A Simple Variant of the Mizuno-Todd-Ye Predictor-Corrector Algorithm and its Objective-Function-Free Complexity

Tomonari Kitahara* Takashi Tsuchiya†

April 2011

Abstract

In this paper, we propose a simple variant of the Mizuno-Todd-Ye predictor-corrector algorithm for linear programming problem (LP). Our variants execute a natural finite termination procedure at each iteration and it is easy to implement the algorithm. Our algorithm admits an objective-function free polynomial-time complexity when it is applied to LPs whose dual feasible region is bounded.

Keywords: Linear programming problem,, Interior point method, Layered least squares interior point method, Iteration complexity, Strong polynomiality.

AMS Mathematical Subject Classification: 90C05

1 Introduction

In this paper, we deal with a simple polynomial-time algorithm for linear programming (LP) with objective-function-free polynomial-time complexity. Let us consider the dual pair of linear programs (LP)

$$\begin{array}{ll} \min & c^T x \\ \text{subject to} & Ax = b, \ x \geq 0, \end{array} \tag{1}$$

and its associated dual problem

$$\begin{array}{ll} \max & b^T y \\ \text{subject to} & A^T y + s = c, \ s \geq 0, \end{array} \tag{2}$$

where $A \in \Re^{m \times n},\ b \in \Re^m,\ c \in \Re^n$ are given, and the vectors $x,\ s \in \Re^n$ and $y \in \Re^m$ are the unknown variables. In the special case where $c = 0$, the problem is called the linear feasibility problem.

The existence of a strongly polynomial-time algorithm for LP is one of the most important open problems in optimization. A seminal result to this problem is due to Tardos who demonstrated that general LPs can be solved in polynomial-time depending just on A [9]. In the context of interior-point algorithms, Vavasis and Ye developed the layered least squares

*Graduate School of Decision Science and Technology, Tokyo Institute of Technology, 2-12-1-W9-62, Ookayama, Meguro-ku, Tokyo, 152-8552, Japan. Tel.: +81-3-5734-2896, Fax: +81-3-5734-2947, E-mail: kitahara.t.ab@m.titech.ac.jp.

†National Graduate Institute for Policy Studies, 7-22-1, Roppongi, Minato-ku, Tokyo, 106-8677, Japan. Tel.& Fax.: +81-3-6439-6130, E-mail: tsuchiya@grips.ac.jp. This author is supported in part with Grant-in-Aid for Scientific Research (B), 2008, 20340024 from the Japan Society for the Promotion of Sciences.

step interior-point algorithm which enjoys the same type of polynomial-time complexity [13]. Vavasis and Ye's algorithm accelerates convergence utilizing the layered least squares step instead of the ordinary interior-point step. See [4] and [6] for variants of the Vavasis and Ye's algorithm. These algorithms are strongly polynomial-time algorithms for combinatorial LPs.

Apart from their theoretical significance, they are hardly practical because of their complication. It is an interesting problem to develop a practical and simple algorithm which enjoys a polynial-time complexity bound depending just on A. In this paper, we make a step to this task.

We show that a simple and natural finite-termination variant of the MTY predictor-corrector algorithm [5] enjoys an objective-function-free computational polynomial-time complexity when applied to LPs with bounded feasible region. This variant utilizes a scaling-invariant finite-termination direction also as an alternative search direction to accelarate convergence at the predictor step. This algorithm would be one of the simplest variants of the MTY predictor-corrector algorithm with finite termination property.

The finite-termination direciton admits an interpretation as a two layered least squares direction. In connection with this, we also show that at most three layer is enough to obtain objective-function-free polynomial-time complexity for general LP problems.

This paper is organized as follows. In Section 2, we review primal-dual interior-point algorithms and state the main results of this paper. In Section 3, we establish a lower bound and an upper bound for the value of dual variables which ensure that they belong to B-index. In Section 4, we will prove the main results based on the result developed in Section 3. In Section 5, we develop some technical results needed in Section 3. Finally in Section 6, we conclude the paper.

2 Linear Programming and MTY Predictor-Corrector algorithms

In this section we review the MTY P-C algorithm for solving the pair of LP problems (1) and (2).

2.1 Assumptions, the central path and its 2-norm neighborhood

In this subsection, we first state our assumptions. After that, we describe the primal-dual central path and its corresponding two-norm neighborhoods.

Given $A \in \Re^{m\times n}$, $b \in \Re^m$ and $c \in \Re^n$, consider the pair of linear programming problems (1) and (2). The set of strictly feasible solutions for these problems are defined as follows.

$$\mathcal{P}^{++} := \{x \in \Re^n : \ Ax = b, \ x > 0\},$$
$$\mathcal{D}^{++} := \{(y, s) \in \Re^m \times \Re^n : \ A^T y + s = c, \ s > 0\}.$$

The followings are standard assumptions on the pair of the problems (1) and (2).

A.1 The rows of A are linearly independent.

A.2 $\mathcal{P}^{++}$ and $\mathcal{D}^{++}$ are nonempty.

Under the above assumptions, it is well-known that for any $\nu > 0$ the system

$$xs = \nu e, \quad (3)$$
$$Ax = b, \ x > 0, \quad (4)$$
$$A^T y + s = c, \ s > 0 \quad (5)$$

has a unique solution $(x(\nu), y(\nu), s(\nu))$. The central path is the set

$$\{(x(\nu), y(\nu), s(\nu)) : \ \nu > 0\}.$$

As ν goes to zero, the path converges to a primal-dual optimal solution (x^*, y^*, s^*) for problems (1) and (2). In our analysis, we further impose the following unique assumptions.

A.3 The system $A^T y = c$ has no solution.

A.4 The set $\{s \in \Re^n : \ s = c - A^T y, y \in \Re^m, s \geq 0\}$ is bounded.

A few comments concerning the assumptions are in order. First, if A.3 does not hold, then any primal feasible solution is an optimal solution for (1), which makes A.3 not so special. Second, under A.1, both (1) and (2) have optimal vertex solutions. The norm of a dual optimal vertex is bounded above by a constant depending only on A and c. Thus, by adding an appropriate constraint to (2) we can make A.4 hold, although the structures of the constraint matrix and the objective vector may change.

Given a point $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$, its duality gap and its normalized duality gap are defined as $x^T s$ and $\mu = \mu(x, s) := x^T s/n$, respectively. We call the point $(x(\mu), y(\mu), s(\mu))$ the central point associated with w. Note that $(x(\mu), y(\mu), s(\mu))$ also has normalized duality gap μ. We define the proximity measure of a point $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ with respect to the central path by

$$\phi(w) = \|xs/\mu - e\|.$$

It is easy to see that $\phi(w) = 0$ if and only if $w = (x(\mu), y(\mu), s(\mu))$. The two-norm neighborhood of the central path with opening β is defined as

$$\mathcal{N}(\beta) := \{w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++} : \ \phi(w) \leq \beta\}.$$

Finally, for any point $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$, we define

$$\delta(w) := s^{1/2} x^{-1/2} \in \Re^n. \tag{6}$$

2.2 Condition numbers

In this subsection we define two condition numbers and state properties of them. One is well known $\bar{\chi}_A$, which is associated with the constraint matrix A. The other is $\zeta_{(A,c)}$, which depends on the constraint matrix A and the objective vector c. $\zeta_{(A,c)}$ is unique to our analysis.

We begin with $\bar{\chi}_A$. Let $\mathcal{D}$ denote the set of all positive definite $n \times n$ diagonal matrices and define

$$\bar{\chi}_A := \sup\{\|A^T (ADA^T)^{-1} AD\| : \ D \in \mathcal{D}\} \tag{7}$$

Finiteness of $\bar{\chi}_A$ is firstly established by Dikin [1]. The condition number and related properties are later studied in [2, 8, 10, 11].

We summarize in the next proposition a few important facts about $\bar{\chi}_A$.

Proposition 2.1 *Let $A \in \Re^{m \times n}$ with full row rank be given. Then the following statements hold:*

(a) $\bar{\chi}_{GA} = \bar{\chi}_A$ for any nonsingular matrix $G \in \Re^{m \times m}$;

(b) $\bar{\chi}_A = \max\{\|G^{-1} A\| : \ G \in \mathcal{G}\}$ where $\mathcal{G}$ denotes the set of all $m \times m$ nonsingular submatrices of A;

(c) *Suppose A includes $m \times m$ identity matrix as its submatrix. Then for any submatrix $\tilde{A}$ of A, we have $\|\tilde{A}\| \le \bar{\chi}_A$.*

(d) *Suppose A includes $m \times m$ identity matrix as its submatrix. Let $\tilde{A} \in \Re^{p\times n}$ be any submatrix of A. Then we have $\bar{\chi}_{\tilde{A}} \le \bar{\chi}_A$.*

(e) *If the entries of A are all integers, then $\bar{\chi}_A$ is bounded above by $2^{\mathcal{O}(L_A)}$, where L_A is the input bit length of A;*

Next we proceed to $\zeta_{(A,c)}$. For $I \subset \{1,\dots,n\}$, consider

$$r(I) := \min_{\tilde{y}\in\Re^m} \|c_I - A_I^T y\|$$

and define

$$\delta_{(A,c)} := \min_{I:\ r(I)\neq 0} r_I. \tag{8}$$

Note that under the assumption A.4, $\delta_{(A,c)}$ is well-defined. From the definition of $\delta_{A,c}$, it is easy to see the following proposition holds.

Proposition 2.2 *For any $I \subset \{1,\dots,n\}$, if there exists y' such that $\|c_I - A_I^T y'\| < \delta_{(A,c)}$, the equation $A_I^T y = c_I$ has a solution.*

The condition number $\zeta_{(A,c)}$ is defined as follows:

$$\zeta_{(A,c)} := \frac{\|c\|}{\delta_{(A,c)}} \tag{9}$$

Note that $\zeta_{(A,c)}$ is invariant under scalings of c.

2.3 Predictor-corrector step and its properties

In this subsection we review the well-known MTY P-C algorithm and its main properties.

Each iteration of the MTY P-C algorithm consists of two steps, namely the predictor (or affine scaling) step and the corrector step (or centrality) step. The search direction used by both steps at a given point $w \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ is the unique solution of the following linear system of equations

$$\begin{aligned} S\Delta x + X\Delta s &= \sigma\mu e - xs, \\ A\Delta x &= 0, \\ A^T\Delta y + \Delta s &= 0, \end{aligned} \tag{10}$$

where $\mu = \mu(x,s)$ and $\sigma \in \Re$ is a prespecified parameter, commonly referred to as the centrality parameter. When $\sigma = 0$, we donote the solution of (10) by $(\Delta x^a, \Delta y^a, \Delta s^a)$ and refer to it as the (primal-dual) affine scaling (AS) direction at w; it is the direction used in the predictor step. When $\sigma = 1$, we denote the solution of (10) by $(\Delta x^c, \Delta y^c, \Delta s^c)$ and refer to it as the centrality direction at w; it is the direction used in the corrector step.

In the following, we describe one iteration of the MTY P-C algorithm. Suppose that a constant $\beta \in (0, 1/4]$ is given. Given a point $w = (x,y,s) \in \mathcal{N}(\beta)$, the algorithm gererates the next point as follows. First move along the direction $(\Delta x^a, \Delta y^a, \Delta s^a)$ until the point hits the boundary of the enlarged neighborhood $\mathcal{N}(2\beta)$. More precisely, the algorithm computes the point $(x^a, y^a, s^a) := (x,y,s) + \alpha_a(\Delta x^a, \Delta y^a, \Delta s^a)$ where

$$\alpha_a := \sup\{\alpha \in [0,1] :\ (x,y,s) + \alpha(\Delta x^a, \Delta y^a, \Delta s^a) \in \mathcal{N}(2\beta).\} \tag{11}$$

This is the predictor step. After that, in the corrector step, the point $w^+ = (x^+, y^+, s^+)$ inside the smaller neighborhood $\mathcal{N}(\beta)$ is generated by taking a unit step along the centrality step at the point w^a, that is, $(x^+, y^+, s^+) = (x^a, y^a, s^a) + (\Delta x^c, \Delta y^c, \Delta s^c) \in \mathcal{N}(\beta)$. Starting from a point $w^0 \in \mathcal{N}(\beta)$ and successively performing iterations as above, the MTY P-C algorithm generates a sequence of points $w^k \subset \mathcal{N}(\beta)$ which converges to a primal-dual optimal solution of problems (1) and (2).

Proposition 2.3 (Predictor step) *Suppose $w = (x, y, s) \in \mathcal{N}(\beta)$ for some constant $\beta \in (0, 1/4]$. Let $\Delta w^a = (\Delta x^a, \Delta y^a, \Delta s^a)$ denote the AS direction at w and let α_a be the step-size computed according to (11). Then the following statements hold:*

(a) the point $w + \alpha \Delta w^a$ has the normalized duality gap $\mu(\alpha) = (1 - \alpha)\mu$ for all $\alpha \in \Re$;

(b) $\alpha_a \geq \max\{1 - \chi/\beta, \sqrt{\beta/n}\}$ where $\chi := \|\Delta x^a \Delta s^a\|/\mu$.

Proposition 2.4 (Corrector step) *Suppose $w = (x, y, s) \in \mathcal{N}(2\beta)$ for some constant $\beta \in (0, 1/4]$ and let $(\Delta x^c, \Delta y^c, \Delta s^c)$ denote the corrector step at w. Then $w + \Delta w^c \in \mathcal{N}(\beta)$. Moreover, the (normalized) duality gap of $w + \Delta w^c$ is the same as that of w.*

For a search direction $\Delta w = (\Delta x, \Delta y, \Delta s)$ at a point $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$, the quantity

$$\begin{aligned} (Rx, Rs) &:= \left(\tfrac{\delta(x+\Delta x)}{\sqrt{\mu}}, \tfrac{\delta^{-1}(s+\Delta s)}{\sqrt{\mu}}\right) \\ &= \left(\tfrac{x^{1/2}s^{1/2}+\delta\Delta x}{\sqrt{\mu}}, \tfrac{x^{1/2}s^{1/2}+\delta^{-1}\Delta s}{\sqrt{\mu}}\right) \end{aligned} \tag{12}$$

where $\delta = \delta(w) = s^{1/2}x^{-1/2}$, appears quite often later in our analysis. We refer to it as the *residual* of Δw. Throughout this paper, we denote the residual of the affine scaling direction $\Delta w^a = (\Delta x^a, \Delta y^a, \Delta s^a)$ at w as $(Rx^a(w), Rs^a(w))$. Note that for the residual of the affine scaling direction, we have

$$Rx^a = -\frac{1}{\sqrt{\mu}}\delta^{-1}\Delta s^a, \;\; Rs^a = -\frac{1}{\sqrt{\mu}}\delta \Delta x^a, \tag{13}$$

and

$$Rx^a + Rs^a = \frac{x^{1/2}s^{1/2}}{\sqrt{\mu}} \tag{14}$$

since $(\Delta x^a, \Delta y^a, \Delta s^a)$ satisfies the first equation of (10) with $\sigma = 0$.

2.4 Important tools

In this subsection, we introduce a few important tools, namely the affine-scaling-bipartition, the affine-scaling-partition, and the crossover event.

First we explain the AS-bipartition and the AS-partition. Suppose $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ is given. Then the AS bipartition at w is defined as follows:

$$\begin{aligned} B(w) &:= \{i : \; |Rs_i^a(w) \leq |Rx_i^a(x)|\}, \\ N(w) &:= \{i : \; |Rs_i^a(w)| > |Rx_i^a(w)|\}. \end{aligned} \tag{15}$$

Remember

$$\begin{aligned} B_* &:= \{i : x_i > 0 \text{ for some optimal solution } x \text{ of } (1)\}, \\ N_* &:= \{i : s_i > 0 \text{ for some optimal solution } (y, s) \text{ of } (2)\}\}, \end{aligned} \tag{16}$$

which is referred to as the optimal partition associated with the pair pf LP problems (1) and (2). The AS-bipartition can be seen as a guess for the optimal partition (B_*, N_*).

Now we define the AS-partition based on the AS-bipartition. Given a point $\mathcal{P}^{++} \times \mathcal{D}^{++}$, we first compute the AS-bipartition $(B, N) = (B(w), N(w))$ by (15). Next, we choose an order $i_1, \dots, i_n$ of index $\{1, \dots, n\}$ such that $\delta_{i_1} \leq \dots \delta_{i_n}$. Then the first block of consecutive indices in the n-tuple $(i_1, \dots, i_n)$ lying in the same set B or N are placed in the first layer $\mathcal{J}_1$, the next block of consecutive indices lying in the other set is placed in $\mathcal{J}_2$, and so on. As an example, assume that $\{i_1, i_2, i_3, i_4, i_5, i_6, i_7\} \in B \times B \times N \times B \times B \times N \times N$. In this case, we have $\mathcal{J}_1 = \{1, 2\}$, $\mathcal{J}_2 = \{3\}$, $\mathcal{J}_3 = \{4, 5\}$ and $\mathcal{J}_4 = \{6, 7\}$. A partition obtained according to the above construction is referred to as the AS-partition, and we denote it by $\mathcal{J}(w)$. Note that the AS-partition is not always unique as there may be multiple ascending orders.

We proceed to a discussion of the crossover event.

Definition 2.1 *For two indices $i, j \in \{1 \dots, n\}$ and a constant $\mathcal{C} \geq 1$, we say that a $\mathcal{C}$-crossover event for the pair (i, j) occurs on the interval $(\nu', \nu]$ if*

$$\begin{array}{l} \text{there exists } \nu_0 \in (\nu', \nu] \text{ such that } \frac{s_j(\nu_0)}{s_i(\nu_0)} \leq \mathcal{C}, \\ \text{and, } \frac{s_j(\tilde{\nu})}{s_i(\tilde{\nu})} > \mathcal{C} \text{ for all } \bar{\nu} \leq \nu'. \end{array} \tag{17}$$

Moreover, we say the interval $(\nu', \nu]$ contains a $\mathcal{C}$-crossover event if (17) holds for some pair (i, j).

From the definition, the crossover event is independent of any algorithm and is a property of the central path only.

We have the following simple but crucial result about the crossover event.

Proposition 2.5 *Let $\mathcal{C}$ be a given constant. There can be at most $n(n-1)/2$ disjoint intervals of the form $(\nu', \nu]$ containing $\mathcal{C}$-crossover events.*

The notion of $\mathcal{C}$-crossover events can be used to define the notion of $\mathcal{C}$-crossover events between two iterates of the MTY P-C algorithm as follows. We say that a $\mathcal{C}$-crossover event occurs between two iterates w^k and w^l, generated by the MTY P-C algorthm if the interval $(\mu(w^l), \mu(w^k)]$ contains a $\mathcal{C}$-crossover event.

2.5 MTY-PC algorithm with a finite termination procedure

In this subsection, we describe a modified MTY-PC algorithm with a finite termination procedure for finding an exact optimal solution of (1) and (2). For the procedure, we need the AS-bipartition $(B(w), N(w))$ at a point $w = (x, y, s)$. Recall that this bipartition is scaling invariant and it is a guess for the optimal partition (B_*, N_*). The finite termination procedure is summarized as follows.

Finite termination (FT) procedure

Given $\beta \in (0, 1/4]$ and $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ such that $w \in \mathcal{N}(\beta)$.
(1) Find the AS-bipartition $(B, N) = (B(w), N(w))$.
(2) Let

$$L_2^x = \arg\min_{Au=0} \|\delta_N(x_N + u_N)\|. \tag{18}$$

Then determine the primal direction Δx^f by solving the following problem.

$$\begin{array}{ll} \min & \|\delta_B(x_B + \Delta x_B)\| \\ \text{subject to} & \Delta x \in L_2^x. \end{array} \tag{19}$$

(3) Let

$$L_1^s = \arg\min_{v \in \mathrm{Im}(A^T)} \|\delta_B^{-1}(s_B + v_B)\|. \tag{20}$$

Then determine the dual direction Δs^f by solving the following problem.

$$\begin{array}{ll} \min & \|\delta_N^{-1}(s_N + \Delta s_N)\| \\ \text{subject to} & \Delta s \in L_1^s. \end{array} \tag{21}$$

We determine Δy^f by solving the equation $A^T \Delta y^f + \Delta s^l = 0$. We call $(\Delta x^f, \Delta y^f, \Delta s^f)$ the FT direction.
(4) Let $\tilde{w} = (x + \Delta x^f, y + \Delta y^f, s + \Delta s^f)$. If $\tilde{w} \in \mathcal{P}$, declare success and output $\tilde{w}$ as an optimal solution. Otherwise declare failure.

With FT procedure above, our modified MTY-PC algorithm is described as follows.

Modified MTY-PC algorithm

Let $\beta \in (0, \frac{1}{4}]$ and $w^0 \in \mathcal{N}(\beta)$ be given. Set $k = 0$.

1. Set $w = w^k$ and compute the AS direction $\Delta w^a = (\Delta x^a, \Delta y^a, \Delta s^a)$ at w.
2. Compute the AS bipartition (B, N) according to (15).
3. Compute the FT direction $\Delta w^f = (\Delta x^f, \Delta y^f, \Delta s^f)$ and conduct the FT procedure. If the procedure succeed, the algorithm stops. Output the optimal solution $\tilde{w}$.
4. Let $w^a = w + \alpha^a \Delta^a$, where $\alpha^a = \sup\{\alpha \in [0, 1] : \ w + \alpha\Delta^a \in N(2\beta)\}$.
5. Let $w^f = w + \alpha^f \Delta^f$, where $\alpha^f = \sup\{\alpha \in [0, 1] : \ w + \alpha\Delta^f \in N(2\beta)\}$.
6. If $\mu(w^f) < \mu(w^a)$, set $w = w^f$. Otherwise set $w = w^a$.
7. If $\mu(w) = 0$, the algorithm stops. In this case, w is an optimal solution.
8. Compute the corrector step Δw^c at w and set $w \leftarrow w + \Delta w^c$.
9. Set $w^{k+1} = w$, increase k by one and go to step 1.

2.6 A new result

In this subsection, we state the result of our convergence analysis of the MTY P-C algorithm.

Theorem 2.1 *Given a termination tolerance $\eta > 0$ for the normalized duality gap and an initial point $w^0 \in \mathcal{N}(\beta)$ with $\beta \in (0, 1/4]$, the MTY P-C algorithm generates an iterate $w^k \in \mathcal{N}(\beta)$ satisfying $\mu(w^k) \le \eta$ in at most*

$$\mathcal{O}(\min\{\sqrt{n}\log(\mu_0/\eta), n^{3.5}\log(\bar{\chi}_A^* + n) + n^2 \log \zeta_{(A,c)}\}), \tag{22}$$

where $\zeta_{(A,c)}$ is defined in (9) and

$$\bar{\chi}_A^* := \inf\{\bar{\chi}_{AD} : \ D \in \mathcal{D}\}.$$

3 Lower and upper bounds for N indices on the path

Recall the AS bipartition is a guess for the optimal partition (B_*, N_*). In this section, we show that on the central path the guess is true as to an index whose corresponding dual variable is sufficiently small or sufficiently large.

We begin with an important result concerning the residual of the affine scaling direction on the central path.

Lemma 3.1 *Suppose $\mu > 0$. Let $w(\mu) = (x(\mu), y(\mu), s(\mu))$ be a point on the central path. Then we have:*

$$Rx^a(w(\mu)) + Rs^a(w(\mu)) = e, \tag{23}$$
$$Rx^a = S(\mu)^{-1}A^T(AS(\mu)^{-2}A^T))^{-1}AS(\mu)^{-1}e = P(s(\mu); A)e. \tag{24}$$

In the following, for $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ we call $P(s; A)$ the projection matrix.

We need to analyze the asymptotic behavior of the projection matrix when some of the dual variables go to zero and the others stay away from zero. Then let $I \subset \{1, \dots, n\}$ be a set of indices whose dual variables are sufficiently small and set $J = \{1, \dots, n\} \setminus I$. Let us split the constraint matrix as follows

$$A^T = \begin{bmatrix} A_I^T \\ A_J^T \end{bmatrix}.$$

To investigate the projection matrix, it is convenient to convert the constraint matrix to a special form called block lower triangular matrix. This technique was established and developed in [3, 4, 12]. The following two lemmas establish convertability of the constraint matrix.

Lemma 3.2 *Let r be the rank of A_I^T. By a sequence of column elementary transformations, the constraint matrix A^T can be converted to a block lower triangular matrix T', which satisfies the following conditions*

(1) T' is a 2×2 block lower triangular matrix, i.e.

$$T' = \begin{bmatrix} T'_{I1} & O \\ T'_{J1} & T'_{J2} \end{bmatrix},$$

where $T'_{I1} \in \Re^{|I| \times r}$ and $T'_{J2} \in \Re^{|J| \times (m-r)}$.

(2) $\mathrm{rank}(T'_{I1}) = r, \ \mathrm{rank}(T'_{J2}) = m - r.$

(3) The image of A^T is unchnaged, that is,

$$\mathrm{Im}\begin{bmatrix} A_I^T \\ A_J^T \end{bmatrix} = \mathrm{Im}\begin{bmatrix} T'_{I1} & O \\ T'_{J1} & T'_{J2} \end{bmatrix}.$$

Proof. Select r vectors from A_I^T which form a basis for Im A_I^T, and move these vectors to the left most side of A_I^T (this can be done by a sequence of column elementary transformations). Then by eliminating the rest of $(m-r)$ vectors by another sequence of column elementary transformations, we can convert A^T to the form of T'.

It is easy to see that the condition (2) holds.

We can convert A^T to T' only by column elementary transformations. This means that T' can be represented as $A^T G'^{-1}$ for some nonsingular matrix $G' \in \Re^{m\times m}$, which proves (3). ■

As the next lemma suggests, the block lower triangular matrix can be transformed further.

Lemma 3.3 *Let r be the rank of A_I^T. By a sequence of column basic elementary transformations, the constraint matrix can be converted to a basis block lower triangular matrix T satisfying the following conditions:*

(1) T is a 2×2 block lower triangular matrix, i.e.

$$T = \begin{bmatrix} T_{I1} & O \\ T_{J1} & T_{J2} \end{bmatrix},$$

where $T_{I1} \in \Re^{|I|\times r}$ and $T_{J2} \in \Re^{|J|\times(m-r)}$.

(2) T include the m-dimensional identity matrix as its submatrix.

(3) Both T_{I1}, T_{J2} include the identity matrix as their submatrices, whose size are r and $m-r$, respectively.

Proof. Suppose we have a block lower triangular matrix T' by the method described in Lemma 3.2. As T'_{I1}, T'_{J2} are column linear independent, we can make a matrix T'' from T', by a sequence of column elementary transformations, such that both T''_{I1} and T''_{J2} have the identity matrix as their submatrix. Then by eliminating non-zero elements in T''_{J1} (this can be done by column elementary transformations), we obtain the desired matrix T.

Note that there is a nonsingular matrix $G \in \Re^{m\times m}$ satisfying $T = A^T G^{-1}$, as T can be constructed from A^T by a sequence of column elementary transformations. ■

The next lemma states the asymptotic behavior of the projection matrix, which is a key result in our analysis. Its proof can be found in Appendix.

Lemma 3.4 *Suppose $w = (x, y, s) \in \mathcal{P}^{++}\times\mathcal{D}^{++}$, $I \subset \{1,\dots,n\}$, $J = \{1,\dots,n\}\setminus I$. Let $A^T = TG$, where $T \in \Re^{n\times m}$ is a basis block lower triangular matrix and $G \in \Re^{m\times m}$ is a nonsingular matrix. Then we have:*

$$P(s;A) = S^{-1}T(T^TS^{-2}T)^{-1}T^TS^{-1} = P(s;T^T), \tag{25}$$

$$P(s;A) = \begin{bmatrix} P(s_I;T_{I1}^T) & O \\ O & P(s_J;T_{J2}^T) \end{bmatrix} + \begin{bmatrix} C_{I1} & C_{I2} \\ C_{I2}^T & C_{J2} \end{bmatrix}, \tag{26}$$

where for $\tau := \|S_I\|\|S_J^{-1}\|$,

$$\begin{array}{l} \|C_{I1}\| \le \bar{\chi}_A^2 \tau^2, \\ \|C_{I2}\| \le 2(1+\bar{\chi}_A^2\tau^2)\bar{\chi}_A\tau, \\ \|C_{J2}\| \le (4+3\bar{\chi}_A^2\tau^2)\bar{\chi}_A^2\tau^2. \end{array} \tag{27}$$

The next lemma is a technical result. For the proof, see Monteiro and Tsuchiya [6].

Lemma 3.5 *Suppose* $\beta \in (0,1)$, $w = (x,y,s) \in \mathcal{N}(\beta)$. *Let* $w(\mu) = (x(\mu), y(\mu), s(\mu))$ *be the central path point associated with* w. *Then we have:*

$$\frac{1-\beta}{1+\beta}s \le \quad s(\mu) \quad \le \frac{1}{1-\beta}s, \tag{28}$$

$$\frac{(1-\beta)^2}{1+\beta}\frac{\delta_i}{\delta_j} \le \quad \frac{s_i(\mu)}{s_j(\mu)} \quad \le \frac{1+\beta}{(1-\beta)^2}\frac{\delta_i}{\delta_j}, \ \forall i,j \in \{1,\dots,n\}. \tag{29}$$

3.1 Lower bound

In this subsection, we prove a lower bound for N variables. First we show a technical lemma.

Lemma 3.6 *Suppose* $w = (x,y,s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$, $I \subset \{1,\dots,n\}$, $J = \{1,\dots,n\} \setminus I$. *Let* $A^T = TG$, *where* $T \in \Re^{n\times m}$ *is a basis block lower triangular matrix and* $G \in \Re^{m\times m}$ *is a nonsingular matrix. If there exists* $y' \in \Re^m$ *such that* $c_I = A_I^T y'$, *then we have*

$$P(s_I; T_{I1}^T)e_I = e_I.$$

Proof. It suffices to show that there exists $\bar{y}$ such that $S_I^{-1}A_I^T\bar{y} = e_I$. Indeed, by substituting $c_I = A_I^T y'$ to $s_I = c_I - A_I^T y$, we have

$$s_I = A_I^T(y' - y),$$

which implies the existence of the desired vector. ■

The following lemma is one of the main results of this section.

Theorem 3.1 *Suppose* $\beta \in (0, 1/4]$ *and* $w = w^k \in \mathcal{N}(\beta)$ *is an iterate of the MTY P-C algorithm. Let* $w(\mu) = (x(\mu), y(\mu), s(\mu))$ *be the point on the central path associated with* w. *If*

$$s_i \le \frac{(1-\beta)\delta(A,c)}{\sqrt{n}(5\bar{\chi}_A\sqrt{n})^n} \tag{30}$$

for some $i \in \{1,\dots,n\}$, *then* $i \in B(w(\mu))$ *or*

$$|Rx_i^a(w(\mu))| \ge |Rs_i^a(w(\mu))|.$$

Proof. Let us permutate the elements of $s(\mu)$ in an ascending order, which yields $s_{i_1}(\mu) \le \cdots \le s_{i_k}(\mu) = s_i(\mu) \le \cdots \le s_{i_n}(\mu)$. When $k = n$ we have

$$s_{i_n}(\mu) \le \frac{1}{1-\beta}s_i \le \frac{\delta(A,c)}{\sqrt{n}(5\bar{\chi}_A\sqrt{n})^n} < \frac{\delta(A,c)}{\sqrt{n}}.$$

where we used Lemma 3.5 for the first inequality. Consider the case $k < n$. If we assume for all $l \ge k$ that $s_{i_{l+1}}(\mu)/s_{i_l}(\mu) < 5\bar{\chi}_A\sqrt{n}$. Then we have

$$\begin{array}{rl} s_{i_n}(\mu) & < (5\bar{\chi}_A\sqrt{n})^{n-k}s_{i_k}(\mu) \\ & \le (5\bar{\chi}_A\sqrt{n})^n s_i(\mu) \\ & \le (5\bar{\chi}_A\sqrt{n})^n \frac{1}{1-\beta}s_i \\ & \le \frac{\delta(A,c)}{\sqrt{n}}. \end{array}$$

Thus in both case we deduce that

$$\|s(\mu)\| \leq \sqrt{n} s_{i_n}(\mu) < \delta(A, c).$$

From the definition of $\delta(A,c)$, this means that there exisits $\tilde{y}$ such that $c - A^T\tilde{y} = 0$, which contradicts the assumption stated in Section 2. Thus we have both $k < n$ and there exists $l \geq k$ such that $s_{i_{l+1}}(\mu)/s_{i_l}(\mu) \geq 5\bar{\chi}_A\sqrt{n}$.

Let l' be the minimum index satisfying the inequality, and partition the set of indices as follows:

$$I = \{i_1, \ldots, i_k = i, \ldots, i_{l'}\},\ J = \{i_{l'+1}, \ldots, i_n\}.$$

By the same argument as the one in the last paragraph, we can show that $\|s_I(\mu)\| < \delta(A,c)$, which guarantees the existence of $\hat{y}$ such that $c_I = A_I^T\hat{y}$. Thus we have $P(s_I; T_{I1}^T)e_I = e_I$ from Lemma 3.6. Note from Lemma 3.1 and Lemma 3.4 that

$$\begin{aligned}
&Rx^a(w(\mu)) + Rs^a(w(\mu)) = e,\\
&Rx^a(w(\mu)) = P(s;A)e = P(s;T^T)e,\\
&(P(s;A)e)_I = P(s_I;T_{I1}^T)e_I + C_{I1}e_I + C_{I2}e_J\\
&\qquad = e_I + C_{I1}e_I + C_{I2}e_J.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
|Rs_i^a(w(\mu))| &= |1 - Rx_i^a(w(\mu))|\\
&= |-(C_{I1}e_I)_i - (C_{I2}e_J)_i|\\
&\leq \|C_{I1}e_I\| + \|C_{I2}e_J\|.
\end{aligned}$$

Now we bound the last term as follows.

$$\begin{aligned}
&\|C_{I1}e_I\| + \|C_{I2}e_J\|\\
&\leq \|C_{I1}\|\|e_I\| + \|C_{I2}\|\|e_J\|\\
&\leq \bar{\chi}_A^2\tau^2 n + 2(1+\bar{\chi}_A^2\tau^2)\bar{\chi}_A\tau\sqrt{n} \quad \text{(from Lemma 3.4)}\\
&\leq \bar{\chi}_A^2(1/5\bar{\chi}_A\sqrt{n})^2 n + 2(1+\bar{\chi}_A^2(1/5\bar{\chi}_A\sqrt{n})^2)\bar{\chi}_A(1/5\bar{\chi}_A\sqrt{n})\sqrt{n}\\
&\quad (\text{as } \tau = \|S_I\|\|S_J^{-1}\| = s_{i'_l}/s_{i_{l'+1}} \leq (1/5\bar{\chi}_A\sqrt{n}))\\
&\leq (\tfrac{1}{5})^2 + 2(1+(\tfrac{1}{5})^2)\tfrac{1}{5}\\
&= \tfrac{57}{125} < 1/2.
\end{aligned}$$

Then we have

$$\begin{aligned}
|Rs_i^a(w(\mu))| &\leq \|C_{I1}e_I\| + \|C_{I2}e_J\|\\
&< 1 - \|C_{I1}e_I\| - \|C_{I2}e_J\|\\
&\leq 1 - |(C_{I1}e_I)_i| - |(C_{I2}e_J)_i|\\
&\leq |1 + (C_{I1}e_I)_i + (C_{I2}e_J)_i| = |Rx_i^a(w(\mu))|,
\end{aligned}$$

thus we conclude that $i \in B(w(\mu))$. ■

3.2 Upper bound

In this subsection, we derive an upper bound for N variables. We first address the case where all slack variables are sufficiently large.

Lemma 3.7 *Suppose $\beta \in (0, 1/4]$ and $w = w^k \in \mathcal{N}(\beta)$ is an iterate of the MTY P-C algorithm. Let $w(\mu) = (x(\mu), y(\mu), s(\mu))$ be the point on the central path associated with w. If $s_i \geq \frac{5\chi_A\|c\|(1+\beta)}{1-\beta}$ for all $i = 1, \ldots, n$, we have $i \in B(w(\mu))$ for all $i = 1, \ldots, n$ or*

$$|Rx_i^a(w(\mu))| \geq |Rs_i^a(w(\mu))|,\ \forall i = 1, \ldots, n.$$

Proof. In the following, we show that $Rx_i^a(w(\mu)) \geq 1/2$ for all $i \in \{1, \dots, n\}$. The statement of the lemma is readily follows from this inequality and the relation $Rx^a(w(\mu)) + Rs^a(w(\mu)) = e$. Note that $\|P(s(\mu); A)e - e\|$ is the optimal value for the next problem.

$$\min_{\Delta y \in \Re^m} \|S(\mu)^{-1} A^T \Delta y - e\|, \tag{31}$$

where $S(\mu) = \mathrm{diag}(s(\mu))$. Let $(\bar{y}, \bar{s})$ be a vertex of the dual problem. It is easy to show $\|\bar{s}\| \leq 2\bar{\chi}_A \|c\|$. Then we have

$$\begin{array}{ll} \|P(s(\mu); A) - e\| & \leq \|S(\mu)^{-1} A^T (\bar{y} - y(\mu)) - e\| \\ & = \|S(\mu)^{-1}(s(\mu) - \bar{s}) - e\| \\ & = \|S(\mu)^{-1} \bar{s}\| \\ & \leq \|S(\mu)^{-1}\| \|\bar{s}\| \\ & \leq (5\bar{\chi}_A \|c\|)^{-1} 2\bar{\chi}_A \|c\| \\ & = 2/5, \end{array}$$

where the last ienquality follows from

$$s(\mu) \geq \frac{1-\beta}{1+\beta} s \geq 5\bar{\chi}_A \|c\| e$$

using the assumption and Lemma 3.5. Therefore for any $i \in \{1, \dots, n\}$ we have

$$\begin{array}{ll} Rx_i^a(w(\mu)) & = [P(s(\mu); A)e]_i \\ & = [e + P(s(\mu); A)e - e]_i \\ & \geq 1 - \|P(s(\mu); A)e - e\| \\ & \geq 3/5, \end{array}$$

which proves the lemma. ∎

The next lemma is a technical result.

Lemma 3.8 *Let $I \subset \{1, \dots, n\}$, $J = \{1, \dots, n\} \setminus I$. Suppose $(\tilde{y}, \tilde{s}_I, \tilde{s}_J) \in \Re^m \times \Re^{|I|} \times \Re^{|J|}$ satisfies the equation*

$$\begin{array}{l} A_I^T \tilde{y} + \tilde{s}_I = c_I, \\ A_J^T \tilde{y} + \tilde{s}_J = c_J. \end{array} \tag{32}$$

Then the equation for (y, s_J)

$$\begin{array}{l} A_I^T y + \tilde{s}_I = c_I, \\ A_J^T y + s_J = c_J, \end{array} \tag{33}$$

has a solution $(\hat{y}, \hat{s}_J)$ with $\|\hat{s}_J\| \leq 2\bar{\chi}_A(\|c\| + \|\tilde{s}_I\|)$.

Proof. Let us remove redundant equations from $A_I^T \tilde{y} + \tilde{s}_I = c_I$. Let I' be the resulting index set. Let $A_K^T \in \Re^{|K| \times n}$ be a submatrix of A_J^T. We can choose A_K^T so that $\tilde{G} = [A_{I'}, A_K]^T$ is nonsingular. Let $\tilde{\beta} = [c_{I'}^T - \tilde{s}_{I'}^T, c_k^T]^T \in \Re^n$ It is easy to see that the pair $\hat{y} = \tilde{G}^{-1}\tilde{\beta}$ and $\hat{s}_J = c_I - A_J^T \hat{y}$ is a solution of the equation (33). Then we have

$$\begin{array}{ll} \|\hat{s}_J\| & = \|c_I - A_J^T \tilde{G}^{-1} \tilde{\beta}\| \\ & \leq \|c\| + \|(\tilde{G^T})^{-1} A\| \|\tilde{\beta}\| \\ & \leq \|c\| + \bar{\chi}_A(\|c\| + \|\tilde{s}_I\|) \\ & \leq 2\bar{\chi}_A(\|c\| + \|\tilde{s}_I\|), \end{array}$$

where the second inequality follows from Proposition 2.1, (b). ∎

Theorem 3.2 *Suppose $\beta \in (0, 1/4]$ and $w = w^k \in \mathcal{N}(\beta)$ is an iterate of the MTY P-C algorithm. Let $w(\mu) = (x(\mu), y(\mu), s(\mu))$ be the point on the central path associated with w. If $s_i \geq \frac{5(1+\beta)^2\bar{\chi}_A(50\sqrt{n}\bar{\chi}_A)^{n-1}\|c\|}{(1-\beta)^3}$ for some $i \in \{1, \ldots, n\}$, then $i \in B(w(\mu))$ or*

$$|Rx_i^a(w(\mu))| \geq |Rs_i^a(w(\mu))|.$$

Proof. Like the proof of Lemma 3.7, we show that $Rx_i^a(w(\mu)) \geq 1/2$. We may assume that there exists $j \in \{1, \ldots, n\}$ such that $s_j \leq \frac{5\bar{\chi}_A\|c\|(1+\beta)}{1-\beta}$, otherwise the conclusion readily follows from Lemma 3.7. Let us permutate the elements of $s(\mu)$ so that

$$s_{i_1}(\mu) \leq \cdots \leq s_{i_k}(\mu) = s_j \leq \cdots \leq s_{i_l}(\mu) = s_i \leq \ldots s_{i_n}(\mu).$$

Note that from Lemma 3.5,

$$\frac{s_i(\mu)}{s_j(\mu)} \geq \frac{(1-\beta)^2 s_j}{(1+\beta)s_i} \geq (50\sqrt{n}\bar{\chi}_A)^{n-1},$$

thus there exists an index $m < l$ such that $s_{i_{m+1}}(\mu)/s_{i_m}(\mu) \geq 50\sqrt{n}\bar{\chi}_A$. Define $I = \{i_1, \ldots i_m\}$, $J = \{i_{m+1}, \ldots, i_n\}$. Note that we may assume either $m + 1 = l$ or $s_{i_{m+1}} > \frac{5\bar{\chi}_A\|c\|(1+\beta)}{1-\beta}$. [1] In the case $m + 1 = l$, we have

$$s_{i_{m+1}} = s_{i_l} = s_i \geq \frac{5(1+\beta)^2\bar{\chi}_A(50\sqrt{n}\bar{\chi}_A)^{n-1}\|c\|}{(1-\beta)^3} > \frac{5\bar{\chi}_A\|c\|(1+\beta)}{1-\beta}.$$

Thus in any case we have

$$s_{i_{m+1}} > \frac{5\bar{\chi}_A\|c\|(1+\beta)}{1-\beta}, \tag{34}$$

$$s_{i_{m+1}}(\mu)/s_{i_m}(\mu) \geq 50\sqrt{n}\bar{\chi}_A, \tag{35}$$

and $i \in J$. We have

$$\begin{aligned} Rx_i^a(w(\mu)) &= [P(s(\mu)_J, T_{J2}^T)e_J + C_{I2}^T e_I + C_{J2}e_J]_i \\ &= [e_J + P(s(\mu)_J, T_{J2}^T)e_J - e_J + C_{I2}^T e_I + C_{J2}e_J]_i \\ &\geq 1 - \|P(s(\mu)_J, T_{J2}^T)e_J - e_J\| - \|C_{I2}^T e_I\| - \|C_{J2}e_J\|. \end{aligned}$$

In the following, we estimate the three norms. Note that the first norm is the optimal value for the next problem.

$$\begin{array}{ll} \min & \|e_J - S(\mu)^{-1}A_J^T\Delta y\| \\ \text{subject to} & A_I^T\Delta y = 0. \end{array} \tag{36}$$

Let $(\hat{y}, \hat{s}_J)$ be a vector implied by Lemma 3.8 for $(\tilde{y}, \tilde{s}) = (y(\mu), s(\mu))$. For this solution we have $\|\hat{s}_J\| \leq 2\bar{\chi}_A(\|c\| + \|s(\mu)_I\|)$. Since $\hat{y} - y(\mu)$ is a feasible solution of the problem (36), we have

$$\begin{aligned} \|P(s(\mu)_J, T_{J2}^T)e_J - e_J\| &\leq \|e_J - S(\mu)_J^{-1}A_J^T(\hat{y} - y(\mu))\| \\ &= \|e_J - S(\mu)_J^{-1}(s(\mu) - \hat{s})\| \\ &= \|S(\mu)_J^{-1}\hat{s}\| \\ &\leq \|S(\mu)_J^{-1}\|\|\hat{s}\| \\ &\leq \|S(\mu)_J^{-1}\|\{2\bar{\chi}_A(\|c\| + \|s(\mu)_I\|\} \\ &\leq 2\bar{\chi}_A\|c\|\|S(\mu)_J^{-1}\| + \sqrt{n}\|S(\mu)_J^{-1}\|\|S(\mu)_I\|\| \\ &\leq \tfrac{2(1-\beta)}{5(1+\beta)} + \tfrac{1}{50\bar{\chi}_A} \\ &\leq \tfrac{2}{5} + \tfrac{1}{50} = \tfrac{21}{50}, \end{aligned}$$

[1] If $m + 1 < l$ and $s_{i_{m+1}} \leq \frac{5\bar{\chi}_A\|c\|(1+\beta)}{1-\beta}$, we repeat the same procedure to obtain m' satisfying $s_{i_{m'+1}}(\mu)/s_{i'_m}(\mu) \geq 50\sqrt{n}\bar{\chi}_A$. This procedure terminates in a finite times, yielding m satisfying the assumption.

where the fifth inequality follows from (34) and (35). For the second norm, from Lemma 3.4 we have

$$\begin{aligned}\|C_{I2}^T e_I\| &\leq \|C_{I2}\|\|e_I\| \\ &\leq 2\sqrt{n}(1+\bar{\chi}_A^2\tau^2)\bar{\chi}_A\tau,\ \ \tau = \|S(\mu)_I\|\|S(\mu)_J^{-1}\| \\ &\leq 2\sqrt{n}(1+\bar{\chi}_A^2\cdot(\tfrac{1}{50\sqrt{n}\bar{\chi}_A})^2)\bar{\chi}_A(\tfrac{1}{50\sqrt{n}\bar{\chi}_A}) \\ &\leq 2(1+\tfrac{1}{2})\cdot\tfrac{1}{50} = \tfrac{3}{50}\end{aligned}$$

Similarly, for the third norm we have

$$\begin{aligned}\|C_{J2}^T e_J\| &\leq \|C_{J2}\|\|e_J\| \\ &\leq \sqrt{n}(4+3\bar{\chi}_A^2\tau^2)\bar{\chi}_A^2\tau^2,\ \ \tau = \|S(\mu)_I\|\|S(\mu)_J^{-1}\| \\ &\leq \sqrt{n}(4+3\bar{\chi}_A^2\cdot(\tfrac{1}{50\sqrt{n}\bar{\chi}_A})^2)\bar{\chi}_A^2(\tfrac{1}{50\sqrt{n}\bar{\chi}_A})^2 \\ &\leq (4+1)\cdot\tfrac{1}{2500} = \tfrac{1}{500}.\end{aligned}$$

Thus we conclude

$$\begin{aligned}Rx_i^a(w(\mu)) &\geq 1 - \|P(s(\mu)_J, T_{J2}^T)e_J - e_J\| - \|C_{I2}^T e_I\| - \|C_{J2}e_J\| \\ &\geq 1 - \tfrac{21}{50} - \tfrac{3}{50} - \tfrac{1}{500} \\ &= \tfrac{259}{500} > \tfrac{1}{2},\end{aligned}$$

which proves the theorem. ■

4 A new convergence analysis of MTY P-C algorithm

In this section, we provide the proof of Theorem 2.1.

Our analysis in this section is strongly motivated by the following result which is proved by Monteiro and Tsuchiya[7].

Lemma 4.1 *Consider the set $\mathcal{K}$ consisting of those indices k such that the iterates w^k and w^{k+1} of the MTY P-C algorithm satisfy the following set of conditions:*

C1) $\text{gap}(\mathcal{J}_k) > 2\bar{g}$ *and* $\epsilon_\infty^a < \tau\bar{g}/(\sqrt{n}\text{gap}(\mathcal{J}_k))$;

C2) $\text{gap}(\mathcal{J}_{k+1} = \delta_i(w^{k+1})/\delta_j(w^{k+1})$ *for some* $i \in N(w^{k+1})$ *and* $j \in B(w^{k+1})$;

C3) $(B(w^{k+1}, N(w^{k+1})) = (B(w^k), N(w^k))$ *and* $\mathcal{J}_{k+1} = \mathcal{J}_k$;

Then if $k \in \mathcal{K}$, we have

$$\text{gap}(\mathcal{J}_{k+1}) \geq 2\ \text{gap}(\mathcal{J}_k), \tag{37}$$

otherwise there exists an iteration index $l > k$ such that $l - k = \mathcal{O}(n^{1.5}\log(\bar{\chi}_A + n))$ and with the property that, either a C-crossover event occurs between w^k and w^l, or the algorithm terminates at or before the l-th iteration.

The following result is from Monteiro and Tsuchiya[7].

Lemma 4.2 *Suppose $\beta \in (0, 1/4]$ and w^k, $w^{k+1} \in \mathcal{N}(\beta)$ are iterates of the MTY P-C algorithm. Let $w : [0,1] \to \mathcal{N}(2\beta)$ be a contituous path such that $w(0) = w^k$, $w(1) = w^{k+1}$ and $\mu(w^{k+1}) \leq \mu(w(t)) \leq \mu(w^k)$. If there exist $t \in [0,1]$ such that $B(w^k) \cap N(w(t)) \neq \emptyset$ or $N(w^k) \cap B(w(t)) \neq \emptyset$, there exists an iteration index $l > k$ such that $l - k = \mathcal{O}(n^{1.5}\log(\bar{\chi}_A + n))$ and with the property that, either a C-crossover event occurs between w^k and w^l, or the algorithm terminates at or before the l-th iteration.*

Let us consider a path which passes through the central path point $w(\mu^k)$ associated with a current point w^k. From Lemma 4.2, if the AS bipartition of the current point w^k is different from that of $w(\mu^k)$, a C-crossover event occurs or the algorithm terminates in $\mathcal{O}(n^{1.5}\log(\bar{\chi}_A + n))$ iterations.

4.1 Bounded case

We first deal with the case where the dual feasible region is bounded, as assumed in A.4. The main result in this subsection is based on Lemma 3.1, which states that there is a lower bound for N variables and it depends on A and c. This fact, combined with the assumption A.4, which guarantees the boundedness of the dual feasible region, means that if the current point w^k satisfies $k \in \mathcal{K}$, the gap of AS partition is bounded above by a constant depending on A and c. Then from (37), the number of consecutive indices in $\mathcal{K}$ depends only on A and c.

Lemma 4.3 *Suppose $\beta \in (0, 1/4]$ and $w = w^k \in \mathcal{N}(\beta)$ is an iterate of the modified MTY P-C algorithm. Let $\zeta_{(A,c)}$ be a constant defined by (9). Then there exists an iteration index $l > k$ such that*

$$l - k = O(n^{1.5} \log(\bar{\chi}_A + n) + \log \zeta_{(A,c)}) \tag{38}$$

and with the property that, either a C-crossover event occurs between w^k and w^l, or the algorithm terminates at or before the l-th iteration.

Proof. Define $\mathcal{M}$ as the set of indices k such that the AS bipartition of w^{k+1} coincides with that of $w(\mu^{k+1})$. Let $\mathcal{K}' = \mathcal{K} \cap \mathcal{M}$. First, from Lemma 4.1 and Lemma 4.2, if $k \notin \mathcal{K}'$, there exists an index l with $l - k = O(n^{1.5} \log(\bar{\chi}_A + n)$ satisfying the property stated earlier.

First we consider the case where there are two layers in the AS partition $\mathcal{J}_h$. In this case, we know that a C-crossover event occurs in $\mathcal{O}(n^{1.5} \log(\bar{\chi}_A + n))$ iterations, since we conduct the finite termination procedure at each iteration and we use, in addition to the AS direction, the finite termination direction as a search direction.

Next we address the case where there are more than two layers in the current AS partition $\mathcal{J}(w^k)$. Assume $k \in \mathcal{K}'$ and $k, k+1, \ldots, l'-1$ be consecutive indices in $\mathcal{K}'$. For $w^h = (x^h, y^h, s^h)$ $(h = k+1, \ldots, l')$ and $j \in \{1, \ldots, n\}$, assume $s_j^h \in N(w^h)$. Let $w(\mu^h)$ be the central path point associated with w^h. If

$$s_j^h < \frac{(1-\beta)\delta(A,c)}{\sqrt{n}(5\bar{\chi}_A\sqrt{n})^n},$$

we have $j \in B(w(\mu^h))$ from Theorem 4.1. Since $h \in \mathcal{M}$, we also have $j \in B(w(\mu^h)) = B(w^h)$, which is a contradiction, as $j \in N(w^h)$. Thus we have

$$s_j^h \geq \frac{(1-\beta)\delta(A,c)}{\sqrt{n}(5\bar{\chi}_A\sqrt{n})^n}. \tag{39}$$

Since the dual feasible region is bounded, we also have

$$\|s^h\| \leq (1 + \bar{\chi}_A)\|c\|,$$

which implies

$$s_j^h \leq (1 + \bar{\chi}_A)\|c\|. \tag{40}$$

Note that this inequality follows if $j \in B(w^h)$. Let $\tilde{B}$ and $\tilde{N}$ be the layers of the AS partition $\mathcal{J}_h$ satisfying

$$\mathrm{gap}(\mathcal{J}_h) = \frac{\min_{k \in \tilde{N}} \delta_k^h}{\max_{k \in \tilde{B}} \delta_k^h}$$

There are two possibilities for the structure of the AS partition $\mathcal{J}_h$. That is, there is a layer below $\tilde{B}$ or not. For the former case, let $\hat{N}$ be a layer of the AS partition $\mathcal{J}_h$ which satisfies

$$\max_{k\in\hat{N}} \delta^h \leq \min_{k\in\tilde{B}} \delta^h_k.$$

We have

$$\begin{aligned}
\mathrm{gap}(\mathcal{J}_h) &= \frac{\min_{k\in\tilde{N}} \delta^h_k}{\max_{k\in\tilde{B}} \delta^h_k} \\
&\leq \frac{\min_{k\in\tilde{B}} \delta^h_k}{\max_{k\in\hat{N}} \delta^k_h} \\
&\leq \frac{(1+\beta)^2}{(1-\beta)^4} \frac{\min_{k\in\hat{N}} s^k_h}{\max_{k\in\tilde{B}} s^k_h} \\
&\leq \frac{(1+\beta)^2}{(1-\beta)^4} \{(1+\bar{\chi}_A\|c\|\}/\{\frac{(1-\beta)\delta(A,c)}{\sqrt{n}(5\bar{\chi}_A\sqrt{n})^n}\} \\
&= \frac{(1+\beta)^2(1+\bar{\chi}_A)(5\bar{\chi}_A)^n n^{(n+1)/2}\|c\|}{(1-\beta)^5\delta(A,c)},
\end{aligned}$$

where the first inequality follows from the definition of $\mathrm{gap}(\mathcal{J}_h)$, the second inequality follows from Lemma 4.5, and the last inequality follows from (39) and (40). Thus we have

$$\mathrm{gap}(\mathcal{J}_h) \leq \frac{(1+\beta)^2(1+\bar{\chi}_A)(5\bar{\chi}_A)^n n^{(n+1)/2}\|c\|}{(1-\beta)^5\delta(A,c)}. \tag{41}$$

By a similar argument, we can prove this inequality for the case there is no layer below $\tilde{B}$.

Now from $\mathrm{gap}(\mathcal{J}_k) > 2\bar{g}$ and (37) of Lemma 4.1, we have

$$\mathrm{gap}(\mathcal{J}_{\bar{l}}) \geq 2^{\bar{l}-k+1}\bar{g}. \tag{42}$$

Thus (41) combined with (48) leads to

$$2^{\bar{l}-k+1}\bar{g} \leq \frac{(1+\beta)^2(1+\bar{\chi}_A)(5\bar{\chi}_A)^n n^{(n+1)/2}\|c\|}{(1-\beta)^5\delta(A,c)},$$

which implies

$$\bar{l} - k = \mathcal{O}(n\log(\bar{\chi}_A + n) + \log\frac{\|c\|}{\delta(A,c)}) \tag{43}$$

As $\bar{l} \notin \mathcal{K}'$, a C-crossover event occurs or the algorithm terminates in $(\mathcal{O}(n^{1.5}\log(\bar{\chi}_A + n))$ iterations. ■

4.2 Unbounded case

In this subsection, by modifying our algorithm, we show that the same complexity can be obtained in the case where the dual feasible region is unbounded.

In our modified MTY-PC-U algorithm, we use another step direction, whose derivation is similar to the one used in the finite termination procedure. Suppose that the AS partition $\mathcal{J}(w)$ at a point $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ consists of three layers $\mathcal{J}_1,\ \mathcal{J}_2,\ \mathcal{J}_3$. Assume also that $\mathcal{J}_1,\ \mathcal{J}_3 \subset B(w)$ and $\mathcal{J}_2 \subset N(w)$, where $(B(w), N(w))$ is the AS-bipartition at w.

We first explain the primal direction. Let $M^x_4 = \ker(A)$. For $k = 3, 2, 1$ define

$$M^x_k = \arg\min_{\Delta x\in M^x_{k+1}} \|\delta_{\mathcal{J}_k}(x_{\mathcal{J}_k} + \Delta x_{\mathcal{J}_k})\| \tag{44}$$

Then we use a vector $\Delta x^l = \Delta x^l_3 \in M^x_1$ for the primal direction.

We move on to the dual direction. Let $M_0^s = \mathrm{Im}(A^T)$. For $k = 1, 2, 3$ define

$$M_k^s = \arg \min_{\Delta x \in M_{k-1}^s} \|\delta_{\mathcal{J}_k}^{-1}(s_{\mathcal{J}_k} + \Delta s_{\mathcal{J}_k})\| \tag{45}$$

Let us choose a vector $\Delta s_3^l \in M_3^s$ and compute Δy_3^l from the relation $A^T \Delta y_3^l + \Delta s_3^l = 0$. We use $(\Delta y^l, \Delta s^l) = (\Delta y_3^l, \Delta s_3^l)$ for the dual direction. We call the direction $\Delta w^l = (\Delta x^l, \Delta y^l, \Delta s^l)$ three-layered direction.

With the three-layered direction, MTY-PC-U algorithm is described as follows.

Modified MTY-PC-U algorithm

Let $\beta \in (0, \frac{1}{4}]$ and $w^0 \in \mathcal{N}(\beta)$ be given. Set $k = 0$.

1. Set $w = w^k$ and compute the AS direction $\Delta w^a = (\Delta x^a, \Delta y^a, \Delta s^a)$ at w.
2. Compute the AS bipartition (B, N) according to (15).
3. Compute the FT direction $\Delta w^f = (\Delta x^f, \Delta y^f, \Delta s^f)$ and conduct the FT procedure. If the procedure succeed, the algorithm stops. Output the optimal solution $\tilde{w}$.
4. Let $w^a = w + \alpha^a \Delta^a$, where $\alpha^a = \sup\{\alpha \in [0, 1] : \ w + \alpha \Delta^a \in N(2\beta)\}$.
5. Let $w^f = w + \alpha^f \Delta^f$, where $\alpha^f = \sup\{\alpha \in [0, 1] : \ w + \alpha \Delta^f \in N(2\beta)\}$.
6. If the AS partition $\mathcal{J}_k = (B^1, N, B^2)$, compute the three-layered direction Δw^l and let $w^l = w + \alpha^f \Delta^f$, where $\alpha^l = \sup\{\alpha \in [0, 1] : \ w + \alpha \Delta^l \in N(2\beta)\}$. Otherwise set $w^l = w$.
7. Set $w = w^q$, where $q = \arg\min_{q \in \{a,f,l\}} \mu(w^q)$.
8. If $\mu(w) = 0$, the algorithm stops. In this case, w is an optimal solution.
9. Compute the corrector step Δw^c at w and set $w \leftarrow w + \Delta w^c$.
10. Set $w^{k+1} = w$, increase k by one and go to step 1.

For the modified MTY-PC-U algorithm, we have the same result as the modified MTY-PC algorithm, as the next lemma suggests.

Lemma 4.4 *Suppose $\beta \in (0, 1/4]$ and $w = w^k \in \mathcal{N}(\beta)$ is an iterate of the modified MTY PC-U algorithm. Let $\zeta_{(A,c)}$ be a constant defined by (9). Then there exists an iteration index $l > k$ such that*

$$l - k = O(n^{1.5} \log(\bar{\chi}_A + n) + \log \zeta_{(A,c)}) \tag{46}$$

and with the property that, either a C-crossover event occurs between w^k and w^l, or the algorithm terminates at or before the l-th iteration.

Proof. Define $\mathcal{M}$ as the set of indices k such that the AS bipartition of w^{k+1} coincides with that of $w(\mu^{k+1})$. Let $\mathcal{K}' = \mathcal{K} \cap \mathcal{M}$. First, from Lemma 4.1 and Lemma 4.2, if $k \notin \mathcal{K}'$, there exists an index l with $i - k = O(n^{1.5} \log(\bar{\chi}_A + n)$ satisfying the property stated earlier.

We first address the case where there are more than one N layers in the current AS partition $\mathcal{J}(w^k)$. Assume $k \in \mathcal{K}'$ and $k, k+1, \ldots, l'-1$ be consecutive indices in $\mathcal{K}'$. For $w^h = (x^h, y^h, s^h)$ $(h = k+1, \ldots, l')$ and $j \in \{1, \ldots, n\}$, assume $s_j^h \in N(w^h)$. Let $w(\mu^h)$ be the central path point associated with w^h. By a similar argument in the proof of Lemma 5.3, we have

$$\frac{(1-\beta)\delta(A,c)}{\sqrt{n}(5\bar{\chi}_A\sqrt{n})^n} \leq s_j^h \leq \frac{5(1+\beta)^2\bar{\chi}_A(50\sqrt{n}\bar{\chi}_A)^n\|c\|}{(1-\beta)^3}. \tag{47}$$

Take N^1 and N^2 from N layers of the AS partition $\mathcal{J}_j$ satisfying $\max_{j\in N^1}\delta_j^h \leq \min_{j\in N^2}\delta_j^h$. From the definition of gap($\mathcal{J}_h$), we have

$$\begin{aligned} \text{gap}(\mathcal{J}_h) &\leq \frac{\min_{j\in N^2}\delta_j^j}{\max_{j\in N_1}\delta_h^j} \\ &\leq \frac{(1+\beta)^2}{(1-\beta)^4}\frac{\min_{j\in N^2} s_h^j}{\max_{j\in N^1} s_h^j} \\ &\leq \frac{5(1+\beta)^4\sqrt{n}(250n\bar{\chi}_A^2)^n\|c\|}{(1-\beta)^8\delta(A,c)}, \end{aligned}$$

where the last inequality follows from (47). Thus we have

$$\text{gap}(\mathcal{J}_h) \leq \frac{5(1+\beta)^4\sqrt{n}(250n\bar{\chi}_A^2)^n\|c\|}{(1-\beta)^8\delta(A,c)}. \tag{48}$$

Now from gap($\mathcal{J}_k$) $> 2\bar{g}$ and (37) of Lemma 4.1, we have

$$\text{gap}(\mathcal{J}_{\bar{l}}) \geq 2^{\bar{l}-k+1}\bar{g}. \tag{49}$$

Thus (48) combined with (49) leads to

$$2^{\bar{l}-k+1}\bar{g} \leq \frac{5(1+\beta)^4\sqrt{n}(250n\bar{\chi}_A^2)^n\|c\|}{(1-\beta)^8\delta(A,c)},$$

which implies

$$\bar{l}-k = \mathcal{O}(n\log(\bar{\chi}_A+n) + \log\frac{\|c\|}{\delta(A,c)}) \tag{50}$$

As $\bar{l} \notin \mathcal{K}'$, a C-crossover event occurs or the algorithm terminates in $(\mathcal{O}(n^{1.5}\log(\bar{\chi}_A+n))$ iterations.

Next we consider the case where there are only one N layer in the AS partition $\mathcal{J}_h$. There are two possibilities of the AS partition $\mathcal{J}_h$, $(B,\ N)$ or $(B^1,\ N,\ B^2)$ (Note: the value of δ_h^j increases from left to right). In both cases, we know that a C-crossover event occurs in $\mathcal{O}(n^{1.5}\log(\bar{\chi}_A+n))$ iterations, since we use, in addition to the AS direction, the finite termination direction and the three-layered direction as a search direction. ■

5 Conclusion

In this paper, we developed a simple finite-termination variant of the Mizuno-Todd-Ye predictor-corrector algorithm with objective-function-free polynomial-time complexity for linear programs with bounded feasible region. We also demonstrated that three layer is enough to obtain the same complexity for general linear programs. An interesting open problem is whether these algorithms have polynomial-time complexity which just depends on the coefficient matrix A.

A The proof of Lemma 3.4

Suppose $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++},\ I \subset \{1, \ldots, n\},\ J = \{1, \ldots, n\} \setminus I$. Let

$$\begin{aligned} A^T &= TG \\ &= \begin{bmatrix} T_{I1} & O \\ T_{J1} & T_{J2} \end{bmatrix} G, \end{aligned}$$

where $T \in \Re^{n\times m}$ is a basis block lower triangular matrix and $G \in \Re^{m\times m}$ is a nonsingular matrix.

Recall T_{I1} has the identity matrix after appropriate row permutations. Therefore we have $T_{J1} = VT_{I1}$ for a matrix $V \in \Re^{|J|\times|I|}$, which consists of column vectors of T_{J1} and zero vectors. Notice $\|V\| \le \|T_{J1}\|$. Thus we have

$$T = \begin{bmatrix} T_{I1} & O \\ VT_{I1} & T_{J2} \end{bmatrix},$$

and

$$S^{-1}T = \begin{bmatrix} S_I^{-1}T_{I1} & O \\ S_J^{-1}T_{J1} & S_J^{-1}T_{J2} \end{bmatrix} = \begin{bmatrix} B & O \\ WB & M \end{bmatrix},$$

where

$$B = S_I^{-1}T_{I1},\ T_{J1} = VT_{I1}, W = S_J^{-1}VS_I,\ M = S_J^{-1}T_{J2}.$$

As $\|V\| \le \|T_{J1}\|$ holds, it is easy to show the following Lemma.

Lemma A.1 *The following relations hold:*

(1) $\|W\| \le \|S_I\|\|S_J^{-1}\|\|T_{J1}\|$.

(2) $\|S_J WB\| = \|T_{J1}\|$.

(3) $\|WB\| = \|S_J^{-1}T_{J1}\| \le \|S_J^{-1}\|\|T_{J1}\|$.

(4) $\|S_J W\| \le \|S_I\|\|T_{J1}\|$.

Lemma A.2 *The following inequality holds:*

$$\|T_{J1}\| \le \bar{\chi}_A. \tag{51}$$

Proof. We can prove the lemma using the properties of $\bar{\chi}_A$ (Proposition 2.1) as follows:

$$\begin{aligned} \|T_{J1}\| = \|T_{J1}^T\| &\le \bar{\chi}_{[T_{I1}^T, T_{J1}^T]} \text{ (from (c))} \\ &\le \bar{\chi}_{T^T} \text{ (from (d))} \\ &= \bar{\chi}_A \text{ (from (a))} \end{aligned}$$

Proof of Lemma 3.4. We have

$$T^T S^{-2} T := \begin{bmatrix} F & G \\ G^T & H \end{bmatrix} = \begin{bmatrix} B^T(I + W^T W)B & B^T W^T M \\ M^T WB & M^T M \end{bmatrix}.$$

From the inversion formula for nonsingular block matrices, it follows that

$$(T^T S^{-2} T)^{-1} = \begin{bmatrix} \tilde{A} & \tilde{B}^T \\ \tilde{B} & \tilde{C} \end{bmatrix} = \begin{bmatrix} L^{-1} & -L^{-1}GH^{-1} \\ -H^{-1}G^T L^{-1} & H^{-1} + H^{-1}G^T L^{-1}GH^{-1} \end{bmatrix},$$

where $L = F - GH^{-1}G^T$.

Next transform $\tilde{A}$ of $(T^TS^{-2}T)^{-1}$ as follows:

$$\begin{aligned}
\tilde{A} &= L^{-1} \\
&= (F - GH^{-1}G^T)^{-1} \\
&= (B^T(I + W^TW)B - B^TW^TM(M^TM)^{-1}M^TWB)^{-1} \\
&= (B^T(I + W^TQ_MW)B)^{-1} \\
&= (B^TB + B^TW^TQ_MQ_MWB)^{-1} \\
&= (B^TB)^{-1} - (B^TB)^{-1}B^TW^TQ_M(I + Q_MWP_BW^TQ_M)^{-1}Q_MWB(B^TB)^{-1},
\end{aligned} \tag{52}$$

where we set $P_B = B(B^TB)^{-1}B^T$, $P_M = M(M^TM)^{-1}M^T$, $Q_M = I - P_M$ and the sixth equality follows from the Sherman-Morrison-Woodbury formula.

Also we have

$$\tilde{B} = -H^{-1}G^TL^{-1} = -(M^TM)^{-1}M^TWB\tilde{A}. \tag{53}$$

By using the equality (52), we further obtain the following:

$$\begin{aligned}
\tilde{C} &= H^{-1} + H^{-1}G^TL^{-1}GH^{-1} \\
&= (M^TM^{-1}) + (M^TM)^{-1}M^TWP_B\times \\
&\quad (I - P_BW^TQ_M(I + Q_MWP_BW^TQ_M)^{-1}Q_MWP_B)P_BW^TM^T(M^TM)^{-1} \\
&= (M^TM)^{-1} + (M^TM)^{-1}M^TWP_B(I + P_BW^TQ_MWP_B)^{-1}P_BW^TM(M^TM)^{-1}.
\end{aligned}$$

We used the Sherman-Morrison-Woodbury formula for the last equality. With the above equalities combined, we have

$$\begin{aligned}
S^{-1}T(TS^{-2}T^T)^{-1}T^TS^{-1} &= \begin{bmatrix} B & O \\ WB & M \end{bmatrix} \begin{bmatrix} \tilde{A} & \tilde{B}^T \\ \tilde{B} & \tilde{C} \end{bmatrix} \begin{bmatrix} B^T & B^TW^T \\ O & M^T \end{bmatrix} \\
&:= \begin{bmatrix} H_{I1} & H_{I2} \\ H_{I2}^T & H_{J2} \end{bmatrix},
\end{aligned}$$

where

$$\begin{aligned}
&H_{I1} = B\tilde{A}B^T, \\
&H_{I2} = B\tilde{A}B^TW^T + B\tilde{B}^TM^T, \\
&H_{J2} = WB\tilde{A}B^TW^T + M\tilde{B}B^TW^T + WB\tilde{B}^TM^T + M\tilde{C}M^T.
\end{aligned}$$

In the following, we evaluate each element of the projection matrix. First let us consider H_{I1}. We have

$$\begin{aligned}
H_{I1} &= B\tilde{A}B^T \\
&= B(B^TB)^{-1}B^T \\
&\quad -B(B^TB)^{-1}B^TW^TQ_M(I + Q_MWP_BW^TQ_M)^{-1}Q_MWB(B^TB)^{-1}B^T.
\end{aligned}$$

From the definition of B, the first term becomes $S_I^{-1}T_{I1}(T_{I1}S_I^{-2}T_{I11})^{-1}T_{I1}S_I^{-1}$, which is the projection matrix $P(s_I; T_{I1}^T)$. As for the second term, we can bound its norm as follows:

$$\begin{aligned}
& \|B(B^TB)^{-1}B^TW^TQ_M(I + Q_MWP_BW^TQ_M)^{-1}Q_MWB(B^TB)^{-1})B^T\| \\
\le\ & \|P(s_I; T_{I1}^T)\|\|W^T\|\|Q_M\|\|(I + Q_MWP_BW^TQ_M)^{-1}\|\|Q_M\|\|W\|\|P(s_I; T_{I1}^T)\| \\
\le\ & 1 \cdot \tau\|T_{J1}\| \cdot 1 \cdot 1 \cdot 1 \cdot \tau\|T_{J1}\| \cdot 1 \\
=\ & \|T_{J1}\|^2\tau^2,
\end{aligned}$$

where $\tau := \|S_I\|\|S_J^{-1}\|$, and we have used Lemma A.1.

Then let us go on to H_{I2}. First notice that the norm of the H_{I1} is bounded above by $1 + \|T_{J1}\|^2\tau^2$. Therefore, for the first term of H_{I2}, we have

$$\begin{aligned}\|B\tilde{A}B^TW^T\| &= \|H_{I1}W^T\| \\ &\leq \|H_{I1}\|\|W^T\| \\ &\leq (1 + \|T_{J1}\|^2\tau^2)\|T_{J1}\|\tau.\end{aligned}$$

We can bound the second term as follows:

$$\begin{aligned}\|B\tilde{B}^TM^T\| &= \|B\tilde{A}B^TW^TM(M^TM)^{-1}M^T\| \\ &\leq \|H_{I1}\|\|W^T\|\|M(M^TM)^{-1}M^T\| \\ &\leq (1 + \|T_{J1}\|^2\tau^2)\|T_{J1}\|\tau.\end{aligned}$$

Overall, $\|H_{I2}\|$ is bounded above by $2(1 + \|T_{J1}\|^2\tau^2)\|T_{J1}\|\tau$.

Finally we address H_{J2}. As to the first term, we have

$$\begin{aligned}\|WB\tilde{A}B^TW^T\| &\leq \|W\|\|H_{I1}\|\|W^T\| \\ &\leq (1 + \|T_{J1}\|^2\tau^2)\|T_{J1}\|^2\tau^2.\end{aligned}$$

Note that the second term is the transpose of the third term. Thus using the result of the previous paragraph, we can bound each term as follow:

$$\begin{aligned}\|WB\tilde{B}^TM^T\| &\leq \|W\|\|B\tilde{B}^TM^T\| \\ &\leq (1 + \|T_{J1}\|^2\tau^2)\|T_{J1}\|^2\tau^2.\end{aligned}$$

Then we handle the fourth term.

We have

$$\begin{aligned}M\tilde{C}M^T &= M(M^TM)^{-1}M^T + \\ &\quad M(M^TM)^{-1}M^TWP_B(I + P_BW^TQ_MWP_B)^{-1}P_BW^TM(M^TM)^{-1}M^T\end{aligned}$$

Notice from the definition of M,

$$M(M^TM)^{-1}M^T = S_J^{-1}T_{J2}(T_{J2}^TS_J^{-2}T_{J2})^{-1}T_{J2}^TS_J^{-1} = P(s_J; T_{J2}^T).$$

For the rest, we have

$$\begin{aligned}&\|M(M^TM)^{-1}M^TWP_B(I + P_BW^TQ_MWP_B)^{-1}P_BW^TM(M^TM)^{-1}M^T\| \\ \leq\;& \|P(s_J; T_{J2})\|\|W\|\|P_B\|\|(I + P_BW^TQ_MWP_B)^{-1}\|\|P_B\|\|W^T\|\|P(s_J; T_{J2})\| \\ \leq\;& \|T_{J1}\|^2\tau^2\end{aligned}$$

As a result, we have shown the following:

$$\|H_{J2} - P(s_J; T_{J2}^T)\| \leq (4 + 3\|T_{J1}\|^2\tau^2)\|T_{J1}\|^2\tau^2.$$

To summarize, we have established the folloing relation:

$$S^{-1}T(T^TD^2T)^{-1}T^TD = \begin{bmatrix} P(s_I; T_{I1}^T) & O \\ O & P(s_J; T_{J2}^T) \end{bmatrix} + \begin{bmatrix} C_{I1} & C_{I2} \\ C_{I2}^T & C_{J2} \end{bmatrix}, \tag{54}$$

where

$$\begin{aligned}&\|C_{I1}\| \leq \|T_{J1}\|^2\tau^2, \\ &\|C_{I2}\| \leq 2(1 + \|T_{J1}\|^2\tau^2)\|T_{J1}\|\tau, \\ &\|C_{J2}\| \leq (4 + 3\|T_{J1}\|^2\tau^2)\|T_{J1}^2\|\tau^2.\end{aligned}$$

From Lemma A.2, we have $\|T_{J1}\| \leq \bar{\chi}_A$. Thus we obtain the desired inequality for the error terms.

References

[1] I. I. Dikin: On the speed of an iterative process. *Upravlyaemye Sistemi*, Vol. 12 (1974), pp. 54-60.

[2] C. C. Gonzaga and H. J. Lara: A note on properties of condition numbers. *Linear Algebra and its Applications*. Vol. 261 (1997), pp. 269–273.

[3] T. Kitahara and T. Tsuchiya: Proximity of Weighted and Layered Least Squares Solutions. *SIAM Journal on Matrix Analysis and Applications*, Vol. 31 (2009), pp.1172-1186.

[4] N. Megiddo, S. Mizuno and T. Tsuchiya: A modified layered-step interior-point algorithm for linear programming. *Mathematical Programming*, Vol. 82 (1998), pp. 339-355.

[5] S. Mizuno, M. J. Todd, and Y. Ye, On adaptive-step primal-dual interior-point algorithms for linear programming. *Mathematics of Operations Research*, Vol. 18 (1993), pp. 964–981.

[6] R. D. C. Monteiro and T. Tsuchiya. A variant of the Vavasis-Ye layered-step interior-point algorithm for linear programming. *SIAM Journal on Optimization*, Vol. 13 (2003), pp. 1054-1079.

[7] R.D.C. Monteiro and T. Tsuchiya: A new iteration-complexity bound for the MTY predictor-corrector algorithm, *SIAM Journal on Optimization* 15 (2004) 319-347.

[8] G. W. Stewart: On scaled projections and pseudoinverses, *Linear Algebra and its Applications*, Vol. 112 (1989) pp. 189–193.

[9] É. Tardos: A strongly polynomial algorithm to solve combinatorial linear programs. *Operations Research* Vol. 34, (1986) pp. 250–256.

[10] M. J. Todd: A Dantzig-Wolfe-like variant of Karmarkar's interior-point linear programming algorithm. *Operations Research*, Vol. 38 (1990) pp. 1006–1018.

[11] M. J. Todd, L. Tunçel and Y. Ye: Characterizations, bounds, and probabilistic analysis of two complexity measures for linear programming problems. *Mathematical Programming*, Vol. 90 (2001), No. 1, pp. 59–70.

[12] T. Tsuchiya: Global convergence property of the affine scaling method for primal degenerate linear programming problems. *Mathematics of Operations Research*, Vol. 17, (1992) pp. 527–557.

[13] S. Vavasis and Y. Ye: A primal-dual accelerated interior point method whose running time depends only on A, *Mathematical Programming* 74 (1996) 79-120.