

# A new iteration-complexity bound for the MTY predictor-corrector algorithm

Renato D. C. Monteiro\*      Takashi Tsuchiya†

October 30, 2002

## Abstract

In this paper we present a new iteration-complexity bound for the Mizuno-Todd-Ye predictor-corrector (MTY P-C) primal-dual interior-point algorithm for linear programming. The analysis of the paper is based on the important notion of crossover events introduced by Vavasis and Ye. For a standard form linear program  $\min\{c^T x : Ax = b, x \geq 0\}$  with decision variable  $x \in \mathbb{R}^n$ , we show that the MTY P-C algorithm started from a well-centered interior-feasible solution with duality gap  $n\mu_0$  finds an interior-feasible solution with duality gap less than  $n\eta$  in  $\mathcal{O}(n^2 \log(\log(\mu_0/\eta)) + n^{3.5} \log(\bar{\chi}_A^* + n))$  iterations, where  $\bar{\chi}_A^*$  is a scaling invariant condition number associated with the matrix  $A$ . More specifically,  $\bar{\chi}_A^*$  is the infimum of all the condition numbers  $\bar{\chi}_{AD}$ , where  $D$  varies over the set of positive diagonal matrices. Under the setting of the Turing machine model, our analysis yields an  $\mathcal{O}(n^{3.5}L_A + n^2 \log L)$  iteration-complexity bound for the MTY P-C algorithm to find a primal-dual optimal solution, where  $L_A$  and  $L$  are the input sizes of the matrix  $A$  and the data  $(A, b, c)$ , respectively. This contrasts well with the classical iteration-complexity bound for the MTY P-C algorithm which depends linearly on  $L$  instead of  $\log L$ .

**Key words:** Interior-point algorithms, primal-dual algorithms, path-following, central path, layered steps, condition number, polynomial complexity, crossover events, scale-invariance, predictor-corrector, affine scaling, strongly polynomial, linear programming.

## 1 Introduction

We consider the LP problem

$$\begin{aligned} & \text{minimize}_x && c^T x \\ & \text{subject to} && Ax = b, \quad x \geq 0, \end{aligned} \tag{1}$$

---

\*School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA (Email: [monteiro@isye.gatech.edu](mailto:monteiro@isye.gatech.edu)). This author was supported in part by NSF Grants CCR-9902010, CCR-0203113 and INT-9910084.

†The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-Ku, Tokyo, 106-8569, Japan. (Email: [tsuchiya@sun312.ism.ac.jp](mailto:tsuchiya@sun312.ism.ac.jp)). This author was supported in part by Japan-US Joint Research Projects of Japan Society for the Promotion of Science “Algorithms for linear programs over symmetric cones” and the Grant-in-Aid for Scientific Research (C) 08680478 of the Ministry of Science, Technology, Education and Culture of Japan.

and its associated dual problem

$$\begin{aligned} & \text{maximize}_{(y,s)} && b^T y \\ & \text{subject to} && A^T y + s = c, \quad s \geq 0, \end{aligned} \tag{2}$$

where  $A \in \mathfrak{R}^{m \times n}$ ,  $c \in \mathfrak{R}^n$  and  $b \in \mathfrak{R}^m$  are given, and the vectors  $x, s \in \mathfrak{R}^n$  and  $y \in \mathfrak{R}^m$  are the unknown variables.

Karmarkar in his seminal paper [5] proposed the first polynomially convergent interior-point method with an  $\mathcal{O}(nL)$  iteration-complexity bound, where  $L$  is the size of the LP instance (1). The first path-following interior-point algorithm was proposed by Renegar in his breakthrough paper [19]. Renegar’s method closely follows the primal central path and exhibits an  $\mathcal{O}(\sqrt{n}L)$  iteration-complexity bound. The first path following algorithm which simultaneously generates iterates in both the primal and dual spaces has been proposed by Kojima, Mizuno and Yoshise [6] and Tanabe [21], based on ideas suggested by Megiddo [9]. In contrast to Renegar’s algorithm, Kojima et al.’s algorithm has an  $\mathcal{O}(nL)$  iteration-complexity bound. A primal-dual path following algorithm with an  $\mathcal{O}(\sqrt{n}L)$  iteration-complexity bound was subsequently obtained by Kojima, Mizuno and Yoshise [7] and Monteiro and Adler [14, 15] independently. Following these developments, many other primal-dual interior-point algorithms for linear programming have been proposed.

An outstanding open problem in optimization is whether there exists a strongly polynomial algorithm for linear programming, that is one whose complexity is bounded by a polynomial of  $m$  and  $n$  only. A major effort in this direction is due to Tardos [22] who developed a polynomial-time algorithm whose complexity is bounded by a polynomial of  $m$ ,  $n$  and  $L_A$ , where  $L_A$  denotes the size of the (integral) matrix  $A$ . Such an algorithm gives a strongly polynomial method for the important class of linear programming problems where the entries of  $A$  are either 1,  $-1$  or 0, e.g., LP formulations of network flow problems. Tardos’ algorithm consists of solving a sequence of “low-sized” LP problems by a standard polynomially-convergent LP method and using their solutions to obtain the solution of the original LP problem.

The development of a method which works entirely in the context of the original LP problem and whose complexity is also bounded by a polynomial of  $m$ ,  $n$  and  $L_A$  is due to Vavasis and Ye [29]. Their method is a primal-dual path-following interior-point algorithm similar to the ones mentioned above except that it uses from time to time a crucial step, namely the least layered square (LLS) direction. They showed that their method has an  $\mathcal{O}(n^{3.5}(\log(\bar{\chi}_A + n)))$  iteration-complexity bound, where  $\bar{\chi}_A$  is a condition number associated with  $A$  which has the property that  $\log \bar{\chi}_A = \mathcal{O}(L_A)$  whenever  $A$  is integral. The number  $\bar{\chi}_A$  was first introduced implicitly by Dikin [1] in the study of primal affine scaling algorithms, and was later studied by several researchers including Vanderbei and Lagarias [28], Todd [23] and Stewart [20]. Properties of  $\bar{\chi}_A$  are studied in [3, 26, 27].

The complexity analysis of Vavasis and Ye’s algorithm is based on the notion of crossover event, a combinatorial event concerning the central path. Intuitively, a crossover event occurs between two variables when one of them is larger than the other at a point in the central path and then becomes smaller asymptotically as the optimal solution set is approached. Vavasis and Ye showed that there can be at most  $n(n-1)/2$  crossover events and that a distinct crossover event occurs every  $\mathcal{O}(n^{1.5}(\log(\bar{\chi}_A + n)))$  iterations, from which they deduced the overall  $\mathcal{O}(n^{3.5}(\log(\bar{\chi}_A + n)))$  iteration-complexity bound. In [13], an LP instance is given where the number of crossover events is  $\Theta(n^2)$ .

One disadvantage of Vavasis and Ye’s method is that it requires the explicit knowledge of  $\bar{\chi}_A$  in order to determine a partition of the variables into layers used in the computation of the LLS

step. This difficulty was remedied in a variant proposed by Megiddo, Mizuno and Tsuchiya [10] which does not require the explicit knowledge of the number  $\bar{\chi}_A$ . They observed that at most  $n$  types of partitions arise as  $\bar{\chi}_A$  varies from 1 to  $\infty$ , and that one of these can be used to compute the LLS step. Based on this idea, they developed a variant which computes the LLS steps for all these partitions and picks the one that yields the greatest duality gap reduction at the current iteration. Moreover, using the argument that, once the first LLS step is computed, the other ones can be cheaply computed by performing rank-one updates, they show that the overall complexity of their algorithm is exactly the same as Vavasis and Ye's algorithm.

Another approach that also remedies the above difficulty was proposed in Monteiro and Tsuchiya [18], where a variant of Vavasis and Ye's algorithm is developed which has the same complexity as theirs and computes only one LLS step per iteration without any explicit knowledge of  $\bar{\chi}_A$ . The method is a predictor-corrector type algorithm like the one described in [12] except that at the predictor stage it takes a step along either the primal-dual affine scaling step or the LLS step. In contrast to the LLS step used in the algorithm of Vavasis and Ye, the partition of variables used in the algorithm of [18] for computing the LLS step is constructed from the information provided by the AS direction and hence does not require any knowledge and/or guess on  $\bar{\chi}_A$ .

In this paper we present a new iteration-complexity bound for the Mizuno-Todd-Ye predictor-corrector (MTY P-C) primal-dual interior-point algorithm for linear programming. Using the notion of crossover events and a few other nontrivial ideas, we show that the MTY P-C algorithm started from a well-centered interior-feasible solution with duality gap  $n\mu_0$  finds an interior-feasible solution with duality gap less than  $n\eta$  in  $\mathcal{O}(n^2 \log(\log(\mu_0/\eta)) + n^{3.5} \log(\bar{\chi}_A^* + n))$  iterations, where  $\bar{\chi}_A^*$  is a scaling invariant condition number associated with the matrix  $A$ . More specifically,  $\bar{\chi}_A^*$  is the infimum of all the conditions numbers  $\bar{\chi}_{AD}$ , where  $D$  varies over the set of positive diagonal matrices. Thus, the derived iteration-complexity bound is scaling-invariant. Furthermore, we show that the MTY algorithm endowed with a certain scaling-invariant finite termination procedure terminates in at most  $\mathcal{O}(n^2 \log(\log(\mu_0/\eta^*)) + n^{3.5} \log(\bar{\chi}_A^* + n))$  iterations, where  $\eta^*$  is a scaling invariant threshold number determined by the data  $(A, b, c)$ . In particular, our results imply that the MTY P-C algorithm solves (1) and (2) in  $\mathcal{O}(n^2 \log L + n^{3.5} L_A)$  iterations under the Turing machine model. This contrasts well with the classical iteration-complexity bound for the MTY P-C algorithm, namely  $\mathcal{O}(\sqrt{n}L)$ , which depends linearly on the input size  $L$  of the data  $(A, b, c)$  instead of the logarithm of  $L$ .

The organization of the paper is as follows. Section 2 consists of five subsections. In Subsection 2.1, we review the notion of the primal-dual central path and its associated two norm neighborhoods. Subsection 2.2 introduces the condition number  $\bar{\chi}_A$  of a matrix  $A$  and describes the properties of  $\bar{\chi}_A$  that will be useful in our analysis. Subsection 2.3 reviews the MTY P-C algorithm and states the main result of this paper which establishes a new scaling-invariant iteration-complexity bound for the MTY P-C algorithm to find a near primal-dual optimal solution of (1) and (2). Subsection 2.4 describes a scaling-invariant finite termination procedure and gives an alternative scaling-invariant iteration-complexity bound for the MTY P-C algorithm to find a primal-dual optimal solution of (1) and (2). Section 3, which consists of four subsections, introduces some basic tools which are used in our convergence analysis. Subsection 3.1 discusses the notion of a crossover event. Subsection 3.2 describes the notion of a least layered squares (LLS) direction and states a proximity result that gives sufficient conditions under which the AS direction can be well approximated by an LLS direction. Subsection 3.3 reviews from a different perspective an important result from Vavasis and Ye [29], which basically provides sufficient conditions for the occurrence of crossover events. Subsection 3.4 describes two ordered partitions of the set of variables which are frequently used in our analysis.

Section 4 is dedicated to the proof of the main result stated in Subsection 2.3. Section 5 deals with a few implications of our main result under the Turing machine model. Section 6 provides some concluding remarks.

The following notation is used throughout our paper. We denote the vector of all ones by  $e$ . Its dimension is always clear from the context. The symbols  $\mathfrak{R}^n$ ,  $\mathfrak{R}_+^n$  and  $\mathfrak{R}_{++}^n$  denote the  $n$ -dimensional Euclidean space, the nonnegative orthant of  $\mathfrak{R}^n$  and the positive orthant of  $\mathfrak{R}^n$ , respectively. The set of all  $m \times n$  matrices with real entries is denoted by  $\mathfrak{R}^{m \times n}$ . If  $J$  is a finite index set then  $|J|$  denotes its cardinality, that is the number of elements of  $J$ . For  $J \subseteq \{1, \dots, n\}$  and  $w \in \mathfrak{R}^n$ , we let  $w_J$  denote the subvector  $[w_i]_{i \in J}$ ; moreover, if  $E$  is an  $m \times n$  matrix then  $E_J$  denotes the  $m \times |J|$  submatrix of  $E$  corresponding to  $J$ . For a vector  $w \in \mathfrak{R}^n$ , we let  $\max(w)$  and  $\min(w)$  denote the largest and the smallest component of  $w$ , respectively,  $\text{Diag}(w)$  denote the diagonal matrix whose  $i$ -th diagonal element is  $w_i$  for  $i = 1, \dots, n$ , and  $w^{-1}$  denote the vector  $[\text{Diag}(w)]^{-1}e$  whenever it is well-defined. For two vectors  $u, v \in \mathfrak{R}^n$ ,  $uv$  denotes their Hadamard product, i.e. the vector in  $\mathfrak{R}^n$  whose  $i$ th component is  $u_i v_i$ . The Euclidean norm, the 1-norm and the  $\infty$ -norm are denoted by  $\|\cdot\|$ ,  $\|\cdot\|_1$  and  $\|\cdot\|_\infty$ , respectively. For a matrix  $E$ ,  $\text{Im}(E)$  denotes the subspace generated by the columns of  $E$  and  $\text{Ker}(E)$  denotes the subspace orthogonal to the rows of  $E$ . The superscript  $T$  denotes transpose.

## 2 Problem and primal-dual predictor-corrector interior-point algorithms

In this section we review the MTY P-C algorithm [12] for solving the pair of LP problems (1) and (2). We also present our main convergence result which establishes a new polynomial iteration-complexity bound for this algorithm, namely  $\mathcal{O}(n^2 \log(\log(\mu_0/\eta)) + n^{3.5} \log(\bar{\chi}_A^* + n))$ , where  $\bar{\chi}_A^*$  is a certain scaling-invariant condition number associated with the matrix  $A$ ,  $n\mu_0$  is the initial duality gap and  $n\eta$  is the required upper bound on the duality gap of the final iterate.

This section is divided into four subsections. In Subsection 2.1, we describe the primal-dual central path and its associated two norm neighborhood. In Subsection 2.2, we define the condition number  $\bar{\chi}_A$  of a matrix  $A$  and review some of its properties. In Subsection 2.3, we review the MTY P-C algorithm and state the main result of this paper which establishes a new scaling-invariant iteration-complexity bound for the MTY P-C algorithm to find a near primal-dual optimal solution of (1) and (2). In Subsection 2.4, we describe a scaling-invariant finite termination procedure which can be invoked at every iterate of the MTY P-C algorithm and then give an alternative scaling-invariant iteration-complexity bound for the MTY P-C algorithm to find a primal-dual optimal solution of (1) and (2).

### 2.1 The problem, the central path and its 2-norm neighborhood

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

Given  $A \in \mathfrak{R}^{m \times n}$ ,  $c \in \mathfrak{R}^n$  and  $b \in \mathfrak{R}^m$ , consider the pairs of linear programs (1) and (2), where  $x \in \mathfrak{R}^n$  and  $(y, s) \in \mathfrak{R}^m \times \mathfrak{R}^n$  are their respective variables. The set of strictly feasible solutions for these problems are

$$\mathcal{P}^{++} \equiv \{x \in \mathfrak{R}^n : Ax = b, x > 0\},$$

$$\mathcal{D}^{++} \equiv \{(y, s) \in \mathfrak{R}^{m \times n} : A^T y + s = c, s > 0\},$$

respectively. Throughout the paper we make the following assumptions on the pair of problems (1) and (2).

**A.1**  $\mathcal{P}^{++}$  and  $\mathcal{D}^{++}$  are nonempty.

**A.2** The rows of  $A$  are linearly independent.

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

$$xs = \nu e, \tag{3}$$

$$Ax = b, \quad x > 0, \tag{4}$$

$$A^T y + s = c, \quad s > 0, \tag{5}$$

has a unique solution  $(x, y, s)$ , which we denote by  $(x(\nu), y(\nu), s(\nu))$ . The central path is the set consisting of all these solutions as  $\nu$  varies in  $(0, \infty)$ . As  $\nu$  converges to zero, the path  $(x(\nu), y(\nu), s(\nu))$  converges to a primal-dual optimal solution  $(x^*, y^*, s^*)$  for problems (1) and (2). 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  $\mu = \mu(x, s) \equiv x^T s/n$ , respectively, and the point  $(x(\mu), y(\mu), s(\mu))$  is said to be the central point associated with  $w$ . Note that  $(x(\mu), y(\mu), s(\mu))$  also has normalized duality gap  $\mu$ . 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

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

Clearly,  $\phi(w) = 0$  if and only if  $w = (x(\mu), y(\mu), s(\mu))$ , or equivalently  $w$  coincides with its associated central point. The two-norm neighborhood of the central path with opening  $\beta > 0$  is defined as

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

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

$$\delta(w) \equiv s^{1/2}x^{-1/2} \in \mathfrak{R}^n. \tag{6}$$

## 2.2 Condition number

In this subsection we define the condition number  $\bar{\chi}_A$  associated with the constraint matrix  $A$  and state the properties of  $\bar{\chi}_A$  which will play an important role in our analysis.

Let  $\mathcal{D}$  denote the set of all positive definite  $n \times n$  diagonal matrices and define

$$\begin{aligned} \bar{\chi}_A &\equiv \sup \left\{ \|A^T(A\tilde{D}A^T)^{-1}A\tilde{D}\| : \tilde{D} \in \mathcal{D} \right\} \\ &= \sup \left\{ \frac{\|A^T y\|}{\|c\|} : y = \operatorname{argmin}_{\tilde{y} \in \mathfrak{R}^n} \|\tilde{D}^{1/2}(A^T \tilde{y} - c)\| \text{ for some } 0 \neq c \in \mathfrak{R}^n \text{ and } \tilde{D} \in \mathcal{D} \right\}. \end{aligned} \tag{7}$$

The parameter  $\bar{\chi}_A$  plays a fundamental role in the complexity analysis of algorithms for linear programming and least square problems (see [29] and references therein). Its finiteness has been firstly established by Dikin [1]. Other authors have also given alternative derivations of the finiteness of  $\bar{\chi}_A$  (see for example Stewart [20], Todd [23] and Vanderbei and Lagarias [28]).

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

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

- a)  $\bar{\chi}_{GA} = \bar{\chi}_A$  for any nonsingular matrix  $G \in \mathfrak{R}^{m \times m}$ ;
- b)  $\bar{\chi}_A = \max\{\|G^{-1}A\| : G \in \mathcal{G}\}$  where  $\mathcal{G}$  denote the set of all  $m \times m$  nonsingular submatrices of  $A$ ;
- c) If the entries of  $A$  are all integers, then  $\bar{\chi}_A$  is bounded by  $2^{\mathcal{O}(L_A)}$ , where  $L_A$  is the input bit length of  $A$ ;
- d)  $\bar{\chi}_A = \bar{\chi}_F$  for any  $F \in \mathfrak{R}^{(n-m) \times n}$  such that  $\text{Ker}(A) = \text{Im}(F^T)$ .

*Proof.* Statement a) readily follows from the definition (7). The inequality  $\bar{\chi}_A \geq \max\{\|G^{-1}A\| : G \in \mathcal{G}\}$  is established in Lemma 3 of [29] while the proof of the reverse inequality is given in [23] (see also Theorem 1 of [24]). Hence, b) holds. The proof of c) can be found in Lemma 24 of [29]. A proof of d) can be found in [3].  $\blacksquare$

We now state a Hoffman-type result for a system of linear equalities whose proof can be found in Lemma 2.3 of [18].

**Lemma 2.2** *Let  $A \in \mathfrak{R}^{m \times n}$  with full row rank be given and let  $(\mathcal{K}, \mathcal{L})$  be an arbitrary bipartition of the index set  $\{1, \dots, n\}$ . Assume that  $\bar{w} \in \mathfrak{R}^{|\mathcal{L}|}$  is an arbitrary vector such that the system  $A_{\mathcal{K}}u = A_{\mathcal{L}}\bar{w}$  is feasible. Then, this system has a feasible solution  $\bar{u}$  such that  $\|\bar{u}\| \leq \bar{\chi}_A \|\bar{w}\|$ .*

### 2.3 Predictor-corrector step and its properties

In this subsection we review the well-known MTY P-C algorithm [12] and its main properties. We also state the main result of this paper which establishes a new scaling-invariant iteration-complexity bound for the MTY P-C algorithm to find a near primal-dual optimal solution of (1) and (2).

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

$$\begin{aligned} S\Delta x + X\Delta s &= \sigma\mu e - xs, \\ A\Delta x &= 0, \\ A^T\Delta y + \Delta s &= 0, \end{aligned} \tag{8}$$

where  $\mu = \mu(x, s)$  and  $\sigma \in \mathfrak{R}$  is a prespecified parameter, commonly referred to as the centrality parameter. When  $\sigma = 0$ , we denote the solution of (8) 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 (8) 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.

To describe an entire iteration of the MTY P-C algorithm, suppose that a constant  $\beta \in (0, 1/4]$  is given. Given a point  $w = (x, y, s) \in \mathcal{N}(\beta)$ , this algorithm generates the next point  $w^+ = (x^+, y^+, s^+) \in \mathcal{N}(\beta)$  as follows. It first moves along the direction  $(\Delta x^a, \Delta y^a, \Delta s^a)$  until it hits

the boundary of the enlarged neighborhood  $\mathcal{N}(2\beta)$ . More specifically, it computes the point  $w^a = (x^a, y^a, s^a) \equiv (x, y, s) + \alpha_a(\Delta x^a, \Delta y^a, \Delta s^a)$  where

$$\alpha_a \equiv \sup \{ \alpha \in [0, 1] : (x, y, s) + \alpha(\Delta x^a, \Delta y^a, \Delta s^a) \in \mathcal{N}(2\beta) \}. \quad (9)$$

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

The convergence analysis of the sequence  $\{w^k\}$  as well as the finite termination of the MTY P-C algorithm has been the subject of study of several papers. In what follows, we review the properties of the P-C iteration which yields the classical polynomial iteration-complexity bound for the MTY P-C algorithm. We also discuss alternative properties of the P-C iteration which will be used in our analysis to derive a new polynomial iteration-complexity bound for MTY P-C algorithm. A scaling-invariant finite termination procedure for the MTY P-C algorithm and its relationship with another well-known finite termination procedure will be discussed in Subsection 2.4.

For a detailed proof of the next two propositions, we refer the reader to [12].

**Proposition 2.3 (Predictor step)** *Suppose that  $w = (x, y, s) \in \mathcal{N}(\beta)$  for some constant  $\beta \in (0, 1/2]$ . Let  $\Delta w^a = (\Delta x^a, \Delta y^a, \Delta s^a)$  denote the AS direction at  $w$  and let  $\alpha_a$  be the stepsize computed according to (9). Then the following statements hold:*

- a) *the point  $w + \alpha \Delta w^a$  has normalized duality gap  $\mu(\alpha) = (1 - \alpha)\mu$  for all  $\alpha \in \mathbb{R}$ ;*
- b)  *$\alpha_a \geq \max\{1 - \chi/\beta, \sqrt{\beta/n}\}$  where  $\chi \equiv \|\Delta x^a \Delta s^a\|/\mu$ .*

*Proof.* It is well-known that a) holds,  $\chi \leq n/2$  and

$$\alpha^a \geq \frac{2}{1 + \sqrt{1 + 4\chi/\beta}}.$$

Using these two inequalities, we see after a simple verification that b) holds. ■

**Proposition 2.4 (Corrector step)** *Suppose that  $w = (x, y, s) \in \mathcal{N}(2\beta)$  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  $(\Delta x, \Delta y, \Delta s)$  at a point  $(x, y, s)$ , the quantity

$$(Rx, Rs) \equiv \left( \frac{\delta(x + \Delta x)}{\sqrt{\mu}}, \frac{\delta^{-1}(s + \Delta s)}{\sqrt{\mu}} \right) = \left( \frac{x^{1/2}s^{1/2} + \delta\Delta x}{\sqrt{\mu}}, \frac{x^{1/2}s^{1/2} + \delta^{-1}\Delta s}{\sqrt{\mu}} \right), \quad (10)$$

where  $\delta \equiv \delta(w)$ , appears quite often in our analysis. We refer to it as the *residual* of  $(\Delta x, \Delta y, \Delta s)$ . 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 if  $(Rx^a, Rs^a) \equiv (Rx^a(w), Rs^a(w))$  then

$$Rx^a = -\frac{1}{\sqrt{\mu}}\delta^{-1}\Delta s^a, \quad Rs^a = -\frac{1}{\sqrt{\mu}}\delta\Delta x^a, \quad (11)$$

and

$$Rx^a + Rs^a = \frac{x^{1/2}s^{1/2}}{\sqrt{\mu}}, \quad (12)$$

due to the fact that  $(\Delta x^a, \Delta y^a, \Delta s^a)$  satisfies the first equation in (8) with  $\sigma = 0$ . The following quantity plays an important role in our analysis:

$$\varepsilon_\infty^a(w) \equiv \max_i \{ \min \{ |Rx_i^a(w)|, |Rs_i^a(w)| \} \}. \quad (13)$$

In terms of this quantity, we have the following bound on the reduction on the duality gap during an iteration of the MTY P-C algorithm.

**Lemma 2.5** *Suppose that  $w = (x, y, s) \in \mathcal{N}(\beta)$  for some constant  $\beta \in (0, 1/4]$  and define  $\varepsilon_\infty^a \equiv \varepsilon_\infty^a(w)$ . Let  $w^+ = (x^+, y^+, s^+)$  denote the point obtained after a single iteration of the MTY P-C algorithm with base point  $w$ . Then,  $w^+ \in \mathcal{N}(\beta)$  and*

$$\mu(w^+) \leq \min \left\{ 1 - \sqrt{\frac{\beta}{n}}, \frac{\varepsilon_\infty^a \sqrt{n}}{\beta} \right\} \mu(w). \quad (14)$$

We end this section by stating the main result of this paper. This result establishes a new iteration-complexity bound for the MTY P-C algorithm.

**Theorem 2.6** *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  $\mu(w^k) \leq \eta$  in at most*

$$\mathcal{O} \left( \min \left\{ \sqrt{n} \log(\mu_0/\eta), n^2 \log(\log(\mu_0/\eta)) + n^{3.5} \log(\bar{\chi}_A^* + n) \right\} \right) \quad (15)$$

iterations, where

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

A few observations are in order at this point. First, the first bound in (15) is the classical one derived in [12] (see also [7, 14, 15, 16]). It follows as an immediate corollary of the first bound on the duality gap reduction obtained in Lemma 2.5. The second bound in (15) is the one which will be established in this paper. Observe that in contrast to the classical iteration-complexity bound which is proportional to  $\log(\mu_0/\eta)$ , the new bound depends linearly on  $\log(\log(\mu_0/\eta))$ . Second, note that the MTY P-C algorithm is scaling-invariant, i.e. if the change of variables  $(x, y, s) = (D\tilde{x}, \tilde{y}, D^{-1}\tilde{s})$  for some  $D \in \mathcal{D}$  is performed on the pair of problems (1) and (2) and the MTY P-C algorithm is applied to the new dual pair of scaled problems, then the sequence of iterates  $\{\tilde{w}^k\}$  generated satisfies  $w^k = D\tilde{w}^k$  for all  $k \geq 1$  as long as the initial iterate  $\tilde{w}^0 \in \mathcal{N}(\beta)$  in the  $\tilde{w}$ -space satisfies  $w^0 = D\tilde{w}^0$ . For this reason, the MTY P-C algorithm should have an iteration-complexity bound which does not depend on the scaled space where the sequence of iterates is generated. Note that the iteration-complexity bound (15) has this property since  $\mu_0$ ,  $\bar{\chi}_A^*$  and the inequality  $\mu(w^k) \leq \eta$  are all scaling-invariant. Third, to establish the iteration-complexity bound stated in Theorem 2.6, it is sufficient to establish that in the scaled space, the iteration-complexity bound is

$$\mathcal{O} \left( \min \left\{ \sqrt{n} \log(\mu_0/\eta), n^2 \log(\log(\mu_0/\eta)) + n^{3.5} \log(\bar{\chi}_{AD} + n) \right\} \right). \quad (16)$$

Since the number of iterations of the MTY P-C algorithm does not depend on the underlying scaled space, it follows that this number is majorized by the infimum of (16) over all  $D \in \mathcal{D}$ , i.e., by the bound (15). Moreover, without any loss of generality, we will consider the MTY P-C algorithm applied to (1) and (2) without any scaling and will establish the iteration-complexity bound (16) with  $D = I$ .

## 2.4 A scaling-invariant finite termination procedure

In this section, we describe a scaling-invariant finite termination procedure which, used in conjunction with the MTY P-C algorithm, allows one to find an exact primal-dual optimal solution of (1) and (2). We also derive an alternative scaling-invariant iteration-complexity bound for the MTY P-C algorithm to find an exact primal-dual optimal solution of (1) and (2).

The finite termination procedure described in this subsection is similar to the one described in Mehrotra and Ye [11] (see also [33]) except that it uses a scaling-invariant scheme for guessing the optimal partition  $(B_*, N_*)$  associated with the pair of LP problems (1) and (2). (Recall that by definition  $B_* \equiv \{i : x_i > 0 \text{ for some } x \in \text{opt}(1)\}$  and  $N_* \equiv \{i : s_i > 0 \text{ for some } (y, s) \in \text{opt}(2)\}$ .) Namely, given a point  $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ , we define the AS-bipartition  $(B(w), N(w))$  at  $w$  as

$$B(w) \equiv \{i : |Rs_i^a(w)| \leq |Rx_i^a(w)|\}, \quad N(w) \equiv \{i : |Rs_i^a(w)| > |Rx_i^a(w)|\}. \quad (17)$$

We now state the following finite termination procedure which uses the AS-bipartition as a guess for the optimal partition.

### FT Procedure:

Given  $\gamma \in (0, 1)$  and  $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$  such that  $xs \geq \gamma\mu(w)$ .

- 1) Find the AS-bipartition  $(B, N) = (B(w), N(w))$  at  $w$ ;
- 2) Solve the following projection problems:

$$x^* \equiv \operatorname{argmin}_{\tilde{x}} \left\{ \|\delta(x - \tilde{x})\|^2 : A\tilde{x} = b, \tilde{x}_N = 0 \right\}, \quad (18)$$

$$(y^*, s^*) \equiv \operatorname{argmin}_{(\tilde{y}, \tilde{s})} \left\{ \|\delta^{-1}(s - \tilde{s})\|^2 : A^T \tilde{y} + \tilde{s} = c, \tilde{s}_B = 0 \right\}; \quad (19)$$

- 3) If  $x_B^* > 0$  and  $s_N^* > 0$  then  $w^* = (x^*, y^*, s^*)$  is optimal; output  $w^*$  and declare success. Otherwise, exit the procedure and declare failure.

### End

Our goal now will be to show that the FT procedure always finds a primal-dual optimal solution of (1) and (2) whenever  $\mu(w)$  is less than a certain positive threshold constant defined in terms of two additional condition numbers associated with the pair of LP problems (1) and (2). We start by defining these two condition numbers. The first condition number is given by

$$\xi(A, b, c) \equiv \min \left\{ \min_{i \in B_*} \xi_i^P, \min_{i \in N_*} \xi_i^D \right\},$$

where for every  $i = 1, \dots, n$ ,

$$\xi_i^P \equiv \max\{\bar{x}_i : \bar{x} \in \text{opt}(1)\}, \quad \xi_i^D \equiv \max\{\bar{s}_i : (\bar{y}, \bar{s}) \in \text{opt}(2)\}.$$

Here,  $\text{opt}(\cdot)$  denotes the set of optimal solutions of the problem  $(\cdot)$ . The second condition number is defined as

$$\zeta(A, (B_*, N_*)) \equiv \max\{\zeta_1, \zeta_2\}, \quad (20)$$

where

$$\begin{aligned} \zeta_1 &\equiv \max_{d_{N_*} \neq 0} \left\{ \min \left\{ \frac{\|d_{B_*}\|}{\|d_{N_*}\|} : d = (d_{B_*}, d_{N_*}) \in \text{Ker}(A) \right\} \right\}, \\ \zeta_2 &\equiv \max_{d_{B_*} \neq 0} \left\{ \min \left\{ \frac{\|d_{N_*}\|}{\|d_{B_*}\|} : d = (d_{B_*}, d_{N_*}) \in \text{Im}(A^T) \right\} \right\}, \end{aligned}$$

Using Lemma 2.2 and Proposition 2.1(d), it is easy to see that  $\zeta(A, (B_*, N_*)) \leq \bar{\chi}_A$ .

The following result states some well-known estimates on the size of the components of a point  $w \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ .

**Lemma 2.7** *Let  $\gamma \in (0, 1)$  be given. Then, for every  $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$  such that  $xs \geq \gamma\mu$ , we have:*

$$\max\{\max(x_{N_*}), \max(s_{B_*})\} \leq \frac{n\mu}{\xi}, \quad (21)$$

$$\min\{\min(x_{B_*}), \min(s_{N_*})\} \geq \frac{\gamma\xi}{n}, \quad (22)$$

$$\|\delta_{B_*}\|_\infty \|\delta_{N_*}^{-1}\|_\infty \leq \frac{n^2\mu}{\gamma\xi^2}, \quad (23)$$

where  $\mu \equiv \mu(w)$ ,  $\delta \equiv \delta(w)$  and  $\xi \equiv \xi(A, b, c)$ .

*Proof.* The inequality (21) follows immediately from the definition of  $\xi(A, b, c)$  and the following identities:  $x^T s = n\mu$ ,  $\bar{x}^T \bar{s} = 0$ ,  $x_i s_i \geq \gamma\mu$  for all  $i = 1, \dots, n$ , and  $(x - \bar{x})^T (s - \bar{s}) = 0$  for all  $\bar{w} = (\bar{x}, \bar{y}, \bar{s}) \in \text{opt}(1) \times \text{opt}(2)$  (see Ye [32] for the details). Moreover, using the assumption that  $xs \geq \gamma\mu$  and (21), we obtain  $s_i \geq \gamma\mu/x_i \geq \gamma\xi/n$  for every  $i \in N_*$  and  $x_i \geq \gamma\mu/s_i \geq \gamma\xi/n$  for every  $i \in B_*$ , from which (22) follows. Inequality (23) follows immediately from (6), (21) and (22). ■

The following result shows not only that the AS-bipartition  $(B(w), N(w))$  is a correct guess for the optimal partition  $(B_*, N_*)$  but also that the FT Procedure yields a primal-dual optimal solution of (1) and (2) whenever the duality gap  $\mu(w)$  is less than a suitable threshold value defined in terms of the conditions numbers  $\xi$  and  $\zeta$  introduced above, the dimension  $n$  and the degree of centrality  $\gamma$  of  $w$ .

**Lemma 2.8** *Suppose that  $\gamma \in (0, 1)$  and  $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$  are such that*

$$xs \geq \gamma\mu, \quad \text{and} \quad \mu < \frac{\gamma^{1.5}\xi^2}{2\zeta n^{2.5}}, \quad (24)$$

where  $\mu = \mu(w)$ ,  $\zeta \equiv \zeta(A, (B_*, N_*))$  and  $\xi \equiv \xi(A, b, c)$ . Then:

$$a) (B(w), N(w)) = (B_*, N_*);$$

b) the FT Procedure yields a strictly complementary primal-dual optimal solution, that is a triple  $w^* = (x^*, y^*, s^*) \in \text{opt}(1) \times \text{opt}(2)$  such that  $x^* + s^* > 0$ .

*Proof.* Suppose  $\gamma \in (0, 1)$  and  $w \in \mathcal{P}^{++} \times \mathcal{D}^{++}$  are such that (24) holds. We will first prove a). Let  $\Delta w^a = (\Delta x^a, \Delta y^a, \Delta s^a)$  denote the AS direction at  $w$ . It is well-known that

$$\max \left\{ \|\delta \Delta x^a\|, \|\delta^{-1} \Delta s^a\| \right\} \leq \sqrt{n\mu}. \quad (25)$$

Moreover, it is easy to see that

$$\Delta x^a = \operatorname{argmin} \left\{ \tilde{s}^T \Delta x + \frac{1}{2} \|\delta \Delta x\|^2 : A \Delta x = 0 \right\},$$

for every  $\tilde{s} \in c + \operatorname{Im}(A^T)$ . Now, fix some  $(\bar{y}, \bar{s}) \in \text{opt}(2)$  and let  $\tilde{s} = \bar{s}$  in the above optimization problem. Splitting the variable  $\Delta x^a$  according to the partition  $(B_*, N_*)$ , noting that  $\bar{s}_{B_*} = 0$ , and fixing the component  $\Delta x_{N_*}$  to  $\Delta x_{N_*}^a$ , we conclude that

$$\Delta x_{B_*}^a = \operatorname{argmin}_{\Delta x_{B_*}} \left\{ \frac{1}{2} \|\delta_{B_*} \Delta x_{B_*}\|^2 : A_{B_*} \Delta x_{B_*} = -A_{N_*} \Delta x_{N_*}^a \right\}. \quad (26)$$

Now, let  $\widetilde{\Delta x}_{B_*}$  denote the minimum norm solution of the system  $A_{B_*} \Delta x_{B_*} = -A_{N_*} \Delta x_{N_*}^a$ . Using (23), (25), (26), and the definition of  $\zeta = \zeta(A, (B_*, N_*))$ , we obtain

$$\begin{aligned} \|\delta_{B_*} \Delta x_{B_*}^a\| &\leq \|\delta_{B_*} \widetilde{\Delta x}_{B_*}\| \leq \|\delta_{B_*}\|_\infty \|\widetilde{\Delta x}_{B_*}\| \leq \zeta \|\delta_{B_*}\|_\infty \|\Delta x_{N_*}^a\| \\ &\leq \zeta \|\delta_{B_*}\|_\infty \|\delta_{N_*}^{-1}\|_\infty \|\delta_{N_*} \Delta x_{N_*}^a\| \leq \sqrt{n\mu} \zeta \|\delta_{B_*}\|_\infty \|\delta_{N_*}^{-1}\|_\infty \\ &\leq \frac{n^{2.5} \zeta \mu^{1.5}}{\gamma \xi^2} < \frac{\sqrt{\gamma \mu}}{2}, \end{aligned}$$

where the last inequality is due to (24). The last inequality together with (24) imply that for every  $i \in B_*$ , we have

$$\left| \frac{\delta_i^{-1} \Delta s_i^a}{\sqrt{\mu}} \right| = \left| \frac{x_i^{1/2} s_i^{1/2}}{\sqrt{\mu}} + \frac{\delta_i \Delta x_i^a}{\sqrt{\mu}} \right| \geq \frac{x_i^{1/2} s_i^{1/2}}{\sqrt{\mu}} - \left| \frac{\delta_i \Delta x_i^a}{\sqrt{\mu}} \right| > \sqrt{\gamma} - \frac{\sqrt{\gamma}}{2} = \frac{\sqrt{\gamma}}{2} \geq \left| \frac{\delta_i \Delta x_i^a}{\sqrt{\mu}} \right|,$$

or equivalently,  $|R x_i^a(w)| > |R s_i^a(w)|$  for every  $i \in B_*$ . Hence,  $B_* \subseteq B(w)$  in view of (17). Similarly, we can show that  $N_* \subseteq N(w)$ . Therefore,  $(B_*, N_*) = (B(w), N(w))$ .

We now prove b). Let  $(x^*, y^*, s^*)$  denote the point determined by (18) and (19). Observe that  $x_{B_*} - x_{B_*}^*$  is a feasible solution of the system  $A_{B_*} \Delta x_{B_*} = -A_{N_*} x_{N_*}$ . Now, let  $\widetilde{\Delta x}_{B_*}$  denote the minimum norm solution of this system. Using (23), (24), (18), and the definition of  $\zeta = \zeta(A, (B_*, N_*))$ , we obtain

$$\begin{aligned} \|x_{B_*}^{-1} (x_{B_*} - x_{B_*}^*)\| &\leq \|x_{B_*}^{-1/2} s_{B_*}^{-1/2}\|_\infty \|\delta_{B_*} (x_{B_*} - x_{B_*}^*)\| \leq \frac{1}{\sqrt{\gamma \mu}} \|\delta_{B_*} (x_{B_*} - x_{B_*}^*)\| \\ &\leq \frac{1}{\sqrt{\gamma \mu}} \|\delta_{B_*} \widetilde{\Delta x}_{B_*}\| \leq \frac{1}{\sqrt{\gamma \mu}} \|\delta_{B_*}\|_\infty \|\widetilde{\Delta x}_{B_*}\| \leq \frac{\zeta}{\sqrt{\gamma \mu}} \|\delta_{B_*}\|_\infty \|x_{N_*}\| \\ &\leq \frac{\zeta}{\sqrt{\gamma \mu}} \|\delta_{B_*}\|_\infty \|\delta_{N_*}^{-1}\|_\infty \|\delta_{N_*} x_{N_*}\| = \frac{\zeta}{\sqrt{\gamma \mu}} \|\delta_{B_*}\|_\infty \|\delta_{N_*}^{-1}\|_\infty \|x_{N_*}^{1/2} s_{N_*}^{1/2}\| \\ &\leq \frac{\zeta \sqrt{n}}{\sqrt{\gamma}} \|\delta_{B_*}\|_\infty \|\delta_{N_*}^{-1}\|_\infty \leq \frac{n^{2.5} \zeta \mu}{\gamma^{1.5} \xi^2} < \frac{1}{2}. \end{aligned}$$

The above inequality clearly implies that  $x_{B^*}^* > 0$ . In a similar way, we can prove that  $s_{N^*}^* > 0$ . Hence, b) follows.  $\blacksquare$

The next result gives an iteration-complexity bound for the MTY P-C algorithm, used in conjunction with the FT procedure, to find a primal-dual optimal solution of (1) and (2).

**Theorem 2.9** *Suppose that  $w^0 \in \mathcal{N}(\beta)$  with  $\beta \in (0, 1/4]$  is given. Then, the version of the MTY P-C algorithm, in which the FT Procedure is invoked at every iterate  $w^k$ , started from  $w^0$  finds a primal-dual strictly complementary optimal solution  $w^*$  in at most*

$$\mathcal{O}\left(\min\left\{\sqrt{n}\log(n\mu_0/\eta_*), n^2\log(\log(n\mu_0/\eta_*)) + n^{3.5}\log(\bar{\chi}_A^* + n)\right\}\right)$$

iterations, where  $\mu_0 = \mu(w^0)$  and

$$\eta_* \equiv \sup\left\{\frac{[\xi(AD, b, Dc)]^2}{\zeta(AD, (B_*, N_*))} : D \in \mathcal{D}\right\}.$$

*Proof.* This result follows immediately from Theorem 2.6, Lemma 2.8 and the scaling-invariance of the algorithm under consideration.  $\blacksquare$

A few implications of this result under the Turing machine model will be discussed in Section 5.

### 3 Basic Tools

In this section we introduce the basic tools that will be used in the proof of Theorem 2.6. The analysis heavily relies on the notion of crossover events due to Vavasis and Ye [29]. In Subsection 3.1, we give a definition of crossover event which is slightly different than the one introduced in [29] and then discuss some of its properties. In Subsection 3.2, we describe the notion of an LLS direction introduced in [29] and then state a proximity result that gives sufficient conditions under which the AS direction can be well approximated by an LLS direction. Subsection 3.3 reviews from a different perspective an important result from [29], namely Lemma 17 of [29], that essentially guarantees the occurrence of crossover events. Since this result is stated in terms of the residual of an LLS step, the use of the proximity result of Subsection 3.2 allows us to obtain a similar result stated in terms of the residual of the AS direction. In Subsection 3.4, we introduce two ordered partitions of the set of variables which play an important role in our analysis.

#### 3.1 Crossover events

In this subsection we discuss the important notion of a crossover event developed by Vavasis and Ye [31].

**Definition:** For two indices  $i, j \in \{1, \dots, n\}$  and a constant  $\mathcal{C} \geq 1$ , a  $\mathcal{C}$ -crossover event for the pair  $(i, j)$  is said to occur on the interval  $(\nu', \nu]$  if

$$\begin{aligned} &\text{there exists } \nu_0 \in (\nu', \nu] \text{ such that } \frac{s_j(\nu_0)}{s_i(\nu_0)} \leq \mathcal{C}, \\ &\text{and, } \frac{s_j(\tilde{\nu})}{s_i(\tilde{\nu})} > \mathcal{C} \text{ for all } \tilde{\nu} \leq \nu'. \end{aligned} \tag{27}$$

Moreover, the interval  $(\nu', \nu]$  is said to contain a  $\mathcal{C}$ -crossover event if (27) holds for some pair  $(i, j)$ .

Hence, the notion of a crossover event is independent of any algorithm and is a property of the central path only. Note that in view of (3), condition (27) can be reformulated into an equivalent condition involving only the primal variable. For our purposes, we will use only (27).

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

**Proposition 3.1** *Let  $\mathcal{C} > 0$  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  $w^l$ ,  $k < l$ , generated by the MTY P-C algorithm if the interval  $(\mu(w^l), \mu(w^k)]$  contains a  $\mathcal{C}$ -crossover event. Note that in view of Proposition 3.1, there can be at most  $n(n-1)/2$  disjoint intervals of this type. We will show in the remaining part of this paper that there exists a constant  $\mathcal{C} > 0$  with the following property: for any index  $k$ , there exists an index  $l > k$  with the following property:  $l - k = \mathcal{O}(\log(\log(\mu_0/\eta)) + n^{1.5} \log(\bar{\chi}_A + n))$  and, if the MTY P-C algorithm does not terminate before or at the  $l$ -th iteration, then a  $\mathcal{C}$ -crossover event must occur between the iterates  $w^k$  and  $w^l$ . Proposition 3.1 and a simple argument then show that the MTY P-C algorithm must terminate within  $\mathcal{O}(n^2 \log(\log(\mu_0/\eta)) + n^{3.5} \log(\bar{\chi}_A + n))$  iterations, yielding Theorem 2.6.

### 3.2 LLS directions and their relationship with the AS direction

In this subsection we describe another type of direction which plays an important role on a criteria which guarantees the occurrence of crossover events (see Lemma 3.3), namely the layered least squares (LLS) direction. We also state a proximity result which describes how the AS direction can be well-approximated by suitable LLS directions.

The LLS direction was first introduced by Vavasis and Ye in [29] and is one of two directions used in their algorithm. While the algorithm in this paper does not rely on this direction, its analysis heavily rely on it by means of the implications of Lemma 3.3.

Let  $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$  and a partition  $(J_1, \dots, J_p)$  of the index set  $\{1, \dots, n\}$  be given and define  $\delta \equiv \delta(w)$ . The primal LLS direction  $\Delta x^{\text{ll}} = (\Delta x_{J_1}^{\text{ll}}, \dots, \Delta x_{J_p}^{\text{ll}})$  at  $w$  with the respect to  $J$  is defined recursively according to the order  $\Delta x_{J_p}^{\text{ll}}, \dots, \Delta x_{J_1}^{\text{ll}}$  as follows. Assume that the components  $\Delta x_{J_p}^{\text{ll}}, \dots, \Delta x_{J_{k+1}}^{\text{ll}}$  have been determined. Let  $\Pi_{J_k} : \mathfrak{R}^n \rightarrow \mathfrak{R}^{J_k}$  denote the projection map defined as  $\Pi_{J_k}(u) = u_{J_k}$  for all  $u \in \mathfrak{R}^n$ . Then  $\Delta x_{J_k}^{\text{ll}} \equiv \Pi_{J_k}(L_k^x)$  where  $L_k^x$  is given by

$$\begin{aligned} L_k^x &\equiv \operatorname{Argmin}_{p \in \mathfrak{R}^n} \left\{ \|\delta_{J_k}(x_{J_k} + p_{J_k})\|^2 : p \in L_{k-1}^x \right\} \\ &= \operatorname{Argmin}_{p \in \mathfrak{R}^n} \left\{ \|\delta_{J_k}(x_{J_k} + p_{J_k})\|^2 : p \in \operatorname{Ker}(A), p_{J_i} = \Delta x_{J_i}^{\text{ll}} \text{ for all } i = k+1, \dots, p \right\}, \end{aligned} \quad (28)$$

with the convention that  $L_0^x = \operatorname{Ker}(A)$ . The slack component  $\Delta s^{\text{ll}} = (\Delta s_{J_1}^{\text{ll}}, \dots, \Delta s_{J_p}^{\text{ll}})$  of the dual LLS direction  $(\Delta y^{\text{ll}}, \Delta s^{\text{ll}})$  at  $w$  with the respect to  $J$  is defined recursively as follows. Assume that the components  $\Delta s_{J_1}^{\text{ll}}, \dots, \Delta s_{J_{k-1}}^{\text{ll}}$  have been determined. Then  $\Delta s_{J_k}^{\text{ll}} \equiv \Pi_{J_k}(L_k^s)$  where  $L_k^s$  is given by

$$\begin{aligned} L_k^s &\equiv \operatorname{Argmin}_{q \in \mathfrak{R}^n} \left\{ \|\delta_{J_k}^{-1}(s_{J_k} + q_{J_k})\|^2 : q \in L_{k-1}^s \right\} \\ &= \operatorname{Argmin}_{q \in \mathfrak{R}^n} \left\{ \|\delta_{J_k}^{-1}(s_{J_k} + q_{J_k})\|^2 : q \in \operatorname{Im}(A^T), q_{J_i} = \Delta s_{J_i}^{\text{ll}} \text{ for all } i = 1, \dots, k-1 \right\}, \end{aligned} \quad (29)$$

with the convention that  $L_0^s = \text{Im}(A^T)$ . Finally, once  $\Delta s^{\text{ll}}$  has been determined, the component  $\Delta y^{\text{ll}}$  is determined from the relation  $A^T \Delta y^{\text{ll}} + \Delta s^{\text{ll}} = 0$ .

It is easy to verify that the AS direction is a special LLS direction, namely the one with respect to the only partition in which  $p = 1$ . Clearly, the LLS direction at a given  $w \in \mathcal{P}^{++} \times \mathcal{D}^{++}$  depends on the partition  $J = (J_1, \dots, J_p)$  used.

A partition  $J = (J_1, \dots, J_p)$  is said to be *ordered* at a point  $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$  if  $\max(\delta_{J_i}) \leq \min(\delta_{J_{i+1}})$  for all  $i = 1, \dots, p-1$ . In this case, the gap of  $J$ , denoted by  $\text{gap}(J)$ , is defined as

$$\text{gap}(J) = \min_{1 \leq i \leq p-1} \left\{ \frac{\min(\delta_{J_{i+1}})}{\max(\delta_{J_i})} \right\} = \frac{1}{\max_{1 \leq i \leq p-1} (\|\delta_{J_i}\|_\infty \|\delta_{J_{i+1}}^{-1}\|_\infty)} \geq 1,$$

with the convention that  $\text{gap}(J) = \infty$  if  $p = 1$ .

In the remaining part of this subsection, we describe how the AS direction at a given  $w \in \mathcal{P}^{++} \times \mathcal{D}^{++}$  can be well-approximated by suitable LLS steps, a result that will be important in our convergence analysis. Another result along this direction has also been obtained by Ye and Vavasis [31]. However, our result is more general and better suited for the development of this paper.

The proximity result below can be proved using the projection decomposition techniques developed in [25]. Another proof using instead the techniques developed in [17] has been given in [18]. The result essentially states that the larger the gap of  $J$  is, the closer the AS direction and the LLS direction with respect to  $J$  will be to one another.

**Proposition 3.2** *Let  $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$  and an ordered partition  $J = (J_1, \dots, J_p)$  at  $w$  be given. Define  $\delta \equiv \delta(w)$  and let  $(Rx^a, Rs^a)$  and  $(Rx^{\text{ll}}, Rs^{\text{ll}})$  denote the residuals of the AS direction at  $w$  and of the LLS direction at  $w$  with respect to  $J$ , respectively. If  $\text{gap}(J) \geq 4p\bar{\chi}_A$ , then*

$$\max \left\{ \left\| Rx^a - Rx^{\text{ll}} \right\|_\infty, \left\| Rs^a - Rs^{\text{ll}} \right\|_\infty \right\} \leq \frac{12\sqrt{n}\bar{\chi}_A}{\text{gap}(J)}.$$

In view of the above result, the AS direction can be well approximated by LLS directions with respect to ordered partitions  $J$  which have large gaps. The LLS direction with  $p = 1$ , which is the AS direction, provides the perfect approximation to the AS direction itself. However, this kind of trivial approximation is not useful for us due to the need of keeping the ‘‘spread’’ of some layers  $J_k$  under control. For an ordered partition  $J = (J_1, \dots, J_p)$  at  $w$ , the spread of the layer  $J_k$ , denoted by  $\text{spr}(J_k)$ , is defined as

$$\text{spr}(J_k) \equiv \frac{\max(\delta_{J_k})}{\min(\delta_{J_k})}, \quad \forall k = 1, \dots, p.$$

### 3.3 Relation between crossover events, the AS step and the LLS directions

In this subsection we develop some variants of Lemma 17 of Vavasis and Ye [29] which are particularly suitable for our analysis in this paper. More specifically, we develop two estimates on the number of iterations that the MTY P-C algorithm needs to perform for some crossover event to occur. While the first estimate essentially depends on the size of the residual of the LLS step and the stepsize at the initial iterate, the second one, derived with the aid of Proposition 3.2, depends only on the size of the residual of the AS direction at the initial iterate.

We start by stating an immediate consequence of Lemma 17 of [29] whose proof can be found in Lemma 3.7 of Monteiro and Tsuchiya [18].

**Lemma 3.3** Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  for some  $\beta \in (0, 1)$  and an ordered partition  $J = (J_1, \dots, J_p)$  at  $w$  be given. Let  $\delta \equiv \delta(w)$ ,  $\mu = \mu(w)$  and  $(Rx^{\text{ll}}, Rs^{\text{ll}})$  denote the residual of the LLS direction  $(\Delta x^{\text{ll}}, \Delta y^{\text{ll}}, \Delta s^{\text{ll}})$  at  $w$  with respect to  $J$ . Then, for any

$$\mathcal{C}_q \geq (1 + \beta) \text{spr}(J_q) / (1 - \beta)^2$$

and for any  $\mu' \in (0, \mu)$  such that

$$\frac{\mu'}{\mu} \leq \frac{\|Rx_{J_q}^{\text{ll}}\|_{\infty} \|Rs_{J_q}^{\text{ll}}\|_{\infty}}{n^3 \mathcal{C}_q^2 \bar{\chi}_A^2},$$

the interval  $(\mu', \mu]$  contains a  $\mathcal{C}_q$ -crossover event.

Using the above result, we can now derive the main result of this subsection which provides two estimates on the number of iterations that the MTY P-C algorithm needs to perform until some crossover event happens. This result is a slight variation of Lemma 3.5 of Monteiro and Tsuchiya [18] which is more suitable for our analysis in this paper.

**Lemma 3.4** Suppose  $\beta \in (0, 1/4]$  and  $J = (J_1, \dots, J_p)$  is an ordered partition at  $w$ . Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  be such that  $\mu(w^{k+1}) \leq \mu(w) \leq \mu(w^k)$ , for some iteration index  $k$  of the MTY P-C algorithm. Let  $(Rx^{\text{ll}}, Rs^{\text{ll}})$  denote the residual of the LLS direction  $(\Delta x^{\text{ll}}, \Delta y^{\text{ll}}, \Delta s^{\text{ll}})$  at  $w$  with respect to  $J$ . Then, for every  $q \in \{1, \dots, p\}$  and every  $\mathcal{C}_q \geq (1 + \beta) \text{spr}(J_q) / (1 - \beta)^2$ , the following statements hold:

a) there exists an iteration index  $l > k$  such that

$$l - k = \mathcal{O} \left( \sqrt{n} \left( \log(\bar{\chi}_A + n) + \log \mathcal{C}_q + \log \left( \frac{\mu(w^{k+1}) / \mu(w)}{\|Rx_{J_q}^{\text{ll}}\|_{\infty} \|Rs_{J_q}^{\text{ll}}\|_{\infty}} \right) \right) \right) \quad (30)$$

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

b) if, in addition,

$$\text{gap}(J) \geq \max \left\{ 4n\bar{\chi}_A, \frac{24\sqrt{n}\bar{\chi}_A}{\varepsilon_{J_q}^a} \right\} \quad (31)$$

where  $\varepsilon_{J_q}^a \equiv \min \left\{ \|Rx_{J_q}^a\|_{\infty}, \|Rs_{J_q}^a\|_{\infty} \right\}$ , then the iteration index  $l$  above satisfies

$$l - k = \mathcal{O} \left( \sqrt{n} \left( \log(\bar{\chi}_A + n) + \log \mathcal{C}_q + \log[(\varepsilon_{J_q}^a)^{-1}] \right) \right). \quad (32)$$

*Proof.* To prove a), assume that the MTY P-C algorithm does not terminate at or before the  $l$ -th iteration. Lemma 3.3 guarantees that the interval  $(\mu(w^l), \mu(w)]$  contains a  $\mathcal{C}_q$ -crossover event, and hence that a  $\mathcal{C}_q$ -crossover event occurs between  $w^k$  and  $w^l$ , whenever

$$\frac{\mu(w^l)}{\mu(w)} = \frac{\mu(w^l)}{\mu(w^{k+1})} \frac{\mu(w^{k+1})}{\mu(w)} \leq \frac{\|Rx_{J_q}^{\text{ll}}\|_{\infty} \|Rs_{J_q}^{\text{ll}}\|_{\infty}}{n^3 \mathcal{C}_q^2 \bar{\chi}_A^2}. \quad (33)$$

Since, by Lemma 2.5,  $\mu(w^{j+1})/\mu(w^j) \leq 1 - \sqrt{\beta/n}$  for all  $j \geq 0$  and  $\mu(w) \leq \mu(w^k)$ , we conclude that (33) holds for any  $l$  satisfying

$$\log\left(\frac{\mu(w^{k+1})}{\mu(w)}\right) + (l - k - 1) \log\left(1 - \sqrt{\frac{\beta}{n}}\right) \leq \log\left[\frac{\|Rx_{J_q}^{\text{ll}}\|_{\infty} \|Rs_{J_q}^{\text{ll}}\|_{\infty}}{n^3 \mathcal{C}_q^2 \bar{\chi}_A^2}\right].$$

Now, using the fact that  $\log(1+x) < x$  for any  $x > -1$ , it is easy to see (30) holds for the smallest  $l$  satisfying the above inequality.

To prove b), it is sufficient to show that the bound in (30) is bounded above by the one in (32) when (31) holds. Indeed, by Proposition 3.2 and (31), it follows that

$$\max\left\{\|Rx^{\text{a}} - Rx^{\text{ll}}\|_{\infty}, \|Rs^{\text{a}} - Rs^{\text{ll}}\|_{\infty}\right\} \leq \frac{12\sqrt{n}\bar{\chi}_A}{\text{gap}(J)} \leq \frac{\varepsilon_{J_q}^{\text{a}}}{2}.$$

Hence, we have

$$\begin{aligned} \min\left\{\|Rx_{J_q}^{\text{ll}}\|_{\infty}, \|Rs_{J_q}^{\text{ll}}\|_{\infty}\right\} &\geq \min\left\{\|Rx_{J_q}^{\text{a}}\|_{\infty} - \|Rx^{\text{a}} - Rx^{\text{ll}}\|_{\infty}, \|Rs_{J_q}^{\text{a}}\|_{\infty} - \|Rs^{\text{a}} - Rs^{\text{ll}}\|_{\infty}\right\} \\ &\geq \min\left\{\|Rx_{J_q}^{\text{a}}\|_{\infty}, \|Rs_{J_q}^{\text{a}}\|_{\infty}\right\} - \frac{\varepsilon_{J_q}^{\text{a}}}{2} = \varepsilon_{J_q}^{\text{a}} - \frac{\varepsilon_{J_q}^{\text{a}}}{2} = \frac{\varepsilon_{J_q}^{\text{a}}}{2}. \end{aligned}$$

Using this estimate in (30) together with the fact that  $\mu(w^{k+1})/\mu(w) \leq 1$ , we conclude that the right hand side of (30) is bounded above by the right hand side of (32).  $\blacksquare$

### 3.4 Two important ordered partitions

In this subsection we describe two ordered partitions which are crucial in the analysis of this paper.

The first ordered partition is due to Vavasis and Ye [29]. Given a point  $w \in \mathcal{P}^{++} \times \mathcal{D}^{++}$  and a parameter  $\bar{g} \geq 1$ , this partition, which we refer to as the *VY  $\bar{g}$ -partition* at  $w$ , is defined as follows. Let  $(i_1, \dots, i_n)$  be an ordering of  $\{1, \dots, n\}$  such that  $\delta_{i_1} \leq \dots \leq \delta_{i_n}$ , where  $\delta = \delta(w)$ . For  $k = 2, \dots, n$ , let  $r_k \equiv \delta_{i_k}/\delta_{i_{k-1}}$  and define  $r_1 \equiv \infty$ . Let  $k_1 < \dots < k_p$  be all the indices  $k$  such that  $r_k > \bar{g}$  for all  $j = 1, \dots, p$ . The VY  $\bar{g}$ -partition  $J$  is then defined as  $J = (J_1, \dots, J_p)$ , where  $J_q \equiv \{i_{k_q}, i_{k_q+1}, \dots, i_{k_{q+1}-1}\}$  for all  $q = 1, \dots, p$ . More generally, given a subset  $I \subseteq \{1, \dots, n\}$ , we can similarly define the *VY  $\bar{g}$ -partition* of  $I$  at  $w$  by taking an ordering  $(i_1, \dots, i_m)$  of  $I$  satisfying  $\delta_{i_1} \leq \dots \leq \delta_{i_m}$  where  $m = |I|$ , defining the ratios  $r_1, \dots, r_m$  as above, and proceeding exactly as in the construction above to obtain an ordered partition  $J = (J_1, \dots, J_p)$  of  $I$ .

It is easy to see that the following result holds for the partition  $J$  described in the previous paragraph.

**Proposition 3.5** *Given a subset  $I \subseteq \{1, \dots, n\}$ , a point  $w \in \mathcal{P}^{++} \times \mathcal{D}^{++}$  and a constant  $\bar{g} \geq 1$ , the VY  $\bar{g}$ -partition  $J = (J_1, \dots, J_p)$  of  $I$  at  $w$  satisfies  $\text{gap}(J) > \bar{g}$  and  $\text{spr}(J_q) \leq \bar{g}^{|J_q|} \leq \bar{g}^n$  for all  $q = 1, \dots, p$ .*

The second ordered partition which is heavily used in our analysis is obtained as follows. Given a point  $w \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ , we first use (17) to compute the AS-bipartition  $(B, N) = (B(w), N(w))$  at  $w$ . Next, an order  $(i_1, \dots, i_n)$  of the index variables is chosen such that  $\delta_{i_1} \leq \dots \leq \delta_{i_n}$ . 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  $\mathcal{J}_1 = \{i_1, i_2\}$ ,  $\mathcal{J}_2 = \{i_3\}$ ,  $\mathcal{J}_3 = \{i_4, i_5\}$  and  $\mathcal{J}_4 = \{i_6, i_7\}$ . A partition obtained according to the above construction is clearly ordered at  $w$ . We refer to it as an *ordered AS-partition*, and denote it by  $\mathcal{J} = \mathcal{J}(w)$ .

Note that an ordered AS-partition is not uniquely determined since there can be more than one  $n$ -tuple  $(i_1, \dots, i_n)$  satisfying  $\delta_{i_1} \leq \dots \leq \delta_{i_n}$ . This situation happens exactly when there are two or more indices  $i$  with the same value for  $\delta_i$ . It can be easily seen that there exists a unique ordered AS-partition at  $w$  if and only if there do not exist  $i \in B(w)$  and  $j \in N(w)$  such that  $\delta_i = \delta_j$ . Hence, if the AS-bipartition  $(B(w), N(w))$  does not have the latter property, there can be multiple ordered AS-partition at  $w$ . In spite of this ambiguity, our analysis in this paper is valid for any chosen ordered AS-partition. So, there is no need for having the notion of ordered AS-partition uniquely defined although this can be easily accomplished.

## 4 Convergence Analysis of the MTY P-C Algorithm

In this section, we provide the proof of Theorem 2.6.

We first introduce some global constants which will be used in the convergence analysis of this section. Let

$$\mathcal{C} \equiv \frac{(1 + \beta)}{(1 - \beta)^2} (2\bar{g})^n, \quad \text{and} \quad \bar{g} \equiv \frac{24 \bar{\chi}_A n}{\tau}, \quad (34)$$

where

$$\tau = \tau(\beta) \equiv \frac{\beta \sqrt{(1 - \beta)(1 - 2\beta)}}{4 \sqrt{(1 + \beta)(1 + 2\beta)}}. \quad (35)$$

Note that in view of (13) and (17), we have

$$\varepsilon_\infty^a(w) = \max \{ \|Rx_N^a(w)\|_\infty, \|Rs_B^a(w)\|_\infty \}. \quad (36)$$

Clearly,  $\varepsilon_\infty^a(w)$  is an upper bound on the absolute value of the small components of the residual  $(Rx^a(w), Rs^a(w))$ . The next result gives a lower bound on the absolute value of the large components of the residual  $(Rx^a(w), Rs^a(w))$ . For a proof of this result, we refer the reader to Lemma 2.6 of Monteiro and Tsuchiya [18].

**Lemma 4.1** *Suppose that  $w = (x, y, s) \in \mathcal{N}(\beta)$  for some  $\beta \in (0, 1/4]$ . Then, we have*

$$\max \{ |Rx_i^a(w)|, |Rs_i^a(w)| \} \geq \frac{\sqrt{1 - \beta}}{2}$$

for all  $i = 1, \dots, n$ , or equivalently, in view of the definition of  $(B, N) \equiv (B(w), N(w))$ , we have

$$\min \left\{ \min_{i \in B} |Rx_i^a(w)|, \min_{i \in N} |Rs_i^a(w)| \right\} \geq \frac{\sqrt{1 - \beta}}{2}.$$

Lemma 3.4 gives a good idea of part of the effort that will be undertaken in this section, namely to find conditions under which the bounds (30) or (32) obtained in this result can be majorized by the quantity  $\mathcal{O}(n \log(\bar{\chi}_A + n) + \log(\log(\mu_0/\eta)))$ . We will break our analysis into the following three cases:

- i)  $\text{gap}(\mathcal{J}) \leq 2\bar{g}$  (Lemma 4.2);
- ii)  $\text{gap}(\mathcal{J}) \geq \bar{g}$  and  $\varepsilon_\infty^a \geq \tau\bar{g}/(\sqrt{n}\text{gap}(\mathcal{J}))$  (Lemma 4.5);
- iii)  $\text{gap}(\mathcal{J}) \geq \bar{g}$  and  $\varepsilon_\infty^a \leq \tau\bar{g}/(\sqrt{n}\text{gap}(\mathcal{J}))$  (Lemmas 4.8 and 4.9),

where  $\varepsilon_\infty^a \equiv \varepsilon_\infty^a(w)$  and  $\mathcal{J}$  is an ordered AS-partition at  $w$ . Observe that the above cases are not disjoint but they cover all possible situations. The motivation for letting the above three cases overlap comes from the proof of the last lemma of this section, namely Lemma 4.9.

We will now give an outline of the approaches used in each of the above cases. In case i), it is easy to see that at least one variable in  $N$  and one variable in  $B$  are in the same layer  $J_q$  of a VY  $2\bar{g}$ -partition  $J = (J_1, \dots, J_p)$  at  $w$ . We call a layer of this type a “mixed VY-layer”. The proof of Lemma 4.2 essentially shows that the existence of a mixed layer implies that the quantity  $\varepsilon_{J_q}^a$  in Lemma 3.4(b) is not too small and that a  $\mathcal{C}$ -crossover event must occur within  $\mathcal{O}(n^{1.5} \log(\bar{\chi}_A + n))$  iterations of the MTY P-C algorithm.

In case ii), the quantity  $\text{gap}(\mathcal{J})$  is sufficiently large to guarantee that the AS direction can be well-approximated by a LLS direction at  $w$  with respect to a suitable ordered partition at  $w$  obtained from an ordered AS-partition at  $w$  by breaking one of its layers into smaller layers. Using the close proximity of these two directions and the fact that a step along this LLS direction followed by at most  $\mathcal{O}(n^{1.5} \log(\bar{\chi}_A + n))$  regular MTY P-C steps yield a  $\mathcal{C}$ -crossover event, we show that case ii) also implies that a  $\mathcal{C}$ -crossover event must occur within  $\mathcal{O}(n^{1.5} \log(\bar{\chi}_A + n))$  iterations of the MTY P-C algorithm.

In case iii), the proximity between the AS direction and the LLS direction mentioned above is not small enough to guarantee that the reduction of the duality gap along these two directions are of the same order of magnitude. However, in this case we manage to show that within  $\mathcal{O}(\log(\log(\mu_0/\eta)))$  iterations of the MTY P-C algorithm we return to a situation in which either a mixed layer arises, or case ii) above holds, or an index changes status (moves from  $B$  to  $N$  or vice-versa) from one iteration to the next. The same kind of techniques used to handle case i) can be used to show that the latter possibility also implies that a  $\mathcal{C}$ -crossover event must occur within  $\mathcal{O}(n^{1.5} \log(\bar{\chi}_A + n))$  iterations of the MTY P-C algorithm. Overall, we conclude that case iii) also implies that a  $\mathcal{C}$ -crossover event must occur within  $\mathcal{O}(n^{1.5} \log(\bar{\chi}_A + n) + \log(\log(\mu_0/\eta)))$  iterations of the MTY P-C algorithm.

The first result below is a slight variation of Lemma 4.2 of [18]. It considers case i) above, which can be handled by applying Lemma 3.3(b) with the ordered partition  $J = (J_1, \dots, J_p)$  chosen to be the VY  $2\bar{g}$ -partition at  $w$ .

**Lemma 4.2** *Suppose  $\beta \in (0, 1/4]$  and  $w = (x, y, s) \in \mathcal{N}(\beta)$  is such that  $\mu(w^{k+1}) \leq \mu(w) \leq \mu(w^k)$  for some iteration index  $k$  of the MTY P-C algorithm. Assume that  $\text{gap}(\mathcal{J}) \leq 2\bar{g}$ , where  $\mathcal{J} = (J_1, \dots, J_r)$  is an ordered AS-partition at  $w$ . Then, 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  $\mathcal{C}$ -crossover event occurs between  $w^k$  and  $w^l$ , or the algorithm terminates at or before the  $l$ -th iteration.*

*Proof.* The proof that the required iteration index  $l$  exists is based on Lemma 3.4(b). Indeed, let  $J = (J_1, \dots, J_p)$  be a VY  $2\bar{g}$ -partition at  $w$ . The assumption  $\text{gap}(\mathcal{J}) \leq 2\bar{g}$  implies the existence of two indices  $i, j$  lying in some layer  $J_q$  of  $J$ , with one contained  $B(w)$  and the other in  $N(w)$ . Without loss of generality, assume that  $i \in B(w)$  and  $j \in N(w)$ . By Lemma 4.1, we then have

$|Rx_i^a| \geq \sqrt{1-\beta}/2$  and  $|Rs_j^a| \geq \sqrt{1-\beta}/2$ . Since  $i, j \in J_q$ , this implies that

$$\varepsilon_{J_q}^a \equiv \min \left\{ \|Rx_{J_q}^a\|_\infty, \|Rs_{J_q}^a\|_\infty \right\} \geq \frac{\sqrt{1-\beta}}{2}. \quad (37)$$

Using the above inequality, the fact that  $\text{gap}(J) \geq 2\bar{g}$  and relations (34) and (35), we easily see that (31) holds. Since by Proposition 3.5 the spread of every layer of a VY  $2\bar{g}$ -partition at  $w$  is bounded above by  $(2\bar{g})^n$ , we conclude that  $\text{spr}(J_q) \leq (2\bar{g})^n$ . Hence, we may set  $\mathcal{C}_q = \mathcal{C} \equiv (1+\beta)(2\bar{g})^n/(1-\beta)^2$  in Lemma 3.4, from which it follows that

$$\log(\mathcal{C}_q) = \mathcal{O}(n \log \bar{g}) = \mathcal{O}(n \log(\bar{\chi}_A + n)), \quad (38)$$

where the last equality is due to (34). The result now follows from Lemma 3.4(b) by noting that the bound in (32) is  $\mathcal{O}(n^{1.5} \log(\bar{\chi}_A + n))$  in view of (37) and (38).  $\blacksquare$

From now on we consider cases ii) and iii), i.e. the situation when  $\text{gap}(\mathcal{J}) \geq \bar{g}$ . For the sake of future reference, we note that when  $\text{gap}(\mathcal{J}) \geq \bar{g}$  we have

$$\frac{\tau \bar{g}}{\sqrt{n} \text{gap}(\mathcal{J})} \leq \tau \leq \frac{\sqrt{1-2\beta}}{4}. \quad (39)$$

The following lemma is a slight variation of Lemma 4.3 of [18]. It provides an upper bound on the right hand side of (30) in terms of the residual of the LLS step with respect to  $\mathcal{J}$ . Note that for this result we assume only that  $\text{gap}(\mathcal{J}) \geq \bar{g}$ .

**Lemma 4.3** *Suppose  $\beta \in (0, 1/4]$  and  $w = (x, y, s) \in \mathcal{N}(\beta)$  is such that  $\mu(w^{k+1}) \leq \mu(w) \leq \mu(w^k)$  for some iteration index  $k$  of the MTY P-C algorithm. Assume that  $\text{gap}(\mathcal{J}) \geq \bar{g}$ , where  $\mathcal{J} = (\mathcal{J}_1, \dots, \mathcal{J}_r)$  is an ordered AS-partition at  $w$ . Let  $(Rx^1, Rs^1)$  denote the residual of the LLS direction at  $w$  with respect to  $\mathcal{J}$  and define*

$$\varepsilon_\infty^1 \equiv \max \left\{ \|Rx_N^1\|_\infty, \|Rs_B^1\|_\infty \right\}. \quad (40)$$

*Then, there exists an iteration index  $l > k$  such that*

$$l - k = \mathcal{O} \left( n^{1.5} \log(\bar{\chi}_A + n) + \sqrt{n} \log \left( \frac{\mu(w^{k+1})/\mu(w)}{\varepsilon_\infty^1} \right) \right) \quad (41)$$

*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.* Assume without loss of generality that  $\varepsilon_\infty^1 = \|Rx_N^1\|_\infty$ ; the case in which  $\varepsilon_\infty^1 = \|Rs_B^1\|_\infty$  can be proved similarly. Then,  $\varepsilon_\infty^1 = |Rx_i^1|$  for some  $i \in N$ . Let  $\mathcal{J}_t$  be the layer of  $\mathcal{J}$  containing the index  $i$  and note that

$$\varepsilon_\infty^1 = |Rx_i^1| = \|Rx_{\mathcal{J}_t}^1\|_\infty \leq \|Rx_{\mathcal{J}_t}^1\|. \quad (42)$$

Now, let  $\mathcal{I} = (\mathcal{I}_1, \dots, \mathcal{I}_p)$  be the VY  $\bar{g}$ -partition of  $\mathcal{J}_t$  at  $w$  and consider the ordered partition  $\mathcal{J}'$  defined as

$$\mathcal{J}' \equiv (\mathcal{J}_1, \dots, \mathcal{J}_{t-1}, \mathcal{I}_1, \dots, \mathcal{I}_p, \mathcal{J}_{t+1}, \dots, \mathcal{J}_r).$$

Let  $(Rx^{\text{ll}}, Rs^{\text{ll}})$  denote the residual of the LLS direction at  $w$  with respect to  $\mathcal{J}'$ . Using the definition of the LLS step, it is easy to see that  $Rx_{\mathcal{J}_j}^1 = Rx_{\mathcal{J}_j}^{\text{ll}}$  for all  $j = t+1, \dots, r$ . Moreover, we have

$\|Rx^1_{\mathcal{J}_t}\| \leq \|Rx^{\text{ll}}_{\mathcal{J}_t}\|$  since  $\|Rx^1_{\mathcal{J}_t}\|$  is the optimal value of the least squares problem which determines the  $\Delta x^1_{\mathcal{J}_t}$ -component of the LLS step with respect to  $\mathcal{J}$ , whereas  $\|Rx^{\text{ll}}_{\mathcal{J}_t}\|$  is the objective value at a certain feasible solution for the same least squares problem. Hence, for some  $q \in \{1, \dots, p\}$ , we have

$$\|Rx^{\text{ll}}_{\mathcal{I}_q}\|_{\infty} = \|Rx^{\text{ll}}_{\mathcal{J}_t}\|_{\infty} \geq \frac{1}{\sqrt{|\mathcal{J}_t|}} \|Rx^{\text{ll}}_{\mathcal{J}_t}\| \geq \frac{1}{\sqrt{n}} \|Rx^{\text{ll}}_{\mathcal{J}_t}\| \geq \frac{1}{\sqrt{n}} \|Rx^1_{\mathcal{J}_t}\|. \quad (43)$$

Combining (42) and (43), we then obtain

$$\|Rx^{\text{ll}}_{\mathcal{I}_q}\|_{\infty} \geq \frac{1}{\sqrt{n}} \varepsilon_{\infty}^1. \quad (44)$$

Let us now bound the quantity  $\|Rs^{\text{ll}}_{\mathcal{I}_q}\|_{\infty}$  from below. Using the triangle inequality for norms, Lemma 4.1 together with the fact that  $\mathcal{I}_q \subseteq N$ , Proposition 3.2 together with the fact that  $\text{gap}(\mathcal{J}') \geq \bar{g} = 24\bar{\chi}_A n / \tau \geq 96\bar{\chi}_A \sqrt{n}$ , where the last inequality is due to (35), we obtain

$$\|Rs^{\text{ll}}_{\mathcal{I}_q}\|_{\infty} \geq \|Rs^{\text{a}}_{\mathcal{I}_q}\|_{\infty} - \|Rs^{\text{ll}}_{\mathcal{I}_q} - Rs^{\text{a}}_{\mathcal{I}_q}\|_{\infty} \geq \frac{1}{4} - \frac{12\sqrt{n}\bar{\chi}_A}{\text{gap}(\mathcal{J}')} \geq \frac{1}{4} - \frac{1}{8} = \frac{1}{8}. \quad (45)$$

Also, note that by (34) we have

$$\log \mathcal{C} = \mathcal{O}(n \log(\bar{\chi}_A + n)). \quad (46)$$

The result now follows from Lemma 3.4(a) with  $J = \mathcal{J}'$  and  $\mathcal{C}_q = \mathcal{C}$ , the observation that Proposition 3.5 and (34) imply that  $\mathcal{C} \geq (1 + \beta)\bar{g}^n / (1 - \beta)^2 \geq (1 + \beta)\text{spr}(\mathcal{I}_q) / (1 - \beta)^2$ , and the fact that the estimates (44)-(46) imply that the bound in (30) is majorized by the one in (41). ■

Our goal now will be to estimate the second logarithm that appears in the iteration-complexity bound (41). The next result gives conditions under which  $\varepsilon_{\infty}^{\text{a}} = \mathcal{O}(\varepsilon_{\infty}^1)$ .

**Lemma 4.4** *Let  $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$  be given and let  $\mathcal{J} = (\mathcal{J}_1, \dots, \mathcal{J}_r)$  denote an ordered AS-partition at  $w$ . Let  $\varepsilon_{\infty}^{\text{a}} \equiv \varepsilon_{\infty}^{\text{a}}(w)$  and  $\varepsilon_{\infty}^1$  be defined according to (40). If*

$$\text{gap}(\mathcal{J}) \geq \max \left\{ 4n\bar{\chi}_A, \frac{24\sqrt{n}\bar{\chi}_A}{\varepsilon_{\infty}^{\text{a}}} \right\} \quad (47)$$

then,  $\varepsilon_{\infty}^{\text{a}} \leq 2\varepsilon_{\infty}^1$ .

*Proof.* Let  $(Rx^{\text{a}}, Rs^{\text{a}})$  and  $(Rx^1, Rs^1)$  denote the residuals of the AS direction at  $w$  and the LLS direction at  $w$  with respect to  $\mathcal{J}$ , respectively. By Proposition 3.2 and condition (47), we have

$$\max \left\{ \|Rx^{\text{a}} - Rx^1\|_{\infty}, \|Rs^{\text{a}} - Rs^1\|_{\infty} \right\} \leq \frac{12\sqrt{n}\bar{\chi}_A}{\text{gap}(\mathcal{J})} \leq \frac{\varepsilon_{\infty}^{\text{a}}}{2}.$$

Hence, we have

$$\begin{aligned} \varepsilon_{\infty}^1 &\equiv \max \left\{ \|Rx^1_N\|_{\infty}, \|Rs^1_B\|_{\infty} \right\} \\ &\geq \max \left\{ \|Rx^{\text{a}}_N\|_{\infty} - \|Rx^{\text{a}}_N - Rx^1_N\|_{\infty}, \|Rs^{\text{a}}_B\|_{\infty} - \|Rs^{\text{a}}_B - Rs^1_B\|_{\infty} \right\} \\ &\geq \max \left\{ \|Rx^{\text{a}}_N\|_{\infty}, \|Rs^{\text{a}}_B\|_{\infty} \right\} - \frac{\varepsilon_{\infty}^{\text{a}}}{2} = \varepsilon_{\infty}^{\text{a}} - \frac{\varepsilon_{\infty}^{\text{a}}}{2} = \frac{\varepsilon_{\infty}^{\text{a}}}{2}. \end{aligned}$$

■

We are now ready to state and prove the result which takes care of case ii).

**Lemma 4.5** *Suppose  $\beta \in (0, 1/4]$  and  $w = w^k \in \mathcal{N}(\beta)$  is an iterate of the MTY P-C algorithm. Assume that  $\text{gap}(\mathcal{J}) \geq \bar{g}$  and  $\varepsilon_\infty^a \geq \tau \bar{g} / (\sqrt{n} \text{gap}(\mathcal{J}))$ , where  $\varepsilon_\infty^a \equiv \varepsilon_\infty^a(w)$  and  $\mathcal{J} = (\mathcal{J}_1, \dots, \mathcal{J}_r)$  is an ordered AS-partition at  $w$ . Then, 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.*

*Proof.* The proof consists of showing that the bound in (41) is  $\mathcal{O}(n^{1.5} \log(\bar{\chi}_A + n))$  under the assumptions of the lemma. Indeed, first note that the condition that  $\varepsilon_\infty^a \geq \tau \bar{g} / \sqrt{n} \text{gap}(\mathcal{J})$  and (34) clearly implies (47). Hence, by Lemma 4.4, it follows that  $\varepsilon_\infty^a \leq 2\varepsilon_\infty^1$ . This together with the fact that  $w = w^k$  and Lemma 2.5 imply that

$$\frac{\mu(w^{k+1})/\mu(w)}{\varepsilon_\infty^1} \leq \frac{\sqrt{n} \varepsilon_\infty^a}{\beta \varepsilon_\infty^1} \leq \frac{2\sqrt{n}}{\beta}.$$

Hence, the bound in (41) is  $\mathcal{O}(n^{1.5} \log(\bar{\chi}_A + n))$  and the result now follows from Lemma 4.3.  $\blacksquare$

We now consider the case in which an iterate  $w = w^k$  of the MTY P-C algorithm satisfies  $\text{gap}(\mathcal{J}) \geq \bar{g}$  and  $\varepsilon_\infty^a \leq \tau \bar{g} / (\sqrt{n} \text{gap}(\mathcal{J}))$ , where  $\mathcal{J} = (\mathcal{J}_1, \dots, \mathcal{J}_r)$  is an ordered AS-partition at  $w$ . By the definition of the quantity  $\text{gap}(\mathcal{J})$ , there exists two indices  $i$  and  $j$ , one lying in  $B(w)$  and the other in  $N(w)$ , such that  $\delta_i/\delta_j = \text{gap}(\mathcal{J})$ . We consider the following two possibilities in our analysis below:

P1) there exist  $i \in B(w)$  and  $j \in N(w)$  such that  $\delta_i/\delta_j = \text{gap}(\mathcal{J})$ ;

P2) there exist  $i \in N(w)$  and  $j \in B(w)$  such that  $\delta_i/\delta_j = \text{gap}(\mathcal{J})$ .

We will first consider possibility P1) since it is the easiest one to analyze. We start with the following technical lemma which studies how the ratio  $\delta_i/\delta_j$  with  $i \in B$  and  $j \in N$  varies from one iteration to the next one.

**Lemma 4.6** *Suppose  $\beta \in (0, 1/4]$  and assume that  $w = w^k$  and  $w^+ = w^{k+1}$  are two consecutive iterates of the MTY P-C algorithm. Assume that  $\text{gap}(\mathcal{J}) \geq \bar{g}$  and  $\varepsilon_\infty^a \leq \sqrt{1 - \beta}/2$ , where  $\varepsilon_\infty^a \equiv \varepsilon_\infty^a(w)$  and  $\mathcal{J} = (\mathcal{J}_1, \dots, \mathcal{J}_r)$  is an ordered AS-partition at  $w$ . Then, for every  $i \in B(w)$  and  $j \in N(w)$ , we have*

$$\frac{\delta_i^+}{\delta_j^+} \leq \frac{\sqrt{n}}{\tau} \frac{\delta_i}{\delta_j} \varepsilon_\infty^a. \quad (48)$$

where  $\delta \equiv \delta(w)$ ,  $\delta^+ \equiv \delta(w^+)$  and  $\tau = \tau(\beta)$  is the constant defined in (35).

*Proof.* Let  $(\Delta x^a, \Delta y^a, \Delta s^a)$  denote the AS direction at  $w$  and let  $(Rx^a, Rs^a) \equiv (Rx^a(w), Rs^a(w))$ . Using Lemma 2.5, (13) and the fact that  $i \in B$ , we obtain

$$\begin{aligned} s_i^+ &= s_i + \alpha_a \Delta s_i^a = (1 - \alpha_a) s_i + \alpha_a (s_i + \Delta s_i^a) = \sqrt{\mu} \delta_i \left[ (1 - \alpha_a) \frac{x_i^{1/2} s_i^{1/2}}{\sqrt{\mu}} + \alpha_a R s_i^a \right] \\ &\leq \sqrt{\mu} \delta_i \left[ \frac{\sqrt{n} \varepsilon_\infty^a}{\beta} \sqrt{1 + \beta} + \varepsilon_\infty^a \right] \leq \sqrt{\mu} \delta_i \varepsilon_\infty^a \frac{2\sqrt{(1 + \beta)n}}{\beta}. \end{aligned} \quad (49)$$

Now, using relations (6), (11) and (36), the fact that  $j \in N$  and  $\alpha_a \leq 1$ , and the assumption that  $\varepsilon_\infty^a \leq \sqrt{1-\beta}/2$ , we obtain

$$\begin{aligned} s_j^+ &= s_j + \alpha_a \Delta s_j^a = \sqrt{\mu} \delta_j \left[ \frac{x_j^{1/2} s_j^{1/2}}{\sqrt{\mu}} + \alpha_a \frac{\delta_j^{-1} \Delta s_j^a}{\sqrt{\mu}} \right] \\ &= \sqrt{\mu} \delta_j \left[ \frac{x_j^{1/2} s_j^{1/2}}{\sqrt{\mu}} - \alpha_a R x_j^a \right] \geq \sqrt{\mu} \delta_j \left[ \sqrt{1-\beta} - \varepsilon_\infty^a \right] \geq \sqrt{\mu} \delta_j \frac{\sqrt{1-\beta}}{2}. \end{aligned} \quad (50)$$

Combining (49) and (50), and using the fact that  $w^+ \in \mathcal{N}(2\beta)$  and relations (6) and (35), we obtain

$$\frac{\delta_i^+}{\delta_j^+} = \frac{(x_j^+ s_j^+)^{1/2}}{(x_i^+ s_i^+)^{1/2}} \frac{s_i^+}{s_j^+} \leq \frac{\sqrt{1+2\beta}}{\sqrt{1-2\beta}} \frac{s_i^+}{s_j^+} \leq \frac{4\sqrt{n} \sqrt{(1+\beta)(1+2\beta)}}{\beta \sqrt{(1-\beta)(1-2\beta)}} \frac{\delta_i}{\delta_j} \varepsilon_\infty^a = \frac{\sqrt{n}}{\tau} \frac{\delta_i}{\delta_j} \varepsilon_\infty^a. \quad \blacksquare$$

**Lemma 4.7** *Suppose that  $\beta \in (0, 1/4]$  and that  $w = w^k \in \mathcal{N}(\beta)$  and  $w^+ = w^{k+1} \in \mathcal{N}(\beta)$  are two consecutive iterates of the MTY P-C algorithm. Define  $(B, N) \equiv (B(w), N(w))$  and  $(B^+, N^+) \equiv (B(w^+), N(w^+))$  and let  $\mathcal{J}$  and  $\mathcal{J}^+$  denote ordered AS-partitions at  $w$  and  $w^+$ , respectively. Assume that  $\text{gap}(\mathcal{J}) \geq \bar{g}$  and  $\varepsilon_\infty^a(w) \leq \sqrt{1-2\beta}/4$ . If either one of the conditions a) or b) below hold then, 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.*

a)  $B \cap N^+ \neq \emptyset$  or  $N \cap B^+ \neq \emptyset$ ;

b) there exist indices  $i$  and  $j$ , one lying in  $B(w)$  and the other in  $N(w)$ , such that  $\delta_i/\delta_j \geq \bar{g}$  and  $\delta_i^+/\delta_j^+ \leq \bar{g}$ .

*Proof.* We will first show that the conclusion of the lemma holds under condition a). Let  $w : [0, 1] \rightarrow \mathcal{N}(2\beta)$  be a continuous path such that  $w(0) = w$ ,  $w(1) = w^+$  and  $\mu(w^+) \leq \mu(w(t)) \leq \mu(w)$  for all  $t \in [0, 1]$ , e.g. consider the path that traces the line segment from  $w$  to  $w + \alpha_a \Delta w^a$  and then the one from  $w + \alpha_a \Delta w^a$  to  $w^+$ . (It is straightforward to verify that these two segments lie in  $\mathcal{N}(2\beta)$ .) We will show more generally that if there exists  $0 < t \leq 1$  such that either  $B \cap N(w(t)) \neq \emptyset$  or  $N \cap B(w(t)) \neq \emptyset$  (for  $t = 1$  this is condition a)), then the conclusion of lemma follows. Indeed, assume that for some  $0 < t \leq 1$ , there exists an index  $j$  in the set  $B \cap N(w(t))$ . (The proof is similar for the case in which  $j \in N \cap B(w(t))$ .) Since  $j \in B$  and, by assumption,  $\varepsilon_\infty^a \leq \sqrt{1-2\beta}/4$ , we have  $|Rs_j^a(w(0))| = |Rs_j^a(w)| \leq \varepsilon_\infty^a \leq \sqrt{1-2\beta}/4$ . Moreover, since  $j \in N(w(t))$ , we have  $|Rs_j^a(w(t))| \geq \sqrt{1-\beta}/2$  in view of Lemma 4.1. The intermediate value theorem applied to the function  $t \rightarrow |Rs_j^a(w(t))|$  implies the existence of some  $\bar{t} \in [0, t]$  such that  $|Rs_j^a(w(\bar{t}))| = \sqrt{1-2\beta}/4$ . Letting  $\bar{w} \equiv w(\bar{t})$ , we have  $\bar{w} \in \mathcal{N}(2\beta)$  and  $|Rs_j^a(\bar{w})| = \sqrt{1-2\beta}/4$ . By Lemma 4.1 with  $\beta$  replaced by  $2\beta$  we have that  $\max\{|Rx_j^a(\bar{w})|, |Rs_j^a(\bar{w})|\} \geq \sqrt{1-2\beta}/2$ . Since  $|Rs_j^a(\bar{w})| = \sqrt{1-2\beta}/4 < \sqrt{1-2\beta}/2$ , we must have  $|Rx_j^a(\bar{w})| \geq \sqrt{1-2\beta}/2$ . We thus proved that  $\varepsilon_\infty^a(\bar{w}) \geq \min\{|Rx_j^a(\bar{w})|, |Rs_j^a(\bar{w})|\} \geq \sqrt{1-2\beta}/4$ . Noting (39), the conclusion of the lemma now follows from Lemma 4.5 with  $w = \bar{w}$ .

We will now show that the conclusion of the lemma holds under condition b). Without loss of generality, assume that  $i \in B$  and  $j \in N$  is such that  $\delta_i/\delta_j \geq \bar{g}$  and  $\delta_i^+/\delta_j^+ \leq \bar{g}$ . The intermediate

value theorem applied to the function  $t \rightarrow \delta_i(w(t))/\delta_j(w(t))$  implies the existence of some  $\bar{t} \in [0, 1]$  such that  $\delta_i(w(\bar{t}))/\delta_j(w(\bar{t})) = \bar{g}$ . If either  $i \in N(w(\bar{t}))$  or  $j \in B(w(\bar{t}))$  then the conclusion of the lemma holds in view of what we have already shown in the previous paragraph. Consider now the case in which  $i \in B(w(\bar{t}))$  and  $j \in N(w(\bar{t}))$ . Since  $\delta_i(w(\bar{t}))/\delta_j(w(\bar{t})) = \bar{g}$ , this case implies that  $\text{gap}(\mathcal{J}(w(\bar{t}))) \leq \bar{g}$ . Thus, the conclusion of the lemma follows from Lemma 4.2 with  $w = w(\bar{t})$ . ■

It is worth noting that Lemma 4.7(a) gives a simple and intuitive scaling-invariant sufficient condition for a crossover event to occur within a reasonable number of iterations, namely that an index changes its status from the current iteration to the next one. Also, note that if neither conditions a) nor b) of Lemma 4.7 occur then we must have  $\mathcal{J} = \mathcal{J}^+$ , i.e. the ordered AS-partition does not change during the iteration from  $w$  to  $w^+$ . This last observation would be used in the proof of Lemma 4.9 below.

The following result which considers case P1) stated just before Lemma 4.6 follows almost as an immediate consequence of the above lemma.

**Lemma 4.8** *Suppose  $\beta \in (0, 1/4]$  and  $w = w^k \in \mathcal{N}(\beta)$  is an iterate of the MTY P-C algorithm. Assume that  $\text{gap}(\mathcal{J}) \geq \bar{g}$  and  $\varepsilon_\infty^a \leq \tau\bar{g}/(\sqrt{n}\text{gap}(\mathcal{J}))$ , where  $\varepsilon_\infty^a \equiv \varepsilon_\infty^a(w)$  and  $\mathcal{J} = (\mathcal{J}_1, \dots, \mathcal{J}_r)$  is an ordered AS-partition at  $w$ . Assume also that case P1) stated just before Lemma 4.6 holds. Then, 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.*

*Proof.* Since we are assuming case P1), there exist  $i \in B$  and  $j \in N$  such that  $\delta_i/\delta_j = \text{gap}(\mathcal{J}) \geq \bar{g}$ . The assumptions on  $\text{gap}(\mathcal{J})$  and  $\varepsilon_\infty^a$  together with (39) imply that  $\varepsilon_\infty^a \leq \sqrt{1 - 2\beta}/4 \leq \sqrt{1 - \beta}/2$ . Hence, by Lemma 4.6 and the assumption on  $\varepsilon_\infty^a$ , we have

$$\frac{\delta_i^+}{\delta_j^+} \leq \frac{\sqrt{n}}{\tau} \frac{\delta_i}{\delta_j} \varepsilon_\infty^a \leq \frac{\sqrt{n}}{\tau} \text{gap}(\mathcal{J}) \varepsilon_\infty^a \leq \bar{g}. \quad (51)$$

The conclusion of the lemma follows from Lemma 4.7(b). ■

We now deal with case P2). In contrast to the other cases we considered above, the bound we derive in this case on the number of iterations needed to guarantee the occurrence of a crossover event depends not only on the quantity  $n^{1.5} \log(\bar{\chi}_A + n)$  but also on the term  $\log(\log(\mu_0/\eta))$ .

**Lemma 4.9** *Suppose  $\beta \in (0, 1/4]$  and  $w = w^k \in \mathcal{N}(\beta)$  is an iterate of the MTY P-C algorithm. Assume that  $\text{gap}(\mathcal{J}) \geq \bar{g}$  and  $\varepsilon_\infty^a \leq \tau\bar{g}/(\sqrt{n}\text{gap}(\mathcal{J}))$ , where  $\varepsilon_\infty^a \equiv \varepsilon_\infty^a(w)$  and  $\mathcal{J} = (\mathcal{J}_1, \dots, \mathcal{J}_r)$  is an ordered AS-partition at  $w$ . Assume also that case P2) stated just before Lemma 4.6 holds. Then, there exists an iteration index  $l > k$  such that  $l - k = \mathcal{O}(n^{1.5} \log(\bar{\chi}_A + n) + \log(\log(\mu_0/\eta)))$  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.* Let  $\hat{k}$  be the first index greater than or equal to  $k$  such that: either  $w^{\hat{k}}$  satisfies the assumptions of one of the Lemmas 4.2, 4.5 or 4.7; or,  $w^{\hat{k}+1}$  satisfies the assumptions of Lemma 4.8. We will show that  $\hat{k} - k = \mathcal{O}(\log(\log(\mu_0/\eta)))$ , from which the lemma immediately follows in view of the contents of these four lemmas. Indeed, let  $k'$  such that  $k \leq k' \leq \hat{k} - 1$  be given. By the definition

of  $\hat{k}$ , we know that  $w^{k'}$  violates the assumptions of Lemmas 4.2, 4.5 or 4.7 simultaneously. Then, it follows that

$$\text{gap}(\mathcal{J}_{k'}) > 2\bar{g}, \quad \varepsilon_\infty^a(w^{k'}) < \frac{\tau\bar{g}}{\sqrt{n} \text{gap}(\mathcal{J}_{k'})}, \quad \text{and} \quad \mathcal{J}_{k'+1} = \mathcal{J}_{k'}, \quad (52)$$

where  $\mathcal{J}_{k'}$  denotes an ordered AS-partition at  $w^{k'}$ . Moreover, the definition of  $\hat{k}$  also implies that  $w^{k'+1}$  violates the assumptions of Lemma 4.8. Hence, case P2) must hold at  $w^{k'+1}$ , i.e. there exist indices  $i = i(w^{k'+1}) \in N$  and  $j = j(w^{k'+1}) \in B$  such that

$$\text{gap}(\mathcal{J}_{k'+1}) = \frac{\delta_i(w^{k'+1})}{\delta_j(w^{k'+1})}.$$

By Lemma 4.6 and (52), we have

$$\begin{aligned} \text{gap}(\mathcal{J}_{k'+1}) &= \frac{\delta_i(w^{k'+1})}{\delta_j(w^{k'+1})} \geq \frac{\tau}{\sqrt{n} \varepsilon_\infty^a(w^{k'})} \frac{\delta_i(w^{k'})}{\delta_j(w^{k'})} \\ &\geq \frac{\tau}{\sqrt{n} \varepsilon_\infty^a(w^{k'})} \text{gap}(\mathcal{J}_{k'}) \geq \frac{1}{\bar{g}} [\text{gap}(\mathcal{J}_{k'})]^2. \end{aligned}$$

Using the fact that the above inequality holds for all  $k'$  such that  $k \leq k' \leq \hat{k} - 1$ , it is easy to verify that for every  $k'$  such that  $k \leq k' \leq \hat{k}$ ,

$$\log\left(\frac{\text{gap}(\mathcal{J}_{k'})}{\bar{g}}\right) \geq 2^{k'-k} \log\left(\frac{\text{gap}(\mathcal{J}_k)}{\bar{g}}\right) \geq 2^{k'-k} \log 2,$$

where the last inequality is due to the first relation in (52) with  $k' = k$ . Now, using Lemma 2.5, the second inequality in (52) with  $k' = \hat{k} - 1$ , the above inequality with  $k' = \hat{k} - 1$  and the fact that  $\tau \leq \beta$ , we obtain

$$\begin{aligned} \log \frac{\mu_0}{\eta} &\geq \log \frac{\mu_{\hat{k}-1}}{\mu_{\hat{k}}} \geq \log\left(\frac{\beta}{\sqrt{n} \varepsilon_\infty^a(w^{\hat{k}-1})}\right) \geq \log\left(\frac{\beta \text{gap}(\mathcal{J}_{\hat{k}-1})}{\tau\bar{g}}\right) \\ &\geq \log\left(\frac{\text{gap}(\mathcal{J}_{\hat{k}-1})}{\bar{g}}\right) \geq 2^{\Delta k} \log 2, \end{aligned}$$

where  $\Delta k \equiv \hat{k} - k - 1$ . Taking logarithms of both sides of the above inequality, we then conclude that  $\hat{k} - k = \Delta k + 1 = \mathcal{O}(\log(\log(\mu_0/\eta)))$ .  $\blacksquare$

Theorem 2.6 is now proved as follows. The assumptions of Lemmas 4.2, 4.5, 4.8 and 4.9 cover all possible cases that could happen at an iterate of the MTY P-C algorithm. The first three cases ensure either termination or occurrence of a  $\mathcal{C}$  crossover event within  $\mathcal{O}(n^{1.5} \log(\bar{\chi}_A + n))$  iterations whereas the fourth case ensures either termination or occurrence of a  $\mathcal{C}$  crossover event within  $\mathcal{O}(\log(\log(\mu^0/\eta)) + n^{1.5} \log(\bar{\chi}_A + n))$  iterations. Since the number of disjoint intervals which contain a  $\mathcal{C}$ -crossover event is  $\mathcal{O}(n^2)$ , the algorithm has to terminate within  $\mathcal{O}(n^2 \log(\log(\mu^0/\eta)) + n^{3.5} \log(\bar{\chi}_A + n))$  iterations.

## 5 Implications of the main result under the Turing Machine Model

Now we consider a few implications of our result under the Turing machine model. Assume that all data of  $A, b$ , and  $c$  are integral. Let  $L$  and  $L_A$  be the input size of  $(A, b, c)$  and  $A$ , respectively. It is well-known that  $\bar{\chi}_A^* \leq \bar{\chi}_A \leq 2^{\mathcal{O}(L_A)}$  and  $\xi(A, b, c) \geq 2^{-\mathcal{O}(L)}$ . As was mentioned before, we have  $\zeta(A, (B_*, N_*)) \leq \bar{\chi}_A$ . Therefore, we have the following corollary which immediately follows from Theorem 2.9.

**Corollary 5.1** *Assume that the data  $(A, b, c)$  is integral, and let  $L$  and  $L_A$  be defined as in the paragraph above. For some  $\beta \in (0, 1/4]$ , suppose that a point  $w^0 \in \mathcal{N}(\beta)$  such that  $\mu(w^0) = 2^{\mathcal{O}(L)}$  is given. Then, the version of the MTY P-C algorithm, in which the FT Procedure is invoked at every iterate  $w^k$ , started from  $w^0$  finds a primal-dual strictly complementary optimal solution  $w^*$  in at most*

$$\mathcal{O}\left(\min\left\{\sqrt{n}L, n^2 \log L + n^{3.5}(L_A + \log n)\right\}\right)$$

*iterations.*

Corollary 5.1 assumes that the initial iterate of the MTY P-C algorithm is a well-centered strictly feasible point whose duality gap is not too large. For a general dual pair of linear programs, even if such a point exists, computing this point is as hard as solving the pair of LP problems. In such a case, an auxiliary pair of dual LP problems is constructed whose optimal solution yields the one of the original pair of LP problems. Using the big M idea, Vavasis and Ye [29] constructs an auxiliary pair of LP problems, which we refer to as the *VY auxiliary LP pair*, associated with the pair of problems (1) and (2) in order to resolve the initialization issue for their least-layered-step algorithm. The VY auxiliary LP pair has the following properties: i) the input size of its coefficient matrix is bounded by  $\mathcal{O}(L_A)$ ; ii) its cost and right hand coefficients are bounded by  $\mathcal{O}(L)$ ; iii) it admits a readily available well-centered initial point whose duality gap is  $n2^{\mathcal{O}(L)}$ , and; iv) if (1) and (2) have a primal-dual optimal solution, then this solution can be easily obtained from an optimal solution of the VY auxiliary LP pair. Therefore, we obtain the following theorem for solving a general pair of LP problems under the Turing machine model.

**Theorem 5.2** *Assume that the data  $(A, b, c)$  is integral, and suppose that (1) and (2) have a primal-dual optimal solution. Then, the MTY P-C algorithm, with the FT Procedure invoked at every iteration, applied to the VY auxiliary LP pair, finds a strict complementary primal-dual optimal solution of (1) and (2) in  $\mathcal{O}(n^2 \log L + n^{3.5}(L_A + \log n))$  iterations.*

## 6 Concluding Remarks

In this paper, we have developed a new iteration-complexity bound for the MTY P-C algorithm using the notion of crossover events due to Vavasis and Ye [29]. In contrast to the iteration-complexity bound developed in [29], ours is scaling-invariant and has an extra but relatively small term, namely  $n^2 \log(\log(\mu_0/\eta))$ .

Note that the second bound in (14), i.e., the inequality  $\mu(w^+)/\mu(w) \leq \varepsilon_\infty^a \sqrt{n}/\beta$ , plays an important role in our analysis. This is also the inequality which plays an important role in establishing that the sequence  $\{\mu(w^k)\}$  generated by the MTY P-C algorithm is quadratically convergent. Observe also that if the sequence  $\{\mu(w^k)\}$  generated by a fictitious algorithm satisfied  $\mu(w^{k+1}) \leq C_1 \mu(w^k)^2$

for all  $k$ , where  $C_1 = \mathcal{O}(1)$  then its iteration-complexity bound would be  $\mathcal{O}(\log(\log(\mu_0/\eta)))$ . The term  $n^2 \log(\log(\mu_0/\eta))$  which appears in our iteration-complexity bound is due to the fact that this type of quadratic convergence happens over  $\mathcal{O}(n^2)$  disjoint finite sets of consecutive iteration indices. A natural conjecture is whether any primal-dual interior-point algorithm which achieves the duality gap reduction given by (14) at every iteration has the same iteration-complexity bound as the one obtained in this paper. More generally, we conjecture whether all interior-point algorithms whose corresponding sequence  $\{\mu(w^k)\}$  converges superlinearly or quadratically is suitable for the same type of analysis performed in this paper.

Note that the iteration-complexity bound obtained in this paper under the real-computation model is with respect to a pair of dual LP problems satisfying Assumptions A.1 and A.2. For a pair of dual LP problems which does not satisfy A.1 or A.2, a natural open question is whether one can develop the same type of scaling-invariant iteration-complexity bound under the real-computation model obtained in this paper. Observe that one of the difficulties in this context is the proper choice of the big M constant in the VY auxiliary LP pair.

Our work was strongly motivated by the work of Vavasis and Ye [29]. Therefore, it is also natural to conjecture whether the MTY P-C algorithm with the FT Procedure has the iteration-complexity bound  $\mathcal{O}(n^{3.5} \log(\bar{\chi}_A^* + n))$ , i.e. the iteration-complexity bound of Vavasis–Ye’s algorithm but with  $\bar{\chi}_A$  replaced by  $\bar{\chi}_A^*$ . The term involving the two log operators in our iteration-complexity bound is due to the possibility of the occurrence of case P2) (see the paragraph preceding Lemma 4.6). When case P2) occurs we have shown that the gap reduces quadratically until some suitable conditions are met which guarantee the occurrence of a crossover event in at most  $\mathcal{O}(\log(\log(\mu_0/\eta)))$  iterations. A very challenging open question is to show that these conditions are met in  $\mathcal{O}(p(n, \bar{\chi}_A))$  iterations, where  $p(x_1, x_2)$  is a polynomial in  $\mathfrak{R}^2$ . Note that this would provide a scaling-invariant iteration-complexity bound for the MTY P-C algorithm depending on  $n$  and  $\bar{\chi}_A^*$  only.

## Acknowledgment

This work was done while the first author was visiting Professor Cornelis Roos at the Information Tech & Systems of Delft University of Technology during the summer of 2001 and the second author at the Institute of Statistical Mathematics in August of 2002, and while the second author was visiting the first author at the School of Industrial and Systems Engineering in June and September of 2002. They are grateful to their host institutions for the financial support and the congenial scientific atmosphere that they have provided during these visits.

## References

- [1] I. I. Dikin and V. I. Zorkalcev, *Iterative Solution of Mathematical Programming Problems: Algorithms for the Method of Interior Points* (in Russian), Nauka, Novosibirsk, USSR, 1980.
- [2] C. C. Gonzaga: Path-following methods for linear programming. *SIAM Review*, Vol. 34, pp. 167–224.
- [3] 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.

- [4] O. Güler and Y. Ye: Convergence behavior of interior-point algorithms. *Mathematical Programming*. Vol. 60 (1993), pp. 215–228.
- [5] N. Karmarkar: A new polynomial-time algorithm for linear programming. *Combinatorica* Vol. 4 (1984), No. 4, pp.373–395.
- [6] M. Kojima, S. Mizuno and A. Yoshise: A primal-dual interior point algorithm for linear programming. In *Progress in Mathematical Programming, Interior-point and Related Methods* (ed. N. Megiddo), Springer-Verlag, New York, 1989, pp. 29–47.
- [7] M. Kojima, S. Mizuno and A. Yoshise: A polynomial-time algorithm for a class of linear complementarity problems. *Mathematical Programming*, Vol. 44 (1989), pp. 1–26.
- [8] L. McLinden: The complementarity problem for maximal monotone multifunction. In *Variational Inequalities and Complementarity Problems* (ed. R.W. Cottle, F. Giannessi, and J.L. Lions), John Wiley, New York, 1980, pp. 251-270.
- [9] N. Megiddo: Pathways to the optimal set in linear programming. *Progress in mathematical programming* (Pacific Grove, CA, 1987), pp. 131–158, Springer, New York-Berlin, 1989.
- [10] 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.
- [11] S. Mehrotra and Y. Ye: Finding an interior point in the optimal face of linear programs. *Mathematical Programming*, Vol. 62 (1993), No. 3, pp. 497–515.
- [12] 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.
- [13] S. Mizuno, N. Megiddo, and T. Tsuchiya: A linear programming instance with many crossover events. *Journal of Complexity* Vol. 12 (1996) pp. 474–479.
- [14] R. D. C. Monteiro and I. Adler: Interior path following primal-dual algorithms. Part I: Linear programming. *Mathematical Programming*, Vol. 44 (1989), pp. 27–41.
- [15] R. D. C. Monteiro and I. Adler. Interior path-following primal-dual algorithms. Part II: Convex quadratic programming. *Mathematical Programming*, Vol. 44 (1989), pp. 43–66.
- [16] R. D. C. Monteiro and I. Adler and Mauricio 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*, Vol. 15 (1990), pp. 191-214.
- [17] R. D. C. Monteiro and T. Tsuchiya: Global convergence of the affine scaling algorithm for convex quadratic programming. *SIAM Journal on Optimization*, Vol. 8 (1998), No. 1, pp. 26–58.
- [18] R. D. C. Monteiro and T. Tsuchiya. A variant of the Vavasis-Ye layered-step interior-point algorithm for linear programming. Manuscript. April, 2001. (To appear in Siam Journal on Optimization.)
- [19] J. Renegar: A polynomial-time algorithm, based on Newton’s method, for linear programming. *Mathematical Programming*, Vol. 40 (1988), No. 1, pp. 59–93.

- [20] G. W. Stewart: On Scaled projections and pseudoinverses, *Linear Algebra and its Applications*, Vol. 112 (1989) pp. 189–193.
- [21] K. Tanabe: Centered Newton method for mathematical programming. *System Modeling and Optimization* (M. Iri and K. Yajima (eds.)), pp. 197-206, Springer-Verlag, Berlin, 1988.
- [22] É. Tardos: A strongly polynomial algorithm to solve combinatorial linear programs. *Operations Research* Vol. 34, (1986) pp. 250–256.
- [23] M. J. Todd: A Dantzig-Wolfe-like variant of Karmarkar’s interior-point linear programming algorithm. *Operations Research*, Vol. 38 (1990) pp. 1006–1018.
- [24] 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.
- [25] 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.
- [26] L. Tunçel: Approximating the complexity measure of Vavasis-Ye algorithm is NP-hard. *Mathematical Programming*, Vol. 86 (1999), No. 1, pp. 219–223.
- [27] L. Tunçel: On the condition numbers for polyhedra in Karmarkar’s form. *Operations Research Letters*, Vol. 24 (1999), No. 4, pp. 149–155.
- [28] R. J. Vanderbei and J. C. Lagarias: I. I. Dikin’s convergence result for the affine-scaling algorithm, Vol. 17, *Contemporary Mathematics* Vol. 114 (1990) pp. 109–119.
- [29] S. Vavasis and Y. Ye: A primal-dual accelerated interior point method whose running time depends only on  $A$ . *Mathematical Programming*, Vol. 74 (1996), pp. 79–120.
- [30] S. Vavasis and Y. Ye: A simplification to “a primal-dual interior point method whose running time depends only on the constraint matrix. High performance optimization, 233–243, Appl. Optim., 33, Kluwer Acad. Publ., Dordrecht, 2000.
- [31] S. Vavasis and Y. Ye: On the relationship between layered least squares and affine scaling steps. In *The Mathematics of Numerical Analysis, Lectures in Applied Mathematics, vol. 32* American Mathematical Society, Providence, RI, 1996, pp. 857–865.
- [32] Y. Ye: On the finite convergence of interior-point algorithms for linear programming. *Mathematical Programming*, Vol. 57 (1992), No. 2, pp. 325–335.
- [33] Y. Ye: *Interior-point Algorithms: Theory and Analysis*. John Wiley and Sons, New York, USA, 1997.