Convergence Analysis of Affine Scaling Method for Linear Programming

Masakazu Muramatsu

Doctor of Philosophy

Graduate University for Advanced Studies,
School of Mathematical and Physical Science,
Department of Statistical Science

1994

Acknowledgements

First I would like to thank Prof. Kunio Tanabe of The Institute of Statistical Mathematics who was my principal academic adviser during the past three years for introducing me into this field of optimization. I learned a lot from discussions with him. His frequent advices were always to the point, and of great value for me. I convince that his research group was the best choice for me.

I also thank Prof. Yoshihiko Miyasato of The Institute of Statistical Mathematics and Prof. Shinji Mizuno of The Institute of Statistical Mathematics for reading manuscript carefully and giving useful suggestions.

Thanks are also due to my academic advisers Prof. Hideaki Itoh and Prof. Genshiro Kitagawa of The Institute of Statistical Mathematics for their supervision. Prof. Kitagawa encouraged and helped me when I had a trouble in research. Whenever I talked with Prof. Itoh, I felt full of inspirations flowing from him.

I would express a deep sense of gratitude to Prof. Masakazu Kojima of Tokyo Institute of Technology, for his kindly advice and suggestions. He also helped me on my trip to the United States in 1993 in all aspects. In particular I was introduced to many researchers through him on the trip, which was a precious experience. The trip is among the best memory of these three years.

Acknowledgement is also given to Dr. Takashi Tsuchiya of The Institute of Statistical Mathematics who, as an expert of affine scaling method, inspired me to this study. He poured large amounts of his time and power into discussions with me, which I always find useful. I really appreciate it.

Thanks are also due to Prof. I. I. Dikin of Siberian Energy Institute who is the founder of affine scaling method. He stayed at The Institute of Statistical Mathematics three months in 1993. His broad view of the affine scaling method frequently surprised me and gave me a lot of idea. In particular, the letter he sent to Dr. Tsuchiya directly inspired us to develop the result on long-step affine scaling method written in Chapter II of this thesis.

I am grateful to Prof. Shun-Ichi Amari of the University of Tokyo who was an adviser of my master thesis for recommending this course.

I also thank my colleagues of The Graduate University for Advanced Studies, Department of Statistical Science and The Institute of Statistical Mathematics. I was very lucky to have such good friends in the university. I have fully enjoyed my school days with them for three years.

Finally, I would like to thank my parents for their support, understanding towards me, and all they have done for me.

Preface

The affine scaling method was proposed by Dikin [11] in 1967, and is well-known as the first and the simplest interior point (IP) method for linear programming (LP) problems. Many implementations and numerical experiments [1, 2, 10, 18, 26, 33, 42, 45, 46, 47] have been done, and they show excellent performance of this method when a long step-size choice is adopted. In that case the next iterate is chosen by proceeding a fixed ratio $\lambda = 0.99$ (typically) in the direction to the boundary of the feasible region. In fact, it is reported that the affine scaling method is comparable or even superior to the simplex method in some cases. Though polynomial complexity has not yet been proved, the affine scaling method has the great advantage of its simplicity. It may be numerically robust, and yield fitness for preconditioning and parallel computation among other properties. The affine scaling method is among the promising IP methods for LP.

On the other hand theoretically, its global convergence with a practical step-size choice has long been an open question. There have been many papers which prove global convergence of the affine scaling method under various nondegeneracy assumptions and/or short step-size choice [3, 8, 11, 14, 20, 38, 58, 59, 57, 62, 63]. See [21] or [23] for a detailed history of the proof of the affine scaling method up to now. In particular, when the LP problem is degenerate, a global convergence proof of the affine scaling method becomes difficult. Many LP problems in the real world however, are tend to be degenerate due to their artificial backgrounds. Therefore, a convergence proof without nondegeneracy assumption is important from the practical point of view, and this is the first subject of this thesis.

In order to start an IP method, we need an initial feasible interior solution. Initializing an IP method is another general important problem. Two major approach have been proposed to overcome this latter difficulty. One approach is to use an IP method by introducing artificial variables like so-called Phase I-Phase II method or Big-$\mathcal{M}$ method. The other and more straight approach is to consider IP methods which allow an infeasible starting point [4, 6, 7, 14, 15, 16, 17, 27, 31, 54, 55]. A method of this type is called an infeasible-IP method (for the naming, see the remark at the end of [27]). It is an interesting question to determine how we can extend the affine scaling method in the framework of the infeasible-IP method, and that is the second subject of this thesis.

This paper consists of four chapters.

The first chapter is an introduction in which we give a general framework of this thesis.

In the second chapter, we deal with the long-step affine scaling method, and give an answer to the first problem. The local Karmarkar potential function developed by Tsuchiya [58] to analyze the global convergence of the affine scaling method plays important role here. We develop a new inequality to estimate the reduction of the local Karmarkar potential function, and succeed in proving that primal sequence will converge to a relative interior point of the optimal face, while dual estimates converge to the analytic center of the optimal face of the dual problem, if we take step-size $\lambda \leq 2/3$. The original contents of this chapter were given in Tsuchiya and Muramatsu [61], but the proof is somewhat simplified.

In the third chapter, we analyze a long-step variant of Karmarkar's projective scaling method [25] by applying the convergence results developed in the first chapter. This variant is essentially the long-step affine scaling method applied to homogeneous LP problem, and in practice, it could be expected to work well. By analyzing the local Karmarkar potential function more precisely, a new polynomiality proof of the projective scaling method is obtained. The original paper is appeared in Muramatsu and Tsuchiya [43].

In the fourth chapter, we challenge the second problem and propose an infeasible-IP method which is an extension of the affine scaling method. The search direction is composed of two affine scaling directions aiming at feasibility and optimality. The combination is defined to keep the scaling invariance of the search direction. The method can start from an arbitrary positive point, and generate a sequence which directly approaches an optimal solution. Again we prove the global convergence of the method by means of the local Karmarkar potential function without any kind of nondegeneracy assumption. When we use an infeasible-IP method, we do not know whether or not the LP problem has an (interior) feasible solution before the algorithm starts. In this point, infeasible-IP methods are quite different from feasible IP methods. We investigate the behavior of the primal and dual sequence in such various cases. The contents of this chapter are the same as Muramatsu and Tsuchiya [44].

Contents

Notation

General

$x \triangleq y$	means that x is defined by y.
x_j	is j-th component of vector x.
$x \geq y$	is component-wise inequalities if x and y are vectors.
$x \not\geq y$	implies that at least one inequality is violated.
$\sigma(x)$	is maximum component of vector x.
$\|x\|$	is Euclidean norm (2-norm) of x.
$\|x\|$	is (component-wise) absolute value of x.
$\|J\|$	is number of elements of J if J is a set, and
J^c	is complement of J with respect to $\{1, \ldots, n\}$.
R^n_+	is the set $\{ x \in R^n \mid x \geq 0 \}$.
$\mathrm{Im}(A)$	means image of A.
$\mathrm{Null}(A)$	means null space of A.
$\mathrm{rank}(A)$	means rank of A.
e	is a vector of all 1's of appropriate dimension.
0	is a vector of all 0's of appropriate dimension.
I	is an identity matrix of appropriate size.
P_A	is an orthogonal projection operator onto $\mathrm{Im}(A^t)$ with respect to Euclidean norm.
$\langle A\rangle, \langle B\rangle, \langle C\rangle, \ldots$	denote optimization problems.
$g^k = \mathcal{O}(f^k)$	means that there exist some constant k_0 and M such that $\|g^k\| \leq M\|f^k\|$ for all $k \geq k_0$. We sometimes use $\mathcal{O}(f^k)$ as a sequence g^k having the above property.
$\mathcal{A}(\subset)\subseteq \mathcal{B}$ and $\mathcal{B}(\supset)\supseteq \mathcal{A}$	mean that $\mathcal{A}$ is a (proper) subset of $\mathcal{B}$.
$\mathcal{A}^\perp$	is the orthogonal complement of linear subspace $\mathcal{A}$.
$\mathrm{Feas}\langle P\rangle$	denotes a feasible region of $\langle P\rangle$.
$\mathrm{Face}(F)$	denotes a face determined by F of indicated polyhedron.
$\mathrm{Int}(\mathcal{A})$	is a set of strictly positive points of $\mathcal{A}$.
$\text{Rel-Int}(\mathcal{A})$	is a set of relative interior points of $\mathcal{A}$.

Subscript expression: For an index set J of $\{1, \ldots, n\}$,

(1) if h is a vector, then h_J is a sub-vector of h composed of components corresponding to J,
(2) if D is a diagonal matrix, then D_J is a $|J| \times |J|$ diagonal sub-matrix of D whose diagonal components are corresponding to J,
(3) if M is an $m \times n$ matrix, then M_J is a sub-matrix of M whose *column* vectors are corresponding to J.

Strictly speaking, M_J is doubly defined if M is a matrix if arbitrary $m \times n$ matrix M happens to be diagonal. But the use of the notation never cause confusion because we always use (2) to the case where D is obviously diagonal.

Chapter II

$\langle P \rangle$	is the primal form LP problem to be solved.
$\langle DP \rangle$	is the dual problem of $\langle P \rangle$.
x^k	is the primal sequence.
s^k	is the dual sequence.
$d(x)$	is an affine scaling direction at x.
λ^k	is the step-size.
λ_{min}	is a lower bound of the step-size.
λ_{max}	is an upper bound of the step-size less than 1.
x^∞	is the limit point of x^k.
E	is an index set satisfying $x_E^\infty = 0$ and $x_{E^c}^\infty > 0$.
$\hat{s}$	is an alternative cost vector. $\hat{s}_{E^c} = 0$.
α^k	is the scaled direction ((2.2.1) on page 30).
$\hat{\alpha}_E^k$	is an approximating vector for α_E^k ((2.2.8) on page 31).
β^k	is $\alpha^k/(c^t x^k - c^t x^\infty)$ ((2.3.2) on page 32).
$\hat{\alpha}_E^k$	is an approximating vector for β_E^k ((2.3.3) on page 32).
$f_E(x_E^k)$	is the local Karmarkar potential function ((2.2.11) on page 31).
Δ^k	is the difference of the local Karmarpar potential function at iteration k.
$\hat{\Delta}^k$	is approximation for Δ^k ((2.3.13) on page 33).

Chapter III

$\langle A \rangle$	is the LP problem to be solved.
$\langle DA \rangle$	is the dual problem of $\langle A \rangle$.
$\langle H \rangle$	is a homogenized problem from $\langle A \rangle$.
$\langle DH \rangle$	is the dual problem of $\langle H \rangle$.
u^k	is the primal sequence for $\langle A \rangle$.
x^k	is the primal sequence for $\langle H \rangle$.
λ	is the step-size.
$d(x)$	is an affine scaling direction at x for $\langle H \rangle$.
g	is an inhomogeneous constraint vector.
$\mathcal{S}_A$	is the optimal face of $\langle A \rangle$.
$\mathcal{S}_H$	is the optimal face of $\langle H \rangle$.
N	is the always-active index set on $\langle H \rangle$.

B	is the complement of N.
$f(x)$	is the Karmarkar potential function.
β^k	is the scaled direction devided by the objective function value ((3.3.5) on page 46).
Δ^k	is the difference of the Karmarpar potential function at iteration k.

Chapter IV

$\langle P \rangle$	is the primal form LP problem to be solved.
$\langle DP \rangle$	is the dual problem of $\langle P \rangle$.
$\langle F \rangle$	is the LP problem to find a feasible solution of $\langle P \rangle$.
$\langle DF \rangle$	is the dual problem of $\langle F \rangle$.
x^k	is the primal sequence.
s^k	is the dual estimate for optimality ((4.5.7) on page 75).
z^k	is the dual estimate for feasibility ((4.8.97) on page 99).
w^k	is an infeasibility criteria ((4.2.20) on page 62).
x^∞	is the limit point of x^k.
w^∞	is the limit point of w^k.
E	is an index set satisfying $x_E^\infty = 0$ and $x_{E^c}^\infty > 0$.
$\tilde{F}$	is $E - F$.
θ	is defined in (4.2.6) on page 60.
$\hat{\delta}$	is defined in (4.2.8).
$(\hat{c}, -\hat{M})$	is an alternative cost vecter. See Proposition 4.6.4 on page 76.
Δx	is the search direction ((4.2.6) on page 60).
$\Delta_a x$	is the optimality direction ((4.2.4) on page 60).
$\Delta_n x$	is the feasibility direction ((4.2.5) on page 60).
μ^k	is the step-size $\lambda^k / \sigma((X^k)^{-1} \Delta x^k)$.
$f_E(x^k)$	is the local Karmarkar potential function for feasibility associated with E ((4.4.36) on page 69).
$g_E(x^k)$	is the local Karmarkar potential function for optimality associated with E ((4.8.59) on page 93).
α^k	is a scaled optimality direction ((4.4.3) on page 65).
β^k	is a scaled feasibility direction ((4.4.4) on page 65).
$\hat{\alpha}_E^k$	is an approximation for α_E^k ((4.6.9) on page 77).
$\hat{\beta}_E^k$	is an approximation for β_E^k ((4.4.6) on page 65).
$\hat{\beta}_{E^c}^k$	is an approximation for $\beta_{E^c}^k$ ((4.4.7) on page 66).
$\hat{\hat{\alpha}}_F^k$	is an approximation for $\hat{\alpha}_F^k$ ((4.8.11) on page 86).
$\hat{\hat{\alpha}}_{\tilde{F}}^k$	is an approximation for $\hat{\alpha}_{\tilde{F}}^k$ ((4.8.12) on page 86).
$\Phi_F(x^k)$, $\Psi_F(x^k)$, $\Lambda_F(x^k)$	are defined on page 81.
$\tilde{\phi}^k$	is $\|X_E^k\|^2 + \Phi_F$. (See page 90.)

CHAPTER I

Introduction to Linear Programming and Affine Scaling Method

In this chapter, we give a general framework of this thesis. In the first section, we introduce linear programming (LP) problem and related concepts which are indispensable to the paper. In the second section, we describe the Dikin's affine scaling method for LP which is the origin of this thesis. In the third section, we observe a relation between the affine scaling method and the Karmarkar's projective scaling method, which is frequently used in the thesis. In the fourth section, we show a few properties of a projection operator used to define the affine scaling method. The readers will also refer to the textbooks [51, 48] or papers [9, 58] for more complete description.

1. Linear Programming

Let $\mathcal{L}$ be an m dimensional subspace of R^n. We call a set which is a parallel transportation of $\mathcal{L}$: $\{\, x \in R^n \,|\, x - a \in \mathcal{L} \,\}$ where $a \in R^n$ an *affine space* $\mathcal{A}$. A polyhedron $\mathcal{P}$ is an intersection of an affine space $\mathcal{A}$ and the positive orthant, namely,

$$\mathcal{P} = \{\, x \in \mathcal{A} \,|\, x \geq 0 \,\}. \tag{1.1}$$

If $\mathcal{P}$ is bounded, $\mathcal{P}$ is called a *polytope*. Since we can express a linear space $\mathcal{L}$ in various ways, we also have as many choice to express a polyhedron $\mathcal{P}$. Among them, the following two expressions are well-known; *primal form*

$$\mathcal{P} = \{\, x \in R^n \,|\, Ax = Aa, \quad x \geq 0 \,\} \tag{1.2}$$

and *dual form*

$$\mathcal{P} = \left\{\, x \in R^n \,|\, x = B^t y - a, \quad x \geq 0 \,\right\} \tag{1.3}$$

where A and B are appropriate matrices. A point $x \in \mathcal{P}$ is called an *interior point* if it is strictly positive. (We use this term even if x is not included in $\mathcal{P}$ in Chapter IV where we treat an infeasible interior-point method.) If

$$\{\, x \in \mathcal{P} \,|\, x_F = 0, \quad x_{F^c} > 0 \,\} \tag{1.4}$$

is nonempty for an index set F, then

$$\mathcal{F} \triangleq \{\, x \in \mathcal{P} \,|\, x_F = 0, \quad x_{F^c} \geq 0 \,\} \tag{1.5}$$

is called a *face* of $\mathcal{P}$ determined by F, and F is the *always active index set* on $\mathcal{F}$. The set (1.4) is *relative interior* of $\mathcal{F}$ and denoted by Rel-Int $\mathcal{F}$. It is easy to see that $\mathcal{F}$ is also a polyhedron in R^n. In particular, $\mathcal{P}$ is a face of itself if it is nonempty. The linear space

$$\mathcal{T} \triangleq \{\, d \in R^n \mid d = \tau(x_1 - x_2), \quad x_1, x_2 \in \mathcal{F}, \quad \tau \in R \,\} \tag{1.6}$$

is called *direction space* or *tangent space* of $\mathcal{F}$. The dimension of face $\mathcal{F}$ is equal to that of its direction space. Therefore, we have in primal form expression

$$\dim \mathcal{F} = \dim \mathrm{Null}(A_{F^c}), \tag{1.7}$$

while in dual form,

$$\dim \mathcal{F} = \dim \left\{ B^t_{F^c} y \mid B^t_F y = 0 \right\}. \tag{1.8}$$

Recalling that the direction space of $\mathcal{P}$ is the original linear space $\mathcal{L}$ which is m dimensional, we see that at least $m - \dim \mathcal{F}$ constraints in dual form must be always active on $\mathcal{F}$. If

$$|F| > m - \dim \mathcal{F}, \tag{1.9}$$

then the face $\mathcal{F}$ is called a *primal-degenerate face*. In particular, if $\mathcal{P}$ is primal-degenerate, then $|F| > 0$ which implies $\mathcal{P}$ does not have an interior point. It is easy to see that if a face $\mathcal{F}$ is primal-degenerate, then a subface of $\mathcal{F}$ determined by an appropriate index set $F' \supset F$ is also primal-degenerate.

We define the *analytic center* of $\mathcal{F}$ by the optimal solution of

$$\begin{cases} \text{minimize} & \sum_{j \in F^c} \log x_j \\ \text{subject to} & x \in \text{Rel-Int}\, \mathcal{F}. \end{cases} \tag{1.10}$$

If and only if $\mathcal{F}$ is a polytope, the optimal solution exists. If it exists, it is unique since the objective function is strictly convex.

A linear programming (LP) problem is an optimization problem on polyhedron whose objective function is linear;

$$\langle P \rangle \begin{cases} \text{minimize} & c^t x \\ \text{subject to} & x \in \mathcal{P}. \end{cases} \tag{1.11}$$

The polyhedron $\mathcal{P}$ is referred to as a *feasible region*, which is denoted by Feas$\langle P \rangle$. A point $x \in \mathcal{P}$ is called a *feasible point* or a *feasible solution*. A feasible solution which attains the optimal value is called an *optimal solution*. We note that an LP problem may not have an optimal solution nor a feasible solution. The whole set of the optimal solution is referred to as the *optimal face*, because it is a face. If the objective function value is constant on a face, then the face is called a *dual-degenerate face*. In particular, the optimal face is a dual-degenerate face and a vertex is always dual-degenerate.

Now we introduce the two nondegeneracy assumptions which are frequently used in analysis of affine scaling method or simplex method, but are never made in this paper;

(1) *Primal-nondegeneracy assumption:*
Feas$\langle P \rangle$ does not have a primal-degenerate face.

(2) *Dual-nondegeneracy assumption:*
Feas$\langle P \rangle$ does not have a dual-degenerate face other than vertices.

We admit that there is no primal-dual relation between (1) and (2), but we follow the tradition.

Assume that $\mathcal{P}$ has a primal expression $\{\, x \in R^n \mid Ax = b, \quad x \geq 0 \,\}$. Then we have a *primal form* or *standard form* LP problem;

$$(1.12) \qquad \langle P \rangle \begin{cases} \text{minimize} & c^t x \\ \text{subject to} & Ax = b, \quad x \geq 0. \end{cases}$$

Given coefficients (A, b, c) of $\langle P \rangle$, we can construct another LP problem;

$$(1.13) \qquad \langle DP \rangle \begin{cases} \text{maximize} & b^t y \\ \text{subject to} & s = c - A^t y, \quad s \geq 0. \end{cases}$$

which is called a *dual problem* of $\langle P \rangle$ [1]. In contrast with the dual problem, we sometimes call $\langle P \rangle$ the *primal problem.* A primal-dual pair of LP problems plays important role in the theory of LP. In the rest of this section, we observe some of the relations between them.

First, we introduce the so-called *weak duality theorem.*

LEMMA 1.1. *If* $x \in \mathrm{Feas}\langle P \rangle$ *and* $s \in \mathrm{Feas}\langle DP \rangle$, *then* $x^t s = c^t x - b^t y \geq 0$. *In particular, if* $x^t s = 0$, *then* x *and* s *are optimal solutions of* $\langle P \rangle$ *and* $\langle DP \rangle$, *respectively.*

The direct consequence of the weak duality theorem is that if the minimum value of $\langle P \rangle$ cannot be attained, then $\langle DP \rangle$ does not have a feasible solution, and vice versa. The proof of Lemma 1.1 is easy, thus we omit it.

Next we introduce two fundamental theorems of LP.

THEOREM 1.2 (DUALITY THEOREM). *If* $\langle P \rangle$ *has an optimal solution, then* $\langle D \rangle$ *also has an optimal solution and the two optimal values coincide.*

THEOREM 1.3 (STRICT COMPLEMENTARITY THEOREM). *If* $\langle P \rangle$ *has an optimal solution, then there exists a pair of optimal solutions* (x, s) *for* $\langle P \rangle$ *and* $\langle DP \rangle$ *which satisfies strict complementarity condition;*

$$(1.14) \qquad x_j = 0 \qquad \text{if and only if} \qquad s_j > 0.$$

One can find thier proof in any textbooks of LP. See, for example, [51]. We remark that these theorems can be proved as a byproduct of the results of Chapter II (See pages 40).

The following lemma is a variant of the well-known *Farkas' lemma.*

LEMMA 1.4. *If* $\langle P \rangle$ *has an interior feasible solution, then every dual-degenerate face of* $\langle DP \rangle$ *is bounded.*

PROOF. Assume that an index set F determines a dual-degenerate face $\mathcal{F}$ of $\mathrm{Feas}\langle DP \rangle$:

$$(1.15) \qquad \left\{ s \in R^n \mid s = c - A^t y, \quad s_F = 0, \quad s_{F^c} \geq 0 \right\},$$

and choose an $\tilde{x} > 0$ such that $A\tilde{x} = b$. If $\mathcal{F}$ is not bounded, then there exists an "infinite" direction d such that

$$(1.16) \qquad A^t_{F^c} d \neq 0, \quad A^t_{F^c} d \leq 0, \quad A^t_F d = 0 \qquad \text{and} \qquad b^t d = 0.$$

[1] We can also define the dual problem in coordinate-free form, but it exceeds the aim of this chapter.

Suppose there exists such d. Then we have

$$b^t d = \tilde{x}^t A^t d = \tilde{x}^t_{F^c} A^t_{F^c} d < 0 \tag{1.17}$$

which contradicts the fact that $b^t d = 0$. □

The last lemma we introduce in this section is a characterization of a dual-degenerate face.

LEMMA 1.5. *A* Face(F) *is dual-degenerate, if and only if there exists an* $\hat{s}$ *such that*

$$\hat{s} = c - A^t \hat{y}, \qquad \hat{s}_{F^c} = 0. \tag{1.18}$$

Furthermore, in this case $c^t x - c^t \hat{x} = \hat{s}^t x = \hat{s}^t_F x_F$ *for all* $x \in \mathrm{Feas}\langle P\rangle$ *and* $\hat{x} \in \mathrm{Face}(F)$.

PROOF. Assume that Face(F) is dual-degenerate. Then for all $d \in \mathcal{T} = \{\, d \in R^n \mid d_F = 0,\ \ A_{F^c} d_{F^c} = 0 \,\}$, $c^t d = c^t_{F^c} d_{F^c} = 0$. Since $\{\, d_{F^c} \mid d \in \mathcal{T} \,\}$ is $\mathrm{Null}(A_{F^c})$, $c_{F^c} \in \mathrm{Im}(A^t_{F^c})$, which implies the existence of $\hat{s}$.

On the contrary, assume that such $\hat{s}$ exists. Choosing a point $\hat{x} \in \mathrm{Face}(F)$, we have for all $x \in \mathrm{Feas}\langle P\rangle$,

$$\begin{aligned} c^t x - c^t \hat{x} &= c^t_F x_F + c^t_{F^c}(x_{F^c} - \hat{x}_{F^c}) \\ &= c^t_F x_F + \hat{y}^t A_{F^c}(x_{F^c} - \hat{x}_{F^c}) \\ &= c^t_F x_F + \hat{y}^t (b - A_F x_F) - \hat{y}^t A_{F^c} \hat{x}_{F^c} \\ &= (c_F - A^t_F \hat{y})^t x_F - \hat{y}^t (A_{F^c} \hat{x}_{F^c} - b) \\ &= \hat{s}^t_F x_F. \end{aligned} \tag{1.19}$$

This implies that for $x \in \mathrm{Face}(F)$, $c^t x = c^t \hat{x}$ which means Face(F) is dual-degenerate, and the latter part of the lemma. □

2. The Affine Scaling Method for Linear Programming

There have been many kinds of affine scaling methods proposed for linear programming [11, 8, 63, 28] and Convex Quadratic programming [52, 60]. A distinct feature of the affine scaling methods is that they are invariant under the change of unit of the coordinates which is called *scaling invariance*. But not all the methods which is scaling invariant cannot be called 'affine scaling method'. In order to obtain the name, they should have something to do with the Dikin's original affine scaling method for LP, which we introduce in this section.

We consider the LP problem $\langle P\rangle$ in (1.11). We assume

ASSUMPTION 1. $\langle P\rangle$ *has an interior feasible solution* x^0.

Most interior point methods need this assumption since they produce a sequence of interior feasible solutions. But it is not easy to find an interior feasible solution in general, and in Chapter IV, we will remove this assumption and discuss how to extend the affine scaling method defined here to allow an infeasible starting point. We also make the following assumption;

ASSUMPTION 2. *The objective function value is not constant on the feasible region.*

If the objective function value is constant, then every feasible point including the initial point x^0 is optimal, and we can check it easily. Therefore, it is reasonable to make this assumption.

For an interior feasible point x, we consider a metric matrix $G(x)$ and a steepest descent direction with respect to $G(x)$ which is the optimal solution of

$$(2.1)\qquad \begin{cases} \text{minimize} & (G^{-1}(x)c-d)^tG(x)(G^{-1}(x)c-d) \\ \text{subject to} & d \in \mathcal{L} \end{cases}$$

where $\mathcal{L}$ is the direction space of $\mathcal{P}$. We define the affine scaling direction $d(x)$ at x as the opposite direction of the steepest descent direction (steepest ascent direction) when we choose the metric $G(x)$ to be X^{-2} where $X = \text{diag}(x)$. We have an explicit formula of the affine scaling direction by using primal form expression for $\mathcal{P}$ as

$$(2.2)\qquad d(x) = X(I - P_{AX})Xc,$$

while by dual form,

$$(2.3)\qquad d(x) = XP_{BX^{-1}}Xc.$$

(This observation means that the so-called primal and dual affine scaling methods are essentially identical.)

Given an initial feasible solution x^0, the affine scaling method produces a sequence of interior feasible solutions $x^k : k = 0, 1, 2, \dots$ as follows:

$$(2.4)\qquad x^{k+1} = x^k - \mu^k d(x^k)$$

where μ^k is a real positive number. We note that $\mu^k < \sigma((X^k)^{-1}d(x^k))^{-1}$ must hold to keep the next iterate positive. If we choose

$$(2.5)\qquad \mu^k = \frac{\lambda^k}{\sigma((X^k)^{-1}d(x^k))}$$

where $\lambda^k < 1$, then the step-size choice is called a *long step-size choice* and the method is called a long-step affine scaling method. On the other hand, if we choose

$$(2.6)\qquad \mu^k = \frac{\tilde{\lambda}^k}{\|(X^k)^{-1}d(x^k)\|},$$

where $\tilde{\lambda}^k \le 1$, then it is called a *short step-size choice* and the method is a short-step affine scaling method. In the Dikin's original affine scaling method, $\tilde{\lambda}$ is taken to be 1. We note that since $\|(X^k)^{-1}d(x^k)\| \ge \sigma((X^k)^{-1}d(x^k))$, the next iterate is always feasible even under a short step-size choice.

The short-step affine scaling method is also derived as follows. For convenience, we consider the primal form. Given an interior feasible point x^k, we define a mapping $h : R^n_+ \to R^n_+$ by

$$(2.7)\qquad h(x) = (X^k)^{-1}x.$$

Then, $\langle P \rangle$ in primal form is transformed to

$$(2.8)\qquad \langle hP \rangle \begin{cases} \text{minimize} & c^tX^k\chi \\ \text{subject to} & AX^k\chi = b, \quad \chi \ge 0, \end{cases}$$

and x^k is mapped to e. This procedure is sometimes called *centering* or *scaling*. We consider an unit ellipsoid around e which is inscribed in $\mathrm{Feas}\langle hP\rangle$ and optimize the objective function on it;

$$\text{(2.9)} \qquad \langle E\rangle \begin{cases} \text{minimize} & (X^k c)^t \chi \\ \text{subject to} & \|\chi - e\|^2 \le 1, \quad AX^k\chi = b. \end{cases}$$

This optimization problem can be easily solved, and the optimal solution is

$$\text{(2.10)} \qquad \chi^* = e - \frac{(I - P_{AX^k})X^k c}{\|(I - P_{AX^k})X^k c\|}.$$

In the original space, we have

$$\text{(2.11)} \qquad x^* = X^k \chi^* = x^k - \frac{X^k(I - P_{AX^k})X^k c}{\|(I - P_{AX^k})X^k c\|},$$

and x^* can be regarded as an approximation for the optimal solution of $\langle P\rangle$. Therefore,

$$\text{(2.12)} \qquad \tilde{d}(x^k) \triangleq \frac{X^k(I - P_{AX^k})X^k c}{\|(I - P_{AX^k})X^k c\|}$$

is sometimes called a short-step affine scaling direction, and the iteration of a short-step affine scaling method is;

$$\text{(2.13)} \qquad x^{k+1} = x^k - \tilde{\lambda}^k \tilde{d}(x^k).$$

We note that a short-step may not be really shorter than a long-step if $\tilde{\lambda} > \lambda$. But in practice, their difference is great (See, for example, page 40).

The affine scaling direction also has the third characterization. Again we use the primal form to discuss it. For given an interior feasible solution x, we define a *dual estimate* $s(x)$ be the optimal solution of

$$\text{(2.14)} \qquad \begin{cases} \text{minimize} & \|Xs\|^2 \\ \text{subject to} & s = c - A^t y. \end{cases}$$

By calculating Karush-Kuhn-Tucker condition, we have

$$\text{(2.15)} \qquad s(x) = X^{-1}(I - P_{AX})Xc = X^{-2}d(x).$$

Therefore, the affine scaling direction $d(x)$ can be calculated as

$$\text{(2.16)} \qquad d(x) = X^2 s(x).$$

We note that $s(x)$ satisfies the equality condition of $\langle DP\rangle$ and this is why $s(x)$ is called dual estimate. If $s(x) \ge 0$ happens then $s(x)$ is a feasible solution of $\langle DP\rangle$. In this sense, the affine scaling method generates not only x^k but also implicitly $s^k = s(x^k)$, which are candidates for a dual optimal solution. In fact, we will prove that the dual estimates converges to the analytic center of the optimal face of $\langle DP\rangle$ under a certain step-size choice in Chapter II. The sequences x^k and s^k are called *primal* and *dual* sequence, respectively.

If $d(x^k) \le 0$ happens, then the objective function value can be decreased to minus infinity, hence we terminate the algorithm and find that $\langle P\rangle$ does not have the optimal solution.

Considering the above situation, we describe the algorithm of a long-step affine scaling method.

LONG STEP AFFINE SCALING METHOD
Initialize $x^0, k := 0$;
while x^k does not satisfy a stopping criteria **do**
  begin
    Compute s^k by using (2.14);
    if $s^k \leq 0$ **then return** UNBOUNDED **endif**
    $d^k := (X^k)^2 s^k$;
    $\mu^k := \lambda^k / \sigma((X^k)^{-1} d^k)$;
    $x^{k+1} := x^k - \mu^k d^k$;
    $k := k + 1$
  end
return x^k;

We will discuss how to determine the step-size λ^k and the stopping criteria in Chapter II.

3. The Projective Scaling Method as an Affine Scaling Method

The projective scaling method was first proposed by Karmarkar [25] in 1984 and well-known as the first polynomial time interior point method. In this section, we show that the projective scaling method is equivalent to an affine scaling method applied to a homogeneous LP problem. The relation has been well-known (See, for example, Bayer and Lagarias [9]), and will be very helpful to understand the analysis of this thesis.

We define the projective scaling method. Consider the following LP problem which is called *Karmarkar canonical form*;

$$\langle K \rangle \begin{cases} \text{minimize} & c^t u \\ \text{subject to} & Au = 0, \quad e^t u = 1, \quad u \geq 0. \end{cases} \tag{3.1}$$

We assume;

(1) An initial interior feasible solution $u^0 > 0$ is given.
(2) The optimal value is known to be 0.

Then the projective scaling method produces a sequence of interior feasible solutions $u^k : k = 0, 1, 2, \ldots$ as follows;

$$u^{k+1} = u^k - \mu_P^k d^P(u^k) \tag{3.2}$$

where

$$d^P(u) \triangleq U(I - P_{AU})Uc - e^t U(I - P_{AU})Ucu = d(u) - e^t d(u) u, \tag{3.3}$$

$U \triangleq \mathrm{diag}(u)$, and μ_P^k is the step-size which is a positive number. The vector $d^P(u)$ is called the *projective scaling direction*.

Next we consider

$$\langle HK \rangle \begin{cases} \text{minimize} & c^t x \\ \text{subject to} & Ax = 0, \quad x \geq 0, \end{cases} \tag{3.4}$$

which is a homogeneous LP problem. Obviously, $\langle HK\rangle$ is derived by removing the constraint $e^t x = 1$ from $\langle K\rangle$. We apply the affine scaling method (2.4) for $\langle HK\rangle$ and get a sequence x^k.

Then we have the following lemma.

LEMMA 3.1. *If we put*

$$\tilde{u}^k \triangleq \frac{x^k}{e^t x^k}, \tag{3.5}$$

then

$$\tilde{u}^{k+1} = \tilde{u}^k - \theta(x^k;\mu^k)d^P(\tilde{u}^k) \tag{3.6}$$

where θ is a positive valued function.

This lemma means that the conically projected affine scaling direction is equivalent to the projective scaling direction as a direction. Therefore, the projective scaling method can be regarded as a kind of affine scaling direction though the step-size choice is changed nonlinearly.

PROOF. Noting that for $\zeta \in R_+$,

$$d(\zeta x^k) = \zeta^2 d(x^k), \tag{3.7}$$

we have

$$\begin{aligned}
\tilde{u}^{k+1} &= \frac{x^{k+1}}{e^t x^{k+1}} = \frac{x^k - \mu^k d(x^k)}{e^t(x^k - \mu^k d(x^k))} \\
&= \frac{x^k - \mu^k (e^t x^k)^2 d(\tilde{u}^k)}{e^t x^k(1 - \mu^k e^t x^k e^t d(\tilde{u}^k))} \\
&= \frac{x^k}{e^t x^k} + \frac{\mu^k e^t x^k (e^t d(\tilde{u}^k) x^k - (e^t x^k) d(\tilde{u}^k))}{e^t x^k (1 - \mu^k e^t x^k e^t d(\tilde{u}^k))} \\
&= \tilde{u}^k - \frac{\mu^k e^t x^k}{1 - \mu^k e^t x^k e^t d(\tilde{u}^k)}(d(\tilde{u}^k) - e^t d(\tilde{u}^k)\tilde{u}^k),
\end{aligned} \tag{3.8}$$

and

$$\begin{aligned}
1 - \mu^k e^t x^k e^t d(\tilde{u}^k) &= \frac{e^t x^k - \mu^k e^t d(x^k)}{e^t x^k} \\
&= \frac{e^t x^{k+1}}{e^t x^k} > 0.
\end{aligned} \tag{3.9}$$

Therefore,

$$\tilde{u}^{k+1} = \tilde{u}^k - \frac{\mu^k (e^t x^k)^2}{e^t x^{k+1}} d^P(\tilde{u}^k), \tag{3.10}$$

and this completes the proof. □

4. On the Projection Operator

We used P_{AX}, a projection operator onto $\mathrm{Im}(XA^t)$, to define the affine scaling direction in (2.2). The projection operator is also used by most of the IP methods proposed to date (see [24]). Therefore, we are motivated to show a few lemmas on this projection matrix in this section.

The first lemma is on the norm of $X^{-1}P_{AX}X$ which appears in the affine scaling direction (2.2).

LEMMA 4.1. *For given $c \in R^n$ and $A \in R^{m\times n}$, there exists a constant M such that for any positive diagonal matrix $X \in R^{n\times n}$,*

$$\|X^{-1}P_{AX}Xc\| \leq M. \tag{4.1}$$

The reader may notice that the above lemma implies the boundedness of dual estimates (See (2.15).). We will follow the proof by Vanderbei and Lagarias [63]. To do that, first we introduce the following basic theorem on determinant.

THEOREM 4.2 (BINET-CAUTHY THEOREM). *For A, $B \in R^{m\times n}$,*

$$\det(AB^t) = \sum_{J\in\mathfrak{S}} \det A_J \det B_J \tag{4.2}$$

where $\mathfrak{S}$ is the set of index sets of $\{1,\dots,n\}$, each of which has m indices arranged in increasing order.

PROOF. Let $C = AB^t$. We use component-wise expression : $c_{ij} = \sum_{p=1}^n a_{ip}b_{jp}$. Since determinant is a linear function of each column, we have

$$\begin{aligned}
\det(C) &= \det\begin{pmatrix} \sum_{p_1=1}^n a_{1p_1}b_{1p_1} & \cdots & \sum_{p_m=1}^n a_{1p_m}b_{mp_m} \\ \vdots & \ddots & \vdots \\ \sum_{p_1=1}^n a_{mp_1}b_{1p_1} & \cdots & \sum_{p_m=1}^n a_{mp_m}b_{mp_m} \end{pmatrix} \\
&= \sum_{p_1,\cdots,p_m} \det\begin{pmatrix} a_{1p_1}b_{1p_1} & \cdots & a_{1p_m}b_{mp_m} \\ \vdots & \ddots & \vdots \\ a_{mp_1}b_{1p_1} & \cdots & a_{mp_m}b_{mp_m} \end{pmatrix} \\
&= \sum_{p_1,\cdots,p_m} \det\begin{pmatrix} a_{1p_1} & \cdots & a_{1p_m} \\ \vdots & \ddots & \vdots \\ a_{mp_1} & \cdots & a_{mp_m} \end{pmatrix} \det\begin{pmatrix} b_{1p_1} & & 0 \\ & \ddots & \\ 0 & & b_{mp_m} \end{pmatrix}.
\end{aligned} \tag{4.3}$$

If, in summation, there exist indices i, j such that $p_i = p_j$, then that term is zero. For $p \in \mathfrak{S}$, let $[p]$ be a set of index sets, each of which is a juxtaposition of p. Then for some

$p \in \mathfrak{S}$, we have

$$
\begin{aligned}
(4.4) \qquad & \sum_{p' \in [p]} \det \begin{pmatrix} a_{1p'_1} & \cdots & a_{1p'_m} \\ \vdots & \ddots & \vdots \\ a_{mp'_1} & \cdots & a_{mp'_m} \end{pmatrix} \det \begin{pmatrix} b_{1p'_1} & & 0 \\ & \ddots & \\ 0 & & b_{mp'_m} \end{pmatrix} \\
&= \sum_{p' \in [p]} \operatorname{sgn} \begin{pmatrix} p \\ p' \end{pmatrix} \det \begin{pmatrix} a_{1p_1} & \cdots & a_{1p_m} \\ \vdots & \ddots & \vdots \\ a_{mp_1} & \cdots & a_{mp_m} \end{pmatrix} \det \begin{pmatrix} b_{1p'_1} & & 0 \\ & \ddots & \\ 0 & & b_{mp'_m} \end{pmatrix} \\
&= \det \begin{pmatrix} a_{1p_1} & \cdots & a_{1p_m} \\ \vdots & \ddots & \vdots \\ a_{mp_1} & \cdots & a_{mp_m} \end{pmatrix} \sum_{p' \in [p]} \operatorname{sgn} \begin{pmatrix} p \\ p' \end{pmatrix} \det \begin{pmatrix} b_{1p'_1} & & 0 \\ & \ddots & \\ 0 & & b_{mp'_m} \end{pmatrix}
\end{aligned}
$$

where

$$
(4.5) \qquad \operatorname{sgn} \begin{pmatrix} p \\ p' \end{pmatrix}
$$

takes 1 if p' becomes p in even number of index changes, and -1 if odd. Now we have

$$
\begin{aligned}
(4.6) \qquad \sum_{p' \in [p]} \operatorname{sgn} \begin{pmatrix} p \\ p' \end{pmatrix} \det \begin{pmatrix} b_{1p'_1} & & 0 \\ & \ddots & \\ 0 & & b_{mp'_m} \end{pmatrix} &= \sum_{p' \in [p]} \operatorname{sgn} \begin{pmatrix} p \\ p' \end{pmatrix} b_{1p'_1} \cdots b_{mp'_m} \\
&= \det \begin{pmatrix} b_{1p_1} & \cdots & b_{1p_m} \\ \vdots & \ddots & \vdots \\ b_{mp_1} & \cdots & b_{mp_m} \end{pmatrix}.
\end{aligned}
$$

Hence we have

$$
\begin{aligned}
(4.7) \qquad \det(C) &= \sum_{p_1, \cdots, p_m} \det \begin{pmatrix} a_{1p_1} & \cdots & a_{1p_m} \\ \vdots & \ddots & \vdots \\ a_{mp_1} & \cdots & a_{mp_m} \end{pmatrix} \det \begin{pmatrix} b_{1p_1} & & 0 \\ & \ddots & \\ 0 & & b_{mp_m} \end{pmatrix} \\
&= \sum_{p \in \mathfrak{S}} \det \begin{pmatrix} a_{1p_1} & \cdots & a_{1p_m} \\ \vdots & \ddots & \vdots \\ a_{mp_1} & \cdots & a_{mp_m} \end{pmatrix} \det \begin{pmatrix} b_{1p_1} & \cdots & b_{1p_m} \\ \vdots & \ddots & \vdots \\ b_{mp_1} & \cdots & b_{mp_m} \end{pmatrix},
\end{aligned}
$$

which completes the proof. □

We are ready to prove Lemma 4.1.

PROOF OF LEMMA 4.1. Without loss of generality, we can assume that A is full rank. Now what we should prove is that

$$
(4.8) \qquad y \triangleq (AX^2A^t)^{-1}AX^2c
$$

is bounded for arbitrary X. Let $U \triangleq AX$, $G \triangleq (AX^2A^t)^{-1} = (UU^t)^{-1}$ and $z \triangleq AX^2c$. By using Cramer's formula, we have

$$
(4.9) \qquad y_j = \det G[j] / \det G
$$

where $G[j]$ is a matrix whose j-th column is z and others are the same as G. Now from Binet-Cauthy theorem, we have

$$\det G = \det(UU^t) = \sum_{J\in\mathfrak{S}} (\det U_J)^2 = \sum_{J\in\mathfrak{S}} (\det A_J)^2(\det X_J)^2 \tag{4.10}$$

where $\mathfrak{S}$ is the set of index sets in Binet-Cauthy's theorem. As for $G[j]$, we have the following component-wise expression;

$$(g[j])_{pq} = \begin{cases} z_p & \text{if } q = j \\ g_{pq} & \text{otherwise.} \end{cases} \tag{4.11}$$

Recalling the definition of G and z, we have

$$(g[j])_{pq} = \begin{cases} \sum_{l=1}^n a_{pl}x_l^2 c_l & \text{if } q = j \\ \sum_{l=1}^n a_{pl}x_l^2 a_{ql} & \text{otherwise.} \end{cases} \tag{4.12}$$

Hence, if we put

$$(a[j])_{pq} = \begin{cases} c_p & \text{if } q = j \\ a_{pq} & \text{otherwise,} \end{cases} \tag{4.13}$$

then we have

$$(g[j])_{pq} = \sum_{l=1}^n a_{pl}x_l^2 (a[j])_{ql}. \tag{4.14}$$

Again from Binet-Cauthy theorem, we have

$$\det G[j] = \sum_{J\in\mathfrak{S}} (\det X_J)^2 \det A_J \det(A[j])_J, \tag{4.15}$$

thus

$$y_j = \frac{\sum_{J\in\mathfrak{S}} (\det X_J)^2 \det A_J \det(A[j])_J}{\sum_{J\in\mathfrak{S}} (\det X_J)^2(\det A_J)^2}. \tag{4.16}$$

To evaluate the above amounts, we invoke the following proposition.

PROPOSITION 4.3.

$$\frac{\left|\sum_{i=1}^N r_i\right|}{\sum_{i=1}^N s_i} \le \max_i \frac{|r_i|}{s_i} \tag{4.17}$$

if $s_i > 0$ for all i.

The proof of this propositoin is easy, thus we omit it.

Now we have from (4.16)

$$|y_j| \le \max_{\substack{J \in \mathfrak{S} \\ \det A_J \neq 0}} \left| \frac{\det(A[j])_J}{\det A_J} \right|, \tag{4.18}$$

which proves that $|y_j|$ is bounded by a constant independent of X. □

Next we investigate an important property of affine scaling direction.

LEMMA 4.4. *For given $c \in R^n$ and $A \in R^{m \times n}$, there exists a constant M such that for any positive diagonal matrix $X \in R^{n \times n}$,*

$$\|X(I - P_{AX})Xc\| \le M\|(I - P_{AX})Xc\|^2. \tag{4.19}$$

The proof is essentially due to Tseng and Luo [57].

PROOF. In view of (2.1), we note that $d(X) \triangleq X(I - P_{AX})Xc$ is a solution of

$$\begin{cases} \text{minimize} & (X^2c - d)^t X^{-2}(X^2c - d) \\ \text{subject to} & Ad = 0, \end{cases} \tag{4.20}$$

and $\|(I - P_{AX})Xc\|^2 = c^t d(X)$. Let

$$\Theta(X) \triangleq \frac{\|d(X)\|}{c^t d(X)}. \tag{4.21}$$

We will prove the lemma by contradiction. Suppose contrary that $\Theta(X)$ is not bounded. In this case, we can choose a sequence X^k such that $\Theta(X^k) \to \infty$. Putting $z^k \triangleq |d(X^k)|/c^t d(X^k)$, we can choose an appropriate subsequence k_t of k such that

$$z_j^{k_t} \to \infty \quad \text{if } j \in I, \tag{4.22}$$

$$z_j^{k_t} \to z_j^* \le M_1 < \infty \quad \text{if } j \in I^c. \tag{4.23}$$

Obviously the index set I is not empty. If $j \in I^c$, then there exists a constant M_1 such that $|d_j(X^{k_t})| \le M_1 c^t d(X^{k_t})$ for all k_t.

Now we consider the following system;

$$c^t d = c^t d(X^{k_t}) \tag{4.24}$$

$$Ad = 0 \tag{4.25}$$

$$d_{I^c} = d_{I^c}(X^{k_t}). \tag{4.26}$$

This system has an obvious solution $d(X^{k_t})$, hence there exists a solution $\hat{d}^{k_t}$ such that

$$\|\hat{d}^{k_t}\| \le M_2(c^t d(X^{k_t}) + \|d_{I^c}(X^{k_t})\|) \le M_2(1 + |I^c|M_1)c^t d(X^{k_t}). \tag{4.27}$$

Therefore, $\|\hat{d}^{k_t}\| < |d_j(X^{k_t})|$ for all $j \in I$ and sufficiently large k_t. This implies that

$$(\hat{d}^{k_t})^t X^{-2}(\hat{d}^{k_t}) < (d(X^{k_t}))^t X^{-2}(d(X^{k_t})) \tag{4.28}$$

which produces, combined with (4.24),

$$(X^2c - \hat{d}^{k_t})^t X^{-2}(X^2c - \hat{d}^{k_t}) < (X^2c - d(X^{k_t}))^t X^{-2}(X^2c - d(X^{k_t})). \tag{4.29}$$

This contradicts the assumption that $d(X^{k_t})$ is the optimal solution of (4.20). □

The next lemma is on the property of $I - P_{AX}$, the projection operator onto Null(AX), when some diagonal components of X tend to 0. This is essentially equivalent to Lemma 4.1 of Tsuchiya [58]. Let β^*, $\hat{\beta}_F$ and $\hat{\beta}_{F^c}$ be the optimal solutions of the following optimization problems respectively:

$$\langle Y \rangle \begin{cases} \text{minimize} & \|\beta - y\|^2 \\ \text{subject to} & AX\beta = 0, \end{cases} \tag{4.30}$$

$$\langle Y_F \rangle \begin{cases} \text{minimize} & \|\beta_F - y_F\|^2 \\ \text{subject to} & A_F X_F \beta_F \in \mathrm{Im}(A_{F^c}), \end{cases} \tag{4.31}$$

$$\langle Y_{F^c} \rangle \begin{cases} \text{minimize} & \|\beta_{F^c} - y_{F^c}\|^2 \\ \text{subject to} & A_{F^c} X_{F^c} \beta_{F^c} = 0, \end{cases} \tag{4.32}$$

where $y \in R^q$, $A \in R^{p \times q}$, and $X \in R^{q \times q}$ which is a positive diagonal matrix. Then, we have the following lemma;

LEMMA 4.5 (PROJECTION DECOMPOSITION LEMMA). *There exists some constants M_1, M_2, M_3 and M_4 which is independent of X and y, and satisfy:*

$$\|\beta^*_F - \hat{\beta}_F\| \le M_1 \|X_{F^c}^{-1}\|^2 \|X_F\|^2 \|y_F\| + M_2 \|X_{F^c}^{-1}\| \|X_F\| \|y_{F^c}\|, \tag{4.33}$$

$$\|\beta^*_{F^c} - \hat{\beta}_{F^c}\| \le M_3 \|X_{F^c}^{-1}\| \|X_F\| \|y_F\| + M_4 \|X_{F^c}^{-1}\|^2 \|X_F\|^2 \|y_{F^c}\|. \tag{4.34}$$

It may be worth noting that when $y = Xc$, β^* is the affine scaling direction in the scaled space. In fact, the above lemma will be applied when $X = X^k = \mathrm{diag}(x^k)$ and $x^k \to x^\infty$. In that case, Lemma 4.5 gives us a good approximation of β^* if $\|(X^k_{F^c})^{-1}\| \|X^k_F\| \to 0$. For further explanation and motivation, see page 30, for example. We also point out that the orthogonal complements of

$$\mathcal{Y}_F \triangleq \{\, \beta_F \in R^{|F|} \mid A_F X_F \beta_F \in \mathrm{Im}(A_{F^c}) \,\} \tag{4.35}$$

and

$$\mathcal{Y}_{F^c} \triangleq \{\, \beta_{F^c} \in R^{|F^c|} \mid A_{F^c} X_{F^c} \beta_{F^c} = 0 \,\} \tag{4.36}$$

are

$$\mathcal{Y}_F^\perp = \{\, \beta_F \in R^{|F|} \mid \beta_F = X_F A_F^t y, \quad A_{F^c}^t y = 0 \,\} \tag{4.37}$$

and

$$\mathcal{Y}_{F^c}^\perp = \{\, \beta_{F^c} \in R^{|F^c|} \mid \beta_{F^c} = X_{F^c} A_{F^c}^t y \,\}, \tag{4.38}$$

respectively.

To prove the lemma, we show several lemmas.

LEMMA 4.6. *If $y_{F^c} = 0$, then the optimal solution of*

$$\langle Y_1 \rangle \begin{cases} \text{minimize} & \|\beta_F - \hat{\beta}_F\|^2 + \|\beta_{F^c}\|^2 \\ \text{subject to} & A_F X_F \beta_F + A_{F^c} X_{F^c} \beta_{F^c} = 0 \end{cases} \tag{4.39}$$

is $(\beta^*_F, \beta^*_{F^c})$.

PROOF. It is easy to see that

$$\|\beta_F - \hat{\beta}_F\|^2 = \|\beta_F - y_F\|^2 - \|\hat{\beta}_F - y_F\|^2 - 2(\beta_F - \hat{\beta}_F)'(\hat{\beta}_F - y_F). \tag{4.40}$$

Here, the last term is always 0 since $\beta_F - \hat{\beta}_F \in \mathcal{Y}_F$ and $\hat{\beta}_F - y_F \in \mathcal{Y}_F^{\perp}$. Then $\langle Y_1 \rangle$ is equivalent to:

$$\langle Y_2 \rangle \begin{cases} \text{minimize} & \|\beta_F - y_F\|^2 + \|\beta_{F^c}\|^2 \\ \text{subject to} & A_F X_F \beta_F + A_{F^c} X_{F^c} \beta_{F^c} = 0, \end{cases} \tag{4.41}$$

which is identical to $\langle Y \rangle$. □

COROLLARY 4.7. *If $y_{F^c} = 0$, then $(\beta_F^* - \hat{\beta}_F, \beta_{F^c}^*)$ is the optimal solution of*

$$\langle Y_3 \rangle \begin{cases} \text{minimize} & \|\beta_F\|^2 + \|\beta_{F^c}\|^2 \\ \text{subject to} & A_F X_F \beta_F + A_{F^c} X_{F^c} \beta_{F^c} = -A_F X_F \hat{\beta}_F. \end{cases} \tag{4.42}$$

LEMMA 4.8. *If $y_{F^c} = 0$, then*

$$\|\beta_{F^c}^*\| \leq M \|X_{F^c}^{-1}\| \|X_F\| \|\beta_F^*\| \leq M \|X_{F^c}^{-1}\| \|X_F\| \|y_F\| \tag{4.43}$$

PROOF. It is easy to see that $\beta_{F^c}^*$ is the optimal solution of

$$\langle Y_4 \rangle \begin{cases} \text{minimize} & \|\beta_{F^c}\|^2 \\ \text{subject to} & A_{F^c} X_{F^c} \beta_{F^c} = -A_F X_F \beta_F^*. \end{cases} \tag{4.44}$$

On the other hand, there exists a solution α' of $A_{F^c}\alpha = -A_F X_F \beta_F^*$ such that $\|\alpha'\| \leq M \|X_F\| \|\beta_F^*\|$ holds. If we set $\beta'_{F^c} \triangleq X_{F^c}^{-1}\alpha'$, then we have

$$\|\beta_{F^c}^*\| \leq \|\beta'_{F^c}\| = \|X_{F^c}^{-1}\alpha'\| \leq M \|X_{F^c}^{-1}\| \|X_F\| \|\beta_F^*\|. \tag{4.45}$$

Recalling that $\|\beta_F^*\| \leq \|y_F\|$, we obtain the lemma. □

LEMMA 4.9. *If $y_{F^c} = 0$, then*

$$\|\beta_F^* - \hat{\beta}_F\| \leq M \|X_{F^c}^{-1}\|^2 \|X_F\|^2 \|y_F\| \tag{4.46}$$

PROOF. Consider the following linear equation

$$A_{F^c} X_{F^c} (\beta_{F^c} - \beta_{F^c}^*) = -A_F X_F (\hat{\beta}_F - \beta_F^*). \tag{4.47}$$

We see that this system has a solution in consideration of the definition of β^* and $\hat{\beta}_F$. Hence there exists a solution $\breve{\beta}_{F^c}$ of (4.47) such that

$$\|\breve{\beta}_{F^c} - \beta_{F^c}^*\| \leq M \|X_{F^c}^{-1}\| \|X_F\| \|\hat{\beta}_F - \beta_F^*\|. \tag{4.48}$$

Then, Lemma 4.8, (4.48) and the fact that $\|\hat{\beta}_F - \beta_F^*\| \leq 2\|y_F\|$ imply that

$$\|\breve{\beta}_{F^c}\| \leq M' \|X_{F^c}^{-1}\| \|X_F\| \|y_F\|. \tag{4.49}$$

On the other hand, noting that (4.47) is actually

$$A_{F^c} X_{F^c} \beta_{F^c} = -A_F X_F \hat{\beta}_F, \tag{4.50}$$

which implies that $(0, \breve{\beta}_{F^c})$ is feasible for $\langle Y_3 \rangle$. So we have

$$\|\beta_F^* - \hat{\beta}_F\|^2 + \|\beta_{F^c}^*\|^2 \leq \|\breve{\beta}_{F^c}\|^2. \tag{4.51}$$

Now we see

(4.52)
$$\begin{aligned}\|\beta^*_F - \hat\beta_F\|^2 &\le \|\check\beta_{F^c}\|^2 - \|\beta^*_{F^c}\|^2\\ &= (\check\beta_{F^c} - \beta^*_{F^c})^t(\check\beta_{F^c} + \beta^*_{F^c})\\ &\le \|\check\beta_{F^c} - \beta^*_{F^c}\|\|\check\beta_{F^c} + \beta^*_{F^c}\|\\ &\le M\|X^{-1}_{F^c}\|\|X_F\|\|\hat\beta_F - \beta^*_F\|\|\check\beta_{F^c} + \beta^*_{F^c}\| \quad \text{(Use (4.48))}\\ &\le MM'\|X^{-1}_{F^c}\|^2\|X_F\|^2\|\hat\beta_F - \beta^*_F\|\|y_F\|\end{aligned}$$

where in the last inequality, we use Lemma 4.8 and (4.49). Then dividing the both sides by $\|\hat\beta_F - \beta^*_F\|$, we obtain the lemma. □

LEMMA 4.10. *If $y_F = 0$, then the optimal solution of*

$$(4.53)\qquad \langle Y_5\rangle \begin{cases} \text{minimize} & \|\beta_{F^c} - \hat\beta_{F^c}\|^2\\ \text{subject to} & A_{F^c}X_{F^c}\beta_{F^c} = -A_FX_F\beta^*_F\end{cases}$$

is $\beta^*_{F^c}$.

PROOF. Since $\hat\beta_{F^c}$ is the optimal solution of $\langle Y_{F^c}\rangle$, by setting up the Lagrangian, we see

$$(4.54)\qquad \hat\beta_{F^c} - y_{F^c} - X_{F^c}A^t_{F^c}\lambda = 0$$

holds for some λ. On the other hand, $\beta^*_{F^c}$ satisfies

$$(4.55)\qquad \beta^*_{F^c} - y_{F^c} - X_{F^c}A^t_{F^c}\lambda' = 0$$

for some λ' since $\beta^*_{F^c}$ is the optimal solution of

$$(4.56)\qquad \langle Y_6\rangle \begin{cases} \text{minimize} & \|\beta_{F^c} - y_{F^c}\|^2\\ \text{subject to} & A_{F^c}X_{F^c}\beta_{F^c} = -A_FX_F\beta^*_F.\end{cases}$$

Comparing (4.54) and (4.55), we have

$$(4.57)\qquad \beta^*_{F^c} - \hat\beta_{F^c} - X_{F^c}A^t_{F^c}(\lambda' - \lambda) = 0$$

which is the Lagrangian equation for $\langle Y_5\rangle$. Now the lemma readily follows since $\beta^*_{F^c}$ is feasible for $\langle Y_5\rangle$. □

LEMMA 4.11. *If $y_F = 0$, then*

$$(4.58)\qquad \|\beta^*_F\| \le M\|X^{-1}_{F^c}\|\|X_F\|\|y_{F^c}\|$$

PROOF. Let P be the projection matrix onto $\mathrm{Null}(AX)$, and decompose it corresponding to the index sets F and F^c, namely,

$$(4.59)\qquad P = \begin{pmatrix} P_a & P_b\\ P^t_b & P_c\end{pmatrix}$$

where $P_a \in R^{|F|\times|F|}$, $P_b \in R^{|F|\times|F^c|}$ and $P_c \in R^{|F^c|\times|F^c|}$. Obviously, $\beta^* = Py$. If $y_{F^c} = 0$, then from Lemma 4.8,

$$(4.60)\qquad \|\beta^*_{F^c}\| = \|P^t_by_F\| \le M\|X^{-1}_{F^c}\|\|X_F\|\|y_F\|$$

holds for any y_F. This implies $\|P_b^t\| \le M\|X_{F^c}^{-1}\|\|X_F\|$. So, if $y_F = 0$, then

$$\|\beta_F^*\| = \|P_b y_{F^c}\| \le M\|X_{F^c}^{-1}\|\|X_F\|\|y_{F^c}\| \tag{4.61}$$

and we completes the proof. □

LEMMA 4.12. *If* $y_F = 0$, *then*

$$\|\beta_{F^c}^* - \hat{\beta}_{F^c}\| \le M\|X_{F^c}^{-1}\|^2\|X_F\|^2\|y_{F^c}\|. \tag{4.62}$$

PROOF. It is easy to see that from Lemma 4.10,

$$\|\beta_{F^c}^* - \hat{\beta}_{F^c}\| \le M\|X_{F^c}^{-1}\|\|X_F\|\|\beta_F^*\|. \tag{4.63}$$

Now the lemma easily follows from Lemma 4.11. □

Now we have all to prove the Projection Decomposition Lemma.

PROOF OF LEMMA 4.5. Since

$$\begin{pmatrix} \beta_F^* \\ \beta_{F^c}^* \end{pmatrix} = P\begin{pmatrix} y_F \\ 0 \end{pmatrix} + P\begin{pmatrix} 0 \\ y_{F^c} \end{pmatrix} \tag{4.64}$$

where P is the projection operator onto Null(AX), the relation (4.33) follows from Lemma 4.9 and Lemma 4.11, whereas (4.34) from Lemma 4.8 and Lemma 4.12. □

In the last of this section, we show one more proposition.

PROPOSITION 4.13. *If* $y_{F^c} = 0$, *then*

$$0 \le \|\hat{\beta}_F\| - \|\beta^*\| \le M\|X_{F^c}^{-1}\|^2\|X_F\|^2\|y_F\| \tag{4.65}$$

PROOF. Lemma 4.6 implies that β^* is a projection of $(\hat{\beta}_F, 0)$. Hence the left inequality readily follows. Further, we have

$$\|\beta^*\|^2 = \hat{\beta}_F^t\beta_F^*. \tag{4.66}$$

Thus we see that

$$\begin{aligned} \|\hat{\beta}_F\|^2 - \|\beta^*\|^2 &= (\|\hat{\beta}_F\| - \|\beta^*\|)(\|\hat{\beta}_F\| + \|\beta^*\|) \\ &= \hat{\beta}_F^t(\hat{\beta}_F - \beta_F^*) \le \|\hat{\beta}_F\|\|\hat{\beta}_F - \beta_F^*\| \\ &\le M\|y_F\|\|\hat{\beta}_F\|\|X_{F^c}^{-1}\|^2\|X_F\|^2. \end{aligned} \tag{4.67}$$

Now we have

$$\begin{aligned} \|\hat{\beta}_F\| - \|\beta^*\| &\le M\|y_F\|\|X_{F^c}^{-1}\|^2\|X_F\|^2\frac{\|\hat{\beta}_F\|}{\|\hat{\beta}_F\| + \|\beta^*\|} \\ &\le M\|y_F\|\|X_{F^c}^{-1}\|^2\|X_F\|^2, \end{aligned} \tag{4.68}$$

which completes the proof. □

CHAPTER II

Global Convergence of a Long-Step Affine Scaling Method

1. Introduction

In this chapter, we deal with a long-step affine scaling method for LP defined in Section 1.2 applied to the primal form LP problem $\langle P \rangle$;

$$
\text{(1.1)} \qquad \langle P \rangle \begin{cases} \text{minimize} & c^t x \\ \text{subject to} & Ax = b, \quad x \geq 0, \end{cases}
$$

where $c, x \in R^n$, $b \in R^m$ and $A \in R^{m \times n}$. We assume Asssumption I-1 and Asssumption I-2 in Section I-2 , but do not make any kind of nondegeneracy assumption nor full rankness of A.

The main theorem proved in this chapter is as follows.

THEOREM 1.1. *Let us apply the long-step affine scaling method on page 18 with step-size*

$$
\text{(1.2)} \qquad 0 < \lambda_{\min} \leq \lambda^k \leq 2/3
$$

where $\lambda_{\min}$ is a constant independent of iteration number k. Suppose that x^k is an infinite sequence and $c^t x^k$ is bounded below. Then the primal sequence x^k converges to a relative interior point of the optimal face of $\langle P \rangle$, while the dual sequence s^k converges to the analytic center of the optimal face of $\langle DP \rangle$.

This chapter is devoted to the proof of Theorem 1.1, which is substantially the same as that of Tsuchiya and Muramatsu [61]. While the dual form LP problem was treated in [61], here we deal with the primal form LP problem $\langle P \rangle$. But we have already seen Section I-2 that there is essentially no difference between the primal and the dual affine scaling methods.

To start the proof, we introduce the following lemma which was first proved by Tseng and Luo [57].

LEMMA 1.2. *If we choose the step-size as*

$$
\text{(1.3)} \qquad 0 < \lambda_{\min} \leq \lambda^k \leq \lambda_{\max} < 1,
$$

and if $c^t x^k$ is bounded below, then the primal sequence x^k converges to a point x^∞ which is a relative interior point of a dual-degenerate face satisfying

$$\frac{c^t x^k - c^t x^\infty}{\|x^k_E\|} \geq \delta > 0 \tag{1.4}$$

where E determines a dual-degenerate face and δ is a positive constant.

We note that the step-size choice (1.3) covers (1.2).

PROOF. Since

$$\begin{aligned} -\infty < \sum_{k=0}^{\infty}(c^t x^{k+1} - c^t x^k) &= -\sum_{k=0}^{\infty} \lambda^k \frac{c^t d(x^k)}{\sigma((X^k)^{-1} d(x^k))} \\ &= -\sum_{k=0}^{\infty} \lambda^k \frac{\|(X^k)^{-1} d(x^k)\|^2}{\sigma((X^k)^{-1} d(x^k))}, \end{aligned} \tag{1.5}$$

we have

$$\sum_{k=0}^{\infty} \lambda^k \frac{\|(X^k)^{-1} d(x^k)\|^2}{\sigma((X^k)^{-1} d(x^k))} < \infty. \tag{1.6}$$

Now we see that

$$\begin{aligned} \sum_{k=0}^{\infty} \|x^{k+1} - x^k\| &= \sum_{k=0}^{\infty} \lambda^k \frac{\|d(x^k)\|}{\sigma((X^k)^{-1} d(x^k))} \\ &\leq \sum_{k=0}^{\infty} \lambda^k \frac{M_1 \|(X^k)^{-1} d(x^k)\|^2}{\sigma((X^k)^{-1} d(x^k))} \quad \text{(Use Lemma 1.4.4)} \\ &< \infty \quad \text{(Use (1.6))}. \end{aligned} \tag{1.7}$$

Therefore, x^k converges. Let x^∞ be the limit point and E be the index set such that $x_j^\infty = 0$ if $j \in E$ and $x_j^\infty > 0$ if $j \in E^c$. Again from (1.6), we note that

$$\infty > \sum_{k=0}^{\infty} \lambda^k \frac{\|(X^k)^{-1} d(x^k)\|^2}{\sigma((X^k)^{-1} d(x^k))} \geq \sum_{k=0}^{\infty} \lambda_{\min} \|(X^k)^{-1} d(x^k)\|, \tag{1.8}$$

thus

$$\|(X^k)^{-1} d(x^k)\| \to 0. \tag{1.9}$$

In view of (1.2.15), (1.9) produces

$$\|X^k s^k\| \to 0. \tag{1.10}$$

Due to Lemma 1.4.1, the dual estimate s^k is bounded, hence has an accumulation point s^*. Now from (1.10), we have

$$s^*_{E^c} = 0, \tag{1.11}$$

which implies, by Lemma 1.1.5, that the face determined by E is dual-degenerate.

Finally, we see that

$$
\begin{aligned}
\|x_E^k\| &\le \|x^k - x^\infty\| \le \sum_{l=k}^{\infty} \|x^{l+1} - x^l\| \\
&\le \sum_{l=k}^{\infty} \lambda^l \frac{\|d(x^l)\|}{\sigma((X^l)^{-1}d(x^l))} \\
&\le \sum_{l=k}^{\infty} \lambda^l M_1 \frac{c^t d(x^l)}{\sigma((X^l)^{-1}d(x^l))} \\
&= M_1(c^t x^k - c^t x^\infty),
\end{aligned}
\tag{1.12}
$$

which completes the proof. □

The rest of this chapter is organized as follows. In Section 2, we investigate asymptotic property of the scaled direction $(X^k)^{-1}d^k$, and introduce the local Karmarkar potential function associated with Face(E). In Section 3, we make a few preliminary observation on the primal sequence. In Section 4, we prove a key inequality which will be used to estimate the difference of the potential function. In Section 5, we prove that x^∞ is a relative interior point of the optimal face, while in Section 6, we prove the global convergence of the dual sequence.

2. Asymptotic Direction

We call

$$\alpha^k \triangleq (X^k)^{-1} d^k \tag{2.1}$$

a *scaled direction*. Due to Lemma 1.1.5, we can choose some $\hat{s}$ such that

$$\hat{s} = c - A^t \hat{y}, \qquad \hat{s}_{E^c} = 0, \tag{2.2}$$

and that

$$c^t x - c^t x^\infty = \hat{s}_E^t x_E \qquad \text{for all } x \in \text{Feas}\langle P \rangle. \tag{2.3}$$

We have the following lemma.

LEMMA 2.1. *α^k is the optimal solution of*

$$\left\{ \begin{array}{ll} \text{minimize} & \|\alpha - X^k \hat{s}\|^2 \\ \text{subject to} & AX^k \alpha = 0. \end{array} \right. \tag{2.4}$$

PROOF. Since

$$X^k c - X^k \hat{s} = X^k A^t \hat{y} \in \text{Im}(X^k A^t), \tag{2.5}$$

the lemma readily follows if we recall α^k is the projection of $X^k c$ onto Null(AX^k). □

Now imagine that x^k is converging to x^∞. In the last stage of the convergence, E^c-components of x^k is relatively much greater than E-components. In the light of (2.1), we may suppose that the inequality constraints $x_{E^c} \ge 0$ in $\langle P \rangle$ have very little effects on α^k. If we completely ignore the inequality constraints, we get

$$\langle P' \rangle \left\{ \begin{array}{ll} \text{minimize} & \hat{s}_E^t x_E \\ \text{subject to} & Ax = b, \quad x_E \ge 0. \end{array} \right. \tag{2.6}$$

Noting that $A_{E^c}x^\infty_{E^c} = b$, we can regard $\langle P' \rangle$ as the following homogeneous LP problem;

$$\text{(2.7)} \qquad \langle H_E \rangle \begin{cases} \text{minimize} & \hat{s}^t_E x_E \\ \text{subject to} & A_E x_E \in \operatorname{Im}(A_{E^c}), \quad x_E \geq 0. \end{cases}$$

Let $\hat{\alpha}^k_E$ be the scaled affine scaling direction for $\langle H_E \rangle$ which is the optimal solution of

$$\text{(2.8)} \qquad \begin{cases} \text{minimize} & \|\alpha_E - X^k_E \hat{s}_E\|^2 \\ \text{subject to} & \alpha_E \in \mathcal{L}_E, \end{cases}$$

where $\mathcal{L}_E \triangleq \{ \alpha_E \in R^{|E|} \mid A_E X^k_E \alpha_E \in \operatorname{Im}(A_{E^c}) \}$. Then we have the following lemma by using Projection Decomposition Lemma (Lemma 1.4.5).

LEMMA 2.2. *There exist some constants M_1 and M_2 such that*

$$\text{(2.9)} \qquad \|\alpha^k_E - \hat{\alpha}^k_E\| \leq M_1 \|X^k_E\|^2 \|\hat{\alpha}^k_E\| \leq M_1 \|\hat{s}_E\| \|X^k_E\|^3$$

$$\text{(2.10)} \qquad \|\alpha^k_{E^c}\| \leq M_2 \|X^k_E\| \|\hat{\alpha}^k_E\| \leq M_2 \|\hat{s}_E\| \|X^k_E\|^2.$$

PROOF. From (2.8), $\|\hat{\alpha}^k_E\| \leq \|X^k_E \hat{s}_E\| \leq \|X^k_E\| \|\hat{s}_E\|$ follows. Due to Lemma 2.1, we can apply Lemma 1.4.5 by substituting $y = X^k \hat{s}$. Noting that the optimal solution of $\langle Y_{F^c} \rangle$ is obviously 0 since $\hat{s}_{E^c} = 0$, and that $\|(X^k_{E^c})^{-1}\|$ is bounded, we obtain the lemma immediately. □

The above lemma means that $(\hat{\alpha}^k_E, 0)$ is a good approximation for α^k if k is sufficiently large. Recall that an affine scaling method applied to the homogeneous LP problem $\langle H_E \rangle$ is essentially Karmarkar's projective scaling method (See Section I-3). Therefore, we may expect that the behavior of x^k_E in large k can be analyzed by using the Karmarkar potential function for $\langle H_E \rangle$;

$$\text{(2.11)} \qquad \begin{aligned} f_E(x^k_E) &\triangleq |E| \log \hat{s}^t_E x^k_E - \sum_{j \in E} \log x^k_j, \\ &= |E| \log(c^t x^k - c^t x^\infty) - \sum_{j \in E} \log x^k_j, \end{aligned}$$

which is called the *local Karmarkar potential function* associated with E. In fact, this function plays very important role to prove the main theorem.

Next we prove a few properties of $\hat{\alpha}^k_E$.

LEMMA 2.3.

$$\text{(2.12)} \qquad \hat{s}^t_E X^k_E \hat{\alpha}^k_E = \|\hat{\alpha}^k_E\|^2,$$

$$\text{(2.13)} \qquad e^t \hat{\alpha}^k_E = \hat{s}^t_E x^k_E.$$

PROOF. If we denote by π an orthogonal projection operator onto $\mathcal{L}_E$, then

$$\text{(2.14)} \qquad \hat{s}^t_E X^k_E \hat{\alpha}^k_E = \hat{s}^t_E X^k_E \pi X^k_E \hat{s}_E = \|\pi X^k_E \hat{s}_E\|^2 = \|\hat{\alpha}^k_E\|^2$$

follows. Further, we have

$$\text{(2.15)} \qquad e^t \hat{\alpha}^k_E = e^t \pi X^k_E \hat{s}_E = e^t X^k_E \hat{s}_E = \hat{s}^t_E x^k_E$$

since $e \in \mathcal{L}_E$. □

3. Property of the Primal Sequence

We see some properties of the primal sequence in this section when we take the step-size

$$0 < \lambda_{\min} \le \lambda^k \le \lambda_{\max} < 1, \tag{3.1}$$

which includes the step-size choice of the main theorem.

We first introduce two new variables;

$$\beta^k \triangleq \frac{\alpha^k}{\hat{s}^t x^k} = \frac{\alpha^k}{c^t(x^k - x^\infty)}, \tag{3.2}$$

$$\hat{\beta}^k_E \triangleq \frac{\hat{\alpha}^k_E}{\hat{s}^t x^k} = \frac{\hat{\alpha}^k_E}{c^t(x^k - x^\infty)}. \tag{3.3}$$

From Lemma 1.2 and Lemma 2.2, we see that

$$\|\beta^k_E - \hat{\beta}^k_E\| = \frac{\|\alpha^k_E - \hat{\alpha}^k_E\|}{\hat{s}^t_E x^k_E} \le \frac{M_1 \|X^k_E\|^3 \|\hat{s}_E\|}{\hat{s}^t_E x^k_E} \le M_1' \|X^k_E\|^2 \tag{3.4}$$

and

$$\|\beta^k_{E^c}\| = \frac{\|\alpha^k_{E^c}\|}{\hat{s}^t_E x^k_E} \le \frac{M_2 \|X^k_E\|^2 \|\hat{s}_E\|}{\hat{s}^t_E x^k_E} \le M_2' \|X^k_E\|. \tag{3.5}$$

Then we have the following lemma.

LEMMA 3.1.

$$\frac{x^{k+1}_j}{x^k_j} = 1 - \lambda^k \frac{\beta^k_j}{\sigma(\beta^k)} \ge 1 - \lambda_{\max} > 0. \tag{3.6}$$

for all $j = 1, \ldots, n$, and

$$\frac{c^t(x^{k+1} - x^\infty)}{c^t(x^k - x^\infty)} = 1 - \lambda^k \frac{\|\beta^k\|^2}{\sigma(\beta^k)} \ge \delta > 0, \tag{3.7}$$

for a positive constant δ.

Note that $\sigma(\beta^k) > 0$ because otherwise, the objective function value goes to $-\infty$.

PROOF. We have

$$\begin{aligned} \frac{x^{k+1}_j}{x^k_j} &= 1 - \lambda^k \frac{\alpha^k_j}{\sigma(\alpha^k)} = 1 - \lambda^k \frac{\beta^k_j}{\sigma(\beta^k)} \\ &\ge 1 - \lambda^k \ge 1 - \lambda_{\max} > 0 \end{aligned} \tag{3.8}$$

which proves (3.6).

Since

$$c^t d^k = c^t X^k (I - P_{AX^k}) X^k c = \|\alpha^k\|^2, \tag{3.9}$$

we have

$$\frac{c^t(x^{k+1} - x^\infty)}{c^t(x^k - x^\infty)} = 1 - \frac{\lambda^k \|\alpha^k\|^2}{c^t(x^k - x^\infty)\sigma(\alpha^k)} = 1 - \lambda^k \frac{\|\beta^k\|^2}{\sigma(\beta^k)} \tag{3.10}$$

which proves the first equality of (3.7). To prove the inequality, we also have

$$
\begin{aligned}
(3.11) \qquad c^t(x^{k+1} - x^\infty) &= \hat{s}_E^t x_E^{k+1} \\
&\geq M\|x_E^{k+1}\| \qquad \text{(Use Lemma 1.2)} \\
&\geq M(1-\lambda_{\max})\|x_E^k\| \qquad \text{(Use (3.6))} \\
&\geq M(1-\lambda_{\max})\frac{\hat{s}_E^t x_E^k}{\|\hat{s}_E\|} \\
&= \frac{M(1-\lambda_{\max})}{\|\hat{s}_E\|}(c^t x^k - c^t x^\infty)
\end{aligned}
$$

and this completes the proof. □

Now we can calculate the difference of the local Karmarkar potential function defined in (2.11) at the k-th iteration which is

$$
\begin{aligned}
(3.12) \qquad \Delta^k &\triangleq f_E(x_E^{k+1}) - f_E(x_E^k) \\
&= |E| \log\left\{1 - \lambda^k \frac{\|\beta^k\|^2}{\sigma(\beta^k)}\right\} - \sum_{j\in E} \log\left\{1 - \lambda^k \frac{\beta_j^k}{\sigma(\beta^k)}\right\}
\end{aligned}
$$

due to Lemma 3.1. To make a bound of Δ^k, we first recall $\langle H_E\rangle$. Given a point $x_E^k \in \text{Feas}\langle H_E\rangle$, if we take a step λ^k of the way to the boundary of $\text{Feas}\langle H_E\rangle$ and if f_E is well-defined on the destination $\hat{x}_E^k$, then the difference of the local Karmarkar potential function can be written as

(3.13)

$$
\begin{aligned}
\hat{\Delta}^k &\triangleq f_E(\hat{x}_E^k) - f_E(x_E^k) \\
&= |E| \log\left\{1 - \frac{\lambda^k \hat{s}_E^t X_E^k \hat{\alpha}_E^k}{\sigma(\hat{\alpha}_E^k)\hat{s}_E^t x_E^k}\right\} - \sum_{j\in E} \log\left\{1 - \frac{\lambda^k \hat{\alpha}_j^k}{\sigma(\hat{\alpha}_E^k)}\right\} \\
&= |E| \log\left\{1 - \frac{\lambda^k \|\hat{\beta}_E^k\|^2}{\sigma(\hat{\beta}_E^k)}\right\} - \sum_{j\in E} \log\left\{1 - \frac{\lambda^k \hat{\beta}_j^k}{\sigma(\hat{\beta}_E^k)}\right\} \qquad \text{(Use (2.12))}.
\end{aligned}
$$

We expect that $\hat{\Delta}^k$ is a good approximation for Δ^k. In general, f_E is not well-defined on $\hat{x}_E$. We however, prove the following lemma.

LEMMA 3.2. *There exist a number $\tilde{K}$ and a positive constant δ such that for $k \geq \tilde{K}$*

$$(3.14) \qquad 1 - \frac{\lambda^k \|\hat{\beta}_E^k\|^2}{\sigma(\hat{\beta}_E^k)} \geq \delta > 0,$$

and

$$(3.15) \qquad \sum_{k\geq \tilde{K}} |\Delta^k - \hat{\Delta}^k| < \infty.$$

PROOF. In view of (3.7), if we can prove

$$(3.16) \qquad \left|\frac{\|\beta^k\|^2}{\sigma(\beta^k)} - \frac{\|\hat{\beta}_E^k\|^2}{\sigma(\hat{\beta}_E^k)}\right| \to 0,$$

then (3.14) readily follows. To prove (3.16), we first observe that $\|\beta^k\|$ is bounded. In fact, due to (3.7),

$$1 - \lambda_{\min}\|\beta^k\| \geq 1 - \lambda^k \frac{\|\beta^k\|^2}{\sigma(\beta^k)} > 0 \tag{3.17}$$

hence we have

$$\|\beta^k\| \leq (\lambda_{\min})^{-1}. \tag{3.18}$$

In view of (3.4), $\hat{\beta}_E^k$ is also bounded.

Then by using (3.4) and (3.5), we have

$$\begin{aligned} \big|\|\beta^k\|^2 - \|\hat{\beta}_E^k\|^2\big| &= \big|\|\beta_E^k\|^2 - \|\hat{\beta}_E^k\|^2 + \|\beta_{E^c}^k\|^2\big| \\ &\leq \|\beta_E^k + \hat{\beta}_E^k\|\|\beta_E^k - \hat{\beta}_E^k\| + \|\beta_{E^c}^k\|^2 \\ &\leq M_1\|X_E^k\|^2. \end{aligned} \tag{3.19}$$

On the other hand, from

$$e^t\hat{\beta}_E^k = \frac{e^t\hat{\alpha}_E^k}{\hat{s}_E^t x_E^k} = 1 \qquad \text{(Use (2.13))}, \tag{3.20}$$

$$\sigma(\hat{\beta}_E^k) \geq 1/|E| \tag{3.21}$$

follows. This implies that there exists a number K_1 such that $\sigma(\beta_E^k)$ is greater than a positive constant δ if $k \geq K_1$, while $\sigma(\beta_{E^c}^k) < \delta$ due to (3.5). Hence, there exist a number $\tilde{K}$ and a constant M_2 such that for $k \geq \tilde{K}$

$$|\sigma(\beta^k) - \sigma(\hat{\beta}_E^k)| = |\sigma(\beta_E^k) - \sigma(\hat{\beta}_E^k)| \leq M_2\|X_E^k\|^2. \tag{3.22}$$

In consideration of (3.21), we have for $k \geq \tilde{K}$,

$$\begin{aligned} \frac{\|\beta^k\|^2}{\sigma(\beta^k)} - \frac{\|\hat{\beta}_E^k\|^2}{\sigma(\hat{\beta}_E^k)} &= \frac{\|\hat{\beta}_E^k\|^2 + \mathcal{O}(\|X_E^k\|^2)}{\sigma(\hat{\beta}_E^k) + \mathcal{O}(\|X_E^k\|^2)} - \frac{\|\hat{\beta}_E^k\|^2}{\sigma(\hat{\beta}_E^k)} \\ &= \frac{\|\hat{\beta}_E^k\|^2}{\sigma(\hat{\beta}_E^k)}(1 + \mathcal{O}(\|X_E^k\|^2)) + \mathcal{O}(\|X_E^k\|^2) - \frac{\|\hat{\beta}_E^k\|^2}{\sigma(\hat{\beta}_E^k)} \\ &= \mathcal{O}(\|X_E^k\|^2). \end{aligned} \tag{3.23}$$

Since $\|X_E^k\| \to 0$, we have (3.16) which proves (3.14). Similarly, we easily have for $j \in E$,

$$\left|\frac{\beta_j^k}{\sigma(\beta^k)} - \frac{\hat{\beta}_j^k}{\sigma(\hat{\beta}_E^k)}\right| = \mathcal{O}(\|X_E^k\|^2). \tag{3.24}$$

From (3.12) and (3.13), we have for $k \geq \tilde{K}$,

(3.25)

$$\Delta^k - \hat{\Delta}^k = |E| \log\left\{1 - \lambda^k \frac{\|\beta^k\|^2}{\sigma(\beta^k)}\right\} - |E| \log\left\{1 - \frac{\lambda^k \|\hat{\beta}^k_E\|^2}{\sigma(\hat{\beta}^k_E)}\right\}$$
$$- \left(\sum_{j\in E} \log\left\{1 - \lambda^k \frac{\beta^k_j}{\sigma(\beta^k)}\right\} - \sum_{j\in E} \log\left\{1 - \frac{\lambda^k \hat{\beta}^k_j}{\sigma(\hat{\beta}^k_E)}\right\}\right).$$

Now we invoke the following proposition.

Proposition 3.3. *If two real sequences $\{p^k\}$ and $\{q^k\}$ satisfy*

$$1 - p^k \geq \delta_1 > 0, \tag{3.26}$$
$$1 - q^k \geq \delta_2 > 0, \tag{3.27}$$
$$|p^k - q^k| \to 0, \tag{3.28}$$

then,

$$|\log(1 - p^k) - \log(1 - q^k)| = \mathcal{O}(|p^k - q^k|). \tag{3.29}$$

In view of Lemma 3.1, (3.14), (3.23), and (3.24), we apply Proposition 3.3 to (3.25) to have

$$|\Delta^k - \hat{\Delta}^k| = \mathcal{O}(\|X^k_E\|^2) \tag{3.30}$$

for $k \geq \tilde{K}$. Note that from (3.7),

$$\sum_{k=\tilde{K}}^{\infty} (c^t x^k - c^t x^\infty) < \infty \tag{3.31}$$

follows. Then we see from Lemma 1.2,

$$\sum_{k=\tilde{K}}^{\infty} |\Delta^k - \hat{\Delta}^k| \leq M_3 \sum_{k=\tilde{K}}^{\infty} \|x^k_E\|^2 \leq M_4 \sum_{k=\tilde{K}}^{\infty} (c^t x^k - c^t x^\infty)^2 < \infty, \tag{3.32}$$

which proves (3.15). This completes the proof. □

Now, by using the same $\tilde{K}$ in Lemma 3.2, we see for $k \geq \tilde{K}$,

(3.33)

$$f_E(x^k_E) = f_E(x^{\tilde{K}}_E) + \sum_{l=\tilde{K}}^{k-1} \Delta^k$$
$$= f_E(x^{\tilde{K}}_E) + \sum_{l=\tilde{K}}^{k-1} \hat{\Delta}^k - \sum_{l=\tilde{K}}^{k-1} (\hat{\Delta}^k - \Delta^k).$$

Lemma 3.2 gives us a bound for $\sum_{l=\tilde{K}}^{k-1} (\hat{\Delta}^k - \Delta^k)$. Now our next target is to bound $\hat{\Delta}^k$.

4. The Key Inequality

In this section, we show an inequality used for bounding $\hat{\Delta}^k$. This inequality is a source of the number 2/3 of the maximal step-size.

LEMMA 4.1. *Let*

$$H_r(\tilde{\beta},\tilde{\nu}) \triangleq r\log(1-\tilde{\nu}\|\tilde{\beta}\|^2) - \sum_{j=1}^{r}\log(1-\tilde{\nu}\tilde{\beta}_j) \tag{4.1}$$

and

$$T_r(\tilde{\beta},\tilde{\nu}) \triangleq \frac{r\tilde{\nu}}{r-\tilde{\nu}}\|\tilde{\beta}-\frac{e}{r}\|\left\{-r+\frac{\tilde{\nu}}{2(1-\tilde{\nu}\sigma(\tilde{\beta}))}\right\}, \tag{4.2}$$

which are well-defined on the set

$$\Omega_0 \triangleq \left\{(\tilde{\beta},\tilde{\nu}) \in R^{r+1} \,|\, e^t\tilde{\beta}=1, \quad \tilde{\nu}\|\tilde{\beta}\|^2<1, \quad 0<\tilde{\nu}\sigma(\tilde{\beta})<1\right\}. \tag{4.3}$$

Then, H_r is bounded as follows over Ω_0:

$$H_r(\tilde{\beta},\tilde{\nu}) \le T_r(\tilde{\beta},\tilde{\nu}). \tag{4.4}$$

Furthermore, if

$$(\tilde{\beta},\tilde{\nu}) \in \Omega_1 \triangleq \left\{(\tilde{\beta},\tilde{\nu}) \in \Omega_0 \,|\, \tilde{\nu}\tilde{\beta} \le 2/3\right\}, \tag{4.5}$$

then we have

$$T_r(\tilde{\beta},\tilde{\nu}) \le 0 \tag{4.6}$$

where $T_r(\tilde{\beta},\tilde{\nu})=0$ holds if and only if $\tilde{\beta}=e/r$.

To prove the above lemma, we introduce the following inequality for summation of logarithm.

PROPOSITION 4.2. *Let h be a vector in R^r such that $h<e$. Then,*

$$-\sum_{j=1}^{r}\log(1-h_j) \le e^t h + \frac{\|h\|^2}{2(1-\sigma(h))}. \tag{4.7}$$

PROOF.

$$\begin{aligned}
(4.8) \qquad \sum_{i=1}^{r}\log(1-h_i) &= \sum_{i:h_i>-\sigma(h)}(-h_i-\frac{h_i^2}{2}-\frac{h_i^3}{3}-\cdots) \\
&\quad + \sum_{i:h_i\le-\sigma(h)}\log(1-h_i) \\
&\ge \sum_{i:h_i>-\sigma(h)}(-h_i-\frac{|h_i|^2}{2}-\frac{|h_i|^3}{2}-\cdots) \\
&\quad + \sum_{i:h_i\le-\sigma(h)}\log(1-h_i) \\
&\ge \sum_{i:h_i>-\sigma(h)}(-h_i-\frac{h_i^2}{2(1-|h_i|)}) \\
&\quad + \sum_{i:h_i\le-\sigma(h)}(-h_i-\frac{h_i^2}{2(1-\sigma(h))}) \\
&\ge -h^te-\frac{\|h\|^2}{2(1-\sigma(h))},
\end{aligned}$$

and this completes the proof. □

Now we prove Lemma 4.1.

PROOF OF LEMMA 4.1. We introduce a new variable $\gamma \triangleq \tilde{\beta} - e/r$. Then we have

$$(4.9) \qquad 1-\tilde{\nu}\tilde{\beta}_i = 1-\frac{\tilde{\nu}}{r}-\tilde{\nu}\gamma_i, \quad 1-\tilde{\nu}\|\tilde{\beta}\|^2 = 1-\frac{\tilde{\nu}}{r}-\tilde{\nu}\|\gamma\|^2.$$

Putting

$$(4.10) \qquad \theta \triangleq \frac{r\tilde{\nu}}{r-\tilde{\nu}}$$

and taking note that $0<\tilde{\nu}\le 2r/3$ in Ω_0 (this follows from $\tilde{\beta}^te=1$), we obtain

$$(4.11) \qquad H_r(\tilde{\beta},\tilde{\nu}) = r\log(1-\theta\|\gamma\|^2) - \sum_{i=1}^{r}\log(1-\theta\gamma_i).$$

It is easy to see that

$$(4.12) \qquad \sigma(\gamma) = \sigma(\tilde{\beta}) - \frac{e}{r}$$

and

$$(4.13) \qquad \gamma^te = 0.$$

The following inequality is also well-known:

$$(4.14) \qquad \log(1-\delta) \le -\delta \qquad \text{for any } \delta < 1.$$

Now we use Proposition 4.2 with the above relations to produce

$$
\begin{aligned}
H_r(\tilde\beta,\tilde\nu) &= r\log(1-\theta\|\gamma\|^2) - \sum_{i=1}^{r}\log(1-\theta\gamma_i) \\
&\le -r\theta\|\gamma\|^2 + \frac{\theta^2\|\gamma\|^2}{2(1-\theta\sigma(\gamma))} \\
&= \theta\|\gamma\|^2(-r+\frac{\theta}{2(1-\theta\sigma(\gamma))})
\end{aligned}
\tag{4.15}
$$

provided that $\theta\sigma(\gamma)<1$ and $\theta\|\gamma\|^2<1$. These conditions are always satisfied on Ω_0 since

$$
\begin{aligned}
\theta\sigma(\gamma)<1 &\Leftrightarrow \frac{r\tilde\nu}{r-\tilde\nu}\left(\sigma(\tilde\beta)-\frac{1}{r}\right)<1 \\
&\Leftrightarrow r\sigma(\tilde\beta)-1<\frac{r-\tilde\nu}{\tilde\nu} \\
&\Leftrightarrow r\sigma(\tilde\beta)<r/\tilde\nu \\
&\Leftrightarrow 1-\tilde\nu\sigma(\tilde\beta)>0
\end{aligned}
\tag{4.16}
$$

and

$$
\begin{aligned}
\theta\|\gamma\|^2<1 &\Leftrightarrow \frac{r\tilde\nu}{r-\tilde\nu}\left(\|\tilde\beta\|^2-\frac{1}{r}\right)<1 \\
&\Leftrightarrow r\|\tilde\beta\|^2-1<\frac{r-\tilde\nu}{\tilde\nu} \\
&\Leftrightarrow \|\tilde\beta\|^2<1/\tilde\nu
\end{aligned}
\tag{4.17}
$$

$$
\Leftrightarrow 1-\tilde\nu\|\tilde\beta\|^2>0. \tag{4.18}
$$

Substituting the definition of γ and θ into the rightmost hand side of (4.15) we obtain

$$
H_r(\tilde\beta,\tilde\nu)\le\frac{r\tilde\nu}{r-\tilde\nu}\|\tilde\beta-\frac{e}{r}\|^2(-r+\frac{\tilde\nu}{2(1-\tilde\nu\sigma(\tilde\beta))})=T_r(\tilde\beta,\tilde\nu). \tag{4.19}
$$

Note that $\sigma(\tilde\beta)\ge 1/r$ follows from $e^t\tilde\beta=1$. Therefore, we have

$$
-r+\frac{\tilde\nu}{2(1-\tilde\nu\sigma(\tilde\beta))}\le -r+\frac{2/(3\sigma(\tilde\beta))}{2(1-2/3)}\le -r+\frac{1}{\sigma(\tilde\beta)}\le 0, \tag{4.20}
$$

which, combined with (4.19), proves (4.6). Since $\sigma(\tilde\beta)=1/r$ occurs if and only if $\tilde\beta=e/r$, the last statement is also obvious. □

5. Global Convergence of the Primal Sequence

Now we return to the analysis of the sequence. Recall that $1-\lambda\|\hat\beta_E^k\|^2/\sigma(\hat\beta_E^k)>0$ for $k\ge\tilde K$. If we set $\hat\nu^k\triangleq\lambda^k/\sigma(\hat\beta_E^k)$, then we have for $k\ge\tilde K$

$$
\hat\Delta^k=|E|\log(1-\hat\nu^k\|\hat\beta_E^k\|^2)-\sum_{j\in E}\log(1-\hat\nu^k\hat\beta_j^k)=H_{|E|}(\hat\beta_E^k,\hat\nu^k) \tag{5.1}
$$

and

$$e^t \hat{\beta}_E^k = 1 \tag{5.2}$$

$$\hat{\nu}^k \|\hat{\beta}_E^k\|^2 < 1 \tag{5.3}$$

$$\hat{\nu}^k \sigma(\hat{\beta}_E^k) = \lambda^k. \tag{5.4}$$

Hence, if λ^k satisfies the step-size choice (1.3), then $(\hat{\beta}_E^k, \hat{\nu}^k) \in \Omega_0$, and if (1.2), then $(\hat{\beta}_E^k, \hat{\nu}^k) \in \Omega_1$. We deal with the case of step-size (1.2) from now on. We apply Lemma 4.1 to get

$$\hat{\Delta}^k \leq T_{|E|}(\hat{\beta}_E^k, \hat{\nu}^k) \leq 0. \tag{5.5}$$

Taking limit of (3.33) in consideration of (3.15) and (5.5), we see that $f_E(x_E^k)$ is bounded above. Furthermore, we have the following lemma;

LEMMA 5.1. *$f_E(x_E^k)$ is bounded below throughout the iteration.*

PROOF. By using the well-known arithmetic and geometric mean, we have

$$\begin{aligned} \exp f_E(x_E^k) &= \frac{(c^t x^k - c^t x^\infty)^{|E|}}{\prod_{j \in E} x_j^k} \geq \left(\frac{\sqrt{|E|}(c^t x^k - c^t x^\infty)}{\|x_E^k\|} \right)^{|E|} \\ &\geq (\sqrt{|E|}\delta)^{|E|} > 0 \qquad \text{(Use Lemma 1.2)}, \end{aligned} \tag{5.6}$$

hence, $f_E(x_E^k)$ is bounded below. □

Now we are ready to prove the following theorem.

THEOREM 5.2. *The primal sequence converges to a relative interior point of the optimal face of the primal problem.*

PROOF. In view of (3.33), (3.14), (5.5) and the above lemma, $\hat{\Delta}^k$ must converge to 0. Hence,

$$\hat{\beta}_E^k \to \frac{e}{|E|} \tag{5.7}$$

follows from Lemma 4.1. This implies that there exists a number K^* such that $\hat{\alpha}_E^k > 0$ holds for $k \geq K^*$.

Choose $k \geq K^*$ and let $\tilde{s}_E^k \triangleq (X_E^k)^{-1}\hat{\alpha}_E^k > 0$. Recalling the remark after Lemma 1.4.5, we see that

$$X_E^k(\tilde{s}_E^k - \hat{s}_E) \in \mathcal{L}_E^\perp = \left\{ \alpha_E \in R^{|E|} \mid \alpha_E = X_E^k A_E^t y, \quad A_{E^c}^t y = 0 \right\}. \tag{5.8}$$

Hence, there exists an appropriate $\tilde{y}^k$ such that

$$\tilde{s}_E^k - \hat{s}_E = A_E^t \tilde{y}^k, \quad A_{E^c}^t \tilde{y}^k = 0, \tag{5.9}$$

and we have

$$\begin{pmatrix} \tilde{s}_E^k \\ 0 \end{pmatrix} = \begin{pmatrix} \hat{s}_E + A_E^t \tilde{y}^k \\ 0 \end{pmatrix} = \begin{pmatrix} c_E - A_E^t(\hat{y} - \tilde{y}^k) \\ 0 \end{pmatrix} = c - A^t(\hat{y} - \tilde{y}^k), \tag{5.10}$$

which means that $(\tilde{s}_E^k, 0)$ satisfies the equality condition of $\langle DP \rangle$. Therefore, $(\tilde{s}_E^k, 0)$ and $x^\infty = (0, x_{E^c}^\infty)$ satisfy the strict complementarity condition, and Face(E) is the optimal face of $\langle P \rangle$. □

This proves the global convergence of the primal sequence. We note that we have got the strict complementarity pair x^∞ and $(\tilde{s}_E^k, 0)$ in the above. Hence, the duality theorem (Theorem 1.1.2) and the strict complementarity theorem (Theorem 1.1.3) are now proved if the primal problem has an interior feasible solution. It is also easy to prove them under the case where no interior feasible solution exists, if one consider a Phase-I problem to find an feasible solution.

The following corollary may be interesting from practical viewpoint.

COROLLARY 5.3. *Under the same assumption of Theorem 1.1, the asymptotic reduction rate of* $c^t x^k - c^t x^\infty$ *is* $1 - \lambda^k$.

PROOF. From (3.4), (3.5) and (5.7), we have

$$\frac{\|\beta^k\|^2}{\sigma(\beta^k)} = \frac{\|\hat{\beta}_E^k\|^2 + \mathcal{O}(\|X_E^k\|^2)}{\sigma(\hat{\beta}_E^k) + \mathcal{O}(\|X_E^k\|^2)} \to 1. \tag{5.11}$$

Therefore, from Lemma 3.7,

$$\frac{c^t(x^{k+1} - x^\infty)}{c^t(x^k - x^\infty)} = 1 - \lambda^k \frac{\|\beta^k\|^2}{\sigma(\beta^k)} \to 1 - \lambda^k \tag{5.12}$$

follows. □

Furthermore, we see that

$$\frac{\|\alpha^k\|}{\sigma(\alpha^k)} = \frac{\|\beta^k\|}{\sigma(\beta^k)} \to \frac{\|e/|E|\|}{\sigma(e/|E|)} = \sqrt{|E|}, \tag{5.13}$$

which implies that the step length of the short step affine scaling method is about $1/\sqrt{|E|}$ of that of the long-step method asymptotically.

6. Global Convergence of the Dual Sequence

Next we prove the global convergence of the dual sequence s^k.

First, we characterize the analytic center of the optimal face of $\langle DP \rangle$.

LEMMA 6.1. *The analytic center of the optimal face of* $\langle DP \rangle$ *is the unique solution of the following system;*

$$A_E S_E^{-1} e \in \mathrm{Im}(A_{E^c}), \tag{6.1}$$

$$s_E + A_E^t y = c_E, \tag{6.2}$$

$$A_{E^c}^t y = c_{E^c}, \quad s_{E^c} = 0 \tag{6.3}$$

where $S_E \triangleq \mathrm{diag}(s_E)$.

PROOF. Recall that the analytic center is the optimal solution of

$$(6.4) \qquad \begin{cases} \text{minimize} & -\sum_{j\in E}\log s_j \\ \text{subject to} & s = c - A^t y, \quad s_E > 0, \quad s_{E^c} = 0. \end{cases}$$

The Karush-Kuhn-Tucker conditions for the above optimization problem are (6.1), (6.2) and (6.3). The existence and uniqueness of the solution follow from Asssumption I-1 and Lemma 1.1.4. □

Now we prove the latter part of the main theorem Theorem 1.1 which is described below.

THEOREM 6.2. *The dual sequence converges to the analytic center of the optimal face of* $\langle DP\rangle$.

PROOF. From (2.10),

$$(6.5) \qquad \|s^k_{E^c}\| = \|(X^k_{E^c})^{-1}\alpha^k_{E^c}\| \le M\|X^k_E\|^2 \to 0.$$

Hence, (6.3) is obvious. Now we analyze s^k_E.

We can choose $\hat{s}_E > 0$ since we have already got strict complementarity solutions in Section 5. We put for $j \in E$,

$$(6.6) \qquad p^k_j \triangleq \frac{x^k_j}{\hat{s}^t_E x^k_E}.$$

Then we have

$$(6.7) \qquad \prod_{j\in E} p^k_j = \exp(-f_E(x^k_E))$$

which is bounded above and below by positive constants. Since p^k_j itself is bounded above by a positive constant $1/(\min_{j\in E}\hat{s}_j)$, p^k_j is also bounded below. We can easily see that $s^k_j = (p^k_j)^{-1}\beta^k_j$ for $j \in E$, hence s^k_E is also bounded and has a strictly positive accumulation point s^*_E. Let k_t be the subsequence convergent to s^*_E. Then,

$$(6.8) \qquad (s^{k_t}_j)^{-1} = p^{k_t}_j(\beta^{k_t}_j)^{-1} \to p^*_j|E| = (s^*_j)^{-1}.$$

On the other hand, we have

$$(6.9) \qquad \begin{aligned} A_E p^k &= \frac{A_E x^k_E}{\hat{s}^t_E x^k_E} = \frac{b - A_{E^c}x^k_{E^c}}{\hat{s}^t_E x^k_E} \\ &= \frac{A_{E^c}(x^\infty_{E^c} - x^k_{E^c})}{\hat{s}^t_E x^k_E} \in \mathrm{Im}(A_{E^c}), \end{aligned}$$

hence,

$$(6.10) \qquad A_E p^* \in \mathrm{Im}(A_{E^c}),$$

which implies that every accumulation point s^*_E satisfies (6.1). Since it is obvious that s^*_E satisfies (6.2), $s = (s^*_E, 0)$ is a solution of (6.1), (6.2) and (6.3). As was shown in Lemma 6.1, the system (6.1), (6.2) and (6.3) has the unique solution hence, every accumulation point of s^k must be identical. Therefore, s^k converges to $(s^*_E, 0)$ which is the analytic center of the optimal face of $\langle DP\rangle$. □

This completes the proof of Theorem 1.1.

CHAPTER III

Convergence Analysis of a Projective Scaling Method

1. Introduction

In this chapter, we apply the results of Chapter II to obtain new convergence results on a long-step variant of the projective scaling method. Let us consider the linear programming problem:

$$(1.1) \qquad \langle A \rangle \begin{cases} \text{minimize} & c^t u \\ \text{subject to} & Au = 0,\ g^t u = 1,\ u \geq 0, \end{cases}$$

where $g, c, u \in R^n$ and $A \in R^{m \times n}$. We make the following assumptions.

ASSUMPTION 1. *An interior feasible solution u^0 is known.*

ASSUMPTION 2. $\text{rank}(A) = m$ *and* $n \geq 3$

ASSUMPTION 3. $g \geq 0$ *but* $g \neq 0$,

ASSUMPTION 4. *The problem $\langle A \rangle$ has an optimal solution whose optimal value is 0.*

It is well-known that $\langle A \rangle$ contains Karmarkar's canonical form and the standard form LP problems. $\langle A \rangle$ has a close relation to the homogeneous problem

$$(1.2) \qquad \langle H \rangle \begin{cases} \text{minimize} & c^t x \\ \text{subject to} & Ax = 0,\ x \geq 0, \end{cases}$$

obtained by removing the inhomogeneous equality constraint $g^t u = 1$. Recalling that Section I-3 , we define the algorithm which generates the iterates by a conical projection from the ones obtained by applying the long-step affine scaling method for $\langle H \rangle$ that moves with a fixed ratio λ upto 2/3 of the way towards the boundary.

Specifically we will show that this variant has an $\mathcal{O}(nL)$ and $\mathcal{O}(n^2L)$ iteration polynomial complexity according to $\lambda < 2/3$ and $\lambda = 2/3$, respectively, generating sequences of primal iterates and dual estimates which converge to a relative interior point of the optimal face of $\langle A \rangle$ and the analytic center of the dual optimal face, respectively. The asymptotic reduction rate of the objective function is exactly $1 - \lambda$. We will also point out that a standard assumption of the boundedness of the optimal set, which usually is assumed in most of the literatures dealing with the projective scaling algorithm and is regarded as inevitable due to its convergence proof, is not necessary to ensure the global convergence.

The proof of polynomiality in this chapter seems to contain some novelty in that it is based on the homogeneous affine scaling direction, which is different from the search direction of Karmarkar's original algorithm when viewed as a search direction for $\langle H \rangle$. Lemma 1.3.1 however, implies that these two directions coincide when projected back to the original problem $\langle A \rangle$. The step-size is parameterized in terms of the ratio towards the boundary of the feasible region of $\langle H \rangle$, and this makes a remarkable contrast with the existing polynomiality proofs where the unit displacement vector is taken on the surface of the unit ellipsoid in scaled space. The proof affords a guaranteed worst-case reduction -0.42 in the potential function when n is sufficiently large by taking the step-size $\lambda = 0.54$, which is 58% of the tight bound on the guaranteed worst-case reduction when we perform exact line search obtained by Anstreicher [5] and McDiarmid [34]. As will be shown in Section 5, this proof can be combined with Todd and Burrell's lower bound updating procedure [56] to relax Assumption 4 of the optimal value being zero without sacrificing complexity. Convergence of the dual estimates and asymptotic convergence rate of the objective function $1 - \lambda(\geq 1/3)$ seem also new to projective scaling methods.

2. Algorithm

We introduce the algorithm. Given an interior feasible solution x of $\langle H \rangle$, the affine scaling direction $d(x)$ is defined as

$$d(x) = X(I - P_{AX})Xc. \tag{2.1}$$

Assume that a feasible solution $u^0 > 0$ of $\langle A \rangle$ is given, and let $x^0 = u^0$. Obviously, x^0 is a feasible solution of $\langle H \rangle$. Let λ be a constant between 0 and 1. With this initialization, we apply the long-step affine scaling method for $\langle H \rangle$ to generate a sequence $\{x^k\}$ of interior feasible solutions, and obtain a sequence $\{u^k\}$ of interior feasible solutions of $\langle A \rangle$ from $\{x^k\}$ by conical projection:

$$\begin{cases} x^{k+1} = x^k - \lambda \dfrac{d(x^k)}{\sigma((X^k)^{-1}d(x^k))} \\ u^{k+1} = \dfrac{x^{k+1}}{g^t x^{k+1}}. \end{cases} \tag{2.2}$$

Taking step-size λ means that we move a fixed fraction λ of the way toward the boundary of the feasible region of $\langle H \rangle$; thus the next iterates x^{k+1}, u^{k+1} are interior points of $\langle H \rangle$ and $\langle A \rangle$ respectively, if $0 < \lambda < 1$.

In the remaining part of this chapter, we use $\mathcal{S}_A$ and $\mathcal{S}_H$ to denote the optimal faces of $\langle A \rangle$ and $\langle H \rangle$, respectively. Further, we denote by N the index set for the always-active constraints on $\mathcal{S}_H$, and by B its complement, respectively. Let N' and B' be the index sets for the always-active constraints on $\mathcal{S}_A$ and its complement, respectively. We have the following propositions.

PROPOSITION 2.1. *If x is a relative interior point of $\mathcal{S}_H$, then $g^t x > 0$.*

PROOF. By contradiction, assume that $g^t x = 0$. Since $x_B > 0$ and $x_N = 0$, we have

$$g^t x = g_B^t x_B = 0, \tag{2.3}$$

which implies

$$g_B = 0. \tag{2.4}$$

Hence, we have, for any optimal point x' of $\langle H \rangle$,

$$g^t x' = g^t_B x'_B = 0. \tag{2.5}$$

This means that $\langle A \rangle$ has no optimal solution, contradicting the Assumption 4. □

PROPOSITION 2.2. *We have $N = N'$ and $B = B'$.*

PROOF. Since an optimal solution of $\langle A \rangle$ is also that of $\langle H \rangle$, $N' \supset N$. To prove $N \supset N'$, assume by contradiction that there exists some index $j \in N'$ which is not contained in N.

Let x be a relative interior point of S_H. Due to Proposition 2.1, $u \triangleq x/g^t x$ is a solution of $\langle A \rangle$ and $u_j > 0$ which contradicts the assumption that $j \in N'$. □

PROPOSITION 2.3. *If $\{x^k\}$ converges to a relative interior point of the optimal face of $\langle H \rangle$, then $\{u^k\}$ converges to a relative interior point of the optimal face of $\langle A \rangle$.*

PROOF. From Proposition 2.1, it is easy to see that u^k converges to $u^* \triangleq x^*/g^t x^*$. In consideration of $u^*_B > 0$ and $u^*_N = 0$, the proposition readily follows from Proposition 2.2. □

From this proposition and Theorem 2.1.1 we have the following result.

THEOREM 2.4. *If λ is taken to be $0 < \lambda \leq 2/3$ in the iteration (2.2), then u^k converges to u^*, a relative interior point of the optimal face of $\langle A \rangle$.*

An interesting point of this theorem is that it does not require boundedness of the optimal face, unlike any other convergence results on the projective scaling method.

3. Evaluation of Potential Reduction

To analyze the polynomialityof the algorithm, we define the Karmarkar potential function :

$$f(x) \triangleq n \log c^t x - \sum_{j=1}^{n} \log x_j. \tag{3.1}$$

THEOREM 3.1. *If λ is taken to be $0 < \lambda < 2/3$ in the iteration (2.2), we have, for all k,*

$$f(x^{k+1}) - f(x^k) < -\frac{\lambda}{3\sqrt{2}}\left(\frac{2}{3} - \lambda\right). \tag{3.2}$$

If $\lambda = 2/3$, we have, for all k,

$$f(x^{k+1}) - f(x^k) < -\frac{\sqrt{2}}{3n}. \tag{3.3}$$

PROOF. The reduction of the potential function is written as follows.

(3.4)

$$\Delta_n(x^k) \triangleq f(x^{k+1}) - f(x^k)$$
$$= n \log \left\{ 1 - \frac{\lambda c^t d^k}{\sigma((X^k)^{-1} d^k) c^t x^k} \right\} - \sum_{j=1}^{n} \log \left\{ 1 - \lambda \frac{(x_j^k)^{-1} d_j^k}{\sigma((X^k)^{-1} d^k)} \right\},$$

where $d^k \triangleq d(x^k)$. Let

(3.5)
$$\beta^k \triangleq \frac{(X^k)^{-1} d^k}{c^t x^k}, \quad \nu^k \triangleq \lambda \frac{c^t x^k}{\sigma((X^k)^{-1} d^k)} = \frac{\lambda}{\sigma(\beta^k)}.$$

Noting that

(3.6)
$$\|\beta^k\|^2 = \frac{c^t d^k}{(c^t x^k)^2},$$

we have

(3.7)
$$\Delta_n(x^k) = n \log \left\{ 1 - \nu^k \|\beta^k\|^2 \right\} - \sum_{j=1}^{n} \log \left\{ 1 - \nu^k \beta_j^k \right\}.$$

Since the objective function is nonnegative even we hit the boundary by taking $\lambda = 1$, we have

(3.8)
$$1 - \sigma(\beta^k) \geq 1 - \frac{\|\beta^k\|^2}{\sigma(\beta^k)} = 1 - c^t \frac{d^k}{c^t x^k \sigma((X^k)^{-1} d^k)} \geq 0.$$

It is easy to verify

(3.9)
$$\nu^k \sigma(\beta^k) = \lambda \leq 2/3,$$

and from (3.8),

(3.10)
$$\nu^k \|\beta^k\|^2 \leq 2/3 < 1.$$

Observe also

(3.11)
$$e^t \beta^k = \frac{e^t (I - P_{AX^k}) X^k c}{c^t x^k} = 1.$$

With (3.9), (3.10) and (3.11), we can apply Lemma 2.4.1 to (3.7) to obtain

(3.12)
$$\Delta_n(x^k) \leq T_n(\beta^k, \nu^k) \leq 0.$$

Let

(3.13)
$$T_n(\beta, \nu) = \phi(\beta) \psi(\beta),$$

(3.14)
$$\phi(\beta) \triangleq \frac{n\nu}{n - \nu} \|\beta - \frac{e}{n}\|^2,$$

(3.15)
$$\psi(\beta) \triangleq -n + \frac{\nu}{2(1 - \nu\sigma(\beta))}.$$

We observe that β^k is not strictly positive for all k. By contradiction, assume that $\beta^k > 0$ for some k. This means that $(X^k)^{-2} d^k$ is a strictly positive feasible solution of the dual problem of $\langle H\rangle$. Hence, Lemma 1.1.4 implies that the optimal face of $\langle H\rangle$ is bounded; thus $x = 0$ is the unique optimal solution of $\langle H\rangle$. This contradicts Assumption 3. Now we have

$$\beta^k \not> 0, \tag{3.16}$$

and hence

$$\sigma(\beta^k) \geq \frac{1}{n-1}. \tag{3.17}$$

From (3.9), (3.10) and (3.11), we see that $\phi(\beta^k) \geq 0$ and $\psi(\beta^k) \leq 0$. Now we concentrate our efforts on finding a lower bound and an upper bound for ϕ and ψ, respectively.

First we deal with ϕ. Due to (3.11), we have

$$\phi(\beta^k) = \frac{n\nu}{n-\nu}\|\beta^k - \frac{e}{n}\|^2 = \frac{\|\beta^k\|^2 - 1/n}{1/\nu - 1/n}. \tag{3.18}$$

We minimize $\|\beta\|^2$ assuming that $\sigma^k \triangleq \sigma(\beta^k)$ is given. By taking account that β satisfies the conditions $\sigma(\beta) = \sigma^k$ and (3.16), we are lead to the following simple quadratic programming problem $\langle G\rangle$.

$$\langle G\rangle \begin{cases} \text{minimize} & \|\tilde\beta\|^2 + (\sigma^k)^2 \\ \text{subject to} & \tilde\beta \in R^{n-1} \\ & e^t\tilde\beta = 1 - \sigma^k \\ & \tilde\beta_1 \leq 0 \\ & \tilde\beta_j \leq \sigma^k \quad (j = 2, \dots, n-1) \end{cases} \tag{3.19}$$

If we denote the optimal solution of $\langle G\rangle$ by $\tilde\beta^*$, then

$$\phi(\beta^k) \geq \frac{(\sigma^k)^2 + \|\tilde\beta^*\|^2 - 1/n}{1/\nu - 1/n}. \tag{3.20}$$

The Karush-Kuhn-Tucker condition for $\langle G\rangle$ is:

$$2\tilde\beta^* + \mu_0 e + \mu = 0, \quad \mu = (\mu_1, \dots, \mu_{n-1})^t \in R^{n-1} \tag{3.21}$$

$$\mu_1\tilde\beta_1^* = 0, \ \mu_1 \geq 0 \tag{3.22}$$

$$\mu_j(\tilde\beta_j^* - \sigma^k) = 0, \ \mu_j \geq 0 \quad (j = 2, \dots, n-1) \tag{3.23}$$

$$\tilde\beta_1^* \leq 0 \tag{3.24}$$

$$\tilde\beta_j^* \leq \sigma^k \quad (j = 2, \dots, n-1) \tag{3.25}$$

$$e^t\tilde\beta^* = 1 - \sigma^k. \tag{3.26}$$

From (3.21), we have

$$\tilde\beta_j^* = -(\mu_0 + \mu_j)/2 \quad (j = 1, \dots, n-1). \tag{3.27}$$

We first consider the case $\tilde\beta_1^* < 0$. Then, from (3.22), we have $\mu_1 = 0$ and from (3.27), $\mu_0 > 0$. This implies that all $\tilde\beta_j^*$ are negative, which contradicts (3.26) recalling that $\sigma^k \leq 1$ due to (3.8).

Now we get $\tilde{\beta}_1^* = 0$ and $\mu_0 \leq 0$. Then it is easy to see that each $\tilde{\beta}_j^*, (j = 2, \ldots, n-1)$ has the same value at the minimum. This and (3.26) produce $\tilde{\beta}_j^* = (1-\sigma^k)/(n-2)$ and the optimal value $(1-\sigma^k)^2/(n-2) + (\sigma^k)^2$.

With this result, now we are ready to come up with the following bound for $\phi(\beta^k)$.

$$\phi(\beta^k) \geq \frac{(\sigma^k)^2 + (1-\sigma^k)^2/(n-2) - 1/n}{\sigma^k/\lambda - 1/n} > \frac{\lambda}{\sqrt{2}n} \tag{3.28}$$

for $n \geq 3$. To simplify the notations, we let $\sigma \triangleq \sigma^k$. We have

$$\begin{aligned} &\frac{\sigma^2 + (1-\sigma)^2/(n-2) - 1/n}{\sigma/\lambda - 1/n} \\ &= \lambda \frac{\sigma^2 + (1-\sigma)^2/(n-2) - 1/n}{\sigma - \lambda/n} \\ &= \left(\frac{\lambda}{n-2}\right) \frac{(n-1)(\sigma - 1/(n-1))^2 - 1/(n-1) + 2/n}{\sigma - \lambda/n} \\ &= \left(\frac{\lambda}{n-2}\right) \frac{(n-1)\zeta^2 + w_1}{\zeta + w_2} \\ &= \left(\frac{\lambda}{n-2}\right) \rho(\zeta) \end{aligned} \tag{3.29}$$

where

$$\zeta \triangleq \sigma - 1/(n-1) \tag{3.30}$$

$$w_1 \triangleq 2/n - 1/(n-1) > 0 \tag{3.31}$$

$$w_2 \triangleq 1/(n-1) - \lambda/n > 0 \quad \text{and} \tag{3.32}$$

$$\rho(\zeta) \triangleq \frac{(n-1)\zeta^2 + w_1}{\zeta + w_2}. \tag{3.33}$$

Now we evaluate $\rho(\zeta)$ for $0 \leq \zeta \leq (n-2)/(n-1)$. Differentiating ρ with respect to ζ, we can easily see that $\rho(\zeta)$ has its minimum at

$$\zeta^* \triangleq -w_2 + \sqrt{w_2^2 + w_1/(n-1)}. \tag{3.34}$$

Hence

$$\rho(\zeta) \geq \frac{(n-1)\left\{-w_2+\sqrt{w_2^2+w_1/(n-1)}\right\}^2+w_1}{\sqrt{w_2^2+w_1/(n-1)}} \tag{3.35}$$

$$> \frac{w_1}{\sqrt{w_2^2+w_1/(n-1)}}$$

$$> \frac{\dfrac{n-2}{n(n-1)}}{\sqrt{\dfrac{1}{(n-1)^2}+\dfrac{n-2}{n(n-1)^2}}}$$

$$= \frac{n-2}{n\sqrt{1+(n-2)/n}}$$

$$> \frac{n-2}{\sqrt{2}n}.$$

Here we used the relation $w_2^2 < 1/(n-1)^2$ which follows from (3.32). From this relation and (3.29), we have

$$\phi(\beta^k) > \frac{\lambda}{\sqrt{2}n}. \tag{3.36}$$

Thus (3.28) was shown.

Next we evaluate ψ. From (3.17), we have $\|\beta^k\|^2 \geq 1/(n-1)$. By using (3.9) and (3.10), ψ can be bounded from above as follows:

$$\psi(\beta^k) = -n + \frac{\nu^k}{2(1-\nu^k\sigma(\beta^k))} \tag{3.37}$$

$$\leq -n+(n-1)-\frac{(1-3\lambda/2)(n-1)}{1-\lambda}$$

$$< -1-(\frac{2}{3}-\lambda)\frac{3}{2}(n-1).$$

Combining the results on $\psi(\beta)$ and $\phi(\beta)$ above, we see, if $\lambda < 2/3$, then

$$T_n(\beta^k,\nu^k) < \left\{-1-\frac{2}{3}(\frac{2}{3}-\lambda)(n-1)\right\}\frac{\lambda}{\sqrt{2}n} \tag{3.38}$$

$$< -\frac{\lambda}{3\sqrt{2}}(\frac{2}{3}-\lambda)$$

and if $\lambda = 2/3$,

$$T_n(\beta^k,\nu^k) < -\frac{\lambda}{\sqrt{2}n} = -\frac{\sqrt{2}}{3n} \tag{3.39}$$

This completes the proof of Theorem 3.1. □

In order to obtain the polynomial complexity results, we stop the iteration when $c^t u^k \leq 2^{-2L}$. Since $f(u^k) = f(x^k)$ for all k, we immediately obtain the following corollary by a standard argument.

COROLLARY 3.2. *Assume that the optimal solution set of* $\langle A \rangle$ *is bounded. Given an initial point such that* $c^t u^0 = \mathcal{O}(2^L)$, *the algorithm terminates in* $\mathcal{O}(nL)$ *iterations if* $\lambda < 2/3$ *in the iteration* (2.2), *and in* $\mathcal{O}(n^2 L)$ *iterations if* $\lambda = 2/3$.

Now we focus our attention on estimating the guaranteed worst-case reduction of the potential function. McDiarmid [34] and Anstreicher [5] proved independently that the potential can be guaranteed to be decreased at least as much as -0.72 per iteration by performing exact linesearch, and that this bound is tight. McDiarmid [34] proved the potential function can be reduced at least about -0.69 where the step-size is taken to be nearly equal to 1 in the meaning of Karmarkar's paper. Here we will show that our proof gives an estimate of the guaranteed worst-case reduction -0.42 with $\lambda = 0.54$, which is 58% of Anstreicher and McDiarmid's bound.

We prove the following theorem.

THEOREM 3.3. *If* n *is sufficiently large, we have*

$$f(x^{k+1}) - f(x^k) \leq -0.42 \tag{3.40}$$

for all k *by taking* $\lambda = 0.54$.

PROOF. Recall that $\Delta_n(x^k)$ denotes reduction of the potential function when we take the step-size λ. Our problem is to find an upper bound for

$$\sup_{x^k} \Delta_n(x^k). \tag{3.41}$$

By (3.12), we have

$$\Delta_n(x^k) \leq T_n(\beta(x^k), \frac{\lambda}{\sigma(\beta^k)}). \tag{3.42}$$

In view of (3.13), (3.14), (3.15) and (3.28), letting

$$f_n(\sigma, \lambda) \triangleq \lambda \frac{n(n-1)\sigma^2 - 2n\sigma + 2}{(n-2)(\sigma - \lambda/n)}, \tag{3.43}$$

$$g_n(\sigma, \lambda) \triangleq -1 + \frac{\lambda}{2(1-\lambda)n\sigma}, \tag{3.44}$$

$$h_n(\sigma, \lambda) \triangleq f_n(\sigma, \lambda) g_n(\sigma, \lambda), \tag{3.45}$$

we have

$$\Delta_n(x^k) \leq T_n(\beta(x^k), \frac{\lambda}{\sigma(\beta^k)}) \leq \max_{1/(n-1) \leq \sigma \leq 1} h_n(\sigma, \lambda). \tag{3.46}$$

From (3.34), $f_n(\sigma, \lambda)$ has its maximum with respect to σ at

$$\begin{aligned}(3.47)\qquad \sigma^* &\triangleq \frac{\lambda}{n} + \sqrt{\left(\frac{1}{n-1} - \frac{\lambda}{n}\right)^2 + \frac{n-2}{n(n-1)^2}} \\ &= \frac{\lambda}{n} + \frac{1}{n-1}\sqrt{\left(1 - \frac{n-1}{n}\lambda\right)^2 + \frac{n-2}{n}}.\end{aligned}$$

Let $\bar{\sigma}_n(\lambda)$ be the σ where h_n takes its minimum. Since $g_n(\sigma, \lambda)$ is monotone decreasing with respect to σ, we see that

$$(3.48)\qquad \frac{1}{n-1} \le \bar{\sigma}_n(\lambda) \le \sigma^* \le \frac{3}{n-1}.$$

Hence putting $\alpha \triangleq (n-1)\sigma$, we have

$$(3.49)\qquad \Delta_n(x^k) \le \max_{1/(n-1)\le\sigma\le 1} h_n(\sigma, \lambda) = \max_{1\le\alpha\le 3} h_n(\alpha/(n-1), \lambda).$$

By evaluating the optimal solution appearing on the right-hand side, we obtain an estimate of the worst-case reduction when the step-size λ is taken. Further, an optimal step-size λ_n^* is obtained as a solution for the following optimization problem:

$$(3.50)\qquad \min_{\lambda} \max_{1\le\alpha\le 3} h_n(\alpha/(n-1), \lambda).$$

We denote by α_n^* the optimal α for this problem.

Now we analyze the behavior of λ_n^* and α_n^* when n is sufficiently large. The domains of α and λ we are interested in are $[1, 3]$ and $[0, 2/3]$ respectively, and it is easy to check that $h_n(\alpha/(n-1), \lambda)$ is uniformly continuous on this domain. Hence $h_n(\alpha/(n-1), \lambda)$ converges uniformly to the function $\bar{h}(\alpha, \lambda)$ on the domain, which is defined as follows:

$$(3.51)\qquad \bar{h}(\alpha, \lambda) \triangleq \lim_{n\to\infty} h_n(\alpha/(n-1), \lambda) = \bar{f}(\alpha, \lambda)\bar{g}(\alpha, \lambda),$$

$$(3.52)\qquad \bar{f}(\alpha, \lambda) \triangleq \lim_{n\to\infty} f_n(\alpha/(n-1), \lambda) = -\frac{\lambda\{(\alpha-1)^2 + 1\}}{\alpha - \lambda} \quad \text{and}$$

$$(3.53)\qquad \bar{g}(\alpha, \lambda) \triangleq \lim_{n\to\infty} g_n(\alpha/(n-1), \lambda) = 1 - \frac{\lambda}{2(1-\lambda)\alpha}.$$

Let (α^*, λ^*) be the optimal solution for the following optimization problem:

$$(3.54)\qquad \min_{\lambda} \max_{1\le\alpha\le 3} \bar{h}(\alpha, \lambda).$$

Due to the uniform convergence of h_n to $\bar{h}$, we have

$$(3.55)\quad \lim_{n\to\infty} \lambda_n^* = \lambda^*, \quad \lim_{n\to\infty} \alpha_n^* = \alpha^*, \quad \lim_{n\to\infty} \min_{\lambda} \max_{1\le\alpha\le 3} h_n(\alpha/(n-1), \lambda) = \bar{h}(\alpha^*, \lambda^*).$$

Unfortunately, it is difficult to obtain exact value of λ^* and α^*. We choose approximately optimal value $0.54 \sim \lambda^*$ by numerical computation, and solve the following optimization problem with respect to α:

$$(3.56)\qquad \max_{1\le\alpha\le 3} \bar{h}(\alpha, 0.54),$$

which is an upper-bound and a good approximation to

$$\bar{h}(\alpha^*, \lambda^*). \tag{3.57}$$

We solve this problem by numerical calculation to obtain an approximate minimum value -0.423 when $\alpha = 1.374$. We can also confirm that exact minimum value is smaller than -0.42 even if we take numerical error into account. Thus we have

$$\Delta_n(x^k) \leq -0.42 \tag{3.58}$$

when $\lambda = 0.54$ if n is sufficiently large. This completes the proof. □

4. Other Convergence Results

In this section, we analyze convergence rate of the objective function value and convergence of dual estimates.

THEOREM 4.1. *If λ is taken to be $0 < \lambda \leq 2/3$ in the iteration (2.2), the asymptotic reduction rate of the objective function value $\{c^t u^k\}$ converges to $1 - \lambda$.*

PROOF. From Corollary 2.5.3, we have

$$\frac{c^t x^{k+1}}{c^t x^k} \to 1 - \lambda. \tag{4.1}$$

Since

$$\lim_{k \to \infty} \frac{g^t x^{k+1}}{g^t x^k} = 1, \tag{4.2}$$

we have

$$\lim_{k \to \infty} \frac{\dfrac{c^t u^{k+1}}{c^t u^k}}{1 - \lambda} = \lim_{k \to \infty} \frac{\dfrac{c^t x^{k+1} g^t x^k}{c^t x^k g^t x^{k+1}}}{1 - \lambda} = 1. \tag{4.3}$$

□

Now we show convergence of the dual estimates. The dual problem of $\langle A \rangle$ is

$$\langle DA \rangle \begin{cases} \text{maximize} & z_{m+1} \\ \text{subject to} & \xi = c - A^t z - z_{m+1} g \geq 0, \ z \in R^m, \ z_{m+1} \in R. \end{cases} \tag{4.4}$$

From duality theorem, the optimal value is 0.

Given an interior feasible solution u^k of $\langle A \rangle$, the dual estimate $(z(u^k), z_{m+1}(u^k))$ for $\langle A \rangle$ is defined as the optimal solution of the following problem:

$$\begin{cases} \text{minimize} & \|U^k(c - A^t z - z_{m+1} g)\|^2 \\ \text{subject to} & (z, z_{m+1}) \in R^{m+1} \end{cases} \tag{4.5}$$

where $U^k \triangleq \operatorname{diag}(u^k)$ (See 13). The explicit formula of the dual estimate is given by

$$z(u^k) = (A(U^k)^2 A^t)^{-1} A(U^k)^2 (c - z_{m+1} g) \tag{4.6}$$

$$z_{m+1}(u^k) = \frac{g^t U^k (I - P_{AU^k}) U^k c}{g^t U^k (I - P_{AU^k}) U^k g}. \tag{4.7}$$

On the other hand, the dual problem of $\langle H\rangle$ is

$$(4.8) \qquad \langle DH\rangle \begin{cases} \text{find} & y \\ \text{subject to} & s = c - A^t y \geq 0, \ y \in R^m. \end{cases}$$

Given an interior feasible solution x^k of $\langle H\rangle$, the dual estimate $y(x^k)$ for $\langle DH\rangle$ is defined as the optimal solution of the following optimization problem:

$$(4.9) \qquad \begin{cases} \text{minimize} & \|X^k(c - A^t y)\|^2 \\ \text{subject to} & y \in R^m. \end{cases}$$

The dual estimate is written explicitly as

$$(4.10) \qquad y(x^k) \triangleq (A(X^k)^2 A^t)^{-1} A(X^k)^2 c.$$

The above definition seems different from (1.2.14). The full rankness of A however, implies that $y(x^k)$ is uniquely determined by $s(x^k)$, hence $y(x^k)$ and $s(x^k)$ have a one-to-one correspondence. We use the term "the dual estimate" for $(z(u^k), z_{m+1}(u^k))$, and call $y(x^k)$ "the homogeneous dual estimate" to avoid confusion.

Next we introduce the analytic center of the optimal faces of $\langle DA\rangle$ and $\langle DH\rangle$. Taking note that z_{m+1} is always 0 on the optimal solution of $\langle DA\rangle$ and that the always-active index set on the dual optimal face is given by B due to Proposition 2.2, we see that the analytic center of the optimal face of $\langle DA\rangle$ is the optimal solution of the following convex optimization problem.

$$(4.11) \qquad \begin{cases} \text{minimize} & -\sum_{j\in N} \log \xi_j \\ \text{subject to} & \xi_N = c_N - A_N^t z - z_{m+1} g_N > 0, \\ & A_B^t z + z_{m+1} g_B = c_B, \quad z_{m+1} = 0. \end{cases}$$

Similarly, it is easy to observe that the analytic center of the optimal face of $\langle DH\rangle$ is given as the optimal solution for the convex optimization problem:

$$(4.12) \qquad \begin{cases} \text{minimize} & -\sum_{j\in N} \log s_j \\ \text{subject to} & s_N = c_N - A_N^t y > 0, \ A_B^t y = c_B. \end{cases}$$

We denote the analytic center of the optimal face of $\langle DA\rangle$ and $\langle DH\rangle$ by (z^*, z_{m+1}^*) and y^*, respectively. Comparing both problems, we have $(z^*, z_{m+1}^*) = (y^*, 0)$.

By applying Theorem 2.1.1 to the algorithm, we immediately obtains the following lemma.

LEMMA 4.2. *If we choose $0 < \lambda \leq 2/3$ in the iteration (2.2), the homogeneous dual estimate $y(x^k)$ converges to y^*, the analytic center of the optimal face of $\langle DH\rangle$.*

Now we are ready to see the convergence of the dual estimates $(z(u^k), z_{m+1}(u^k))$ to the analytic center of the dual optimal face of $\langle A\rangle$.

THEOREM 4.3. *If λ is taken to be $0 < \lambda \leq 2/3$ in the iteration (2.2), the dual estimate $(z(u^k), z_{m+1}(u^k))$ converges to $(y^*, 0)$, the analytic center of the optimal face of $\langle DA\rangle$.*

PROOF. From (4.6),

$$(4.13) \qquad z(u^k) = y(u^k) - z_{m+1}(u^k)(A(U^k)^2 A^t)^{-1} A(U^k)^2 g.$$

Due to Lemma 1.4.1, $A^t(AU^2A^t)^{-1}AU^2g = U^{-1}P_{AU}Ug$ is bounded for any diagonal matrix U which has positive diagonal components. The full rankness of A also assures that $(AU^2A^t)^{-1}AU^2g$ is bounded. Hence if

$$z_{m+1}(u^k) \to 0, \tag{4.14}$$

then using the fact that $y(u^k) = y(x^k)$ and Lemma 4.2, we have

$$z(u^k) \to y^*, \tag{4.15}$$

which proves the theorem.

Below we show (4.14). We can write (4.7) as follows:

$$z_{m+1}(u^k) = \frac{g^t(U^k)^2(c - A^t y(u^k))}{\|(I - P_{AU^k})U^k g\|^2}. \tag{4.16}$$

Since $e^t U^k g = 1$ and $U^k g \geq 0$, $U^k g$ is bounded for all k. We also have

$$U^k(c - A^t y(u^k)) = \frac{X^k(c - A^t y(x^k))}{g^t x^k} \to 0 \tag{4.17}$$

from complementarity condition. These facts imply the numerator of (4.16) converges to zero. Hence, to show that $z_{m+1}(u^k) \to 0$, it is enough to prove the denominator of (4.16) is bounded below by a positive constant. Since

$$e^t(I - P_{AU^k})U^k g = e^t U^k g - e^t P_{AU^k} U^k g = 1, \tag{4.18}$$

we have

$$\|(I - P_{AU^k})U^k g\|^2 \geq 1/n, \tag{4.19}$$

and this completes the proof. □

5. Todd-Burrell Lower Bound Updating Procedure

In this section we replace Assumption 4 with the following assumption:

Assumption 4* *A lower bound for the optimal value of $\langle A \rangle$ is known.*

We will show that an analogue of Todd-Burrell lower bound updating procedure [56] is applied to the algorithm without sacrificing the complexity analysis.

We denote by z^* the optimal value of $\langle A \rangle$. For $z \leq z^*$, we consider the following problem modified from $\langle A \rangle$:

$$\langle A(z) \rangle \begin{cases} \text{minimize} & (c - zg)^t u \\ \text{subject to} & Au = 0, \quad g^t u = 1, \quad u \geq 0. \end{cases} \tag{5.1}$$

and the associated homogeneous problem $\langle H \rangle$:

$$\langle H(z) \rangle \begin{cases} \text{minimize} & (c - zg)^t x \\ \text{subject to} & Ax = 0, \quad x \geq 0. \end{cases} \tag{5.2}$$

Note that the optimal value of $\langle A(z^*) \rangle$ is 0 and that if we know z^*, we can apply the original algorithm to $\langle A(z^*) \rangle$.

Given a lower bound z for the optimal value and an interior feasible point x of $\langle H(z)\rangle$, we define the Karmarkar potential function as

$$\tilde{f}(x,z) \triangleq n\log((c-zg)^t x) - \sum_{j=1}^{n}\log x_j. \tag{5.3}$$

By using the standard argument, we can solve $\langle H\rangle$ by finding (x,z) such that

$$\tilde{f}(x,z) = -2L, \tag{5.4}$$

where L is the input size of the problem. Since we can easily derive an optimal solution of $\langle A\rangle$ from that of $\langle H\rangle$, we now intend to reduce $\tilde{f}$ until (5.4) is satisfied.

It is not difficult to see that

$$\tilde{f}(x,z_1) \le \tilde{f}(x,z_0) \tag{5.5}$$

if $z_0 \le z_1 \le z^*$.

Due to these observations, if we can generate a sequence $\{(x^k,z^k)\}$ of the pair of an interior feasible solution for $\langle H\rangle$ and a lower bound for the optimal value such that

$$\tilde{f}(x^{k+1},z^{k+1}) \le \tilde{f}(x^k,z^k) - \delta, \quad z^k \le z^{k+1} \tag{5.6}$$

for all k, we can solve $\langle H\rangle$ and $\langle A\rangle$ with $\mathcal{O}(nL)$ and $\mathcal{O}(n^2L)$ iteration complexity when $\delta = \mathcal{O}(1)$ and $\delta = \mathcal{O}(1/n)$, respectively.

Todd-Burrell update is a procedure to generate a sequence with this property. At the k-th each step it first computes

$$\tilde{d}(x^k,z^k) = X^k(I - P_{AX^k})X(c^k - z^k g) \tag{5.7}$$

as a candidate for the search direction.

If $\tilde{d}(x^k,z^k) \not> 0$, we can show that moving to the direction $\tilde{d}$ gives a sufficient decrease in the potential function $\tilde{f}(\cdot,z^k)$, hence set

$$z^{k+1} = z^k, \tag{5.8}$$

and adopt

$$\tilde{d}(x^k,z^k)(= \tilde{d}(x^k,z^{k+1})) \tag{5.9}$$

as the search direction, which yields a sufficient decrease in $\tilde{f}$.

On the other hand, if (5.7) is positive, the procedure generates z^{k+1} such that $z^k < z^{k+1} \le z^*$ with which the search direction $\tilde{d}(x^k,z^{k+1})$ can decrease the potential sufficiently. This is an outline of the algorithm. Below we explain it in more detail.

Suppose $\tilde{d}(x^k,z^k) \not> 0$. Then we put $z^{k+1} = z^k$. By using the direction $\tilde{d}(x^k,z^k)(= \tilde{d}(x^k,z^{k+1}))$, the difference of the modified potential can be written as

$$\begin{aligned} \tilde{\Delta}_k &\triangleq \tilde{f}(x^{k+1};z^{k+1}) - \tilde{f}(x^k;z^k) \\ &= n\log\left\{1 - \frac{\lambda}{\sigma((X^k)^{-1}\tilde{d}^k)}\frac{(c-z^k g)^t\tilde{d}^k}{(c^t - z^k g)^t x^k}\right\} \\ &\quad - \sum_{j=1}^{n}\log\left\{1 - \frac{\lambda(x_j^k)^{-1}\tilde{d}_j^k}{\sigma((X^k)^{-1}\tilde{d}^k)}\right\}. \end{aligned} \tag{5.10}$$

Putting

$$\tilde{\beta}^k \triangleq \frac{(X^k)^{-1}\tilde{d}^k}{(c-zg)^t x^k} \tag{5.11}$$

and

$$\tilde{\nu}^k \triangleq \frac{\lambda((c-zg)^t x^k)}{\sigma((X^k)^{-1}\tilde{d}^k)} = \frac{\lambda}{\sigma(\tilde{\beta}^k)}, \tag{5.12}$$

we have

$$\|\tilde{\beta}^k\|^2 = \frac{(c-z^k g)^t \tilde{d}^k}{((c-z^k g)^t x^k)^2} \tag{5.13}$$

and

$$e^t\tilde{\beta}^k = \frac{e^t(I-P_{AX^k})X^k(c-z^k g)}{(c-z^k g)x^k} = \frac{(c-z^k g)^t x^k}{(c-z^k g)x^k} = 1. \tag{5.14}$$

We rewrite (5.10) as

$$\tilde{\Delta}_k = n\log\left\{1-\tilde{\nu}^k\|\tilde{\beta}^k\|^2\right\} - \sum_{j=1}^{n}\log\left\{1-\tilde{\nu}^k\tilde{\beta}_j^k\right\}, \tag{5.15}$$

which has the same form as (3.7). It is easy to see that $(\tilde{\beta}^k, \tilde{\nu}^k)$ satisfies (3.9), (3.10) and (3.11). Since sign of $\tilde{\beta}^k$ is the same as $\tilde{d}^k$, $\tilde{d}^k \not> 0$ implies $\tilde{\beta}^k$ satisfies (3.16), and we can see that the sufficient reduction is obtained in the value of the potential function by using the same argument given in the proof of Theorem 2.

On the other hand, if $\tilde{d}(x^k; z^k) > 0$, we have

$$\tilde{\beta}^k > 0 \Leftrightarrow c - A^t(A(X^k)^2A^t)A(X^k)^2(c-z^k g) - z^k g > 0. \tag{5.16}$$

Letting

$$\tilde{y}^k \triangleq (A(X^k)^2A^t)A(X^k)^2(c-z^k g), \tag{5.17}$$

$\tilde{d}(x^k; z^k) > 0$ implies that $(\tilde{y}^k, z^k)$ is a feasible solution of $\langle DA \rangle$, the dual problem of $\langle A \rangle$. Further, if we put

$$z^{k+1} \triangleq \min_{j:g_j>0} \frac{c_j - a_j^t\tilde{y}^k}{g_j} \tag{5.18}$$

where a_j is the j-th column vector of A, then $c - A^t\tilde{y}^k - z^{k+1}g \geq 0$ and $(\tilde{y}^k, z^{k+1})$ is another feasible solution of $\langle DA \rangle$. Obviously, $z^{k+1} > z^k$ and we get a better lower bound for z^*. Now we use $\tilde{d}(x^k, z^{k+1})$ as the search direction. From its construction, it is easy to see that the $\tilde{d}(x^k, z^{k+1})$ is not positive, namely, (3.16) holds, and, again, in a similar manner as in the proof of Theorem 2, we see that the potential $\tilde{f}$ can be reduced sufficiently by moving in the direction of $\tilde{d}(x^k, z^{k+1})$.

Now we obtain the following algorithm and theorem on its complexity.

Modified Algorithm

Initialize $z^0 (< z^*), x^0 := u^0, k := 0$;

while $c^t u^k - z^k$ *is not sufficiently small* **do**

begin

$\tilde{d}^k := X^k(I - P_{AX^k})X^k(c - z^k g)$;

$\tilde{y}^k := (A(X^k)^2 A^t)^{-1} A(X^k)^2 (c - z^k g)$;

if $\tilde{d}^k > 0$ **then**

begin

$z^{k+1} := \min_{j:g_j>0} (c_j - a_j^t y^k)/g_j$;

$\tilde{d}^k := X^k(I - P_{AX^k})X^k(c - z^{k+1} g)$

end

else $z^{k+1} := z^k$;

endif

$x^{k+1} := x^k - \dfrac{\lambda \tilde{d}^k}{\sigma((X^k)^{-1}\tilde{d}^k)}$;

$u^{k+1} := \dfrac{x^{k+1}}{g^t x^{k+1}}$;

$k := k + 1$

end

return u^k;

Theorem 5.1. *Assume that Assumption 1, 2 and 3* hold, and that the optimal face of $\langle A \rangle$ is bounded. Given a lower bound for the optimal value and an initial interior point u^0 such that $c^t u^0 = \mathcal{O}(2^L)$, the modified algorithm terminates in $\mathcal{O}(nL)$ iterations if $\lambda < 2/3$ and in $\mathcal{O}(n^2 L)$ iterations if $\lambda = 2/3$.*

CHAPTER IV

An Affine Scaling Method with an Infeasible Starting Point

1. Introduction

As was already pointed out in the preface, the original affine scaling method as well as most of the early interior point (IP) methods has the problem of initialization. In this chapter, we challenge this problem and propose an infeasible-IP method which is an extension of the affine scaling method. In contrast with the method proposed here, the original affine scaling method which uses an interior feasible solution is sometimes referred to as *feasible* affine scaling method.

The search direction is composed of two directions, one of which aims for feasibility and the other for optimality. Each of the directions is an affine scaling direction of a certain LP problem. We define the search direction by combining these two directions so that it has scaling invariance property. The method can be viewed as an extension of the affine scaling method in this sense. In fact, if the starting point happens to be feasible, then the method is reduced to the feasible long-step affine scaling method whose global convergence has already been proved in Chapter II.

As for history, many infeasible-IP methods have been already proposed. Most of them are classified into two types: primal-dual infeasible-IP methods (e.g. Kojima, Megiddo and Mizuno [27]) and combined Phase I-Phase II methods (Anstreicher [6], Anstreicher [7], Freund [15],Freund [16],Freund [17], Lustig [31], Todd [54], Todd [55]). Our method works on the primal problem only, thus it is not a primal-dual infeasible-IP method and is similar to the combined methods. While they are based on projective scaling methods [6, 54], path-following methods [31], barrier-function method [15, 17], or affine potential reduction methods [7, 16, 55], our method is based on the original affine scaling method. Another similar algorithm was developed by Dikin and Zorkaltsev [14], in which they used the two same directions as ours. Recently, Andersensen [4] also proposed an infeasible-IP method which uses the two directions, and left some numerical experiments. But the combinations of them are different from ours, and no global convergence proof has been made with their algorithms. On the other hand, we succeed to invent such a nice combination that we can prove the global convergence of our method. We also point out that our method is related to feasibility-improving gradient-acute-projection (FIGAP) methods developed by Tanabe [53]. FIGAP methods are a certain family of algorithms for nonlinear programming, and our method can be regarded as a kind of FIGAP methods applied for linear programming.

We prove that if the original LP problem has an optimal solution, then the sequence converges to a relative interior point of the optimal face. The objective function value may not decrease monotonically since the initial objective function value may be less than the optimal value. Instead, we define a certain infeasibility criteria, which decreases monotonically. As a result, the method finds an interior feasible solution in a finite number of iterations if it exists. Then, as was already mentioned, the method is reduced to the feasible affine scaling method. If no optimal solution exists, then due to the duality theorem (Theorem 1.1.2), the original problem and/or its dual problem does not have a feasible solution, in which case the method detects at least one of the infeasibilities of the primal or the dual problem. Since we have two directions for feasibility and optimality, we can construct two types of dual estimates correspondingly. If the primal problem has an optimal solution, then dual estimates for optimality converge to the analytic center of the optimal face of the dual problem, while the dual estimates for feasibility is used to detect infeasibility of the primal problem.

We will prove the global convergence of the proposed method under very weak assumptions by means of local Karmarkar potential function. In fact, we do not assume any kind of nondegeneracy nor boundedness while assuming that the coefficient matrix is full rank for convenience.

This chapter consists of ten sections and one appendix. In Section 2, we introduce the problem, define the algorithm and state the main theorem of this chapter. In Section 3, we prove that if the objective function value is bounded below, the sequence is convergent. In Section 4, we prove that the limit point is feasible if the original LP problem has a feasible solution. We also deal with the case where no feasible region exists in this section. Section 5 to Section 9 are devoted to prove the optimality of the limit point. In Section Section 5, we prove that the limit point is in a dual degenerate face. In Section 6 and Section 7, we make preliminary analysis for the main theorems which are proved in Section 8 and Section 9. Section 10 is concluding remarks where the theorems proved in this chapter are summarized. Section 10 is an appendix where we prove a few technical lemmas used in the chapter.

2. The Algorithm and the Main Theorem

2.1. The Algorithm

We consider the standard form linear programming problem

$$\langle P\rangle \left\{ \begin{array}{ll} \text{minimize} & c^t x \\ \text{subject to} & Ax = b, \quad x \geq 0, \end{array} \right. \tag{2.1}$$

where $c, x \in R^n, b \in R^m$ and $A \in R^{m\times n}$. The dual problem of $\langle P\rangle$ is

$$\langle DP\rangle \left\{ \begin{array}{ll} \text{maximize} & b^t y \\ \text{subject to} & A^t y \leq c. \end{array} \right. \tag{2.2}$$

We denotes the set of feasible solution by Feas$\langle P\rangle$. We make the following assumption.

Assumption 1. rank$(A) = m$.

We do not make any other assumptions on nondegeneracy nor boundedness. The problem $\langle P\rangle$ may not have an optimal solution nor even a feasible solution. Our analysis, however, deals with all these irregular situations.

The algorithm starts from an arbitrary positive point x^0, and make a sequence of positive points. The whole algorithm can be roughly sketched as follows.

ALGORITHM 0
Initialize $x^0 > 0, k := 0$;
while x^k *does not satisfy the stopping criteria* **do**
  begin
    Compute Δx^k;
    $x^{k+1} := x^k - \mu^k \Delta x^k$;
    $k := k+1$
  end
return x^k;

Here, Δx^k is the search direction and μ^k is a step-size. To keep the iterates always positive,

$$\mu^k < \sigma((X^k)^{-1}\Delta x^k)^{-1} \tag{2.3}$$

must hold where $X^k \triangleq \mathrm{diag}(x^k)$.

The starting point x^0 may not satisfy the equality constraints of $\langle P\rangle$, hence the algorithm should improve not only optimality but also feasibility. In fact, the search direction is composed of two differently originated directions for optimality and feasibility. We define the optimality direction at x by

$$\Delta_a x \triangleq \frac{X(I - P_{AX})Xc}{\|(I - P_{AX})Xc\|}, \tag{2.4}$$

and feasibility direction by

$$\Delta_n x = X^2 A^t (AX^2A^t)^{-1}(Ax - b) \tag{2.5}$$

where $X \triangleq \mathrm{diag}(x)$. We will motivate these two directions and investigate their properties in the next subsection. The search direction is defined by:

$$\Delta x^k \triangleq \theta \Delta_a x^k + \Delta_n x^k, \tag{2.6}$$

where $0 < \theta < 1$. We set $\hat{\epsilon} \triangleq 1 - \theta$.

If we choose $\mu^k = 1$, then

$$Ax^{k+1} = A(x^k - \Delta x^k) = Ax^k - A\Delta_n x^k = Ax^k - (Ax^k - b) = b, \tag{2.7}$$

which means that the next iterate satisfies the equality constraints of $\langle P\rangle$. Hence in view of (2.3), if $\sigma((X^k)^{-1}\Delta x^k) < 1$, then by taking step-size $\mu^k = 1$, we obtain an interior feasible point. In consideration of the above observation, we adopt the following step-size choice.

Step Size Choice 1 *While x^k is infeasible, if $\sigma^k \triangleq \sigma((X^k)^{-1}\Delta x^k) < 1$ then choose $\mu^k = 1$ (we find an interior feasible solution.), else choose λ^k satisfying*

$$\lambda_{\min} \le \lambda^k \le \min\left(\frac{2}{3} - \hat{\delta}, \frac{\sigma^k}{1+\sigma^k}\right) \tag{2.8}$$

and let

$$\mu^k = \lambda^k/\sigma^k \tag{2.9}$$

where $\lambda_{\min}$ *and* $\hat{\delta}$ *are predefined positive constants. Once* x^k *becomes feasible, then we choose the step-size as* (2.9) *with*

$$\lambda_{\min} \le \lambda^k \le 2/3. \tag{2.10}$$

The step-size choice (2.9) means to move with the ratio λ^k in the direction to the boundary of the positive orthant. Note that since $\sigma^k \ge 1$ in (2.8),

$$\frac{\sigma^k}{1+\sigma^k} \ge \frac{1}{2} \tag{2.11}$$

holds. We also point out that if x^k is feasible, Algorithm 1 is nothing but a feasible long-step affine scaling method, which can be easily verified.

Now we describe the complete algorithm:

ALGORITHM 1
Initialize $x^0 > 0, k := 0$;
while x^k *does not satisfy the stopping criteria* **do**
  begin
    Compute Δx^k;
    if x^k *is not feasible*
      then if $\sigma((X^k)^{-1}\Delta x^k) < 1$
        then $\mu^k := 1$;
        else $\mu^k := \lambda^k/\sigma((X^k)^{-1}\Delta x^k)$; # λ^k *is chosen by* (2.8).
      endif
      else $\mu^k := \lambda^k/\sigma((X^k)^{-1}\Delta x^k)$; # $\lambda_{\min} \le \lambda^k \le 2/3$
    endif
    $x^{k+1} := x^k - \mu^k \Delta x^k$;
    $k := k+1$
  end
return x^k;

The main iteration is written by using projection operator as

$$x^{k+1} = x^k - \mu^k \left\{ \theta \frac{X^k(I - P_{AX^k})X^k c}{\|(I - P_{AX^k})X^k c\|} + X^k P_{AX^k}(X^k)^{-1}(x^k - \hat{x}) \right\} \tag{2.12}$$

where $\hat{x}$ is a solution of $Ax = b$.

2.2. Optimality Direction and Feasibility Direction

In this subsection, we motivate the optimality and feasibility directions. We first discuss the optimality direction. Given a point $x^k > 0$, consider the following LP problem.

$$\langle O(x^k) \rangle \begin{cases} \text{minimize}_x & c^t x \\ \text{subject to} & Ax - b = r^k, \quad x \ge 0 \end{cases} \tag{2.13}$$

where $r^k \triangleq Ax^k - b$ is the residual at x^k. Noting that x^k itself is an interior feasible point of $\langle O(x^k)\rangle$, we define the affine scaling direction at x^k by using the unit ellipsoid in the scaled space, which is the optimality direction (2.4). In other words, this is a short-step affine scaling direction (See (1.2.12)).

As to the feasibility direction, we can interpret it as follows. Given a point x^k, we define a mapping $h : R^n_+ \mapsto R^n_+$ by

$$h(x) = (X^k)^{-1}x. \tag{2.14}$$

Then $h(x^k) = e$, and $\{\, x \mid Ax = b \,\}$ is mapped to $\{\, \chi \mid AX^k\chi = b \,\}$. Let $\tilde{\chi}$ be the nearest point in $\{\, \chi \mid AX^k\chi = b \,\}$ from e, which is the optimal solution of

$$\begin{cases} \text{minimize} & \|\chi - e\|^2 \\ \text{subject to} & AX^k\chi = b. \end{cases} \tag{2.15}$$

It is easy to verify that the feasibility direction is equal to

$$X^k(e - \tilde{\chi}). \tag{2.16}$$

We can also view the feasibility direction as an affine scaling direction. Before showing it, we make some observations. First, we introduce the following LP problem to find a feasible point of $\langle P\rangle$:

$$\langle F\rangle \begin{cases} \text{minimize}_{(x,w)} & w \\ \text{subject to} & Ax - wr^0 = b, \quad x \geq 0, \end{cases} \tag{2.17}$$

where $w \in R$ and $r^0 \triangleq Ax^0 - b$ is the initial residual. Note that w is a free variable. The dual problem of $\langle F\rangle$ is:

$$\langle DF\rangle \begin{cases} \text{minimize} & b^ty \\ \text{subject to} & A^ty \geq 0, \quad (r^0)^ty = 1. \end{cases} \tag{2.18}$$

Obviously, $(x, w) = (x^0, 1)$ is an interior feasible solution of $\langle F\rangle$. If we succeed to reduce w to 0, then we find a feasible solution of $\langle P\rangle$. In this sense, w-component expresses infeasibility. Recall that the residual at x^k is $r^k = Ax^k - b$. In view of (2.4) and (2.5),

$$\begin{aligned} r^{k+1} &= Ax^{k+1} - b = Ax^k - b - \mu^k A\Delta x^k \\ &= r^k - \mu^k A\Delta_n x^k = r^k - \mu^k(Ax^k - b) \\ &= (1 - \mu^k)r^k \end{aligned} \tag{2.19}$$

holds. Hence, if we define

$$w^k \triangleq \prod_{l=0}^{k-1}(1 - \mu^l), \tag{2.20}$$

then

$$r^k = w^k r^0 \tag{2.21}$$

holds and (x^k, w^k) is a feasible solution of $\langle F\rangle$. In particular, if $x^k > 0$ and $w^k > 0$, then (x^k, w^k) is an interior feasible solution of $\langle F\rangle$.

Given an interior feasible point (x, w) of $\langle F\rangle$, let us define $(\tilde{\Delta}x, \tilde{\Delta}w)$ be the affine scaling direction at (x, w) which is scaled so that $\tilde{\Delta}w = w$ holds. Then we have an explicit formula for $\tilde{\Delta}x$ as follows.

LEMMA 2.1.

$$\tilde{\Delta}x = wXP_{AX}X^{-1}(x^0 - \hat{x}) \tag{2.22}$$

where $\hat{x}$ is a solution of $Ax = b$.

Lemma 2.1 is proved in Appendix. Now we show that the feasibility direction is an affine scaling direction.

LEMMA 2.2. *The feasibility direction at x^k is equal to x-components of the affine scaling direction for $\langle F \rangle$ at (x^k, w^k) which is re-scaled so that the w-component is equal to w^k.*

PROOF. We have

$$\begin{aligned} \tilde{\Delta}x^k &= w^k X^k P_{AX^k}(X^k)^{-1}(x^0 - \hat{x}) \\ &= w^k (X^k)^2 A^t (A(X^k)^2 A^t)^{-1} r^0 \\ &= (X^k)^2 A^t (A(X^k)^2 A^t)^{-1}(Ax^k - b) = \Delta_n x^k, \end{aligned} \tag{2.23}$$

which shows that the affine scaling direction for $\langle F \rangle$ is the feasibility direction. □

In view of Lemma 2.1 and Lemma 2.2, we can write the main iteration (2.12) as

$$x^{k+1} = x^k - \mu^k \left\{ \theta \frac{X^k(I - P_{AX^k})X^k c}{\|(I - P_{AX^k})X^k c\|} + w^k X^k P_{AX^k}(X^k)^{-1}(x^0 - \hat{x}) \right\}. \tag{2.24}$$

Let w^* be the optimal value of $\langle F \rangle$. Then the following facts are easily observed by strict complementarity.

PROPOSITION 2.3. (1) *If $w^* > 0$, then* $\mathrm{Feas}\langle P \rangle$ *is empty.*
(2) *If $w^* = 0$, then* $\mathrm{Feas}\langle P \rangle$ *is nonempty but* $\mathrm{Int\,Feas}\langle P \rangle$ *is empty.*
(3) *If $w^* < 0$, then* $\mathrm{Int\,Feas}\langle P \rangle$ *is nonempty.*

2.3. Main Theorem

The following is the main theorem on the primal sequence of the algorithm derived in this paper.

THEOREM 2.4. *If $\langle P \rangle$ has an optimal solution, then the sequence x^k produced by the Algorithm 1 under Step Size Choice 1 converges to a relative interior point of the optimal face. In particular, if $\langle P \rangle$ has interior feasible solutions, the algorithm finds one of them in a finite number of iterations, and then, becomes identical to the feasible long-step affine scaling method with step-size λ^k.*

If $c^t x^k$ is bounded below and $\langle P \rangle$ does not have a feasible solution, then the sequence converges to a relative interior point of the optimal face of $\langle F \rangle$.

In addition to this, we will prove several theorems on two kinds of dual sequence defined later. We discuss the behaviors of dual sequences in Section 8.5 and in Section 8.6, and summarize the results in concluding remarks in Section 10.

We remark that if $c^t x^k$ is not bounded below, then the dual problem $\langle DP \rangle$ does not have a feasible solution, which means $\langle P \rangle$ cannot have an optimal solution. Therefore, Algorithm 1 terminates by

(1) finding an optimal solution,
(2) detecting infeasibility of $\langle P \rangle$, or
(3) detecting infeasibility of $\langle DP \rangle$.

If $c \in \mathrm{Im}(A^t)$, then the optimality direction is always 0 by definition. In this case, we can regard our algorithm as a long-step affine scaling method for finding feasibility and it is easy to see that the above theorem holds in view of the results of Chapter II. To get rid of this trivial case, we make the following assumption throughout the analysis.

ASSUMPTION 2. $c \notin \mathrm{Im}(A^t)$.

If $\mathrm{Feas}\langle P\rangle$ is nonempty, then the above assumption implies that c^tx is not constant on $\mathrm{Feas}\langle P\rangle$.

3. Convergence of the Sequence

We begin the analysis of our algorithm by proving the following theorem.

THEOREM 3.1. *If c^tx^k is bounded below, x^k converges.*

PROOF. Since c^tx^k is bounded below, there exists a positive constant M_1 which satisfies the following:

(3.1)

$$\begin{aligned}
-M_1 &< \sum_{k=0}^{\infty}(c^tx^{k+1} - c^tx^k) \\
&= -\sum_{k=0}^{\infty}\left\{\mu^k\theta\|(I - P_{AX^k})X^kc\| + \mu^kw^kc^tX^kP_{AX^k}(X^k)^{-1}(x^0 - \hat{x})\right\} \\
&\le -\sum_{k=0}^{\infty}\mu^k\theta\|(I - P_{AX^k})X^kc\| + \|c\|M_2\sum_{k=0}^{\infty}\mu^kw^k \quad \text{(Use Lemma 1.4.1)} \\
&\le -\sum_{k=0}^{\infty}\mu^k\theta\|(I - P_{AX^k})X^kc\| + \|c\|M_2\sum_{k=0}^{\infty}(w^k - w^{k+1}) \\
&\le -\sum_{k=0}^{\infty}\mu^k\theta\|(I - P_{AX^k})X^kc\| + \|c\|M_2.
\end{aligned}$$

This implies

$$\sum_{k=0}^{\infty}\mu^k\theta\|(I - P_{AX^k})X^kc\| < M_1 + \|c\|M_2. \tag{3.2}$$

Using the above inequality, we have

$$\begin{aligned}
\sum_{k=0}^{\infty}\|x^{k+1} - x^k\| &\le \sum_{k=0}^{\infty}\mu^k\theta\frac{\|X^k(I - P_{AX^k})X^kc\|}{\|(I - P_{AX^k})Xc\|} + \\
&\qquad \sum_{k=0}^{\infty}\mu^kw^k\|X^kP_{AX^k}(X^k)^{-1}(x^0 - \hat{x})\| \\
&\le M_3\sum_{k=0}^{\infty}\mu^k\theta\|(I - P_{AX^k})Xc\| + M_2\sum_{k=0}^{\infty}\mu^kw^k \\
&\qquad \text{(Use Lemma 1.4.4 and Lemma 1.4.1)} \\
&\le M_3(M_1 + \|c\|M_2) + M_2 \\
&< \infty.
\end{aligned} \tag{3.3}$$

Hence x^k converges. □

If $\langle DP\rangle$ has a feasible solution, $c^t x^k$ is bounded below due to the duality theorem. Theorem 3.1 implies that the sequence will converge independently of the existence of feasible region of $\langle P\rangle$ in this case.

4. Feasibility of the Limit Point

Up to now, we have seen that x^k converges if $c^t x^k$ is bounded below. Let x^∞ be the limit point. The next problem is whether x^∞ is feasible or not. We give an answer to the problem in this section. Let

$$(4.1) \qquad E \triangleq \left\{ j \in \{1,\dots,n\} \mid x_j^\infty = 0 \right\} \quad \text{and} \quad E^c \triangleq \left\{ j \in \{1,\dots,n\} \mid x_j^\infty > 0 \right\}.$$

It is easy to see that the limit point x^∞ cannot be an interior point, thus E is not empty.

This section consists of two subsections. In the first subsection, we investigate properties of feasibility direction in the scaled space. Lemma 4.1 is the main lemma of this subsection, which is used frequently in the consecutive analysis. In the second subsection, we prove that the limit point is feasible if the problem $\langle P\rangle$ has a feasible region (Theorem 4.5), and that the limit point is an optimal solution of $\langle F\rangle$ satisfying strict complementarity if $\langle P\rangle$ has no feasible region (Theorem 4.12).

4.1. Asymptotic Direction for Feasibility

We multiply both sides of the iteration (2.12) by $(X^k)^{-1}$ to obtain

$$(4.2) \qquad (X^k)^{-1}x^{k+1} = e - \mu^k \left\{ \frac{\theta(I - P_{AX^k})X^k c}{\|(I - P_{AX^k})X^k c\|} + P_{AX^k}(X^k)^{-1}(x^k - \hat{x}) \right\}.$$

Since this expresses the iteration in the space transformed by h defined in (2.14), we call it the scaled iteration. Let

$$(4.3) \qquad \alpha^k \triangleq (I - P_{AX^k})X^k c \qquad \text{and}$$

$$(4.4) \qquad \beta^k \triangleq P_{AX^k}(X^k)^{-1}(x^k - \hat{x}).$$

Then we can rewrite the scaled iteration (4.2) as

$$(4.5) \qquad (X^k)^{-1}x^{k+1} = e - \mu^k(\theta \frac{\alpha^k}{\|\alpha^k\|} + \beta^k).$$

The aim of this subsection is to construct vectors which approximate β^k when $x^k \to x^\infty$ by means of Projection Decomposition Lemma (Lemma 1.4.5). (The approximate vector for α^k also plays important role in the consecutive analysis, and will be considered later in Section 6.)

We will prove the following.

LEMMA 4.1. *Assume that w^k is an infinite sequence of positive numbers which converges to $w^\infty \geq 0$, and let $\hat{\beta}_E^k$ and $\hat{\beta}_{E^c}^k$ be the optimal solutions of the following optimization problem:*

$$(4.6) \qquad \langle B_E\rangle \begin{cases} \text{minimize} & \|\beta_E - e\|^2 \\ \text{subject to} & \beta_E = X_E^k A_E^t y, \quad A_{E^c}^t y = 0, \end{cases}$$

and

$$(4.7) \qquad \langle B_{E^c} \rangle \begin{cases} \text{minimize} & \|\beta_{E^c} - (X^k_{E^c})^{-1}(x^k_{E^c} - x^\infty_{E^c})\|^2 \\ \text{subject to} & \beta_{E^c} = X^k_{E^c} A^t_{E^c} y, \end{cases}$$

respectively. Then we have

$$(4.8) \qquad \|\frac{w^k - w^\infty}{w^k}\beta^k_E - \hat{\beta}^k_E\| = \mathcal{O}(\|X^k_E\|^2) + \mathcal{O}(\|X^k_E\|\|x^k_{E^c} - x^\infty_{E^c}\|)$$

and

$$(4.9) \qquad \|\frac{w^k - w^\infty}{w^k}\beta^k_{E^c} - \hat{\beta}^k_{E^c}\| = \mathcal{O}(\|X^k_E\|) + \mathcal{O}(\|X^k_E\|^2\|x^k_{E^c} - x^\infty_{E^c}\|).$$

It might be a helpful observation for understanding Lemma 4.1 that the feasible regions of $\langle B_E \rangle$ and $\langle B_{E^c} \rangle$ are the orthogonal complements of $\{\, \alpha_E \in R^{|E|} \mid A_E X^k_E \alpha_E \in \mathrm{Im}(A_{E^c}) \,\}$ and $\{\, \alpha_{E^c} \in R^{|E^c|} \mid A_{E^c} X^k_{E^c} \alpha_{E^c} = 0 \,\}$, respectively. In fact, as for $\langle B_E \rangle$, let $\{v_1, \ldots, v_{\hat{l}}\}$ be a basis of $\mathrm{Null}(A^t_{E^c})$ where $\hat{l} \triangleq \dim \mathrm{Null}(A^t_{E^c})$, and N be a matrix whose j-th row vector is v^t_j, namely, $N^t = (v_1, \ldots, v_{\hat{l}})$. Putting

$$(4.10) \qquad \tilde{A}_E \triangleq N A_E,$$

we see that

$$(4.11) \qquad \left\{ \alpha_E \mid A_E X^k_E \alpha_E \in \mathrm{Im}(A_{E^c}) \right\} = \left\{ \alpha_E \mid \tilde{A}_E X^k_E \alpha_E = 0 \right\}.$$

Hence, the orthogonal complement of the above is

$$(4.12) \qquad \left\{ \beta_E \mid \beta_E = X^k_E \tilde{A}^t_E y' \right\} = \left\{ \beta_E \mid \beta_E = X^k_E A^t_E y,\ A^t_{E^c} y = 0 \right\}.$$

The above observation suggests that Lemma 1.4.5 will be used to prove Lemma 4.1.

Before proving Lemma 4.1, we characterize β^k as follows.

LEMMA 4.2. *If w^k is an infinite sequence of positive numbers which converges to $w^\infty \geq 0$, then β^k is the optimal solution of*

$$(4.13) \qquad \begin{cases} \text{minimize} & \|\beta - \dfrac{w^k}{w^k - w^\infty}(X^k)^{-1}(x^k - x^\infty)\|^2 \\ \text{subject to} & \beta = X^k A^t y. \end{cases}$$

PROOF. From the discussion in Section 2.2, it is easy to see that β^k is $e - \tilde{\chi}$ (See (2.16)). Since $\tilde{\chi}$ is the optimal solution of (2.15), β^k is the optimal solution of

$$(4.14) \qquad \begin{cases} \text{minimize} & \|\beta\|^2 \\ \text{subject to} & A X^k \beta = w^k r^0. \end{cases}$$

On the other hand, since (x^k, w^k) is a feasible solution of $\langle F \rangle$, we have

$$(4.15) \qquad A x^k - b = w^k r^0$$

and from the assumption,

$$(4.16) \qquad A x^\infty - b = w^\infty r^0.$$

Subtracting (4.16) from (4.15), we have

$$A(x^k - x^\infty) = (w^k - w^\infty)r^0, \tag{4.17}$$

which yields

$$w^k r^0 = \frac{w^k A(x^k - x^\infty)}{w^k - w^\infty}. \tag{4.18}$$

Hence, (4.14) can be written as

$$\begin{cases} \text{minimize} & \|\beta\|^2 \\ \text{subject to} & AX^k\beta = \dfrac{w^k A(x^k - x^\infty)}{w^k - w^\infty}. \end{cases} \tag{4.19}$$

By solving Karush-Kuhn-Tucker conditions of the above optimization problem, we have

$$\beta^k = \frac{w^k}{w^k - w^\infty} P_{AX^k}(X^k)^{-1}(x^k - x^\infty). \tag{4.20}$$

Obviously this is the optimal solution of (4.13). □

Now we prove Lemma 4.1.

PROOF OF LEMMA 4.1. Let $\bar{\gamma}^k$ be the optimal solution of

$$\begin{cases} \text{minimize} & \|\gamma - (X^k)^{-1}(x^k - x^\infty)\|^2 \\ \text{subject to} & AX^k\gamma = 0. \end{cases} \tag{4.21}$$

Furthermore, let $\hat{\gamma}_E^k$ and $\hat{\gamma}_{E^c}^k$ be the optimal solutions of

$$\begin{cases} \text{minimize} & \|\gamma_E - e\|^2 \\ \text{subject to} & A_E X_E^k \gamma_E \in \mathrm{Im}(A_{E^c}) \end{cases} \tag{4.22}$$

and

$$\begin{cases} \text{minimize} & \|\gamma_{E^c} - (X_{E^c}^k)^{-1}(x_{E^c}^k - x_{E^c}^\infty)\|^2 \\ \text{subject to} & A_{E^c} X_{E^c}^k \gamma_{E^c} = 0, \end{cases} \tag{4.23}$$

respectively. We note that the feasible region of (4.21) is the orthogonal complement of that of (4.13). In view of Lemma 4.2, it is easy to see:

$$\frac{w^k - w^\infty}{w^k}\beta^k + \bar{\gamma}^k = e - (X^k)^{-1}x^\infty. \tag{4.24}$$

Recalling that the remark after Lemma 4.1, it is also easy to see:

$$\hat{\beta}_E^k + \hat{\gamma}_E^k = e, \tag{4.25}$$

and

$$\hat{\beta}_{E^c}^k + \hat{\gamma}_{E^c}^k = e - (X_{E^c}^k)^{-1}x_{E^c}^\infty. \tag{4.26}$$

On the other hand, it follows from Lemma 1.4.5 that

$$\|\bar{\gamma}_E^k - \hat{\gamma}_E^k\| = \mathcal{O}(\|X_E^k\|^2) + \mathcal{O}(\|X_E^k\|\|(X_{E^c}^k)^{-1}(x_{E^c}^k - x_{E^c}^\infty)\|) \tag{4.27}$$

and

$$\|\bar{\gamma}^k_{E^c} - \hat{\gamma}^k_{E^c}\| = \mathcal{O}(\|X^k_E\|) + \mathcal{O}(\|X^k_E\|^2 \|(X^k_{E^c})^{-1}(x^k_{E^c} - x^\infty_{E^c})\|). \tag{4.28}$$

Thus the lemma easily follows by substituting the relations (4.24), (4.25) and (4.26) into (4.27) and (4.28). □

COROLLARY 4.3. *If $w^\infty = 0$, then*

$$\|\beta^k_E - \hat{\beta}^k_E\| = \mathcal{O}(\|X^k_E\|^2) + \mathcal{O}(\|X^k_E\| \|x^k_{E^c} - \hat{x}_{E^c}\|) \tag{4.29}$$

and

$$\|\beta^k_{E^c} - \hat{\beta}^k_{E^c}\| = \mathcal{O}(\|X^k_E\|) + \mathcal{O}(\|X^k_E\|^2 \|x^k_{E^c} - \hat{x}_{E^c}\|). \tag{4.30}$$

The next lemma shows a basic property of $\hat{\beta}^k_E$.

LEMMA 4.4. *If $\hat{\beta}^k_E \geq 0$ for some k, then x^∞ is the optimal solution of $\langle F \rangle$. In particular, if $\hat{\beta}^k_E > 0$, then x^∞ is a relative interior point of the optimal face of $\langle F \rangle$.*

PROOF. Let $(\hat{\beta}^k_E, \hat{y}^k)$ be the optimal solution of $\langle B_E \rangle$. By using (4.18), we have

$$\begin{aligned} w^k (r^0)^t \hat{y}^k &= \frac{w^k (x^k - x^\infty)^t}{w^k - w^\infty} A^t \hat{y}^k \\ &= \frac{w^k (x^k_E)^t A^t_E \hat{y}^k}{w^k - w^\infty} \\ &= \frac{w^k e^t X^k_E A^t_E \hat{y}^k}{w^k - w^\infty} \\ &= \frac{w^k \|\hat{\beta}^k_E\|^2}{w^k - w^\infty}. \end{aligned} \tag{4.31}$$

Putting

$$\bar{y} \triangleq \frac{(w^k - w^\infty)\hat{y}^k}{\|\hat{\beta}^k_E\|^2}, \tag{4.32}$$

we have

$$(r^0)^t \bar{y} = 1 \qquad \text{and} \qquad A^t_{E^c} \bar{y} = 0. \tag{4.33}$$

If $\hat{\beta}^k_E \geq 0$, then we have $A^t_E \bar{y} \geq 0$, which implies that x^∞ and $\bar{y}$ satisfy the complementarity condition of $\langle F \rangle$ and $\langle DF \rangle$. If $\hat{\beta}^k_E > 0$, then the strict complementarity is satisfied. □

4.2. Proof of Feasibility of the Limit Point

We will prove the following theorem.

THEOREM 4.5. *If x^k converges and $\langle P \rangle$ has a feasible region, then $w^k \to 0$.*

The case where the feasible region of $\langle P \rangle$ is empty will be also analyzed in the end of this subsection.

To prove Theorem 4.5, we show the following two lemmas.

LEMMA 4.6. *If $x^k \to x^\infty$ and $w^k \to w^\infty > 0$, then there exists a constant M such that for all k,*

$$\|x_E^k\| \leq M(w^k - w^\infty). \tag{4.34}$$

PROOF.

(4.35)

$$\begin{aligned}
\|x_E^k\| &\leq \sum_{l=k}^{\infty} \|x_E^{l+1} - x_E^l\| \\
&\leq \sum_{l=k}^{\infty} \|x^{l+1} - x^l\| \\
&\leq \sum_{l=k}^{\infty} \mu^l \theta \frac{\|X^l(I - P_{AX^l})X^l c\|}{\|(I - P_{AX^l})X^l c\|} + \sum_{l=k}^{\infty} \mu^l w^l \|X^l P_{AX^l}(X^l)^{-1}(x^0 - \hat{x})\| \\
&\leq M_1 \sum_{l=k}^{\infty} \mu^l \|(I - P_{AX^l})X^l c\| + M_2 \sum_{l=k}^{\infty} \mu^l w^l \\
&\leq \sum_{l=k}^{\infty} \mu^l w^l \left(\frac{M_1 \|(I - P_{AX^l})X^l c\|}{w^\infty} + M_2 \right) \\
&\leq \sum_{l=k}^{\infty} \mu^l w^l M \qquad \text{(since } \|x^k\| \text{ is bounded)} \\
&= M(w^k - w^\infty)
\end{aligned}$$

which completes the proof. □

LEMMA 4.7. *If $x^k \to x^\infty$, $w^k \to w^\infty > 0$ and $\langle P \rangle$ has a feasible region, then the local Karmarkar potential function for feasibility defined by*

$$f_E(x^k) \triangleq |E| \log(w^k - w^\infty) - \sum_{j \in E} \log x_j^k \tag{4.36}$$

diverges to $-\infty$.

Once Lemma 4.7 is proved, the proof of Theorem 4.5 is easy.

PROOF OF THEOREM 4.5. The proof is by contradiction. Suppose that $w^k \to w^\infty > 0$. Then from Lemma 4.6, using the well-known inequality between arithmetic and geometric mean, we obtain

$$\exp(f_E(x^k)) = \frac{(w^k - w^\infty)^{|E|}}{\prod_{j\in E} x_j^k} \geq \left(|E|\frac{w^k - w^\infty}{e^t x^k}\right)^{|E|} \tag{4.37}$$
$$\geq \left(\sqrt{|E|}\frac{w^k - w^\infty}{\|x_E\|}\right)^{|E|} \geq \frac{1}{M} > 0 \qquad \text{for all } k.$$

On the other hand, Lemma 4.7 implies that $\exp f_E(x^k) \to 0$. These two facts contradicts each other, hence the assumption that $w^\infty > 0$ should be denied. □

To prove Lemma 4.7, we prepare a few lemmas.

LEMMA 4.8. *If $x^k \to x^\infty$ and $w^k \to w^\infty > 0$, then $\sigma((X^k)^{-1}\Delta x) = \sigma((X_E^k)^{-1}\Delta x_E)$ holds for sufficiently large k.*

PROOF. Since $w^\infty > 0$, we see that $\mu^k \to 0$ in view of (2.20). This implies that $\sigma((X^k)^{-1}\Delta x) \to \infty$. Obviously, $\alpha^k/\|\alpha^k\|$ is bounded. Since $\Delta_n x$ is bounded,

$$\beta_{E^c}^k = (X_{E^c}^k)^{-1}\Delta_n x_{E^c} \tag{4.38}$$

is also bounded. Thus $\sigma(\beta_E^k)$ must diverge to infinity. Now the lemma readily follows. □

LEMMA 4.9. *If $x^k \to x^\infty$ and $w^k \to w^\infty > 0$, then there exist some constants δ_1 and δ_2 such that*

$$1 \geq \frac{\mu^k w^k}{w^k - w^\infty} \geq \delta_1 > 0, \qquad \textit{and} \qquad \sigma(\hat{\beta}_E^k) \geq \delta_2 > 0 \tag{4.39}$$

for sufficiently large k. In addition,

$$\frac{\mu^k w^k}{w^k - w^\infty} = \frac{\lambda^k}{\sigma(\hat{\beta}_E^k)} + \mathcal{O}(w^k - w^\infty). \tag{4.40}$$

PROOF. Since $w^k - w^\infty \geq w^k - w^{k+1} = \mu^k w^k$,

$$1 \geq \frac{\mu^k w^k}{w^k - w^\infty} \tag{4.41}$$

follows.

We have by using Lemma 4.1 and Lemma 4.8,

$$\mu^k = \frac{\lambda^k}{\sigma(\theta\alpha^k/\|\alpha^k\| + \beta^k)} = \frac{\lambda^k}{\sigma(\theta\alpha_E^k/\|\alpha^k\| + \beta_E^k)} \tag{4.42}$$
$$= \frac{\lambda^k}{\sigma\left(\theta\alpha_E^k/\|\alpha^k\| + \dfrac{w^k}{w^k - w^\infty}(\hat{\beta}_E^k + \mathcal{O}(\|X_E^k\|))\right)}$$

for sufficiently large k. Taking note that $\|X_E^k\| = \mathcal{O}(w^k - w^\infty)$ due to Lemma 4.6, we see that

$$1 \geq \frac{\mu^k w^k}{w^k - w^\infty} = \frac{\lambda^k}{\sigma(\hat{\beta}_E^k) + \mathcal{O}(w^k - w^\infty)} > 0. \tag{4.43}$$

Therefore,

$$\sigma(\hat{\beta}_E^k) + \mathcal{O}(w^k - w^\infty) \geq \lambda^k \geq \lambda_{\min} > 0 \tag{4.44}$$

holds and we have

$$\sigma(\hat{\beta}_E^k) \gtrsim \lambda_{\min}/2 > 0 \tag{4.45}$$

for sufficiently large k.

On the other hand, since $\hat{\beta}_E^k$ is the optimal solution of $\langle B_E \rangle$, $\hat{\beta}_E^k$ is bounded, thus (4.40) and the former part of (4.39) readily follow from (4.43). □

LEMMA 4.10. *Assume that* $\nu \in R$, $\eta \in R^q$, $\Delta_j \in R$, $j = 0, \ldots, q$, *and* $\Delta_0' \in R$ *satisfy:*

$$e^t\eta = 1 + \Delta_0, \tag{4.46}$$

$$\nu\sigma(\eta) \leq 2/3, \tag{4.47}$$

$$1 - \nu\|\eta\|^2 + \Delta_0' > 0, \tag{4.48}$$

$$|\Delta_j| \leq 1/4, \qquad \text{for } j = 0, \ldots, q. \tag{4.49}$$

Then the following inequality holds:

(4.50)

$$\begin{aligned} &q \log(1 - \nu\|\eta\|^2 + \Delta_0') - \sum_{j=1}^{q} \log(1 - \nu\eta_j + \Delta_j) \\ &\leq \frac{q\nu}{q-\nu}\|\eta - \frac{e}{q}\|^2 \left\{-q + \frac{\nu}{2(1-\nu\sigma(\eta))}\right\} + M(|\Delta_0| + |\Delta_0'| + \sum_{j=1}^{q} |\Delta_j|) \end{aligned}$$

where M is a positive constant depending only on q.

Note that the left hand side is well-defined due to (4.48), (4.47) and (4.49). The above lemma is slightly stronger than what we need to prove Lemma 4.7. However, we will use this lemma again in the proof of Lemma 8.18, and there we need Lemma 4.10 in its form.

We will prove the above lemma in Appendix.

The following lemma is substantial to the proof of Lemma 4.7.

LEMMA 4.11. *If $x^k \to x^\infty$ and $w^k \to w^\infty > 0$, then the difference of the local Karmarkar potential function for feasibility is bounded as follows for sufficiently large k:*

(4.51)

$$\begin{aligned} \Delta f_E(x^k) &\triangleq f_E(x^{k+1}) - f_E(x^k) \\ &\leq \frac{|E|\bar{\nu}^k}{|E| - \bar{\nu}^k}\|\bar{\eta}^k - \frac{e}{|E|}\|^2 \left\{-|E| + \frac{\bar{\nu}^k}{2(1 - \bar{\nu}^k\sigma(\bar{\eta}^k))}\right\} + \mathcal{O}(w^k - w^\infty) \\ &\leq -\epsilon\|\bar{\eta}^k - \frac{e}{|E|}\|^2 + \mathcal{O}(w^k - w^\infty) \end{aligned}$$

where

$$\bar{\eta}_E^k \triangleq \frac{\hat{\beta}_E^k}{\|\hat{\beta}_E^k\|^2}, \qquad \bar{\nu}^k \triangleq \frac{\lambda^k\|\hat{\beta}_E^k\|^2}{\sigma(\hat{\beta}_E^k)} = \frac{\lambda^k}{\sigma(\bar{\eta}_E^k)} \tag{4.52}$$

and ϵ is a positive constant.

PROOF. The E-components of the scaled iteration can be written by using Lemma 4.1 as follows:

$$(4.53)\qquad \begin{aligned}(X_E^k)^{-1}x_E^{k+1} &= e - \mu^k\left\{\theta\frac{\alpha_E^k}{\|\alpha^k\|} + \frac{w^k}{w^k - w^\infty}(\hat\beta_E^k + \mathcal{O}(\|X_E^k\|))\right\}\\ &= e - \frac{\mu^k w^k}{w^k - w^\infty}(\hat\beta_E^k + \mathcal{O}(w^k - w^\infty))\\ &= e - \frac{\lambda^k\hat\beta_E^k}{\sigma(\hat\beta_E^k)} + \mathcal{O}(w^k - w^\infty).\end{aligned}$$

In the last equality, we used (4.40) of Lemma 4.9. Since

$$(4.54)\qquad e^t\bar\eta_E^k = 1, \qquad \text{and}$$

$$(4.55)\qquad \bar\nu^k\sigma(\bar\eta_E^k) = \lambda^k \le 2/3 - \hat\delta,$$

we apply Lemma 4.10 with $\Delta_0 = 0$ as follows.

(4.56)

$$\begin{aligned}\Delta f_E(x^k) &= |E|\log\left\{1 - \frac{\mu^k w^k}{w^k - w^\infty}\right\} - \sum_{j\in E}\log\left\{1 - \frac{\mu^k w^k}{w^k - w^\infty}(\hat\beta_j^k + \mathcal{O}(w^k - w^\infty))\right\}\\ &= |E|\log\left\{1 - \frac{\lambda^k}{\sigma(\hat\beta_E^k)} + \mathcal{O}(w^k - w^\infty)\right\} - \sum_{j\in E}\log\left\{1 - \frac{\lambda^k\hat\beta_j^k}{\sigma(\hat\beta_E^k)} + \mathcal{O}(w^k - w^\infty)\right\}\\ &= |E|\log\left\{1 - \bar\nu^k\|\bar\eta_E^k\|^2 + \mathcal{O}(w^k - w^\infty)\right\} - \sum_{j\in E}\log\left\{1 - \bar\nu^k\bar\eta_j^k + \mathcal{O}(w^k - w^\infty)\right\}\\ &\le \frac{|E|\bar\nu^k}{|E| - \bar\nu^k}\|\bar\eta^k - \frac{e}{|E|}\|^2\left\{-|E| + \frac{\bar\nu^k}{2(1 - \bar\nu^k\sigma(\bar\eta^k))}\right\} + \mathcal{O}(w^k - w^\infty).\end{aligned}$$

This proves the first inequality of (4.51). Furthermore, from (4.54)

$$(4.57)\qquad \sigma(\bar\eta_E^k) \ge 1/|E|$$

follows. Hence

$$(4.58)\qquad \begin{aligned}-|E| + \frac{\bar\nu^k}{2(1 - \bar\nu^k\sigma(\bar\eta_E^k))} &= -|E| + \frac{\lambda^k}{2\sigma(\bar\eta_E^k)(1 - \lambda^k)}\\ &< -|E| + |E|\frac{2/3 - \hat\delta}{2/3 + 2\hat\delta} \le -\delta' < 0\end{aligned}$$

holds for a constant δ' under *Step Size Choice 1*. We also have

$$(4.59)\qquad |E| - \bar\nu^k \ge |E|/3$$

and $\bar\nu^k$ is bounded below by a positive constant since $\sigma(\hat\beta_E^k)$ is bounded below by a positive constant due to (4.39). Now the lemma follows from (4.56) □

Now we are ready to prove Lemma 4.7.

PROOF OF LEMMA 4.7. Due to Proposition 2.3, w^∞ cannot be the optimal value of $\langle F\rangle$ (recall that we assumed the existence of a feasible solution of $\langle P\rangle$ which has smaller value of w), and E does not determine the optimal face of $\langle F\rangle$. Hence, from Lemma 4.4, we see that $\hat\beta_E^k \not> 0$. If $|E| = 1$, then from (4.54) $\hat\beta_E^k > 0$ follows, which contradicts the above observation. Therefore, we assume $|E| \geq 2$. In this case, from (4.54)

$$\|\bar\eta_E^k - e/|E|\|^2 \geq \frac{1}{|E|(|E|-1)} \tag{4.60}$$

follows. Then from Lemma 4.11 we see that the difference of $\Delta f_E(x^k)$ must be smaller than a negative constant for sufficiently large k. Therefore, $f_E(x^k) \to -\infty$ as $k \to \infty$, and this completes the proof. □

Next, we deal with the case when the feasible region of $\langle P\rangle$ is empty.

THEOREM 4.12. *If $\langle P\rangle$ has no feasible solution and $c^t x^k$ is bounded below, then $w^k \to w^*$ and $x^k \to x^\infty$ where x^∞ is a relative interior point of the optimal face of $\langle F\rangle$.*

PROOF. In this case, Lemma 4.6 holds, thus the local Karmarkar potential function is bounded below. From (4.39),

$$\frac{w^{k+1} - w^\infty}{w^k - w^\infty} = 1 - \frac{\mu^k w^k}{w^k - w^\infty} \leq 1 - \delta_1 \tag{4.61}$$

holds for sufficiently large k. This implies

$$\sum_{k=0}^{\infty}(w^k - w^\infty) < \infty. \tag{4.62}$$

In view of Lemma 4.11 (cf. (4.51)), if $\bar\eta_E^k$ does not converge to $e/|E|$, then $f_E(x^k) \to -\infty$ which contradicts Lemma 4.6. Hence, $\bar\eta_E^k \to e$ follows, thus there exist some k such that $\hat\beta_E^k > 0$. Now the theorem follows from Lemma 4.4. □

5. Convergence to a Point on a Dual-Degenerate Face

We have proved that if $c^t x^k$ is bounded below and $\langle P\rangle$ is infeasible, then the sequence converges to an optimal solution of $\langle F\rangle$ (Theorem 4.12). When $c^t x^k$ is not bounded from below, we find the infeasibility of $\langle DP\rangle$. In both cases, there is no optimal solution of $\langle P\rangle$.

We have also proved that if $c^t x^k$ is bounded below and $\langle P\rangle$ is feasible, then the limit point of the sequence is a feasible solution of $\langle P\rangle$. Therefore, our concern naturally turns to the optimality of the limit point in this case. In the rest of the paper we deal with the case where x^k converges to a feasible point. In other words, we assume

ASSUMPTION 3. $w^k \to 0$

in the consecutive analysis except in Theorem 8.25 where we discuss the convergence of dual sequence when $\langle P\rangle$ is infeasible.

If the algorithm finds an interior feasible point in a finite number of iterations, then the proposed method becomes identical to a long-step affine scaling method, thus the optimality of the limit point and the global convergence of the dual estimates have been already established in Chapter II. Hence, we assume

ASSUMPTION 4. *the algorithm produces an infinite sequence of infeasible points*

from now on. Recalling *Step Size Choice 1*, we can easily see that Assumption 4 implies

$$\sigma((X^k)^{-1}\Delta x^k) \geq 1 \qquad \text{for all } k. \tag{5.1}$$

A face of Feas$\langle P\rangle$ is called dual-degenerate if the objective function value is constant on the face. The aim of this section is to prove the following lemma.

LEMMA 5.1. *If x^k converges to a feasible point of $\langle P\rangle$, then x^∞ is contained in a dual-degenerate face.*

To prove Lemma 5.1, we first show the following lemma.

LEMMA 5.2. *$\sigma((X^k)^{-1}\Delta x)$ is bounded above.*

PROOF. It is obvious that the first term of $(X^k)^{-1}\Delta x^k$ in the expression of (4.5),

$$\theta\frac{(I-P_{AX^k})X^kc}{\|(I-P_{AX^k})X^kc\|} \tag{5.2}$$

is bounded by 1. Since the second term can be written as

$$w^kP_{AX^k}(X^k)^{-1}(x^0-\hat{x}) = w^k(X^k)^{-1}X^kP_{AX^k}(X^k)^{-1}(x^0-\hat{x}) \tag{5.3}$$

and $X^kP_{AX^k}(X^k)^{-1}(x^0-\hat{x})$ is bounded due to Lemma 1.4.1, we see that $(X^k_{E^c})^{-1}\Delta_n x^k_{E^c} \to 0$.

From Corollary 4.3, we have

$$\sigma((X^k_E)^{-1}\Delta_n x^k_E) = \sigma(\hat{\beta}^k_E) + \mathcal{O}(\|x^k_E\|) \tag{5.4}$$

where $\hat{\beta}^k_E$ is the projection of c. This shows that $\sigma((X^k_E)^{-1}\Delta_n x^k_E)$ is bounded. Now the lemma readily follows. □

COROLLARY 5.3. *There exists a constant δ such that $\mu^k \geq \delta > 0$ for all k.*

Next we show:

LEMMA 5.4.

$$\|(I-P_{AX^k})X^kc\| \to 0 \tag{5.5}$$

PROOF. We make the following inequality due to the boundedness of objective function value:

$$
\begin{aligned}
(5.6) \qquad -\infty < c^t(x^\infty - x^0) &= \sum_{k=0}^{\infty}(c^t x^{k+1} - c^t x^k) \\
&\leq -\sum_{k=0}^{\infty} \mu^k \theta \|(I - P_{AX^k})X^k c\| \\
&\quad - \sum_{k=0}^{\infty} \mu^k w^k c^t X^k P_{AX^k}(X^k)^{-1}(x^0 - \hat{x}) \\
&\leq -M_1 \sum_{k=0}^{\infty} \|(I - P_{AX^k})X^k c\| + M_2 \\
&\qquad \text{(Use Lemma 1.4.1 and Corollary 5.3)}
\end{aligned}
$$

where M_1 and M_2 are positive constants. This implies $\sum_{k=0}^{\infty} \|(I - P_{AX^k})X^k c\| < \infty$, thus the lemma follows. □

Now we prove Lemma 5.1.

PROOF OF LEMMA 5.1. Let

$$
(5.7) \qquad s^k \triangleq (X^k)^{-1}(I - P_{AX^k})X^k c,
$$

which is the dual estimate for optimality. We can also express it by using an appropriate y^k as

$$
(5.8) \qquad s^k = c - A^t y^k.
$$

Since Lemma 1.4.1 implies s^k is bounded, s^k has a convergent subsequence. Let s^* be an accumulation point of s^k. Then, from Lemma 5.4,

$$
(5.9) \qquad \|X^k s^k\| \to 0,
$$

thus $s^*_{E^c} = 0$ follows. This implies $c_{E^c} \in \operatorname{Im} A^t_{E^c}$, which means that the face defined by E is dual-degenerate. □

6. Convergence Rates of Infeasibility Criteria and Objective Function

We investigate convergence rates of w^k and $c^t x^k - c^t x^\infty$ compared with $\|x^k_E\|$ in this section. As was stated in the previous section, we deal with the case where x^k is an infinite sequence of infeasible points and converges to a feasible point. We also know that the limit point is in a dual degenerate face due to Lemma 5.1.

We will prove the following two lemmas:

LEMMA 6.1.

$$
(6.1) \qquad \frac{w^k}{\|x^k_E\|} \to 0.
$$

LEMMA 6.2. *There exist some constants δ_1 and δ_2 such that*

$$\frac{c^t x^k - c^t x^\infty}{\|x_E^k\|} \geq \delta_1 > 0 \quad \text{and} \quad \frac{c^t x^k - c^t x^\infty}{\|x_{E^c}^k - x_{E^c}^\infty\|} \geq \delta_2 > 0 \tag{6.2}$$

hold for sufficiently large k.

Note that $c^t x^k$ may be smaller than $c^t x^\infty$ in general. Lemma 6.2 however, implies that $c^t x^k$ approaches $c^t x^\infty$ from above in the final stage of the convergence.

We will prove the following lemma.

LEMMA 6.3.

$$f_E(x^k) \to -\infty \tag{6.3}$$

where

$$\begin{aligned} f_E(x^k) &= |E| \log(w^k - w^\infty) - \sum_{j \in E} \log x_j^k \\ &= |E| \log w^k - \sum_{j \in E} \log x_j^k \end{aligned} \tag{6.4}$$

is the local Karmarkar potential function for feasibility.

Once Lemma 6.3 is proved, then Lemma 6.1 follows. Lemma 6.2 will be derived from Lemma 6.1. Thus our first goal is to prove Lemma 6.3.

One may think that Lemma 6.3 is similar to Lemma 4.7. An important difference is that w^∞ is zero in Lemma 6.3 while w^∞ is assumed to be positive in Lemma 4.7. Because of this, we have a new difficulty in proving Lemma 6.3 that we should take care of α_E^k which we could ignore in (4.53) in the proof of Lemma 4.7.

The rest of this section consists of three subsections. In the first subsection, after proving a basic proposition about the objective function $c^t x$ on Feas$\langle F \rangle$ (Proposition 6.4), we construct an approximate vector for α^k (Lemma 6.6) and express the scaled iteration in terms of the approximate vector (Lemma 6.8). These results will be used in the consecutive sections as well as in the proof of Lemma 6.3 and Lemma 6.2. We prove Lemma 6.3 in the second subsection, and Lemma 6.2 in the third subsection.

6.1. Preliminary Observation

First, we consider the objective function value $c^t x$ on Feas$\langle F \rangle$.

PROPOSITION 6.4. *There exist a vector $\hat{c}_E \in R^{|E|}$ and a real number $\hat{M}$ such that*

$$c^t x - c^t x^\infty = \hat{c}_E^t x_E - w\hat{M} \tag{6.5}$$

for all $(x, w) \in$ Feas$\langle F \rangle$.

PROOF. Let $\grave{y}$ be a vector such that $c_{E^c} = A_{E^c}^t \grave{y}$, whose existence is guaranteed by Lemma 5.1. Then we see:

$$\begin{aligned} c^t x &= c_E^t x_E + c_{E^c}^t x_{E^c} = c_E^t x_E + \grave{y}^t A_{E^c} x_{E^c} \\ &= c_E^t x_E + \grave{y}^t (b + w r^0 - A_E x_E) \\ &= (c_E - A_E^t \grave{y})^t x_E + w \grave{y}^t r^0 + \grave{y}^t b \end{aligned} \tag{6.6}$$

for any $(x, w) \in \mathrm{Feas}\langle F\rangle$. Letting $x = x^k$ and $w = w^k$, and taking their limits, we obtain $c^t x^\infty = \grave{y}^t b$. Setting $\hat{c}_E = c_E - A^t_{E^c}\grave{y}$ and $\hat{M} = -\grave{y}^t r^0$, we obtain the lemma. □

Due to the above proposition, we observe the following.

LEMMA 6.5. *α^k is the optimal solution of*

$$\left\{\begin{array}{ll}\text{minimize} & \|\alpha - \begin{pmatrix} X^k_E \hat{c}_E \\ 0 \end{pmatrix}\|^2 \\ \text{subject to} & AX^k\alpha = 0.\end{array}\right. \tag{6.7}$$

PROOF. Since

$$\begin{aligned} X^k c - \begin{pmatrix} X^k_E \hat{c}_E \\ 0 \end{pmatrix} &= \begin{pmatrix} X^k_E A^t_E \grave{y} \\ X^k_{E^c} c_{E^c} \end{pmatrix} \\ &= X^k A^t \grave{y} \in \mathrm{Im}(X^k A^t), \end{aligned} \tag{6.8}$$

and α^k is a projection of $X^k c$ onto $\mathrm{Null}(AX^k)$, the lemma readily follows. □

Let $\hat{\alpha}^k_E$ be the optimal solution of

$$\langle A^k_E\rangle \left\{\begin{array}{ll}\text{minimize} & \|\alpha_E - X^k_E \hat{c}_E\|^2 \\ \text{subject to} & A_E X^k_E \alpha_E \in \mathrm{Im}(A_{E^c}).\end{array}\right. \tag{6.9}$$

We recall that the feasible region of $\langle A^k_E\rangle$ is the orthogonal complement of that of $\langle B^k_E\rangle$ (See the remark following Lemma 4.1). Therefore, $(\hat{\beta}^k_E)^t\hat{\alpha}^k_E = 0$. We also have:

LEMMA 6.6.

$$\alpha^k_E = \hat{\alpha}^k_E + \mathcal{O}(\|X^k_E\|^2\|\hat{\alpha}^k_E\|), \qquad \textit{and} \tag{6.10}$$

$$\alpha^k_{E^c} = \mathcal{O}(\|X^k_E\|\|\hat{\alpha}^k_E\|). \tag{6.11}$$

PROOF. From Lemma 6.5 and Lemma 1.4.5, the lemma readily follows. □

LEMMA 6.7.

$$\frac{\alpha^k_E}{\|\alpha^k\|} = \frac{\hat{\alpha}^k_E}{\|\hat{\alpha}^k_E\|} + \mathcal{O}(\|X^k_E\|^2), \tag{6.12}$$

PROOF. Lemma 6.6 implies

$$\begin{aligned} \|\alpha^k\|^2 = \|\alpha^k_E\|^2 + \|\alpha^k_{E^c}\|^2 &= \|\hat{\alpha}^k_E\|^2 + \mathcal{O}(\|X^k_E\|^2\|\hat{\alpha}^k_E\|^2) \\ &= \|\hat{\alpha}^k_E\|^2(1 + \mathcal{O}(\|X^k_E\|^2)). \end{aligned} \tag{6.13}$$

Hence,

$$\frac{\|\alpha^k\|}{\|\hat{\alpha}^k_E\|} = 1 + \mathcal{O}(\|X^k_E\|^2). \tag{6.14}$$

Therefore we have

$$\frac{\alpha_E^k}{\|\alpha^k\|} = \frac{\hat{\alpha}_E^k + \mathcal{O}(\|X_E^k\|^2\|\hat{\alpha}_E^k\|)}{\|\hat{\alpha}_E^k\|(1+\mathcal{O}(\|X_E^k\|^2))} \tag{6.15}$$

$$= \frac{\hat{\alpha}_E^k}{\|\hat{\alpha}_E^k\|}(1-\mathcal{O}(\|X_E^k\|^2)) + \frac{\mathcal{O}(\|X_E^k\|^2)}{1+\mathcal{O}(\|X_E^k\|^2)}$$

$$= \frac{\hat{\alpha}_E^k}{\|\hat{\alpha}_E^k\|} + \mathcal{O}(\|X_E^k\|^2), \tag{6.16}$$

which completes the proof. □

Next, we check the scaled iteration.

LEMMA 6.8. *The E-components of the scaled iteration is written as*

$$(X_E^k)^{-1}x_E^{k+1} = e - \mu^k\left\{\frac{\theta\hat{\alpha}_E^k}{\|\hat{\alpha}_E^k\|} + \hat{\beta}_E^k + \mathcal{O}(\|X_E^k\|)\right\}. \tag{6.17}$$

PROOF. From Corollary 4.3, we have

$$\hat{\beta}_E^k = \beta_E^k + \mathcal{O}(\|X_E^k\|). \tag{6.18}$$

Combining (6.18) and Lemma 6.7, we obtain the lemma. □

6.2. Proof of Lemma 6.3

We first observe that $\hat{\beta}_E^k \not> 0$ for all k. Suppose contrary that $\hat{\beta}_E^k > 0$ for some k. Then from Lemma 4.4, x^∞ must be a relative interior point of the optimal face of $\langle F\rangle$. If there exists an interior feasible solution, it cannot occur. (To see this, recall that the optimal value w^* of $\langle F\rangle$ is negative while $w^\infty = 0$. Thus, (x^∞, w^∞) cannot be an optimal solution of $\langle F\rangle$.) Even if there exists no interior feasible solution, the limit point is contained in a dual-degenerate face of $\langle P\rangle$ which is a proper subset of the optimal face of $\langle F\rangle$ due to Assumption 2. Hence, the limit point cannot be a relative interior point of the optimal face of $\langle F\rangle$. Therefore, we have $\hat{\beta}_E^k \not> 0$. This implies that $\|e - \hat{\beta}_E^k\| \geq 1$.

Let

$$\gamma_E^k \triangleq \frac{\theta\hat{\alpha}_E^k}{\|\hat{\alpha}_E^k\|} - (e - \hat{\beta}_E^k). \tag{6.19}$$

Since $\hat{\beta}_E^k$ is a projection of e and $(\hat{\beta}_E^k)^t\hat{\alpha}_E^k = 0$ (see the remark following (6.9)), we have

$$e^t\gamma_E^k = \frac{\theta e^t\hat{\alpha}_E^k}{\|\hat{\alpha}_E^k\|} - \|e - \hat{\beta}_E^k\|^2, \tag{6.20}$$

and

$$\|\gamma_E^k\|^2 = \theta^2 + \|e - \hat{\beta}_E^k\|^2 - 2\theta\frac{e^t\hat{\alpha}_E^k}{\|\hat{\alpha}_E^k\|}. \tag{6.21}$$

The fact that $\|\gamma_E^k\|^2 \geq 0$ implies that

$$\frac{\theta e^t\hat{\alpha}_E^k}{\|\hat{\alpha}_E^k\|} - \|e - \hat{\beta}_E^k\|^2 \leq \frac{1}{2}(\theta^2 - \|e - \hat{\beta}_E^k\|^2) < 0. \tag{6.22}$$

From Lemma 6.8, the scaled iteration can be written by using γ_E^k as follows:

$$(6.23)\qquad (X_E^k)^{-1}x_E^{k+1} = (1-\mu^k)\left\{e - \frac{\mu^k}{1-\mu^k}\gamma_E^k + \mathcal{O}(\|X_E^k\|)\right\}.$$

Noting that $\mu^k \le 2/3 - \hat{\delta}$ (due to *Step Size Choice 1* and (5.1)), we give a bound for the difference of the potential function for feasibility as follows.

$$(6.24)\qquad \begin{aligned}\Delta f_E(x^k) &\triangleq f_E(x^{k+1}) - f_E(x^k)\\ &= |E|\log(w^{k+1}/w^k) - \sum_{j\in E}\log(x_j^{k+1}/x_j^k)\\ &= -\sum_{j\in E}\log\left\{1 - \frac{\mu^k}{1-\mu^k}\gamma_j^k + \mathcal{O}(\|X_E^k\|)\right\}\\ &\le \frac{\mu^k}{1-\mu^k}e^t\gamma_E^k + \frac{\left(\frac{\mu^k}{1-\mu^k}\right)^2\|\gamma_E^k\|^2}{2\left(1 - \frac{\mu^k}{1-\mu^k}\sigma(\gamma_E^k) + \mathcal{O}(\|X_E^k\|)\right)}\\ &\quad + \mathcal{O}(\|X_E^k\|) \qquad \text{(Use Lemma 1.4.4)}\\ &= \frac{\mu^k}{1-\mu^k}\left\{e^t\gamma_E^k + \frac{\mu^k\|\gamma_E^k\|^2}{2(1-\mu^k\sigma(\gamma_E^k+e) + \mathcal{O}(\|X_E^k\|))}\right\}\\ &\quad + \mathcal{O}(\|X_E^k\|).\end{aligned}$$

Since $\|\gamma_E^k\|^2$ is bounded and

$$(6.25)\qquad \begin{aligned}1-\mu^k\sigma(\gamma_E^k+e) &= 1-\mu^k\sigma(\theta\hat{\alpha}_E^k/\|\hat{\alpha}_E^k\| + \hat{\beta}_E^k)\\ &= 1-\mu^k\sigma(\theta\alpha^k/\|\alpha^k\| + \beta^k) + \mathcal{O}(\|X_E^k\|)\\ &= 1-\lambda^k + \mathcal{O}(\|X_E^k\|) \ge 1/3\end{aligned}$$

holds for sufficiently large k, we have

$$(6.26)\qquad \Delta f_E(x^k) \le \frac{\mu^k}{1-\mu^k}\left\{e^t\gamma_E^k + \frac{\mu^k\|\gamma_E^k\|^2}{2(1-\lambda^k)}\right\} + \mathcal{O}(\|X_E^k\|).$$

We define

$$(6.27)\qquad \zeta(\lambda^k) \triangleq \frac{\lambda^k}{(1-\lambda^k)\sigma((X^k)^{-1}\Delta x^k)} = \frac{\mu^k}{(1-\lambda^k)}.$$

Then obviously, $\zeta(\lambda^k) \ge 0$. Furthermore,

$$(6.28)\qquad \begin{aligned}\zeta(\lambda^k) \le 1 &\Leftrightarrow \lambda^k \le (1-\lambda^k)\sigma((X^k)^{-1}\Delta x^k)\\ &\Leftrightarrow (\sigma((X^k)^{-1}\Delta x^k)+1)\lambda^k \le \sigma((X^k)^{-1}\Delta x^k)\\ &\Leftrightarrow \lambda^k \le \frac{\sigma((X^k)^{-1}\Delta x^k)}{1+\sigma((X^k)^{-1}\Delta x^k)}\end{aligned}$$

hence, we have

$$\zeta(\lambda^k) \leq 1 \tag{6.29}$$

due to *Step Size Choice 1*. (See (2.8).)

Substituting (6.20) and (6.21) into (6.26), we have

(6.30)

$$\begin{aligned}
\Delta f_E(x^k) &\leq \frac{\mu^k}{1-\mu^k}\Bigg\{\frac{\theta e^t \hat{\alpha}_E^k}{\|\hat{\alpha}_E^k\|} - \|e - \hat{\beta}_E^k\|^2 \\
&\quad + \frac{\zeta(\lambda^k)}{2}\left(\theta^2 + \|e - \hat{\beta}_E^k\|^2 - \frac{2\theta e^t \hat{\alpha}_E^k}{\|\hat{\alpha}_E^k\|}\right)\Bigg\} + \mathcal{O}(\|X_E^k\|) \\
&= \frac{\mu^k}{1-\mu^k}\Bigg\{(1-\zeta(\lambda^k))\left(\theta\frac{e^t \hat{\alpha}_E^k}{\|\hat{\alpha}_E^k\|} - \|e - \hat{\beta}_E^k\|^2\right) \\
&\quad + \frac{\zeta(\lambda^k)}{2}(\theta^2 - \|e - \hat{\beta}_E^k\|^2)\Bigg\} + \mathcal{O}(\|X_E^k\|) \\
&\leq \frac{\mu^k}{1-\mu^k}\Bigg\{\frac{1-\zeta(\lambda^k)}{2}(\theta^2 - \|e - \hat{\beta}_E^k\|^2) \\
&\quad + \frac{\zeta(\lambda^k)}{2}(\theta^2 - \|e - \hat{\beta}_E^k\|^2)\Bigg\} + \mathcal{O}(\|X_E^k\|) \quad \text{(Use (6.22).)} \\
&\leq \frac{\mu^k}{2(1-\mu^k)}(\theta^2 - \|e - \hat{\beta}_E^k\|^2) + \mathcal{O}(\|X_E^k\|).
\end{aligned}$$

Recalling that $\|e - \hat{\beta}_E^k\| \geq 1$, $\theta < 1$, $\mu^k < 2/3$ due to (5.1) and that μ^k is bounded below by a positive constant (Corollary 5.3), we see from (6.30) that the difference is smaller than a negative constant for sufficiently large k. Now the lemma follows. □

6.3. Proof of Lemma 6.2

From Proposition 6.4, we have

$$\begin{aligned}
\hat{c}_E^t x_E^k - w^k \hat{M} = c^t x^k - c^t x^\infty &= \sum_{l=k}^{\infty}(c^t x^l - c^t x^{l+1}) \\
&= \sum_{l=k}^{\infty} \mu^l \theta \|(I - P_{AX^l})X^l c\| \\
&\quad + \sum_{l=k}^{\infty} \mu^l w^l c^t X^l P_{AX^l}(X^l)^{-1}(x^0 - \hat{x}).
\end{aligned} \tag{6.31}$$

Recall that

$$\sum_{l=k}^{\infty} \mu^k w^k = \sum_{l=k}^{\infty}(w^{k+1} - w^k) = w^k. \tag{6.32}$$

We have from (6.31),

$$\sum_{l=k}^{\infty}\mu^l\theta\|(I-P_{AX^l})X^lc\| \le \hat{c}_E^t x_E^k - w^k\hat{M} + \sum_{l=k}^{\infty}\mu^l w^l M_1 \tag{6.33}$$
$$= \hat{c}_E^t x_E^k + M_1' w^k$$

where M_1 is an appropriate constant whose existence is guaranteed by Lemma 1.4.1, and $M_1' = M_1 + |\hat{M}|$. Then we have the following inequality.

$$\begin{aligned}\|x^k - x^\infty\| &\le \sum_{l=k}^{\infty}\|x^{l+1}-x^l\| \\ &\le \sum_{l=k}^{\infty}\mu^l\theta\frac{\|X^l(I-P_{AX^l})X^lc\|}{\|(I-P_{AX^l})X^lc\|} \\ &\quad + \sum_{l=k}^{\infty}\mu^l w^l\|X^lP_{AX^l}(X^l)^{-1}(x^0-\hat{x})\| \\ &\le M_2\sum_{l=k}^{\infty}\mu^l\theta\|(I-P_{AX^l})X^lc\| + \sum_{l=k}^{\infty}\mu^l w^l M_3 \\ &\qquad \text{(Use Lemma 1.4.1 and Lemma 1.4.4)} \\ &\le M_2(\hat{c}_E^t x_E^k + w^k M_1') + w^k M_3 \\ &= M_2\hat{c}_E^t x_E^k + (M_1'M_2 + M_3)w^k \\ &= M_2(\hat{c}_E^t x_E^k - w^k\hat{M}) + (M_1'M_2 + M_3 + M_2\hat{M})w^k \\ &\le M_2(c^t x^k - c^t x^\infty) + (M_1'M_2 + M_3 + M_2|\hat{M}|)w^k.\end{aligned}\tag{6.34}$$

By dividing both sides by $\|x^k - x^\infty\|$, we see from Lemma 6.1 that (6.34) implies

$$\frac{c^t x^k - c^t x^\infty}{\|x^k - x^\infty\|} \ge \delta > 0 \tag{6.35}$$

for sufficiently large k. Now the lemma readily follows. □

7. Asymptotic Property of the Sequence

Before proving further convergence results, we make a few observations in this section. Note that we deal with the case where x^∞ is a feasible solution in a dual degenerate face and x^k is an infinite sequence of infeasible points. (See the remark in the beginning of Section 5.)

We define for a nonempty index set $F \subseteq E$,

$$\Phi_F(x^k) \triangleq \|(X_{F^c}^k)^{-1}\|\|X_F^k\|, \tag{7.1}$$

$$\Psi_F(x^k) \triangleq \frac{w^k}{\|x_F^k\|}, \tag{7.2}$$

$$\Lambda_F(x^k) \triangleq \|(X_{F^c}^k)^{-1}\|w^k. \tag{7.3}$$

We sometimes use Φ_F^k, Ψ_F^k, and Λ_F^k as abbreviation for $\Phi_F(x^k), \Psi_F(x^k)$, and $\Lambda_F(x^k)$, respectively. We remark that $\Phi_F(x^k)$ already appeared in Lemma 1.4.5, and $\Psi_E(x^k)$ in Lemma 6.1.

The aim of this section is to prove the following lemma.

LEMMA 7.1. *One and only one of the following two cases occurs on the sequence* $\{x^k\}$*:*

(1) *there exists a nonempty index set* $F \subset E$ *such that*

$$(\Phi_F^k, \Lambda_F^k) \to 0 \quad \text{and} \quad M \geq \Psi_F^k \geq \delta > 0 \tag{7.4}$$

where M *and* δ *are positive constants which do not depend on* k *and* Λ_F^k *converges asymptotically linearly; or*

(2) *there exists a nonempty index set* $F \subset E$ *such that* (Φ_F^k, Λ_F^k) *has* 0 *as an accumulation point and* Ψ_F^k *is not bounded above.*

We say that an index set J *determines* a face of $\mathrm{Feas}\langle P\rangle$ if there exists an $x \in \mathrm{Feas}\langle P\rangle$ such that

$$x_J = 0 \quad \text{and} \quad x_{J^c} > 0. \tag{7.5}$$

We denote by $\mathrm{Face}(J)$ a face of $\mathrm{Feas}\langle P\rangle$ determined by index set J.

The meaning of Lemma 7.1 may become clear in view of the following lemma which is proved in Appendix.

LEMMA 7.2. *If* $(\Phi_F(x^k), \Lambda_F(x^k))$ *has* 0 *as an accumulation point, then* F *determines a face of* $\mathrm{Feas}\langle P\rangle$.

Due to the above lemma, we see that the index set F identified in the cases (1) and (2) of Lemma 7.1 determines a face. Intuitively speaking, in both cases of Lemma 7.1, there exists a face $\mathrm{Face}(F)$ larger than $\mathrm{Face}(E)$ and that the sequence approaches the limit point in $\mathrm{Face}(E)$ tangentially to $\mathrm{Face}(F)$.

To prove Lemma 7.1, we use the following two lemmas:

LEMMA 7.3. *There exists an index set* $F \subset E$ *such that* (Φ_F^k, Λ_F^k) *has* 0 *as an accumulation point and* $(\Phi_F^k, \Lambda_F^k, \Psi_F^k)$ *does not have* 0 *as an accumulation point.*

LEMMA 7.4. *Assume that there exists an index set* F *which satisfies*

(1) $F \subset E$.
(2) Ψ_F^k *is bounded above.*
(3) Φ_F^k *has* 0 *as an accumulation point.*
(4) (Φ_F^k, Ψ_F^k) *does not have* 0 *as an accumulation point.*

Then, Φ_F^k *converges to* 0 *and* Λ_F^k *converges asymptotically linearly to* 0.

Based on these two lemmas, Lemma 7.1 is shown as follows.

PROOF OF LEMMA 7.1. Due to Lemma 7.3, there exists a nonempty index set F such that (Φ_F^k, Λ_F^k) has 0 as an accumulation point. Suppose that the case (2) of Lemma 7.1 does not occur, that is, Ψ_F^k is bounded above. Then it is easy to check that F satisfies the condition (1), (2), and (3) of Lemma 7.4.

Suppose that $(\Phi_F^{k_l}, \Psi_F^{k_l}) \to 0$ holds for a subsequence. Since Ψ_F^k is bounded above, we see that $\Lambda_F^k \leq |F|\Phi_F^k\Psi_F^k \leq M\Phi_F^k$ holds, thus if $\Phi_F^{k_l} \to 0$, then $\Lambda_F^{k_l} \to 0$. Therefore, we have $(\Phi_F^{k_l}, \Lambda_F^{k_l}, \Psi_F^{k_l}) \to 0$ which contradicts Lemma 7.3. Hence, 0 cannot be an accumulation point of $(\Phi_{F'}^k, \Psi_{F'}^k)$, and condition (4) of Lemma 7.4 holds. Now applying Lemma 7.4, we see that the case (1) of Lemma 7.1 holds. □

Lemma 7.4 is proved in Appendix. We prove Lemma 7.3 in the remaining part of this section. To begin with, we prove the following lemma.

LEMMA 7.5. *$w^k(X^k)^{-1}e$ does not have 0 as an accumulation point.*

PROOF. We have

(7.6)
$$\begin{aligned}\sigma((X^k)^{-1}\Delta x^k) &= \sigma(\frac{\theta\alpha^k}{\|\alpha^k\|} + w^k(X^k)^{-1}X^k P_{AX^k}(X^k)^{-1}(x^0 - x^\infty)) \qquad \text{(Use (2.22))}\\ &\leq \theta + Mw^k\sigma((X^k)^{-1}e) \qquad \text{(Use Lemma 1.4.1).}\end{aligned}$$

Suppose contrary that 0 is an accumulation point of $w^k(X^k)^{-1}e$. Then, $\sigma((X^k)^{-1}\Delta x^k) < 1$ occurs for some k, which contradicts (5.1). □

Since $w^k(X_{E^c}^k)^{-1}e$ converges to 0, the above lemma implies that $w^k(X_E^k)^{-1}e$ does not have 0 as an accumulation point.

Next, we prove the following lemma.

LEMMA 7.6. *Assume that $(\Phi_{E'}^k, \Lambda_{E'}^k, \Psi_{E'}^k)$ has 0 as an accumulation point for an index set E'. Then, there exists a nonempty index set $F' \subset E'$ such that $(\Phi_{F'}^k, \Lambda_{F'}^k)$ has 0 as an accumulation point.*

PROOF. Let $(\Phi_{E'}^{k_t}, \Lambda_{E'}^{k_t}, \Psi_{E'}^{k_t})$ be a subsequence convergent to 0. Let

$$\tilde{u}_{E'}^k \triangleq \frac{x_{E'}^k}{w^k}. \tag{7.7}$$

Then $\|\tilde{u}_{E'}^{k_t}\| \to \infty$. We can choose an appropriate subsequence $k_{t'}$ of k_t such that

$$\tilde{u}_{F_1}^{k_{t'}} \to \tilde{u}_{F_1}^* < \infty, \tag{7.8}$$

$$\tilde{u}_{\tilde{F}_1}^{k_{t'}} \to \infty, \tag{7.9}$$

$$F_1 + \tilde{F}_1 = E'. \tag{7.10}$$

Suppose that F_1 is empty. It means that for all $j \in E'$,

$$w^{k_{t'}}/x_j^{k_{t'}} \to 0. \tag{7.11}$$

On the other hand, for $j \notin E'$,

$$w^{k_{t'}}/x_j^{k_{t'}} \leq \Lambda_{E'}^{k_{t'}} \to 0 \tag{7.12}$$

also holds. These facts contradict Lemma 7.5. Thus F_1 cannot be empty.

Since $\tilde{F}_1$ is also not empty, we have $F_1 \subset E'$ and

$$\Phi_{F_1}^{k_{t'}} = \|(X_{F_1^c}^{k_{t'}})^{-1}\|\|X_{F_1}^{k_{t'}}\| \to 0. \tag{7.13}$$

Next, let

$$u_{E'}^k \triangleq \frac{x_{E'}^k}{\|x_{E'}^k\|}. \tag{7.14}$$

Then we can choose an appropriate subsequence $k_{t''}$ of k_t such that

$$u_{F_2}^{k_{t''}} \to 0, \tag{7.15}$$

$$u_{\tilde{F}_2}^{k_{t''}} \to u_{\tilde{F}_2}^* > 0, \tag{7.16}$$

$$F_2 + \tilde{F}_2 = E'. \tag{7.17}$$

Due to the existence of F_1, we can assume that F_2 is not empty. Obviously, $\tilde{F}_2$ is not empty too, thus $F_2 \subset E'$. There exists a positive constant δ such that

$$u_j^{k_{t''}} \geq \delta > 0 \tag{7.18}$$

holds for all $j \in \tilde{F}_2$ and all $k_{t''}$. Hence we have

$$\Lambda_{F_2}^{k_{t''}} = \frac{w^{k_{t''}}}{\|x_{E'}^{k_{t''}}\|} \frac{\|x_{E'}^{k_{t''}}\|}{\min_{j \notin F_2} x_j^{k_{t''}}} \leq \Psi_{E'}^{k_{t''}}/\delta = \mathcal{O}(\Psi_{E'}^{k_{t''}}) \to 0. \tag{7.19}$$

We also have

$$\begin{aligned} \Phi_{F_2}^{k_{t''}} &= \|(X_{F_2{}^c}^{k_{t''}})^{-1}\| \|X_{F_2}^{k_{t''}}\| \\ &= \max(\|(X_{\tilde{F}_2}^{k_{t''}})^{-1}\|, \|(X_{E'^c}^{k_{t''}})^{-1}\|)\|X_{F_2}^{k_{t''}}\| \to 0. \end{aligned} \tag{7.20}$$

Now the index set F_2 is what we sought. □

Finally, we prove Lemma 7.3.

Proof of Lemma 7.3. From Lemma 6.1, we have $\Psi_E^k \to 0$, which implies $(\Phi_E^k, \Lambda_E^k, \Psi_E^k) \to 0$. We can apply Lemma 7.6 and see that there exists an index set $E_1 \subset E$ such that $(\Phi_{E_1}^k, \Lambda_{E_1}^k)$ has 0 as an accumulation point. If $(\Phi_{E_1}^k, \Lambda_{E_1}^k, \Psi_{E_1}^k)$ does not have 0 as an accumulation point, then the lemma follows. Suppose that 0 is an accumulation point of $(\Phi_{E_1}^k, \Lambda_{E_1}^k, \Psi_{E_1}^k)$. Then we can apply Lemma 7.6 again with $E' = E_1$. We can apply the above procedure recursively and produce a sequence of index sets $\{ E_j \mid j = 1, 2, \ldots \}$, where

$$E \supset E_1 \supset E_2 \supset \cdots . \tag{7.21}$$

Since E is a finite set, this procedure should end up with some subset E_N. Since E_N is produced by Lemma 7.6, E_N is nonempty and $(\Phi_{E_N}^k, \Lambda_{E_N}^k)$ has 0 as an accumulation point. Furthermore, since E_N does not satisfy the assumption of Lemma 7.6, $(\Phi_{E_N}^k, \Lambda_{E_N}^k, \Psi_{E_N}^k)$ does not have 0 as an accumulation point. Now E_N is what we want. □

8. Case (1) of Lemma 7.1

In this section, we investigate the behavior of the primal sequence and dual estimates in the case (1) of Lemma 7.1 extensively. Therefore, we assume that

$$(\Phi_F^k, \Lambda_F^k) \to 0 \qquad \text{and} \qquad M \geq \Psi_F^k \geq \delta > 0 \tag{8.1}$$

and that Λ_F^k converges to 0 asymptotically linearly throughout the section. From the boundedness of Ψ_F^k, we see that

$$\Lambda_F^k \leq |F| \Psi_F^k \Phi_F^k = \mathcal{O}(\Phi_F^k). \tag{8.2}$$

We let $\tilde{F} \triangleq E - F$. Since

$$w^k = \Lambda_F^k \min_{j \notin F} x_j^k \leq \Lambda_F^k \|X_{\tilde{F}}^k\| \tag{8.3}$$

holds for sufficiently large k, we have

$$w^k = \mathcal{O}(\Phi_F^k \|X_{\tilde{F}}^k\|). \tag{8.4}$$

The case (2) will be discussed in the next section. We will show that $\langle P \rangle$ has no interior feasible solution in the case (1) of Lemma 7.1. In other words, we will prove:

THEOREM 8.1. *If the case (1) of Lemma 7.1 occurs, then F determines the optimal face of $\langle F \rangle$.*

This section consists of six subsections. First three subsections are devoted to proving the following lemma:

LEMMA 8.2. *Assume that the face determined by F is not the optimal face of $\langle F \rangle$. Then, $\Psi_F^k \to 0$.*

Once this lemma is shown, Theorem 8.1 readily follows by taking account of the second relation of (8.1). We will prove Lemma 8.2 by using the local Karmarkar potential function developed in Chapter II. To evaluate the local Karmarkar potential function, we prepare a few lemmas on behaviors of β_E^k and $\hat{\alpha}_E^k$ in Subsections 8.1 and 8.2, respectively. The lemmas and corollaries derived in these subsections are used in the proof of Lemma 8.2 as well as in the consecutive subsections. In Subsection 8.4, we prove global convergence of the primal sequence when the case (1) of Lemma 7.1 occurs by using the local Karmarkar potential function. Subsections 8.5 and 8.6 are devoted to proving global convergence of dual estimates for optimality and feasibility, respectively.

8.1. Behavior of β_E^k

The first lemma we introduce here is:

LEMMA 8.3. $\|\beta_{\tilde{F}}^k\| = \mathcal{O}(\Phi_F^k)$.

PROOF. In view of Lemma 2.1 and Lemma 1.4.1, it is easy to see that $\|\beta_{\tilde{F}}^k\| = \mathcal{O}(\Lambda_F^k) = \mathcal{O}(\Phi_F^k)$. □

We denote by $\bar{\beta}_F^k$ the optimal solution of

$$\langle \bar{B}_F \rangle \left\{ \begin{array}{ll} \text{minimize} & \|\beta_F - e\|^2 \\ \text{subject to} & \beta_F = X_F^k A_F^t y, \quad A_{F^c}^t y = 0. \end{array} \right. \tag{8.5}$$

Then we have the following lemma.

LEMMA 8.4.

$$\|\bar{\beta}_F^k - \beta_F^k\| = \mathcal{O}(\Phi_F^k). \tag{8.6}$$

This lemma means that $\bar{\beta}_F^k$ approximates β_F^k when $\Phi_F^k \to 0$. The proof is almost the same as that of Lemma 4.1 if we recall that $w^\infty = 0$, hence we omit the proof.

Next we show the following property of $\bar{\beta}_F^k$.

LEMMA 8.5. *If $\bar{\beta}_F^k > 0$ happens for some k, then* Face(F) *is the optimal face of $\langle F \rangle$.*

PROOF. From (8.1) and Lemma 7.2, F determines a face of $\langle P\rangle$, and there exists a point $\bar{x} \in R^n$ such that

$$A_{F^c}\bar{x}_{F^c} = b, \quad \bar{x}_{F^c} > 0, \quad \bar{x}_F = 0. \tag{8.7}$$

Let $(\bar{\beta}_F^k, \tilde{y}^k)$ be the optimal solution of $\langle \bar{B}_F\rangle$. Putting

$$\bar{y} \triangleq \frac{w^k \tilde{y}^k}{\|\bar{\beta}_F^k\|^2}, \tag{8.8}$$

we have

$$\begin{aligned}(r^0)^t\bar{y} &= \frac{w^k(r^0)^t\tilde{y}^k}{\|\bar{\beta}_F^k\|^2} = \frac{(x^k - x^\infty)^t A^t \tilde{y}^k}{\|\bar{\beta}_F^k\|^2} \\ &= \frac{(x_F^k - x_F^\infty)^t A_F^t \tilde{y}^k}{\|\bar{\beta}_F^k\|^2} = \frac{(x_F^k - x_F^\infty)^t (X_F^k)^{-1} X_F^k A_F^t \tilde{y}^k}{\|\bar{\beta}_F^k\|^2} \\ &= \frac{e^t\bar{\beta}_F^k}{\|\bar{\beta}_F^k\|^2} = \frac{\|\bar{\beta}_F^k\|^2}{\|\bar{\beta}_F^k\|^2} = 1.\end{aligned} \tag{8.9}$$

Hence $\bar{y}$ satisfies the equality constraints of $\langle DF\rangle$.

Since $A_F^t\bar{y} > 0$ follows from $\bar{\beta}_F^k > 0$, we see that $\bar{y}$ and $(\bar{x}, 0)$ satisfy strict complementarity condition for $\langle F\rangle$ and $\langle DF\rangle$. □

8.2. Behavior of $\hat{\alpha}_E^k$

We recall that $\hat{\alpha}_E^k$ is the optimal solution of

$$\langle A_E\rangle \begin{cases} \text{minimize} & \|\alpha_E - X_E^k \hat{c}_E\|^2 \\ \text{subject to} & \tilde{A}_E X_E^k \alpha_E = 0, \end{cases} \tag{8.10}$$

where $\tilde{A}_E$ is defined by (4.10). Let $\hat{\alpha}_F^k$ and $\hat{\hat{\alpha}}_{\tilde{F}}^k$ be the optimal solutions of

$$\langle \tilde{A}_F\rangle \begin{cases} \text{minimize} & \|\alpha_F - X_F^k \hat{c}_F\|^2 \\ \text{subject to} & \tilde{A}_F X_F^k \alpha_F \in \mathrm{Im}(\tilde{A}_{\tilde{F}}), \end{cases} \tag{8.11}$$

and

$$\langle \tilde{A}_{\tilde{F}}\rangle \begin{cases} \text{minimize} & \|\alpha_{\tilde{F}} - X_{\tilde{F}}^k \hat{c}_{\tilde{F}}\|^2 \\ \text{subject to} & \tilde{A}_{\tilde{F}} X_{\tilde{F}}^k \alpha_{\tilde{F}} = 0, \end{cases} \tag{8.12}$$

respectively. Then we see from Lemma 1.4.5 that there exist some constants $\tilde{M}_1$, $\tilde{M}_2$, $\tilde{M}_3$, and $\tilde{M}_4$ such that

$$\|\hat{\alpha}_F^k - \hat{\hat{\alpha}}_F^k\| \le \tilde{M}_1(\Phi_F(x^k))^2\|\hat{\hat{\alpha}}_F^k\| + \tilde{M}_2\Phi_F(x^k)\|x_{\tilde{F}}^k\|, \tag{8.13}$$

$$\|\hat{\alpha}_{\tilde{F}}^k - \hat{\hat{\alpha}}_{\tilde{F}}^k\| \le \tilde{M}_3\Phi_F(x^k)\|\hat{\hat{\alpha}}_F^k\| + \tilde{M}_4(\Phi_F(x^k))^2\|x_{\tilde{F}}^k\|. \tag{8.14}$$

We will prove the following three lemmas in the remaining part of this subsection.

LEMMA 8.6. *$\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k$ is positive for sufficiently large k, and*

$$\frac{e^t\hat{\hat{\alpha}}_{\tilde{F}}^k}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k} = 1 + \mathcal{O}(\Phi_F^k). \tag{8.15}$$

LEMMA 8.7.

$$\frac{\|\hat{\alpha}^k_F\|}{\|\hat{\alpha}^k_E\|} = \mathcal{O}(\Phi^k_F). \tag{8.16}$$

LEMMA 8.8.

$$\frac{\hat{\alpha}^k_{\tilde{F}}}{\|\hat{\alpha}^k_E\|} = \frac{\hat{\hat{\alpha}}^k_{\tilde{F}}}{\|\hat{\hat{\alpha}}^k_{\tilde{F}}\|} + \mathcal{O}(\Phi^k_F). \tag{8.17}$$

The above lemmas, combined with Lemma 8.3 and Lemma 8.4, suggest that F-component of the scaled direction is almost $\bar{\beta}^k_F$, while $\tilde{F}$-component is close to $\hat{\hat{\alpha}}^k_{\tilde{F}}$.

To prove the above lemmas, we prepare a few lemmas. The following lemma is substantial to the proof of Lemma 8.6.

LEMMA 8.9. *Let $\hat{\hat{\gamma}}^k_{\tilde{F}}$ be the optimal solution of*

$$\langle G_{\tilde{F}}\rangle \begin{cases} \text{minimize} & \|\gamma_{\tilde{F}} - e\|^2 \\ \text{subject to} & \tilde{A}_{\tilde{F}} X^k_{\tilde{F}} \gamma_{\tilde{F}} = 0. \end{cases} \tag{8.18}$$

Then,

$$\|\hat{\hat{\gamma}}^k_{\tilde{F}} - e\| = \mathcal{O}(\Phi_F(x^k)). \tag{8.19}$$

PROOF. Recall the matrix N defined after Lemma 4.1. We have

$$NA_{E^c} = 0 \tag{8.20}$$

by definition. Since

$$\begin{aligned} A_E x^k_E + A_{E^c}(x^k_{E^c} - x^\infty_{E^c}) &= w^k r^0 = w^k(Ax^0 - b) \\ &= w^k(A_E x^0_E + A_{E^c} x^0_{E^c} - A_{E^c} x^\infty_{E^c}), \end{aligned} \tag{8.21}$$

multiplying both sides by N, we have, due to (8.20),

$$\tilde{A}_E X^k_E(e - w^k (X^k_E)^{-1} x^0_E) = 0. \tag{8.22}$$

Let $\hat{\gamma}^k_E = e - w^k (X^k_E)^{-1} x^0_E$. From (8.22), we have

$$\tilde{A}_{\tilde{F}} X^k_{\tilde{F}} \hat{\gamma}^k_{\tilde{F}} = -\tilde{A}_F X^k_F \hat{\gamma}^k_F = -\tilde{A}_F(x^k_F - w^k x^0_F). \tag{8.23}$$

Then there exists a $\bar{\gamma}_{\tilde{F}}$ which is a solution of the system $\tilde{A}_{\tilde{F}} X^k_{\tilde{F}} \gamma_{\tilde{F}} = 0$ such that

$$\|\hat{\gamma}^k_{\tilde{F}} - \bar{\gamma}_{\tilde{F}}\| \le M_1 \|(X^k_{\tilde{F}})^{-1}\| \|x^k_F - w^k x^0_F\| = \mathcal{O}(\Phi_F(x^k) + \Lambda_F(x^k)). \tag{8.24}$$

Noting that $\Lambda^k_F = \mathcal{O}(\Phi^k_F)$ since Ψ^k_F is bounded, we have

$$\|e - w^k (X^k_{\tilde{F}})^{-1} x^0_{\tilde{F}} - \bar{\gamma}^k_{\tilde{F}}\| = \mathcal{O}(\Phi_F(x^k)), \tag{8.25}$$

hence

$$\|e - \bar{\gamma}^k_{\tilde{F}}\| = \mathcal{O}(\Phi_F(x^k)). \tag{8.26}$$

On the other hand, since $\bar{\gamma}^k_{\tilde{F}}$ is feasible for $\langle G_{\tilde{F}}\rangle$ and $\hat{\hat{\gamma}}^k_{\tilde{F}}$ is optimal for $\langle G_{\tilde{F}}\rangle$, we have

$$\|\hat{\hat{\gamma}}^k_{\tilde{F}} - e\| \le \|\bar{\gamma}^k_{\tilde{F}} - e\| = \mathcal{O}(\Phi_F(x^k)) \tag{8.27}$$

and this completes the proof. $\square$

COROLLARY 8.10.

$$e^t \hat{\alpha}_{\tilde{F}}^k = \hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k + \mathcal{O}(\|X_{\tilde{F}}^k\| \Phi_F^k). \tag{8.28}$$

PROOF. Since $\hat{\alpha}_{\tilde{F}}^k$ and $\hat{\gamma}_{\tilde{F}}^k$ are projections onto the same linear space $\mathrm{Null}(\tilde{A}_{\tilde{F}} X_{\tilde{F}}^k)$ of $X_{\tilde{F}}^k \hat{c}_{\tilde{F}}$ and e, respectively,

$$(\hat{\gamma}_{\tilde{F}}^k)^t \hat{\alpha}_{\tilde{F}}^k = (\hat{\gamma}_{\tilde{F}}^k)^t X_{\tilde{F}}^k \hat{c}_{\tilde{F}} \tag{8.29}$$

holds. Then, in view of the previous lemma, we have

$$\begin{aligned} e^t \hat{\alpha}_{\tilde{F}}^k &= (\hat{\gamma}_{\tilde{F}}^k)^t \hat{\alpha}_{\tilde{F}}^k + \mathcal{O}(\|\hat{\alpha}_{\tilde{F}}^k\| \Phi_F^k) \\ &= (\hat{\gamma}_{\tilde{F}}^k)^t X_{\tilde{F}}^k \hat{c}_{\tilde{F}} + \mathcal{O}(\|\hat{\alpha}_{\tilde{F}}^k\| \Phi_F^k) \\ &= e^t X_{\tilde{F}}^k \hat{c}_{\tilde{F}} + \mathcal{O}(\|X_{\tilde{F}}^k\| \Phi_F^k) \\ &= \hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k + \mathcal{O}(\|X_{\tilde{F}}^k\| \Phi_F^k) \end{aligned} \tag{8.30}$$

and this completes the proof. □

LEMMA 8.11.

$$\frac{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k}{\|x_E^k\|} \geq \delta > 0 \tag{8.31}$$

for sufficiently large k.

PROOF. In view of Lemma 6.1, Lemma 6.2 and Proposition 6.4, we see that

$$\begin{aligned} 0 < \delta' &\leq \frac{\hat{c}_E^t x_E^k}{\|x_E^k\|} \\ &\leq \frac{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k}{\|x_E^k\|} + \frac{\|\hat{c}_F\| \|x_F^k\|}{\|x_E^k\|} \end{aligned} \tag{8.32}$$

holds for sufficiently large k. Now the lemma readily follows from the fact that $\Phi_F(x^k) \to 0$. □

COROLLARY 8.12.

$$\frac{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k}{\|x_{\tilde{F}}^k\|} \geq \delta > 0 \tag{8.33}$$

for sufficiently large k.

PROOF. Since $\|x_E^k\| \geq \|x_{\tilde{F}}^k\|$, the corollary is obvious in view of Lemma 8.11. □

Now we are ready to prove the main lemmas.

PROOF OF LEMMA 8.6. Lemma 8.11 implies that $\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k$ is positive for sufficiently large k. Then, from Corollary 8.10 and Corollary 8.12,

$$\frac{e^t \hat{\alpha}_{\tilde{F}}^k}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k} = \frac{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k + \mathcal{O}(\|X_{\tilde{F}}^k\| \Phi_F^k)}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k} = 1 + \mathcal{O}(\Phi_F^k). \tag{8.34}$$

□

COROLLARY 8.13.

$$\frac{\|\hat{\alpha}_{\tilde{F}}^k\|}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k} \geq \frac{1}{|\tilde{F}|} + \mathcal{O}(\Phi_F^k) \qquad \textit{for sufficiently large } k. \tag{8.35}$$

PROOF. This is obvious from Lemma 8.6. □

COROLLARY 8.14.

$$\frac{\|\hat{\alpha}_{\tilde{F}}^k\|}{\|x_E^k\|} \geq \delta > 0 \qquad \textit{for sufficiently large } k. \tag{8.36}$$

PROOF.

$$\frac{\|\hat{\alpha}_{\tilde{F}}^k\|}{\|x_E^k\|} = \frac{\|\hat{\alpha}_{\tilde{F}}^k\| \hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k \|x_E^k\|} \geq \delta > 0. \tag{8.37}$$

□

PROOF OF LEMMA 8.7. From (8.13), we have

$$\begin{aligned} \|\hat{\alpha}_F^k\| &= \|\hat{\alpha}_{\tilde{F}}^k\| + \mathcal{O}((\Phi_F^k)^2 \|\hat{\alpha}_{\tilde{F}}^k\| + \Phi_F^k \|X_{\tilde{F}}^k\|) \\ &= \mathcal{O}(\|\hat{\alpha}_{\tilde{F}}^k\| + \Phi_F^k \|X_{\tilde{F}}^k\|) \\ &= \mathcal{O}(\|x_{\tilde{F}}^k\| + \Phi_F^k \|X_{\tilde{F}}^k\|). \end{aligned} \tag{8.38}$$

Thus,

$$\begin{aligned} \frac{\|\hat{\alpha}_F^k\|}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k} &= \mathcal{O}\left(\frac{\|x_{\tilde{F}}^k\| + \Phi_F^k \|X_{\tilde{F}}^k\|}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k}\right) \\ &= \mathcal{O}\left(\frac{\|x_{\tilde{F}}^k\| \|X_{\tilde{F}}^k\|}{\|x_{\tilde{F}}^k\| \hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k} + \Phi_F^k\right) \qquad \text{(Use Corollary 8.12)} \\ &= \mathcal{O}(\Phi_F^k) \qquad \text{(Use Corollary 8.12 again)} \end{aligned} \tag{8.39}$$

for sufficiently large k. Then we have from (8.14) that

$$\begin{aligned} \frac{\|\hat{\alpha}_{\tilde{F}}^k\|}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k} &= \frac{\|\hat{\alpha}_{\tilde{F}}^k\| + \mathcal{O}(\Phi_F^k \|\hat{\alpha}_F^k\| + (\Phi_F^k)^2 \|X_{\tilde{F}}^k\|)}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k} \\ &= \frac{\|\hat{\alpha}_{\tilde{F}}^k\|}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k} + \mathcal{O}\left(\frac{\Phi_F^k \|X_{\tilde{F}}^k\| (\|(X_{\tilde{F}}^k)^{-1}\| \|\hat{\alpha}_F^k\| + \Phi_F^k)}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k}\right) \\ &= \frac{\|\hat{\alpha}_{\tilde{F}}^k\|}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k} + \mathcal{O}((\Phi_F^k)^2) \qquad \text{(Use Corollary 8.12)} \\ &\geq \delta' > 0 \qquad \text{(Use Corollary 8.13)}. \end{aligned} \tag{8.40}$$

Hence, from (8.39) and (8.40),

$$\frac{\|\hat{\alpha}_F^k\|}{\|\hat{\alpha}_E^k\|} \leq \frac{\|\hat{\alpha}_F^k\|}{\|\hat{\alpha}_{\tilde{F}}^k\|} = \frac{\|\hat{\alpha}_F^k\| \hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k \|\hat{\alpha}_{\tilde{F}}^k\|} = \mathcal{O}(\Phi_F^k). \tag{8.41}$$

□

PROOF OF LEMMA 8.8. Lemma 8.7 implies that

$$\|\hat{\alpha}_F^k\|^2 = \|\hat{\alpha}_E^k\|^2 + \mathcal{O}((\Phi_F^k)^2\|\hat{\alpha}_E^k\|^2). \tag{8.42}$$

Therefore,

$$\|\hat{\alpha}_F^k\| = \|\hat{\alpha}_E^k\|(1 + \mathcal{O}((\Phi_F^k)^2)) \tag{8.43}$$

holds. On the other hand, (8.14) implies that

$$\begin{aligned} \big|\|\hat{\alpha}_F^k\| - \|\hat{\bar{\alpha}}_F^k\|\big| &\le \|\hat{\alpha}_F^k - \hat{\bar{\alpha}}_F^k\| \\ &= \mathcal{O}(\Phi_F^k\|\hat{\bar{\alpha}}_F^k\| + (\Phi_F^k)^2\|x_F^k\|). \end{aligned} \tag{8.44}$$

From (8.43) and (8.44), we have

$$\big|\|\hat{\bar{\alpha}}_F^k\| - \|\hat{\alpha}_E^k\|\big| = \mathcal{O}(\Phi_F^k\|x_E^k\|). \tag{8.45}$$

Then by using (8.14) we have

$$\begin{aligned} \frac{\hat{\alpha}_F^k}{\|\hat{\alpha}_E^k\|} &= \frac{\hat{\bar{\alpha}}_F^k + \mathcal{O}(\Phi_F^k\|x_E^k\|)}{\|\hat{\bar{\alpha}}_F^k\| + \mathcal{O}(\Phi_F^k\|x_E^k\|)} \\ &= \frac{\hat{\bar{\alpha}}_F^k + \mathcal{O}(\Phi_F^k\|x_E^k\|)}{\|\hat{\bar{\alpha}}_F^k\|\left(1 + \mathcal{O}(\Phi_F^k\frac{\|x_E^k\|}{\|\hat{\bar{\alpha}}_F^k\|})\right)} \\ &= \frac{\hat{\bar{\alpha}}_F^k + \mathcal{O}(\Phi_F^k\|x_E^k\|)}{\|\hat{\bar{\alpha}}_F^k\|(1 + \mathcal{O}(\Phi_F^k))} \qquad \text{(Use Corollary 8.14)} \\ &= \frac{\hat{\bar{\alpha}}_F^k}{\|\hat{\bar{\alpha}}_F^k\|} + \mathcal{O}(\Phi_F^k), \qquad \text{(Use Corollary 8.14 again)} \end{aligned} \tag{8.46}$$

and this completes the proof. □

8.3. Proof of Lemma 8.2

From Lemma 6.7, Lemma 8.7 and Lemma 8.4, we see that the scaled iteration associated with F can be written as

$$(X_F^k)^{-1}x_F^{k+1} = e - \mu^k\bar{\beta}_F^k + \mathcal{O}(\Phi_F^k + \|X_E^k\|^2) \tag{8.47}$$

where $\bar{\beta}_F^k$ is the optimal solution of $\langle\bar{B}_F\rangle$ in (8.5).

We consider reduction of the local Karmarkar potential function for feasibility:

$$f_F(x^k) = |F|\log w^k - \sum_{j\in F}\log x_j^k. \tag{8.48}$$

Let $\tilde{\phi}_F^k \triangleq \Phi_F^k + \|X_E^k\|^2$. Then we have the following lemma which is an substantial part of the proof of Lemma 8.2.

LEMMA 8.15.

(8.49)
$$\begin{aligned}\Delta f_F(x^k) &\triangleq f_F(x^{k+1}) - f_F(x^k)\\ &\le \left(\frac{\mu^k}{1-\mu^k}\right)\|e-\bar\beta_F^k\|^2\left\{-1+\frac{\mu^k}{2(1-\mu^k\sigma(\bar\beta_F^k))}\right\} + \mathcal{O}(\tilde\phi_F^k)\\ &\le -\epsilon\|e-\bar\beta_F^k\|^2 + \mathcal{O}(\tilde\phi_F^k)\end{aligned}$$

for sufficiently large k and for a positive constant ϵ.

PROOF. We have

(8.50)
$$\begin{aligned}\Delta f_F(x^k) &= |F|\log\frac{w^{k+1}}{w^k} - \sum_{j\in F}\log\frac{x_j^{k+1}}{x_j^k}\\ &= |F|\log(1-\mu^k) - \sum_{j\in F}\log(1-\mu^k)(1+\frac{\mu^k}{1-\mu^k}(1-\bar\beta_j^k)+\mathcal{O}(\tilde\phi_F^k))\\ &= -\sum_{j\in F}\log(1+\frac{\mu^k}{1-\mu^k}(1-\bar\beta_j^k)+\mathcal{O}(\tilde\phi_F^k))\\ &\le -\left(\frac{\mu^k}{1-\mu^k}\right)\|e-\bar\beta_F^k\|^2 + \frac{\left(\frac{\mu^k}{1-\mu^k}\right)^2\|e-\bar\beta_F^k\|^2}{2\left(1-\frac{\mu^k}{1-\mu^k}\sigma(-(e-\bar\beta_F^k))+\mathcal{O}(\tilde\phi_F^k)\right)}\\ &\qquad + \mathcal{O}(\tilde\phi_F^k) \qquad \text{(Use Proposition 2.4.2)}\\ &= \left(\frac{\mu^k}{1-\mu^k}\right)\|e-\bar\beta_F^k\|^2\left\{-1+\frac{\mu^k}{2(1-\mu^k\sigma(\bar\beta_F^k)+\mathcal{O}(\tilde\phi_F^k))}\right\} + \mathcal{O}(\tilde\phi_F^k).\end{aligned}$$

Furthermore, we have

(8.51)
$$\begin{aligned}\mu^k\sigma(\bar\beta_F^k) &= \mu^k\sigma(\beta_F^k) + \mathcal{O}(\Phi_F^k) \qquad \text{(Use Lemma 8.4)}\\ &= \mu^k\sigma(\theta\alpha_F^k/\|\alpha^k\| + \beta_F^k) + \mathcal{O}(\Phi_F^k + \|X_E^k\|^2)\\ &\qquad \text{(Use Lemma 6.7 and Lemma 8.7)}\\ &\le \mu^k\sigma(\theta\alpha^k/\|\alpha^k\| + \beta^k) + \mathcal{O}(\Phi_F^k + \|X_E^k\|^2)\\ &= \lambda^k + \mathcal{O}(\Phi_F^k + \|X_E^k\|^2)\\ &\le 2/3 - \hat\delta + \mathcal{O}(\Phi_F^k + \|X_E^k\|^2).\end{aligned}$$

Therefore,

(8.52)
$$2/3 - \hat\delta/2 \ge \mu^k\sigma(\bar\beta_F^k)$$

holds for sufficiently large k. Also, we note that $\mu^k = \lambda^k/\sigma((X^k)^{-1}\Delta x) \le 2/3 - \hat\delta$. From these observation, we see

$$\frac{\mu^k}{2(1-\mu^k\sigma(\bar\beta^k_F)+\mathcal{O}(\tilde\phi^k_F))}=\frac{\mu^k}{2(1-\mu^k\sigma(\bar\beta^k_F))}+\mathcal{O}(\tilde\phi^k_F). \tag{8.53}$$

Now substituting (8.53) for (8.50) and noting that $\bar\beta^k_F$ is bounded by definition, we have the first inequality of (8.49).

The inequality (8.52) implies

$$-1+\frac{\mu^k}{2(1-\mu^k\sigma(\bar\beta^k_F))}\le -1+\frac{2/3-\hat\delta}{2/3+\hat\delta}\le-\delta'<0. \tag{8.54}$$

Hence from Corollary 5.3 and the fact that $\mu^k<2/3$, the lemma readily follows. □

Now we prove Lemma 8.2.

PROOF OF LEMMA 8.2. By assumption, $\bar\beta^k_F\not>0$ from Lemma 8.5. Hence we have $\|e-\bar\beta^k_F\|\ge 1$. From Lemma 8.15, $\Delta f_F(x^k)\le-\epsilon'<0$ for sufficiently large k. This implies $f_F(x^k)\to-\infty$, which completes the proof. □

8.4. Global Convergence of the Primal Sequence

In this subsection, we prove global convergence of the primal iterates.

THEOREM 8.16. *If the case (1) of Lemma 7.1 occurs, then x^∞ is a relative interior point of the optimal face.*

First, as we already know that Face(F) is the optimal face of $\langle F\rangle$, we consider the following LP problem by removing F-components of $\langle P\rangle$:

$$\langle P_{F^c}\rangle\begin{cases}\text{minimize} & c^t_{F^c}x_{F^c}\\ \text{subject to} & A_{F^c}x_{F^c}=b,\quad x_{F^c}\ge 0.\end{cases} \tag{8.55}$$

Note that $\langle P_{F^c}\rangle$ has an interior feasible solution and that the optimal face of $\langle P_{F^c}\rangle$ is derived by ignoring F components of the optimal face of $\langle P\rangle$ which are all zeros. Now we prove the following lemma:

LEMMA 8.17. *If $\hat\alpha^k_{\tilde F}>0$ for some k, then* Face($\tilde F$) *is the optimal face of $\langle P\rangle$.*

PROOF. By setting up Lagrangian of $\langle\tilde A_{\tilde F}\rangle$ in (8.12), we see that

$$\hat\alpha^k_{\tilde F}-X^k_{\tilde F}\hat c_{\tilde F}-X^k_{\tilde F}\tilde A^t_{\tilde F}\tilde y=0 \tag{8.56}$$

for an appropriate $\tilde y$. Then letting $\hat s^k_{\tilde F}\triangleq(X^k_{\tilde F})^{-1}\hat\alpha^k_{\tilde F}$, we have

$$\hat s^k_{\tilde F}=\hat c_{\tilde F}-\tilde A^t_{\tilde F}\tilde y=\hat c_{\tilde F}-A^t_{\tilde F}N\tilde y \tag{8.57}$$

where N is defined in Subsection 8.1. Hence,

$$\begin{pmatrix}\hat s^k_{\tilde F}\\ 0\end{pmatrix}=\begin{pmatrix}\hat c_{\tilde F}\\ 0\end{pmatrix}-A^t_{F^c}Ny=c_{F^c}-A^t_{F^c}y' \tag{8.58}$$

holds for an appropriate y'. If $\hat\alpha^k_{\tilde F}>0$, then $\hat s^k_{\tilde F}>0$ follows thus $(\hat s^k_{\tilde F},0)$ and $(0,x^\infty_{E^c})$ satisfy strict complementarity condition for $\langle P_{F^c}\rangle$ and its dual $\langle DP_{F^c}\rangle$. Now by recalling the remark following (8.55), the lemma readily follows. □

Next we define the local Karmarkar potential function for optimality:

$$g_{\tilde{F}}(x^k) \triangleq |\tilde{F}| \log(c^t x^k - c^t x^\infty) - \sum_{j \in \tilde{F}} \log x_j^k. \tag{8.59}$$

We point out that $c^t x^k$ may be less than $c^t x^\infty$ in general. In this case, $g_{\tilde{F}}(x^k)$ cannot be defined. Lemma 6.2, however, guarantees that there exists a number K such that if $k \geq K$, then $c^t x^k > c^t x^\infty$ and $g_{\tilde{F}}(x^k)$ is well-defined. Since our main concern is convergence property of the sequence, we are not interested in the case where k is so small that $g_{\tilde{F}}(x^k)$ is not well-defined. It is sufficient to deal with the case where $k \geq K$, and we assume it in the consecutive analysis until the end of this section.

To prove Theorem 8.16, we prove the following lemma.

LEMMA 8.18. *If* $\mathrm{Face}(\tilde{F})$ *is not the optimal face of* $\langle P \rangle$, *then* $g_{\tilde{F}}(x^k) \to -\infty$.

If we can prove Lemma 8.18, then the theorem is proved as follows.

PROOF OF THEOREM 8.16. Lemma 6.2 implies that there exists a positive constant δ such that

$$\frac{c^t x^k - c^t x^\infty}{\|x_{\tilde{F}}^k\|} \geq \frac{c^t x^k - c^t x^\infty}{\|x_E^k\|} \geq \delta > 0. \tag{8.60}$$

Since

$$\exp(g_{\tilde{F}}(x^k)) = \frac{(c^t x^k - c^t x^\infty)^{|\tilde{F}|}}{\prod_{j \in \tilde{F}} x_j^k} \geq \left(\sqrt{|\tilde{F}|} \frac{(c^t x^k - c^t x^\infty)}{\|x_{\tilde{F}}^k\|} \right)^{|\tilde{F}|}, \tag{8.61}$$

we see that $g_{\tilde{F}}(x^k)$ is bounded below. Suppose that $\tilde{F}$ is not the optimal face. Then from Lemma 8.18, $\exp(g_{\tilde{F}}(x^k)) \to 0$. This contradicts (8.61). Hence, $\mathrm{Face}(\tilde{F})$ must be the optimal face. □

We show a few lemmas to prove Lemma 8.18.

LEMMA 8.19.

$$\frac{c^t x^{k+1} - c^t x^\infty}{c^t x^k - c^t x^\infty} = 1 - \mu^k \theta \frac{\|\hat{\alpha}_{\tilde{F}}^k\|}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k} + O(\tilde{\phi}_F^k) \tag{8.62}$$

where $\tilde{\phi}_F^k = \Phi_F^k + \|X_E^k\|^2$.

PROOF. We have

$$\begin{aligned} \frac{c^t x^{k+1} - c^t x^\infty}{c^t x^k - c^t x^\infty} &= \frac{\hat{c}_E^t x_E^{k+1} - w^{k+1} \hat{M}}{\hat{c}_E^t x_E^k - w^k \hat{M}} \qquad \text{(Use Proposition 6.4)} \\ &= 1 - \mu^k \frac{\hat{c}_E^t X_E^k(\theta \alpha_E^k / \|\alpha^k\| + \beta_E^k) - w^k \hat{M}}{\hat{c}_E^t x_E^k - w^k \hat{M}}. \end{aligned} \tag{8.63}$$

Now we see:

$$\begin{aligned}
(8.64) \qquad \frac{\hat{c}_E^t X_E^k \alpha_E^k}{\|\alpha^k\|} &= \hat{c}_E^t X_E^k \left(\frac{\hat{\alpha}_E^k}{\|\hat{\alpha}_E^k\|} + \mathcal{O}(\|X_E^k\|^2) \right) \qquad \text{(Use Lemma 6.7)} \\
&= \frac{\hat{c}_{\tilde{F}}^t X_{\tilde{F}}^k \hat{\alpha}_{\tilde{F}}^k}{\|\hat{\alpha}_{\tilde{F}}^k\|} + \mathcal{O}(\|X_{\tilde{F}}^k\| \Phi_F^k + \|X_E^k\|^3) \\
&\qquad \text{(Use Lemma 8.7 and Lemma 8.8)} \\
&= \|\hat{\alpha}_{\tilde{F}}^k\| + \mathcal{O}(\|X_E^k\| \tilde{\phi}_F^k) \\
&\qquad \text{(Recall the definition of } \hat{\alpha}_{\tilde{F}}^k)
\end{aligned}$$

and

$$\begin{aligned}
\hat{c}_E^t X_E^k \beta_E^k &= \mathcal{O}(\|x_F^k\|) + \mathcal{O}(\|X_{\tilde{F}}^k\| \Phi_F^k) \qquad \text{(Use Lemma 8.3)} \\
&= \mathcal{O}(\|X_{\tilde{F}}^k\| \Phi_F^k).
\end{aligned}$$

Hence, we have

$$\begin{aligned}
(8.65) \qquad \frac{c^t x^{k+1} - c^t x^\infty}{c^t x^k - c^t x^\infty} &= 1 - \mu^k \frac{\theta \|\hat{\alpha}_{\tilde{F}}^k\| + \mathcal{O}(\|X_E^k\| \tilde{\phi}_F^k + \|X_{\tilde{F}}^k\| \Phi_F^k) - w^k \hat{M}}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k + \hat{c}_F x_F^k - w^k \hat{M}} \\
&= 1 - \mu^k \theta \frac{\|\hat{\alpha}_{\tilde{F}}^k\| + \mathcal{O}(\|X_E^k\| \tilde{\phi}_F^k)}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k (1 + \mathcal{O}(\Phi_F^k))} \\
&\qquad \text{(Use (8.1), (8.4) and Lemma 8.11.)} \\
&= 1 - \mu^k \theta \frac{\|\hat{\alpha}_{\tilde{F}}^k\|}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k} \left(1 + \mathcal{O}(\Phi_F^k)\right) + \mathcal{O}(\tilde{\phi}_F^k) \\
&\qquad \text{(Use Lemma 8.11)} \\
&= 1 - \mu^k \theta \frac{\|\hat{\alpha}_{\tilde{F}}^k\|}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k} + \mathcal{O}(\tilde{\phi}_F^k). \\
&\qquad \text{(Use Lemma 8.11 and Corollary 8.14.)}
\end{aligned}$$

□

The following lemma is essential to the proof of Lemma 8.18.

LEMMA 8.20. *The difference of $g_{\tilde{F}}(x^k)$ at each iteration is evaluated as*

(8.66)

$$\begin{aligned}
\Delta g_{\tilde{F}}(x^k) &\triangleq g_{\tilde{F}}(x^{k+1}) - g_{\tilde{F}}(x^k) \\
&\le \frac{|\tilde{F}|\nu^k}{|\tilde{F}| - \nu^k} \|\hat{\eta}_{\tilde{F}}^k - \frac{e}{|\tilde{F}|}\|^2 \left\{ -|\tilde{F}| + \frac{\nu^k}{2(1 - \nu^k \sigma(\hat{\eta}_{\tilde{F}}^k))} \right\} + \mathcal{O}(\tilde{\phi}_F^k) \\
&\le -\epsilon \|\hat{\eta}_{\tilde{F}}^k - \frac{e}{|\tilde{F}|}\|^2 + \mathcal{O}(\tilde{\phi}_F^k)
\end{aligned}$$

for sufficiently large k where

$$\hat{\eta}_{\tilde{F}}^k \triangleq \frac{\hat{\alpha}_{\tilde{F}}^k}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k}, \qquad \nu^k \triangleq \mu^k\theta / \|\hat{\eta}_{\tilde{F}}^k\| \tag{8.67}$$

and ϵ is a positive constant.

PROOF. By using Lemma 8.19, we have

(8.68)

$$\begin{aligned}\Delta g_{\tilde{F}}(x^k) &= |\tilde{F}| \log\left\{\frac{c^t x^{k+1} - c^t x^\infty}{c^t x^k - c^t x^\infty}\right\} - \sum_{j\in\tilde{F}} \log x_j^{k+1}/x_j^k \\ &= |\tilde{F}| \log\left\{1 - \mu^k\theta \frac{\|\hat{\alpha}_{\tilde{F}}^k\|}{\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k} + \mathcal{O}(\tilde{\phi}_F^k)\right\} - \sum_{j\in\tilde{F}} \log(1 - \mu^k(\theta\alpha_j^k/\|\alpha^k\| + \beta_j^k)).\end{aligned}$$

In view of Lemma 6.7 and Lemma 8.8, we see that

$$\frac{\alpha_{\tilde{F}}^k}{\|\alpha^k\|} = \frac{\hat{\alpha}_{\tilde{F}}^k}{\|\hat{\alpha}_E^k\|} + \mathcal{O}(\|X_E^k\|^2) = \frac{\hat{\alpha}_{\tilde{F}}^k}{\|\hat{\alpha}_{\tilde{F}}^k\|} + \mathcal{O}(\|X_E^k\|^2 + \Phi_F(x^k)). \tag{8.69}$$

Since $\|\beta_{\tilde{F}}^k\| = \mathcal{O}(\Phi_F^k)$ due to 8.3, we have

$$\begin{aligned}\Delta g_{\tilde{F}}(x^k) &= |\tilde{F}| \log\left\{1 - \mu^k\theta\|\hat{\eta}_{\tilde{F}}^k\| + \mathcal{O}(\tilde{\phi}_F^k)\right\} \\ &\quad - \sum_{j\in\tilde{F}} \log\left\{1 - \mu^k\theta\frac{\hat{\alpha}_j^k}{\|\hat{\alpha}_{\tilde{F}}^k\|} + \mathcal{O}(\tilde{\phi}_F^k)\right\} \\ &= |\tilde{F}| \log\left\{1 - \frac{\mu^k\theta}{\|\hat{\eta}_{\tilde{F}}^k\|}\|\hat{\eta}_{\tilde{F}}^k\|^2 + \mathcal{O}(\tilde{\phi}_F^k)\right\} \\ &\quad - \sum_{j\in\tilde{F}} \log\left\{1 - \frac{\mu^k\theta}{\|\hat{\eta}_{\tilde{F}}^k\|}\hat{\eta}_j^k + \mathcal{O}(\tilde{\phi}_F^k)\right\} \\ &= |\tilde{F}| \log(1 - \nu^k\|\hat{\eta}_{\tilde{F}}^k\|^2 + \mathcal{O}(\tilde{\phi}_F^k)) \\ &\quad - \sum_{j\in\tilde{F}} \log(1 - \nu^k\hat{\eta}_j^k + \mathcal{O}(\tilde{\phi}_F^k)).\end{aligned} \tag{8.70}$$

Due to the facts that $\nu^k\sigma(\hat{\eta}_{\tilde{F}}^k) \le \mu^k\theta < 2/3$ and Lemma 8.6, we apply Lemma 4.10 to obtain the following bound

$$\Delta g_{\tilde{F}}(x^k) \le \frac{|\tilde{F}|\nu^k}{|\tilde{F}| - \nu^k}\|\hat{\eta}_{\tilde{F}}^k - \frac{e}{|\tilde{F}|}\|^2\left\{-|\tilde{F}| + \frac{\nu^k}{2(1-\nu^k\sigma(\hat{\eta}_{\tilde{F}}^k))}\right\} + \mathcal{O}(\tilde{\phi}_F^k) \tag{8.71}$$

which proves the first inequality of (8.66).

From Corollary 8.12 and the definition of $\hat{\alpha}_{\tilde{F}}^k$, we have

$$\|\hat{\eta}_{\tilde{F}}^k\| = \frac{\|\hat{\alpha}_{\tilde{F}}^k\|\|X_{\tilde{F}}^k\|}{\|X_{\tilde{F}}^k\|\hat{c}_{\tilde{F}}^t x_{\tilde{F}}^k} < \infty \tag{8.72}$$

for sufficiently large k. Furthermore, Corollary 5.3 tells us that μ^k is bounded below by a positive constant. Hence, ν^k is also bounded below by a positive constant.

By using Corollary 8.13, we have

$$
\begin{aligned}
(8.73) \qquad |\tilde{F}| - \nu^k &= |\tilde{F}| - \mu^k\theta/\|\hat{\eta}^k_{\tilde{F}}\| \\
&\geq |\tilde{F}| - \mu^k\theta|\tilde{F}| + \mathcal{O}(\Phi^k_{\tilde{F}}) \geq |\tilde{F}|/3 + \mathcal{O}(\Phi^k_{\tilde{F}}) \geq \epsilon' > 0
\end{aligned}
$$

for sufficiently large k. Finally, we have

$$
\begin{aligned}
(8.74) \qquad -|\tilde{F}| + \frac{\nu^k}{2(1-\nu^k\sigma(\hat{\eta}^k_{\tilde{F}}))} &= -|\tilde{F}| + \frac{\mu^k\theta/\|\hat{\eta}^k_{\tilde{F}}\|}{2(1-\mu^k\theta\sigma(\hat{\eta}^k_{\tilde{F}})/\|\hat{\eta}^k_{\tilde{F}}\|)} \\
&\leq -|\tilde{F}| + \frac{(2/3-\hat{\delta})|\tilde{F}|}{2/3+2\hat{\delta}} \\
&\leq -\delta' < 0.
\end{aligned}
$$

Now the lemma follows from (8.71) and the above observations. □

Now we are ready to prove Lemma 8.18.

PROOF OF LEMMA 8.18. By assumption, Face($\tilde{F}$) is not the optimal face. Note that if $|\tilde{F}| = 1$, then for sufficiently large k, $\hat{\eta}^k_{\tilde{F}} > 0$ follows from Lemma 8.6. Thus, from Lemma 8.17, Face($\tilde{F}$) is the optimal face which contradicts the assumption. Thus we assume $|\tilde{F}| \geq 2$.

Due to Lemma 8.17, $\hat{\alpha}^k_{\tilde{F}}$ (hence, $\hat{\eta}^k_{\tilde{F}}$) has at least one non-positive component. This and Lemma 8.6 imply

$$
(8.75) \qquad \|\hat{\eta}^k_{\tilde{F}} - \frac{e}{|\tilde{F}|}\| \geq \frac{1}{|\tilde{F}|(|\tilde{F}|-1)} + \mathcal{O}(\Phi^k_{\tilde{F}}).
$$

Hence, from Lemma 8.20, we have

$$
(8.76) \qquad \Delta g_{\tilde{F}}(x^k) \leq -\frac{\epsilon}{2|F|(|F|-1)} < 0
$$

for sufficiently large k where ϵ is identified in (8.66). Therefore, we have

$$
(8.77) \qquad g_{\tilde{F}}(x^k) \to -\infty,
$$

which completes the proof. □

8.5. Global Convergence of Dual Estimates for Optimality

Next we investigate the limiting behavior of the dual estimate for optimality s^k which is defined by (5.7). We have already seen in Section 5 that s^k is bounded and $s^k_{E^c} \to 0$.

We define the analytic center of the optimal face of $\langle DP_{F^c}\rangle$, the dual problem of $\langle P_{F^c}\rangle$ by the optimal solution of

$$
(8.78) \qquad \langle CDP_{F^c}\rangle \left\{ \begin{array}{ll} \text{minimize} & -\sum_{j\in\tilde{F}} \log s_j \\ \text{subject to} & s_{\tilde{F}} + A^t_{\tilde{F}}y = c_{\tilde{F}}, \quad s_{\tilde{F}} > 0, \\ & A^t_{E^c}y = c_{E^c}, \quad s_{E^c} = 0. \end{array} \right.
$$

The Karush-Kuhn-Tucker condition for $\langle CDP_{F^c}\rangle$ is

$$A_{\tilde{F}}\tilde{S}_{\tilde{F}}^{-1}e - A_{E^c}\tilde{\xi} = 0, \tag{8.79}$$

$$\tilde{s}_{\tilde{F}} + A_{\tilde{F}}^t\tilde{y} = c_{\tilde{F}}, \tag{8.80}$$

$$A_{E^c}^t\tilde{y} = c_{E^c}, \tag{8.81}$$

$$\tilde{s}_{\tilde{F}} > 0, \tag{8.82}$$

where $\tilde{S}_{\tilde{F}} \triangleq \mathrm{diag}(\tilde{s_F})$. The convexity of logarithm and the boundedness of the optimal face of $\langle DP_{F^c}\rangle$ imply that $\tilde{s}_{\tilde{F}}$ is unique.

We have the following lemmas.

LEMMA 8.21.

$$\sum_{k=0}^{\infty}\tilde{\phi}_F^k < \infty \tag{8.83}$$

PROOF. From Lemma 8.19 and Corollary 8.13, we see that $c^tx^k - c^tx^\infty$ converges to 0 asymptotically linearly. Thus we see that for sufficiently large k, there exists a positive constant M_1 such that

$$\sum_{l=k}^{\infty}\|x_E^l\| \le M_1\sum_{l=k}^{\infty}(c^tx^l - c^tx^\infty) < \infty \tag{8.84}$$

in consideration of Lemma 6.2. Since Λ_F^k converges to 0 asymptotically linearly, we have

$$\sum_{l=k}^{\infty}\Phi_F(x^l) \le M_2\sum_{l=k}^{\infty}\Lambda_F^l < \infty, \tag{8.85}$$

which, combined with (8.84), shows (8.83). □

LEMMA 8.22. (1) *$g_{\tilde{F}}(x^k)$ is bounded above.*

(2) $\hat{\eta}_{\tilde{F}}^k \to \dfrac{e}{|\tilde{F}|}$.

PROOF. Noting that $g_{\tilde{F}}(x^k)$ is bounded below due to Lemma 6.2, the lemma readily follows from Lemma 8.20 and Lemma 8.21. □

Now we are ready to prove the theorem.

THEOREM 8.23. *If the case (1) of Lemma 7.1 occurs, then $s_{F^c}^k \to \tilde{s}_{F^c}$, which is the analytic center of the optimal face of $\langle DP_{F^c}\rangle$.*

We do not mention the behavior of the F-components of s^k. In fact we do not even tell whether s_F^k converges or not. One may think that the above theorem is meaningless from practical viewpoint since F is not known. In the next subsection however, we prove that we can detect the index F by observing dual estimates for feasibility. Hence, Theorem 8.23 still gives us a good information on the dual problem.

PROOF OF THEOREM 8.23. We use $\hat{s}_{\tilde{F}}^k = (X_{\tilde{F}}^k)^{-1}\hat{\alpha}_{\tilde{F}}^k$ which is defined in the proof of Lemma 8.17 again. The proof has the following two steps.

(STEP 1) Prove $\hat{s}_{\tilde{F}}^k \to \tilde{s}_{\tilde{F}}$.
(STEP 2) Prove $\|\hat{s}_{\tilde{F}}^k - s_{\tilde{F}}^k\| \to 0$.

Then, recalling that $s^k_{E^c}$ converges to 0 due to Proposition 5.1, we have the theorem.
(STEP 1) $\hat{s}^k_{\tilde{F}} \to \tilde{s}_{\tilde{F}}$.

From (8.58), we see that $\hat{s}^k_{\tilde{F}}$ satisfies (8.80) and (8.81).

We see from Proposition 6.4 that

$$(8.86) \qquad c^t_{F^c} x^k_{F^c} - c^t_{F^c} x^\infty_{F^c} = \hat{c}^t_{\tilde{F}} x^k_{\tilde{F}} + \mathcal{O}(w^k + \|x^k_F\|).$$

Since Proposition 6.4 also holds if $w = 0$, we may assume that $\hat{c}_{\tilde{F}} > 0$ by strict complementarity of $\langle P_{F^c}\rangle$ and $\langle DP_{F^c}\rangle$.

For $j \in \tilde{F}$,

$$(8.87) \qquad \begin{aligned} p^k_j \triangleq \frac{c^t x^k - c^t x^\infty}{x^k_j} &= \frac{\hat{c}^t_{\tilde{F}} x^k_{\tilde{F}} + \hat{c}^t_F x^k_F - w^k \hat{M}}{x^k_j} \\ &\geq \min_{j \in \tilde{F}} \hat{c}_j + \mathcal{O}(\Phi^k_F + \Lambda^k_F) \geq \delta > 0 \end{aligned}$$

holds for sufficiently large k and a positive constant δ. On the other hand, (1) of Lemma 8.22 implies that there exists a constant M_1 such that

$$(8.88) \qquad \prod_{j \in \tilde{F}} p^k_j = \frac{(c^t x^k - c^t x^\infty)^{|\tilde{F}|}}{\prod_{j \in \tilde{F}} x^k_j} \leq M_1.$$

The inequalities (8.87) and (8.88) imply that there exists a constant M_2 such that

$$(8.89) \qquad p^k_j \leq M_2.$$

Hence p^k_j is bounded and every accumulation point of p^k_j is positive. Then every accumulation point of $\hat{s}^k_{\tilde{F}}$ is also positive since

$$(8.90) \qquad \hat{s}^k_j = \frac{\hat{\alpha}^k_j}{x^k_j} = p^k_j \hat{\eta}^k_j$$

and $\hat{\eta}^k_j \to 1/|\tilde{F}|$ due to (2) of Lemma 8.22.

Let

$$(8.91) \qquad q^k \triangleq \frac{x^k - x^\infty}{c^t x^k - c^t x^\infty}.$$

Lemma 6.2 implies that q^k is bounded above for sufficiently large k. Let k_l be a subsequence of k such that $q^{k_l} \to q^*$ which is an accumulation point. Then we see that p^{k_l} and $\hat{s}^{k_l}_{\tilde{F}}$ are also convergent to p^* and $\hat{s}^*_{\tilde{F}}$, respectively and that for $j \in \tilde{F}$, $|\tilde{F}| q^*_j = |\tilde{F}|(p^*_j)^{-1} = (\hat{s}^*_j)^{-1}$. Since

$$(8.92) \qquad Aq^k = \frac{A_F x^k_F}{c^t x^k - c^t x^\infty} + \frac{A_{\tilde{F}} x^k_{\tilde{F}}}{c^t x^k - c^t x^\infty} + \frac{A_{E^c}(x^k_{E^c} - x^\infty_{E^c})}{c^t x^k - c^t x^\infty} = \frac{w^k r^0}{c^t x^k - c^t x^\infty},$$

we have

(8.93)

$$
\begin{aligned}
A_{\tilde F} q_{\tilde F}^{k_t} &= -\frac{A_{E^c}(x_{E^c}^{\infty} - x_{E^c}^{k_t})}{c^t x^{k_t} - c^t x^{\infty}} + \mathcal{O}\Big(\frac{w^{k_t} + \|x_F^{k_t}\|}{c^t x^{k_t} - c^t x^{\infty}}\Big) \\
&= -\frac{A_{E^c}(x_{E^c}^{\infty} - x_{E^c}^{k_t})}{c^t x^{k_t} - c^t x^{\infty}} + \mathcal{O}\Big(\frac{w^{k_t} + \|x_F^{k_t}\|}{\|x_E^{k_t}\|} \frac{\|x_E^{k_t}\|}{c^t x^{k_t} - c^t x^{\infty}}\Big) \\
&= -\frac{A_{E^c}(x_{E^c}^{\infty} - x_{E^c}^{k_t})}{c^t x^{k_t} - c^t x^{\infty}} + \mathcal{O}(\Phi_F^{k_t} + \Psi_E^{k_t}). \qquad \text{(Use Lemma 6.1 and Lemma 6.2).}
\end{aligned}
$$

Taking limit of the above, we have

$$
A_{\tilde F} q_{\tilde F}^{*} = A_{E^c} q_{E^c}^{*}, \tag{8.94}
$$

which implies that $\hat{s}_{\tilde F}^{*}$ satisfies (8.79). Since every accumulation point of $\hat{s}_{\tilde F}^{k}$ satisfies (8.79)-(8.82) and $\tilde{s}_{\tilde F}$ is unique, we conclude that $\hat{s}_{\tilde F}^{k} \to \tilde{s}_{\tilde F}$.

(STEP 2) $\|\hat{s}_{\tilde F}^{k} - s_{\tilde F}^{k}\| \to 0$.

Since p_j^k is bounded above and below by a positive constant, we see

$$
\infty > M_3 \geq \frac{\max_{j \in \tilde F} p_j^k}{\min_{j \in \tilde F} p_j^k} = \|(X_{\tilde F}^k)^{-1}\|\|X_{\tilde F}^k\|. \tag{8.95}
$$

Thus we have

$$
\begin{aligned}
\|\hat{s}_{\tilde F}^k - s_{\tilde F}^k\| &= \|(X_{\tilde F}^k)^{-1}(\hat{\alpha}_{\tilde F}^k - \alpha_{\tilde F}^k)\| \\
&\leq \|(X_{\tilde F}^k)^{-1}\|(\|\alpha_{\tilde F}^k - \hat{\alpha}_{\tilde F}^k\| + \|\hat{\alpha}_{\tilde F}^k - \hat{\hat{\alpha}}_{\tilde F}^k\|) \\
&\leq M_1\|(X_{\tilde F}^k)^{-1}\|\|X_E^k\|^2\|\hat{\alpha}_E^k\| + M_2\|(X_{\tilde F}^k)^{-1}\|\|\Phi_F^k\|x_{\tilde F}^k\| \\
&\qquad \text{(Use Lemma 6.6 and (8.14))} \\
&\leq M_1 M_3 \|X_E^k\|\|\hat{\alpha}_E^k\| + M_2 M_3 \Phi_F^k \to 0
\end{aligned} \tag{8.96}
$$

and this completes the proof. □

8.6. Global Convergence of Dual Estimates for Feasibility

We consider the following optimization problem;

$$
\begin{cases} \text{minimize} & \|X^k z\|^2 \\ \text{subject to} & z = A^t y, \quad (r^0)^t y = 1. \end{cases} \tag{8.97}
$$

Let $z^k = A^t y^k$ be the optimal solution. We call z^k a dual estimate for feasibility.

By calculating Karush-Kuhn-Tucker condition, it is easy to see that

$$
z^k = w^k \frac{(X^k)^{-1}\beta^k}{\|\beta^k\|^2}. \tag{8.98}
$$

The optimal face of $\langle DF \rangle$ is

$$
\mathcal{D} \triangleq \left\{ z \,|\, z = A^t y, \quad z_{F^c} = 0, \quad z_F \geq 0, \quad (r^0)^t y = 1 \right\}. \tag{8.99}
$$

We define the analytic center of the optimal face of $\langle DF\rangle$ by the optimal solution of

$$(8.100)\qquad \langle CDF\rangle \begin{cases} \text{minimize} & -\sum_{j\in F}\log z_j \\ \text{subject to} & z\in\mathcal{D},\quad z_F>0. \end{cases}$$

The Karush-Kuhn-Tucker condition for $\langle CDF\rangle$ is

$$(8.101)\qquad \tilde{z}=A^t\tilde{y},$$

$$(8.102)\qquad A_F^t\tilde{Z}_F^{-1}e-A_{F^c}\tilde{\xi}_{F^c}-\tilde{\xi}'r^0=0,\quad (\tilde{Z}_F\triangleq \mathrm{diag}(\tilde{z}_F)),$$

$$(8.103)\qquad \tilde{z}_F>0,$$

$$(8.104)\qquad \tilde{z}_{F^c}=0,$$

$$(8.105)\qquad (r^0)^t\tilde{y}=1.$$

Since $\langle F\rangle$ has an interior feasible solution, the optimal face of $\langle DF\rangle$ is bounded, thus $\tilde{z}$ is unique.

Then we have the following theorem.

THEOREM 8.24. *If the case (1) of Lemma 7.1 occurs, then* $z^k\to\tilde{z}$.

PROOF. The equations (8.101) and (8.105) follow from the definition.

From (8.1), $f_F(x^k)$ is bounded below. Then $\bar{\beta}_F^k\to e$ in view of (8.49) and Lemma 8.21. This and Lemma 8.4 imply that

$$(8.106)\qquad \beta_F^k\to e.$$

Hence $\|\beta^k\|$ is bounded below. Since $\Lambda_F^k\to 0$, it is easy to see that $z_{F^c}^k\to 0$ from (8.98), which proves (8.104).

We also easily see that $f_F(x^k)$ is bounded above due to (8.49) and Lemma 8.21.

Let $q_j^k\triangleq x_j^k/w_j^k$ for $j\in F$. Since F determines the optimal face of $\langle F\rangle$, we can write $w^k=\hat{z}_F^t x_F^k$ where $\hat{z}_F>0$ by strict complementarity. Hence $\|q^k\|$ is bounded above.

On the other hand,

$$(8.107)\qquad \prod_{j\in F}(q_j^k)^{-1}=\exp(f_F(x^k))$$

is bounded below by a positive constant and above since $f_F(x^k)$ is bounded below and above. Hence, q_j is also bounded below by a positive constant. Therefore, q_F^k has an accumulation point q_F^* which is strictly positive. Let k_t be a subsequence of k with which $q_F^{k_t}\to q_F^*$. Then for $j\in F$, $(z_j^{k_t})^{-1}\to |F|q_j^*>0$ holds from (8.98) and (8.106).

Now

$$(8.108)\qquad A_Fq_F^{k_t}=\frac{A_Fx_F^{k_t}}{w^{k_t}}=\frac{1}{w^{k_t}}\left(w^{k_t}r^0-A_{F^c}(x_{F^c}^{k_t}-\hat{x}_{F^c})\right)$$
$$=r^0-\frac{A_{F^c}(x_{F^c}^{k_t}-\hat{x}_{F^c})}{w^{k_t}}.$$

Since left hand side is convergent, the right hand side is also convergent and there exists some $q_{F^c}^*$ such that

$$(8.109)\qquad A_Fq_F^*=r^0-A_{F^c}q_{F^c}^*.$$

This implies that z_F^* satisfies (8.102) by setting $\tilde{\xi}' = |F|$. In addition, obviously $z_F^* > 0$. Hence z^* satisfies (8.101)-(8.105). Now from the uniqueness of $\tilde{z}$, we have $z^k \to \tilde{z}$. $\square$

It is worth noting that we can construct strict complementarity solutions of $\langle P \rangle$ and $\langle DP \rangle$ due to Theorem 8.24. The key observation is that we already have an "infinite direction" of the optimal face of $\langle DP \rangle$, $\tilde{z}$. To show this, recall that s^k is bounded due to Lemma 1.4.1 thus there exists a constant M such that $|s_j^k| < M$ for all $j \in F$. Let s_F^* be an accumulation point of s_F^k. Then,

$$(8.110) \qquad \begin{pmatrix} \zeta_F^* \\ \zeta_{\tilde{F}}^* \\ 0 \end{pmatrix} \triangleq \begin{pmatrix} s_F^* \\ \tilde{s}_{\tilde{F}} \\ 0 \end{pmatrix} + \frac{M}{\min_{j \in F} \tilde{z}_j} \begin{pmatrix} \tilde{z}_F \\ 0 \\ 0 \end{pmatrix}$$

and $(0, x_E^\infty)$ satisfy the strict complementarity condition.

The dual estimate for feasibility z^k also plays important role when we detect infeasibility of $\langle P \rangle$. We have the following theorem.

THEOREM 8.25. *If $c^t x^k$ is bounded below and $\langle P \rangle$ is infeasible, then $z^k \to \tilde{z}$ which is the analytic center of the optimal face of $\langle DF \rangle$.*

Due to the above theorem, if x^∞ and z^∞ satisfy strict complementarity, we know that $\langle P \rangle$ does not have a feasible solution. In view of the proof of Theorem 4.12, the proof of the above theorem is almost same as that of Theorem 8.24, hence we omit the proof.

9. Case (2) of Lemma 7.1

In this section, we deal with the case (2) of Lemma 7.1. Surprisingly, the main lemma is as follows.

LEMMA 9.1. *The case (2) of Lemma 7.1 cannot occur under Assumption 4.*

Recall that we make Assumption 4 in the analysis up to now and that $\langle P \rangle$ does not have an interior feasible solution in the case (1) of Lemma 7.1 due to Lemma 8.1. Therefore, Lemma 9.1 implies that if $\langle P \rangle$ has an interior feasible solution, then Assumption 4 must break, thus the proposed method finds an interior feasible solution in a finite number of iterations. This proves, combined with Theorem 8.16 and Theorem 4.12, Theorem 2.4.

To prove Lemma 9.1, we will prove the following lemma.

LEMMA 9.2. *If the case (2) of Lemma 7.1 occurs, then $\beta_F^k \to 0$.*

Once the above lemma is proved, Lemma 9.1 can be proved as follows.

PROOF OF LEMMA 9.1. We first see that for $j \notin F$,

$$(9.1) \qquad \frac{w^{k_t}}{x_j^{k_t}} \leq \Lambda_F^{k_t} \to 0$$

for some subsequence k_t. Since

$$(9.2) \qquad \beta^k = w^k (X^k)^{-1} X^k P_{AX^k} (X^k)^{-1} (x^0 - x^\infty)$$

and $X^k P_{AX^k} (X^k)^{-1} (x^0 - x^\infty)$ is bounded due to Lemma 1.4.1, (9.1) implies $\sigma(\beta_{F^c}^{k_t}) \to 0$. This and Lemma 9.2 imply $\sigma(\beta^{k_t}) \to 0$. On the other hand, obviously $\sigma(\theta \alpha^k / \|\alpha^k\|) \leq$

$\theta < 1$. Hence we have $\sigma((X^{k_t})^{-1}\Delta x^{k_t}) < 1$ for sufficiently large t, which contradicts (5.1) and Assumption 4. □

Hence, all we should prove is Lemma 9.2. We use $\bar{\beta}_F^k$ again which is defined by the optimal solution of $\langle \bar{B}_F \rangle$ in (8.5). It is easy to see that Lemma 8.4 also holds in the case, thus to prove Lemma 9.2, it is sufficient to show that $\bar{\beta}_F^k \to 0$. In fact, we give the following lemma which is stronger than we need.

LEMMA 9.3. *If Ψ_F^k is not bounded above, then $\bar{\beta}_F^k = 0$ for all k.*

To prove Lemma 9.3, we give the following two lemmas:

LEMMA 9.4. *One and only one of the following two cases occurs:*
(1) *$\bar{\beta}_F^k = 0$ for all k; or*
(2) *$\bar{\beta}_F^k \neq 0$ for all k.*

LEMMA 9.5. *If $\bar{\beta}_F^k \neq 0$, then Ψ_F^k is bounded above by a constant depending only on A, b, x^0.*

Then, Lemma 9.3 is proved as follows.

PROOF OF LEMMA 9.3. Suppose contrary that $\bar{\beta}_F^k \neq 0$ for some k. Then Lemma 9.4 imply that $\bar{\beta}_F^k \neq 0$ for all k, and Lemma 9.5 imply Ψ_F^k is bounded above by a constant for all k. This contradicts the assumption. □

To prove Lemma 9.4, the following lemma is essential.

LEMMA 9.6.

$$\bar{\beta}_F^k = 0 \Leftrightarrow A_F x_F^k \in \mathrm{Im}(A_{F^c}). \tag{9.3}$$

PROOF. The Karush-Kuhn-Tucker condition for $\langle \bar{B}_F \rangle$ is

$$A_F (X_F^k)^2 A_F^t y - A_F x_F^k - A_{F^c}\xi = 0, \tag{9.4}$$

$$A_{F^c}^t y = 0. \tag{9.5}$$

Since $\bar{\beta}_F^k = X_F^k A_F^t y$, necessity is obvious. To prove sufficiency, we assume that there exists a vector ξ' such that $A_F x_F^k = -A_{F^c}\xi'$. Then, $(y, \xi) = (0, \xi')$ satisfies (9.4) and (9.5), and we get $\bar{\beta}_F^k = 0$. □

Then we can prove Lemma 9.4.

PROOF OF LEMMA 9.4. Since

$$\begin{aligned} A_F x_F^k &= w^k r^0 + b - A_{F^c} x_{F^c}^k \\ &= w^k A_F x_F^0 + A_{F^c}\{w^k x_{F^c}^0 + (1 - w^k)\hat{x}_{F^c} - x_{F^c}^k\}, \end{aligned} \tag{9.6}$$

we see

$$A_F x_F^k \in \mathrm{Im}(A_{F^c}) \Leftrightarrow A_F x_F^0 \in \mathrm{Im}(A_{F^c}). \tag{9.7}$$

Now the lemma readily follows from Lemma 9.6. □

Next, we prove Lemma 9.5. We set for x^k such that $\bar{\beta}_F^k \neq 0$,

$$\bar{z}_F^k \triangleq \frac{w^k (X_F^k)^{-1} \bar{\beta}_F^k}{\|\bar{\beta}_F^k\|^2}. \tag{9.8}$$

Then we have the following lemma.

LEMMA 9.7. *$\bar{z}_F^k$ is bounded by a constant depending only on A, b, x^0.*

PROOF. Since (x^k, w^k) is a feasible solution of $\langle F \rangle$, we have

$$A_F x_F^k + A_{F^c} x_{F^c}^k - A_{F^c} \hat{x}_{F^c} = w^k (A_F x_F^0 + A_{F^c} x_{F^c}^0 - A_{F^c} \hat{x}_{F^c}). \tag{9.9}$$

Let $\{\bar{v}_1, \dots, \bar{v}_{\bar{l}}\}$ be a basis of $\mathrm{Null}(A_{F^c}^t)$ where $\bar{l} \triangleq \dim \mathrm{Null}(A_{F^c}^t)$. Since $\bar{\beta}_F^k \neq 0$, $\mathrm{Null}(A_{F^c}^t)$ is not $\{0\}$. We denote by $\bar{N}$ the matrix whose j-th row vector is $\bar{v}_j^t$, that is, $\bar{N}^t = (\bar{v}_1, \dots, \bar{v}_{\bar{l}})$. We multiply both sides of (9.9) by $\bar{N}$ and get

$$\bar{A}_F x_F^k = w^k \bar{A}_F x_F^0 \tag{9.10}$$

where $\bar{A}_F \triangleq \bar{N} A_F$.

It is easy to see that $\langle \bar{B}_F \rangle$ can be written as

$$\begin{cases} \text{minimize} & \|\beta_F - e\|^2 \\ \text{subject to} & \beta_F = X_F^k \bar{A}_F y. \end{cases} \tag{9.11}$$

Let $(\bar{y}, \bar{z})$ be the optimal solution of the above problem. Then $\bar{y}$ is a solution of

$$\bar{A}_F (X_F^k)^2 \bar{A}_F^t y - \bar{A}_F x_F^k = 0. \tag{9.12}$$

Due to (9.10), the above equation becomes

$$\bar{A}_F (X_F^k)^2 \bar{A}_F^t y = w^k \bar{A}_F X_F^k (X_F^k)^{-1} x_F^0 \tag{9.13}$$

which is a normal equation and we have

$$\bar{\beta}_F^k = X_F^k \bar{A}_F^t \bar{y} = w^k P_{\bar{A}_F X_F^k} (X_F^k)^{-1} x_F^0. \tag{9.14}$$

There exists a matrix B such that $\mathrm{Im}(X_F^k \bar{A}_F^t) = \mathrm{Null}(B (X_F^k)^{-1})$. By using B, we have the following inequality:

(9.15)
$$\begin{aligned} \|(X_F^k)^{-1} P_{\bar{A}_F X_F^k} (X_F^k)^{-1} x_F^0\| &= \|(X_F^k)^{-1} (I - P_{B(X_F^k)^{-1}}) (X_F^k)^{-1} x_F^0\| \\ &\leq M \|(I - P_{B(X_F^k)^{-1}}) (X_F^k)^{-1} x_F^0\|^2 \\ &= M \|P_{\bar{A}_F X_F^k} (X_F^k)^{-1} x_F^0\|^2 \qquad \text{(Use Lemma 1.4.4)} \end{aligned}$$

where $M = M(\bar{A}_F, x_F^0)$. Then we have

$$\begin{aligned} w^k \|(X_F^k)^{-1} \bar{\beta}_F^k\| &= (w^k)^2 \|(X_F^k)^{-1} P_{\bar{A}_F X_F^k} (X_F^k)^{-1} x_F^0\| \\ &\leq (w^k)^2 M \|P_{\bar{A}_F X_F^k} (X_F^k)^{-1} x_F^0\|^2 \\ &= M \|\bar{\beta}_F^k\|^2. \end{aligned} \tag{9.16}$$

Now the lemma readily follows. □

Now we are ready to prove Lemma 9.5.

PROOF OF LEMMA 9.5. Noting that $\bar{\beta}_F^k$ is a projection of e, we have

$$(9.17) \qquad (x_F^k)^t \bar{z}_F^k = \frac{w^k e^t \bar{\beta}_F^k}{\|\bar{\beta}_F^k\|^2} = w^k.$$

Then by using Lemma 9.7, we see that

$$(9.18) \qquad \frac{w^k}{\|x_F^k\|} = \frac{(x_F^k)^t \bar{z}_F^k}{\|x_F^k\|} \le \|\bar{z}_F^k\| \le M,$$

which completes the proof. □

10. Summary

We summarize the convergence results of Algorithm 1 derived in this paper.

When $\langle DP \rangle$ has a feasible solution, $c^t x^k$ is bounded below, thus we see from Theorem 3.1 that x^k converges to x^∞. The convergence results of this case are divided into three cases.

(1) (Theorem 2.4) When $\langle P \rangle$ has an interior feasible solution, x^∞ is a relative interior point of the optimal face of $\langle P \rangle$, and dual estimates for optimality s^k converge to the analytic center of the optimal face of $\langle DP \rangle$. There exists a finite number K such that for $k \ge K$, x^k is feasible. Once x^k is feasible, the proposed method is reduced to a feasible affine scaling method.

(2) (Theorem 2.4, Theorem 8.23 and Theorem 8.24) When $\langle P \rangle$ has a feasible solution but no interior feasible solution, x^∞ is a relative interior point of the optimal face of $\langle P \rangle$. Dual estimates for feasibility z^k converge to the analytic center of the optimal face of $\langle DF \rangle$ while F^c components of dual estimates for optimality $s_{F^c}^k$ converge to the analytic center of the optimal face of $\langle DP_{F^c} \rangle$ which is the dual problem of $\langle P_{F^c} \rangle$ defined in (8.55). Here, F is the index set which decides the feasible region of $\langle P \rangle$ and F^c is its complement.

(3) (Theorem 4.12 and Theorem 8.25) When $\langle P \rangle$ has no feasible solution, then x^∞ is a relative interior point of the optimal face of $\langle F \rangle$ and dual estimates for feasibility z^k converge to the analytic center of the optimal face of $\langle DF \rangle$.

When $\langle DP \rangle$ has no feasible solution, we have two cases.

(1) (Theorem 2.4) When $\langle P \rangle$ has a feasible solution, $c^t x^k$ diverges to minus infinity and x^k also diverges.

(2) (Theorem 2.4, Theorem 4.12 and Theorem 8.25) When $\langle P \rangle$ has no feasible solution, we do not tell whether $c^t x^k$ is bounded below or not. If $c^t x^k$ is bounded below, then the primal sequence converges to a relative interior point of the optimal face of $\langle F \rangle$ and dual estimates for feasibility converge to the analytic center of the optimal face of $\langle DF \rangle$. Otherwise, x^k diverges.

The above is all what we get in this chapter. But since we do not know about the feasibility of $\langle P \rangle$ and $\langle DP \rangle$ before the algorithm starts, when we use the algorithm, what we need is the classification of the results on the behavior of the sequence. From the viewpoint, we have the following three cases.

(1) *The algorithm finds an interior feasible solution in a finite number of iterations.* In the case, initialization of interior point method is succeeded, and we can start any kind of feasible interior point method for solving linear programming problem.

A natural method in the context of our method is a long-step affine scaling method. The reason is found in Section 2.

(2) *The algorithm does not find an interior feasible solution, but the primal sequence converges.*
In the case, if the limit point of the primal sequence is feasible, then it is optimal. The original LP problem $\langle P \rangle$ does not have an interior feasible solution, and dual estimates for feasibility tells us the indices which is always-active on the feasible region. Combining the dual estimates for feasibility and optimality, we get a strict complementarity solutions of $\langle P \rangle$ and $\langle DP \rangle$.
If the limit point is not feasible, then $\langle P \rangle$ is not feasible.

(3) *The algorithm does not find an interior feasible solution and the primal sequence diverges.*
This implies that $\langle DP \rangle$ is not feasible, hence $\langle P \rangle$ does not have optimal solutions.

We can neither tell whether $\langle DP \rangle$ has a feasible solution or not when x^k converges to an infeasible point nor whether $\langle P \rangle$ has a feasible solution or not when x^k diverges. Convergence analysis in these cases deserves further research.

Appendix: Proof of Technical Lemmas

A.1. Proof of Lemma 2.1

We choose a basis of Null(A) as $b^1, \ldots, b^{n-m}$, and let $B^t \triangleq (b^1, \ldots, b^{n-m})$. Then,

$$AB^t = 0, \qquad \text{and} \qquad \mathrm{Im}(B^t) = \mathrm{Null}(A). \tag{A.1}$$

Let $\tilde{u}$ be a solution of $Au = r^0$. Since

$$\begin{aligned} Ax - wr^0 = b &\Leftrightarrow A(x - \hat{x} - w\tilde{u}) = 0, \\ &\Leftrightarrow x - \hat{x} - w\tilde{u} \in \mathrm{Null}(A), \end{aligned} \tag{A.2}$$

we can express Feas$\langle F \rangle$ as

$$\left\{ x \geq 0 \,|\, x = B^t y + w\tilde{u} + \hat{x} \right\}. \tag{A.3}$$

Then we can convert $\langle P \rangle$ to dual form by using the matrix $\tilde{B}^t \triangleq (B^t, \tilde{u})$, as

$$\begin{cases} \text{minimize} & w \\ \text{subject to} & \tilde{B}^t \begin{pmatrix} y \\ w \end{pmatrix} \geq -\hat{x}. \end{cases} \tag{A.4}$$

Now the formula of affine scaling direction for dual form LP can be applied and after some calculation, we have the affine scaling direction $d(x)$ as:

$$\begin{aligned} d(x) &= \tilde{B}^t(\tilde{B}X^{-2}\tilde{B}^t)^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = (B^t, \tilde{u}) \begin{pmatrix} BX^{-2}B^t & BX^{-2}\tilde{u} \\ \tilde{u}^t X^{-2} B^t & \tilde{u}^t X^{-2} \tilde{u} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= (B^t, \tilde{u}) \begin{pmatrix} -\pi(BX^{-2}B^t)^{-1} BX^{-2}\tilde{u} \\ \pi \end{pmatrix}, \end{aligned} \tag{A.5}$$

where

$$\pi \triangleq (\tilde{u}^t X^{-1}(I - P_{BX^{-1}})X^{-1}\tilde{u})^{-1}. \tag{A.6}$$

Continuing the calculation, we have

$$\begin{aligned} d(x) &= -\pi(B^t(BX^{-2}B^t)^{-1}BX^{-2}\tilde{u} - \tilde{u}) \\ &= \pi X(I - P_{BX^{-1}})X^{-1}\tilde{u} \\ &= \frac{X(I - P_{BX^{-1}})X^{-1}\tilde{u}}{\|(I - P_{BX^{-1}})X^{-1}\tilde{u}\|^2}. \end{aligned} \tag{A.7}$$

Note that the direction of w is π itself. To rescale the direction so that the length is w, we should multiply w/π to the direction and get

$$wX(I - P_{BX^{-1}})X^{-1}\tilde{u}. \tag{A.8}$$

We convert it to the primal form. Note that

$$I - P_{BX^{-1}} = P_{AX}. \tag{A.9}$$

Then we have

$$\Delta_n x = wXP_{AX}X^{-1}\tilde{u}. \tag{A.10}$$

From (A.2), we have

$$A(x^0 - \hat{x} - \tilde{u}) = 0, \tag{A.11}$$

so

$$P_{AX}X^{-1}\tilde{u} = P_{AX}X^{-1}(x^0 - \hat{x}). \tag{A.12}$$

Substituting (A.12) from (A.10), we have the lemma. □

A.2. Proof of Lemma 4.10

Due to (4.47), we have

$$\begin{aligned}
&\sum_{j=1}^{q} - \log(1 - \nu\eta_j + \Delta_j) \\
&= \sum_{j=1}^{q} - \log(1 - \nu\eta_j) + \sum_{j=1}^{q} - \log\left\{1 + \frac{\Delta_j}{1 - \nu\eta_j}\right\}.
\end{aligned} \tag{A.13}$$

Then,

(A.14)

$$\begin{aligned}
\sum_{j=1}^{q} - \log\left\{1 + \frac{\Delta_j}{1 - \nu\eta_j}\right\} &\leq \sum_{j=1}^{q} - \log(1 - 3|\Delta_j|) \\
&\leq 3\sum_{j=1}^{q} |\Delta_j| + \frac{9\sum_{j=1}^{q} \Delta_j^2}{2(1 - 3/4)} \qquad \text{(Use Proposition 2.4.2.)} \\
&\leq 3\sum_{j=1}^{q} |\Delta_j| + 18\sum_{j=1}^{q} \Delta_j^2 \\
&\leq 21\sum_{j=1}^{q} |\Delta_j|.
\end{aligned}$$

On the other hand, from (4.46) and (4.49),

$$\sigma(\eta) \geq (1 + \Delta_0)/q > 0 \tag{A.15}$$

follows. Hence, we have from (4.47) and (4.49), $\nu \leq 2q/(3(1 - |\Delta_0|)) \leq 8q/9$ and

$$q - \nu \geq q/9 > 0. \tag{A.16}$$

We set

$$\rho \triangleq \eta - e/q. \tag{A.17}$$

It is easy to verify:

$$e^t\rho = \Delta_0 \qquad \text{and} \tag{A.18}$$

$$\|\rho\|^2 = \|\eta\|^2 - 1/q - 2\Delta_0/q. \tag{A.19}$$

Now we make the following evaluation:

(A.20)

$$
\begin{aligned}
& q\log(1-\nu\|\eta\|^2+\Delta_0') - \sum_{j=1}^{q}\log(1-\nu\eta_j+\Delta_j) \\
& \le q\log(1-\nu\|\eta\|^2+\Delta_0') - \sum_{j=1}^{q}\log(1-\nu\eta_j) + 21\sum_{j=1}^{q}|\Delta_j| \qquad \text{(Use (A.14))} \\
& = q\log\left\{1-\nu(\|\rho\|^2+1/q-2\Delta_0/q)+\Delta_0'\right\} \\
& \qquad - \sum_{j=1}^{q}\log\{1-\nu(\rho_j+1/q)\} + 21\sum_{j=1}^{q}|\Delta_j| \qquad \text{(Use (A.17) and (A.19))} \\
& = q\log\left\{\frac{q-\nu}{q}-\nu\|\rho\|^2+2\nu\Delta_0/q+\Delta_0'\right\} - \sum_{j=1}^{q}\log\left\{\frac{q-\nu}{q}-\nu\rho_j\right\} + 21\sum_{j=1}^{q}|\Delta_j| \\
& = q\log\left\{1-\frac{q\nu}{q-\nu}\|\rho\|^2+\frac{(2\nu\Delta_0+q\Delta_0')}{q-\nu}\right\} \\
& \qquad - \sum_{j=1}^{q}\log\left\{1-\frac{q\nu}{q-\nu}\rho_j\right\} + 21\sum_{j=1}^{q}|\Delta_j| \qquad \text{(Use (A.16))} \\
& \le -\frac{q^2\nu}{q-\nu}\|\rho\|^2 + \frac{q}{q-\nu}(2\nu\Delta_0+q\Delta_0') + \frac{q\nu}{q-\nu}\Delta_0 + \frac{\left(\frac{q\nu}{q-\nu}\right)^2\|\rho\|^2}{2\left(1-\left(\frac{q\nu}{q-\nu}\right)\sigma(\rho)\right)} \\
& \qquad + 21\sum_{j=1}^{q}|\Delta_j| \qquad \text{(Use Proposition 2.4.2 and (A.17))} \\
& \le \frac{q\nu}{q-\nu}\|\eta-\frac{e}{q}\|^2\left\{-q+\frac{\nu}{2(1-\nu\sigma(\eta))}\right\} + 24q|\Delta_0| + 9q|\Delta_0'| + 21\sum_{j=1}^{q}|\Delta_j|,
\end{aligned}
$$

and this completes the proof. □

A.3. Proof of Lemma 7.2

There exists $d_{F^c}^k$ which is the solution of

(A.21)
$$A_{F^c}(d_{F^c}-x_{F^c}^k) = -w^k r^0 + A_F x_F^k,$$

and satisfy:

(A.22)
$$\|d_{F^c}^k - x_{F^c}^k\| \le M_0\|w^k r^0 - A_F x_F^k\| \le M_1(w^k+\|x_F^k\|).$$

Let $(\Phi_F(x^{k_t}), \Lambda_F(x^{k_t})) \to 0$ as $t\to\infty$. Then

(A.23)
$$A_{F^c} d_{F^c}^{k_t} = A_F x_F^{k_t} + A_{F^c} x_{F^c}^{k_t} - w^{k_t} r^0 = b$$

holds for all t and we have

(A.24)
$$|x_j^{k_t} - d_j^{k_t}| \le M_1(\|x_F^{k_t}\| + w^{k_t}).$$

If $j \in F^c$,

$$d_j^{k_t} \geq x_j^{k_t} - M_1(\|x_F^{k_t}\| + w^{k_t}) > 0. \tag{A.25}$$

holds for sufficiently large t, and $(0, d_{F^c}^{k_t})$ satisfies (7.5). □

A.4. Proof of Lemma 7.4

We first give the following lemma.

LEMMA A.1. *If $\{(\Phi^k, \Lambda^k) | k = 0, 1, \dots\}$, pairs of real positive numbers satisfy the following conditions;*

(1) *The sequence has 0 as an accumulation point.*
(2) *There exist some constants $\tilde{\epsilon} > 0, 0 < M_1, M_2, M_3 < \infty$ and $\delta > 0$ such that if $\Phi^k < \tilde{\epsilon}$, then*

$$\Phi^k \leq M_1 \Lambda^k \leq M_2 \Phi^k \tag{A.26}$$

$$\Lambda^{k+1} \leq (1 - \delta)\Lambda^k \tag{A.27}$$

$$\Phi^{k+1} \leq M_3 \Phi^k, \tag{A.28}$$

then $(\Phi^k, \Lambda^k) \to 0$.

PROOF. Given any positive $\epsilon < \tilde{\epsilon}$, let

$$\Omega(\epsilon) \triangleq \left\{ k \,|\, \Phi^k \leq \epsilon, \Lambda^k \leq \epsilon \right\}. \tag{A.29}$$

Choose $\hat{\epsilon}$ such that $\hat{\epsilon} < \epsilon / \max(M_2, M_3)$. Without loss of generality, we assume that $M_1, M_2, M_3 > 1$. From assumption (1), there exists a number K such that

$$\Phi^K < \hat{\epsilon} \qquad \text{and} \qquad M_1 \Lambda^K < \hat{\epsilon}. \tag{A.30}$$

Obviously, $K \in \Omega(\hat{\epsilon}) \subseteq \Omega(\epsilon)$. We now see from assumption (2) that

$$\Phi^{K+1} \leq M_3 \Phi^K < M_3 \hat{\epsilon} < \epsilon < \tilde{\epsilon}, \tag{A.31}$$

which means the condition of (2) is also satisfied at the $(K+1)$-th element. Then using (A.26), (A.27) and (A.30), we have

$$\Phi^{K+1} \leq M_1 \Lambda^{K+1} \leq M_1(1 - \delta)\Lambda^K \leq (1 - \delta)\hat{\epsilon} < \hat{\epsilon} < \epsilon. \tag{A.32}$$

This implies that $K + 1 \in \Omega(\hat{\epsilon}) \subset \Omega(\epsilon)$ and (A.30) also holds at $(K+1)$-th element. Applying the above procedure recursively, we see that $k \in \Omega(\epsilon)$ if $k \geq K$. □

Now we are ready to prove Lemma 7.4.

PROOF OF LEMMA 7.4. We confirm the condition to use Lemma A.1.
First, choosing a subsequence $\{ k_t \,|\, t = 0, 1, \dots \}$ of k such that $\Phi_F^{k_t} \to 0$, we have

$$\Lambda_F^{k_t} = \mathcal{O}(\Phi_F^{k_t} \Psi_F^{k_t}) = \mathcal{O}(\Phi_F^{k}) \tag{A.33}$$

since Ψ_F^k is bounded above. This means (Φ_F^k, Λ_F^k) has 0 as an accumulation point.

Next from the assumption (4), there exists a small positive number ϵ such that if $\Phi_F^k < \epsilon$, then $\Psi_F^k \geq \delta$ for a positive constant δ. In the case,

$$\Phi_F^k \leq (\Psi_F^k)^{-1}\Lambda_F^k \leq \frac{1}{\delta}\Lambda_F^k \qquad \text{and} \tag{A.34}$$

$$\Lambda_F^k \leq |F|\Phi_F^k\Psi_F^k \leq M\Phi_F^k \tag{A.35}$$

hold.

Further, from the definition of the iteration, we have the following expression:

$$\frac{x_j^{k_t+1}}{x_j^{k_t}} = 1 - \mu^{k_t}\left(\theta\frac{\alpha_j^{k_t}}{\|\alpha^k\|} + \frac{w^{k_t}}{x_j^{k_t}}\xi_j^{k_t}\right) \quad \text{for } j = 1,\ldots,n \tag{A.36}$$

where $\xi^k = X^k P_{AX^k}(X^k)^{-1}(x^0 - \hat{x})$. Note that ξ is bounded from Lemma 1.4.1. Noting also that $\theta < 1$, we see that there exists a positive constant ϵ' such that

$$\frac{x_j^{k_t+1}}{x_j^{k_t}} \geq 1 - \mu^{k_t} + \epsilon' \tag{A.37}$$

for $j \notin F$ and sufficiently large t since $w^{k_t}/x_j^{k_t} \to 0$. Then we have the following estimation:

$$\begin{aligned}\frac{\Lambda^{k_t+1}}{\Lambda^{k_t}} &\leq \frac{w^{k_t+1}}{w^{k_t}}\left(\min_{j\notin F}\frac{x_j^{k_t+1}}{x_j^{k_t}}\right)^{-1} \\ &\leq (1-\mu^k)(1-\mu^k+\epsilon')^{-1} \\ &\leq 1-\epsilon'',\end{aligned} \tag{A.38}$$

for sufficiently large t.

Next in consideration of Corollary 4.3, we see that β_F^k is bounded. Therefore, there exists a constant M' such that

$$\frac{x_j^{k+1}}{x_j^k} = 1 - \mu^k\left(\theta\frac{\alpha_j^k}{\|\alpha^k\|} + \beta_j^k\right) \leq M' \tag{A.39}$$

for $j \in F$. Hence we have

$$\begin{aligned}\Phi_F^{k+1} &= \frac{\|X_F^{k+1}\|}{\min_{j\notin F} x_j^{k+1}} \leq \frac{M'\|X_F^k\|}{(1-\lambda^k)\min_{j\notin F} x_j^k} \\ &\leq M''\Phi_F^k.\end{aligned} \tag{A.40}$$

Now due to (A.34),(A.35),(A.38) and (A.40), we can apply Lemma A.1 which shows $\Phi_F^k \to 0$. Since (A.27) holds if Φ_F^k is sufficiently small, Λ_F^k converges to 0 asymptotically linearly, and this completes the proof. □

Conclusion and Direction in Further Research

In Chapter II, we give an answer for the first problem (identified in the preface) on the global convergence of the affine scaling method in a practical setting. The result is that the long step-size 2/3 is admissible for ensuring the global convergence of the primal and dual sequences. To date, this is the best result on the global convergence of the affine scaling method. The bound 2/3 is tight in the sense that if $\lambda > 2/3$, then we can construct an example showing that the dual estimates do not converge to the analytic center of the optimal face of the dual problem [61, 22]. We also prove that the asymptotic reduction rate of the objective function value is $1 - \lambda$, which is independent of the dimension of the LP problem. Since 2/3 is tight for the dual sequence, an interesting direction in further research is to find the tight bound for the global convergence of the primal sequence. (Recall that $\lambda = 0.99$ is adopted in most of the implementations.) Very recently we received a paper by Mascarenhas [32] showing an example that by taking $\lambda = 0.999$, the long-step affine scaling method fails to converge to an optimal solution. Searching a tight true bound for the primal sequence is still an open problem.

In Chapter III, the results of Chapter II are applied to obtain a new complexity proof of a long-step variant of the projective scaling method. In this variant, we use a fixed step-size $\lambda \leq 2/3$. By analyzing the inequality of the potential reduction more precisely, we give a new proof of the polynomiality with complexities of $O(nL)$ and $O(n^2L)$ iterations for $\lambda < 2/3$ and $\lambda = 2/3$, respectively. The asymptotic reduction rate of the objective function is shown to be $1 - \lambda$. We also prove that the primal sequence converges to a relative interior point of the optimal face, and that the dual estimates converge to the analytic center of the optimal face of the dual problem without the assumption of the boundedness of the optimal face. These convergence results are new to the projective scaling algorithm. Finally, we apply the well-known Todd and Burrell's lower bound updating procedure [56] to our variant without affecting the complexity results.

In Chapter IV, we propose an extension of the affine scaling method to allow an infeasible starting point which is a solution to the second problem identified in the preface. We discuss the case where there exists an interior feasible point as well as the case where no interior feasible point or no feasible point exists. It is shown that (i) if the optimal solution of the LP problem exists, then the primal sequence converges to a relative interior point of the optimal face, (ii) if there exists an interior feasible region, then an interior feasible solution is obtained in a finite number of iterations, and then the method becomes identical to the long-step affine scaling method, (iii) dual estimates for optimality converge to the analytic center of the optimal face of the dual problem, (iv) dual estimates for feasibility give us useful information about the feasible region. All of the convergence results are obtained without nondegeneracy assumptions. A weak point of the proposed method however, is that the optimality direction is a short-step affine scaling direction and might be too short in practice. It is desirable to lengthen the optimality direction. Convergence analysis of this variant as well as the numerical experience deserves further research.

Bibliography

[1] I. Adler, M. Resende, G. Veiga and N. Karmarkar : An implementation of Karmarkar's algorithm for linear programming. *Mathematical Programming,* Vol. 44 (1989), pp. 297-335.

[2] I. Adler, N. Karmarkar, M. Resende and G. Veiga : Data structures and programming techniques for the implementation of Karmarkar's algorithm. *ORSA Journal on Computing,* Vol. 1, No. 2 (1989), pp. 84-106.

[3] I. Adler and R. D. C. Monteiro: Limiting behavior of the affine scaling continuous trajectories for linear programming problems. *Mathematical Programming,* Vol. 50, pp. 29-51.

[4] K. D. Andersensen : An infeasible dual affine scaling method for linear programming. *Technical Report,* Institut for Matematik og Datalogi, Odense Universitet, Denmark (1993).

[5] K. M. Anstreicher : The worst-case step in Karmarkar's algorithm. *Mathematics of Operations Research,* Vol. 14 (1989), No. 2, pp. 294-302.

[6] K. M. Anstreicher : A combined phase I-phase II projective algorithm for linear programming, *Mathematical Programming* Vol. 43, (1989), pp. 425-453.

[7] K. M. Anstreicher : A combined phase I-phase II scaled potential algorithm for linear programming, *Mathematical Programming* Vol. 52, (1991), pp. 429-439.

[8] E. R. Barnes : A variation on Karmarkar's algorithm for solving linear programming problems. *Mathematical Programming* Vol. 36 (1986), pp. 174-182.

[9] D. A. Bayer and J. C. Lagarias : The nonlinear geometry of linear programming: I. affine and projective trajectories. *Transactions of the American Mathematical Society,* Vol. 314, No. 2 (1989), pp. 499-526.

[10] Yun-Chan Cheng, David J. Houck, Jun-Min Liu, Marc S. Meketon, Lev Slutsman, Robert J. Vanderbei and Pyng Wang : The AT&T KORBX system. *AT&T Technical Journal,* Vol. 68, No. 3, pp. 7-19.

[11] I. I. Dikin : Iterative solution of problems of linear and quadratic programming. *Soviet Mathematics Doklady,* Vol. 8 (1967), pp. 674-675.

[12] I. I. Dikin : About the convergence of an iterative process (in Russian). *Upravlyaemye Sistemi,* Vol. 12 (1974), pp. 54-60.

[13] I. I. Dikin: The convergence of dual variables. *Technical Report,* Siberian Energy Institute, Irkutsk, Russia (1991).

[14] I. I. Dikin and V. I. Zorkaltsev : Iterative solution of mathematical programming problems (Algorithms for the method of interior points). Nauka, Novosibirsk, USSR (1980).

[15] R. M. Freund : Theoretical efficiency of a shifted barrier function algorithm for linear programming. *Linear Algebra and its Applications* Vol. 152 (1991), pp.19-41.

[16] R. M. Freund : A potential reduction algorithm with user-specified Phase I-Phase II balance, for solving a linear program from an infeasible warm start. *Mathematical Programming* Vol. 52 (1991), pp.441-466.

[17] R. M. Freund : An infeasible-Start algorithm for linear programming whose complexity depends on the distance from the starting point to the Optimal solution. *Working Paper,* Sloan School of Management, Massachusetts Institute of Technology, USA (1993).

[18] D. Gay : Stopping tests that compute optimal solutions for interior-point linear programming algorithms. Numerical Analysis Manuscript 89-11, AT&T Bell Laboratories, NJ, USA (1989).

[19] C. C. Gonzaga : Conical projection algorithms for linear programming. *Mathematical Programming,* Vol. 43 (1989), pp. 151-173.

[20] C. C. Gonzaga : Convergence of the large step primal affine-scaling algorithm for primal non-degenerate linear programs. *Technical Report,* Department of Systems Engineering and Computer Sciences, COPPE-Federal University of Rio de Janeiro, Brazil (1990).

[21] O. Güler, D. den Hertog, C. Roos, T. Terlaky and T. Tsuchiya : Degeneracy in interior point methods for linear programming. *Technical Report,* Faculty of Mathematics and Computer Science, Delft University of Technology, Delft, Netherlands (1991).

[22] L. A. Hall and R. J. Vanderbei : Two-thirds is sharp for affine scaling, *Operations Research Letters* Vol. 13 (1993), pp. 197-201.

[23] D. den Hertog : Interior point approach to linear, quadratic and convex programming. Ph.D thesis, Faculty of Mathematics and Computer Science, Delft University of Technology, Delft, Netherlands, (1992).

[24] D. den Hertog and C. Roos : A survey of search directions in interior point methods for linear programming. *Mathematical Programming* Vol. 52 (1991) 481-509.

[25] N. Karmarkar : A new polynomial-time algorithm for linear programming. *Combinatorica,* Vol. 4, No. 4 (1984), pp. 373-395.

[26] N. Karmarkar and K. Ramakrishnan : Computational results of an interior point algorithm for large scale linear programming. *Mathematical Programming* Vol. 52 (1991), pp.555-586.

[27] M. Kojima, N. Megiddo and S. Mizuno : A primal-dual infeasible-interior-point algorithm for linear programming. *Mathematical Programming* Vol. 61 (1993) pp. 263-280.

[28] M. Kojima, N. Megiddo, T. Noma and A. Yoshise : A unified approach to interior point algorithms for linear complementarity problems. *Lecture Notes in Computer Science* 538.

[29] M. Kojima, S. Mizuno and A. Yoshise : A primal-dual interior point algorithm for linear programming. *Progress in Mathematical Programming*, N. Megiddo (Ed.), Springer Verlag, New York, USA (1989), pp. 29-47.

[30] M. Kojima, S. Mizuno and A. Yoshise : A polynomial-time algorithm for a class of linear Complementarity Problem. *Mathematical Programming*, Vol. 44 (1989), pp. 1-26.

[31] I. Lustig : Feasibility issues in an interior point method for linear programming. *Technical Report,* SOR 88-9, Princeton University, USA (1988).

[32] W. F. Mascarenhas : The affine scaling algorithm fails for $\lambda = 0.999$. *Technical Report,* Universidade Estadual de Campinas, Campinas S.S., Brazil (1993).

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

[34] C. McDiarmid : On the improvement per iteration in Karmarkar's algorithm for linear programming. *Mathematical Programming*, Vol. 46 (1990), pp. 299-320.

[35] R. D. C. Monteiro, T. Tsuchiya and Y. Wang : A simplified global convergence proof of affine scaling algorithm. *Technical Report,* University of Arizona, Tucson, USA (1992).

[36] R. D. C. Monteiro and T. Tsuchiya : Limiting behavior of the derivatives of certain trajectories associated with a monotone horizontal linear complementarity problem. *Technical Report,* University of Arizona, Tucson, USA (1992).

[37] N. Megiddo : Pathways to the optimal set in linear programming. *Progress in Mathematical Programming*, N. Megiddo (Ed.), Springer Verlag, New York, USA (1989), pp. 131-158.

[38] N. Megiddo and M. Shub : Boundary behavior of interior point algorithms for linear programming. *Mathematics of Operations Research*, Vol. 14, No. 1 (1989), pp. 97-146.

[39] S. Mehrotra : Implementations of affine scaling methods: Approximate solutions of system of linear equations using preconditioned conjugate gradient methods. *Technical Report,* Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208, USA (1989).

[40] R. Monterio and I. Adler : Interior path-following primal-dual algorithms, Part I: linear programming. *Mathematical Programming*, Vol. 44 (1989), pp. 27-41.

[41] R. Monterio, I. Adler and M. Resende : A polynomial-time primal-dual affine scaling algorithms for linear and convex quadratic programming and its power series extension. *Mathematics of Operations Research*, Vol. 15 (1990), pp. 191-214.

[42] C. L. Monma and A. J. Morton : Computational experience with a dual affine variant of Karmarkar's method for linear programming. *Operations Research Letters*, Vol. 6 (1987), pp. 261-267.

[43] M. Muramatsu and T. Tsuchiya : A convergence analysis of a long-step variant of the projective scaling algorithm. *Research Memorandum,* No.454, The Institute of Statistical Mathematics, Tokyo, Japan (1992).

[44] M. Muramatsu and T. Tsuchiya : An affine scaling method with an infeasible starting point. *Research Memorandum,* No.490, The Institute of Statistical Mathematics, Tokyo, Japan (1993).

[45] M. G. C. Resende and G. Veiga : Computational study of two implementations of the dual affine scaling algorithm. *Technical Report,* AT&T Bell Laboratories, Murray Hill, NJ, USA, 1990.

[46] M. G. C. Resende and G. Veiga : An implementation of the dual affine scaling algorithm for minimum cost flow on bipartite uncapaciated networks. *SIAM Journal on Optimization* Vol. 3, pp.516-537.

[47] M. G. C. Resende and G. Veiga : An efficient implementation of a network interior point method, *Technical Report,* AT&T Bell Laboratories, Murray Hill, NJ, USA (1992). *To appear in DIMACS Series in Discrete Mathematics and Theoretical Computer Science.*

[48] R. T. Rockafellar : Convex analysis. Princeton University Press, Princeton, N.J. USA (1970).

[49] R. Saigal : A simple proof of primal affine scaling method. *Technical Report,* University of Michigan, Ann Arbor, USA (1992).

[50] R. Saigal : An efficient variant of the affine scaling method. *Technical Report,* University of Michigan, Ann Arbor, USA (1993).

[51] A. Schrijver : Theory of linear and integer programming. Wiley international series in discrete mathematics and optimization, USA (1986).

[52] J. Sun : A convergence proof of an affine-scaling algorithm for convex quadratic programming without nondegeneracy assumptions. *Mathematical Programming* Vol. 60 (1993), pp.69-79.

[53] K. Tanabe : Feasibility-Improving Gradient-Acute-Projection methods: A unified approach to nonlinear programming. *Lecture Notes in Numerical Application and Analysis,* Vol.3, pp.57-76 (1981).

[54] M. J. Todd : On Anstreicher's combined phase I-phase II projective algorithm for linear programming. *Mathematical Programming,* Vol.55 (1992), pp.1-15.

[55] M. J. Todd : Combining phase I and phase II in a potential reduction algorithm for linear programming. *Technical Report* No.907, Cornell University, Ithaca, USA (1989). *To appear in SIAM Journal on Optimization.*

[56] M. J. Todd and B. P. Burrell: An extension of Karmarkar's algorithm for linear programming using dual variables. *Technical Report,* Cornell University, Ithaca, New York, USA (1985).

[57] P. Tseng and Z. Q. Luo : On the convergence of the affine scaling algorithm. *Mathematical Programming,* Vol. 56, (1992), pp.301-319.

[58] T. Tsuchiya : Global convergence property of the affine scaling method for primal degenerate linear programming problems. *Mathematics of Operations Research,* Vol. 17, No. 3 (1992), pp. 527-557.

[59] T. Tsuchiya : Global convergence of the affine scaling method for degenerate linear programming problems. *Mathematical Programming,* Vol. 52 (1992), pp. 377-404.

[60] T. Tsuchiya : Global convergence of the affine scaling algorithm for the primal degenerate strictly convex quadratic programming problems. *Research Memorandum,* No.417, The Institute of Statistical Mathematics, Tokyo, Japan (1991).

[61] T. Tsuchiya and M. Muramatsu : Global convergence of a long-step affine scaling algorithm for degenerate linear programming problems. *Research Memorandum,*

No. 423, The Institute of Statistical Mathematics, Tokyo, Japan, (1992). *To appear in SIAM Journal on Optimization.*

[62] R. J. Vanderbei, M. S. Meketon and B. A. Freedman : A modification of Karmarkar's linear programming algorithm. *Algorithmica*, Vol. 1 (1986), pp. 395-407.

[63] R. J. Vanderbei and J. C. Lagarias : I. I. Dikin's convergence result for the affine-scaling algorithm. *Contemporary Mathematics*, Vol. 114 (1988), pp. 109-119.