{\overline{B}_x F_x^0} \\
&\quad - R^T_{\overline{B}_x F_x^0} \overline{Z}^T_{12} R_{B_x F_x^0} - R^T_{\overline{B}_x F_x^0} \overline{Z}_{22} R_{\overline{B}_x F_x^0}.
\end{aligned}
$$

We analyze the blocks of the matrix Ξ individually to establish the statement of the lemma. Let us start with the upper left corner. Recalling that $\overline{Z}_{11} = W - W R_{B_x \overline{N}_x} U R^T_{B_x \overline{N}_x} W$ and $W = (R_{B_x F_x^+} R^T_{B_x F_x^+})^{-1}$ we have:

$$
\begin{aligned}
\Xi_{F_x^+ F_x^+} &= \left[I - R^T_{B_x F_x^+} \left(R_{B_x F_x^+} R^T_{B_x F_x^+} \right)^{-1} R_{B_x F_x^+} \right] \\
&\quad + R^T_{B_x F_x^+} W R_{B_x \overline{N}_x} U R^T_{B_x \overline{N}_x} W R_{B_x F_x^+}. \qquad (94)
\end{aligned}
$$

The term in the square brackets is the orthogonal projection matrix Ξ_+ defined at the end of Section 4.1. Also note that

$$
\overline{\Xi}_+ = R^T_{B_x F_x^+} \left(R_{B_x F_x^+} R^T_{B_x F_x^+} \right)^{-1} R_{B_x F_x^+}.
$$

Define

$$
R_{F_x^+ \overline{N}_x} = \begin{bmatrix} R_{B_x \overline{N}_x} \\ 0 \end{bmatrix}, \qquad (95)
$$

where the zero matrix is $|N_x| \times |\overline{N}_x|$. Then, from (60) we have $R_{B_x \overline{N}_x} = R_{B_x F_x^+} R_{F_x^+ \overline{N}_x}$. Now, the second term in (94), which we will call $\Delta \Xi_{F_x^+ F_x^+}$, can be written as:

$$
\Delta \Xi_{F_x^+ F_x^+} = R^T_{B_x F_x^+} W R_{B_x \overline{N}_x} U R^T_{B_x \overline{N}_x} W R_{B_x F_x^+} \quad = \quad \overline{\Xi}_+ R_{F_x^+ \overline{N}_x} U R^T_{F_x^+ \overline{N}_x} \overline{\Xi}_+.
$$

Since $\|U\| \leq 1$, $\|\overline{\Xi}_+\| = 1$, and

$$
\|R_{F_x^+ \overline{N}_x}\| = \|R_{B_x \overline{N}_x}\| \quad = \quad \mathcal{O}(\|X^{-1}_{B_x}\| \|X_{\overline{N}_x}\|), \qquad (96)
$$

we have

$$\|\Delta\Xi_{F_x^+F_x^+}\| \;=\; \mathcal{O}(\|X_{B_x}^{-1}\|^2\|X_{\overline{N}_x}\|^2). \tag{97}$$

Next, we analyze the off-diagonal blocks:

$$\begin{aligned}\Xi_{F_x^+F_x^0} &= -(R^T_{B_xF_x^+}\overline{Z}_{11}R_{B_xF_x^0} + R^T_{B_xF_x^+}\overline{Z}_{12}R_{\overline{B}_xF_x^0}) \\ &= -\left(R^T_{B_xF_x^+}WR_{B_xF_x^0}\right) + \left(R^T_{B_xF_x^+}WR_{B_x\overline{N}_x}UR^T_{B_x\overline{N}_x}WR_{B_xF_x^0}\right) \\ &\quad + \left(R^T_{B_xF_x^+}WR_{B_x\overline{N}_x}UR^T_{\overline{B}_x\overline{N}_x}R_{\overline{B}_xF_x^0}\right).\end{aligned} \tag{98}$$

Since $R_{B_xF_x^0} = \begin{bmatrix} 0 & R_{B_x\overline{N}_x} \end{bmatrix}$ and $R_{B_x\overline{N}_x} = R_{B_xF_x^+}R_{F_x^+\overline{N}_x}$, we have

$$R_{B_xF_x^0} = R_{B_xF_x^+}\begin{bmatrix} 0 & R_{F_x^+\overline{N}_x} \end{bmatrix} \tag{99}$$

where the zero matrix is $|F_x^+| \times |\overline{B}_x|$. Therefore,

$$R^T_{B_xF_x^+}WR_{B_xF_x^0} \;=\; R^T_{B_xF_x^+}WR_{B_xF_x^+}\begin{bmatrix} 0 & R_{F_x^+\overline{N}_x} \end{bmatrix} = \overline{\Xi}_+\begin{bmatrix} 0 & R_{F_x^+\overline{N}_x} \end{bmatrix},$$

and therefore, $\|R^T_{B_xF_x^+}WR_{B_xF_x^0}\| = \mathcal{O}(\|X_{B_x}^{-1}\|\|X_{\overline{N}_x}\|)$. Similarly,

$$R^T_{B_xF_x^+}WR_{B_x\overline{N}_x}UR^T_{B_x\overline{N}_x}WR_{B_xF_x^0} \;=\; \overline{\Xi}_+R_{F_x^+\overline{N}_x}UR^T_{F_x^+\overline{N}_x}\overline{\Xi}_+\begin{bmatrix} 0 & R_{F_x^+\overline{N}_x} \end{bmatrix},$$

which indicates that $\|R^T_{B_xF_x^+}WR_{B_x\overline{N}_x}UR^T_{B_x\overline{N}_x}WR_{B_xF_x^0}\| = \mathcal{O}(\|X_{B_x}^{-1}\|^3\|X_{\overline{N}_x}\|^3)$.

The last term on the right hand side of (98) is

$$R^T_{B_xF_x^+}WR_{B_x\overline{N}_x}UR^T_{\overline{B}_x\overline{N}_x}R_{\overline{B}_xF_x^0} \;=\; \Xi_+R_{F_x^+\overline{N}_x}\begin{bmatrix} UR^T_{\overline{B}_x\overline{N}_x} & UR^T_{\overline{B}_x\overline{N}_x}R_{\overline{B}_x\overline{N}_x} \end{bmatrix} \tag{100}$$

Next, we will bound the two terms in square brackets through a more careful look at the U matrix. Recall that

$$U \;=\; \left(I + R^T_{B_x\overline{N}_x}WR_{B_x\overline{N}_x} + R^T_{\overline{B}_x\overline{N}_x}R_{\overline{B}_x\overline{N}_x}\right)^{-1}.$$

Note that

$$\begin{aligned}R^T_{B_x\overline{N}_x}WR_{B_x\overline{N}_x} &= R^T_{F_x^+\overline{N}_x}R^T_{B_xF_x^+}WR_{B_xF_x^+}R_{F_x^+\overline{N}_x} \\ &= R^T_{F_x^+\overline{N}_x}\overline{\Xi}_+R_{F_x^+\overline{N}_x} = R^T_{F_x^+\overline{N}_x}\overline{\Xi}_+^2R_{F_x^+\overline{N}_x}.\end{aligned}$$

The last equality follows from the fact that $\overline{\Xi}_+$ is a projection matrix. Using the Sherman-Morrison-Woodbury formula in (93) (with $E = I + R^T_{\overline{B}_x\overline{N}_x} R_{\overline{B}_x\overline{N}_x}$) and defining $V = (I + R^T_{\overline{B}_x\overline{N}_x} R_{\overline{B}_x\overline{N}_x})^{-1}$, we can write U as follows:

$$U \;=\; V - VR^T_{F_x^+\overline{N}_x}\overline{\Xi}_+\left(I + \overline{\Xi}_+ R_{F_x^+\overline{N}_x} V R^T_{F_x^+\overline{N}_x}\overline{\Xi}_+\right)^{-1}\overline{\Xi}_+ R_{F_x^+\overline{N}_x} V \quad (101)$$
$$\;=\; V - V\mathcal{O}(\|X_{B_x}^{-1}\|^2\|X_{\overline{N}_x}\|^2)V. \quad (102)$$

The second equation above holds because $\|(I + \Xi_+ R_{F_x^+\overline{N}_x} V R^T_{F_x^+\overline{N}_x}\Xi_+)^{-1}\| \leq 1$, $\|\Xi_+\| = 1$, and $\|R_{F_x^+\overline{N}_x}\| = \mathcal{O}(\|X_{B_x}^{-1}\|\|X_{\overline{N}_x}\|)$.

Since $VR^T_{\overline{B}_x\overline{N}_x} R_{\overline{B}_x\overline{N}_x} = I - V$, and the pd matrix V is less than or equal to the identity matrix in the Löwner sense, $\|VR^T_{\overline{B}_x\overline{N}_x} R_{\overline{B}_x\overline{N}_x}\| \leq 1$. If the singular values of $R_{\overline{B}_x\overline{N}_x}$ are denoted by λ_i, singular values of the matrix $VR^T_{\overline{B}_x\overline{N}_x}$ are $\frac{\lambda_i}{1+\lambda_i^2}$, all of which are between $-\frac{1}{2}$ and $\frac{1}{2}$. Therefore, $\|VR^T_{\overline{B}_x\overline{N}_x}\| \leq \frac{1}{2}$, and $\| \; UR^T_{\overline{B}_x\overline{N}_x} \quad UR^T_{\overline{B}_x\overline{N}_x} R_{\overline{B}_x\overline{N}_x} \; \| = \mathcal{O}(1)$. Combining this observation with (96) and (100), we obtain

$$\|R^T_{B_xF_x^+} W R_{B_x\overline{N}_x} U R^T_{\overline{B}_x\overline{N}_x} R_{\overline{B}_xF_x^0}\| \;=\; \mathcal{O}(\|X_B^{-1}\|\|X_{\overline{N}_x}\|). \quad (103)$$

Therefore,

$$\|\Xi_{F_x^+F_x^0}\| = \|\Xi_{F_x^0F_x^+}\| \;=\; \mathcal{O}(\|X_{B_x}^{-1}\|\|X_{\overline{N}_x}\|). \quad (104)$$

Next, we analyze the lower right corner:

$$\Xi_{F_x^0F_x^0} \;=\; I - \left(R^T_{B_xF_x^0}\overline{Z}_{11}R_{B_xF_x^0} + R^T_{B_xF_x^0}\overline{Z}_{12}R_{\overline{B}_xF_x^0} + R^T_{\overline{B}_xF_x^0}\overline{Z}^T_{12}R_{B_xF_x^0} + R^T_{\overline{B}_xF_x^0}\overline{Z}_{22}R_{\overline{B}_xF_x^0}\right)$$

Using (99), we have

$$R^T_{B_xF_x^0}\overline{Z}_{11}R_{B_xF_x^0} \;=\; \begin{bmatrix} 0 & R_{F_x^+\overline{N}_x} \end{bmatrix}^T (R^T_{B_xF_x^+}\overline{Z}_{11}R_{B_xF_x^+}) \begin{bmatrix} 0 & R_{F_x^+\overline{N}_x} \end{bmatrix},$$

where the middle term in parentheses is exactly the matrix $I - \Xi_{F_x^+F_x^+} = \mathcal{O}(1)$. Now, recalling (95) and (96), we have

$$\|R^T_{B_xF_x^0}\overline{Z}_{11}R_{B_xF_x^0}\| \;=\; \mathcal{O}(\|X_{B_x}^{-1}\|^2\|X_{\overline{N}_x}\|^2). \quad (105)$$

Similarly,

$$R^T_{B_xF_x^0}\overline{Z}_{12}R_{\overline{B}_xF_x^0} \;=\; \begin{bmatrix} 0 & R_{F_x^+\overline{N}_x} \end{bmatrix}^T R^T_{B_xF_x^+}\overline{Z}_{12}R_{\overline{B}_xF_x^0}.$$

From the analysis of the off-diagonal blocks we have that

$$\|R^T_{B_xF_x^+}\overline{Z}_{12}R_{\overline{B}_xF_x^0}\| \;\; = \;\; \mathcal{O}(\|X^{-1}_{B_x}\|\|X_{\overline{N}_x}\|),$$

and therefore,

$$\|R^T_{B_xF_x^0}\overline{Z}_{12}R_{\overline{B}_xF_x^0}\| = \|R^T_{\overline{B}_xF_x^0}\overline{Z}^T_{12}R_{B_xF_x^0}\| = \mathcal{O}(\|X^{-1}_{B_x}\|^2\|X_{\overline{N}_x}\|^2). \tag{106}$$

Finally, using (102) we have

$$\begin{aligned} & R^T_{\overline{B}_xF_x^0}\overline{Z}_{22}R_{\overline{B}_xF_x^0} = R^T_{\overline{B}_xF_x^0}(I - R_{\overline{B}_x\overline{N}_x}UR^T_{\overline{B}_x\overline{N}_x})R_{\overline{B}_xF_x^0} \\ & = \;\; R^T_{\overline{B}_xF_x^0}(I - R_{\overline{B}_x\overline{N}_x}VR^T_{\overline{B}_x\overline{N}_x})R_{\overline{B}_xF_x^0} \\ & = \;\; +R^T_{\overline{B}_xF_x^0}R_{\overline{B}_x\overline{N}_x}V\mathcal{O}(\|X^{-1}_{B_x}\|^2\|X_{\overline{N}_x}\|^2)VR^T_{\overline{B}_x\overline{N}_x}R_{\overline{B}_xF_x^0}. \end{aligned} \tag{107}$$

Note that,

$$\begin{aligned} I - R_{\overline{B}_x\overline{N}_x}VR^T_{\overline{B}_x\overline{N}_x} \;\; & = \;\; I - R_{\overline{B}_x\overline{N}_x}(I + R^T_{\overline{B}_x\overline{N}_x}R_{\overline{B}_x\overline{N}_x})^{-1}R^T_{\overline{B}_x\overline{N}_x} \\ & = \;\; (I + R_{\overline{B}_x\overline{B}_x}R^T_{\overline{B}_x\overline{N}_x})^{-1} = (R_{\overline{B}_xF_x^0}R^T_{\overline{B}_xF_x^0})^{-1}. \end{aligned}$$

The last two equalities above follow from (93) and (62). Therefore, the matrix

$$R^T_{\overline{B}_xF_x^0}(I - R_{\overline{B}_x\overline{N}_x}VR^T_{\overline{B}_x\overline{N}_x})R_{\overline{B}_xF_x^0} = R^T_{\overline{B}_xF_x^0}(R_{\overline{B}_xF_x^0}R^T_{\overline{B}_xF_x^0})^{-1}R_{\overline{B}_xF_x^0} = \Xi_0$$

is the orthogonal projection matrix defined in (65).

Recall from the analysis for the off-diagonal blocks that $\|VR^T_{\overline{B}_x\overline{N}_x}R_{\overline{B}_xF_x^0}\| \leq 1$. Accordingly, the second term on the right-hand-side of (107) has norm $\mathcal{O}(\|X^{-1}_{B_x}\|^2\|X_{\overline{N}_x}\|^2)$. Aggregating all the results for different components of the lower right corner, we arrive at the following conclusion:

$$\Xi_{F_x^0F_x^0} \;\; = \;\; \Xi_0 + \mathcal{O}(\|X^{-1}_{B_x}\|^2\|X_{\overline{N}_x}\|^2). \tag{108}$$

To summarize, we have

$$\Xi \;\; = \;\; \begin{bmatrix} \Xi_+ & \\ & \Xi_0 \end{bmatrix} + \begin{bmatrix} \mathcal{O}(\|X^{-1}_{B_x}\|^2\|X_{\overline{N}_x}\|^2) & \mathcal{O}(\|X^{-1}_{B_x}\|\|X_{\overline{N}_x}\|) \\ \mathcal{O}(\|X^{-1}_{B_x}\|\|X_{\overline{N}_x}\|) & \mathcal{O}(\|X^{-1}_{B_x}\|^2\|X_{\overline{N}_x}\|^2) \end{bmatrix}, \tag{109}$$

concluding the proof of the first portion of the statement of the lemma. (69) is proved identically. □

References

[1] K. M. Anstreicher, "Potential reduction algorithms," in: T. Terlaky, ed., *Interior point methods of mathematical programming*, (Kluwer Academic Publishers, Dordrecht, Netherlands, 1996) pp. 125–158.

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

[3] M. Iri and H. Imai, "A multiplicative barrier function method for linear programming," *Algorithmica*, 1 (1986) 455–82.

[4] M. Iri, "A proof of the polynomiality of the Iri-Imai method," *Journal of Complexity*, 9 (1993) 269–290.

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

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

[7] K. Tanabe, "Centered Newton method for mathematical programming," in: M. Iri and K. Yajima, eds., *Lecture Notes in Control and Information Sciences,* (Springer-Verlag, Berlin, 1988) pp. 197–206.

[8] M. J. Todd, "Potential-reduction methods in mathematical programming," *Mathematical Programming,* 76 (1997) 3–45.

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

[10] T. Tsuchiya, "Global convergence property of the affine scaling methods for primal degenerate linear programming problems," *Mathematics of Operations Research*, 17 (1992) 527–557.

[11] T. Tsuchiya, "Quadratic convergence of the Iri-Imai algorithm for degenerate linear programming problems", *Journal of Optimization Theory and Applications*, 87 (1995) 703-726.

[12] R. H. Tütüncü, "A primal-dual variant of the Iri-Imai Algorithm for Linear Programming", Research Report 98-217, Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA, 1998.

[13] S. Wright, *Primal-dual interior-point methods* (SIAM, Philadelphia, 1997).

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