Theoretical Convergence of Large-Step Primal-Dual Interior Point Algorithms for Linear Programming[1]

Masakazu Kojima[†], Nimrod Megiddo[#] and Shinji Mizuno[*]

November 1990
Revised September 1991

Abstract. This paper proposes two sets of rules, Rule G and Rule P, for controlling step lengths in a generic primal-dual interior point method for solving the linear programming problem in standard form and its dual. Theoretically, Rule G ensures the global convergence, while Rule P, which is a special case of Rule G, ensures the $O(nL)$ iteration polynomial-time computational complexity. Both rules depend only on the lengths of the steps from the current iterates in the primal and dual spaces to the respective boundaries of the primal and dual feasible regions. They rely neither on neighborhoods of the central trajectory nor on potential function. These rules allow large steps without performing any line search. Rule G is especially flexible enough for implementation in practically efficient primal-dual interior point algorithms.

Key words: Primal-Dual Interior Point Algorithm, Linear Program, Large Step, Global Convergence, Polynomial-Time Convergence

Abbreviated Title: Large-Step Primal-Dual Interior Point Algorithms

† Department of Information Sciences, Tokyo Institute of Technology
2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan

\# IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120-6099, and School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel

* The Institute of Statistical Mathematics
4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan

[1]Part of this research was done when M. Kojima and S. Mizuno visited at the IBM Almaden Research Center. Partial support from the Office of Naval Research under Contracts N00014-87-C-0820 and N00014-91-C-0026 is acknowledged.

1. Introduction

Let $\boldsymbol{A} \in R^{m\times n}$, $\boldsymbol{b} \in R^m$ and $\boldsymbol{c} \in R^n$. Consider the primal-dual pair of linear programming problems:

$$
\begin{array}{lll}
\text{(P)} & \text{Minimize} & \boldsymbol{c}^T\boldsymbol{x} \\
 & \text{subject to} & \boldsymbol{x} \in P = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geq \boldsymbol{0}\}. \\
\text{(D)} & \text{Maximize} & \boldsymbol{b}^T\boldsymbol{y} \\
 & \text{subject to} & (\boldsymbol{y}, \boldsymbol{z}) \in D = \{(\boldsymbol{y}, \boldsymbol{z}) : \boldsymbol{A}^T\boldsymbol{y} + \boldsymbol{z} = \boldsymbol{c},\ \boldsymbol{z} \geq \boldsymbol{0}\}.
\end{array}
$$

Define

$$
\begin{aligned}
& P_{++} = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} > \boldsymbol{0}\}, \\
& D_{++} = \{(\boldsymbol{y}, \boldsymbol{z}) : \boldsymbol{A}^T\boldsymbol{y} + \boldsymbol{z} = \boldsymbol{c},\ \boldsymbol{z} > \boldsymbol{0}\}, \\
& S_{++} = P_{++} \times D_{++} = \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) : \boldsymbol{x} \in P_{++},\ (\boldsymbol{y}, \boldsymbol{z}) \in D_{++}\}.
\end{aligned}
$$

We assume throughout that the rank of the matrix $\boldsymbol{A}$ is m.

This paper studies a class of primal-dual interior point algorithms which originated from a fundamental analysis by Megiddo [18] on the central trajectory for the problems (P) and (D). The central trajectory is defined as the set $S_{cen} = \{(\boldsymbol{x}(\mu), \boldsymbol{y}(\mu), \boldsymbol{z}(\mu)) : \mu > 0\}$ of solutions $(\boldsymbol{x}(\mu), \boldsymbol{y}(\mu), \boldsymbol{z}(\mu))$ to the system of equations with a parameter $\mu > 0$:

$$
\boldsymbol{X}\boldsymbol{z} = \mu\boldsymbol{e},\ \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \ \boldsymbol{A}^T\boldsymbol{y} + \boldsymbol{z} = \boldsymbol{c},\ \ \boldsymbol{x} \geq \boldsymbol{0}\ \text{ and } \boldsymbol{z} \geq \boldsymbol{0}. \tag{1}
$$

Here $\boldsymbol{X} = \mathrm{diag}(\boldsymbol{x}) \in R^{n\times n}$ denotes a diagonal matrix with the coordinates of a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$, and $\boldsymbol{e} = (1, \ldots, 1)^T \in R^n$. By definition, $(\boldsymbol{x}(\mu), \boldsymbol{y}(\mu), \boldsymbol{z}(\mu)) \in S_{++}$ for every $\mu > 0$. It was shown by Megiddo [18] that, as the parameter $\mu > 0$ tends to zero, $\boldsymbol{x}(\mu) \in P_{++}$ and $(\boldsymbol{y}(\mu), \boldsymbol{z}(\mu)) \in D_{++}$ converge to optimal solutions of (P) and (D), respectively. We also refer to the book [7] by Fiacco and McCormick, who described the trajectory in terms of a logarithmic penalty function, and discussed some properties of the trajectory. Thus, optimal solutions of (P) and (D) can be approximated by tracing the trajectory S_{cen} until the parameter μ becomes sufficiently small. The first polynomial-time algorithm based on this idea was given by Kojima, Mizuno and Yoshise [11].

We describe a generic primal-dual interior point method (abbreviated to the GPD method), which provides a general framework for many existing primal-dual interior point algorithms [4; 9; 11; 12; 13; 16; 17; 21; 22; 23; 32; 33]. The GPD method generates a sequence $\{(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k)\} \subset S_{++}$. Assuming we have obtained the kth iterate $(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k)$, we will show how the method computes the $(k+1)$th iterate $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1}) \in S_{++}$. Let $f^k = (\boldsymbol{x}^k)^T\boldsymbol{z}^k/n$. If $(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k)$ lies on the central trajectory S_{cen}, then $x_i^k z_i^k = \mu$ $(i = 1, 2, \ldots, n)$ for some μ; hence $f^k = \mu$. We can easily verify that $nf^k = (\boldsymbol{x}^k)^T\boldsymbol{z}^k$ coincides with the duality gap $\boldsymbol{c}^T\boldsymbol{x}^k - \boldsymbol{b}^T\boldsymbol{y}^k$; hence

$$
f^k = \frac{(\boldsymbol{x}^k)^T\boldsymbol{z}^k}{n} = \frac{\boldsymbol{c}^T\boldsymbol{x}^k - \boldsymbol{b}^T\boldsymbol{y}^k}{n}. \tag{2}
$$

We now consider the Newton direction $(\boldsymbol{\Delta x}, \boldsymbol{\Delta y}, \boldsymbol{\Delta z})$ at the current iterate $(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k) \in S_{++}$ for approximating a point $(\boldsymbol{x}(\mu), \boldsymbol{y}(\mu), \boldsymbol{z}(\mu))$ with $\mu = \beta f^k$ on the central trajectory

S_{cen}, where $\beta \in [0,1]$ denotes a parameter to be specified later. More precisely, the direction $(\boldsymbol{\Delta x}, \boldsymbol{\Delta y}, \boldsymbol{\Delta z})$ is given as a unique solution of the Newton equation

$$\left.\begin{array}{lcl} \boldsymbol{Z}^k\boldsymbol{\Delta x} + \boldsymbol{X}^k\boldsymbol{\Delta z} & = & \beta f^k\boldsymbol{e} - \boldsymbol{X}^k\boldsymbol{z}^k, \\ \boldsymbol{A}\boldsymbol{\Delta x} & = & \boldsymbol{0}, \\ \boldsymbol{A}^T\boldsymbol{\Delta y} + \boldsymbol{\Delta z} & = & \boldsymbol{0}. \end{array}\right\} \tag{3}$$

Finally, we choose step lengths α_p and α_d to generate a new iterate $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1})$ such that

$$\left.\begin{array}{lcl} \boldsymbol{x}^{k+1} & = & \boldsymbol{x}^k + \alpha_p\boldsymbol{\Delta x} \in P_{++}, \\ (\boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1}) & = & (\boldsymbol{y}^k, \boldsymbol{z}^k) + \alpha_d(\boldsymbol{\Delta y}, \boldsymbol{\Delta z}) \in D_{++}. \end{array}\right\} \tag{4}$$

The GPD method depends on three parameters: a search direction parameter β, a primal step length α_p and a dual step length α_d. If we choose an initial solution $(\boldsymbol{x}^0, \boldsymbol{y}^0, \boldsymbol{z}^0) \in S_{++}$ and assign appropriate values to these parameters at each iteration of the GPD method, we obtain a particular primal-dual interior point algorithm, which we abbreviate as a PD algorithm.

The parameter β determines a target point $(\boldsymbol{x}(\beta f^k), \boldsymbol{y}(\beta f^k), \boldsymbol{z}(\beta f^k))$ on the central trajectory S_{cen} which we want to approximate by the new iterate $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1}) \in S_{++}$. When we take $\beta = 1$, the target point $(\boldsymbol{x}(f^k), \boldsymbol{y}(f^k), \boldsymbol{z}(f^k)) \in S_{cen}$ minimizes the Euclidean distance $\|\boldsymbol{X}\boldsymbol{z} - \boldsymbol{X}^k\boldsymbol{z}^k\|$ from the current iterate $(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k) \in S_{++}$ to points $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ on S_{cen}. Thus, the resulting Newton direction $(\boldsymbol{\Delta x}, \boldsymbol{\Delta y}, \boldsymbol{\Delta z}) = (\boldsymbol{\Delta x}^c, \boldsymbol{\Delta y}^c, \boldsymbol{\Delta z}^c)$ may be regarded as a "centering" direction. On the other hand, when we take $\beta = 0$, the system (1) with $\mu = \beta f^k = 0$ turns out to be the necessary and sufficient Karush-Kuhn-Tucker optimality condition for the problems (P) and (D):

$$\boldsymbol{X}\boldsymbol{z} = \boldsymbol{0},\ \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \ \boldsymbol{A}^T\boldsymbol{y} + \boldsymbol{z} = \boldsymbol{c},\ \ \boldsymbol{x} \geq \boldsymbol{0}\ \text{ and }\ \boldsymbol{z} \geq \boldsymbol{0}.$$

Hence the Newton direction $(\boldsymbol{\Delta x}, \boldsymbol{\Delta y}, \boldsymbol{\Delta z}) = (\boldsymbol{\Delta x}^a, \boldsymbol{\Delta y}^a, \boldsymbol{\Delta z}^a)$ from the current point $(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k)$ aims at optimal solutions $\boldsymbol{x}$ of (P) and $(\boldsymbol{y}, \boldsymbol{z})$ of (D). We call the direction $(\boldsymbol{\Delta x}^a, \boldsymbol{\Delta y}^a, \boldsymbol{\Delta z}^a)$ an "affine scaling direction" [15; 25] since, as in the primal affine scaling algorithm [3; 2; 30], it does not involve any centering direction. In general, each direction $(\boldsymbol{\Delta x}, \boldsymbol{\Delta y}, \boldsymbol{\Delta z})$ with a $\beta \in [0,1]$ can be represented as a convex combination of the centering direction $(\boldsymbol{\Delta x}^c, \boldsymbol{\Delta y}^c, \boldsymbol{\Delta z}^c)$ and the affine scaling direction $(\boldsymbol{\Delta x}^a, \boldsymbol{\Delta y}^a, \boldsymbol{\Delta z}^a)$ such that

$$(\boldsymbol{\Delta x}, \boldsymbol{\Delta y}, \boldsymbol{\Delta z}) = (1-\beta)(\boldsymbol{\Delta x}^a, \boldsymbol{\Delta y}^a, \boldsymbol{\Delta z}^a) + \beta(\boldsymbol{\Delta x}^c, \boldsymbol{\Delta y}^c, \boldsymbol{\Delta z}^c).$$

The parameters α_p and α_d determine step lengths in the primal and dual spaces, respectively. Kojima, Mizuno and Yoshise [11] showed that if the same step length $\alpha_p = \alpha_d = \alpha$ is chosen in the primal and dual spaces, then the new duality gap $(\boldsymbol{x}^{k+1})^T\boldsymbol{z}^{k+1} = \boldsymbol{c}^T\boldsymbol{x}^{k+1} - \boldsymbol{b}^T\boldsymbol{y}^{k+1} = \boldsymbol{c}^T(\boldsymbol{x}^k + \alpha\boldsymbol{\Delta x}) - \boldsymbol{b}^T(\boldsymbol{y}^k + \boldsymbol{\Delta y})$ satisfies

$$\boldsymbol{c}^T(\boldsymbol{x}^k + \alpha\boldsymbol{\Delta x}) - \boldsymbol{b}^T(\boldsymbol{y}^k + \alpha\boldsymbol{\Delta y}) = (1 - \alpha(1-\beta))(\boldsymbol{c}^T\boldsymbol{x}^k - \boldsymbol{b}^T\boldsymbol{y}^k). \tag{5}$$

See Lemma 3.1. It follows from this equality that

(i) β has to be less than or equal to 1 in order for the duality gap not to increase,
(ii) the smaller β and the larger α, the larger the reduction in the duality gap.

Theoretically, we can choose any $\beta \in [0, 1]$. On the other hand, in order to keep the new iterate $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1})$ in S_{++}, we need to take the step lengths α_p and α_d smaller than

$$\hat{\alpha}_p = \max\{\alpha : \boldsymbol{x}^k + \alpha\boldsymbol{\Delta x} \geq 0\} \text{ and } \hat{\alpha}_d = \max\{\alpha : \boldsymbol{z}^k + \alpha\boldsymbol{\Delta z} \geq 0\},$$

respectively. Hence, if we choose $\alpha_p = \alpha_d = \alpha$, it is bounded from above by

$$\hat{\alpha} = \min\{\hat{\alpha}_p,\ \hat{\alpha}_d\} = \max\{\alpha : \boldsymbol{x}^k + \alpha\boldsymbol{\Delta x} \geq 0,\ \boldsymbol{z}^k + \alpha\boldsymbol{\Delta z} \geq 0\}.$$

The bounds $\hat{\alpha}_p$, $\hat{\alpha}_d$ and $\hat{\alpha}$ depend on the location of the current iterate $(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k)$ and the value of the search direction parameter β. It is known that we can guarantee a large $\hat{\alpha}$ when the current point $(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k) \in S_{++}$ is not too far from the central trajectory S_{cen}. See Lemma 3.3. In fact, many of the existing PD algorithms [4; 9; 11; 12; 22; 21; 23] generate a sequence $\{(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k)\}$ in a prescribed horn neighborhood of the central trajectory S_{cen}, so that the bound $\hat{\alpha}$ remains larger than some positive constant.

Kojima, Mizuno and Yoshise [11] used $\beta = 1/2$ and $\alpha_p = \alpha_d = \alpha$ to generate a sequence $\{(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^p)\} \subset S_{++}$ which eventually moves into a prescribed neighborhood of the central trajectory S_{cen}. The algorithm runs in $O(nL)$ iterations. In other papers, Kojima Mizuno and Yoshise [12] and Monteiro and Adler [24] improved the complexity $O(nL)$ to $O(\sqrt{n}L)$. Their algorithms assign in advance small neighborhoods to the central trajectory S_{cen}, and generate sequences $\{(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k)\} \in S_{++}$ in the neighborhood by taking artificial initial solutions $(\boldsymbol{x}^0, \boldsymbol{y}^0, \boldsymbol{z}^0)$ in the neighborhood and the parameters $\beta = 1 - \delta/\sqrt{n}$ and $\alpha_p = \alpha_d = 1$ for some positive number $\delta \leq 1$. Their theoretical improvements, however, do not lead directly to improvements in the practical efficiency of PD algorithms. In fact, from (5) we see that the duality gap reduces at least linearly with a ratio of $(1 - \delta/\sqrt{n})$ in every iteration. Hence, the convergence of the duality gap to zero is too slow in practice when n is large.

In view of the above, a smaller search direction parameter β seems necessary to increase the efficiency of the GPD method. Lustig [14] discussed a region in the space of the search direction and step length parameters in which the GPD method converges globally. Mizuno, Todd and Ye [21] proposed an $O(nL)$ iteration PD algorithm where they took $\beta = \beta'$ in every iteration with an arbitrary fixed $\beta' \in (0, 1)$ and a larger neighborhood of the central trajectory S_{cen} than the ones used in [11; 12; 23]. If β' were of the form n^{-p} for some positive number p, we could prove that their PD algorithm would require $O(n^{p+1}L)$ total iterations.

In all the PD algorithms mentioned so far, the same step length α is chosen in the primal and dual spaces, so that the new iterate $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1}) \in S_{++}$ remains in a certain horn neighborhood of the central trajectory S_{cen}. The notion of a neighborhood plays a key role in gaining sufficient reduction in the duality gap at each iteration to ensure the polynomial-time convergence. There has been another development in PD algorithms, namely, an $O(\sqrt{n}L)$ iteration PD potential reduction algorithm given by Kojima, Mizuno and Yoshise [13]. They have taken a search direction parameter $\beta = n/(n + \sqrt{n})$ and a step length α such that in each iteration there is at least a constant reduction in the primal-dual potential function of Todd and Ye [29], rather than the duality gap. Kojima, Megiddo, Noma and Yoshise [9] generalized the PD potential reduction algorithm in a unified way to a wider class of primal-dual potential reduction algorithms including a globally convergent affine scaling PD algorithm. Ye [32; 33] investigated the range of the search direction parameter β which guarantees the polynomial-time convergence of the PD potential reduction algorithm.

The introduction of the potential function in the GPD method has opened up the possibilities of taking a larger step length α because the new iterate $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1}) \in S_{++}$ is not required anymore to be in a given neighborhood of the central trajectory S_{cen}. Ye [32] discussed the use of different step lengths in the primal and dual spaces for primal-dual potential reduction algorithms. Theoretically, however, it is not clear how large a step length α we can take even if we perform a line search along a search direction to gain a big reduction in the potential function.

McShane, Monma, and Shanno [17] proposed taking $\beta = 1/n$ and different step lengths α_p and α_d in the primal and dual spaces such that

$$\alpha_p = \theta\hat{\alpha}_p \text{ and } \alpha_d = \theta\hat{\alpha}_d \tag{6}$$

for $\theta = 0.9995$. They reported that the GPD method using this choice of the parameters solved the NETLIB set of test problems very efficiently. See also [16].

The authors feel that there still remain the following differences between the theoretical PD algorithms [4; 9; 11; 12; 13; 22; 21; 23] which enjoy the global and/or polynomial-time convergence and the practically efficient implementations [16; 17] of PD algorithms:

(i) Values of the search direction parameter β in the practically efficient implementations are smaller than those in the theoretical algorithms.
(ii) Most of the theoretical algorithms use the same step length α in the primal and dual spaces, while the practical implementations take different step lengths α_p in the primal space and α_d in the dual space given by (6).
(iii) The different step lengths α_p and α_d are usually much larger than the common step length α used in the theoretical algorithms.
(iv) The PD potential reduction algorithm [9; 13] requires a line search to get a larger reduction in the potential function.
(v) There has been no proof for the global convergence of the practical implementation [16; 17].

The goal of the present paper is to fill these gaps. We propose two sets of rules for controlling the parameters β, α_p and α_d in the GPD method. The first, Rule G, ensures global convergence (Theorem 3.2), while the second, Rule P, ensures polynomial-time complexity (Theorem 4.1). Both rules depend only on the step lengths $\hat{\alpha}_p$ and $\hat{\alpha}_d$ from the current iterates to the boundaries of the primal and dual feasible regions, respectively. They rely neither on any neighborhood of the central trajectory S_{cen} nor on the potential function. These rules allow taking large steps without performing any line search.

Under Rule G, we choose constants $\bar{\beta}$, $\bar{\omega}, \bar{\theta}$, θ^* and α^* in advance so that

$$0 \leq \bar{\beta} < 1, \ 0 < \bar{\omega} \leq 1, \ 0 < \bar{\theta} \leq \theta^* < 1 \ \text{ and } 0 < \alpha^*. \tag{7}$$

These constants can depend arbitrarily on n. At each iteration we choose a search direction parameter $\beta \in [0, \bar{\beta}]$. We have two independent conditions for choosing step length parameters α_p and α_d. One is

$$\left. \begin{array}{l} 0 \leq \alpha_p \leq \theta^* \hat{\alpha}_p, \ 0 \leq \alpha_d \leq \theta^* \hat{\alpha}_d, \\[2ex] \dfrac{\boldsymbol{c}^T \boldsymbol{x}^{k+1} - \boldsymbol{b}^T \boldsymbol{y}^{k+1}}{\boldsymbol{c}^T \boldsymbol{x}^k - \boldsymbol{b}^T \boldsymbol{y}^k} = \dfrac{\boldsymbol{c}^T(\boldsymbol{x}^k + \alpha_p \boldsymbol{\Delta x}) - \boldsymbol{b}^T(\boldsymbol{y}^k + \alpha_d \boldsymbol{\Delta y})}{\boldsymbol{c}^T \boldsymbol{x}^k - \boldsymbol{b}^T \boldsymbol{y}^k} \leq 1 - \bar{\omega}, \end{array} \right\} \tag{8}$$

and the other is

$$\left. \begin{array}{l} \bar{\theta} \leq \theta \leq \theta^*, \\ \alpha_p = \alpha_d = \alpha = \begin{cases} \theta \hat{\alpha} & \text{if } \hat{\alpha} \geq \alpha^*, \\ \theta \hat{\alpha}^2 / \alpha^* & \text{otherwise.} \end{cases} \end{array} \right\} \tag{9}$$

In a practical implementation of the GPD method, (8) should be preferred over (9). We can easily check whether there exist some α_p and α_d satisfying (8) by calculating a pair (α_p, α_d) which minimizes the duality gap

$$\boldsymbol{c}^T(\boldsymbol{x}^k + \alpha_p \boldsymbol{\Delta x}) - \boldsymbol{b}^T(\boldsymbol{y}^k + \alpha_d \boldsymbol{\Delta y}) \tag{10}$$

subject to the constraints

$$0 \leq \alpha_p \leq \theta^* \hat{\alpha}_p \ \text{ and } 0 \leq \alpha_d \leq \theta^* \hat{\alpha}_d.$$

We take the same step length $\alpha_p = \alpha_d = \alpha$ determined by (9) only when the minimizer (α_p, α_d) does not satisfy the last inequality of (8). Theoretically, however, the global convergence of the GPD method using Rule G holds even if we always use the same step length α. It should be noted that the same step length α always satisfies $\alpha < \hat{\alpha}$; hence, the resulting new iterate $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1})$ lies in S_{++}.

As a simple illustrative example of Rule G, we may take

$$\beta = \bar{\beta} = \frac{1}{n}, \ \bar{\omega} = 10^{-4}, \ \theta = \bar{\theta} = \theta^* = 0.9995 \ \text{ and } \alpha^* = 10^{-4}.$$

In this case the conditions (8) and (9) turn out to be

$$\left. \begin{array}{l} \alpha_p = 0.9995 \hat{\alpha}_p, \quad \alpha_d = 0.9995 \hat{\alpha}_d, \\[2ex] \dfrac{\boldsymbol{c}^T \boldsymbol{x}^{k+1} - \boldsymbol{b}^T \boldsymbol{y}^{k+1}}{\boldsymbol{c}^T \boldsymbol{x}^k - \boldsymbol{b}^T \boldsymbol{y}^k} = \dfrac{\boldsymbol{c}^T(\boldsymbol{x}^k + \alpha_p \boldsymbol{\Delta x}) - \boldsymbol{b}^T(\boldsymbol{y}^k + \alpha_d \boldsymbol{\Delta y})}{\boldsymbol{c}^T \boldsymbol{x}^k - \boldsymbol{b}^T \boldsymbol{y}^k} \leq 0.9999, \end{array} \right\} \tag{11}$$

and

$$\alpha_p = \alpha_d = \alpha = \begin{cases} 0.9995 \hat{\alpha} & \text{if } \hat{\alpha} \geq 10^{-4}, \\ 9995 \hat{\alpha}^2 & \text{otherwise,} \end{cases} \tag{12}$$

respectively.

If we take the same step length $\alpha_p = \alpha_d = \alpha$ in the primal and dual spaces, we know from (5) that the duality gap (10) decreases as the step length α increases. If, however, we take different step lengths α_p and α_d, then the duality gap (10) does not necessarily decrease with either α_p or α_d. We show in the Appendix that the duality gap may increase with α_p (or α_d). Furthermore, the step length $\alpha_p = \theta\hat{\alpha}_p$ (or $\alpha_d = \theta\hat{\alpha}_d$), used by McShane, Monma and Shanno [17] in the implementation of the GPD method, is not always well-defined because $\hat{\alpha}_p$ (or $\hat{\alpha}_d$) can be infinite. Therefore, the last inequality of the condition (8) (or (11)) works as a reasonable safeguard against such a difficulty.

The condition (8) in Rule G is moderate and flexible. One can expect that there exist step lengths α_p and α_d satisfying (8) whenever we take $\bar{\omega}$ sufficiently small and the current iterate $(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k) \in S_{++}$ is not too close to the boundary of S_{++}. In fact, we know from the definition of $\hat{\alpha} = \min\{\hat{\alpha}_p, \hat{\alpha}_d\}$ and (5) that if we take $\alpha = \theta^*\hat{\alpha}$ then

$$0 \leq \alpha \leq \theta^*\hat{\alpha}_p,\ 0 \leq \alpha \leq \theta^*\hat{\alpha}_d,$$

$$\frac{\boldsymbol{c}^T(\boldsymbol{x}^k + \alpha\boldsymbol{\Delta x}) - \boldsymbol{b}^T(\boldsymbol{y}^k + \alpha\boldsymbol{\Delta y})}{\boldsymbol{c}^T\boldsymbol{x}^k - \boldsymbol{b}^T\boldsymbol{y}^k} \leq 1 - \theta^*\hat{\alpha}(1-\bar{\beta}).$$

Hence, if $\hat{\alpha} \geq \bar{\omega}/(\theta^*(1-\bar{\beta}))$, then the same step length $\alpha_p = \alpha_d = \alpha = \theta^*\hat{\alpha}$ satisfies (8). If, in addition, the duality gap (10) decreases in both α_p and α_d, we may take $\alpha_p = \theta^*\hat{\alpha}_p$ and $\alpha_d = \theta^*\hat{\alpha}_d$ as in (6) used in [17].

Now, suppose that some step lengths α_p and α_d do not satisfy (8) with a small $\bar{\omega}$. Then, $\hat{\alpha} = \max\{\alpha : \boldsymbol{x} + \alpha\boldsymbol{\Delta x} \geq \boldsymbol{0},\ \boldsymbol{z} + \alpha\boldsymbol{\Delta z} \geq \boldsymbol{0}\}$ must be smaller than $\bar{\omega}/(\theta^*(1-\bar{\beta}))$ because otherwise the common step length $\alpha = \theta^*\hat{\alpha}$ would satisfy the condition (8) as we have observed above. Hence, we know that the current iterate $(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k)$ lies near the boundary of S_{++}. Then, Rule G tells us to take a common step length α of the order of $\hat{\alpha}^2$. In such a case, however, it might be better in practice to try another direction $(\boldsymbol{\Delta x}, \boldsymbol{\Delta y}, \boldsymbol{\Delta z})$ with a larger β so as to move away from the boundary, although Rule G certainly ensures the global convergence.

Under Rule G we can even take $\beta = 0$ in every iteration. In this case we have a globally convergent affine scaling PD algorithm.

A remark on generalizing Rule G. In the GPD method using Rule G described above, after choosing a search direction parameter $\beta \in [0, \bar{\beta}]$, we try to find a new iterate $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1}) \in S_{++}$ "along the search directions $\boldsymbol{\Delta x}$ in the primal space and $(\boldsymbol{\Delta y}, \boldsymbol{\Delta z})$ in the dual space" so as to decrease the duality gap at least by a constant factor $1 - \bar{\omega}$. See (8). This part can be generalized significantly by eliminating the restriction "along the search directions ... in the dual space." That is, we can take a new iterate $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1})$ anywhere in S_{++} as long as the condition

$$\frac{\boldsymbol{c}^T\boldsymbol{x}^{k+1} - \boldsymbol{b}^T\boldsymbol{y}^{k+1}}{\boldsymbol{c}^T\boldsymbol{x}^k - \boldsymbol{b}^T\boldsymbol{y}^k} \leq 1 - \bar{\omega}$$

is fulfilled. The global convergence of the GPD method using Rule G (Theorem 3.2) and its proof remain valid under this generalization. This generalization makes it possible to incorporate various techniques to increase the practical efficiency of the interior point algorithms such as higher order predictor methods [1; 19; 25; 26] and multidimensional searches [5; 21]. Also, we can incorporate some techniques [8; 34; 35] developed for accelerating the local convergence of the GPD method.

We now describe Rule P. Let

$$0 < \beta^* \leq 0.5, \; 0.5 \leq \theta^* < 1 \text{ and } 0 < \alpha^* \leq 1. \tag{13}$$

We can allow β^*, θ^* and α^* to depend on n as long as $1/(1-\theta^*)$, $1/\beta^*$ and $1/\alpha^*$ are polynomial in n. At each iteration, we take parameters β, α_p and α_d such that

$$\beta^* \leq \beta \leq 0.5, \; 0.5 \leq \theta \leq \theta^*, \tag{14}$$

$$\alpha = \alpha_p = \alpha_d = \begin{cases} \theta\hat{\alpha} & \text{if } \hat{\alpha} \geq \alpha^*, \\ \theta\hat{\alpha}^2/\alpha^* & \text{otherwise.} \end{cases} \tag{15}$$

For example, Rule P allows us to take

$$\beta^* = \beta = \frac{1}{n}, \; \theta^* = \theta = 0.9995 \; \text{ and } \alpha^* = 10^{-4}.$$

Then, the common step length α is given as in (12).

Obviously, Rule P is a special case of Rule G. Compared with Rule G and (6) used by McShane, Monma and Shanno [17] in the implementation of the GPD method, Rule P is restrictive in the sense that it always requires taking a common step length in the primal and dual spaces. In Section 4 we establish the polynomial-time convergence of the GPD method using Rule P.

In [4; 9; 12; 13; 22] PD algorithms were presented for the complementarity problem, rather than pairs of primal and dual linear programs. All the results obtained there can be easily adapted to the primal-dual pair of linear programs (P) and (D). See the concluding remarks of [13]. Many interior point algorithms have been proposed which work on the primal-dual pair of problems (P) and (D) but are not covered by the GPD method. Among others, we refer to the following:

(i) An $O(n^3L)$ algorithm using a sequence [20].
(ii) An $O(\sqrt{n}L)$ iteration potential reduction algorithm [31]. See also [6].
(iii) A potential reduction algorithm for the linear complementarity problem with P_0 matrices [10].

We also mention that Tanabe's centered Newton method [27; 28] is closely related to the GPD method.

2. Notation

We summarize below the notation which we use throughout.

$(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k)$: the kth iterate of the GPD method.
$f^k = \dfrac{(\boldsymbol{x}^k)^T \boldsymbol{z}^k}{n} = \dfrac{\boldsymbol{c}^T \boldsymbol{x}^k - \boldsymbol{b}^T \boldsymbol{y}^k}{n}$.
$(\boldsymbol{\Delta x}, \boldsymbol{\Delta y}, \boldsymbol{\Delta z})$: the search direction at the kth iterate.
$\hat{\alpha}_p = \sup\{\alpha : \boldsymbol{x}^k + \alpha \boldsymbol{\Delta x} \geq 0\}$.
$\hat{\alpha}_d = \sup\{\alpha : \boldsymbol{z}^k + \alpha \boldsymbol{\Delta z} \geq 0\}$.
$\hat{\alpha} = \min\{\hat{\alpha}_p,\ \hat{\alpha}_d\}$.
α : a common step length in the primal and dual spaces; $0 \leq \alpha \leq \hat{\alpha}$.
α_p : a step length in the primal space; $0 \leq \alpha_p \leq \hat{\alpha}_p$.
α_d : a step length in the dual space; $0 \leq \alpha_d \leq \hat{\alpha}_d$.
$\beta \in [0, 1]$: a search direction parameter.
$\bar{\beta}$, $\bar{\omega}$, $\bar{\theta}$, θ^*, α^* β^* : constants fixed in Rules G and P. See (7) and (13).
θ : a parameter used in Rules G and P. See (9), (14) and (15).
$\pi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \min\left\{ \dfrac{x_j z_j}{\boldsymbol{x}^T \boldsymbol{z}/n} : j = 1, 2, \ldots, n \right\}$ for every $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in S_{++}$.
$\pi^k = \pi(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k)$.

In general, the superscript k stands for the kth iterate. The values of $\hat{\alpha}$, $\hat{\alpha}_p$, $\hat{\alpha}_d$, β, α, α_p, α_d and θ can vary from one iteration to another, but we usually omit their dependence on k.

3. Global convergence of the GPD method using Rule G

Throughout this section, we assume that the parameters $\bar{\beta}$, $\bar{\omega}$, $\bar{\theta}$, θ^* and α^* associated with Rule G satisfy (7). We also assume $(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k) \in S_{++}$, and that $\beta \in [0, \bar{\beta}]$. We begin by proving the equality (5) which we have used in our discussion in the Introduction. The following results were essentially due to [11].

Lemma 3.1. *Assume that we take a common step length $\alpha = \alpha_p = \alpha_d < \hat{\alpha}$. Then,*

$$\begin{aligned}
&\boldsymbol{c}^T \boldsymbol{x}^{k+1} - \boldsymbol{b}^T \boldsymbol{y}^{k+1} = (1 - \alpha(1-\beta))(\boldsymbol{c}^T \boldsymbol{x}^k - \boldsymbol{b}^T \boldsymbol{y}^k) > 0, \\
&f^{k+1} = (1 - \alpha(1-\beta)) f^k > 0, \\
&0 < 1 - \alpha(1-\beta) < 1.
\end{aligned}$$

Proof: By (4), $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1}) \in S_{++}$ and $\alpha < \hat{\alpha}$, we have

$$\begin{aligned}
0 &< \boldsymbol{c}^T \boldsymbol{x}^{k+1} - \boldsymbol{b}^T \boldsymbol{y}^{k+1} \\
&= (\boldsymbol{x}^{k+1})^T \boldsymbol{z}^{k+1} \\
&= (\boldsymbol{x}^k + \alpha \boldsymbol{\Delta x})^T (\boldsymbol{z}^k + \alpha \boldsymbol{\Delta z}) \\
&= (\boldsymbol{x}^k)^T \boldsymbol{z}^k + \alpha((\boldsymbol{z}^k)^T \boldsymbol{\Delta x} + (\boldsymbol{x}^k)^T \boldsymbol{\Delta z}) + \alpha^2 \boldsymbol{\Delta x}^T \boldsymbol{\Delta z} \\
&= (\boldsymbol{c}^T \boldsymbol{x}^k - \boldsymbol{b}^T \boldsymbol{y}) + \alpha((\boldsymbol{z}^k)^T \boldsymbol{\Delta x} + (\boldsymbol{x}^k)^T \boldsymbol{\Delta z}) + \alpha^2 \boldsymbol{\Delta x}^T \boldsymbol{\Delta z}.
\end{aligned}$$

Here $(\boldsymbol{\Delta x}, \boldsymbol{\Delta y}, \boldsymbol{\Delta z})$ is a solution of the system (3) of linear equations. Hence the second and third terms in the last line above can be rewritten as

$$\begin{aligned} \alpha((\boldsymbol{z}^k)^T \boldsymbol{\Delta x} + (\boldsymbol{x}^k)^T \boldsymbol{\Delta z}) &= \alpha \boldsymbol{e}^T(\boldsymbol{Z}^k \boldsymbol{\Delta x} + \boldsymbol{X}^k \boldsymbol{\Delta z}) \\ &= \alpha \boldsymbol{e}^T(\beta f^k \boldsymbol{e} - \boldsymbol{X}^k \boldsymbol{z}^k) \\ &= \alpha(n\beta f^k - (\boldsymbol{x}^k)^T \boldsymbol{z}^k) \\ &= -\alpha(1-\beta)(\boldsymbol{c}^T \boldsymbol{x}^k - \boldsymbol{b}^T \boldsymbol{y}^k), \end{aligned}$$

and

$$\alpha^2 \boldsymbol{\Delta x}^T \boldsymbol{\Delta z} = \alpha^2 \boldsymbol{\Delta x}^T(-\boldsymbol{A}^T \boldsymbol{\Delta y}) = -\alpha^2 \boldsymbol{\Delta y}^T \boldsymbol{A} \boldsymbol{\Delta x} = 0,$$

respectively. Thus we have shown the first relation of the lemma. The second relation follows from the first and the definition (2) of f^k. Since $\boldsymbol{c}^T \boldsymbol{x}^k - \boldsymbol{b}^T \boldsymbol{y}^k > 0$, we obtain the last inequality of the lemma. ∎

The theorem below establishes the global convergence of the GPD method using Rule G.

Theorem 3.2. *Suppose that* $(\boldsymbol{x}^0, \boldsymbol{y}^0, \boldsymbol{z}^0) \in S_{++}$. *Let* $\{(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k)\} \subset S_{++}$ *be a sequence generated by the GPD method using Rule G. Then, the duality gap* $\boldsymbol{c}^T \boldsymbol{x}^k - \boldsymbol{b}^T \boldsymbol{y}^k$ *converges to* 0 *as* k *tends to infinity.*

The remainder of this section is devoted to proving the theorem. We need to introduce a quantity $\pi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ to measure a deviation from the central path S_{cen} at each $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in S_{++}$:

$$\pi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \min \left\{ \frac{x_j z_j}{\boldsymbol{x}^T \boldsymbol{z}/n} : j = 1, 2, \ldots, n \right\} \text{ for every } (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in S_{++}. \tag{16}$$

Obviously, π is a continuous function in $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in S_{++}$. It is easily verified that

$$\begin{aligned} &0 < \pi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \leq 1 \text{ for every } (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in S_{++}, \\ &\pi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = 1 \text{ if and only if } (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in S_{cen}. \end{aligned}$$

We may say that $\pi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ decreases from 1 to zero as $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in S_{++}$ moves away from the central path S_{cen} and approaches the boundary of S_{++}. Thus, $1 - \pi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ represents a deviation from S_{cen}. For simplicity of notation, we use π^k for $\pi(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k)$.

Remark. Kojima, Mizuno and Yoshise [11] and Lustig [14] used $1/\pi$, which they denoted by π, to measure a deviation from the central path S_{cen}. See also [9] for some other quantities to measure a deviation from the central path S_{cen} and their relation to π.

It follows from the Newton equation (3), which is satisfied by $(\boldsymbol{\Delta x}, \boldsymbol{\Delta y}, \boldsymbol{\Delta z})$, that

$$\sum_{j=1}^{n} \Delta x_j \Delta z_j = \boldsymbol{\Delta x}^T \boldsymbol{\Delta z} = 0, \tag{17}$$

$$z_j^k \Delta x_j + x_j^k \Delta z_j = \beta f^k - x_j^k z_j^k \text{ for every } j = 1, 2, \ldots, n. \tag{18}$$

These inequalities will be utilized in the lemmas below. The next Lemma is a modification of Lemma 5 in [21]

Lemma 3.3.

$$\hat{\alpha}^2 \geq \min\left\{\frac{1}{4},\ \frac{2(\pi^k)^2}{(\beta^2 - 2\beta\pi^k + \pi^k)n}\right\} \geq \min\left\{\frac{1}{4},\ \frac{2(\pi^k)^2}{n}\right\}.$$

Proof: It follows from (18) that, for every $j = 1, 2, \ldots, n$,

$$(z_j^k \Delta x_j)(x_j^k \Delta z_j) \leq \left(\frac{\beta f^k - x_j^k z_j^k}{2}\right)^2,$$

which implies

$$\Delta x_j \Delta z_j \leq \frac{(\beta f^k - x_j^k z_j^k)^2}{4 x_j^k z_j^k}.$$

Using (17) and the inequality above, we have that

$$\begin{aligned} -\Delta x_i \Delta z_i &= \sum_{j \neq i} \Delta x_j \Delta z_j \\ &\leq \sum_{j \neq i} \frac{(\beta f^k - x_j^k z_j^k)^2}{4 x_j^k z_j^k} \\ &\leq \frac{1}{4} \sum_{j=1}^{n} \left(\frac{(\beta f^k)^2}{x_j^k z_j^k} - 2\beta f^k + x_j^k z_j^k \right) \\ &\leq \frac{1}{4} \left(\frac{n(\beta f^k)^2}{\pi^k f^k} - 2n\beta f^k + n f^k \right); \end{aligned}$$

hence

$$-\Delta x_i \Delta z_i \leq \frac{(\beta^2 - 2\beta\pi^k + \pi^k) n f^k}{4\pi^k}. \tag{19}$$

On the other hand, by the definition of $\hat{\alpha}$, we can find an index i such that

$$\begin{aligned} 0 &= (x_i^k + \hat{\alpha}\Delta x_i)(z_i^k + \hat{\alpha}\Delta z_i) \\ &= x_i^k z_i^k + \hat{\alpha}(x_i^k \Delta z_i + z_i^k \Delta x_i) + \hat{\alpha}^2 \Delta x_i \Delta z_i \\ &= x_i^k z_i^k + \hat{\alpha}(\beta f^k - x_i^k z_i^k) + \hat{\alpha}^2 \Delta x_i \Delta z_i \qquad \text{(by (18))} \\ &= (1 - \hat{\alpha}) x_i^k z_i^k + \hat{\alpha}\beta f^k + \hat{\alpha}^2 \Delta x_i \Delta z_i, \end{aligned}$$

If $\hat{\alpha} \leq 1/2$, the equality above implies

$$\begin{aligned} 0 &\geq (1 - \hat{\alpha}) x_i^k z_i^k + \hat{\alpha}^2 \Delta x_i \Delta z_i \\ &\geq \frac{1}{2} x_i^k z_i^k + \hat{\alpha}^2 \Delta x_i \Delta z_i \\ &\geq \frac{1}{2} \min\{x_j^k z_j^k : j = 1, 2, \ldots, n\} + \hat{\alpha}^2 \Delta x_i \Delta z_i \\ &= \frac{1}{2} \pi^k f^k + \hat{\alpha}^2 \Delta x_i \Delta z_i. \end{aligned}$$

Hence we see that

$$\hat{\alpha}^2 \geq \min\left\{\frac{1}{4},\ \frac{\pi^k f^k}{-2\Delta x_i \Delta z_i}\right\}.$$

Substituting the inequality (19) in the inequality above, we obtain the first inequality of the lemma. Since $0 < \pi^k \leq 1$ and $0 \leq \beta \leq \bar{\beta} < 1$,

$$\begin{aligned} 0 &< (\beta - \pi^k)^2 + \pi^k(1-\pi^k) \\ &= \beta^2 - 2\beta\pi^k + \pi^k \\ &\leq \beta - \beta\pi^k + \pi^k \\ &= 1 - (1-\beta)(1-\pi^k) \\ &\leq 1. \end{aligned}$$

Thus the second inequality of the lemma follows. ∎

Lemma 3.4. *Assume that* $\alpha = \alpha_p = \alpha_d < \hat{\alpha}$, *and that*

$$1 - \alpha - (1-\hat{\alpha})\left(\frac{\alpha}{\hat{\alpha}}\right)^2 \geq 0.$$

Then

$$x_j^{k+1} z_j^{k+1} \geq \left(1 - \alpha - (1-\hat{\alpha})\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)\pi^k f^k + \left(\alpha - \hat{\alpha}\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)\beta f^k$$

for every $j = 1, 2, \ldots, n$.

Proof: Let j be fixed. Then,

$$\begin{aligned} 0 &\leq (x_j^k + \hat{\alpha}\Delta x_j)(z_j^k + \hat{\alpha}\Delta z_j) \\ &= x_j^k z_j^k + \hat{\alpha}(z_j^k \Delta x_j + x_j^k \Delta z_j) + \hat{\alpha}^2 \Delta x_j \Delta z_j \\ &= x_j^k z_j^k + \hat{\alpha}(\beta f^k - x_j^k z_j^k) + \hat{\alpha}^2 \Delta x_j \Delta z_j. \qquad \text{(by (18))} \end{aligned}$$

Hence

$$\Delta x_j \Delta z_j \geq -\frac{x_j^k z_j^k + \hat{\alpha}(\beta f^k - x_j^k z_j^k)}{\hat{\alpha}^2}.$$

By the definition of π^k, we also see $x_j^k z_j^k \geq \pi^k f^k$. It follows that

$$\begin{aligned} x_j^{k+1} z_j^{k+1} &= (x_j^k + \alpha\Delta x_j)(z_j^k + \alpha\Delta z_j) \\ &= x_j^k z_j^k + \alpha(\beta f^k - x_j^k z_j^k) + \alpha^2 \Delta x_j \Delta z_j \qquad \text{(by (18))} \\ &\geq x_j^k z_j^k + \alpha(\beta f^k - x_j^k z_j^k) + \left(\frac{\alpha}{\hat{\alpha}}\right)^2\left(-x_j^k z_j^k - \hat{\alpha}(\beta f^k - x_j^k z_j^k)\right) \\ &= \left(1 - \alpha - (1-\hat{\alpha})\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right) x_j^k z_j^k + \left(\alpha - \hat{\alpha}\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)\beta f^k \\ &\geq \left(1 - \alpha - (1-\hat{\alpha})\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)\pi^k f^k + \left(\alpha - \hat{\alpha}\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)\beta f^k. \end{aligned}$$

∎

Lemma 3.5. *Define*

$$\kappa = \min\left\{1,\ \alpha^*,\ \frac{\alpha^*\sqrt{1-\bar{\beta}}}{6\theta^*}\right\} \quad \textit{and} \quad \gamma = \frac{\theta^*}{\alpha^*(1-\bar{\beta})}. \tag{20}$$

Assume that $\pi^k \leq 0.5$ *and* $\hat{\alpha} \leq \kappa$. *Take a common step length* $\alpha = \alpha_p = \alpha_d$ *according to (9). Then*

$$\pi^{k+1} \geq (1-\alpha(1-\beta))^{2\gamma}\pi^k.$$

Proof: By the assumption, $\alpha = \theta\hat{\alpha}^2/\alpha^* < \hat{\alpha} \leq \kappa \leq 1$. It follows that

$$\begin{aligned} 1-\alpha-(1-\hat{\alpha})\left(\frac{\alpha}{\hat{\alpha}}\right)^2 &\geq 1-\hat{\alpha}-(1-\hat{\alpha}) = 0, \\ \alpha - \hat{\alpha}\left(\frac{\alpha}{\hat{\alpha}}\right)^2 &= \alpha - \frac{\alpha^2}{\hat{\alpha}} \geq 0. \end{aligned} \tag{21}$$

Let j be fixed. By the inequalities above and Lemma 3.4, we see that

$$x_j^{k+1}z_j^{k+1} \;\geq\; \left(1-\alpha-(1-\hat{\alpha})\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)\pi^k f^k + \left(\alpha-\hat{\alpha}\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)\beta f^k.$$

By Lemma 3.1, $f^{k+1} = (1-\alpha(1-\beta))f^k$. Hence

$$\frac{x_j^{k+1}z_j^{k+1}}{f^{k+1}} \geq \varphi(\beta) = \frac{\psi(\beta)}{\chi(\beta)},$$

where $\psi : [0,1] \to R$ and $\chi : [0,1] \to R$ are functions such that

$$\begin{aligned} \psi(\xi) &= \left(1-\alpha-(1-\hat{\alpha})\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)\pi^k + \left(\alpha-\hat{\alpha}\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)\xi, \\ \chi(\xi) &= 1-\alpha(1-\xi). \end{aligned}$$

We now prove that $\varphi(\beta) \geq \varphi(0)$ by showing that

$$\varphi'(\xi) = \frac{\psi'(\xi)\chi(\xi)-\chi'(\xi)\psi(\xi)}{\chi(\xi)^2} \geq 0 \quad \text{for every } \xi \in [0,\beta]$$

whenever the assumptions of the lemma are satisfied. By Lemma 3.1, the denominator $\chi(\xi)^2$ is positive for every $\xi \in [0,\beta]$. Evaluating the numerator $\psi'(\xi)\chi(\xi)-\chi'(\xi)\psi(\xi)$ for each $\xi \in [0,\beta]$, we have

$$\begin{aligned} &\psi'(\xi)\chi(\xi)-\chi'(\xi)\psi(\xi) \\ = \;& \left(\alpha-\hat{\alpha}\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)(1-\alpha(1-\xi)) \\ & -\alpha\left\{\left(1-\alpha-(1-\hat{\alpha})\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)\pi^k + \left(\alpha-\hat{\alpha}\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)\xi\right\} \\ \geq \;& \left(\alpha-\hat{\alpha}\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)(1-\alpha) + \left(\alpha-\hat{\alpha}\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)\alpha\xi - \alpha\left(\pi^k+\alpha\xi\right) \end{aligned}$$

$$
\begin{aligned}
& \qquad\qquad (\text{by } 0 < \alpha < \hat\alpha \le \kappa \le 1,\ 0 < \pi^k \text{ and } 0 \le \xi) \\
\ge\ & \alpha\left(1 - \frac{\alpha}{\hat\alpha}\right)(1-\alpha) - \alpha(\pi^k + \alpha) \qquad (\text{by (21) and } 0 \le \xi < 1) \\
\ge\ & \alpha\left(1 - \frac{\alpha}{\hat\alpha} - \alpha - \pi^k - \alpha\right) \\
\ge\ & \alpha\left(\frac{1}{2} - \frac{3\alpha}{\hat\alpha}\right) \qquad (\text{since } \hat\alpha \le \kappa \le 1 \ \text{ and } \pi^k \le 0.5) \\
\ge\ & \alpha\left(\frac{1}{2} - \frac{3\theta^*\hat\alpha}{\alpha^*}\right) \qquad (\text{since } \alpha/\hat\alpha = \theta\hat\alpha/\alpha^* \le \theta^*\hat\alpha/\alpha^*) \\
\ge\ & 0. \qquad (\text{since } \hat\alpha \le \kappa \le \left(\alpha^*\sqrt{1-\bar\beta}\right)/(6\theta^*) \le \alpha^*/(6\theta^*))
\end{aligned}
$$

Thus, we have shown $\varphi'(\xi) \ge 0$ for all $\xi \in [0, \bar\beta]$. Hence

$$
\begin{aligned}
\frac{x_j^{k+1} z_j^{k+1}}{f^{k+1}} &\ge \varphi(\beta) \ge \varphi(0) \\
&= \frac{\left(1 - \alpha - (1-\hat\alpha)\left(\frac{\alpha}{\hat\alpha}\right)^2\right)\pi^k}{1-\alpha} \\
&\ge \left(1 - \left(\frac{\alpha}{\hat\alpha}\right)^2\right)\pi^k \qquad (\text{since } 1 - \alpha \ge 1 - \hat\alpha) \\
&\ge \left(1 - \frac{\theta^*\alpha}{\alpha^*}\right)\pi^k \qquad (\text{since } \alpha = \theta\hat\alpha^2/\alpha^* \le \theta^*\hat\alpha^2/\alpha^* \ ) \\
&\ge \left(1 - \frac{(1-\beta)\alpha\theta^*}{(1-\bar\beta)\alpha^*}\right)\pi^k \qquad (\text{since } 1 - \beta \ge 1 - \bar\beta) \\
&= (1 - (1-\beta)\alpha\gamma)\,\pi^k.
\end{aligned}
$$

Hence, we have shown

$$
\frac{x_j^{k+1} z_j^{k+1}}{f^{k+1}} \ge (1 - \gamma\alpha(1-\beta))\,\pi^k.
$$

This inequality holds for every $j = 1, 2, \ldots, n$. From the definition

$$
\pi^{k+1} = \pi(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1}) = \min\left\{\frac{x_j^{k+1} z_j^{k+1}}{f^{k+1}} : j = 1, 2, \ldots, n\right\}
$$

of π^{k+1}, we obtain

$$
\pi^{k+1} \ge (1 - \gamma\alpha(1-\beta))\,\pi^k.
$$

On the other hand, $0 < \alpha(1-\beta) < 1$ by Lemma 3.1. We also see

$$
\begin{aligned}
\alpha(1-\beta) &\le \alpha \qquad (\text{since } 0 \le \beta \le \bar\beta < 1) \\
&\le \frac{\theta^*\hat\alpha^2}{\alpha^*} \qquad (\text{since } \alpha = \theta\hat\alpha^2/\alpha^* \le \theta^*\hat\alpha^2/\alpha^*) \\
&\le \frac{\alpha^*(1-\bar\beta)}{36\theta^*} \qquad (\text{since } \hat\alpha \le \kappa \le \left(\alpha^*\sqrt{1-\bar\beta}\right)/(6\theta^*)) \\
&\le \frac{1}{36\gamma}. \qquad (\text{since } \alpha^*(1-\bar\beta)/\theta^* = 1/\gamma)
\end{aligned}
$$

Finally, using the inequality

$$(1-\gamma\xi) \geq (1-\xi)^{2\gamma}$$

with $0 \leq \xi = \alpha(1-\beta) \leq \min\{1, 1/(2\gamma)\}$, we obtain

$$\pi^{k+1} \geq (1-\gamma\alpha(1-\beta))\pi^k \geq (1-\alpha(1-\beta))^{2\gamma}\pi^k.$$

∎

Now we are ready to prove Theorem 3.2.

Proof of Theorem 3.2:

We need to designate the dependence of α, α_p, α_d, $\hat{\alpha}$, β and θ, which are used in Rule G, on the iteration explicitly in the proof below. So we will write α^k, α_p^k, α_d^k, $\hat{\alpha}^k$, β^k and θ^k. Since the duality gap $\boldsymbol{c}^T\boldsymbol{x}^k - \boldsymbol{b}^T\boldsymbol{y}^k$ decreases monotonically, the duality gap apparently converges to 0 if we can take different step lengths α_p^k and α_d^k satisfying (8) for infinitely many k's. Hence, we only have to deal with the case where the same step length $\alpha^k = \alpha_p^k = \alpha_d^k$ is chosen according to (9) for every $k \geq \bar{k}$ and some $\bar{k}$. Assume that $\hat{\alpha}^k \geq \kappa$ or $\pi^k \geq \bar{\pi}$ for some $\bar{\pi} \in (0, 0.5]$, where κ is defined by (20). If we denote

$$\delta = \min\left\{\frac{1}{4},\ \frac{2\bar{\pi}^2}{n},\ \kappa^2\right\} = \min\left\{\frac{1}{4},\ \frac{2\bar{\pi}^2}{n},\ (\alpha^*)^2,\ \left(\frac{\alpha^*\sqrt{1-\bar{\beta}}}{6\theta^*}\right)^2\right\},$$

we see by Lemmas 3.3 that $(\hat{\alpha}^k)^2 \geq \delta$. Hence

$$\begin{aligned}
&\frac{\boldsymbol{c}^T\boldsymbol{x}^{k+1} - \boldsymbol{b}^T\boldsymbol{y}^{k+1}}{\boldsymbol{c}^T\boldsymbol{x}^k - \boldsymbol{b}^T\boldsymbol{y}^k} \\
= {} & 1-(1-\beta^k)\alpha^k && \text{(by Lemma 3.1)} \\
\leq {} & 1-(1-\bar{\beta})\bar{\theta}\left(\min\left\{\hat{\alpha}^k, \frac{(\hat{\alpha}^k)^2}{\alpha^*}\right\}\right) && \text{(by } \beta^k \leq \bar{\beta},\ 0 < \bar{\theta} \leq \theta \text{ and (9))} \\
\leq {} & 1-(1-\bar{\beta})\bar{\theta}\left(\min\left\{\alpha^*, \frac{(\hat{\alpha}^k)^2}{\alpha^*}\right\}\right) && \text{(since } \hat{\alpha}^k \geq (\hat{\alpha}^k)^2/\alpha^* \text{ if } \alpha^* \geq \hat{\alpha}^k) \\
\leq {} & 1-(1-\bar{\beta})\bar{\theta}\left(\min\left\{\alpha^*, \frac{\delta}{\alpha^*}\right\}\right). && \text{(by } (\hat{\alpha}^k)^2 \geq \delta)
\end{aligned}$$

If the inequality above holds for infinitely many k's, the duality gap $\boldsymbol{c}^T\boldsymbol{x}^k - \boldsymbol{b}^T\boldsymbol{y}^k$ converges to 0. So we may further restrict ourselves to the case where

$$\lim_{k\to\infty} \pi^k = 0,\ \ \hat{\alpha}^k \leq \kappa \ \text{ and } \pi^k \leq 0.5 \text{ for every } k \geq \ell \text{ and some } \ell \geq \bar{k}.$$

Applying Lemma 3.5, we now obtain

$$\pi^{k+1} \geq (1-\alpha^k(1-\beta^k))^{2\gamma}\pi^k$$

for every $k \geq \ell$. It follows that

$$
\begin{aligned}
\pi^{\ell+r} &\geq \prod_{k=\ell}^{\ell+r-1} (1-\alpha^k(1-\beta^k))^{2\gamma}\pi^\ell \\
&= \left(\prod_{k=\ell}^{\ell+r-1} (1-\alpha^k(1-\beta^k))\right)^{2\gamma} \pi^\ell \\
&= \left(\prod_{k=\ell}^{\ell+r-1} \frac{\boldsymbol{c}^T\boldsymbol{x}^{k+1}-\boldsymbol{b}^T\boldsymbol{y}^{k+1}}{\boldsymbol{c}^T\boldsymbol{x}^{k}-\boldsymbol{b}^T\boldsymbol{y}^{k}}\right)^{2\gamma} \pi^\ell \qquad \text{(by Lemma 3.1)} \\
&= \left(\frac{\boldsymbol{c}^T\boldsymbol{x}^{\ell+r}-\boldsymbol{b}^T\boldsymbol{y}^{\ell+r}}{\boldsymbol{c}^T\boldsymbol{x}^{\ell}-\boldsymbol{b}^T\boldsymbol{y}^{\ell}}\right)^{2\gamma} \pi^\ell
\end{aligned}
$$

for every $r = 1, 2, \ldots$. Thus, we obtain

$$
\frac{\pi^{\ell+r}(\boldsymbol{c}^T\boldsymbol{x}^\ell-\boldsymbol{b}^T\boldsymbol{y}^\ell)^{2\gamma}}{\pi^\ell} \geq (\boldsymbol{c}^T\boldsymbol{x}^{\ell+r}-\boldsymbol{b}^T\boldsymbol{y}^{\ell+r})^{2\gamma}
$$

for every $r = 1, 2, \ldots$. Since $\lim_{r\to\infty} \pi^{\ell+r} = 0$, we can conclude from the inequality above that the duality gap $\boldsymbol{c}^T\boldsymbol{x}^k-\boldsymbol{b}^T\boldsymbol{y}^k$ converges to 0 as k tends to infinity. This completes the proof of Theorem 3.2.

4. Polynomial-time convergence of the GPD method with Rule P

Throughout this section, we assume that the parameters β^*, α^* and θ^* associated with Rule P satisfy (13).

Theorem 4.1. *Let $\epsilon > 0$. Suppose that $(\boldsymbol{x}^0, \boldsymbol{y}^0, \boldsymbol{z}^0) \in S_{++}$. Define*

$$
\begin{aligned}
\eta &= \log\left(\frac{\boldsymbol{c}^T\boldsymbol{x}^0-\boldsymbol{b}^T\boldsymbol{y}^0}{\epsilon}\right), \\
\sigma &= \min\left\{\frac{(1-\theta^*)\alpha^*\beta^*}{2},\ \pi^0\right\}, \\
\tilde{\alpha} &= \min\left\{\frac{\alpha^*}{2}, \frac{\sigma^2}{n\alpha^*}\right\}.
\end{aligned}
$$

Let $\{(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k)\} \subset S_{++}$ be a sequence generated by the GPD method using Rule P. Then

(i) $\pi^k \geq \sigma$ *for all* $k = 0, 1, 2, \ldots$,
(ii) $\boldsymbol{c}^T\boldsymbol{x}^r-\boldsymbol{b}^T\boldsymbol{y}^r \leq \epsilon$ *if* $r \geq 2\eta/\tilde{\alpha}$.

From (i), the generated sequence $\{(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{z}^k)\}$ lies in a neighborhood $\{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in S_{++} : \pi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \geq \sigma\}$ of the central trajectory S_{cen} although Rule P does not explicitly require the sequence to remain in any prescribed neighborhood of S_{cen}. If we let $\beta = 1-\sigma$, the neighborhood $\{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in S_{++} : \pi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \geq \sigma\}$ coincides with the neighborhood

$\mathcal{N}_\infty^-(\beta) = \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in S_{++} : \|\boldsymbol{X}\boldsymbol{z} - \mu\boldsymbol{e}\|_\infty^- \leq \beta\mu, \mu = \boldsymbol{x}^T\boldsymbol{z}/n\}$ introduced by Mizuno, Todd and Ye [21]. Once we know that the sequence lies in the neighborhood $\mathcal{N}_\infty^-(\beta)$, we can deduce the computational complexity of the algorithm from the results of [21]. But we will show the entire proof of Theorem 4.1.

The theorem also shows that if all the quantities $1/(1-\theta^*)$, $1/\beta^*$, $1/\alpha^*$ and $1/\pi^0$ are polynomial in n, then the GPD method with Rule P computes approximate optimal solutions $\boldsymbol{x}^r$ of (P) and $(\boldsymbol{y}^r, \boldsymbol{z}^r)$ of (D) with the duality gap not greater than ϵ, within time polynomial in n. Specifically, when they are of the order $O(1)$ we obtain the following corollary.

Corollary 4.2. *In addition to the assumption in Theorem 4.1, suppose that $1/(1-\theta^*)$, $1/\beta^*$, $1/\alpha^*$ and $1/\pi^0$ are all $O(1)$. Then, the duality gap $\boldsymbol{c}^T\boldsymbol{x}^r - \boldsymbol{b}^T\boldsymbol{y}^r$ becomes not greater than ϵ in $O(n\eta)$ iterations.*

Preparing an artificial initial point $(\boldsymbol{x}^0, \boldsymbol{y}^0, \boldsymbol{z}^0) \in S_{++}$ such that

$$\log(\boldsymbol{c}^T\boldsymbol{x}^0 - \boldsymbol{b}^T\boldsymbol{y}^0) = O(L) \text{ and } \pi^0 = \pi(\boldsymbol{x}^0, \boldsymbol{y}^0, \boldsymbol{z}^0) = O(1),$$

and setting $\epsilon > 0$ such that $\log\epsilon = -O(L)$, we can prove under the conditions of Corollary 4.2 that the GPD method using Rule P enjoys the $O(nL)$ iteration complexity, where L denotes the input size of the problem (P). But the details are omitted here. See Kojima, Mizuno, and Yoshise [12] and Monteiro and Adler [23] for such an artificial initial point.

We need the following lemma to prove Theorem 4.1.

Lemma 4.3. *Let $\sigma \in (0, (1-\theta^*)\alpha^*\beta^*/2]$ be a constant. If $\pi^k \geq \sigma$ then $\pi^{k+1} \geq \sigma$.*

Proof: By Lemma 3.1, $(1 - \hat{\alpha}(1-\beta)) \geq 0$. By assumption (14), $\beta \leq 0.5$, so $\hat{\alpha} \leq 1/(1-\beta) \leq 2$. We also know $\alpha < \hat{\alpha}$ by the assumption. Hence,

$$\begin{aligned} 1 - \alpha - (1-\hat{\alpha})\left(\frac{\alpha}{\hat{\alpha}}\right)^2 &= \left(1 - \frac{\alpha}{\hat{\alpha}}\right)\left(1 + \frac{\alpha}{\hat{\alpha}} - \alpha\right) \\ &\geq \left(1 - \frac{\alpha}{\hat{\alpha}}\right)\left(\frac{2\alpha}{\hat{\alpha}} - \alpha\right) \qquad (\text{since } \alpha < \hat{\alpha}) \\ &= \left(1 - \frac{\alpha}{\hat{\alpha}}\right)\left(\frac{2}{\hat{\alpha}} - 1\right)\alpha \\ &\geq 0. \qquad (\text{since } \hat{\alpha} \leq 2 \text{ and } \alpha < \hat{\alpha}) \end{aligned}$$

Thus we can apply Lemma 3.4, and obtain

$$x_j^{k+1} z_j^{k+1} \geq \left(1 - \alpha - (1-\hat{\alpha})\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)\pi^k f^k + \left(\alpha - \hat{\alpha}\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right)\beta f^k$$

for $j = 1, 2, \ldots, n$. By Lemma 3.1, we also know that

$$f^{k+1} = (1 - \alpha(1-\beta))f^k.$$

It follows from the two relations above and $\pi^k \geq \sigma$ that

$$\begin{aligned}
& x_j^{k+1} z_j^{k+1} - \sigma f^{k+1} \\
\geq\ & \left(1 - \alpha - (1-\hat{\alpha}) \left(\frac{\alpha}{\hat{\alpha}}\right)^2\right) \pi^k f^k - (1 - \alpha(1-\beta))\sigma f^k + \left(\alpha - \hat{\alpha}\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right) \beta f^k \\
\geq\ & \left(-\alpha\beta - (1-\hat{\alpha}) \left(\frac{\alpha}{\hat{\alpha}}\right)^2\right) \sigma f^k + \left(\alpha - \hat{\alpha}\left(\frac{\alpha}{\hat{\alpha}}\right)^2\right) \beta f^k.
\end{aligned}$$

If $\hat{\alpha} \geq \alpha^*$ then $\alpha = \theta\hat{\alpha}$; hence

$$\begin{aligned}
& x_j^{k+1} z_j^{k+1} - \sigma f^{k+1} \\
\geq\ & \left(-\theta\hat{\alpha}\beta - (1-\hat{\alpha})\theta^2\right) \sigma f^k + \left(\theta\hat{\alpha} - \hat{\alpha}\theta^2\right) \beta f^k \\
=\ & \left((\theta - \beta)\theta\hat{\alpha} - \theta^2\right) \sigma f^k + (1-\theta)\theta\hat{\alpha}\beta f^k \\
\geq\ & -\theta^2 \sigma f^k + (1-\theta)\theta\alpha^*\beta f^k \qquad \text{(since } \beta \leq 0.5 \leq \theta \text{ and } \alpha^* \leq \hat{\alpha}) \\
\geq\ & -\theta^2 \cdot \frac{(1-\theta)\alpha^*\beta}{2} \cdot f^k + (1-\theta)\theta\alpha^*\beta f^k \\
& \qquad\qquad (\text{since } \sigma \leq \frac{(1-\theta^*)\alpha^*\beta^*}{2} \leq \frac{(1-\theta)\alpha^*\beta}{2}) \\
=\ & (1 - 0.5\theta)(1-\theta)\theta\alpha^*\beta f^k \\
\geq\ & 0. \qquad\qquad (\text{since } \theta \leq \theta^* < 1)
\end{aligned}$$

On the other hand, if $\hat{\alpha} < \alpha^*$ then $\alpha = \theta\hat{\alpha}^2/\alpha^*$; hence

$$\begin{aligned}
& x_j^{k+1} z_j^{k+1} - \sigma f^{k+1} \\
\geq\ & \left(-\frac{\beta\theta\hat{\alpha}^2}{\alpha^*} - \left(\frac{\theta\hat{\alpha}}{\alpha^*}\right)^2\right) \sigma f^k + \left(\frac{\theta\hat{\alpha}^2}{\alpha^*} - \hat{\alpha}\left(\frac{\theta\hat{\alpha}}{\alpha^*}\right)^2\right) \beta f^k \\
\geq\ & -(\beta\alpha^* + \theta)\theta \left(\frac{\hat{\alpha}}{\alpha^*}\right)^2 \sigma f^k + (\alpha^* - \theta\alpha^*)\theta \left(\frac{\hat{\alpha}}{\alpha^*}\right)^2 \beta f^k \quad (\text{since } \alpha^* > \hat{\alpha}) \\
\geq\ & -2\theta \left(\frac{\hat{\alpha}}{\alpha^*}\right)^2 \sigma f^k + (1-\theta)\alpha^*\theta \left(\frac{\hat{\alpha}}{\alpha^*}\right)^2 \beta f^k \quad (\text{since } 0 \leq \beta,\ \theta,\ \alpha^* \leq 1) \\
=\ & (-2\sigma + (1-\theta)\alpha^*\beta)\theta \left(\frac{\hat{\alpha}}{\alpha^*}\right)^2 f^k \\
\geq\ & 0. \qquad\qquad (\text{since } \sigma \leq \frac{(1-\theta^*)\alpha^*\beta^*}{2} \leq \frac{(1-\theta)\alpha^*\beta}{2})
\end{aligned}$$

Thus we have shown the inequality

$$x_j^{k+1} z_j^{k+1} - \sigma f^{k+1} \geq 0$$

both for the case of $\hat{\alpha} \geq \alpha^*$ and for the case of $\hat{\alpha} < \alpha^*$. This inequality holds for $j = 1, 2, \ldots, n$. Therefore,

$$\pi^{k+1} = \min\left\{\frac{x_j^{k+1} z_j^{k+1}}{f^{k+1}} : j = 1, 2, \ldots, n\right\} \geq \sigma.$$

∎

Proof of Theorem 4.1

As in the proof of Theorem 3.2, we will use the symbols β^k, $\hat{\alpha}^k$, α^k and θ^k instead of β, $\hat{\alpha}$, α and θ, respectively. By Lemma 4.3 and the definition of σ, we obtain (i). By the assumption (13) on the parameters α^*, β^* and θ^*, we have

$$\sigma \leq \frac{(1-\theta^*)\alpha^*\beta^*}{2} \leq \frac{1}{4}.$$

By Lemma 3.3, we have

$$(\hat{\alpha}^k)^2 \geq \min\left\{\frac{1}{4}, \frac{2(\pi^k)^2}{n}\right\} \geq \min\left\{\frac{1}{4}, \frac{2\sigma^2}{n}\right\} = \frac{2\sigma^2}{n}.$$

Hence, the step length α^k determined by (15) satisfies

$$\begin{aligned}
\alpha^k &= \min\left\{\theta^k\hat{\alpha}^k,\ \frac{\theta^k(\hat{\alpha}^k)^2}{\alpha^*}\right\} \\
&\geq \min\left\{\theta^k\alpha^*,\ \frac{\theta^k(\hat{\alpha}^k)^2}{\alpha^*}\right\} && (\text{since } \hat{\alpha}^k \geq (\hat{\alpha}^k)^2/\alpha^* \text{ if } \alpha^* \geq \hat{\alpha}^k) \\
&\geq \min\left\{\theta^k\alpha^*,\ \frac{2\theta^k\sigma^2}{\alpha^* n}\right\} && (\text{since } (\hat{\alpha}^k)^2 \geq 2\sigma^2/n) \\
&\geq \min\left\{\frac{\alpha^*}{2},\ \frac{\sigma^2}{\alpha^* n}\right\} && (\text{since } \theta^k \geq 0.5) \\
&= \tilde{\alpha}.
\end{aligned}$$

Thus we have shown $\alpha^k \geq \tilde{\alpha}$. To prove (ii), assume that $r \geq 2\eta/\tilde{\alpha}$. Then

$$\begin{aligned}
&\boldsymbol{c}^T\boldsymbol{x}^{r+1} - \boldsymbol{b}^T\boldsymbol{y}^{r+1} \\
&= (\boldsymbol{c}^T\boldsymbol{x}^0 - \boldsymbol{b}^T\boldsymbol{y}^0)\prod_{k=1}^{r}(1-\alpha^k(1-\beta^k)) && (\text{by Lemma 3.1}) \\
&\leq (\boldsymbol{c}^T\boldsymbol{x}^0 - \boldsymbol{b}^T\boldsymbol{y}^0)\prod_{k=1}^{r}(1-0.5\tilde{\alpha}) && (\text{since } \beta^k \leq 0.5 \text{ and } \tilde{\alpha} \leq \alpha^k) \\
&= (\boldsymbol{c}^T\boldsymbol{x}^0 - \boldsymbol{b}^T\boldsymbol{y}^0)(1-0.5\tilde{\alpha})^r.
\end{aligned}$$

Hence

$$\begin{aligned}
\log(\boldsymbol{c}^T\boldsymbol{x}^{r+1} - \boldsymbol{b}^T\boldsymbol{y}^{r+1}) &\leq \log(\boldsymbol{c}^T\boldsymbol{x}^0 - \boldsymbol{b}^T\boldsymbol{y}^0) + r\log(1-0.5\tilde{\alpha}) \\
&\leq \log(\boldsymbol{c}^T\boldsymbol{x}^0 - \boldsymbol{b}^T\boldsymbol{y}^0) - 0.5\tilde{\alpha}r \\
&\leq \log(\boldsymbol{c}^T\boldsymbol{x}^0 - \boldsymbol{b}^T\boldsymbol{y}^0) - 0.5\tilde{\alpha}\cdot\frac{2\eta}{\tilde{\alpha}} \\
&\leq \log(\boldsymbol{c}^T\boldsymbol{x}^0 - \boldsymbol{b}^T\boldsymbol{y}^0) - \eta \\
&= \log(\boldsymbol{c}^T\boldsymbol{x}^0 - \boldsymbol{b}^T\boldsymbol{y}^0) - \log\left(\frac{\boldsymbol{c}^T\boldsymbol{x}^0 - \boldsymbol{b}^T\boldsymbol{y}^0}{\epsilon}\right) \\
&= \log\epsilon.
\end{aligned}$$

Thus the assertion (ii) follows. This completes the proof of Theorem 4.1.

Appendix. Inconsistency of the step length control rule (6).

We will show by an example that the step length control rule (6) used by McShane, Monma and Shanno [17] in the implementation of the GPD method is theoretically incomplete. We consider linear programming problems (P) and (D) with $n = 2$, $m = 1$, $\boldsymbol{A} = (1, -1)$, $\boldsymbol{b} = -0.9$ and $\boldsymbol{c} = (1, 1)^T$. Let $\boldsymbol{x}^k = (0.1, 1) \in P_{++}$ and $(\boldsymbol{y}^k, \boldsymbol{z}^k) = (0, 1, 1) \in D_{++}$. Take $\beta = 1/n = 0.5$ as in [17]. Then the Newton direction calculated as the solution of the system (3) turns out to be

$$\begin{aligned} \boldsymbol{\Delta x} &= (41/440, 41/440)^T, \\ \boldsymbol{\Delta y} &= -9/11, \\ \boldsymbol{\Delta z} &= (9/11, -9/11)^T. \end{aligned}$$

Since $\boldsymbol{\Delta x} \geq 0$, we have $\hat{\alpha}_p = \infty$. Hence the primal step length $\alpha_p = \theta\hat{\alpha}_p$ determined by (6) is infinite. We also see $\boldsymbol{c}^T\boldsymbol{\Delta x} > 0$, so that the duality gap $\boldsymbol{c}^T\boldsymbol{x}^{k+1} - \boldsymbol{b}^T\boldsymbol{y}^{k+1}$ given in (10) increases monotonically as the primal step length α_p increases.

References

[1] I. Adler, M. G. C. Resende, G. Veiga and N. Karmarkar, "An implementation of Karmarkar's algorithm for linear programming," *Mathematical Programming* **44** (1989) 297–335.

[2] E. R. Barnes, "A variation on Karmarkar's algorithm for solving linear programming problems," *Mathematical Programming* **36** (1986) 174–182.

[3] I. I. Dikin, "Iterative solution of problems of linear and quadratic programming," *Soviet Mathematics Doklady* **8** (1967) 674–675.

[4] J. Ding and T.-Y. Li, "A polynomial-time predictor-corrector algorithm for linear complementarity problems," *SIAM Journal on Optimization* **1** (1991) 83–92.

[5] P. D. Domich, P. T. Boggs, J. R. Donaldson and C. Witzgall, "Optimal 3-dimensional methods for linear programming," NISTIR89-4225, U.S. Department of Commerce, National Institute of Standards and Technology (Gaithersburg, MD, 1989).

[6] R. M. Freund, "Polynomial-time algorithms for linear programming based only on primal scaling and projected gradients of a potential function", Report OR 182-88, Sloan School of Management, Massachusetts Institute of Technology (Cambridge, Massachusetts, 1988).

[7] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Technique* (John Wiley and Sons, New York, 1968).

[8] M. Kojima, N. Megiddo and T. Noma, "Homotopy continuation methods for nonlinear complementarity problems," *Mathematics of Operations Research*, to appear.

[9] M. Kojima, N. Megiddo, T. Noma and A. Yoshise, "A unified approach to interior point algorithms for linear complementarity problems," Research Report RJ 7493 (70008), IBM Almaden Research Center (San Jose, CA, 1990).

[10] M. Kojima, N. Megiddo and Y. Ye, "An interior point potential reduction algorithm for the linear complementarity problem," *Mathematical Programming*, to appear.

[11] M. Kojima, S. Mizuno and A. Yoshise, "A primal-dual interior point algorithm for linear programming," in N. Megiddo, ed., *Progress in Mathematical Programming: Interior-Point and Related Methods* (Springer-Verlag, New York, 1989) 29–47.

[12] M. Kojima, S. Mizuno and A. Yoshise, "A polynomial-time algorithm for a class of linear complementary problems," *Mathematical Programming* **44** (1989) 1–26.

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

[14] I. J. Lustig, "A generic primal-dual interior point algorithm," Technical Report SOR 88-3, Program in Statistics and Operations Research, Department of Civil Engineering and Operations Research, Princeton University (Princeton, New Jersey, 1988).

[15] I. J. Lustig, Private communication (1989).

[16] R. Marsten, R. Subramanian, M. Saltzman, I. J. Lustig and D. F. Shanno, "Interior point methods for linear programming: Just call Newton, Lagrange and Fiacco and McCormick!" *Interfaces* **20** (1990) 105–116.

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

[18] N. Megiddo, "Pathways to the optimal set in linear programming," in N. Megiddo, ed., *Progress in Mathematical Programming: Interior-Point and Related Methods* (Springer-Verlag, New York, 1989) 131–158.

[19] S. Mehrotra, "On the implementation of a (primal-dual) interior point method," Technical Report 90-03, Department of Industrial Engineering and Management Sciences, Northwestern University (Evanston, IL, 1990).

[20] S. Mizuno, "An $O(n^3L)$ algorithm using a sequence for a linear complementarity problem," *J. Operations Research Soc. of Japan* **33** (1990) 66–75.

[21] S. Mizuno, M. J. Todd and Y. Ye, "On adaptive-step primal-dual interior-point algorithms for linear programming," Technical Report No. 944, School of Operations Research and Industrial Engineering, Cornell University (Ithaca, New York, USA) 1989.

[22] S. Mizuno, A. Yoshise and T. Kikuchi, "Practical polynomial time algorithms for linear complementarity problems," *J. Operations Research Soc. of Japan* **32** (1989) 75–92.

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

[24] R. D. C. Monteiro and I. Adler, "Interior path following primal-dual algorithms, Part II: Convex quadratic programming," *Mathematical Programming* **44** (1989) 43–66.

[25] R. D. C. Monteiro, I. Adler and M. G. C. Resende, "A polynomial-time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series extension," *Mathematics of Operations Research* **15** (1990) 191–214.

[26] G. Sonnevend and J. Stoer, "Global ellipsoidal approximations and homotopy methods for solving convex analytic programs," Report No. 40, Institut für Angewandte Mathematik und Statistik, Universität Würzburg (Würzburg, Germany, 1988).

[27] K. Tanabe, "Complementarity-enforcing centered Newton method for mathematical programming," in K. Tone, ed., *New Methods for Linear Programming*, The Institute of Statistical Mathematics (Tokyo 106, 1987) 118–144.

[28] K. Tanabe, "Centered Newton method for mathematical programming," in M. Iri and K. Yajima, ed., *Systems Modeling and Optimization* (Springer-Verlag, New York, 1988) 197–206.

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

[30] R. J. Vanderbei, M. S. Meketon and B. A. Freedman, "A modification of Karmarkar's linear programming algorithm," *Algorithmica* **1** (1986) 395–407.

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

[32] Y. Ye, "Line search in potential reduction algorithms for linear programming," Technical report, Department of Management Sciences, The University of Iowa (Iowa City, Iowa, 1989).

[33] Y. Ye, K. O. Kortanek, J. A. Kaliski and S. Huang, "Near-boundary behavior of primal-dual potential reduction algorithms for linear programming," Working Paper Series No. 90-9, College of Business Administration, The University of Iowa (Iowa City, Iowa, 1990).

[34] Y. Zhang, R. A. Tapia and J. E. Dennis, "On the superlinear and quadratic convergence of primal-dual interior point linear programming algorithms," Technical report, Department of Mathematical Sciences, Rice University (Houston, Texas, 1990).

[35] Y. Zhang, R. A. Tapia and F. Potra, "On the superlinear convergence of interior point algorithms for a general class of problems," TR90-9, Department of Mathematical Sciences, Rice University (Houston, Texas, 1990).