# Effective Finite Termination Procedures in Interior-Point Methods for Linear Programming

*Pamela Joy Williams*

**CRPC-TR98756-S**
**April 1998**

Center for Research on Parallel Computation
Rice University
6100 South Main Street
CRPC - MS 41
Houston, TX 77005

Submitted August 1998; Also available as Rice CAAM TR 98-17

RICE UNIVERSITY

# Effective Finite Termination Procedures in Interior-Point Methods for Linear Programming

by

**Pamela Joy Williams**

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

**Doctor of Philosophy**

Approved, Thesis Committee:

Richard A. Tapia, Chairman
Noah Harding Professor of Computational and Applied Mathematics

Amr S. El-Bakry, Co-Chairman
Visiting Research Associate, Center for Research on Parallel Computation

David W. Scott
Professor of Statistics

Yin Zhang
Associate Professor of Computational and Applied Mathematics

Houston, Texas

April, 1998

# Abstract

## Effective Finite Termination Procedures in Interior-Point Methods for Linear Programming

by

Pamela Joy Williams

Due to the structure of the solution set, an exact solution to a linear program cannot be computed by an interior-point algorithm without adding features, such as finite termination procedures, to the algorithm. Finite termination procedures attempt to compute an exact solution in a finite number of steps. The addition of a finite termination procedure enables interior-point algorithms to generate highly accurate solutions for problems in which the ill-conditioning of the Jacobian in the neighborhood of the solution currently precludes such accuracy.

The main ingredients of finite termination are activating the procedure, predicting the optimal partition, formulating a simple mathematical model to compute a solution and developing computational techniques to solve the model. Each of these issues are studied in turn in this thesis. First, the current optimal face identification models are extended to bounded variable linear programming problems. Distance to the lower and upper bounds are incorporated into the model to prevent the computed solution from violating the bound constraints. Theory in the literature is extended to the new model. Empirical evidence shows that the proposed model reduces the number of projection attempts needed to find an exact solution. When early termination is the

goal, projection from a pure composite Newton step is advocated. However, the cost may exceed the benefits due to the average need of more than one projection attempt to find an exact solution.

Variants of Mehrotra's predictor-corrector primal-dual interior-point algorithm provide the foundation for most practical interior-point codes. To take advantage of all available algorithmic information, we investigate the behavior of the Tapia predictor-corrector indicator, which incorporates the corrector step, to identify the optimal partition. Globally, the Tapia predictor-corrector indicator behaves poorly as do all indicators, but locally exhibits fast convergence.

# Acknowledgments

This thesis is dedicated to my parents, W. J. and Jardia Williams, who nurtured my love of mathematics from early childhood until today, encouraged me in the bleakest of hours, gave me wings to fly and always welcomed me back with loving arms.

My sister, Alethea, deserves a medal for putting up with me while I weathered the ups and downs of graduate school. I am not the easiest person to live with even under the best circumstances.

The members of my committee gave generously of their time and energy to mold me into a scientist. For that, I thank them.

My senior advisor, Dr. Richard Tapia, provided me with a solid foundation in optimization, especially interior-point methods. He gave mathematics a human face by not only conveying the fundamental concepts but also relaying insights about the person behind the theory or algorithm. I want to thank him for taking me under his wings and for glimpsing my potential as a mathematician.

Thanks to Dr. Amr El-Bakry for his infectious excitement during the course of this research and his willingness to listen to my often muddled explanations of new ideas. His critique of the rough drafts of this thesis and recommended changes improved the readability by a factor of one hundred. I am indebted to Dr. Yin Zhang for his comments and questions about the research. His queries forced me to look at the big picture, justify every statement, and curtail my sloppy tendencies. I thank Dr. David Scott for his patience and understanding.

I thank the AT&T Labs Cooperative Research Fellowship Program for its financial support throughout my graduate career. Special thanks goes to my CRFP mentor, Dr. Patricia Wirth, for sharing her graduate school and professional experiences with me.

It is inspiring to witness her accomplishments first at AT&T Bell Laboratories and now at AT&T Labs. Moreover, thanks to David Houck and the rest of the Teletraffic and Performance Analysis Department for their encouragement throughout the years.

The Rice graduate and undergraduate students who have positively impacted my career are too numerous to list here. However, I want to acknowledge the Spend the Summer with a Scientist program (1992-1997) whose participants have provided much needed moral support over the years.

Cassandra McZeal, Dr. Anthony Kearsley, and Monica Martinez have all contributed in some fashion to my success. Cassandra McZeal has been a confidante, friend, study partner, and a great source of support over the last six years. Without her assistance in a myriad ways I could not have completed this work. Tony Kearsley was and is a wonderful mentor. From the day I came to visit Rice's campus as an undergraduate to the present, he has dispensed advice about classes and life beyond the hedges. (Tony - Because of you, I will always remember Baltimore.) Monica Martinez pushed me to excel in the classroom.

Thanks to Theresa Chatman for the use of Macintosh computer resources whenever needed, leftovers from CRPC meetings, and her friendship.

I want to thank Daria Lawrence and Fran Moshiri for tracking down my stipend on more than one occasion, answering all my questions no matter how inane, and offering help with a smile. Linda Neyra made conference travel and graduate life, in general, less complicated. She is a tremendous asset to the Optimization Group and the department.

Last but not least, I thank my first grade teacher, Miss Arletta Bacon, for not sending me home on the many occasions that I feigned illness. Although I resented your tough love then, I now realize you helped define who I am today.

*To W.J. and Jardia Williams*

# Contents

# Illustrations

# Tables

# Chapter 1

# Introduction

Many aspects of everyday life can be modeled as linear programming problems. Examples include the design and restoration of fiber optic networks in the telecommunications industry, portfolio selections in the financial sector, and diet plans in the burgeoning weight loss industry. In 1947, Dantzig [8] developed the simplex method to solve linear programming problems. To search for an optimal solution, the simplex method selects a sequence of vertices. Although the simplex method possesses the finite termination property, the method has in the worst case an exponential running time complexity.

Interior-point methods emerged as computational tools to solve linear programs in the early 1980s. The introduction of Karmarkar's method [19] propelled research activity in the area of interior-point methods. Karmarkar's projective algorithm is a polynomial time algorithm for solving linear programs that was reportedly significantly faster than the simplex method. Interior-point methods generate iteration sequences that travel through the interior of the feasible region and, under proper assumptions, converge to the solution set. Since the introduction of Karmarkar's algorithm, interior-point algorithms have been developed with strong theoretical properties and excellent numerical performance.

One of the few questions that remains to be resolved for interior-point methods is the development of effective finite termination procedures, techniques that compute an exact solution in a finite number of steps. For linear programming problems, the duality gap is exactly zero. Interior-point algorithms generate an approximate solution in polynomial time; however, because of the structure of the solution set the

algorithms cannot produce an exact solution in polynomial time without the addition of finite termination procedures.

The basic idea for a finite termination procedure is as follows. Once the iterates become close enough to the solution set, the interior-point method can be suspended and the zero-nonzero structure of the solution set can be estimated and used to obtain a solution through some finite procedure whose arithmetic complexity is bounded by a polynomial in the size of the input data. Assuming infinite precision arithmetic, the computed duality gap would be zero. If the output of a finite termination procedure satisfies the feasibility and optimality conditions, then an exact solution has been recovered. Otherwise, the interior-point algorithm resumes.

Research in finite termination can be categorized into two areas, optimal face identification (Tardos [46], Mehrotra [30], Mehrotra and Ye [35], and Ye [52], [54], [55]) and optimal basis identification (Andersen [1], Andersen and Ye [2], Bixby and Saltzman [5], Marsten, Saltzman, Shanno, Pierce and Ballintijn [24], Tapia and Zhang [45], Vavasis and Ye [47], Ye and Todd [50], and Ye [51]). Optimal face identification techniques identify the face upon which the objective function attains its optimal value. The optimal face is uniquely defined by the set of variables which are zero at the solution. Once the zero variables have been identified, the solution to the linear program can be obtained by calculating an interior feasible point on the face. Optimal basis identification methods construct an optimal basis from the zero-nonzero structure of the solution set.

In 1989, Gay [13] proposed stopping tests that computed optimal solutions for interior-point methods for linear programming problems. While these tests do not constitute a finite termination procedure, they are clearly predecessors of current optimal face identification techniques. Gay's idea was to use the zero-nonzero partition of the variables to find the solution of linear programs. It was his belief that the

output of the stopping tests could be used as stopping criteria for the algorithm. Furthermore, he thought that the early stopping tests would decrease numerical difficulties associated with Jacobians which were necessarily singular on the solution set.

Gay solved two linear feasibility problems to obtain interior points on the optimal primal and dual faces. The linear feasibility problems were defined to take advantage of the Cholesky factorization which was already a part of the underlying interior-point algorithm. Instead of factoring the coefficient matrix of a scaled linear system, Gay formed the associated normal equations and then used a Cholesky factorization to decompose the normal equations matrix. However, Gay used an iterative, not direct method to find a feasible point. Hence, his technique cannot be categorized as a finite termination procedure.

Gay's influence can be seen in Mehrotra and Ye [35] where the authors solve a linear feasibility problem via Gaussian elimination. In Chapter 5, we propose linear feasibility problems similar to the ones that appeared in Gay [13].

Adding finite termination procedures to the interior-point framework would lead to definitive stopping criteria, computational savings, and highly accurate solutions. For degenerate problems, the Jacobian is necessarily singular at a solution. Therefore, we expect that the Jacobian will be ill-conditioned close the solution set. The ill-conditioned Jacobian may produce step directions which prevent the problem from being solved to a high accuracy. With a finite termination procedure, we can avoid, to some degree, the effects of ill-conditioning.

In this work, we extend the current optimal face identification models to linear programs with upper bound constraints. We also propose new criteria for when to first attempt the calculation of an exact solution.

Identification of the active set is an important component of a finite termination procedure. The active set corresponds to variables that are zero at the solution. A commonly used indicator is the Tapia indicator (see Tapia [43]), which consists of the step direction and the current iterate. A natural extension of the Tapia indicator is obtained by using the predictor and centering-corrector directions generated by predictor-corrector interior-point methods. We investigate this new indicator that takes advantage of the supplemental information provided by the centering-corrector direction.

The thesis is organized as follows. Section 1.1 gives essential background material on the linear programming problem. In Section 1.2, we describe a generic finite termination procedure. Chapter 2 provides an overview of finite termination procedures that incorporate optimal face identification. In Chapter 3, theoretical results regarding the addition of a weighted projection model to infeasible primal-dual interior-point algorithms are presented. We also determine the optimal choice for a weighting matrix. Numerical comparisons are presented to show how a weighting matrix affects the performance of a finite termination procedure. Variants of the Mehrotra-Ye procedure are the subject of Chapter 4. In Chapter 5, fast local convergence of the duality gap is used to determine when to attempt to compute an exact solution. In Chapter 6, we extend the Tapia indicator to include corrector information from the Mehrotra primal-dual predictor-corrector algorithm. We give concluding remarks in Chapter 7.

## 1.1 Background and Notation

We consider linear programs in the standard form:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \geq 0, \end{array} \tag{1.1}$$

where $c, x \in \mathbf{R}^n$, $b \in \mathbf{R}^m$, $A \in \mathbf{R}^{m \times n}$ $(m \leq n)$ and $A$ has full rank $m$.

The optimality conditions for (1.1) are

$$F(x,y,z) = \begin{pmatrix} Ax - b \\ A^T y + z - c \\ XZe \end{pmatrix} = 0, \quad (x,z) \geq 0, \tag{1.2}$$

where $y \in \mathbf{R}^m$ are the Lagrange multipliers corresponding to the equality constraints, $z \in \mathbf{R}^n$ are the Lagrange multipliers corresponding to the inequality constraints, $X = \mathrm{diag}(x)$, $Z = \mathrm{diag}(z)$ and $e$ is the $n$-vector of all ones.

The Jacobian of (1.2) is

$$F'(x,y,z) = \begin{pmatrix} A & 0 & 0 \\ 0 & A^T & I \\ Z & 0 & X \end{pmatrix}. \tag{1.3}$$

The primal-dual feasibility set is defined as

$$\mathcal{F} = \{(x,y,z) : Ax = b, \quad A^T y + z = c, \quad (x,z) \geq 0\}.$$

The strict feasibility set of the primal and dual is

$$\mathcal{F}^0 = \{(x,y,z) : Ax = b, \quad A^T y + z = c, \quad (x,z) > 0\}.$$

We denote the solution set of (1.2) as

$$\mathcal{S} = \{(x,y,z) : F(x,y,z) = 0, \quad (x,z) \geq 0\}.$$

If a solution satisfies

$$x + z > 0,$$

in addition to $XZe = 0$, then this solution is said to satisfy the strict complementarity condition or strict complementarity.

Given feasible iterates, we see that $\|F(x,y,z)\|_1 = x^T z$. It can be shown that the expression $x^T z$ is equal to the duality gap, which is the difference between the primal and dual objective function values. At optimality, $x^T z = 0$.

The strict complementarity condition is not restrictive. For linear programming problems, Goldman and Tucker [14], proved that among all optimal solutions there exists at least one solution that satisfies strict complementarity. Thus for nondegenerate problems, the unique solution satisfies strict complementarity.

If $\mathcal{S} \neq \emptyset$, then the relative interior of $\mathcal{S}$, $ri(\mathcal{S})$, is nonempty. In this case, the solution set $\mathcal{S}$ has the following structure (see El-Bakry, Tapia, and Zhang [10] for a proof): (i) all points in the relative interior satisfy strict complementarity; (ii) the zero-nonzero pattern of points in the relative interior is invariant. Therefore, for any $(x^*, y^*, z^*) \in ri(\mathcal{S})$,

$$\mathcal{B} = \{j : x_j^* > 0,\ 1 \le j \le n\} \ \text{ and } \ \mathcal{N} = \{j : z_j^* > 0,\ 1 \le j \le n\}.$$

For more details, see Güler and Ye [17] and McLinden [25]. Moreover,

$$\mathcal{B} \bigcup \mathcal{N} = \{1, \ldots . n\} \ \text{ and } \ \mathcal{B} \bigcap \mathcal{N} = \emptyset.$$

Thus, the sets $\mathcal{B}$ and $\mathcal{N}$ define the optimal partition of the set $\{1, \ldots, n\}$.

The optimal primal face is defined as

$$\Theta_p = \{x : Ax = b,\ x \ge 0,\ x_j = 0\ j \in \mathcal{N}\}.$$

Similarly, the optimal dual face is

$$\Theta_d = \{(y,z) : A^T y + z = c,\ z \ge 0,\ z_j = 0\ j \in \mathcal{B}\}.$$

In the following chapters, the columns of $A$ corresponding to the indices of $\mathcal{B}$ comprise the matrix $B$. The matrix $N$ is formed in an analogous manner. The

components of the vector $x$ whose indices are in $\mathcal{B}$ are denoted by $x_{\mathcal{B}}$. Unless otherwise denoted, $\|\cdot\|$ is the Euclidean norm. The cardinality of set $\mathcal{B}$ is denoted by $|\mathcal{B}|$. We use the notation

$$\min u = \min_{1\leq i\leq n} u_i \text{ for } u \in \mathbf{R}^n.$$

The central path parameterized by $\mu$, see Megiddo [26], is defined as

$$\mathcal{C} = \{(x, y, z) \in \mathcal{F}^0 : XZe = \mu e \;\; \text{where } \mu = x^T z/n\}. \tag{1.4}$$

We define a neighborhood of the central path as

$$\mathcal{N}_{-\infty}(\gamma) = \{(x, z)| \;\; \min(XZe) \geq \gamma\mu\} \tag{1.5}$$

where $\mu = (x^T z)/n,\ \gamma \in (0, 1)$. Restricting the iterates to $\mathcal{N}_{-\infty}(\gamma)$ prevents them from prematurely getting too close to the boundary of the nonnegative orthant.

## 1.2 Finite Termination

Consider a nondegenerate problem (i.e., an unique solution exists). Within the interior-point framework, assume that the active set is known. Then for the standard linear programming problem, $n$-$m$ variables that are zero at the solution and $m$ variables that are positive at the solution have been identified. After setting the variables in the active set to zero, the reduced square system can be solved for the remaining variables. If optimality and feasibility conditions are satisfied for both the primal and dual problems, the solution has been identified and the interior-point algorithm terminates.

However, in practice most problems are degenerate, which results in a rectangular coefficient matrix of the reduced system. If the linear system is underdetermined, there exists infinitely many solutions. Consequently, there is no straightforward way

to compute a nonnegative solution of the reduced system. Several models have been proposed to recover a nonnegative solution.

We now describe a generic finite termination procedure.

**Procedure 1** *(A Finite Termination Procedure)*

**(1)** At some iteration $k$, suspend the interior-point algorithm and estimate $(\mathcal{B}, \mathcal{N})$.

**(2)** Set $x_{\mathcal{N}} = 0$ and $z_{\mathcal{B}} = 0$.

**(3)** Solve a mathematical model for the vectors $(x_{\mathcal{B}}, y, z_{\mathcal{N}})$.

**(4)** If $x_{\mathcal{B}} \geq 0$, $z_{\mathcal{N}} \geq 0$, and we satisfy the primal and dual constraints, stop. Since

$$x^T z = x_{\mathcal{B}}^T z_{\mathcal{B}} + x_{\mathcal{N}}^T z_{\mathcal{N}} = 0,$$

we have found an exact solution.

Else, return to the interior-point algorithm.

Below, we graphically show steps 1 and 3 of a finite termination procedure.

From the outline above, we see that four important questions must be addressed in a finite termination procedure. When do we attempt to compute an exact solution (Chapter 5)? How do we identify the optimal partition of a linear program (Chapter 6)? In the presence of degeneracy, what mathematical model should be used to attempt to find a solution (Chapters 3, 4)? Given the model, what is the most computationally efficient way to solve the model (Chapters 3, 4)? Our research focuses on these issues.

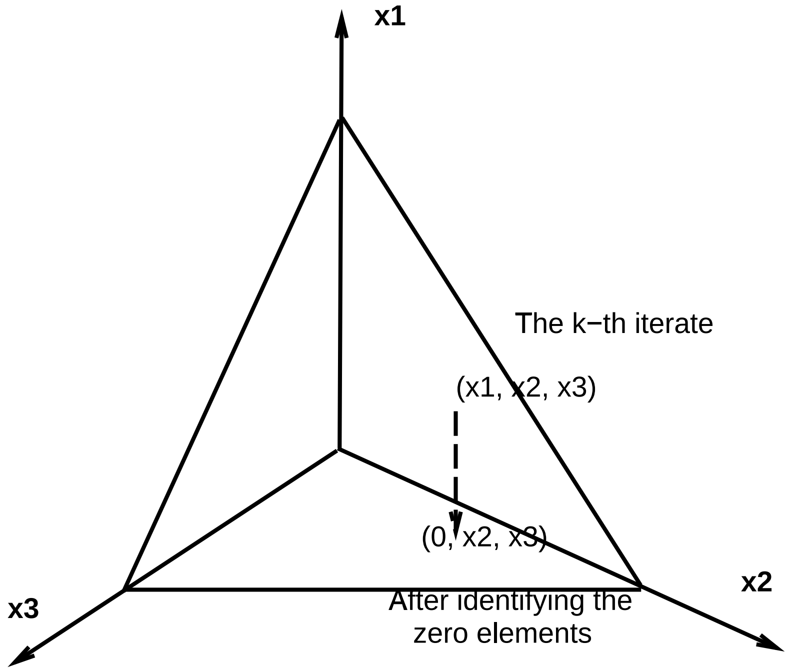


**Figure 1.1** Step 1 - Identifying the optimal partition of a linear program: $\mathcal{B} = \{2, 3\}$ and $\mathcal{N} = \{1\}$

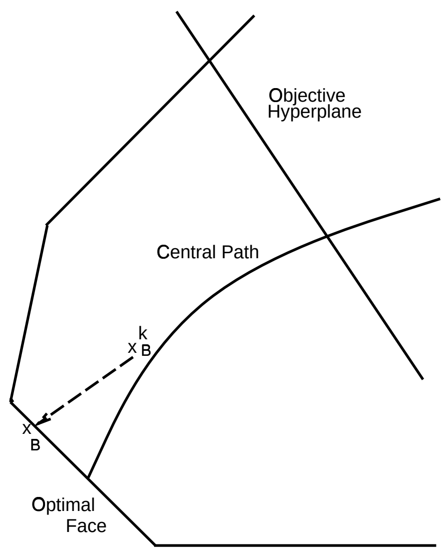


**Figure 1.2** Step 3 - Projection of $x_{\mathcal{B}}^k$ onto the optimal primal face

# Chapter 2

# Contributions from the Literature

Much work has been done in the area of finite termination for feasible interior-point methods for linear programming, see Mehrotra [30], [31] and Ye [52], [54], [55]. Finite termination procedures in infeasible interior-point methods for linear programming have been studied by Potra [38] and Anstreicher, Ji, Potra, and Ye in [3], where a probabilistic analysis was given. Monteiro and Wright [36], as well as Ji and Potra [18] investigated finite termination procedures in infeasible interior-point algorithms for degenerate monotone linear complementarity problems (LCPs). Resende and Veiga [39] identified the optimal dual face and generated a primal basic solution to derive robust stopping criteria for minimum cost network flow problems. The authors also introduced an optimal stopping criteria that required solving a maximum flow problem to determine a primal solution. Subsequent research into identifying the optimal dual face for network flow problems appeared in Portugal, Resende, Veiga, and Judice [37] and Resende, Tsuchiya, and Veiga [40]. Ye [55] proved that given a homogeneous self-dual linear program a finite termination procedure can generate an optimal solution to the standard linear program or detect infeasibility in finite time. He also showed that the finite termination procedure can find a feasible point for the homogeneous linear system.

In the following three sections of this chapter, we give a detailed description of existing models and techniques used to find an interior feasible point on the optimal primal and dual faces. The models and techniques differ in the incorporation of the inequalities and the method by which an interior feasible point is computed.

Recall that the optimal primal face can be written as

$$\Theta_p = \{x : Bx_{\mathcal{B}} = b, \;\; x_{\mathcal{B}} \geq 0, \;\; x_{\mathcal{N}} = 0\},$$

and the optimal dual face as

$$\Theta_d = \{(y, z) : B^T y = c_{\mathcal{B}}, \;\; c_{\mathcal{N}} - N^T y = z_{\mathcal{N}}, \;\; z_{\mathcal{N}} \geq 0\}.$$

Finite Termination Problem: Given that the partition $(\mathcal{B}, \mathcal{N})$ is correct, the vector $u \in \mathbf{R}^{|\mathcal{B}|}$ is close to the linear manifold

$$Bx_{\mathcal{B}} = b,$$

and $v \in \mathbf{R}^{n-|\mathcal{B}|}$ is close to the linear manifold

$$A^T y + z = c,$$

the actual problem is

$$\begin{aligned} &\text{Find } \Delta u \;\; \text{s.t.} \;\; u + \Delta u \in \Theta_p \\ &\text{and } (\Delta v, \Delta w) \;\; \text{s.t.} \;\; (v + \Delta v, w + \Delta w) \in \Theta_d. \end{aligned} \tag{2.1}$$

A solution of the primal subproblem can be obtained by solving

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{2}\|x_{\mathcal{B}} - u\|^2 \\ &\text{subject to} \quad Bx_{\mathcal{B}} = b \qquad x_{\mathcal{B}} \geq 0. \end{aligned} \tag{2.2}$$

Similarly, a solution of the dual subproblem can be obtained by solving

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{2}\|y - v\|^2 \\ &\text{subject to} \quad A^T y + z = c \quad z_{\mathcal{N}} = c_{\mathcal{N}} - N^T y \geq 0. \end{aligned} \tag{2.3}$$

There exist many iterative methods that can solve (2.1) or (2.2) and (2.3). However, finite termination procedures require the solution of the problem by a direct, not indirect, method. Kincaid and Cheney [20] define direct methods as methods that proceed

through a finite number of steps and produce a solution that would be completely accurate if not for roundoff errors. In general, the presence of nonnegativity constraints render direct methods ineffective for solving problems (2.1) or (2.2). However, if the vectors $u$ and $v$ are close enough to the solution set, we can solve approximation models of problems (2.1) and (2.2) by a direct method.

## 2.1 Orthogonal Projection

Ye [52] was probably the first to study finite termination in interior-point methods for linear programming. He was motivated by the fact that the simplex method for linear programming has the finite termination property and also by research activity in efficient algorithmic termination techniques.

Ye computed a point in the interior of the optimal primal face by solving the following least squares problem,

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|x_{\mathcal{B}} - x_{\mathcal{B}}^k\|^2 \\ \text{subject to} & Bx_{\mathcal{B}} = b. \end{array} \tag{2.4}$$

The dual least squares problem can be written as

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|y - y^k\|^2 \\ \text{subject to} & B^T y = c_{\mathcal{B}}. \end{array} \tag{2.5}$$

No rank assumptions were made on the matrix $B$.

The solutions of (2.4) and (2.5) are the orthogonal projections (i.e, the solutions are the closest points to $x_{\mathcal{B}}^k$ and $y^k$ on the respective linear manifolds). The advantages of Ye's orthogonal projection model are twofold. First, the solutions $x_{\mathcal{B}}$ and $y$ are unique. Second, the cost of solving both the primal and dual formulations is equivalent to the cost of one interior-point iteration.

However, the model has some drawbacks. In particular, Ye's model does not include the nonnegativity constraints. If $x_{\mathcal{B}}^k$ is close to the boundary, the orthogonal

projection can produce points outside the positive feasible region. As a result, the subsequent $x_{\mathcal{B}}$ is rejected and control is returned to the interior-point algorithm.

Ye [52] proved, under certain conditions, that a finite termination procedure added to a feasible interior-point algorithm yields an exact solution in polynomial time. Analogous results were proven by Anstreicher, Ji, Potra, and Ye [3] and Potra [38] for an infeasible primal-dual interior-point method for linear programming and by Monteiro and Wright [36] and Ji and Potra [18] for degenerate monotone linear complementarity problems.

Portugal, Resende, Veiga, and Judice [37], Resende and Veiga [39], and Resende, Tsuchiya, and Veiga [40] used the orthogonal projection model to find dual solutions for minimum cost network flow problems. The block triangular structure of the constraint matrix led to computing the solution by orthogonally projecting onto each individual subspace. This resulted in a significant reduction in the cost of the orthogonal projection model. A point on the optimal dual face can be computed in $O(p)$ operations as opposed to $O(p^3)$, where $p$ is the number of columns in the constraint matrix.

## 2.2 Mehrotra-Ye Procedure

In [46], Tardos proposed the use of Gaussian elimination to calculate a feasible point on the optimal face of an integer linear program. Mehrotra and Ye [35] studied the effectiveness of this factorization in computing an interior point on the optimal face of a linear program. Instead of solving a least squares problem, the authors used Gaussian elimination to find basic solutions to

$$B\Delta x_{\mathcal{B}} = b - Bx_{\mathcal{B}}^k \quad \text{and} \quad B^T\Delta y = c_{\mathcal{B}} - B^T y^k.$$

Linearly dependent rows and columns were dropped as they were encountered during the elimination process. Components of the solution vectors corresponding to linearly dependent rows/columns were set equal to zero. The nonnegativity constraints on $x_{\mathcal{B}}$ and $z_{\mathcal{N}}$ are ignored.

Whereas Ye's formulation produces an unique solution, the Mehrotra-Ye model potentially has an infinite number of solutions. One would think that ignoring the inequalities would be disastrous. On the contrary, the numerical results in [35] are quite impressive. For 74 out of 86 *netlib* problems, the authors were able to find an interior point on the optimal face in one attempt when the procedure is activated when the relative gap is less than or equal to $10^{-8}$.

In the presence of degeneracy, the Gaussian elimination approach has two major deficiencies. First, the approach does not incorporate information about the inequalities. All components of the current iterate are treated equally. There is no penalty for radical movement of the small components of the current iterate. Second, success of the procedure is dependent on the choice of the basis matrix, i.e., a nonsingular submatrix of $B$. Different basis matrices produce dissimilar solutions, some of which may lie outside the positive orthant. However, the procedure is computationally inexpensive. The cost for the Mehrotra-Ye procedure is one matrix factorization and four back substitutions for the primal and dual solutions.

## 2.3 Weighted Projection

The weighted projection model incorporates information about the inequalities into the objective function by weighting the distance between the current iterate and the solution of the least squares problem. The model restricts movement in components that can least afford to deviate from $x^k_{\mathcal{B}}$ by placing large weights on the smaller components. Large movements in the components of small magnitude are penalized

more heavily than movements in components of larger magnitude. The primal model, proposed by Ye [55], is

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{2}\|(X_{\mathcal{B}}^k)^{-1}(x_{\mathcal{B}} - x_{\mathcal{B}}^k)\|^2 \\ &\text{subject to} \quad Bx_{\mathcal{B}} = b. \end{aligned} \tag{2.6}$$

Ye defined the dual model as

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{2}\|(Z_{\mathcal{N}}^k)^{-1}N^T(y - y^k)\|^2 \\ &\text{subject to} \quad B^T y = c_{\mathcal{B}}. \end{aligned} \tag{2.7}$$

Since $N^T y + z_{\mathcal{N}} = c_{\mathcal{N}}$, problem (2.7) is equivalent to

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{2}\|(Z_{\mathcal{N}}^k)^{-1}(z_{\mathcal{N}} - z_{\mathcal{N}}^k - r_{d_{\mathcal{N}}}^k)\|^2 \\ &\text{subject to} \quad B^T y = c_{\mathcal{B}} \\ &\qquad\qquad\quad N^T y + z_{\mathcal{N}} = c_{\mathcal{N}}, \end{aligned} \tag{2.8}$$

where $r_{d_{\mathcal{N}}}^k = c_{\mathcal{N}} - N^T y^k - z_{\mathcal{N}}^k$. Clearly $r_{d_{\mathcal{N}}}^k = 0$ in feasible interior-point algorithms. Ye [55] proved that the solutions of problems (2.6) and (2.7) are interior points on the optimal primal and dual faces. Moreover, the solutions can be obtained in finite time when included in a feasible primal-dual interior-point algorithm.

We can rewrite (2.6) as

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{2}\|dx_{\mathcal{B}}\|^2 \\ &\text{subject to} \quad (BX_{\mathcal{B}}^k)dx_{\mathcal{B}} = b - Bx_{\mathcal{B}}^k, \end{aligned} \tag{2.9}$$

where $dx_{\mathcal{B}} = (X_{\mathcal{B}}^k)^{-1}(x_{\mathcal{B}} - x_{\mathcal{B}}^k)$.

Similarly, the dual can be rewritten as

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{2}\|(Z_{\mathcal{N}}^k)^{-1}N^T\Delta y\|^2 \\ &\text{subject to} \quad B^T\Delta y = c_{\mathcal{B}} - B^T y^k, \end{aligned} \tag{2.10}$$

where $\Delta y = y - y^k$.

Note that the primal constraint matrix depends on the primal weighting matrix, but the dual constraint matrix does not depend on the dual weighting matrix. Thus, two matrix factorizations are required to find feasible points on the optimal primal and dual faces, making this model twice as expensive as the previous two.

Samuelsen [41] referred to the choice of weight in problem (2.6) as the Dikin-Karmarkar weight. In [41], the author showed that the Dikin-Karmarkar norm, unlike the Euclidean norm, does not suffer from the short step syndrome. Freund [12] measured proximity to the feasible region by formulations similar to (2.6) and (2.7). However, he did not drop columns of the matrix $A$ and components of the vectors $x$ and $z$ corresponding to zeros on the solution set.

In the next chapter, we extend the existing theory to infeasible primal-dual interior-point algorithms. Furthermore, we provide numerical experimentation to compare the effectiveness of the weighted projection model with the orthogonal projection model.

# Chapter 3

# Weighted Projections

In this chapter, we prove that the weighted projection model implemented in an infeasible interior-point method produces, in finite time, a solution that satisfies the strict complementarity condition. First, we provide an algorithmic framework for the problem and show that the optimal partition can be identified when the residual $\|F(x,y,z)\|_1$ is sufficiently small. Then, we present a technical lemma which will be relevant in our analysis of the finite termination procedure, give an optimal choice for weighting matrix, extend the weighted projection model to incorporate upper bound information, establish an arithmetic complexity result for the given algorithm, and describe our numerical experiments.

## 3.1 Algorithmic Framework

Kojima, Mizuno, and Yoshise [21] proposed the first feasible primal-dual interior-point method for linear programming. It is well-known that their method can be viewed as perturbed and damped Newton's method on the first order optimality conditions. In this section, we describe a generic infeasible primal-dual interior-point method, which is also based on Newton's method.

**Algorithm 1** *(Infeasible Primal-Dual Algorithm)*

Given $(x^0, y,^0, z^0)$ with $(x^0, z^0) > 0$, for $k = 0, 1, 2, \ldots$, do

**(1)** Choose $\sigma^k \in (0,1)$ and set $\mu^k = ((x^k)^T z^k)/n$.

**(2)** Solve for the step $(\Delta x^k, \Delta y^k, \Delta z^k)$

$$F'(x^k, y^k, z^k) \begin{pmatrix} \Delta x^k \\ \Delta y^k \\ \Delta z^k \end{pmatrix} = \begin{pmatrix} b - Ax^k \\ c - A^T y^k - z^k \\ \sigma^k \mu^k e - X^k Z^k e \end{pmatrix}.$$

**(3)** Choose $\tau^k \in (0,1)$ and set $\alpha^k = min(1, \tau^k \hat{\alpha}^k)$, where

$$\hat{\alpha}^k = \frac{-1}{\min((X^k)^{-1}\Delta x^k, (Z^k)^{-1}\Delta z^k)}.$$

**(4)** Let $(x^{k+1}, y^{k+1}, z^{k+1}) := (x^k, y^k, z^k) + \alpha^k(\Delta x^k, \Delta y^k, \Delta z^k)$.

**(5)** Test for convergence.

The optimality conditions (1.2) are perturbed to keep the iterates in the interior of the nonnegative orthant. If $\sigma^k = 0$ (i.e., no perturbation), global convergence may be precluded. See Proposition 3.1 of Gonzalez-Lima [22] for a proof. The iteration sequence is damped to maintain the nonnegativity requirement.

The algorithm is considered an infeasible interior-point method since we do not require feasibility with respect to the linear constraints. The terminology, infeasible interior-point methods, was first introduced by Zhang [56].

The following lemma provides a theoretical basis for finite termination procedures.

**Lemma 3.1** (Güler and Ye [17]) Let $\{(x^k, y^k, z^k)\}$ be an iteration sequence generated by an interior-point algorithm that converges to the solution set of a linear program. Furthermore, let $x^k$ and $z^k$ satisfy

$$\frac{min(X^k Z^k e)}{(x^k)^T z^k / n} \geq \gamma \tag{3.1}$$

where $\gamma > 0$ and is independent of $k$. Then every limit point of $\{(x^k, z^k)\}$ satisfies the strict complementarity condition.

Güler-Ye [17] showed that iterates of feasible path-following algorithms (e.g., Kojima, Mizuno, and Yoshise [21], Mizuno, Todd, and Ye [34]) satisfy inequality (3.1), which is one of many centrality measures used in linear programming.

All points in the relative interior of the solution set satisfy the strict complementarity condition. Therefore, Lemma 3.1 is sufficient to guarantee that all limit points of the iteration sequence are in the relative interior of the solution set, see Güler and Ye [17]. It is well-known that in the relative interior the nonzero-zero pattern of points is invariant, see El-Bakry, Tapia, and Zhang [10]. Consequently, the optimal primal and dual faces are uniquely defined.

We now present some definitions that are essential to the development of the theory for infeasible interior-point methods. First, we define a central path neighborhood that includes infeasible points as

$$\begin{aligned} \mathcal{N}_{-\infty}(\gamma,\beta) = \{(x,y,z) | \quad & \|(r_p, r_d)\| \le [\|(r_p^0, r_d^0)\|/\mu^0]\beta\mu, \\ & (x,z) > 0, \quad \min(XZe) \ge \gamma\mu\} \end{aligned} \tag{3.2}$$

where $\gamma \in (0,1)$, $\beta \ge 1$, $\mu = x^T z/n$, $r_p = b - Ax$, and $r_d = c - A^T y - z$. The first inequality of (3.2) is known as the Feasibility Priority Principle, see Zhang [56]. It requires that infeasibility decreases at least as fast as complementarity.

We also define the following quantities:

$$\begin{aligned} \|r^k\| &= \|(r_p^k, r_d^k)\| \\ \psi_k &= (1-\alpha^{k-1})\psi_{k-1} = \textstyle\prod_{i=0}^{k-1}(1-\alpha^i) \\ \Theta_p &= \{x : Ax = b,\ x \ge 0,\ x_j = 0 \text{ for } j \in \mathcal{N}\} \\ \Theta_d &= \{(y,z) : A^T y + z = c,\ z \ge 0,\ z_j = 0 \ \text{ for } j \in \mathcal{B}\} \end{aligned} \tag{3.3}$$

If $\mathcal{F}^0 \neq \emptyset$, then $\Theta_p$ and $\Theta_d$ are nonempty and bounded. Hence, $\xi_p^*$ and $\xi_d^*$ are bounded, where

$$\xi_p^* = \min_{j\in\mathcal{B}} \{\max x_j, \text{ s.t. } x \in \Theta_p\},$$

$$\xi_d^* = \min_{j \in \mathcal{N}} \{\max z_j, \text{ s.t. } (y, z) \in \Theta_d\}, \text{ and}$$

$$\xi^* = \min(\xi_p^*, \xi_d^*). \tag{3.4}$$

We assume that the iteration sequence generated by Algorithm 1 stays in $\mathcal{N}_{-\infty}(\gamma, \beta)$. This condition is necessary for the proof of global convergence of the algorithm.

## 3.2 Optimal Face Identification Theory

The following theorem shows that the optimal partition can be identified when the residual $\|F(x, y, z)\|_1$ falls below some threshold value and also provides bounds on the iteration sequence $\{(x^k, z^k)\}$. The proof follows the outline of Anstreicher et al. (Lemma 3.2 [3]), Mizuno [33], Monterio and Wright (Lemma 2.1 [36]), Potra (Lemma 3.2 [38]), and Wright (Lemma 6.3 [49]).

Assuming all algorithmic parameters are the same, the upper bound imposed on the duality gap in Theorem 3.1 is more stringent than the bound which appears in Proposition 5.2 of Potra [38]. For our defined value of $\tau$, a smaller duality gap is required before the optimal partition can be identified. The constant $\tau$ depends on the value of $\beta$ given in (3.2). For his analysis, Potra sets $\beta$ equal to one.

**Theorem 3.1** Let $(x^*, y^*, z^*) \in \mathcal{S}$, and $\mathcal{B}^k = \{x_j^k > z_j^k\}$. Then $\mathcal{B}^k = \mathcal{B}$, when

$$(x^k)^T z^k < \gamma(\xi^*)^2/\tau^2 n,$$

for some $\tau > 0$.

**Proof** We have

$$Ax^k - b = (1 - \alpha^{k-1})(Ax^{k-1} - b) = \psi_k(Ax^0 - b) \tag{3.5}$$

by definition of the algorithm and (3.3). Similarly,

$$A^T y^k + z^k - c = \psi_k (A^T y^0 + z^0 - c). \tag{3.6}$$

Thus, by (3.5) and (3.6),

$$\frac{A(x^k - \psi_k x^0)}{(1-\psi_k)} = b \text{ and } \frac{A^T(y^k - \psi_k y^0) + (z^k - \psi_k z^0)}{(1-\psi_k)} = c.$$

Then $Ax^* = b$ and $A^T y^* + z^* = c$ imply $A\bar{x} = 0$ and $A^T \bar{y} + \bar{z} = 0$ where

$$\bar{x} = \psi_k x^0 + (1-\psi_k)x^* - x^k \text{ and } \bar{z} = \psi_k z^0 + (1-\psi_k)z^* - z^k.$$

Due to the orthogonality of $\bar{x}$ and $\bar{z}$,

$$\bar{x}^T \bar{z} = (\psi_k x^0 + (1-\psi_k)x^* - x^k)^T (\psi_k z^0 + (1-\psi_k)z^* - z^k) = 0$$

which can be written as

$$\begin{aligned} \bar{x}^T \bar{z} = \psi_k^2 ({x^0}^T z^0) + (1-\psi_k)^2 {x^*}^T z^* + \psi_k (1-\psi_k)({x^0}^T z^* + {z^0}^T x^*) + \\ (x^k)^T z^k - \psi_k ({x^0}^T z^k + {z^0}^T x^k) - (1-\psi_k)((x^k)^T z^* + {z^k}^T x^*). \end{aligned} \tag{3.7}$$

Hence the nonnegativity of the iteration sequence implies

$$\begin{aligned} \bar{x}^T \bar{z} \le \psi_k^2 ({x^0}^T z^0) + \psi_k (1-\psi_k)({x^0}^T z^* + {z^0}^T x^*) + \\ (x^k)^T z^k - (1-\psi_k)((x^k)^T z^* + {z^k}^T x^*). \end{aligned} \tag{3.8}$$

Therefore, we have

$$\begin{aligned} (1-\psi_k)((x^k)^T z^* + {z^k}^T x^*) &\le \psi_k^2 ({x^0}^T z^0) + \psi_k (1-\psi_k)({x^0}^T z^* + {z^0}^T x^*) + (x^k)^T z^k \\ &= \psi_k^2 (n\mu^0) + \psi_k (1-\psi_k)({x^0}^T z^* + {z^0}^T x^*) + n\mu^k, \end{aligned} \tag{3.9}$$

from the definition of $\mu^k$.

We now divide both sides of the inequality by $(1-\psi_k)$ to obtain,

$$
\begin{aligned}
(x^k)^T z^* + {z^k}^T x^* &\leq \tfrac{\psi_k^2 n\mu^0}{1-\psi_k} + \tfrac{n\mu^k}{1-\psi_k} + \psi_k({x^0}^T z^* + {z^0}^T x^*) \\
&\leq \tfrac{\psi_k(\mu^0)^{-1}(\beta\mu^k)n\mu^0 + n\mu^k}{1-\psi_k} + \psi_k({x^0}^T z^* + {z^0}^T x^*) \\
&= \tfrac{(\psi_k\beta+1)(x^k)^T z^k}{1-\psi_k} + \psi_k({x^0}^T z^* + {z^0}^T x^*) \\
&\leq \tfrac{(\psi_k\beta+1)(x^k)^T z^k}{1-\psi_k} + \beta((x^k)^T z^k / {x^0}^T z^0)({x^0}^T z^* + {z^0}^T x^*) \\
&\leq \tfrac{(1+\beta(1-\alpha^0))(x^k)^T z^k}{\alpha^0} + \beta((x^k)^T z^k / {x^0}^T z^0)({x^0}^T z^* + {z^0}^T x^*) \\
&= \tfrac{\beta(1/\beta+(1-\alpha^0))(x^k)^T z^k}{\alpha^0} + \beta((x^k)^T z^k / {x^0}^T z^0)({x^0}^T z^* + {z^0}^T x^*) \\
&\leq \beta\left[\tfrac{2-\alpha^0}{\alpha^0} + \tfrac{{x^0}^T z^* + {z^0}^T x^*}{{x^0}^T z^0}\right](x^k)^T z^k \text{ since } \beta \geq 1.
\end{aligned} \tag{3.10}
$$

Let

$$
\tau = \beta\left[\frac{2-\alpha^0}{\alpha^0} + \frac{{x^0}^T z^* + {z^0}^T x^*}{{x^0}^T z^0}\right]. \tag{3.11}
$$

Thus, we have $z_j^k x_j^* \leq {z^k}^T x^* \leq \tau((x^k)^T z^k)$. The value of $\tau$ is identical to what appears in Potra [38], with the exception that here the constant $\beta$ is greater than or equal to one. Observe that,

$$
\frac{x_j^*}{x_j^k}\gamma((x^k)^T z^k)/n \leq \frac{x_j^*}{x_j^k}(x_j^k z_j^k)
$$

which implies that

$$
\frac{x_j^*}{x_j^k}\gamma((x^k)^T z^k)/n \leq \tau((x^k)^T z^k) \text{ or equivalently } x_j^k \geq \gamma x_j^*/\tau n.
$$

From the definition of $\xi^*$,

$$
x_j^k \geq (\gamma\xi^*)/\tau n, \text{ for all } j \in \mathcal{B}
$$

and

$$
z_j^k \leq (\tau(x^k)^T z^k)/\xi^*, \text{ for all } j \in \mathcal{B}.
$$

Similarly,

$$
z_j^k \geq (\gamma\xi^*)/\tau n \text{ and } x_j^k \leq (\tau(x^k)^T z^k)/\xi^*, \text{ for all } j \in \mathcal{N}.
$$

Given $\mathcal{B}^k = \{x_j^k > z_j^k\}$, the partition is optimal when

$$x_j^k \geq (\gamma\xi^*)/\tau n > (\tau(x^k)^T z^k)/\xi^* \geq z_j^k.$$

This concludes the proof. $\square$

Notice that the separation of variables depends on the size of the $(-\infty)$ neighborhood and the initial iterate. A large neighborhood (i.e, $\gamma$ close to zero) could seriously impact the effectiveness of any indicator in identifying the optimal partition in a timely fashion. Moreover, the larger $\xi^*$, the fewer interior-point iterations needed to identify the optimal partition. Conversely, as $\xi^*$ decreases, the number of interior-point iterations required to determine the optimal partition increases. An important fact to remember is that $\xi^*$ depends on the problem data.

### 3.2.1 Technical Result

Subsequent theory requires bounds on the primal and dual residuals. Rather than repeatedly deriving the bounds, we establish them at this juncture. Note that the bound depends on the current duality gap, the initial duality gap, and the minimum positive value in the solution set, which is not known a priori.

**Lemma 3.2** Let $\{(x^k, y^k, z^k)\}$ be the iteration sequence generated by Algorithm 1. Further, assume that $\mathcal{B}^k = \mathcal{B}$. Then,

$$\begin{aligned} \|b - Bx_{\mathcal{B}}^k\| &\leq (n\mu^0\xi^*)^{-1} \left(\xi^*\beta\|r^0\| + \mu^0\tau n\sqrt{n}\|N\|\right)(x^k)^T z^k \\ \|c_{\mathcal{B}} - B^T y^k\| &\leq (n\mu^0\xi^*)^{-1}(\xi^*\beta\|r^0\| + \mu^0\tau n\sqrt{n})(x^k)^T z^k. \end{aligned} \tag{3.12}$$

**Proof** We have

$$\begin{aligned} \|b - Bx_{\mathcal{B}}^k\| &= \|b - Bx_{\mathcal{B}}^k - Nx_{\mathcal{N}}^k + Nx_{\mathcal{N}}^k\| \\ &\leq \|r_p^k\| + \|Nx_{\mathcal{N}}^k\| \\ &\leq (\mu^0)^{-1}\mu^k\beta\|r^0\| + \sqrt{n}\max(x_{\mathcal{N}}^k)\|N\| \\ &\leq (\mu^0)^{-1}\mu^k\beta\|r^0\| + \tau\sqrt{n}(\xi^*)^{-1}\|N\|(x^k)^T z^k. \end{aligned} \tag{3.13}$$

Similarly,

$$\begin{aligned} \|c_{\mathcal{B}} - B^T y^k\| &= \|c_{\mathcal{B}} - B^T y^k - z^k_{\mathcal{B}} + z^k_{\mathcal{B}}\| \\ &\leq \|r^k_d\| + \|z^k_{\mathcal{B}}\| \\ &\leq (\mu^0)^{-1}\mu^k\beta\|r^0\| + \tau\sqrt{n}(\xi^*)^{-1}(x^k)^T z^k. \end{aligned} \tag{3.14}$$

Substituting $\mu^k = (x^k)^T z^k/n$, completes the proof. □

Observe that the primal residual depends on the norm of $N$, but the dual does not. The bounds on the residuals are valid once the optimal partition has been identified but not before.

### 3.2.2 Positivity of the Finite Termination Solution

Next, we demonstrate that the finite termination procedure described in Section 2.3 produces a solution that satisfies the strict complementarity condition.

First, we present theory for the case where $B$ has fewer rows than columns. The theoretical result is given for an arbitrary scaling matrix $D$ instead of $X^k_{\mathcal{B}}$ as proposed in (2.6). The following notation is needed for the proof. Let $B_1$ denote a set of maximal independent rows of the matrix $B$. Then $A_1, N_1, b_1, y^k_1,$ are the corresponding submatrices of $A$ and $N$ and components of the vectors, $b$ and $y^k$.

**Theorem 3.2** Let $\{(x^k, y^k, z^k)\}$ be generated by Algorithm 1. Assume

(1) $\mathcal{B}^k = \mathcal{B}$.

(2) The matrix $B_1$ is full rank.

(3) $D$ is a diagonal matrix such that $d_{jj} \leq x^k_j$ for all $j$ in $\mathcal{B}$ and

$$\min(d_{jj}) > \|B_1^T(B_1B_1^T)^{-1}\| \left[(\mu^0 n\xi^*)^{-1}(\xi^*\beta\|r^0\| + \mu^0\tau n\sqrt{n}\|N\|)\right](x^k)^T z^k.$$

(4) $Z^k_{\mathcal{N}}$ is a diagonal matrix such that

$$\begin{aligned}\min(z^k_{\mathcal{N}}) \quad > \quad & (n\mu^0\xi^*)^{-1}[\beta\xi^*\left(\|N^T\|\|(B_1B_1^T)^{-1}B_1\|+1\right)\|r^0\| \\ & + n\sqrt{n}\tau\mu^0\|N^T\|\|(B_1B_1^T)^{-1}B_1\|](x^k)^Tz^k.\end{aligned}$$

Then the solution $x_{\mathcal{B}}$ obtained from solving

$$\begin{array}{ll}\text{minimize} & \frac{1}{2}\|D^{-1}(x_{\mathcal{B}}-x^k_{\mathcal{B}})\|^2 \\ \text{subject to} & Bx_{\mathcal{B}}=b.\end{array} \tag{3.15}$$

satisfies $x_{\mathcal{B}}>0$. Moreover, $z_{\mathcal{N}}=c_{\mathcal{N}}-N^Ty$ is positive.

**Proof** Obviously (3.15) is equivalent to

$$\begin{array}{ll}\text{minimize} & \frac{1}{2}\|D^{-1}(x_{\mathcal{B}}-x^k_{\mathcal{B}})\|^2 \\ \text{subject to} & B_1x_{\mathcal{B}}=b_1.\end{array}$$

Let $dx_{\mathcal{B}}=D^{-1}\Delta x_{\mathcal{B}}$, where $\Delta x_{\mathcal{B}}=x_{\mathcal{B}}-x^k_{\mathcal{B}}$. Therefore,

$$\begin{aligned}\|dx_{\mathcal{B}}\| &= \|DB_1^T(B_1D^2B_1^T)^{-1}(b_1-B_1x^k_{\mathcal{B}})\| \\ &= \|DB_1^T(B_1D^2B_1^T)^{-1}B_1DD^{-1}[B_1^T(B_1B_1^T)^{-1}(b_1-B_1x^k_{\mathcal{B}})]\| \\ &\le \|D^{-1}\|\|B_1^T(B_1B_1^T)^{-1}\|\|(b_1-B_1x^k_{\mathcal{B}})\| \\ &\le \|D^{-1}\|\|B_1^T(B_1B_1^T)^{-1}\|\|(b-Bx^k_{\mathcal{B}})\|.\end{aligned} \tag{3.16}$$

From Lemma 3.2, we have

$$\|dx_{\mathcal{B}}\| \le \|D^{-1}\|\|B_1^T(B_1B_1^T)^{-1}\|\left[(n\mu^0\xi^*)^{-1}(\xi^*\beta\|r^0\|+\mu^0\tau n\sqrt{n}\|N\|)\right](x^k)^Tz^k.$$

Consequently, $\|dx_{\mathcal{B}}\|_\infty \le \|dx_{\mathcal{B}}\|_2 < 1$. The first inequality holds from the fact that for any vector $v$, $\|v\|_\infty \le \|v\|_2$, the second inequality from the assumption that $\min(d_{jj}) > \|B_1^T(B_1B_1^T)^{-1}\|\left[(n\mu^0\xi^*)^{-1}(\xi^*\beta\|r^0\|+\mu^0\tau n\sqrt{n}\|N\|)\right](x^k)^Tz^k$. Hence

$$-1 < \frac{\Delta x_j}{d_{jj}} < 1 \Rightarrow \quad x^k_j - d_{jj} < x^k_j + \Delta x_j < x^k_j + d_{jj} \quad \text{for all } j \in \mathcal{B}.$$

The additional assumption, $d_{jj} \leq x_j$ , ensures the solution is positive.

First, components of $\Delta y$ that correspond to dependent rows of $B$ are set equal to zero. Then, we solve the following least squares problem for the remaining components, $\Delta y_1$. Formally, let $\Delta y_1$ be the solution of

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|(Z_{\mathcal{N}}^k)^{-1} N_1^T \Delta y_1\|^2 \\ \text{subject to} & B_1^T \Delta y_1 = c_{\mathcal{B}} - B_1^T y_1^k \end{array} \tag{3.17}$$

Then,

$$\Delta y_1 = (B_1 B_1^T)^{-1} B_1 (c_{\mathcal{B}} - B_1^T y_1^k). \tag{3.18}$$

Let $y = y^k + \Delta y$ and $dz_{\mathcal{N}} = (Z_{\mathcal{N}}^k)^{-1}\Delta z_{\mathcal{N}} = (Z_{\mathcal{N}}^k)^{-1}(N^T(y - y^k) - r_{d_{\mathcal{N}}}^k)$. Recall that $r_d^k = c - A^T y^k - z^k$. It follows that the residual for the equations that correspond to the variables $z_{\mathcal{N}}$ is $r_{d_{\mathcal{N}}}^k = c_{\mathcal{N}} - N^T y^k - z_{\mathcal{N}}^k$. We also have, $z_{\mathcal{N}} = c_{\mathcal{N}} - N^T y$. By direct substitution, we arrive at an expression for $dz_{\mathcal{N}}$. Therefore,

$$\begin{aligned} \|dz_{\mathcal{N}}\| &= \|(Z_{\mathcal{N}}^k)^{-1}[N^T(y - y^k) - r_{d_{\mathcal{N}}}^k]\| \\ &\leq \|(Z_{\mathcal{N}}^k)^{-1}\| \left(\|N^T\|\|(B_1B_1^T)^{-1}B_1(c_{\mathcal{B}} - B_1^T y_1^k)\| + \|r_{d_{\mathcal{N}}}^k\|\right) \\ &\leq \|(Z_{\mathcal{N}}^k)^{-1}\| \left(\|N^T\|\|(B_1B_1^T)^{-1}B_1^T\|\|c_{\mathcal{B}} - B_1^T y_1^k\| + \|r_d^k\|\right) \\ &= \|(Z_{\mathcal{N}}^k)^{-1}\| \left(\|N^T\|\|(B_1B_1^T)^{-1}B_1^T\|\|c_{\mathcal{B}} - B_1^T y_1^k - z_{\mathcal{B}}^k + z_{\mathcal{B}}^k\| + \|r_d^k\|\right) \\ &\leq \|(Z_{\mathcal{N}}^k)^{-1}\| \left((\|N^T)\|(B_1B_1^T + 1)^{-1}B_1^T\|\|r_d^k\| + \|N^T\|\|(B_1B_1^T)^{-1}B_1^T\|\|z_{\mathcal{B}}^k\|\right). \end{aligned} \tag{3.19}$$

From the Feasibility Priority Principle and the bound on $z_{\mathcal{B}}^k$, we have

$$\begin{aligned} \|dz_{\mathcal{N}}\| \;\; \leq \;\; & (n\mu^0\xi^*)^{-1}\|(Z_{\mathcal{N}}^k)^{-1}\|[\beta\xi^*\left(\|N^T\|\|(B_1B_1^T)^{-1}B_1^T\| + 1\right)\|r^0\| \\ & + n\sqrt{n}\tau\mu^0\|N^T\|\|(B_1B_1^T)^{-1}B_1^T\|](x^k)^T z^k. \end{aligned}$$

By assumption (3), we have $\|dz_{\mathcal{N}}\|_2 < 1$. Combining the above inequality with the fact that for any vector $v$, $\|v\|_\infty \leq \|v\|_2$, we obtain $\|dz_{\mathcal{N}}\|_\infty \leq \|dz_{\mathcal{N}}\|_2 < 1$. Thus

$$-1 < \frac{\Delta z_j}{z_j^k} < 1 \Rightarrow \quad z_j^k - z_j^k < z_j^k + \Delta z_j < z_j^k + z_j^k \quad \text{for all } j \in \mathcal{N}.$$

□

We can state an analogous result when matrix $B$ has more columns than rows. Define $B_2$ as a maximal matrix of independent columns of $B$. Let $D_2$ be the non-singular, diagonal matrix whose diagonal entries correspond to the columns of the matrix $B_2$ and $c_2$ the corresponding elements of the objective function coefficients.

**Theorem 3.3** Let $\{(x^k, y^k, z^k)\}$ be generated by Algorithm 1. Assume

(1) $\mathcal{B}^k = \mathcal{B}$.

(2) The matrix $B_2$ is full rank.

(3) $D_2$ is a diagonal matrix such that $d_{jj} \leq x_j^k$ for all $j \in \mathcal{B}$ and

$$\min(d_{jj}) > \|(B_2^T B_2)^{-1} B_2^T\| (n\mu^0\xi^*)^{-1}(\xi^*\beta\|r^0\| + \mu^0\tau n\sqrt{n}\|N\|)(x^k)^T z^k.$$

(4) $Z_{\mathcal{N}}^k$ is a diagonal matrix such that

$$\begin{aligned}\min(z_{\mathcal{N}}^k) \quad > \quad & (n\mu^0\xi^*)^{-1}[\beta\xi^*\left(1 + \|N^T\|\|B_2(B_2^T B_2)^{-1}\|\right)\|r^0\| \\ & + n\sqrt{n}\tau\mu^0\|N^T\|\|B_2(B_2^T B_2)^{-1}\|](x^k)^T z^k.\end{aligned}$$

Then the solution $x_{\mathcal{B}}$ obtained by solving the following least squares problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|D^{-1}(x_{\mathcal{B}} - x_{\mathcal{B}}^k)\|^2 \\ \text{subject to} & Bx_{\mathcal{B}} = b. \end{array} \tag{3.20}$$

satisfies $x_{\mathcal{B}} > 0$. Moreover, $z_{\mathcal{N}} = c_{\mathcal{N}} - N^T y$ is positive.

**Proof** Recall that $\Delta x_{\mathcal{B}} = x_{\mathcal{B}} - x_{\mathcal{B}}^k$. Now, let us partition $\Delta x_{\mathcal{B}}$ into

$$\begin{pmatrix} \Delta x_1 \\ \Delta x_2 \end{pmatrix},$$

where $\Delta x_1$ and $\Delta x_2$, respectively, correspond to the dependent and independent columns of $B$. After setting $\Delta x_1 = 0$, (3.20) is equivalent to the following formulation

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{2}\|D_2^{-1}(x_2 - x_2^k)\|^2 \\ &\text{subject to} \quad B_2 x_2 = b, \end{aligned} \tag{3.21}$$

where $x_2, x_2^k$ are the subvectors of $x_{\mathcal{B}}$ and $x_{\mathcal{B}}^k$ corresponding to independent columns of $B$. Let $dx_{\mathcal{B}} = D^{-1}\Delta x_{\mathcal{B}}$. Therefore,

$$\begin{aligned} \|dx_{\mathcal{B}}\| &= \|(D_2 B_2^T B_2 D_2)^{-1} D_2 B_2^T (b - B_2 x_2^k)\| \\ &= \|(D_2)^{-1}(B_2^T B_2)^{-1} B_2^T (b - B_2 x_2^k)\| \text{ due to the nonsingularity of } D_2 \\ &\le \|D_2^{-1}\|\|(B_2^T B_2)^{-1} B_2^T\|\|(b - B_2 x_2^k)\| \\ &\le \|D_2^{-1}\|\|(B_2^T B_2)^{-1} B_2^T\|(n\mu^0\xi^*)^{-1}\left(\xi^*\beta\|r^0\| + \mu^0\tau n\sqrt{n}\|N\|\right)(x^k)^T z^k. \end{aligned} \tag{3.22}$$

The last inequality holds from Lemma 3.2. By assumption (2), $x_{\mathcal{B}} > 0$.

For the dual slack update, we have

$$\begin{aligned} \|dz_{\mathcal{N}}\| &= \|(Z_{\mathcal{N}}^k)^{-1}\left[N^T B_2 (B_2^T B_2)^{-1}(c_2 - B_2^T y^k) - r_d^k\right]\| \\ &\le \|(Z_{\mathcal{N}}^k)^{-1}\|\left(\|N^T\|\|B_2(B_2^T B_2)^{-1}\|\|c_2 - B_2^T y^k\| + \|r_d^k\|\right). \end{aligned} \tag{3.23}$$

Then

$$\begin{aligned} \|dz_{\mathcal{N}}\| \;\le\; & (n\mu^0\xi^*)^{-1}\|(Z_{\mathcal{N}}^k)^{-1}\|[\beta\xi^*\left(1 + \|N^T\|\|B_2(B_2^T B_2)^{-1}\|\right)\|r^0\| \\ & + n\sqrt{n}\tau\mu^0\|N^T\|\|B_2(B_2^T B_2)^{-1}\|](x^k)^T z^k. \end{aligned}$$

The last inequality holds from applying Lemma 3.2 and the Feasibility Priority Principle. Then by assumption (3), the theorem is complete. □

**Remark:** Since we assume the matrix $B_2$ has full column rank, the solution of the linear system of (3.20) is unique. Thus, the result presented in Theorem 3.3 can be obtained by solving the linear system without introducing a weighting matrix.

## 3.3 Choice of Scale

In Theorems 3.2 and 3.3, $\|dx_{\mathcal{B}}\|_{\infty} < 1$ was a necessary and sufficient condition to guarantee the positivity of the computed $x_{\mathcal{B}}$ on the optimal primal face. The theorems are applicable for any arbitrary diagonal matrix, $D$, that satisfies assumption (2). An obvious choice for weighting matrix $D$ is $d_{jj} = min(x_{\mathcal{B}}^k) \quad \text{for all } j \in \mathcal{B}$, which satisfies the assumptions of Theorems 3.2 and 3.3. However, an uniform weight for the diagonal elements $d_{jj}$ does not change the geometry and hence the solution of the least squares problem. The optimal value will change but not the solution.

**Remark:** Given our method of proof, $d_{jj} = x_j^k$ for $j \in \mathcal{B}$ are the maximum weights that will lead to a positive solution of problem (2.5).

## 3.4 Dual Formulation

As previously mentioned, one drawback of the weighted projection method is that the proposed formulations require separate matrix factorizations for the optimal primal and dual face identification problems. To reduce the total cost of the finite termination procedure, we suggest using the following optimization problem:

$$\min_{y} \quad \frac{1}{2}\|X_{\mathcal{B}}^k(B^T y - c_{\mathcal{B}})\|^2 \tag{3.24}$$

to find an interior point on the optimal dual face. This particular formulation allows us to reduce the work of a finite termination procedure which uses a weighted projection model to one matrix factorization which is the same cost a finite termination procedure that uses the orthogonal projection model or Mehrotra-Ye technique described in Chapter 2. Moreover, for feasible iterates, the problem is equivalent to minimizing the complementarity equation,

$$\min_{z_{\mathcal{B}}} \quad \frac{1}{2}\|X_{\mathcal{B}}^k z_{\mathcal{B}}\|^2. \tag{3.25}$$

If we denote the solution of (2.5) as $y_{WLS}$ and the solution to the above problem as $y_{X_\mathcal{B}}$, then

$$y_{X_\mathcal{B}} - y_{WLS} = (BX_\mathcal{B}^{k^2}B^T)^{-1}B(X_\mathcal{B}^{k^2} - I)(c_B - B^T y_{WLS}). \qquad (3.26)$$

Here we assume the matrix $B$ has full row rank. If $c_\mathcal{B} \in range(B^T)$, (i.e., the linear system is consistent), then $y_X = y_{WLS}$ Golub and Van Loan [15].

**Lemma 3.3** Let $\{(x^k, y^k, z^k)\}$ be generated by Algorithm 1. Assume

**(1)** $\mathcal{B}^k = \mathcal{B}$,

**(2)** $B_1$ has full row rank,

**(3)** $\min(z_\mathcal{N}^k) > (n\mu^0\xi^*)^{-1}[\beta\xi^*\left(1 + \|N^T\|\|(B_1B_1^T)^{-1}B_1\|\right)\|r^0\|$

$$+ n\sqrt{n}\tau\mu^0\|(B_1B_1^T)^{-1}B_1\|](x^k)^T z^k, \text{ and}$$

**(4)** $c_\mathcal{B} \in range(B_1^T)$.

Then $z_\mathcal{N} = c_\mathcal{N} - N^T y$ is positive.

**Proof** Let $\Delta y = y - y^k$. Then

$$\begin{aligned} \Delta y &= (B_1(X_\mathcal{B}^k)^2B_1^T)^{-1}B_1(X_\mathcal{B}^k)^2(c_\mathcal{B} - B_1^T y_1) \\ &= (B_1B_1^T)^{-1}B_1(c_\mathcal{B} - B_1^T y_1) \text{ since } c_\mathcal{B} \in range(B_1^T). \end{aligned} \qquad (3.27)$$

From the second part of Theorem 3.2, $z_\mathcal{N}$ is positive. □

Ye proposed a related least squares problem in [51] to update the lower bound of a potential function and the associated dual solution. His formulation included the entire dual residual and the dual objective function.

## 3.5 Complexity Theory

First we show that with proper choice of algorithmic parameters, Algorithm 1 obtains $\epsilon-$complementarity in polynomial time. A polynomial bound for Algorithm 1 coupled with the results in the preceding sections implies that an exact solution of the linear programming problem can be found in finite time.

To establish polynomial complexity for Algorithm 1, the initial iterate, the centering parameter, and the step length parameter must satisfy certain conditions. Let

$$\rho_0 = \min\{\|(u_0, v_0)\| : Au_0 = b, \;\; A^T\pi_0 + v_0 = c \text{ for some } \pi_0\}.$$

Then set

$$(x^0, y^0, z^0) = (\rho e, 0, \rho e), \text{ for some scalar } \rho > 0,$$

such that

$$\rho \geq \rho_0 \text{ and } \rho \geq \rho^*/\sqrt{n}, \tag{3.28}$$

where $\rho^* = \|(x^*, z^*)\|$ for an optimal solution $(x^*, y^*, z^*)$. Zhang [56] specified this particular initial point to establish polynomiality in infeasible primal-dual interior-point methods. Next, the centering parameter must be bounded away from zero. Given $\sigma_{min}, \sigma_{max}$, and $0 < \sigma_{min} < \sigma_{max} \leq 1/2$ , choose

$$\sigma^k \in [\sigma_{min}, \sigma_{max}]. \tag{3.29}$$

Finally, the step length parameter, $\alpha^k$ is chosen as the largest value of $\alpha \in [0, 1]$ such that

$$\begin{gathered}(x(\alpha), y(\alpha), z(\alpha)) \in \mathcal{N}_{-\infty}(\gamma, \beta) \\ \text{and} \\ (x^{k+1})^T z^{k+1} \leq [1 - \delta(\alpha^k)](x^k)^T z^k,\end{gathered} \tag{3.30}$$

where

$$x(\alpha) = x + \alpha\Delta x, \qquad x^{k+1} = x^k + \alpha^k\Delta x$$

and

$$\delta(\alpha) = \alpha \left( 1 - \sigma^k + \alpha \frac{\Delta x^T \Delta z}{x^T z} \right).$$

Now we can establish a polynomial complexity bound for Algorithm 1. Such a polynomial bound for infeasible interior-point algorithms was first established in Zhang [56].

**Theorem 3.4** (see for example, Wright [49]) Let $\epsilon > 0$ be given. Let $(x^0, y^0, z^0) = (\rho e, 0, \rho e)$ where $\rho$ satisfies (3.28). Assume

$$\rho^2 \leq C/\epsilon^\kappa$$

for some positive constants $C$ and $\kappa$. Furthermore, $\sigma^k$ is chosen according to (3.29) and $\alpha^k$ is chosen according to (3.30). Then there is an index $K$ with

$$K = O(n^2 \log(1/\epsilon))$$

such that the iteration sequence $\{(x^k, y^k, z^k)\}$ generated by Algorithm 1 satisfies

$$\mu^k \leq \epsilon \text{ for all } k \geq K.$$

We now present the main result; Algorithm 1 combined with the weighted projection procedure generates an exact solution in finite time.

Note that assumption (2) of Theorems 3.2 and 3.3 is equivalent to $\min(x_{\mathcal{B}}) > \|\Delta x_{\mathcal{B}}\|$. Therefore, we can restate the assumption in a more succinct manner. We also state the theorem in terms of the matrix $B^\dagger$, the pseudoinverse of $B$.

**Theorem 3.5** Let $(x^0, y^0, z^0) = (\rho e, 0, \rho e)$. Assume

(1) $x^k_{\mathcal{B}} > \frac{\gamma \xi^*}{\tau n}$ and $z^k_{\mathcal{N}} > \frac{\gamma \xi^*}{\tau n}$ for $\xi^*$ defined in (3.4),

(2) $\min(x^k_{\mathcal{B}}) > \|B^\dagger\| \left[ (\mu^0 n \xi^*)^{-1} (\xi^* \beta \|r^0\| + \mu^0 \tau n \sqrt{n} \|N\|) \right] (x^k)^T z^k$,

(3) $\min(z^k_{\mathcal{N}}) > (n\mu^0\xi^*)^{-1}[\beta\xi^*(1+\|N^T\|\|(B^\dagger)^T\|)\|r^0\|$

$$+ n\sqrt{n}\tau\mu^0\|N^T\|\|(B^\dagger)^T\|](x^k)^T z^k.$$

Then Algorithm 1 combined with the weighted projection model defined by problems (2.6) and (2.7) generates a solution in

$$O\left(n^2\left(\log n + \log\left(\frac{\tau(\max(\rho,\phi,\eta))}{\gamma(\xi^*)^2}\right)\right)\right) \text{ iterations}$$

for $\tau$ as defined in (3.11), $\phi = \|B^\dagger\|\left(\xi^*\beta\|r^0\| + \mu^0\tau n\sqrt{n}\|N\|\right)$, and $\eta = \beta\xi^*\left(1+\|N^T\|\|(B^\dagger)^T\|\right)\|r^0\| + n\sqrt{n}\tau\mu^0\|N^T\|\|(B^\dagger)^T\|$.

**Proof** From assumptions (1) and (2)

$$(x^k)^T z^k < \frac{\gamma(\xi^*)^2\mu^0}{\tau\|B^\dagger\|(\xi^*\beta\|r^0\| + \mu^0\tau n\sqrt{n}\|N\|)}. \tag{3.31}$$

For ease of notation, we set $\phi = \|B^\dagger\|\left(\xi^*\beta\|r^0\| + \mu^0\tau n\sqrt{n}\|N\|\right)$. Similarly from assumptions (1) and (3)

$$(x^k)^T z^k < \frac{\gamma(\xi^*)^2\mu^0}{\tau\left[\beta\xi^*\left(1+\|N^T\|\|(B^\dagger)^T\|\right)\|r^0\| + n\sqrt{n}\mu^0\tau\|N^T\|\|(B^\dagger)^T\|\right]}. \tag{3.32}$$

We then set $\eta = \beta\xi^*\left(1+\|N^T\|\|(B^\dagger)^T\|\right)\|r^0\| + n\sqrt{n}\tau\mu^0\|N^T\|\|(B^\dagger)^T\|$.

From Theorem 7.1 of Zhang [56], we know that the duality gap sequence $\{(x^k)^T z^k\}$ generated by Algorithm 1 converges $Q-linearly$ to zero, i.e., there exist $\delta$ such that

$$(x^{k+1})^T z^{k+1} \le (1-\delta)(x^k)^T z^k,$$

where $\delta(\alpha^k) \ge \delta$. Given the initial point defined in (3.28), Zhang (Lemma 7.2, [56]) proved that $1/\delta = O(n^2)$, i.e., there exists some positive constant $C_2$ such that $1/\delta \le C_2 n^2$.

Thus

$$(x^k)^T z^k \le (1 - C_3/n^2)^k {x^0}^T z^0, \tag{3.33}$$

where $C_3 = 1/C_2$ and ${x^0}^T z^0 = n\rho^2$.

We first derive an expression for the number of iterations that is needed to satisfy (3.31). Taking the logarithms of both sides of (3.33), we obtain

$$\begin{aligned} \log((x^k)^T z^k) &\leq k \log(1 - C_3/n^2) + \log({x^0}^T z^0) \\ &\leq k(-C_3/n^2) + \log({x^0}^T z^0) \end{aligned} \tag{3.34}$$

since $\log(1-\theta) \leq -\theta$ for $\theta < 1$.

To satisfy (3.31), we need

$$k(-C_3/n^2) + \log({x^0}^T z^0) \leq \log\gamma + 2log\xi^* + \log(\mu^0) - [\log\tau + \log\phi],$$

which implies

$$k(-C_3/n^2) + \log({x^0}^T z^0) \leq \log\gamma + 2\log\xi^* + \log({x^0}^T z^0) - \log n - (\log\tau + \log\phi).$$

Thus, $k \geq (n^2/C_3)(\log n + \log\tau + \log\phi - \log\gamma - 2\log\xi^*)$. Then the inequality (3.31) holds for any

$$k \geq K_1 = O\left(n^2 \log n + n^2 \log(\tau\phi/\gamma(\xi^*)^2)\right).$$

Similarly, we can show that inequality (3.32) holds for any

$$k \geq K_2 = O\left(n^2 \log n + n^2 \log(\tau\eta/\gamma(\xi^*)^2)\right).$$

From Theorem 3.1, we know $\mathcal{B}^k = \mathcal{B}$ when $(x^k)^T z^k < \frac{\gamma(\xi^*)^2}{\tau^2 n}$. Consequently, any

$$k \geq K_3 = O\left(n^2 \log n + n^2 \log(\tau\rho/\gamma(\xi^*)^2)\right)$$

satisfies that inequality. The strict complementary solution can be calculated when $k = \max(K_1, K_2, K_3)$. This concludes the proof. $\square$

## 3.6 Bounded Variable Linear Programs

In this section, we extend the analysis of the previous sections to linear programs with upper bound constraints.

Thus far we have only considered optimal face identification for linear programs of the form (1.1). We now turn our attention to the more general linear programming problem with upper bound constraints.

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & l \leq x \leq u \end{array} \tag{3.35}$$

where $l \in \mathbf{R}^n$ represents the vector of lower bounds and $u \in \mathbf{R}^n$ represents the vector of upper bounds for the vector $x$. Without loss of generality, we assume all the variables have lower bounds of zero and finite upper bounds. The above problem rewritten in standard form is

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x + s = u \\ & x, s \geq 0 \end{array} \tag{3.36}$$

where $s \in \mathbf{R}^n$ is the primal slack vector.

The corresponding dual problem is

$$\begin{array}{ll} \text{maximize} & b^T y - u^T w \\ \text{subject to} & A^T y \ + \ w \ - \ z = c \\ & w, z \geq 0 \end{array} \tag{3.37}$$

where $y \in \mathbf{R}^m$ are the Lagrange multipliers corresponding to the $Ax = b$ constraints, $z \in \mathbf{R}^n$ are the Lagrange multipliers corresponding to the nonnegativity constraints

$(x \geq 0)$, and $w \in \mathbf{R}^n$ are the Lagrange multipliers corresponding to the upper bound constraints.

The perturbed optimality conditions for (3.36) are

$$F_\mu(x,y,z,s,w) = \begin{pmatrix} Ax - b \\ x + s - u \\ A^T y + z - w - c \\ XZe - \mu e \\ SWe - \mu e \end{pmatrix} = 0, \quad (x,z,s,w) \geq 0, \; \mu > 0 \tag{3.38}$$

where $S = \mathrm{diag}(s)$ and $W = \mathrm{diag}(w)$.

The Jacobian of (3.38) is

$$F'_\mu(x,y,z,s,w) = \begin{pmatrix} A & 0 & 0 & 0 & 0 \\ I & 0 & 0 & I & 0 \\ 0 & A^T & I & -I & 0 \\ Z & 0 & X & 0 & 0 \\ 0 & 0 & 0 & W & S \end{pmatrix}. \tag{3.39}$$

Given feasible iterates, we see that $\|F(x,y,z,s,w)\|_1 = x^T z + s^T w$. It can be shown that the expression $x^T z + s^T w$ is equal to the duality gap, which is the difference between the primal and dual objective function values of problems (3.36) and (3.37).

Below, we present an interior-point algorithm for (3.36) which follows the framework of Algorithm 1. The duality gap contains the additional term, $s^T w$.

**Algorithm 2** Given $v^0 = (x^0, y,^0, z^0, s^0, w^0)$ with $(x^0, z^0, s^0, w^0) > 0$, for $k = 0, 1, \ldots$, do

**(1)** Choose $\sigma^k \in (0,1)$ and set $\mu^k = ((x^k)^T z^k + (s^k)^T w^k)/2n$.

**(2)** Solve for the step $\Delta v^k$

$$F'_\mu(v^k)\Delta v^k = -F_\mu(v^k).$$

**(3)** Choose $\tau^k \in (0,1)$ and set $\alpha^k = min(1, \tau^k \hat{\alpha}^k)$, where

$$\hat{\alpha}^k = \frac{-1}{\min((X^k)^{-1}\Delta x^k, (Z^k)^{-1}\Delta z^k, (S^k)^{-1}\Delta s^k, (W^k)^{-1}\Delta w^k)}.$$

**(4)** Let $v^{k+1} := v^k + \alpha^k \Delta v^k$.

**(5)** Test for convergence.

For notational convenience in the statements and proofs of the theory in the next sections, we introduce the following notation:

$$\tilde{x} = (x, s) \in \mathbf{R}^{2n}$$

$$\tilde{z} = (z, w) \in \mathbf{R}^{2n}.$$

It is obvious that $\tilde{x}^T\tilde{z} = x^Tz + s^Tw$, which is the duality gap.

The central path of (3.36) is defined as the set

$$\tilde{\mathcal{C}} = \{(\tilde{x}, y, \tilde{z}) \in \mathcal{F}^0 : \tilde{X}\tilde{Z} = \mu e \text{ where } \mu = (\tilde{x}^T\tilde{z})/2n.\}$$

For any $(\tilde{x}^*, y^*, \tilde{z}^*)$ in the relative interior of the solution set of (3.36), we define the index sets $\tilde{\mathcal{B}}$ and $\tilde{\mathcal{N}}$ as

$$\tilde{\mathcal{B}} = \{j : \tilde{x}_j^* > 0,\ 1 \le j \le 2n\} \text{ and } \tilde{\mathcal{N}} = \{j : \tilde{x}_j^* = 0,\ 1 \le j \le 2n\}.$$

The optimal primal face is

$$\tilde{\Theta}_p = \{\tilde{x} : Ax = b,\ x + s = u,\ \tilde{x} \ge 0,\ \tilde{x}_j = 0 \text{ for } j \in \tilde{\mathcal{N}}\}.$$

Similarly, the optimal dual face is

$$\tilde{\Theta}_d = \{(y, \tilde{z}) : A^Ty + z - w = c,\ \tilde{z} \ge 0,\ \tilde{z}_j = 0 \text{ for } j \in \tilde{\mathcal{B}}\}.$$

If the set of strictly feasible points is nonempty, then $\tilde{\Theta}_p$ and $\tilde{\Theta}_d$ are nonempty and bounded. Hence, $\tilde{\xi}_p^*$ and $\tilde{\xi}_d^*$ are bounded, where

$$\tilde{\xi}_p^* = \min_{j \in \tilde{\mathcal{B}}} \{\max \tilde{x}_j : \tilde{x} \in \Theta_p\}$$

$$\tilde{\xi}_d^* = \min_{j \in \tilde{\mathcal{N}}} \{\max \tilde{z}_j, : (y, \tilde{z}) \in \Theta_d\}$$

$$\tilde{\xi}^* = \min(\tilde{\xi}_p^*, \tilde{\xi}_d^*).$$

### 3.6.1 Technical Results

We revisit results from Sections 3.1 and 3.2 that have been extended to include the iteration sequence and solution of Algorithm 2.

A centrality measure for problem (3.36) can be written as

$$\frac{min(\tilde{X}^k \tilde{Z}^k e)}{(\tilde{x}^k)^T \tilde{z}^k / 2n} \geq \gamma \tag{3.40}$$

where $\gamma > 0$ and is independent of $k$. Inequality (3.40) requires that the pairwise products of the iterates decrease at a controlled rate. No pairwise product can converge to zero faster than the others.

A central path neighborhood that includes infeasible points can be defined as

$$\begin{aligned} \mathcal{N}_{-\infty}(\gamma, \beta) = \{(\tilde{x}, y, \tilde{z})| \quad & \|(r_p, r_u, r_d)\| \leq [\|(r_p^0, r_u^0, r_d^0)\|/\mu^0]\beta\mu, \\ & (\tilde{x}, \tilde{z}) > 0, \quad \min(\tilde{X}\tilde{Z}e) \geq \gamma\mu, \quad \mu = (x^T z + s^T w)/2n\} \end{aligned} \tag{3.41}$$

where $\gamma \in (0, 1)$, $\beta \geq 1$, $r_p = b - Ax$, $r_u = u - x - s$, and $r_d = c - A^T y - z + w$.

Theorem 3.6 shows that the optimal partition can be identified in infeasible interior-point algorithms once the residual $\|F(\tilde{x}, y, \tilde{z})\|_1$ is sufficiently small.

**Theorem 3.6** Let $(\tilde{x}^*, y^*, \tilde{z}^*) \in \mathcal{S}$, and $\tilde{\mathcal{B}}^k = \{\tilde{x}_j^k > \tilde{z}_j^k\}$. Then $\tilde{\mathcal{B}}^k = \tilde{\mathcal{B}}$, when

$$(\tilde{x}^k)^T \tilde{z}^k < \gamma(\tilde{\xi}^*)^2 / 2\tilde{\tau}^2 n,$$

for

$$\tilde{\tau} = \beta \left[ \frac{2-\alpha^0}{\alpha^0} + \frac{(\tilde{x}^0)^T \tilde{z}^* + (\tilde{z}^0)^T \tilde{x}^*}{(\tilde{x}^0)^T \tilde{z}^0} \right] > 0.$$

**Proof** Substitute $\tilde{x}, \tilde{z},$ and $2n$ for $x$,$z$, and $n$ in Theorem 3.1. □

Subsequent theory requires bounds on the primal and dual residuals. Rather than repeatedly deriving the bounds, we establish them at this juncture. Note that the bound depends on the current duality gap, the initial duality gap, and the minimum positive value in the solution set, which is not known a priori.

**Lemma 3.4** Let $\{(\tilde{x}^k, y^k, \tilde{z}^k)\}$ be the iteration sequence generated by Algorithm 2. Further, assume that $\tilde{\mathcal{B}}^k = \tilde{\mathcal{B}}$. Then,

$$\|b - Bx_{\mathcal{B}}^k\| \le (2n\mu^0\tilde{\xi}^*)^{-1}\left(\tilde{\xi}^*\beta\|r^0\| + 2n\sqrt{2n}\mu^0\tilde{\tau}\|N\|\right)\left((x^k)^T z^k + (s^k)^T w^k\right)$$

$$\|c_{\mathcal{B}} - B^T y^k\| \le (2n\mu^0\tilde{\xi}^*)^{-1}(\tilde{\xi}^*\beta\|r^0\| + 4n\sqrt{2n}\mu^0\tilde{\tau})\left((x^k)^T z^k + (s^k)^T w^k\right).$$

**Proof** We omit the first part of the proof because it follows directly from Lemma 3.1. The dual feasibility constraint includes an additional variable, $w$. It is interesting to see how introducing the variable $w$ affects the bound on $\|c_{\mathcal{B}} - B^T y^k\|$.

$$\begin{aligned}
\|c_{\mathcal{B}} - B^T y^k\| &= \|c_{\mathcal{B}} - B^T y^k - z_{\mathcal{B}}^k + z_{\mathcal{B}}^k + w_{\mathcal{B}}^k - w_{\mathcal{B}}^k\| \\
&\le \|r_d^k\| + \|z_{\mathcal{B}}^k - w_{\mathcal{B}}^k\| \\
&\le \|r_d^k\| + \|z_{\mathcal{B}}^k\| + \|w_{\mathcal{B}}^k\| \qquad (3.42) \\
&\le \|r_d^k\| + 2\sqrt{2n}\max(\tilde{z}_{\mathcal{B}}^k) \\
&\le (\mu^0)^{-1}\mu^k\beta\|r^0\| + \tilde{\tau}2\sqrt{2n}(\tilde{\xi}^*)^{-1}\left((x^k)^T z^k + (s^k)^T w^k\right).
\end{aligned}$$

Substituting $\mu^k = \left((x^k)^T z^k + (s^k)^T w^k\right)/2n$, completes the proof. □

### 3.6.2 Primal Model

Generating an exact solution for the bounded variable linear program is complicated by the fact that a solution of any finite termination procedure must not only satisfy $x_{\mathcal{B}} \geq 0$ but also $x_{\mathcal{B}} \leq u_{\mathcal{B}}$, where $u_{\mathcal{B}}$ is the subvector of $u$ corresponding to $x_{\mathcal{B}}$. In this section, we consider two strategies for incorporating upper bound information into a finite termination procedure.

#### Extension of the Orthogonal Projection Model

One approach to incorporating upper bound information is to explicitly include the upper bound constraints as equalities in the problem formulation. The resulting model is

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\left\| \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} x_{\mathcal{B}} - x_{\mathcal{B}}^k \\ s_{\mathcal{B}} - s_{\mathcal{B}}^k \end{pmatrix} \right\|^2 \\ \text{subject to} & Bx_{\mathcal{B}} = b \\ & x_{\mathcal{B}} + s_{\mathcal{B}} = u_{\mathcal{B}}, \end{array} \tag{3.43}$$

where $s_{\mathcal{B}}$ is the subvector of $s$ corresponding to $x_{\mathcal{B}}$. From Chapter 1, recall that

$$\mathcal{B} = \{j : x_j^* > 0,\ 1 \leq j \leq n\}$$

and the columns of $A$ whose indices lie in the set $\mathcal{B}$ comprise $B$.

Problem (3.43) is a natural extension of problem (2.4). In fact, when the interior-point iterates are feasible, we can show that the solution of (3.43) is the same as the solution of (2.4). For infeasible interior-point algorithms,

$$\hat{x}_{\mathcal{B}} = x_{\mathcal{B}} + \frac{1}{2} P_B(r_u^k)$$

where $\hat{x}_{\mathcal{B}}$ is the solution of problem (3.43), $x_{\mathcal{B}}$ is the solution of problem (2.4), and $P_B$ is the projection of the upper bound residual onto $\mathcal{N}(B)$, the null space of $B$. Despite

the presence of the upper bound constraints in (3.43), the solution of the model is not affected, unless the model is implemented in an infeasible interior-point algorithm. Moreover, (3.43) suffers from the same drawbacks as the orthogonal projection model.

### Extension of the Weighted Projection Model

The bounded variable linear program (3.36) contains two sets of inequalities involving the vector $x$ ($x \geq 0$ and $x \leq u$). The component-wise distance of $x^k$ to its lower bound is

$$(x_j^k - 0) \text{ or simply } x_j^k.$$

Similarly,

$$(u_j - x_j^k)$$

is the component-wise distance of $x^k$ to its upper bound.

The role of the weighting matrix $(X_{\mathcal{B}}^k)^{-1}$ in problem (2.6) is to incorporate information about the inequalities, $x_{\mathcal{B}} \geq 0$, into the model. The weight is sufficient for problems of the form (1.1). However, for bounded variable linear programs, one-half of the inequalities are ignored. Specifically, the model does not incorporate the inequalities corresponding to the upper bounds of problem (3.36). It would seem natural to incorporate the upper bound inequalities into the model in the same manner as the lower bound inequalities appear. Consequently,

$$\operatorname{diag}(u_j - x_j^k) \text{ for } j \in \mathcal{B}$$

is another potential weighting matrix.

How should we include this information into the model? Clearly replacing $X_{\mathcal{B}}^k$ in problem (2.6) with $U_{\mathcal{B}} - X_{\mathcal{B}}^k = \operatorname{diag}(u_{\mathcal{B}} - x_{\mathcal{B}}^k)$ does not resolve the issue of incorporating all bound information. This substitution robs the model of lower bound information.

We introduce a diagonal weighting matrix $D$, where

$$d_{jj} = \min(x_j^k, u_j - x_j^k) \quad \text{for } j \in \mathcal{B},$$

to incorporate both lower and upper bound information into a finite termination procedure. We select the minimum of the distances to a bound as a weight to penalize long steps of a component of $x_{\mathcal{B}}^k$ towards its nearest bound. Whereas the weighted projection model described in Section 2.3 only penalizes movement of variables close to zero, weighting the objective function by $D$ penalizes the movement of variables in the direction of their nearest bound. Therefore, if $x_j^k$ for $j \in \mathcal{B}$ is close to its upper bound, the weight in (3.44) prevents the *jth* component of the solution vector $x_{\mathcal{B}}$ from violating its upper bound as well as its lower bound which is the desired result.

We propose the following weighted projection model

$$\begin{aligned} \text{minimize} \quad & \tfrac{1}{2}\|D^{-1}(x_{\mathcal{B}} - x_{\mathcal{B}}^k)\|^2 \\ \text{subject to} \quad & Bx_{\mathcal{B}} = b \end{aligned} \tag{3.44}$$

where

$$d_{jj} = \min(x_j^k, u_j - x_j^k) \quad \text{for } j \in \mathcal{B}. \tag{3.45}$$

When $x_j^k = u_j$ for $j \in \mathcal{B}$, we remove the corresponding columns from B and update the right-hand side. Then we solve an equivalent form of the model.

The following theorem is a restatement of Theorem 3.2 for the diagonal matrix $D$ defined in (3.45). Note that assumption 2 still requires the diagonal element to be less than or equal to the current iterate. Recall that the matrix $B_1$ denotes a set of maximal independent rows of the matrix $B$.

**Theorem 3.7** Let $\{(\tilde{x}^k, y^k, \tilde{z}^k)\}$ be generated by Algorithm 2. Assume

(1) $\tilde{\mathcal{B}}^k = \tilde{\mathcal{B}}$.

(2) The matrix $B_1$ has full row rank.

(3) $D$ is a nonsingular matrix such that

$$\begin{aligned}\min(d_{jj}) \;&>\; \left[(2n\mu^0\tilde{\xi}^*)^{-1}(\tilde{\xi}^*\beta\|r^0\| + 2n\sqrt{2n}\mu^0\tilde{\tau}\|N\|)\right] \\ &\quad \|B_1^T(B_1B_1^T)^{-1}\|(\tilde{x}^k)^T\tilde{z}^k.\end{aligned}$$

Then the solution $x_{\mathcal{B}}$ obtained from solving

$$\begin{aligned}&\text{minimize} \quad \tfrac{1}{2}\|D^{-1}(x_{\mathcal{B}} - x_{\mathcal{B}}^k)\|^2 \\ &\text{subject to} \quad Bx_{\mathcal{B}} = b\end{aligned} \tag{3.46}$$

satisfies

$$0 < x_{\mathcal{B}} \le u_{\mathcal{B}} \text{ and } 0 \le s_{\mathcal{B}} < u_{\mathcal{B}}.$$

**Proof** See proof of Theorem 3.2. □

From assumption (2) and the fact that $\|dx_{\mathcal{B}}\|_\infty \le \|dx_{\mathcal{B}}\|_2 < 1$, we have the following description of the solution. Recall that $\Delta x_{\mathcal{B}} = x_{\mathcal{B}} - x_{\mathcal{B}}^k$ and $dx_{\mathcal{B}} = D^{-1}\Delta x_{\mathcal{B}}$.

Case 1: If $x_j^k < u_j - x_j^k$ for all $j \in \mathcal{B}$, then

$$-1 < \frac{\Delta x_j}{x_j^k} < 1.$$

Hence,

$$x_j^k - x_j^k < x_j^k + \Delta x_j < x_j^k + x_j^k \text{ and consequently, } 0 < x_{\mathcal{B}} < 2x_{\mathcal{B}} < u_{\mathcal{B}}.$$

Case 2: If $u_j - x_j^k < x_j^k$ for all $j \in \mathcal{B}$, then

$$-1 < \frac{\Delta x_j}{u_j - x_j^k} < 1.$$

Hence,

$$x_j^k - (u_j - x_j^k) < x_j^k + \Delta x_j < x_j^k + (u_j - x_j^k) \text{ and consequently, } 0 < x_{\mathcal{B}} < u_{\mathcal{B}}.$$

This particular choice of weighting matrix satisfies both the lower and upper bound constraints. If we let $D = X_{\mathcal{B}}^k$ and Case 2 holds, it is trivial to show that the solution $x_{\mathcal{B}}$ satisfies the nonnegativity constraints but the upper bound constraints may be violated. In Section 3.7, we present numerical evidence to support the theory.

### 3.6.3 Dual Model

Now, let us consider the optimal dual face identification problem.

To find a feasible point on the optimal dual face for bounded variable problems, we solve

$$\min_{y} \quad \frac{1}{2}\|D^{-1}(B^T y - c_{\mathcal{B}})\|^2 \tag{3.47}$$

for $\Delta y$ and then update the vectors $w$ and $z$. Here,

$$d_{jj} = \min(x_j^k, u_j - x_j^k) \text{ for } j \in \mathcal{B}.$$

With this formulation only one matrix factorization is needed to solve both the optimal primal and dual face identification problems.

To conclude that our finite termination procedure is successful, we must now show that the dual variables $z_{\mathcal{N}}$ and $w_{\mathcal{N}}$ are nonnegative.

**Lemma 3.5** Let $\{(\tilde{x}^k, y^k, \tilde{z}^k)\}$ be generated by Algorithm 2. Assume

**(1)** $\tilde{\mathcal{B}}^k = \tilde{\mathcal{B}}$,

**(2)** $B_1$ has full row rank,

**(3)** $\min(\tilde{z}_{\mathcal{N}}^k) > (2n\mu^0\xi^*)^{-1}[\beta\xi^*\left(1 + \|N^T\|\|(B_1B_1^T)^{-1}B_1\|\right)\|r^0\|$

$$+ 4n\sqrt{2n}\tau\mu^0\|N^T\|\|(B_1B_1^T)^{-1}B_1\|](\tilde{x}^k)^T\tilde{z}^k, \text{ and}$$

**(4)** $c_{\mathcal{B}} \in range(B_1^T)$.

Then $z_{\mathcal{N}}$ and $w_{\mathcal{N}}$ are nonnegative.

**Proof** Let $\Delta y = y - y^k$. Then

$$\begin{aligned} \Delta y &= (B_1 D^2 B_1^T)^{-1} B_1 D^2 (c_{\mathcal{B}} - B_1^T y_1) \\ &= (B_1 B_1^T)^{-1} B_1 (c_{\mathcal{B}} - B_1^T y_1) \text{ since } c_{\mathcal{B}} \in range(B_1^T). \end{aligned} \tag{3.48}$$

From problem (3.37), we have

$$A^T y + z - w = c \text{ which gives us}$$

$$c_{\mathcal{N}} - N^T y = z_{\mathcal{N}} - w_{\mathcal{N}}.$$

If $c_{\mathcal{N}} - N^T y < 0$, then $z_{\mathcal{N}} = 0$ and $w_{\mathcal{N}} > 0$. Therefore, we have to check for optimality(i.e., that $w_{\mathcal{N}}$ is positive). From the dual constraint, it is easy to see that

$$\begin{aligned} \|dw_{\mathcal{N}}\| &= \|N^T(y - y^k) - r^k_{d_{\mathcal{N}}}\| \\ &\leq \|N^T\| \|(B_1 B_1^T)^{-1} B_1 (c_{\mathcal{B}} - B_1^T y_1^k)\| + \|r^k_{d_{\mathcal{N}}}\| \\ &\leq \|N^T\| \|(B_1 B_1^T)^{-1} B_1^T\| \|c_{\mathcal{B}} - B_1^T y_1^k\| + \|r^k_d\|. \end{aligned} \tag{3.49}$$

From Lemma 3.1 and the bound on $\tilde{z}^k_{\mathcal{B}}$, we have

$$\begin{aligned} \|dw_{\mathcal{N}}\| \leq\ & (2n\mu^0\xi^*)^{-1}[\beta\xi^* \left(\|N^T\| \|(B_1 B_1^T)^{-1} B_1^T\| + 1\right) \|r^0\| \\ & + 4n\sqrt{2n}\tau\mu^0 \|N^T\| \|(B_1 B_1^T)^{-1} B_1^T\|](\tilde{x}^k)^T \tilde{z}^k. \end{aligned}$$

From the third assumption of the lemma, we have $\|dw_{\mathcal{N}}\|_2 < 1$. Combining the above inequality with the fact that for any vector $v$, $\|v\|_\infty \leq \|v\|_2$, we obtain $\|dw_{\mathcal{N}}\|_\infty \leq \|dw_{\mathcal{N}}\|_2 < 1$. Thus

$$-1 < \frac{\Delta w_j}{w_j^k} < 1 \Rightarrow \quad w_j^k - w_j^k < w_j^k + \Delta w_j < w_j^k + w_j^k \quad \text{for all } j \in \mathcal{N}.$$

If $c_{\mathcal{N}} - N^T y > 0$, then $z_{\mathcal{N}} > 0$ and $w_{\mathcal{N}} = 0$. Hence, we only have to check that $z_{\mathcal{N}}$ is positive. The proof that $z_{\mathcal{N}} > 0$ follows the same format as the preceding proof that $w_{\mathcal{N}} > 0$. $\square$

## 3.7 Numerical Results

The numerical experiments were conducted on a Sun workstation with 64 bit arithmetic. We used the LIPSOL - Linear programming Interior-Point SOLver- package developed under the MATLAB* environment. The software package, written by Zhang [57], implements an infeasible primal-dual predictor-corrector interior point method. The *netlib* suite of linear programming problems comprises the test set. We tested 87 out of 96 problems available in *netlib*.

The initial matrix is scaled in an attempt to achieve row/column equilibration. Preprocessing deletes fixed variables, deletes zero rows and columns from the matrix $A$, solves equations of one variable, and shifts nonzero lower bounds to zero. For problem *greenbea*, preprocessing deletes fixed variables, deletes zero columns from the matrix $A$, and shifts nonzero lower bounds to zero. No other preprocessing is performed.

The transformed linear systems that define the optimal face are solved in the least squares sense. Depending on the dimensions of the matrix $B$, we form the lower dimensional normal equations matrix. If the linear system is underdetermined, we factor $BD^2B^T$, where $D$ is as proposed in previous sections. Likewise, if the linear system is overdetermined, we factor $B^TB$.

The LIPSOL code implements the Cholesky-Infinity method to factor the coefficient matrix at each iteration. The Cholesky-Infinity factorization [58] allows the decomposition of positive semidefinite matrices. In the presence of degeneracies, the standard Cholesky factorization breaks down. When a negligible pivot is encountered in the Cholesky-Infinity factorization, the corresponding $L(j,j)$ element is set to a

*MATLAB is a registered trademark of The MathWorks, Inc.

large number, where $LL^T$ is the Cholesky factorization of $BD^2B^T$ (or $BB^T$). For a more thorough discussion of the Cholesky-Infinity factorization, see Zhang [58].

Dividing by $L(j,j)$ in the solution phase essentially zeros out the right-hand side. Therefore, the least squares solution is a basic solution. That is, the components of the solution vector corresponding to dependent columns of the matrix are set to zero.

We implemented the following optimal face identification procedure.

**Procedure 2** *(Optimal Face Identification - Finite Termination Procedure)*

**(1)** If
$$\frac{|c^T x^k - b^T y^k|}{1 + |b^T y^k|} \leq 10^{-8},$$
and the number of projection attempts $\leq 6$, set
$$\mathcal{B}^k = \{j : z_j^k \leq 1.e-14 \quad \text{or} \quad |\Delta^p x^k|/x_j^k \leq |\Delta^p z^k|/z_j^k\}.$$
Here variables and the Tapia indicators are used to identify the optimal partition.

**(2)** Solve the optimal primal and dual face identification problems by any of the projection models described in the previous sections.

**(3)** Update,
$$x = \begin{cases} x_j^k + \Delta x_j & j \in \mathcal{B} \\ 0 & j \in \mathcal{N} \end{cases}$$
$y = y^k + \Delta y$ and $z = c - A^T y$.

**(4)** If upper bounds exist, set $s = u - x$, $\delta = c - A^T y$,
$$z = \begin{cases} 0 & \text{if } \delta < 0 \\ \delta & \text{else} \end{cases} \quad \text{and } w = \begin{cases} -\delta & \text{if } \delta < 0 \\ 0 & \text{else} \end{cases}$$

**(5)** We set dual bound infeasibility $(dbi) = \max(0, -z_N, -z_B)$. If the computed solution is complementary and satisfies

$$\max\left(\frac{\|Ax-b\|}{1+\|b\|}, \frac{\|A^Ty+z-c\|}{1+\|c\|}, \frac{|c^Tx-b^Ty|}{1+|b^Ty|}\right) \leq 10^{-11}$$

and $dbi < 10^{-9}$, we terminate the algorithm with a solution. If not, we repeat the finite termination procedure at the next interior-point iteration.

The update formula for the dual variables is the same strategy used by Resende, et al in [37], [39], and [40], to generate feasible dual variables.

The first column in Table 3.1 represents the number of misses before the optimal face identification problem was solved to the desired accuracy. A miss occurs if a finite termination procedure generates a point which does not satisfy both the feasibility and optimality tolerances. The second column gives the number of problems solved by the orthogonal projection (OP) model for the given number of misses. The third column contains the computational results of Ye's weighted projection (WP) model and the fourth column shows the number of problems per miss for our modified weighted projection (MWP) model with weighting matrix $D = \min(X_{\mathcal{B}}^k, U_{\mathcal{B}} - X_{\mathcal{B}}^k)$.

| | Subproblems | | |
|---|---|---|---|
| # of misses | OP | WP | MWP |
| 0 | 64 | 65 | 69 |
| 1 | 13 | 13 | 12 |
| 2 | 3 | 3 | 5 |
| 3 | 2 | 3 | 1 |
| 4 | 2 | 0 | 0 |
| 5 | 1 | 1 | 0 |
| more than 5 | 2 | 2 | 0 |
| TOTAL misses | 48 | 43 | 25 |

**Table 3.1** Comparison of missed projections per model

Solving the three models produce identical results for problems with no upper bound constraints, with the exception of problem *scrs8*. The problem *scrs8* is a staircase problem that after preprocessing has 485 rows and 1270 columns. The problem requires more projection attempts when the matrix $B$ is weighted by the current iterate than when it is not. The matrix $B$ has 305 rows and 317 columns. The values of $x_{\mathcal{B}}$ range from 7.71e+05 to 7.77e-04. The condition number of the matrix $BX_{\mathcal{B}}^2B^T$ is 1.3e+29 and the MATLAB calculated rank is 169. The extra projection may be the result of the ill-conditioned matrix. The corresponding estimates for the matrix $BB^T$ are a condition number of 1.98e+17 and MATLAB calculated rank of 301. When we weight by the minimum of the current iterate only one projection is needed to obtain a solution. The condition number of the corresponding matrix is 2.91e+18; the numerical rank is 301. For this particular problem, weighting by the minimum of the current iterate produces similar numerical effects as weighting by the identity matrix.

To illustrate the effectiveness of weighting, let us look at how the misses per model compare for problems with upper bounds. Column 1 gives the number of misses by a finite termination procedure. Columns 2 through 4 give the number of problems solved by the three respective projection models.

With the modified weighting matrix, we are able to compute interior points on the optimal face for all problems in our test set. The other two models fail to deliver a solution for two problems, *greenbea* and *nesm*. These two problems have upper bound constraints. If we weight the constraint matrix of problem *greenbea* by the modified weighting matrix, we can compute an interior point on the optimal face in one projection attempt. The solution agrees to thirteen digits with the CPLEX† reoptimized

†CPLEX is a trademark of CPLEX Optimization, Inc.

| | Subproblems | | |
|---|---|---|---|
| # of misses | OP | WP | MWP |
| 0 | 13 | 15 | 19 |
| 1 | 13 | 12 | 11 |
| 2 | 2 | 2 | 4 |
| 3 | 2 | 3 | 1 |
| 4 | 2 | 0 | 0 |
| 5 | 1 | 1 | 0 |
| more than 5 | 2 | 2 | 0 |
| TOTAL misses | 46 | 40 | 22 |

**Table 3.2** Problems with upper bound constraints

objective function value reported in Table II of Bixby [6]. The most accurate solutions are obtained when we weight by the matrix $D$, where $d_{jj} = \min(x_j^k, u_j - x_j^k)$ for $j \in \mathcal{B}$. Weighting by the current iterate, $X_{\mathcal{B}}^k$, produces the least accurate solutions.

Problem *scsd6* exemplifies the instance where the partition has been correctly identified, but the model cannot verify its correctness by computing an interior solution on the optimal primal and dual faces. For problem *scsd6*, the finite termination procedure fails to compute a positive $z_{\mathcal{N}}$ vector on its first two attempts. However, the partition $\mathcal{B}^k$ for $k \geq 12$ is invariant. The table below shows that assumption 3 of Theorem 3.2 ( $\min(z_{\mathcal{N}}) > \|\Delta z_{\mathcal{N}}\|$) is violated on the first two calls of the procedure.

| Iteration | $\|\Delta z_{\mathcal{N}}\|$ | $\min(z_{\mathcal{N}}^k)$ |
|---|---|---|
| 12 | 7.93e-08 | 1.28e-09 |
| 13 | 3.74e-08 | 9.27e-10 |
| 14 | 8.44e-10 | 9.28e-10 |

**Table 3.3** Problem SCSD6, Weighted Projection model

# Chapter 4

# Variants of the Mehrotra-Ye Procedure

In this chapter, we study the effectiveness of using Gaussian elimination to solve the linear feasibility problems associated with the optimal primal and dual faces. Recall that one of the major drawbacks of the Mehrotra-Ye procedure was that, unlike the orthogonal projection methods, it does not incorporate nonnegativity information from the current iterate into the model. We propose two strategies that explicitly include the current interior-point iterate into the model. The goal is to bias the Gaussian elimination so that the columns corresponding to the smallest components of $x^k_{\mathcal{B}}$ are chosen as pivots last. The first idea is to column scale the matrix $B$ by the current iterate. In [42], Skeel provided a theoretical basis for column scaling as an effective tool to achieve numerical stability. The second strategy is to order the columns of $B$, where the ordering is a function of the current iterate.

The following example illustrates the benefits of the new strategies.

**Example:** Consider the optimal primal face defined by the single equation

$$10x_1 + x_2 = 1,$$

where $x_1^k = .05$ and $x_2^k = 1.002$. To find a feasible point on the optimal primal face, we perform Gaussian elimination with partial pivoting on the transpose of the constraint matrix.

For a single constraint, we can examine all the possible bases in $O(|\mathcal{B}|)$ time and select the one that generates a positive solution. For larger problems, we would have to inspect a combinatorial number of bases which would be extremely costly as the dimensions of the matrix $B$ increase.

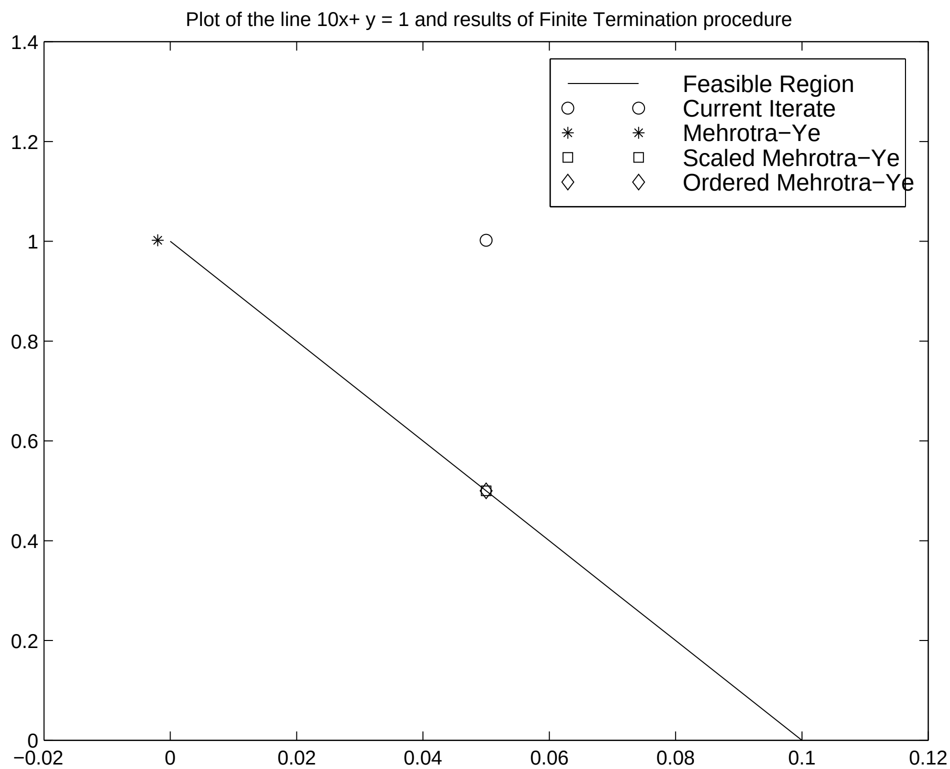


**Figure 4.1** Comparison of variants of Mehrotra-Ye

| Method | System | Basis Matrix | Solution |
|---|---|---|---|
| Mehrotra-Ye | $10\Delta x_1 + \Delta x_2 = -.502$ | 10 | $x = (-.002, 1.0002)^T$ |
| Scaled | $.5\Delta x_1 + \Delta x_2 = -.502$ | 1 | $x = (.05, .5)^T$ |
| Ordered | $10\Delta x_1 + \Delta x_2 = -.502$ | 1 | $x = (.05, .5)^T$ |

**Table 4.1** Comparison of variants of the Mehrotra-Ye technique

## 4.1 Scaled Columns

Skeel [42] proved that Gaussian elimination with row pivoting is numerically stable when the matrix is column scaled by $D = \text{diag}\,|\hat{v}|$, where $\hat{v}$ is the computed solution of $Av = g$. Unfortunately, the value of $\hat{v}$ is not known when the factorization begins. Row pivoting is defined as the interchange of columns so that each pivot is the largest in its row. The theory assumes that columns of $A$ corresponding to negligible components of $\hat{v}$ are selected as pivots last.

Skeel's theorem motivated us to consider finding a feasible point on the optimal primal and dual faces, respectively, by solving the following linear systems.

$$(BX_{\mathcal{B}}^k)dx_{\mathcal{B}} = b - Bx_{\mathcal{B}}^k \quad \text{and} \quad (B^TY^k)dy = c_{\mathcal{B}} - B^Ty^k \tag{4.1}$$

where

$$Y_j^k = \begin{cases} y_j^k & y_j^k \neq 0 \\ 1 & \text{else} \end{cases}$$

Here, we approximate the computed solution with the value of the current interior-point iterate.

We now prove that the solutions of the linear systems just described leads to the calculation of strictly positive vectors, $x_{\mathcal{B}}$ and $z_{\mathcal{N}}$. First, we define $\bar{B}$ as an arbitrary nonsingular submatrix of $B$ with maximal rank. Similarly, $\bar{N}$ and $\bar{Y}^k$ denote the corresponding submatrices of $N$ and $Y$.

**Theorem 4.1** Consider the iteration sequence $\{(x^k, y^k, z^k)\}$ generated by Algorithm 1. Assume

(1) $\mathcal{B}^k = \mathcal{B}$.

(2) $\bar{B}$ and $\bar{Y}^k$ are nonsingular matrices.

(3) $\min(x_{\mathcal{B}}^k) > \|\bar{B}^{-1}\|\,[(\mu^0 n\xi^*)^{-1}(\xi^*\beta\|r^0\| + \mu^0\tau n\sqrt{n}\|N\|)]\,(x^k)^Tz^k$.

(4) $\min(z^k_{\mathcal{N}}) > (n\mu^0\xi^*)^{-1}[\beta\xi^*\left(1+\|N^T\|\|\bar{B}^{-T}\|\right)\|r^0\|$

$$+\, n\sqrt{n}\tau\mu^0\|N^T\|\|\bar{B}^{-T}\|](x^k)^T z^k.$$

Then $x_{\mathcal{B}} = x^k_{\mathcal{B}} + X^k_{\mathcal{B}} dx_{\mathcal{B}} > 0$, where $dx_{\mathcal{B}}$ is the solution of

$$(BX^k_{\mathcal{B}})dx_{\mathcal{B}} = b - Bx^k_{\mathcal{B}}. \tag{4.2}$$

Moreover, $z_{\mathcal{N}} > 0$, where

$$(B^T Y^k)dy = c_{\mathcal{B}} - B^T y^k \tag{4.3}$$

and $y = y^k + Y^k dy,\ z_{\mathcal{N}} = z^k_{\mathcal{N}} + \Delta z_{\mathcal{N}}$.

**Proof** Assume $\bar{B}$ is a nonsingular submatrix of $B$. Let $dx_{\bar{\mathcal{B}}} = (X^k_{\bar{\mathcal{B}}})^{-1}\Delta x_{\bar{\mathcal{B}}}$. Then

$$\begin{aligned}
\|\Delta x_{\bar{\mathcal{B}}}\| &= \|X^k_{\bar{\mathcal{B}}} dx_{\bar{\mathcal{B}}}\| \\
&= \|X^k_{\bar{\mathcal{B}}}(\bar{B}X^k_{\bar{\mathcal{B}}})^{-1}(\bar{b} - \bar{B}x^k_{\bar{\mathcal{B}}})\| \\
&= \|\bar{B}^{-1}(\bar{b} - \bar{B}x^k_{\bar{\mathcal{B}}})\| \text{ since } (X^k_{\bar{\mathcal{B}}})^{-1} \text{ is nonsingular by construction.} \\
&\leq \|\bar{B}^{-1}\|(\mu^0 n\xi^*)^{-1}\left(\xi^*\beta\|r^0\| + \mu^0\tau n\sqrt{n}\|N\|\right)(x^k)^T z^k.
\end{aligned} \tag{4.4}$$

The last inequality holds from an application of (3.12). From assumption (2), the solution $x_{\bar{\mathcal{B}}}$ is positive. Let $dx_j = 0$, for $j \in \mathcal{B} \setminus \bar{\mathcal{B}}$ when column $B_{.j}$ is linearly dependent on columns of $\bar{B}$. Hence the entire vector $x_{\mathcal{B}}$ is positive.

We now prove that the dual solution is positive. First, we compute the solution of the scaled dual Mehrotra-Ye formulation. Let $\bar{dy} = (\bar{Y}^k)^{-1}\Delta\bar{y}$, where $\bar{Y}^k$ corresponds to the linearly independent rows of the matrix $\bar{B}$. The remaining components of $dy$ are set equal to zero. Then

$$\Delta\bar{y} = \bar{Y}^k\bar{dy} = \bar{Y}^k(\bar{B}^T\bar{Y}^k)^{-1}(c_{\bar{\mathcal{B}}} - \bar{B}^T\bar{y}^k).$$

Given $z_{\mathcal{N}} = c_{\mathcal{N}} - N^T y$, we have

$$
\begin{aligned}
\|\Delta z_{\mathcal{N}}\| &= \|N^T(y - y^k) - r^k_{d_{\mathcal{N}}}\| \\
&\leq \|N^T(y - y^k)\| + \|r^k_{d_{\mathcal{N}}}\| \\
&\leq \|N^T\| \|\bar{Y}^k(\bar{B}^T\bar{Y}^k)^{-1}(c_{\bar{\mathcal{B}}} - \bar{B}^T\bar{y}^k)\| + \|r^k_{d_{\mathcal{N}}}\| \\
&\leq \|N^T\| \|\bar{B}^{-T}(c_{\mathcal{B}} - \bar{B}^T y^k)\| + \|r^k_d\| \\
&\leq (n\mu^0\xi^*)^{-1}\left[\beta\xi^*\left(1 + \|N^T\|\|\bar{B}^{-T}\|\right) r^0\| + n\sqrt{n}\tau\mu^0\|N^T\|\|\bar{B}^{-T}\|\right]((x^k)^T z^k).
\end{aligned} \tag{4.5}
$$

The last inequality follows from (3.12) and the Feasibility Priority Principle. Hence by assumption (3), $z_{\mathcal{N}} > 0$. □

We state the corresponding result for bounded variable linear programs. Recall that

$$
\tilde{x} = (x, s) \in \mathbf{R}^{2n}
$$
$$
\tilde{z} = (z, w) \in \mathbf{R}^{2n}.
$$

**Theorem 4.2** Consider the iteration sequence $\{(\tilde{x}^k, y^k, \tilde{z}^k)\}$ generated by Algorithm 2. Assume

(1) $\tilde{\mathcal{B}}^k = \tilde{\mathcal{B}}$,

(2) $\bar{B}$, $\bar{Y}^k$, and $\bar{D}$ are nonsingular matrices, where

$$
\bar{D} = \min(X^k_{\bar{\mathcal{B}}}, U_{\bar{\mathcal{B}}} - X^k_{\bar{\mathcal{B}}}),
$$

(3) $\min(\bar{d}_{jj}) > \|\bar{B}^{-1}\|\left[(2n\mu^0\xi^*)^{-1}(\xi^*\beta\|r^0\| + \mu^0\tau 2n\sqrt{2n}\|N\|)\right](\tilde{x}^k)^T\tilde{z}^k$, and

(4) $\min(\tilde{z}^k_{\mathcal{N}}) > (2n\mu^0\xi^*)^{-1}[\beta\xi^*\left(1 + \|N^T\|\|\bar{B}^{-T}\|\right)\|r^0\|$

$$
+ 4n\sqrt{2n}\tau\mu^0\|N^T\|\|\bar{B}^{-T}\|](\tilde{x}^k)^T\tilde{z}^k.
$$

Then $0 < x_{\mathcal{B}} = x_{\mathcal{B}}^k + Ddx_{\mathcal{B}} \le u_{\mathcal{B}}$, where $dx_{\mathcal{B}}$ is the solution of

$$(BD)dx_{\mathcal{B}} = b - Bx_{\mathcal{B}}^k, \tag{4.6}$$

and $0 \le s_{\mathcal{B}} < u_{\mathcal{B}}$. Moreover, $z_{\mathcal{N}}$ and $w_{\mathcal{N}}$ are nonnegative, where

$$(B^T Y^k)dy = c_{\mathcal{B}} - B^T y^k \tag{4.7}$$

and $y = y^k + Y^k dy,\ z_{\mathcal{N}} - w_{\mathcal{N}} = c_{\mathcal{N}} - N^T y$.

**Proof** The proof that $x_{\mathcal{B}}$ satisfies its lower and upper bounds is omitted because it is the same as the proof of Theorem 4.1 with $X_{\mathcal{B}}^k$ replaced by $D$.

If $c_{\mathcal{N}} - N^T y > 0$, then $z_{\mathcal{N}} > 0$ and $w_{\mathcal{N}} = 0$. The proof follows directly from Theorem 4.1, since the scale factor for the dual linear system does not change if upper bound constraints are present, since the dual variables $y$ are not explicitly dependent on the upper bound constraints.

If $c_{\mathcal{N}} - N^T y > 0$, then $z_{\mathcal{N}} = 0$ and $w_{\mathcal{N}} > 0$. Substitute $w_{\mathcal{N}}$ for $z_{\mathcal{N}}$ in the proof of Theorem 4.1. $\square$

We have shown that the scaled Mehrotra-Ye procedure produces a solution that satisfies strict complementarity. Now, we show that this solution can be generated in finite time when coupled with the partitioning strategy,

$$\mathcal{B}^k = \{j : x_j^k > z_j^k\}.$$

**Theorem 4.3** Let $(x^0, y^0, z^0) = (\rho e, 0, \rho e)$. Assume

(1) $\mathcal{B}^k = \mathcal{B}$.

(2) $x_{\mathcal{B}}^k > \frac{\gamma\xi^*}{\tau n}$ and $z_{\mathcal{N}}^k > \frac{\gamma\xi^*}{\tau n}$, where $\xi^*$ is as defined in (3.4).

(3) $\min(x_{\mathcal{B}}^k) > \|\bar{B}^{-1}\| \left[(\mu^0 n \xi^*)^{-1}(\xi^*\beta\|r^0\| + \mu^0 \tau n\sqrt{n}\|N\|)\right](x^k)^T z^k$.

(4) $\min(z_{\mathcal{N}}^k) > (n\mu^0\xi^*)^{-1}[\beta\xi^*\left(1 + \|N^T\|\|\bar{B}^{-T}\|\right)\|r^0\|$

$$+ n\sqrt{n}\tau\mu^0\|N^T\|\|\bar{B}^{-T}\|](x^k)^T z^k.$$

Then Algorithm 1 combined with the scaled Mehrotra-Ye procedure generates a solution in

$$O\left(n^2\left(\log n + \log\left(\frac{\tau(\max(\rho,\phi,\eta))}{\gamma\xi^*}\right)\right)\right) \text{ iterations,}$$

where $\tau$ is as defined in (3.11), $\phi = \|\bar{B}^{-1}\|\left(\xi^*\beta\|r^0\| + \mu^0\tau n\sqrt{n}\|N\|\right)$, and $\eta = \beta\xi^*\left(1 + \|N^T\|\|\bar{B}^{-T}\|\right)\|r^0\| + n\sqrt{n}\tau\mu^0\|N^T\|\|\bar{B}^{-T}\|$.

**Proof** The proof is similar to the proof given in Chapter 3 for the weighted projection model. □

## 4.2 Ordered Columns

Another modification to Mehrotra-Ye is to order the columns according to the magnitude of the vector $x_{\mathcal{B}}^k$. First, the components of $x_{\mathcal{B}}^k$ are sorted in nonincreasing order. We then order the columns of $B$ so that $B_{.1}$ corresponds to the largest element of the approximate solution vector and $B_{.k}$ corresponds to the *k-th* largest element. We want columns of $B$ that correspond to large components of the vector $x_{\mathcal{B}}^k$ to enter the basis first. Dependent columns are removed from the matrix when negligible pivots are encountered. The column ordering scheme has the theoretical advantage of not increasing the condition number of the constraint matrix.

In a related work when the linear programming problem has an unique primal degenerate solution, Asic, Kovacevic-Vujcic, and Radosavljevic-Nikolic [4] performed a similar column sort to transform the matrix $A$ into near upper triangular (trapezoidal) form. This transformation allows the matrix $A(X^k)^2A^T$ to be inverted by the inversion of well-conditioned matrices.

## 4.3 Numerical Results

We compare the effectiveness of three solution techniques in finding an interior point on the optimal face. The techniques are our implementation of Mehrotra-Ye, scaled Mehrotra-Ye, and ordered Mehrotra-Ye. The numerical experiments were conducted on a Sun workstation with 64 bit arithmetic. We use the LIPSOL - Linear programming Interior-Point SOLver- package developed under the MATLAB[‡] environment, see [57]. The tests were run on MATLAB version 4.2c. Because of the dense implementation of the Gaussian elimination routine, we were only able to solve problems where the matrix $B$ had approximately 500 rows and columns. Consequently, our test set consisted of 55 problems from the *netlib* suite. The three largest problems tested were *maros* with 835 rows and 1921 columns, *scsd8* with 397 rows and 2750 columns, and *ship08l* with 688 rows and 4339 columns. The removal of columns corresponding to zero variables combined with the elimination of zero rows reduced the original matrix $A$ to the desired dimensions.

The initial matrix $A$ is scaled in an attempt to achieve row and column equilibration. Preprocessing deletes fixed variables, zero rows, and columns from the matrix A, solves equations of only one variable, and shifts nonzero lower bounds to zero. We first attempt to compute an exact solution when

$$\left(\frac{|c^T x^k - b^T y^k|}{1 + |b^T y^k|}\right) \leq 10^{-8}. \tag{4.8}$$

The partition is determined by the Tapia indicators in tandem with variables as indicators. The interior-point algorithm was terminated after six attempts to find a feasible point on the optimal face.

The actual linear systems solved were

$$(BD)dx_{\mathcal{B}} = b - Bx^k_{\mathcal{B}} \text{ and } DB^T dy = D(c_{\mathcal{B}} - B^T y^k) \tag{4.9}$$

[‡]MATLAB is a registered trademark of The MathWorks, Inc.

where

$$d_{jj} = \begin{cases} \min(x_j^k, u_j - x_j^k) & \text{for } j \in \mathcal{B} \text{ and upper bounds exist} \\ x_j^k & \text{for } j \in \mathcal{B} \text{ else.} \end{cases}$$

The dual model is not what we proposed in Section 4.1. The solutions of the proposed model in Section 4.1 require two matrix factorizations. This is twice the computational expense of an interior-point iteration. However, with formulation (4.9) a single matrix factorization suffices to compute primal and dual points on the respective optimal faces.

For the column scaling approach, we factored the matrix $DB^T$ using Gaussian elimination with partial pivoting and row interchanges. The matrix, $DB^T$, was factored instead of its transpose to take advantage of the column scaling. Partial pivoting with row interchanges of $BD$ is equivalent to partial pivoting of $B$. In the column ordering approach, we order by the vector $d$ where $d = \text{diag}(D)$ if upper bounds are present.

If a negligible pivot was encountered, the column was removed from the matrix. The pivot tolerance was

$$max(m, |B|) * \|B\|_1 * 10^{-16}.$$

This is precisely the default tolerance used in MATLAB to determine the numerical rank of a matrix. We did not pivot to minimize fill-in of the triangular factors, $L$ and $U$. Zero rows were removed before the Gaussian elimination subroutine started. At the completion of the factorization any remaining zero rows were deleted. Components of the solution vector corresponding to dependent rows and columns were set to zero. Consequently, we computed a basic solution of the reduced linear system.

Table 4.2 shows the results of our numerical experiments. Column 1 gives the number of calls to the finite termination procedure. Columns 2 through 4 give the

number of problems solved by the three respective variants of the Mehrotra-Ye procedure. The results of Column 2 were obtained by computing an $LU-$factorization of $B^T$ and using those factors to solve systems (2.2). If we factor $B$ instead, the projection count changes slightly. In fact, additional projections are required to find exact solutions for some problems.

| | Techniques | | |
|---|---|---|---|
| # of misses | Mehrotra-Ye | Scaled | Ordered |
| 0 | 44 | 47 | 46 |
| 1 | 7 | 6 | 5 |
| 2 | 2 | 1 | 1 |
| 3 | 1 | 1 | 2 |
| 4 | 0 | 0 | 0 |
| 5 | 1 | 0 | 0 |
| more than five | 0 | 0 | 1 |
| TOTAL misses | 19 | 11 | 19 |

**Table 4.2** Misses per technique

In terms of the number of attempts needed to find a solution, the three solution techniques perform comparably. Scaling and ordering the columns of the matrix saved one interior-point iteration for problems, *boeing2, kb2,* and *seba.* Two interior-point iterations are saved for problems *etamacro, finnis* and *stair.* All six problems have upper bound constraints. When we implemented the standard Mehrotra-Ye approach, six attempts were needed to find feasible primal and dual points for problem *etamacro.*

However, the standard Mehrotra-Ye procedure generates the most accurate solutions. For 91 percent of the problems, the objective function value agrees to thirteen digits with the reoptimized CPLEX objective function value that was reported in [6]. The thirteen digit agreement is 89 percent for scaled and ordered Mehrotra-Ye.

Column scaling and ordering are equivalent for all but three problems - *agg3, grow15,* and *scrs8.* When the columns are scaled in problem *scrs8*, two attempts

are necessary to find a feasible point on the optimal face. The ordered model only needs one attempt. When column scaled the coefficient matrix of *scrs8* is highly ill-conditioned. This degeneracy may be reflected in the additional projection needed to obtain an exact solution. The rank of the matrix $B$ is 301; the rank of $BX_B^k$ is 169. The matrix has 305 rows and 317 columns. The difference in the maximum and minimum components of $x_B^k$ is nine orders of magnitude.

In the following, we attempt to provide a plausible explanation for the behavior of the ordered columns technique for problems *agg3* and *grow15*. According to Skeel, Gaussian elimination without scaling leads to numerical instability. Instability manifested itself in problems *agg3* and *grow15*. After one call to the finite termination procedure, solutions were computed for the other problems in the *grow* and *agg* classes. No solution was found for problem *agg3*.

| Problem | Standard | Scaled | Ordered |
|---|---|---|---|
| grow7 | 1 | 1 | 1 |
| grow15 | 1 | 1 | 4 |

**Table 4.3** Technique comparison for class GROW

| Problem | Standard | Scaled | Ordered |
|---|---|---|---|
| agg | 1 | 1 | 1 |
| agg2 | 1 | 1 | 1 |
| agg3 | 1 | 1 | - |

**Table 4.4** Technique comparison for class AGG

Problem *grow15* has staircase structure and upper bound constraints; problem *agg3* does not. *Grow15* is highly degenerate. The associated matrix $B$ has 300 rows and 573 columns. *Agg3* is only mildly degenerate. The full rank matrix, $B$, has

512 rows and 551 columns. The condition numbers of the matrices are 5.64e+00 for *grow15* and 2.35e+03 for *agg3*.

# Chapter 5

# The Role of Fast Local Convergence

Thus far, we have studied various models and techniques for generating a feasible point on the optimal primal and dual faces. Now, we turn our attention to another important question in finite termination procedures - when should we first attempt to generate an exact solution. Numerous issues impact the first attempt. Foremost, is the researcher's goal. Does he/she want to save computational expense or simply generate an exact solution post optimally? If the choice is the former, the researcher has to worry about indicator reliability when the iterate is far from the solution set and the tradeoffs between the costs of multiple projections and the cost of the standard interior-point iteration.

In [35], Mehrotra and Ye projected the current point onto the optimal faces when the relative gap is less than or equal to $10^{-8}$. In a subsequent numerical study, Mehrotra [30], solved for a point on the optimal faces when the current iterate satisfied

$$\max\left(\frac{\|Ax^k - b\|}{1+\|b\|}, \frac{\|A^Ty^k + z^k - c\|}{1+\|c\|}, \frac{|c^Tx^k - b^Ty^k|}{1+b^Ty^k}\right) \leq 10^{-8}, \tag{5.1}$$

which is more commonly known as the eight digits of relative precision criterion. Both criteria essentially initiate the finite termination procedure at the solution. For many interior-point algorithms (e.g., HOPDM [16], LIPSOL [57], PCx [7] , etc.), inequality (5.1) is the default stopping tolerance. Projecting when the iterates satisfy inequality (5.1) does not generally save computational expense. When Mehrotra [30] first observed superlinear convergence of the duality gap sequence $\{(x^k)^T z^k\}$ to zero, he attempted to find an feasible point on the optimal face of the linear program.

On the other hand, a more liberal strategy may not be advisable, especially if a significant number of projection attempts is needed. Recall that the cost of a projection attempt is equivalent to one interior-point iteration.

For many problems in our test set, the optimal face can be identified and verified before inequality (5.1) is satisfied. Consequently, we propose a more aggressive criterion to test the optimality of the partition. We extend Mehrotra's idea of using fast convergence of the duality gap sequence $\{(x^k)^T z^k\}$ as an indication of when to proceed from an approximate solution to an exact solution. We propose projecting from a full pure composite Newton step when the full Newton step gives $Q$-quadratic convergence of the duality gap sequence $\{(x^k)^T z^k\}$.

In Section 5.1, we present the composite Newton method. A general feasible primal-dual predictor-corrector interior-point method is described in Section 5.2. Convergence theory for the duality gap sequence $\{(x^k)^T z^k\}$ is the topic of Section 5.3. In Section 5.4, we describe a weighted projection model that uses the pure composite Newton step. Numerical results comprise the last section.

## 5.1 The Composite Newton Method

The composite Newton method is the modification of Newton's method where the step consists of the sum of a Newton step and a simplified Newton step. The previous Jacobian is used in the simplified Newton equation. Below, we describe the level-$m$ composite Newton method for the nonlinear system of equations, $F(x) = 0$.

$$\begin{array}{lll} \text{Solve} & F'(x_k)(\Delta x_i) = -F(x_k + \Delta x_0 + \Delta x_1 + \cdots + \Delta x_{i-1}) & \text{for } \Delta x_i, i = 0, \ldots, m \\ \text{set} & x_{k+1} = x_k + \alpha^k(\Delta x_0 + \Delta x_1 + \cdots + \Delta x_m), & k = 0, 1, \ldots \end{array} \tag{5.2}$$

Under the standard Newton assumptions the level-1 composite Newton method has a $Q$-cubic convergence rate.

## 5.2 Predictor-Corrector Algorithm

In this section, we describe a general primal-dual predictor-corrector interior-point method. The primal-dual predictor-corrector interior-point method was suggested by Mehrotra [31], and was shown to be equivalent to a perturbed and damped level-one composite Newton method by Tapia, Zhang, Saltzman, and Weiser [44].

Each iteration of this primal-dual method consists of a predictor step and a corrector step. The predictor step is a step towards optimality; the centering-corrector step goes toward the central path. Both the predictor and corrector steps satisfy a linear system of the form

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ Z & 0 & X \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ r \end{pmatrix}, \tag{5.1}$$

where $r = -Xz$ for the predictor step and $r = \sigma\mu e - \bar{X}\bar{z}$ for the corrector step, where $(\bar{x}, \bar{y}, \bar{z}) = (x + \Delta^p x^k, y + \Delta^p y^k, z + \Delta^p z^k)$. Since for any feasible point $(x, y, z)$, the first two blocks of linear equations in $F(x, y, z) = 0$, see (1.2), are always satisfied; it is easy to see that the matrix in the left-hand-side of (5.1) is $F'(x, y, z)$ and the right-hand-side is $-F(x, y, z)$ for the predictor step and $\sigma\mu e - F(\bar{x}, \bar{y}, \bar{z})$ for the corrector step.

**Algorithm 3** *(Predictor-Corrector Algorithm)*

Given a strictly feasible point $(x^0, y^0, z^0)$ for $k = 0, 1, 2, \ldots$, do

**(1)** Solve for the predictor step $(\Delta^p x^k, \Delta^p y^k, \Delta^p z^k)$ from (5.1) with

$$(x, y, z) = (x^k, y^k, z^k) \quad \text{and} \quad r = -X^k z^k.$$

**(2)** For $\sigma^k \in [0,1)$ and $\mu^k = ((x^k)^T z^k)/n$, solve for the corrector step $(\Delta^{cc}x^k, \Delta^{cc}y^k, \Delta^{cc}z^k)$ from (5.1) with

$$(\bar{x}, \bar{y}, \bar{z}) := (x^k + \Delta^p x^k, y^k + \Delta^p y^k, z^k + \Delta^p z^k), \quad r = \sigma^k \mu^k e - \bar{X}^k \bar{z}^k.$$

**(3)** Set $(\Delta x^k, \Delta y^k, \Delta z^k) := (\Delta^p x^k, \Delta^p y^k, \Delta^p z^k) + (\Delta^{cc}x^k, \Delta^{cc}y^k, \Delta^{cc}z^k)$.

**(4)** Choose $\tau^k \in (0,1)$ and set $\alpha^k = min(1, \tau^k \hat{\alpha}^k)$, where

$$\hat{\alpha}^k = \frac{-1}{\min((X^k)^{-1}\Delta x^k, (Z^k)^{-1}\Delta z^k)}.$$

**(5)** Let $(x^{k+1}, y^{k+1}, z^{k+1}) := (x^k, y^k, z^k) + \alpha^k(\Delta x^k, \Delta y^k, \Delta z^k)$.

**(6)** Test for convergence.

The dampening and perturbing of interior-point iterates preclude the standard proof of cubic convergence. Tapia, Zhang, Saltzman, Weiser [44] proved that for nondegenerate problems that cubic convergence can be restored locally (in a neighborhood of the solution) by taking the full pure composite Newton step, (i.e., no centering or dampening). We present Theorem 3.1 of Tapia, Zhang, Saltzman, and Weiser [44] which has been tailored to suit our purposes.

**Theorem 5.1** ( Tapia, Zhang, Saltzman, Weiser [44]) Consider the iteration sequence $\{x^k, y^k, z^k\}$ produced by Algorithm 2. Assume

(1) strict complementarity

(2) $x^*$ is a nondegenerate vertex

(3) $(x^k, y^k, z^k)$ converges to $(x^*, y^*, z^*)$.

If the choice of $\sigma^k$ satisfies

$$0 \leq \sigma^k \leq \min(\sigma, C((x^k)^T z^k)^2)$$

and for large $k$

$$\alpha^k = 1,$$

then the convergence is $Q$-cubic, i.e. there exists $\gamma_3 > 0$ such that for $k$ large

$$\|(x^{k+1}, y^{k+1}, z^{k+1}) - (x^*, y^*, z^*)\| \leq \gamma_3 \|(x^k, y^k, z^k) - (x^*, y^*, z^*)\|^3.$$

## 5.3 Fast Convergence of the Duality Gap

When the interior-point algorithm exhibits fast local convergence we attempt to generate an exact solution rather than waiting for an approximate solution to attain eight digits of relative precision.

Projecting early in the iterative process has its drawbacks. First, the variables may not have separated enough for correct identification of the active set. Second given the correct partition, the interior-point iterate may be far from the solution set. Therefore, a finite termination procedure could generate a solution that, while nonnegative, may fail to satisfy the prescribed linear feasibility and optimality tolerances. Incorrect identification of the active set dooms a finite termination procedure; infeasibility of the iterate does not.

Now let us consider fast local convergence of the duality gap sequence From step 5 of Algorithm 3, we have the following expression for the duality gap at iteration $k+1$,

$$\begin{aligned} {x^{k+1}}^T z^{k+1} &= (x^k + \alpha^k \Delta x^k)^T (z^k + \alpha^k \Delta z^k) \\ &= (1 - \alpha^k (1 - \sigma^k))\, {x^k}^T z^k \end{aligned} \tag{5.2}$$

The $Q_1$ factor for the sequence $\{{x^k}^T z^k\}$ is

$${x^{k+1}}^T z^{k+1} / {x^k}^T z^k = (1 - \alpha^k (1 - \sigma^k)).$$

To obtain superlinear convergence of the sequence $\{x^{k^T} z^k\}$, we need

$$(1 - \alpha^k(1 - \sigma^k)) \to 0 \text{ or } \alpha^k(1 - \sigma^k) \to 1,$$

which is achieved by letting $\alpha^k \to 1$ and $\sigma^k \to 0$.

The centering term $\sigma^k$ is an algorithmic parameter, but the step length $\alpha^k$ is not. However, the step length $\alpha^k$ is dependent upon the algorithmic parameter $\tau^k$. Hence, the theory for $Q$-superlinear convergence of the duality gap sequence $\{(x^k)^T z^k\}$ is stated in terms of $\sigma^k$ and $\tau^k$.

**Theorem 5.2** (Zhang, Tapia, Dennis [60]) Let $\{(x^k, z^k)\}$ be generated by Algorithm 1 and $(x^k, y^k, z^k) \to (x^*, y^*, z^*)$. Assume

**(i)** strict complementarity,

**(ii)** the sequence $(x^k)^T z^k / (n \min(X^k Z^k e))$ is bounded,

**(iii)** $\tau^k \to 1$ and $\sigma^k \to 0$.

Then the duality gap sequence $\{(x^k)^T z^k\}$ converges to zero $Q$-superlinearly. That is, the $Q_1$-factor

$$Q_1 = \lim_{k \to \infty} \sup \frac{x^{k+1^T} z^{k+1}}{x^{k^T} z^k} = 0.$$

Zhang and Tapia [59] improved Theorem 5.2 by replacing the convergence of the iteration sequence $\{(x^k, y^k, z^k)\}$ assumption with the assumption that the duality gap sequence $\{(x^k)^T z^k\}$ converges to zero.

**Lemma 5.1** (Zhang, Tapia, Potra [61]) Under the assumptions of Theorem 5.2,

$$\lim_{k \to \infty} \alpha^k = 1.$$

Mehrotra [30] used observed fast local convergence as an indication to project onto the optimal primal and dual faces. The optimal partition is identified at the $k-th$ iteration if the following conditions hold

$$\begin{aligned} {x^{k+1}}^T z^{k+1} / {x^k}^T z^k &\leq .01 \\ \alpha^k &\geq .95. \end{aligned} \tag{5.3}$$

Mehrotra concluded that superlinear convergence is detected after the optimal face and a point on it can be identified.

We propose a more aggressive criterion to test the optimality of the partition, which is based on fast local convergence of the duality gap sequence $\{(x^k)^T z^k\}$ which only consists of the predictor step. Under standard assumptions, Newton's method is known to give $Q$-quadratic convergence. We try to exploit this fact by forming the full Newton iterate, $\bar{x} = x^k + \Delta^p x^k$. Recall that the predictor step is the Newton step for problem (1.2). This particular update formula may violate the nonnegativity constraints. We then use the intermediate duality gap, $\bar{x}^T \bar{z}$, which only consists of the full Newton step(predictor step) as a indicator for estimating the optimal partition.

Specifically, when we observe $Q$-quadratic convergence

$$\bar{x}^T \bar{z} / {x^k}^T z^k \leq C_4({x^k}^T z^k), \quad \text{for some constant } C_4 > 0 \tag{5.4}$$

we could estimate the optimal partition. However, in practice, it is difficult to determine a value for the constant $C_4$.

## 5.4 Projection from a Pure Composite Newton Step

All models in the literature project from a strict interior point. We investigate the idea of projecting from a full pure composite Newton step when the full Newton step gives $Q$-quadratic convergence of the duality gap sequence $\{(x^k)^T z^k\}$. We denote the

full pure composite Newton iterate as $\bar{x}$, where

$$\bar{x} = x^k + \Delta^p x^k + \Delta^{cc} x^k$$

and the centering step, $\Delta^{cc} x^k$, is computed in step 3 of Algorithm 2 with $\sigma^k = 0$. We know that quadratic convergence of the duality gap sequence $\{(x^k)^T z^k\}$ to zero in the full Newton step implies cubic convergence of the duality gap to zero in the pure composite Newton step.

In an earlier numerical experiment, we computed a full pure composite Newton iterate in the neighborhood of the solution. We observed fast local convergence of the duality gap sequence $\{(x^k)^T z^k\}$ to zero, but overall convergence was stymied by an increase in constraint infeasibility when we set predicted zero variables to zero. Note we used an infeasible primal-dual predictor-corrector interior-point algorithm to perform the previously mentioned tests. Therefore, we combine the full pure composite Newton iterate with the finite termination procedure to recover linear feasibility. If we a finite termination procedure fails to generate a solution that satisfies the feasibility and optimality tolerances, we discard the full pure composite Newton iterate and proceed with the predictor-corrector algorithm by forming the damped Newton iterate.

The new weighted projection model can be written as follows

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|\bar{D}^{-1}(x_{\mathcal{B}} - \bar{x}_{\mathcal{B}})\|^2 \\ \text{subject to} & Bx_{\mathcal{B}} = b, \end{array} \tag{5.5}$$

where $\bar{x} = x^k + \Delta^p x^k + \Delta^{cc} x^k$ and $\bar{d}_{jj} = diag(\min(\bar{x}_j, u_j - \bar{x}_j))$ for $j \in \mathcal{B}$.

The dual model is

$$\min_y \frac{1}{2}\|\bar{D}^{-1}(B^T y - c_{\mathcal{B}})\|^2. \tag{5.6}$$

It is important to note that identification of the optimal partition does not require the formation of the full pure composite Newton iterate. We can combine inequality (5.4) with the projection models and techniques discussed in Chapters 3 and 4.

## 5.5 Numerical Results

The numerical experiments were conducted on a Sun workstation with 64 bit arithmetic. We used the LIPSOL - Linear programming Interior-Point SOLver- package. In this numerical study, we tested 80 problems from the *netlib* set. Larger problems such as *fit2d, marosr7, standgub, pilots, etc* were not included in this particular experiment due to the prohibitive cost of an interior-point iteration. We used models (5.5) and (5.6) to find feasible points on the optimal faces. The experiment is outlined below.

The study is similar to the one proposed by Mehrotra in [30].

**Procedure 3** *(Projecting From Pure Composite Newton Step)*

At every iteration of Algorithm 3,

**(1)** Set $(\bar{x}, \bar{y}, \bar{z}) := (x^k, y^k, z^k) + (\Delta^p x^k, \Delta^p y^k, \Delta^p z^k.)$

**(2)** If

$$\Big(\|r^k\| + (x^k)^T z^k \leq 1 \text{ and } \|\bar{r}\| + \bar{x}^T \bar{z} \leq (\|r^k\| + (x^k)^T z^k)^2\Big)$$
$$\text{or } \frac{|c^T x^k - b^T y^k|}{1 + |b^T y^k|} \leq 10^{-8}$$

continue. Otherwise, return to interior-point algorithm. The duality gap of the pure Newton step is calculated as $\bar{x}^T \bar{z} := |\bar{x}|^T |\bar{z}|$. The absolute value of the iterates must be taken since the unit step does not ensure positive iterates. Recall that

$$\|r^k\| = \|Ax^k - b, A^T y^k + z^k - c\|.$$

**(3)** Solve for the pure corrector step $(\Delta^c x^k, \Delta^c y^k, \Delta^c z^k)$ with right-hand side $= -\bar{X}\bar{z}$.

**(4)** Set $(\bar{x}, \bar{y}, \bar{z}) := (\bar{x}, \bar{y}, \bar{z}) + (\Delta^c x^k, \Delta^c y^k, \Delta^c z^k)$.

**(5)** Partition the variables into the sets $\mathcal{B}$ and $\mathcal{N}$.

**(6)** Check the negativity of $(\bar{x}, \bar{z})$. If $|\bar{x_j}| \leq x_j^k$ for $j \in \mathcal{N}$, then proceed with the finite termination procedure. Otherwise, return control to the interior-point algorithm. Perform the same test for $\bar{z}_j$, j $\in \mathcal{B}$.

**(7)** Find the closet points to $(\bar{x_{\mathcal{B}}}, \bar{y}, \bar{z_{\mathcal{N}}})$ that lie on the optimal primal and dual faces.

**(8)** Test for positivity of $(x_{\mathcal{B}}, z_{\mathcal{N}})$ and convergence.

**(9)** If the convergence criteria is not satisfied, return control to the standard interior-point algorithm. Else, terminate the algorithm with a solution.

In step (2), we delay the procedure until

$$\|r^k\| + (x^k)^T z^k \leq 1. \tag{5.7}$$

If the $k$-th iterate is feasible, the value of the duality gap dominates (5.7). Given $(x^k)^T z^k > 1$ and the unit step, $\bar{x}^T \bar{z} < (x^k)^T z^k < ((x^k)^T z^k)^2$. Unfortunately, the procedure would be activated at each iteration. Empirical evidence shows that inequality (5.7) criteria can be relaxed for some problems.

As stated, the procedure requires three back substitutions per iteration instead of two. Thus $O(n^2)$ additional work is needed per iteration. To save computational expense, the pure corrector step can be replaced with the centering-corrector step when $\sigma^k = \min\left(\sigma, C_5((x^k)^T z^k)^2\right)$ where $\sigma \in (0, 1)$. From the statement of Theorem 5.1, we can expect $Q$-cubic convergence for nondegenerate problems.

Step six tests for the degree of negativity of the vector $(\bar{x}, \bar{y}, \bar{z})$. Since we take an undamped step, the iterate may have negative components. For predicted nonzero variables which are negative, we reflect through the origin. That is,

$$\begin{aligned} \bar{x_j} &:= |\bar{x_j}| \quad j \in \mathcal{B} \\ \bar{z_j} &:= |\bar{z_j}| \quad j \in \mathcal{N}. \end{aligned} \tag{5.8}$$

which increases the residual

$$\|B|\bar{x_{\mathcal{B}}}| - b\|.$$

The projection procedure is asked to compensate for the additional infeasibility.

One can argue that the success of this procedure is directly attributed to the fact that $(\bar{x}, \bar{y}, \bar{z})$ approximates $(x^{k+1}, y^{k+1}, z^{k+1})$. This is true - the undamped pure composite Newton update does approximate the $k+1$st iterate. However, we are using asynchronous indicator information. The partition generated in the $k+1$-st iteration is based on

$$\frac{x_j^{k+2}}{x_j^{k+1}} \text{ not } \frac{x_j^{k+1}}{x_j^k}.$$

Unless the indicators have converged by the $k$-th iteration, the data is inconsistent.

Tables 5.1 and 5.2 show the decrease of the duality gap when the undamped pure composite Newton step is taken. The first two columns report the primal and dual infeasibilities, respectively. The third column gives the value of the absolute duality gap. We see cubic decrease in the duality gap, but the infeasibility does not decrease at the same rate. It is important to note that the problems listed are degenerate.

| Step | $\|Ax-b\|$ | $\|A^Ty+z-c\|$ | $x^Tz$ |
|---|---|---|---|
| $k-$th iterate | 6.23e-12 | 1.02e-15 | 3.42e-02 |
| Full Newton | 1.81e-12 | 1.21e-15 | 1.09e-06 |
| Composite Newton | 1.23e-13 | 1.88e-15 | 1.13e-10 |

**Table 5.1** Problem AFIRO, m= 27, n = 51

| Step | $\|Ax-b\|$ | $\|A^Ty+z-c\|$ | $x^Tz$ |
|---|---|---|---|
| $k-$th iterate | 3.19e-05 | 3.34e-12 | 1.36e-01 |
| Full Newton | 2.05e-11 | 3.78e-12 | 8.03e-04 |
| Composite Newton | 5.01e-13 | 3.26e-12 | 5.53e-05 |

**Table 5.2** Problem DEGEN2, m = 444, n = 757

Tables 5.3 and 5.4 illustrate the ability to terminate far from the solution set. The second column reports the absolute primal infeasibility associated with the indicated step. The third column gives the value of the absolute dual infeasibility. The last column records the absolute duality gap. We projected from the pure composite Newton step rather than the $k-$th iteration.

| Step | $\|Ax-b\|$ | $\|A^Ty+z-c\|$ | $x^Tz$ |
|---|---|---|---|
| $k-$th iterate | 3.19e-05 | 3.34e-12 | 1.36e-01 |
| Composite Newton | 5.01e-13 | 3.26e-12 | 5.53e-05 |
| Finite Termination | 8.52e-15 | 1.66e-12 | 9.74e-12 |

**Table 5.3** Problem DEGEN2, Relative Gap = 7.46e-05

Observe that when the finite termination procedure is initiated there are only three digits of accuracy in the solution of problem *sc105*.

| Step | $\|Ax-b\|$ | $\|A^Ty+z-c\|$ | $x^Tz$ |
|---|---|---|---|
| $k-$th iterate | 3.22e-08 | 7.28e-12 | 2.80e-01 |
| Composite Newton | 1.11e-11 | 2.70e-16 | 5.88e-04 |
| Finite Termination | 4.15e-13 | 1.32e-16 | 5.05e-15 |

**Table 5.4** Problem SC105, Relative Gap = 5.20e-03

Unfortunately, projecting from the pure composite Newton step only saves six percent (87/1547) of the total iterations. We averaged 1.92 projection attempts per

problem, while only saving 1.24 iterations. Thus it appears that our projection criterion is more time-consuming. However, we did save five iterations on problem *degen3*, which is highly degenerate.

Projecting from the pure composite Newton step generates highly accurate solutions. For 95 percent of the (76 out of 80) problems, our objective function value agrees to thirteen digits with the objective function value reported in [6]. Moreover, we gained two digits of accuracy for problem *gfrdpnc* when compared to the experiments conducted in the previous chapters.

# Chapter 6

# Indicators

The term indicator denotes a function that identifies constraints that are active at the solution of a constrained optimization problem, see Tapia [43] and El-Bakry [9]. Indicators play an important role in finite termination procedures. After deciding when to first attempt to compute an exact solution, we must estimate which variables are active at the solution

Commonly used indicators include variables as indicators, the primal-dual indicator, and the Tapia indicator. See El-Bakry [9] and El-Bakry, Gonzalez-Lima, Tapia, and Zhang [11] for a thorough study of indicators. El-Bakry, Tapia, and Zhang [10] showed that identification of the set of variables that are zero at the solution can lead to reduction of problem size and computational savings.

The primal-dual predictor-corrector interior-point algorithm is the workhorse for most numerically efficient interior-point implementations. The increased practical efficiency of predictor-corrector methods can be directly attributed to the generation of the corrector step. Lustig, Marsten, and Shanno [23] provided empirical evidence that in terms of the iteration count, the predictor-corrector method outperformed the primal-dual method when tested on the *netlib* suite of linear programming problems. However, indicators in the literature do not incorporate the corrector step.

To take advantage of all available algorithmic information, we extend the definition of the Tapia indicators to the predictor-corrector algorithmic framework. The centering-corrector step as well as the predictor step comprise this new indicator. In this chapter, we investigate its theoretical properties. We also examine the numerical effectiveness of this indicator in identifying the optimal partition.

## 6.1 The Tapia Indicators

Tapia [43] used the following indicators to determine the active set in nonlinear constrained optimization problems. The Tapia indicators are

$$T_p(x_j^k) = \frac{x_j^{k+1}}{x_j^k} \qquad \text{and} \qquad T_d(z_j^k) = \frac{z_j^{k+1}}{z_j^k},$$

where $x_j^{k+1} = x_j^k + \Delta^p x_j^k$ and $z_j^{k+1} = z_j^k + \Delta^p z_j^k$. In [10] El-Bakry, Tapia, and Zhang showed the Tapia indicators have a 0-1 separation property and converge $R - superlinearly$ to their limiting values.

In the context of interior-point methods, Mehrotra [28] suggested the use of the relative change of variables as indicators, which are a simple restatement of the Tapia indicators. The relative change in variables as indicators are

$$R_p(x_j^k) = \frac{\Delta^p x_j^k}{x_j^k} = T_p(x_j^k) - 1 \qquad \text{and} \qquad R_d(z_j^k) = \frac{\Delta^p z_j^k}{z_j^k} = T_d(z_j^k) - 1.$$

Mehrotra and Ye in [35], Mehrotra in [29] and [30], as well as Andersen and Ye [2] used the Tapia indicators to identify the optimal partition. Specifically, they defined

$$\mathcal{B}^k = \{j : \frac{|\Delta^p x_j^k|}{x_j^k} \leq \frac{|\Delta^p z_j^k|}{z_j^k}.\} \tag{6.1}$$

Mehrotra and Ye [35] proved that when $\mathcal{B}^k$ was defined as in (6.1) the optimal partition could be identified in finite time for algorithms that generate iteration sequences that satisfy centrality measure (3.1).

## 6.2 The Tapia Predictor-Corrector Indicators

We define the Tapia predictor-corrector indicators as

$$PC_p(x_j^k) = \frac{x_j^{k+1}}{x_j^k} \qquad \text{and} \qquad PC_d(z_j^k) = \frac{z_j^{k+1}}{z_j^k},$$

where $x_j^{k+1} = x_j^k + \Delta^p x_j^k + \Delta^{cc} x_j^k$, $z_j^{k+1} = z_j^k + \Delta^p z_j^k + \Delta^{cc} z_j^k$, and $\Delta^p x_j^k$ and $\Delta^{cc} x_j^k$ are computed as in steps 1 and 2 of Algorithm 3, respectively. The Tapia predictor-corrector indicators differ from the Tapia indicators in the addition of the centering-corrector step.

In the following proposition, we provide an expression for the sum of the primal and dual Tapia predictor-corrector indicators. In subsequent results, we show, with appropriate choices for the centering and step length parameters, the limit of the right-hand side is one.

**Proposition 6.1** Assume that the sequence of iterates $\{(x^k, y^k, z^k)\}$ has been generated by Algorithm 3. Then

$$(X^k)^{-1} x^{k+1} + (Z^k)^{-1} z^{k+1} =$$

$$(2 - \alpha^k)e - \alpha^k \Delta^p X^k \Delta^p Z^k (X^k Z^k)^{-1} e + \alpha^k \sigma^k \mu^k (X^k Z^k)^{-1} e, \qquad (6.2)$$

where $\Delta^p X^k = \mathrm{diag}(\Delta^p x^k)$ and $\Delta^p Z^k = \mathrm{diag}(\Delta^p z^k)$.

**Proof** Consider the linearized complementarity equation for the predictor step

$$X^k \Delta^p z^k + Z^k \Delta^p x^k = -X^k Z^k e$$

and the linearized complementarity equation for the centering-corrector step

$$X^k \Delta^{cc} z^k + Z^k \Delta^{cc} x^k = \sigma^k \mu^k e - \Delta^p X^k \Delta^p Z^k e.$$

Combining the two complementarity equations yields

$$X^k(\Delta^p z^k + \Delta^{cc} z^k) + Z^k(\Delta^p x^k + \Delta^{cc} x^k) = -X^k Z^k e - \Delta^p X^k \Delta^p Z^k e + \sigma^k \mu^k e.$$

Thus,

$$X^k(z^{k+1}) + Z^k(x^{k+1}) = (2 - \alpha^k) X^k Z^k e - \alpha^k \Delta^p X^k \Delta^p Z^k e + \alpha^k \sigma^k \mu^k e.$$

Multiply both sides by $(X^k Z^k)^{-1}$ and this completes the proof. $\square$

The following lemma shows that the $(x, z)$ components of the iteration sequence generated by a path following algorithm are bounded. The lower limit on the predicted nonzero variables permit us to bound the ratio of the steps direction to the current iterate.

**Lemma 6.1** (Güler and Ye, [17]) Let $\mu^0 > 0$ and $\gamma \in (0,1)$. Then for all points $\{(x^k, y^k, z^k)\}$ with

$$\{(x^k, y^k, z^k)\} \in \mathcal{N}_{-\infty}(\gamma) \subset \mathcal{F}^0, \quad \mu^k \le \mu^0$$

there exists a constant $C_0$ such that $\|(x, z)\| \le C_0$ and

$$\begin{array}{ll} 0 < x_j^k \le ((x^k)^T z^k)/\xi^* \quad (j \in \mathcal{N}) & \qquad 0 < z_j^k \le ((x^k)^T z^k)/\xi^* \quad (j \in \mathcal{B}) \\ \quad x_j^k \ge (\xi^* \gamma)/n \quad (j \in \mathcal{B}) & \qquad \quad z_j^k \ge (\xi^* \gamma)/n \quad (j \in \mathcal{N}). \end{array} \tag{6.3}$$

The next proposition is used in the separation proof for the Tapia predictor-corrector indicator. Proposition 6.2 enables us to prove that the Tapia predictor-corrector indicator retains the 0-1 separation property of the Tapia indicator. Moreover, it shows, under certain conditions, that in the limit the Tapia predictor-corrector indicator equals the Tapia indicator. To the best of our knowledge, it is not in the literature.

**Proposition 6.2** Consider a sequence of iterates $\{(x^k, y^k, z^k)\}$ generated by Algorithm 3. Assume that

1. $(x^k)^T z^k \to 0$
2. $\min(X^k Z^k e) \ge \gamma \mu^k$, for all $k$ and some $\gamma \in (0,1)$
3. The algorithmic parameters are chosen such that

$$\sigma^k \to 0 \text{ and } \tau^k \to 1.$$

Then for $j = 1, \ldots, n$

$$\lim_{k\to\infty} \frac{\Delta^{cc} x_j^k}{x_j^k} = 0 \qquad \text{and} \qquad \lim_{k\to\infty} \frac{\Delta^{cc} z_j^k}{z_j^k} = 0.$$

**Proof** Consider the linearized complementarity equation for the centering-correction direction,

$$X^k \Delta^{cc} z^k + Z^k \Delta^{cc} x^k = \sigma^k \mu^k e - \Delta^p X^k \Delta^p Z^k e.$$

Divide both sides by $(X^k Z^k)$. Component-wise, we have

$$\frac{\Delta^{cc} x_j^k}{x_j^k} + \frac{\Delta^{cc} z_j^k}{z_j^k} = \sigma^k \Big( \frac{(x^k)^T z^k / n}{x_j^k z_j^k} \Big) - \frac{\Delta^p x_j^k \Delta^p z_j^k}{x_j^k z_j^k}. \tag{6.4}$$

Let's consider the last term on the right-hand side. From Theorem 3.1 Ye, Güler, Tapia, and Zhang [53], Mehrotra [32], and Theorem 7.4 of Wright [49], we know $\|\Delta^p x^k, \Delta^p z^k\| = O(\mu^k)$. Combining $\|\Delta^p x^k, \Delta^p z^k\| = O(\mu^k)$ with (6.3), we obtain

$$\frac{|\Delta^p x^k|}{x_j^k} = O(\mu^k), \quad j \in \mathcal{B} \qquad \text{and} \qquad \frac{|\Delta^p z^k|}{z_j^k} = O(\mu^k), \quad j \in \mathcal{N}. \tag{6.5}$$

If we look at the linearized complementarity equation for the predictor step, we see

$$\frac{\Delta^p x_j^k}{x_j^k} + \frac{\Delta^p z_j^k}{z_j^k} = -1. \tag{6.6}$$

Equations (6.5) and(6.6) imply

$$\begin{array}{lclcl} \lim_{k\to\infty} \frac{\Delta^p x_j^k}{x_j^k} = \ 0 & \quad \text{and} \quad & \lim_{k\to\infty} \frac{\Delta^p z_j^k}{z_j^k} = -1, & j \in \mathcal{B} \\ \lim_{k\to\infty} \frac{\Delta^p x_j^k}{x_j^k} = -1 & \quad \text{and} \quad & \lim_{k\to\infty} \frac{\Delta^p z_j^k}{z_j^k} = \ 0, & j \in \mathcal{N}. \end{array} \tag{6.7}$$

From the observations made above and the assumptions of the proposition, Equation 6.4 takes the form

$$\lim_{k\to\infty} \Big( \frac{\Delta^{cc} x_j^k}{x_j^k} + \frac{\Delta^{cc} z_j^k}{z_j^k} \Big) = 0 \ \ j = 1, \ldots, n. \tag{6.8}$$

Given that $\|\Delta^{cc} x^k, \Delta^{cc} z^k\| = O(\mu^k + \sigma^k)$, see Wright [48], we have

$$\frac{|\Delta^{cc} x^k|}{x_j^k} = O(\mu^k + \sigma^k), \quad j \in \mathcal{B} \qquad \text{and} \qquad \frac{|\Delta^{cc} z^k|}{z_j^k} = O(\mu^k + \sigma^k), \quad j \in \mathcal{N}.$$

From (6.8) this implies

$$\lim_{k\to\infty} \frac{\Delta^{cc} x_j^k}{x_j^k} = 0 \qquad \text{and} \qquad \lim_{k\to\infty} \frac{\Delta^{cc} z_j^k}{z_j^k} = 0, \quad j \in \mathcal{B}$$

and a similar relation holds for $j \in \mathcal{N}$. $\square$

Now we formally present the limiting values of the Tapia predictor-corrector indicator. As a direct consequence of Proposition 6.2, we show that the 0-1 separation property of the Tapia indicator is retained.

**Proposition 6.3** Consider a sequence of iterates $\{(x^k, y^k, z^k)\}$ generated by Algorithm 3. Assume

1. $(x^k)^T z^k \to 0$
2. $\min(X^k Z^k e) \geq \gamma \mu^k$, for all $k$ and for some $\gamma \in (0,1)$
3. The algorithmic parameters are chosen such that

$$\sigma^k \to 0 \text{ and } \tau^k \to 1.$$

Then for $j = 1, \ldots, n$

$$\lim_{k\to\infty} \frac{x_j^{k+1}}{x_j^k} = \begin{cases} 0 & j \in \mathcal{N} \\ 1 & j \in \mathcal{B} \end{cases}$$

$$\lim_{k\to\infty} \left(1 - \frac{z_j^{k+1}}{z_j^k}\right) = \begin{cases} 0 & j \in \mathcal{N} \\ 1 & j \in \mathcal{B} \end{cases}$$

where $x^{k+1} = x^k + \beta^k(\Delta^p x^k + \Delta^{cc} x^k)$ and $z^{k+1} = z^k + \beta^k(\Delta^p z^k + \Delta^{cc} z^k)$ for any $\beta^k \in [\alpha^k, 1]$ with $\alpha^k$ given in step 4 of the Algorithm 3.

**Proof** Consider

$$X^k(\Delta^p z^k + \Delta^{cc} z^k) + Z^k(\Delta^p x^k + \Delta^{cc} x^k) = -X^k Z^k e - \Delta^p X^k \Delta^p Z^k e + \sigma^k \mu^k e.$$

Clearly,

$$\begin{gathered}X^k(z^k+\beta^k(\Delta^p z^k+\Delta^{cc}z^k))+Z^k(x^k+\beta^k(\Delta^p x^k+\Delta^{cc}x^k))=\\(2-\beta^k)X^kZ^ke-\beta^k\Delta^pX^k\Delta^pZ^ke+\beta^k\sigma^k\mu^ke.\end{gathered}\tag{6.9}$$

Divide both sides by $(X^kZ^k)$. Then

$$\begin{gathered}(X^k)^{-1}x^{k+1}+(Z^k)^{-1}z^{k+1}=\\(2-\beta^k)e-\beta^k\Delta^pX^k\Delta^pZ^k(X^kZ^k)^{-1}e+\beta^k\sigma^k\tfrac{(x^k)^Tz^k}{n}(X^kZ^k)^{-1}e.\end{gathered}\tag{6.10}$$

From Proposition 6.2,

$$\lim_{k\to\infty}\frac{\Delta^{cc}x_j^k}{x_j^k}=0\quad\text{for all } j\qquad\text{and}\qquad\lim_{k\to\infty}\frac{\Delta^{cc}z_j^k}{z_j^k}=0\quad\text{for all } j.$$

Thus,

$$\begin{array}{lcl}\lim_{k\to\infty}\frac{\Delta^px_j^k+\Delta^{cc}x_j^k}{x_j^k}=0 & \text{and} & \lim_{k\to\infty}\frac{\Delta^pz_j^k+\Delta^{cc}z_j^k}{z_j^k}=-1,\quad j\in\mathcal{B}\\ \lim_{k\to\infty}\frac{\Delta^px_j^k+\Delta^{cc}x_j^k}{x_j^k}=-1 & \text{and} & \lim_{k\to\infty}\frac{\Delta^pz_j^k+\Delta^{cc}z_j^k}{z_j^k}=0,\quad j\in\mathcal{N}\end{array}\tag{6.11}$$

From (6.11) and the definition of $\hat{\alpha}^k$, it follows that $\hat{\alpha}^k\to 1$. Since $\tau^k\to 1$, it follows that $\alpha^k\to 1$ and therefore $\beta^k\to 1$. The result follows from (6.10) and assumptions 2 and 3. $\square$

We now show that the Tapia predictor-corrector indicators converge *R-superlinearly* to their limits. This result is somewhat disappointing; we conjectured that the addition of the centering-corrector step would produce faster local convergence of the indicators for variables in the active set than what has been proven for the Tapia indicators.

**Proposition 6.4** Consider a sequence of iterates $\{(x^k,y^k,z^k)\}$ generated by Algorithm 3. Assume that

1. $(x^k)^Tz^k\to 0$

2. $\min(X^k Z^k e) \geq \gamma \mu^k$ for all $k$ and some $\gamma \in (0,1)$

3. The algorithmic parameters $\sigma^k$ and $\tau^k$ have been chosen so that

$$\sigma^k = O(((x^k)^T z^k)^\lambda) \text{ and } \tau^k = 1 - O(((x^k)^T z^k)^\lambda) \quad \lambda \in (0,1).$$

Then for $j = 1, \ldots, n$

$$\lim_{k\to\infty} \frac{x_j^{k+1}}{x_j^k} = \begin{cases} 0 & j \in \mathcal{N} \\ 1 & j \in \mathcal{B} \end{cases}$$

$$\lim_{k\to\infty} \Big(1 - \frac{z_j^{k+1}}{z_j^k}\Big) = \begin{cases} 0 & j \in \mathcal{N} \\ 1 & j \in \mathcal{B} \end{cases}$$

with an R-rate of convergence $1+\lambda$, where $x^{k+1} = x^k + \beta^k(\Delta^p x^k + \Delta^{cc} x^k)$ and $z^{k+1} = z^k + \beta^k(\Delta^p z^k + \Delta^{cc} z^k)$ for any $\beta^k \in [\alpha^k, 1]$ with $\alpha^k$ given in step 4 of the algorithm.

**Proof** For $j \in \mathcal{B}$,

$$|PC(x_j^k) - 1| \leq \beta^k |\frac{\Delta x_j^k}{x_j^k}| \leq \beta^k \Big(|\frac{\Delta^p x_j^k}{x_j^k}| + |\frac{\Delta^{cc} x_j^k}{x_j^k}|\Big).$$

From Theorem 3.1 of Ye, Güler, Tapia, and Zhang [53], Mehrotra [32], Theorem 7.4 of Wright [49], and Lemma 6.1 we know

$$\frac{|\Delta^p x_j^k|}{x_j^k} = O(\mu^k).$$

From the analysis in Wright [48] of steps generated by Mehrotra's predictor-corrector algorithm and Lemma 6.1,

$$\frac{|\Delta^{cc} x_j^k|}{x_j^k} = O(\mu^k + \sigma^k).$$

Therefore,

$$|PC(x_j^k) - 1| \leq O(\mu^k + \sigma^k).$$

As a result of Proposition 6.1,

$$PC(x_j^k) = (2-\beta^k)e - \beta^k \Delta^p X^k \Delta^p Z^k (X^k Z^k)^{-1} e + \beta^k \sigma^k \mu^k (X^k Z^k)^{-1} e - PC(z_j^k)$$

which implies for $j \in \mathcal{N}$,

$$|PC(x_j^k)| \leq \sigma^k \mu^k (XZ)^{-1} e + |PC(z_j^k) - 1| = O(\mu^k + \sigma^k).$$

From assumptions 2 and 3 the indicators are bounded above by a sequence that tends to zero with $Q-$rate $1+\lambda$.

By symmetry, we can prove a similar result for the dual slack indicator. This completes the proof. $\square$

We now present a surprising result regarding the relationship of the scaled ratio of the predictor step for nonzero primal variables to the scaled ratio of the centering-corrector step for zero dual slack variables.

**Theorem 6.1** Consider a sequence of iterates $\{(x^k, y^k, z^k)\}$ generated by Algorithm 3. Assume

1. $(x^k)^T z^k \to 0$
2. $\min(X^k Z^k e) \geq \gamma \mu^k, \quad$ for all $k$ and for some $\gamma \in (0,1)$
3. The algorithmic parameters are chosen such that

$$\sigma^k \to 0 \text{ and } \tau^k \to 1.$$

Then

$$\lim_{k\to\infty} \frac{\Delta^p z_j^k}{z_j^k} = \lim_{k\to\infty} \frac{\Delta^{cc} x_j^k}{x_j^k}, \quad j \in \mathcal{N}$$

and

$$\lim_{k\to\infty} \frac{\Delta^p x_j^k}{x_j^k} = \lim_{k\to\infty} \frac{\Delta^{cc} z_j^k}{z_j^k}, \quad j \in \mathcal{B}.$$

**Proof** The predictor step for the dual slack is $\Delta^p z^k = -A^T(A(Z^k)^{-1}X^kA^T)^{-1}AXe$. The corrector step for the primal variable is

$$\begin{aligned}\Delta^{cc}x^k &= -X^k(Z^k)^{-1}A^T(A(Z^k)^{-1}X^kA^T)^{-1}A(Z^k)^{-1}(\sigma^k\mu^k e - \Delta^p X^k \Delta^p Z^k e) \\ &\quad + (Z^k)^{-1}(\sigma^k\mu^k e - \Delta^p X^k \Delta^p Z^k e).\end{aligned}$$

Clearly, the scaled dual predictor step has the form

$$(Z^k)^{-1}\Delta^p z^k = -(Z^k)^{-1}A^T(A(Z^k)^{-1}X^kA^T)^{-1}AXe$$

and the scaled primal corrector step can be written as

$$\begin{aligned}(X^k)^{-1}\Delta^{cc}x^k &= -(Z^k)^{-1}A^T(A(Z^k)^{-1}X^kA^T)^{-1}A(Z^k)^{-1}(\sigma^k\mu^k e - \Delta^p X^k \Delta^p Z^k e) \\ &\quad + (X^kZ^k)^{-1}(\sigma^k\mu^k e - \Delta^p X^k \Delta^p Z^k e).\end{aligned}$$

Let's consider the first term of $(X^k)^{-1}\Delta^{cc}x^k$. After distributing terms and substituting for the scaled dual predictor step, we obtain

$$\begin{aligned}(X^k)^{-1}\Delta^{cc}x^k &= -\sigma^k\mu^k(Z^k)^{-1}A^T(A(Z^k)^{-1}X^kA^T)^{-1}A(Z^k)^{-1}e \\ &\quad + \sigma^k\mu^k(X^kZ^k)^{-1}e + (X^kZ^k)^{-1}(\sigma^k\mu^k e - \Delta^p X^k \Delta^p Z^k e).\end{aligned}$$

Let $A_j$ denote the $j-th$ column of the matrix $A$. Component-wise, we have

$$\lim_{k\to\infty}\frac{\Delta^{cc}x_j^k}{x_j^k} = \lim_{k\to\infty}\frac{\sigma^k\mu^k A_j^T(A_j(z_j^k)^{-1}x_j^kA_j^T)^{-1}A_j}{(z_j^k)^2}\lim_{k\to\infty}+\frac{2\sigma^k\mu^k}{x_j^kz_j^k} - \lim_{k\to\infty}\frac{\Delta^p x_j^k}{x_j^k}\Big(\frac{\Delta^p z_j^k}{z_j^k}\Big).$$

The theorem's assumptions guarantee $\Delta^p x_j^k/x_j^k \to -1$ for $j \in \mathcal{N}$. The remainder of the proof follows from Assumptions 2 and 3. The second part of the proof can be proven in a similar manner. $\square$

The result hinges on the centering-corrector's step dependence on the product of the predictor steps. The theorem is also applicable for algorithms that center the predictor step instead of the corrector step, see Zhang and Zhang [62]. However, it

is not applicable to primal-dual methods where the centering step is not an explicit function of the centering parameter and the predictor (Newton) step.

Now we extend Theorem 2 of Mehrotra and Ye [35] to show that the Tapia predictor-corrector indicator identifies the optimal partition in finite time. Unlike their result, we allow the step length to vary from one iteration to another. To prove the theorem, we need a different set of assumptions than what we used in the previous propositions. The new assumptions are needed to establish global (polynomial) convergence for Mehrotra-type predictor-corrector interior-point algorithms. The previous assumptions guaranteed fast local convergence.

First, we provide an algorithmic framework. Then we present the main result of this chapter, polynomiality of the Tapia predictor-corrector indicator.

**Algorithm 4** Given a strictly feasible point $(x^0, y^0, z^0)$, for $k = 0, 1, 2, \ldots$, do

**(1)** Solve for the predictor step $(\Delta^p x^k, \Delta^p y^k, \Delta^p z^k)$ from (5.1) with

$$(x, y, z) = (x^k, y^k, z^k) \quad \text{and} \quad r = -X^k z^k.$$

**(2)** For $\sigma^k \in (0,1)$ and $\mu^k = ((x^k)^T z^k)/n$, solve for the corrector step $(\Delta^{cc} x^k, \Delta^{cc} y^k, \Delta^{cc} z^k)$ from (5.1) with

$$(\bar{x}, \bar{y}, \bar{z}) := (x^k + \Delta^p x^k, y^k + \Delta^p y^k, z^k + \Delta^p z^k), \quad r = \sigma^k \mu^k e - \bar{X}^k \bar{z}^k.$$

**(3)** Choose the largest $\alpha^k \in [0,1]$ such that

$$(x^{k+1}, y^{k+1}, z^{k+1}) \in \mathcal{N}_{-\infty}(\gamma),$$

then update accordingly

$$\begin{aligned} x^{k+1} &:= x^k + \alpha^k \Delta^p x^k + (\alpha^k)^2 \Delta^{cc} x^k \\ y^{k+1} &:= y^k + \alpha^k \Delta^p y^k + (\alpha^k)^2 \Delta^{cc} y^k \\ z^{k+1} &:= z^k + \alpha^k \Delta^p z^k + (\alpha^k)^2 \Delta^{cc} z^k. \end{aligned} \tag{6.12}$$

**(4)** Test for convergence.

Algorithm 4 differs from the previously presented predictor-corrector algorithm in the calculation of the step length, $\alpha^k$, and the iteration update.

**Theorem 6.2** Consider the iteration sequence $\{(x^k, y^k, z^k)\}$ generated by Algorithm 4. Assume that

(1) $(x^k)^T z^k \to 0$

(2) $\min(X^k Z^k e) \geq \gamma\mu^k$, for all $k$ and for some $\gamma \in (0,1)$

(3) The algorithmic parameter $\sigma^k = \sigma$ where $0 < \sigma < \gamma < 1$.

(4) $\|\Delta^p x^k\|, \|\Delta^p z^k\|, \|\Delta^{cc} x^k\|$, and $\|\Delta^{cc} z^k\|$ are $O(\mu^k)$ for $k \in \tilde{K}$, where $\tilde{K}$ is a subsequence of the natural numbers.

Define

$$\mathcal{B}^k = \{j : |x_j^{k+1} - x_j^k|/x_j^k \leq |z_j^{k+1} - z_j^k|/z_j^k\}.$$

Then there exists a finite $K$ such that for all $k \geq K$ and $k \in \tilde{K}$

$$\mathcal{B}^k = \mathcal{B}.$$

**Proof** Let $k \in \tilde{K}$. Then, from the linearized complementarity equations for the predictor and centering-corrector steps we obtain

$$\frac{\Delta x_j^k}{x_j^k} + \frac{\Delta z_j^k}{z_j^k} = \sigma^k \frac{\mu^k}{x_j^k z_j^k} - \frac{\Delta^p x_j^k}{x_j^k}\frac{\Delta^p z_j^k}{z_j^k} - 1.$$

From $x_j^k \geq \xi^* \gamma/n$, see Lemma 5.13 in Wright [49], Güler and Ye [17], and assumption 5, we have the following inequality,

$$\lim_{k\to\infty} \frac{|x_j^{k+1} - x_j^k|}{x_j^k} = \lim_{k\to\infty} \frac{|\alpha^k \Delta^p x_j^k + (\alpha^k)^2 \Delta^{cc} x_j^k|}{x_j^k} \leq \lim_{k\to\infty} \frac{\alpha^k |\Delta x_j^k|}{x_j^k} = 0 \quad \text{for } j \in \mathcal{B}.$$

Due to the nonnegativity of $\alpha^k$ and $x_j^k$, the limit must be zero. Now we must show that $|z_j^{k+1} - z_j^k|/z_j^k$ is bounded below by a number strictly greater than zero.

From the complementarity equations, the preceding limit, and centrality condition (1.6),

$$\limsup_{k\to\infty} \frac{\Delta z_j^k}{z_j^k} = \limsup_{k\to\infty} \left(\sigma^k \frac{\mu^k}{x_j^k z_j^k} - 1 - \frac{\Delta x_j^k}{x_j^k} - \frac{\Delta^p x_j^k}{x_j^k}\frac{\Delta^p z_j^k}{z_j^k}\right) \leq \frac{\sigma}{\gamma} - 1 \quad \text{for } j \in \mathcal{B}.$$

Note that the ratio $|\Delta^p z_j^k|/z_j^k$ is bounded below by $x_j^k/\gamma$ and

$$\lim_{k\to\infty} \frac{|\Delta^p x_j^k|}{x_j^k} = 0 \text{ for } j \in \mathcal{B}$$

which gives us the preceding inequality. The lim sup is negative, since $\sigma < \gamma$. Thus

$$\liminf_{k\to\infty} \frac{|z_j^{k+1} - z_j^k|}{z_j^k} \geq \liminf_{k\to\infty} (\alpha^k)^2 \frac{|\Delta z_j^k|}{z_j^k} \geq (\alpha^k)^2 \frac{(\gamma - \sigma)}{\gamma} > 0 \quad \text{for } j \in \mathcal{B}.$$

Then there exists a $K$ such that for all $k \geq K$ and $k \in \tilde{K}$, we obtain

$$\frac{|x_j^{k+1} - x_j^k|}{x_j^k} < (\alpha^k)^2 \frac{(\gamma - \sigma)}{\gamma} \leq \frac{|z_j^{k+1} - z_j^k|}{z_j^k} \text{ for } j \in \mathcal{B}.$$

We can prove the analogous result for $j \in \mathcal{N}$. $\square$

## 6.3 Numerical Results

The numerical experiments were conducted on a Sun workstation. Once again, we used LIPSOL, the software package, written by Zhang [57], which implements an infeasible primal-dual predictor-corrector interior point method. The tests were run on a subset of the *netlib* linear programming suite of problems.

In this experiment, the Tapia indicators and the Tapia predictor-corrector indicators with step length $\alpha^k$ equal one are calculated and compared at each iteration of the interior-point algorithm. The interior-point algorithm terminates when

$$\max\left(\frac{\|Ax - b\|_2}{1 + \|b\|_2}, \frac{\|A^T y + z - c\|_2}{1 + \|c\|_2}, \frac{|c^T x - b^T y|}{1 + b^T y}\right) \leq 10^{-8}.$$

The numerical experiments show the Tapia indicator has better global behavior than the Tapia predictor-corrector indicator. The global behavior can be attributed to the Tapia indicator's lack of dependence on the centering parameter. The poor global behavior of the Tapia predictor-corrector indicator is an immediate consequence of its dependence on the centering parameter, $\sigma^k$. As evidence, the Tapia indicator for zero variables approaches its limiting value earlier than the Tapia predictor-corrector indicator. However, the Tapia predictor-corrector indicator exhibits good local behavior. The behavior of the two indicators for nonzero variables is nearly identical.

The following tables track the behavior of the two indicators for a single variable as the iteration sequence converges to the solution set. The first column gives the iteration count. The second column contains the value of the centering parameter. The next four columns give the absolute values of the Tapia predictor-corrector indicator for primal variables, Tapia indicator for primal variables, the Tapia predictor-corrector indicator for dual variables, and the Tapia indicator for dual variables, respectively. The last column reports the maximum relative error at that iteration. Clearly, the semi-local behavior of the Tapia predictor-corrector indicator is undesirable.

| Iteration | $\sigma^k$ | $PC_p(x_j)$ | $T_p(x_j)$ | $PC_d(z_j)$ | $T_d(z_j)$ | Relative Error |
|---|---|---|---|---|---|---|
| 5 | 1.59e-03 | 6.34e+00 | 2.96e-02 | 9.80e-01 | 9.70e-01 | 3.54e-02 |
| 6 | 2.93e-08 | 5.60e-08 | 2.86e-04 | 1.00e+00 | 1.00e+00 | 7.35e-05 |
| 7 | 7.33e-17 | 0.00e+00 | 1.43e-08 | 1.00e+00 | 1.00e+00 | 3.68e-09 |

**Table 6.1** Problem AFIRO, $x_j^* = 0$, $z_j^* = 3.45e-01$

For problem *afiro*, the Tapia indicators identified the optimal partition at iteration 4. The Tapia predictor-corrector indicators took two additional iterations to generate the same partition. This typifies the behavior of the Tapia predictor-corrector indicator.

In Figure 6.1, we plot the values of $\log(1-PC(x^k))$ on the $y-$axis and the iteration count on the $x-$axis. Therefore in the plots, the Tapia predictor-corrector indicators for the active set converge to zero and the Tapia predictor-corrector indicators for nonzero variables tend to negative infinity.

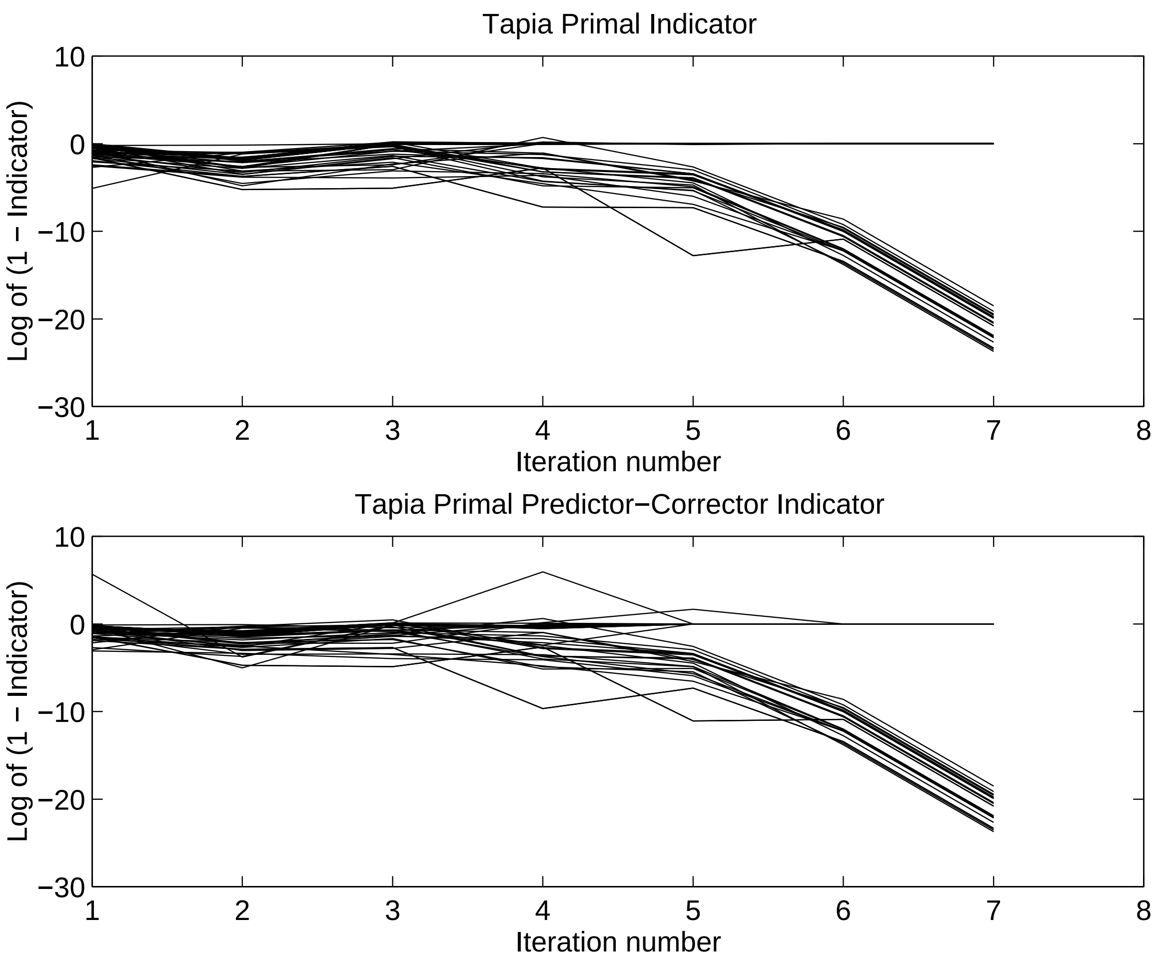


**Figure 6.1** Indicator Plots for Problem Afiro

From the numerical experiments, it appears that the fast local convergence of the Tapia predictor-corrector indicator to zero coincides with a sharp decrease in the value of the centering parameter, $\sigma^k$.

| Iteration | $\sigma^k$ | $PC_p(x_j)$ | $T_p(x_j)$ | $PC_d(z_j)$ | $T_d(z_j)$ | Relative Error |
|---|---|---|---|---|---|---|
| 11 | 1.00e-03 | 3.68e+00 | 7.49e-11 | 1.02e+00 | 1.00e+00 | 1.67e-05 |
| 12 | 1.71e-08 | 4.52e-08 | 4.04e-13 | 1.00e+00 | 1.00e+00 | 5.73e-08 |
| 13 | 2.74e-22 | 2.53e-16 | 5.06e-16 | 1.00e+00 | 1.00e+00 | 4.97e-14 |

**Table 6.2** Problem BEACONFD, $x_j^* = 0$, $z_j^* = 4.36e+03$

| Iteration | $\sigma^k$ | $PC_p(x_j)$ | $T_p(x_j)$ | $PC_d(z_j)$ | $T_d(z_j)$ | Relative Error |
|---|---|---|---|---|---|---|
| 10 | 4.40e-03 | 1.09e+00 | 1.09e+00 | 3.07e+00 | 8.62e-02 | 1.81e-04 |
| 11 | 2.40e-10 | 1.00e+00 | 1.00e+00 | 5.28e-10 | 1.44e-05 | 9.13e-07 |
| 12 | 1.00e-11 | 1.00e+00 | 1.00e+00 | 1.06e-11 | 1.60e-11 | 8.88e-13 |

**Table 6.3** Problem BLEND, $x_j^* = 2.49e-01$, $z_j^* = 0$

The study shows that the Tapia predictor-corrector indicator must be closer to the solution set than the Tapia indicator to correctly identify the partition. For 46 out of 87 problems, the Tapia predictor-corrector indicator needed an additional iteration to identify the optimal partition. For 30 problems, the two indicators identified the optimal partition at the same iteration. Eight digits of accuracy in the solution was insufficient for either indicator to identify the optimal partition for the remainder of the test set.

| Iteration | $\sigma^k$ | $PC_p(x_j)$ | $T_p(x_j)$ | $PC_d(z_j)$ | $T_d(z_j)$ | Relative Error |
|---|---|---|---|---|---|---|
| 7 | 1.01e-02 | 2.40e+00 | 1.09e-01 | 8.48e-01 | 8.91e-01 | 5.24e-03 |
| 8 | 4.06e-05 | 7.50e-05 | 4.63e-03 | 9.95e-01 | 9.95e-01 | 7.47e-05 |
| 9 | 7.85e-12 | 1.51e-11 | 2.27e-06 | 1.00e+00 | 1.00e+00 | 4.14e-08 |
| 10 | 2.59e-24 | 0.00e+00 | 2.27e-13 | 1.00e+00 | 1.00e+00 | 1.46e-13 |

**Table 6.4** Problem SC105, $x_j^* = 0$, $z_j^* = 1.80e-03$

# Chapter 7

# Concluding Remarks

At the beginning of the thesis, we posed four questions that must be addressed in all finite termination procedures. We answer each of them in turn. Then we propose a numerically effective finite termination procedure and describe future work.

## 7.1 When do we first attempt to compute an exact solution?

If computational savings is not a pressing concern, project from the interior-point iterate when total relative error $\leq 10^{-8}$. When early termination is the goal, projection from a pure composite Newton step is advocated. However for the activation criteria suggested in Chapter 5, the cost may exceed the benefits due to the average need of more than one projection attempt to find an exact solution. On the other hand, projecting from the pure composite Newton step when total relative error falls below $10^{-5}$ appears to be a promising idea to generate highly accurate solutions and reduce the iteration count.

## 7.2 How do we determine the partition of variables into their respective zero-nonzero sets ?

Variants of Mehrotra's predictor-corrector primal-dual interior-point algorithm provide the foundation for most practical interior-point codes. To take advantage of all available algorithmic information, we propose the Tapia predictor-corrector indicator, which naturally includes the corrector step, to identify the optimal partition. Globally, the Tapia predictor-corrector indicator behaves poorly, but locally exhibits

fast convergence. Therefore, the Tapia indicator should be used to identify the optimal partition early in the iterative process. At later iterations, the Tapia indicator and Tapia predictor-corrector indicator can be used in tandem.

## 7.3 Which mathematical model and what is the best way to solve it?

Before answering the last two questions, we must compare the orthogonal projection models with the Mehrotra-Ye procedures. Let us see how the procedures compare numerically.

**Observation 1**: For the common 55 problems tested, the Mehrotra-Ye procedure and the orthogonal projection method found the solution at the same iteration.

**Observation 2**: When the matrices were scaled, the weighted projection method needed one more iteration for problems *boeing2* and *finnis* than did scaled Mehrotra-Ye. Depending on the problem, the scale equals either $x_j^k$ or $\min(x_j^k, u_j - x_j^k)$ for $j \in \mathcal{B}$.

Even in the presence of degeneracy, the weighted projection model can be used to find an exact solution on the primal and dual optimal faces. When the coefficient matrix is rank deficient, or ill-conditioned the Cholesky-Infinity factorization implemented in LIPSOL calculates a basic solution of the linear system. Numerical instability arising is no longer a problem. Assuming your interior-point algorithm solves a normal equation at each iteration, no additional linear solver needs to be implemented. If the interior-point algorithm solves a symmetric indefinite augmented system at each iteration instead of reducing the KKT conditions to the normal equations, we would favor the scaled Mehrotra-Ye approach for the same reason.

## 7.4 Recommended Finite Termination Procedure

From our numerical experiments, we recommend the following procedure. We assume that a normal equation is solved at each iteration.

**Procedure 4** *(A Practical Finite Termination Procedure)*

**(1)** If **total relative error** $\leq 10^{-8}$ , use the **Tapia indicators** to partition the variables into $(\mathcal{B}, \mathcal{N})$.

**(2)** Set $x_{\mathcal{N}} := 0$ and $z_{\mathcal{B}} := 0$.

**(3)** Solve the **primal weighted projection** model with

$$d_{jj} = \begin{cases} \min(x_j^k, u_j - x_j^k) & \text{for } j \in \mathcal{B} \text{ and upper bounds exist} \\ x_j^k & \text{for } j \in \mathcal{B} \text{ else} \end{cases}$$

and the following unconstrained least squares problem

$$\min_{y} \quad \frac{1}{2}\|D(B^T y - c_{\mathcal{B}})\|^2$$

for the vectors $(x_{\mathcal{B}}, y, z_{\mathcal{N}}, s, w)$.

**(4)** If $x_{\mathcal{B}} \geq 0$, $z_{\mathcal{N}} \geq 0$, $s \geq 0$, $w \geq 0$, and we satisfy the primal and dual constraints, stop. Else, return to the interior point algorithm.

## 7.5 Future Work

Areas of further research include

1. The investigation of a nonlinear weighted projection model of the form

$$\min \quad \tfrac{1}{2}\|D(x_{\mathcal{B}}, x_{\mathcal{B}}^k)(x_{\mathcal{B}} - x_{\mathcal{B}}^k)\|^2$$

$$\text{s.t.} \qquad Bx_{\mathcal{B}} = b,$$

where $D = (X_{\mathcal{B}}^{1/2} X_{\mathcal{B}}^{k\,1/2})^{-1}$ to generate a feasible point on the optimal primal and dual faces,

2. further examination of when to first attempt to generate a feasible point on the optimal primal and dual faces,

3. implementation of a sparse $LU$-factorization to adequately ascertain the numerical effectiveness of column scaling or column ordering,

4. experimentation with a weighted projection and a scaled Mehrotra-Ye procedure added to interior-point methods for linear complementarity problems, and

5. the addition of a weighted projection model to a parallel interior-point algorithm.

# Bibliography

[1] E. Andersen. On exploiting problem structure in a basis identification procedure for linear programming. Technical Report, Department of Management, Odense University, Denmark, 1996.

[2] E. Andersen and Y. Ye. Combining interior-point and pivoting algorithms for linear programming. Technical Report, Department of Management Sciences, University of Iowa, 1994.

[3] K. Anstreicher, J. Ji, F. Potra, and Y. Ye. Probabilistic analysis of an infeasible interior-point algorithm for linear programming. Reports on Computational Mathematics, No. 27, Department of Mathematics, University of Iowa, 1993.

[4] M. Asic, V. Kovacevic-Vujcic, and M. Radosavljevic-Nikolic. Asymptotic Behaviour of Karmarkar's Method for Linear Programming. *Mathematical Programming*, 46:173–190, 1990.

[5] R. Bixby and M. Saltzman. Recovering an Optimal LP Basis from an Interior-Point Solution. Technical Report, TR91-32, Department of Mathematical Sciences, Rice University, 1991.

[6] R. Bixby. Implementing the Simplex Method: The Initial Basis *ORSA Journal on Computing*, 4:267–284, 1992.

[7] J. Czyzyk, S. Mehrotra, and S. Wright. PCx User Guide. Technical Report, OTC 96-01, Optimization Technology Center, Argonne National Laboratories, 1996.

[8] G. Dantzig. *Linear Programming and Extensions*, Princeton University Press, Princeton, NJ, 1963.

[9] A. El-Bakry. On the role of indicators in identifying zero variables in linear programming. PhD thesis, Department of Mathematical Sciences, Rice University, 1991.

[10] A. El-Bakry, R. Tapia, and Y. Zhang. A study of indicators for identifying zero variables in interior-point methods. *SIAM Review*, 36:45–72, 1994.

[11] A. El-Bakry, M. Gonzalez-Lima, R. Tapia, and Y. Zhang. On the identification of subgroups of variables in interior-point methods. Working Paper, Department of Computational and Applied Mathematics, Rice University, 1997.

[12] R. Freund. An Infeasible-Start Algorithm for Linear Programming whose Complexity depends on the Distance from the Starting Point to the Optimal Solution. Technical Report, Sloan W.P. No.3559-93, Sloan School of Management Massachusetts Institute of Technology, 1993.

[13] D. Gay. Stopping tests that compute optimal solutions for interior-point linear programming algorithms. In S. Gómez, J. Hennart, and R. Tapia, eds., *Advances in Numerical Partial Differential Equations and Optimization: Proceedings of the Fifth Mexico-United States Workshop*, pages 17–42. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1991.

[14] A. Goldman and A. Tucker. Theory of Linear Programming. In H.W. Kuhn and A.W. Tucker, eds., *Linear Inequalties and Related Systems*, pages 53–97. Princeton University Press, Princeton, NJ, 1956.

[15] G. Golub and C. Van Loan. *Matrix Computations*, Second Edition, John Hopkins University Press, Baltimore, MD, 1989.

[16] J. Gondzio. HOPDM(version 2.12) - a Fast LP Solver Based on a Primal-Dual Interior Point Method. *European Journal of Operational Research*,85:221 -225, 1995.

[17] O. Güler and Y. Ye. Convergence behavior of interior-point algorithms. *Mathematical Programming*, 60:215-228, 1993.

[18] J. Ji and F. Potra. Tapia indicators and finite termination of infeasible interior-point methods for degenerate LCP. Technical Report No. 81/1995, Department of Mathematics, University of Iowa, 1995.

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

[20] D. Kincaid and W. Cheney. *Numerical Analysis*,Brooks/Cole Publishing Company, Pacific Grove, CA 1991.

[21] M. Kojima, S. Mizuno, and A. Yoshise. A primal-dual interior-point method for linear programming. In N. Megiddo, ed., *Progress in Mathematical Programming, Interior Point and Related Methods*, pages 29–47. Springer–Verlag, New York, NY, 1989.

[22] M. Gonzalez-Lima. Effective Computation of the Analytic Center of the Solution Set in Linear Programming Using Primal-Dual Interior-Point Methods. PhD thesis, Department of Computational and Applied Mathematics, Rice University, 1994.

[23] I. Lustig, R. Marsten, and D. Shanno. On implementing Mehrotra's predictor-corrector interior-point method for linear programming. *SIAM Journal on Optimization*, 2(3):435–449, 1992.

[24] R. Marsten, M. Saltzman, D. Shanno, G. Pierce, and J. Ballintijn. Implementation of a Dual Affine Interior-Point Algorithm for Linear Programming. *ORSA Journal on Computing*, 1:4, 287–297, 1989.

[25] L. McLinden. An analogue of Moreau's proximation theorem with application to nonlinear complementarity problem. *Pacific Journal Math*, 88:101–161, 1988.

[26] N. Megiddo. Pathways to the Optimal Set in Linear Programming. In N. Megiddo, ed., *Progress in Mathematical Programming, Interior Point and Related Methods*, pages 29–47. Springer–Verlag, New York, NY, 1989.

[27] N. Megiddo. On Finding Primal- and Dual-Optimal Bases. *ORSA Journal on Computing*, 3:1, 63–65, 1991.

[28] S. Mehrotra. Implementations of affine scaling methods: towards faster implementations with complete Cholesky factor in use. Technical Report TR89-15, Dept. of Industrial Engineering and Management Sciences, Northwestern University, 1989.

[29] S. Mehrotra. On Finding a Vertex Solution Using Interior-Point Methods. *Linear Algebra and its Applications*, 152:233–253, 1991.

[30] S. Mehrotra. Finite Termination and Superlinear Convergence in Primal-Dual Methods. Technical Report TR91-13, Dept. of Industrial Engineering and Management Sciences, Northwestern University, 1991.

[31] S. Mehrotra. On the implementation of a primal-dual interior-point method. *SIAM Journal on Optimization*, 2(4):575–601, 1992.

[32] S. Mehrotra. Quadratic Convergence in a Primal-Dual Method. *Mathematics of Operations Research*, 18(3):741–751, 1993.

[33] S. Mizuno. A Predictor-Corrector Infeasible Interior-Point Algorithm for Linear Programming. *Operations Research Letters*,16(2):61–66, 1994.

[34] S. Mizuno, M. Todd, and Y. Ye. On adaptive primal-dual interior-point algorithms for linear programming. *Mathematics of Operations Research*, 18:964-981, 1993.

[35] S. Mehrotra and Y. Ye. Finding an Interior Point in the Optimal Face of Linear Programs. *Mathematical Programming*, 62:497–515, 1993.

[36] R. Monteiro and S. Wright. Local Convergence of Interior-Point Algorithms for Degenerate Monotone LCP. *Computational Optimization and Applications*, 3: 131–155, 1994.

[37] L. Portugal, M. Resende, G. Veiga, and J. Judice. A Truncated Primal-Infeasible Dual-Feasible Network Interior-Point Method. 1994.

[38] F. Potra. A quadratically convergent predictor-corrector method for solving linear programs from infeasible starting points. *Mathematical Programming*, 67:383–406, 1994.

[39] M. Resende and G. Veiga. An Efficient Implementation of a Network Interior-Point Method. In D. Johnson and C. McGeoch, eds., *Network Flows and Matching: First DIMACS Implementation Challenge, Volume 12 of DIMACS Series in*

*Discrete Mathematics and Theoretical Computer Science*, pages 299–348. American Mathematical Society, 1993.

[40] M. Resende, T. Tsuchiya, and G. Veiga. Identifying the Optimal Face of a Network Linear Program with a Globally Convergent Interior-Point Method. In W. Hager, D. Hearn, and P. Pardalos, eds., *Large-Scale Optimization: State of the Art*, pages 362–387. Kluwer Academic Publishers, 1994.

[41] C. M. Samuelsen. The Dikin-Karmarkar Principle for Steepest Descent. Ph.D. thesis, Department of Mathematical Sciences, Rice University, 1992.

[42] R. Skeel. Scaling for Numerical Stability in Gaussian Elimination. *Journal of the Association for Computing Machinery*, 26(3):494–526, 1979.

[43] R. A. Tapia. On the role of slack variables in quasi-Newton methods for constrained optimization. In L.C.W. Dixon and G.P. Szegö, eds., *Numerical Optimization of Dynamic Systems*, pages 235–246. North Holland, 1980.

[44] R. A. Tapia, Y. Zhang, M. Saltzman and A. Weiser. The Mehrotra Predictor-Corrector Interior-Point Method as a Perturbed Composite Newton Method. *SIAM Journal of Optimization*, 6(1): 47–56, 1996.

[45] R. A. Tapia and Y. Zhang. An Optimal Basis Identification Technique for Interior-Point Linear Programming Algorithms. *Linear Algebra and Its Applications*, 152:343–363, 1991.

[46] E. Tardos. A Strongly Polynomial Algorithm to Solve Combinatorial Linear Programs. *Operations Research*, 36(2):250–256, 1986.

[47] S. Vavasis and Y. Ye. Identifying an optimal basis in linear programming. Cornell University, 1995.

[48] S. Wright. Modified Cholesky Factorizations in Interior-Point Algorithms for Linear Programming. Preprint ANL/MCS-P600-0596, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, 60439.

[49] S. Wright. *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, Pennsylvania, 1997.

[50] Y. Ye and M. Todd. Containing and Shrinking Ellipsoids in the Path-Following Algorithm. *Mathematical Programming*, 39:305-317, 1987.

[51] Y. Ye Recovering Optimal Basic Variables in Karmarkar's Polynomial Algorithm for Linear Programming. *Mathematics of Operations Research*, 15(3):564-572, 1990.

[52] Y. Ye. On the Finite Convergence of Interior-Point Algorithms for Linear Programming. *Mathematical Programming( Series B)*, 57:325-335,1992.

[53] Y. Ye, O. Güler, R. A. Tapia, and Y. Zhang. A quadratically convergent $O(\sqrt{n}L)$-iteration algorithm for linear programming. *Mathematical Programming*, 59:151-162, 1993.

[54] Y. Ye Toward Probabilistic Analysis of Interior-Point Algorithms for Linear Programming. *Mathematics of Operations Research*, 19(1):38-52, 1994.

[55] Y. Ye. *Interior-Point Algorithm: Theory and Analysis*, John Wiley & Sons, New York, New York, 1997.

[56] Y. Zhang. On the Convergence of a class of Infeasible Interior-Point Methods for the Horizontal Linear Complementarity Problem. *SIAM Journal of Optimization*, 4:208-227, 1994.

[57] Y. Zhang. LIPSOL beta version 2.1. Department of Mathematics and Statistics, University of Maryland Baltimore County, 1995.

[58] Y. Zhang. Solving Large-Scale Linear Programs by Interior-Point Methods Under the MATLAB Environment. Department of Mathematics and Statistics, University of Maryland Baltimore County, 1996.

[59] Y. Zhang and R. Tapia. Superlinear and Quadratic Convergence of Primal-Dual Interior-Point for Linear Programming Revisited. *Journal of Optimization Theory and Applications*, 73:229-242, 1992.

[60] Y. Zhang, R. Tapia, and J. Dennis. On the Superlinear and Quadratic Convergence of Primal-Dual Interior-Point Linear Programming Algorithms. *SIAM Journal of Optimization*, 2:304-324, 1992.

[61] Y. Zhang, R. Tapia, and F. Potra. On the Superlinear Convergence of Interior-Point Algorithms for a General Class of Problems. *SIAM Journal of Optimization*, 3:413-422, 1993.

[62] Y. Zhang and D. Zhang. On Polynomiality of the Mehrotra-type Predictor-Corrector Interior-Point Algorithms. *Mathematical Programming*, 68:303-318, 1995.