# ON THE LIMITING PROPERTIES OF THE AFFINE-SCALING DIRECTIONS

Hugo J. Lara *    Clovis C. Gonzaga†    Levent Tunçel‡

August 5, 2003; revised: June 2004

**Abstract**

We study the limiting properties of the affine-scaling directions for linear programming problems. The worst-case angle between the affine-scaling directions and the objective function vector provides an interesting measure that has been very helpful in convergence analyses and in understanding the behaviour of various interior-point algorithms. We establish new relations between this measure and some other complexity measures which are used in the complexity analyses of algorithms for linear programming. We also provide a new characterization of the smallest large variable complexity measure of Ye.

**Keywords:** Affine-scaling direction, linear programming, complexity measures, interior-point methods

**AMS Subject Classification:** 90C05, 90C60, 49K40, 90C31, 90C51

*(hugol@ucla.edu.ve) Decanato de Ciencias. Universidad CentroOccidental Lisandro Alvarado (UCLA). Apdo 400. Barquisimeto, 3001, Venezuela. Research supported in part by CDCHT-UCLA, Venezuela; also supported in part by a Discovery Grant from NSERC and a PREA of the third author.

†Department of Mathematics, Federal University of Santa Catarina, Florianopolis, SC, Brazil, e-mail clovis@mtm.ufsc.br. Research partially supported by CAPES, Brazil.

‡(ltuncel@math.uwaterloo.ca) Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. Research of this author was supported in part by a Discovery Grant from NSERC and a PREA from Ontario, Canada.

# 1 Introduction

Consider the primal and dual linear programming (LP) problems

$$\begin{array}{llrcl} (P) & \min & c^T x & & \\ & & Ax & = & b, \\ & & x & \geq & 0, \\ (D) & \max & b^T y & & \\ & & A^T y & \leq & c, \end{array}$$

where $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ has full row rank (this will be assumed throughout the paper).

On many occasions, it will be more convenient for us to consider the equivalent formulations which are phrased in terms of subspaces: $\mathcal{N}(A)$ (the null space of $A$) and its orthogonal complement $\mathcal{R}(A^T)$ (the range space of $A^T$) in $\mathbb{R}^n$ as follows.

$$\begin{array}{llrcl} (P) & \min & \bar{s}^T x & & \\ & & x & \in & (S + \bar{x}), \\ & & x & \geq & 0, \\ (D) & \min & \bar{x}^T s & & \\ & & s & \in & \left(S^\perp + \bar{s}\right), \\ & & s & \geq & 0, \end{array}$$

where we can take $S := \mathcal{N}(A)$ (thus, $S^\perp = \mathcal{R}(A^T)$), $\bar{x} \in \mathbb{R}^n$ such that $A\bar{x} = b$ and $\bar{s} \in \mathbb{R}^n$ such that $(\bar{s} - c) \in \mathcal{R}(A^T)$. Then the new pair is equivalent to the previous one in that the corresponding optimal solution sets (in the appropriate space for $(D)$) are identical and the objective values can be related by trivial transformations.

The subspace transformation reveals other equivalence classes in the data space for LP problems in this standard form. Two full row rank matrices $A$, $\bar{A}$ of the same dimension such that $\mathcal{N}(A) = \mathcal{N}(\bar{A})$ get mapped to the same subspace pair $S, S^\perp$ in the subspace formulation. Moreover, many $\bar{x}$ vectors correspond to the same $b$ (namely, $\{x \in \mathbb{R}^n : Ax = b\}$) and many $c$ vectors can be reduced to the same $\bar{s}$ (namely, $\left\{c \in \mathbb{R}^n : (\bar{s} - c) \in \mathcal{R}\left(A^T\right)\right\}$).

If $c \in \mathcal{R}\left(A^T\right)$ then we can take $\bar{s} := 0$ which proves that in $(P)$, every feasible solution is optimal (the objective function is constant over the feasible region). Similarly, if $\bar{x} \in \mathcal{N}(A)$ (i.e., $b = 0$) then every feasible solution of $(D)$ is optimal. We will exclude such special cases. Even the least sophisticated algorithms will do the right thing in such cases. For instance, *the affine-scaling algorithm* (whose search direction is one of the main objects of study here) always takes a step along the *scaled* (with respect to the metric defined by the current interior-point iterate) steepest-descent direction. This direction is simply a scaled projection of scaled $c$ (in case of $(D)$, $\bar{x}$) onto a linear subspace and it will always result in the zero vector for $(P)$, if $c \in \mathcal{R}\left(A^T\right)$ (and similary, it will always result in the zero vector for $(D)$ if $b = 0$—i.e., $\bar{x} \in \mathcal{N}(A)$).

Let $P_A$ denote the orthogonal projection onto $\mathcal{N}(A)$. For $x \in \mathbb{R}^n$, $X$ denotes the $n \times n$ diagonal matrix whose $ii^{\text{th}}$ entry is $x_i$. For convenience, suppose that we are given $x^0 \in \mathbb{R}^n_{++} := \{x \in \mathbb{R}^n : x > 0\}$ such that $Ax^0 = b$ and that the optimal objective value is known to be $v$. We want to get within $\epsilon$ (given) of $v$. $e$ denotes the vector of all ones (of appropriate size, determined by the context). Then, the affine-scaling algorithm can be described as follows:

$k := 0;$
WHILE $c^T x^k - v \geq \epsilon$ DO
    $\bar{A} := AX_k, \ \bar{c} := X_k c, \ \bar{d} := -P_{\bar{A}}\bar{c};$
    choose $\alpha > 0$ such that $(e + \alpha\bar{d}) > 0;$
    $x^{k+1} := X_k(e + \alpha\bar{d});$
    $k := k + 1;$
END{WHILE};

Following Karmarkar's algorithm [11], the affine-scaling algorithm was proposed as a simplification of it. These proposals came from Barnes [2] and Vanderbei, Meketon and Friedman [24], independently. It turned out, however, that Dikin [5] proposed it almost 20 years prior to its rediscovery. Current best convergence results (without any assumptions on the nondegeneracy of $(P)$ or $(D)$) are due to Tsuchiya and Muramatsu [23]. Let $\alpha_{\max}$ denote the maximum step size for which the next iterate is feasible. They prove that taking the step size as $\frac{2}{3}\alpha_{\max}$ guarantees the convergence of the algorithm. (This result is somewhat tight, in that for larger step sizes the dual iterates need not converge —see Hall and Vanderbei [9].) However, currently it is not known whether there exists a scheme for choosing the step size such that the affine-scaling algorithm becomes a polynomial time algorithm for LP. On a related issue, Megiddo and Shub [15] proved that the affine-scaling trajectories can get arbitrarity close to the combinatorial paths on the boundary of the Klee-Minty cube which is used to show that various variants of the *simplex method* are exponential time algorithms. Affine-scaling trajectories were also studied by Adler and Monteiro [1].

Affine-scaling algorithm is the simplest and one of the most fundamental of the interior-point algorithms. Deep understanding of the behaviour of the affine-scaling algorithm usually has important consequences for many other, more sophisticated interior-point algorithms.

The notion of *local metric* which is profoundly important in interior-point methods, is the key in the affine-scaling algorithm. The affine-scaling search direction can be interpreted as the displacement between the current point $\bar{x}$ and the minimizer of $c^T x$ over the following ellipsoid centered at $\bar{x}$:

$$\left\{x \in \mathbb{R}^n : (x-\bar{x})^T \bar{X}^{-2}(x-\bar{x}) \leq 1\right\}.$$

This ellipsoid (called *the Dikin Ellipsoid*) is completely contained in the feasible region when restricted to the affine subspace $\{x \in \mathbb{R}^n : Ax = b\}$. Note that the shape and the size of the ellipsoid are defined by the positive definite matrix $\bar{X}^{-2}$ which changes as the current point changes (hence the term *local metric*).

Let $X_k \bar{d}$ denote the affine-scaling direction at the point $x^k$, as in the statement of the algorithm. Then

$$x^{k+1} := x^k + \alpha X_k \bar{d}$$

and

$$\begin{aligned} c^T x^{k+1} &= c^T x^k + \alpha c^T X_k \bar{d} \\ &= c^T x^k - \alpha \bar{c}^T P_{\bar{A}} \bar{c} \\ &= c^T x^k - \alpha \left(\bar{c}^T P_{\bar{A}}\right)\left(P_{\bar{A}}^T \bar{c}\right) \\ &= c^T x^k - \alpha \|P_{\bar{A}} \bar{c}\|^2 \end{aligned}$$

(where we used the facts: $P_{\bar{A}}^T = P_{\bar{A}}$ and $P_{\bar{A}}^2 = P_{\bar{A}}$). We see that the objective value of $(P)$ strictly improves from one iteration to the next. Moreover, if $P_{\bar{A}}\bar{c} = 0$ (same as $c \in \mathcal{R}(A^T)$) then every feasible solution of $(P)$, including $x^k$ is optimal.

Usually, the worst-case behaviour of interior-point algorithms for LP is measured in terms of the larger of the dimensions (that is $n$) of the problem and the desired accuracy $\epsilon \in (0,1)$ of the final solution (relative to the starting point). Moreover, we are usually content with just focusing on the bounds on the number of iterations required. The current best polynomial bound of this sort is $O(\sqrt{n}\ln(1/\epsilon))$. However, if we study the algorithms more deeply, we can come up with complexity analyses which depend on the data $(A, b, c)$ in more sophisticated ways (using complexity measures other than $n$ and $\ln(1/\epsilon)$) and expose more specific information about the performance of the underlying algorithms. Once such knowledge is exposed, then we can go back to the development of the algorithms and try to improve them in a way that the overall complexity bounds as well as the practical performance of the algorithms get better.

In the current paper, our focus is on the affine-scaling direction. Almost all search directions in interior-point methods can be expressed as a linear combination of the affine-scaling and *centering* directions. Since the affine-scaling direction is the component responsible for the improvement in the objective function value, its analysis is very important for most interior-point algorithms.

In the next section, we discuss oblique projections and the complexity measure $\bar{\chi}(A)$. Section 3 introduces some of the existing results about the affine-scaling directions. In Section 4, using geometric concepts (together with the algebraic concept of minimal linear dependence), we characterize the worst-case angle between the affine-scaling directions and the objective function vector (see Theorem 4.2). In Section 5, we represent a new characterization of Ye's complexity measure (this extends an earlier partial characterization by Lara and Gonzaga [12]); see Theorem 5.1. Using our geometric characterization of the worst-case angle from Section 4, we provide lower bounds on this angle in terms of certain complexity measures of the data, including $\bar{\chi}(A)$ (see Theorem 5.2). We conclude with a strengthening of this last result in the interesting case that we are near the optimal face (see Theorem 5.3).

## 2 Oblique projections, pseudo-inverses and $\bar{\chi}(A)$

The complexity measure $\bar{\chi}(A)$ can be defined as the suprema of the norms of all oblique projection operators onto $\mathcal{N}(A)$. This measure has been studied by Stewart [19], O'Leary [18], Todd [20], Vavasis and Ye [27] and others. $\bar{\chi}(A)$ has been used to study the computational complexity of some interior-point path-following algorithms for linear programming. Most notably, see Vavasis and Ye [26] and the recent and very nice analysis of Monterio and Tsuchiya [17] (who used a scale-invariant version of $\bar{\chi}(A)$). $\bar{\chi}(A)$ was also used in the analysis of a generalization of Tardos' scheme (see Ho and Tunçel [10]); but recently a better complexity measure replaced $\bar{\chi}(A)$, see, Lara and Tunçel [14].

Let $v^0 \in \mathbb{R}^n$ and a vector $d$ in the positive orthant $\mathbb{R}^n_{++}$ be given. We define the oblique projection of $v^0 \in \mathbb{R}^n$ onto $\mathcal{R}(A^T)$ with respect to $d \in \mathbb{R}^n_{++}$ (and denote it by $\tilde{Q}_{A,d}v^0$) as the unique solution of the problem

$$\text{minimize } \{\|D(v - v^0)\| \ : \ v = A^T w \text{ for some } w \in \mathbb{R}^m\},$$

where $D := \text{Diag}(d)$ is the diagonal matrix whose diagonal elements are the entries of $d$. Similarly, the oblique projection $Q_{A,d}v^0$ of $v^0$ onto $\mathcal{N}(A)$, the null space of $A$, with respect to $d \in \mathbb{R}^n_{++}$ is the unique minimizer of

$$\text{minimize } \{\|D^{-1}(u - v^0)\| \ : \ Au = 0\}.$$

Under the assumption that $A$ is of full row rank, it is easy to show that

$$Q_{A,d}v^0 = \left[I - D^2A^T(AD^2A^T)^{-1}A\right]v^0 \text{ and } \tilde{Q}_{A,d}v^0 = A^T(AD^2A^T)^{-1}AD^2v^0.$$

The matrix $A_d^+ := (AD^2A^T)^{-1}AD^2$ is frequently named as *the weighted pseudo-inverse of* $A$ (because $A_d^+A^T = I$). Even though this formula does not hold without the assumption of full row rank, $A_d^+$ always exists. In general, we can write

$$\tilde{Q}_{A,d}v^0 = A^T(A_d^+)v^0 \text{ and } Q_{A,d}v^0 = \left[I - \left(A_d^+\right)^T A\right]v^0.$$

Consider now the set $\mathcal{B}(A)$ of column indices associated with maximally linearly independent columns of $A$. If $J \in \mathcal{B}(A)$ then the columns of $A_J$ are linearly independent and if we add any other column of $A$ to $A_J$, then the linear independence is broken. Under the assumption of full row rank, the elements $J$ of $\mathcal{B}(A)$ define $m \times m$ nonsingular submatrices of $A$. The following results are devoted to calculating the weighted pseudo-inverses: The oblique projection $\bar{v}$ of the origin onto the set $\{v \ : \ v = c - A^Ty, y \in \mathbb{R}^m\}$, is the unique solution of the problem $\min\{\|Dv\| \ : \ v = c - A^Ty, y \in \mathbb{R}^m\}$. Using the optimality conditions, we can write $\bar{v} = c - A^T\bar{y}$, where $\bar{y} = A_d^+c$. Similarly, the oblique projection $\bar{w}$ of the origin onto the affine subspace defined by $Ax = b$ is the unique minimizer of the problem $\min\{\|D^{-1}w\| \ : \ Aw = b\}$. By the optimality conditions, $\bar{w} = (A_d^+)^Tb$. The following result, due to Dikin [4], calculates $\bar{y}$ and $\bar{w}$:

**Lemma 2.1** *(Dikin [4]) Let* $A \in \mathbb{R}^{m\times n}$*,* $c \in \mathbb{R}^n$*,* $b \in \mathbb{R}^m$*, and* $D := \mathrm{Diag}(d)$*, for some* $d \in \mathbb{R}^n_{++}$*. Then for* $\lambda_J(d) := \frac{\det(D_J^2)\det(A_J)^2}{\sum_{K\in\mathcal{B}(A)}\det(D_K^2)\det(A_K)^2}$*,*

*(a)*

$$A_d^+c = \sum_{J\in\mathcal{B}(A)} \lambda_J(d)\,(A_J)^{-T}\,c_J,$$

*(b)*

$$(A_d^+)^Tb = \sum_{J\in\mathcal{B}(A)} \lambda_J(d)w^{(J,b)},$$

*where* $w_J^{(J,b)} := (A_J)^{-1}\,b$ *and the remaining entries of* $w^{(J,b)}$ *are zero.*

**Proof.** A proof for part (a) can be found in Ben-Tal and Teboulle [3]. For part (b), note that for every $c \in \mathbb{R}^n$, we have

$$\begin{aligned} c^T(A_d^+)^Tb &= \textstyle\sum_{J\in\mathcal{B}(A)} \lambda_J(d)c_J^T\,(A_J)^{-1}\,b \\ &= \textstyle\sum_{J\in\mathcal{B}(A)} \lambda_J(d)c^Tw^{(J,b)} \\ &= c^T\textstyle\sum_{J\in\mathcal{B}(A)} \lambda_J(d)w^{(J,b)}. \end{aligned}$$

Since $(A_d^+)^T b$ is unique, and the above identities hold for every $c \in \mathbb{R}^n$, we have our claim. □

Now, we specialize these ideas to fit our needs. Consider a partition of the column indices set $[B, N]$ and define for $\mathcal{B}(A_B)$ an extended $\bar{\mathcal{B}}(A_B)$ such that for each $J \in \mathcal{B}(A_B)$ we choose an extended $\bar{J} \supseteq J$ such that $\bar{J} \in \mathcal{B}(A)$. For some $\bar{b} \in \mathcal{R}(A_B)$ we can calculate basic solutions of $A_B w_B = \bar{b}$ by using the basic extensions given by the elements of $\bar{\mathcal{B}}(A_B)$. So, a basic solution $w^{(J,\bar{b})}$ can be calculated by $w_{\bar{J}}^{(J,\bar{b})} = (A_{\bar{J}})^{-1}\bar{b}$ (and setting the other entries of $w^{(J,\bar{b})}$ to zero).

**Theorem 2.1** *Let $A \in \mathbb{R}^{m\times n}$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and $D := \mathrm{Diag}(d)$, for some $d \in \mathbb{R}^n_{++}$. Also let $[B, N]$ be a partition of the index set. Then for*

$$\lambda_J(d_B) := \frac{\det(D_J^2)\det(A_J)^2}{\sum_{K\in\bar{\mathcal{B}}(A_B)}\det(D_K^2)\det(A_K)^2},$$

*we have*

$$(A_B)^+_{d_B}\, c = \sum_{J\in\bar{\mathcal{B}}(A_B)} \lambda_J(d_B)\,(A_J)^{-T}\, c_J \tag{1}$$

*and*

$$\left[(A_B)^+_{d_B}\right]^T b = \sum_{J\in\bar{\mathcal{B}}(A_B)} \lambda_J(d_B) w^{(J,\bar{b})}, \tag{2}$$

*where $w_J^{(J,\bar{b})} := (A_J)^{-1}\bar{b}$ (and the other entries of $w^{(J,\bar{b})}$ are set to zero).*

**Proof.** The solution given in (1) is a convex combination of the basic solutions for $A_B^T y \le c_B$. These basic solutions can be calculated by using the extended basis in $\bar{\mathcal{B}}(A_B)$, and the scalars $\lambda_J(d_B)$ in the same way as in the lemma above. The proof of equation (2) is analogous. □

We define the complexity measure $\bar{\chi}(A)$ as

$$\bar{\chi}(A) := \sup\{\|\tilde{Q}_{A,d}v\| \;:\; \|v\| = 1, d \in \mathbb{R}^n_{++}\}.$$

We list in the following lemma some of the properties of $\bar{\chi}(A)$. First, we define the sets $\mathcal{S} := \{s \in \mathcal{R}(A^T) \;:\; \|s\| = 1\}$ and $\mathcal{X} := \{x \in \mathcal{N}(AD) \;:\; d \in \mathbb{R}^n_{++}\}$.

**Lemma 2.2** *Consider the $m \times n$ matrix $A$, $b \in \mathcal{R}(A)$, the matrix $Z$ whose rows form a basis of $\mathcal{N}(A)$. Then:*

*(a)* $\frac{1}{\bar{\chi}(A)} = \rho(A) := \inf\{\|s - x\| \ : \ s \in \mathcal{S}, x \in \mathcal{X}\}$.

*(b)* $\|A_d^+ b\| \leq \bar{\chi}(A)\|A^+ b\|$, *where $A^+$ is the right-pseudo-inverse of $A$.*

*(c)* $\bar{\chi}(A) = \bar{\chi}(Z)$.

*(d)* $\bar{\chi}(A) = \max\{\|A^T (A_J)^{-T}\| \ : \ J \in \mathcal{B}(A)\}$, *where $\mathcal{B}(A)$ is the set of column indices associated with nonsingular $m \times m$ submatrices of $A$.*

**Proof.** Stewart [19] and O'Leary [18] demonstrated part $(a)$. A proof of part $(b)$ can be found in Vavasis and Ye [27], while Gonzaga and Lara [8] proved part $(c)$. Finally, for part $(d)$ we refer to Todd [22], Vavasis and Ye [26] and Todd, Tunçel and Ye [21]. □

# 3 The Affine-Scaling Direction

Consider $d \in \mathbb{R}^n_{++}$. The primal affine-scaling direction $u(d)$ is defined as the unique solution for the problem

$$\begin{array}{ll} \text{minimize} & -c^T u + \frac{1}{2}\|D^{-1}u\|^2 \\ \text{subject to} & Au = 0. \end{array} \tag{3}$$

Analogously, we define the dual affine-scaling direction in terms of the right hand side $b$ as follows: First we get some vector $\bar{x}$ satisfying $A\bar{x} = b$. Then the dual affine-scaling direction $v(d)$ is the unique solution for the problem

$$\begin{array}{ll} \text{minimize} & -\bar{x}^T v + \frac{1}{2}\|Dv\|^2 \\ \text{subject to} & v = A^T w. \end{array} \tag{4}$$

We want to study lower bounds for the cosine of the angle between $u(d)$ and $c$ for all $d \in \mathbb{R}^n_{++}$. Tseng and Luo [22] show that this infimum is positive. Tseng and Luo's demonstration is an indirect proof, so the infimum is not calculated in a constructive way. We shall study such limiting angles and their relationship with other known complexity measures. A proof of the following result can also be found in Monteiro and Tsuchiya [16].

**Lemma 3.1** *(Tseng and Luo [22])*
*Consider $A, b$ and $c$ as before, with $c \notin \mathcal{R}(A^T)$. Then there exist positive constants $\tau_{A,c}$ and $\tau_{A,b}$ such that*

$$\tau_{A,c} = \inf_{d \in \mathbb{R}^n_{++}} \left\{ \frac{c^T u(d)}{\|u(d)\|} \right\},$$
$$\tau_{A,b} = \inf_{d \in \mathbb{R}^n_{++}} \left\{ \frac{\bar{x}^T v(d)}{\|v(d)\|} \right\}.$$

Our aim is to find lower bounds on these infima in terms of some complexity measures of $(P)$ and $(D)$.

The direction $u(d)$ satisfies the optimality conditions for some $y \in \mathbb{R}^m$:

$$Au = 0, \tag{5}$$
$$u = D^2(c - A^T y). \tag{6}$$

Using these equations we obtain the following equivalent compact formulae

$$u(d) = D P_{AD} D c, \tag{7}$$
$$u(d) = D^2 \left[ c - A^T (AD^2A^T)^{-1} AD^2 c \right] \tag{8}$$
$$u(d) = D^2 \left[ c - A^T A_d^+ c \right] \text{ or} \tag{9}$$
$$u(d) = (I - (A_d^+)^T A) D^2 c \tag{10}$$

for the full row rank matrix $A$.

Consider an arbitrary set $I \subseteq \{1, 2, \ldots, n\}$. We will denote by $[u(d)]_I$ the restriction of $u(d)$ to the indicies in $I$; and $u(d_I)$ stands for the affine-scaling direction calculated with the restricted data $A_I, c_I, d_I$. Then $[u(d_I)]_I = u(d_I)$ satisfies for some $y$:

$$A_I u_I = 0, \tag{11}$$
$$u_I = D_I^2(c_I - A_I^T y). \tag{12}$$

By (10) we have
$$[u(d)]_I = (I - (A_{d_I}^+)^T A_I) D_I^2 c_I.$$

Consider an arbitrary partition $[B, N]$ of the column index set. Assume that we know $[u(d)]_N$, the part of $u(d)$ indexed by $N$. Then the remainder, $[u(d)]_B$, can be obtained as the unique solution of

$$\begin{array}{ll} \text{minimize} & -c_B^T u_B + \frac{1}{2} \| (D_B)^{-1} u_B \|^2 \\ & A_B u_B = -A_N [u(d)]_N . \end{array}$$

This solution satisfies the optimality conditions

$$-c_B + D_B^{-2} u_B + A_B^T \lambda = 0$$
$$A_B u_B = -A_N \left[u(d)\right]_N$$

which give us the form

$$\left[u(d)\right]_B = u(d_B) - \left[(A_B)^+_{d_B}\right]^T A_N \left[u(d)\right]_N .$$

Using Lemma 2.1, we can write

$$\left[u(d)\right]_B = u(d_B) + \sum_{J \in \bar{\mathcal{B}}(A_B)} \lambda_J(d_B) \left[w^{\left(J, -A_N[u(d)]_N\right)}\right]_B , \tag{13}$$

where

$$\lambda_J(d_B) := \frac{\det(D_J^2) \det(A_J)^2}{\sum_{K \in \bar{\mathcal{B}}(A_B)} \det(D_K^2) \det(A_K)^2},$$

$\left[w^{\left(J, -A_N[u(d)]_N\right)}\right]_J := -\left(A_J\right)^{-1} A_N \left[u(d)\right]_N$, and all other entries of $w^{\left(J, -A_N[u(d)]_N\right)}$ are set to zero. The vector $w^{(J, \tilde{b})} \in \mathbb{R}^n$ is defined based on the input $J \subseteq \{1, 2, \dots, n\}$ such that $|J| = m$, $A_J$ is nonsingular, and the vector $\tilde{b} \in \mathbb{R}^m$. We simply set

$$\begin{aligned} \left[w^{(J, \tilde{b})}\right]_J &:= \left[A_J\right]^{-1} \tilde{b} \text{ and} \\ \left[w^{(J, \tilde{b})}\right]_j &:= 0, \text{ for every } j \notin J. \end{aligned}$$

Equation (13) is equivalent to

$$\left[u(d)\right]_B = u(d_B) + \sum_{J \in \bar{\mathcal{B}}(A_B)} \sum_{i \in N} \lambda_J(d_B) \left[u(d)\right]_i \left[w^{(J, -A_i)}\right]_B \tag{14}$$

where $\left[w^{(J, -A_i)}\right]_J := -\left(A_J\right)^{-1} A_i$ and all the other entries of $w^{(J, -A_i)}$ are set to zero. Note that if $i \in N \cap J$ for some $J \in \bar{\mathcal{B}}(A_B)$ then $\left[w^{(J, -A_j)}\right]_i = 0$.

For $x \in \mathbb{R}^n$, let

$$\begin{aligned} J_-(x) &:= \{j \in \{1, 2, \dots, n\} \ : \ x_j < 0\}, \\ J_+(x) &:= \{j \in \{1, 2, \dots, n\} \ : \ x_j > 0\}, \\ J_0(x) &:= \{j \in \{1, 2, \dots, n\} \ : \ x_j = 0\}, \\ J(x) &:= J_-(x) \cup J_+(x). \end{aligned}$$

We want to study the map $u(d)$ when $d \to \bar{d} \geq 0$ ($\bar{d} \neq 0$). Given $\bar{d}$, consider the partition $[B, N]$ defined as $B := J_+(\bar{d})$ and $N := J_0(\bar{d})$ (note that $J_-(\bar{d}) = \emptyset$). The components of

$d$ indexed by $B$ are called "the large variables" and the other ones "the small variables". Megiddo and Shub [15] and Gonzaga and Tapia [7] studied the behavior of the large variables in the mapping $u(d)$ when $d \to \bar{d}$. We quote the result here, and give a simple proof which uses Lemma 2.1 and (14):

**Theorem 3.1** *(Megiddo and Shub [15])*
*Consider $u(d)$, $d \to \bar{d}$ and the partition $[B, N]$ as defined above. Then*

*(i)* $[u(d)]_B \to u(\bar{d}_B) = \bar{D}_B P_{A_B \bar{D}_B} \bar{D}_B c_B$.

*(ii)* $[u(d)]_N \to 0$*; moreover,* $(P_{AD} Dc)_N \to 0$.

**Proof.** From (9) we have

$$\begin{aligned} [u(d)]_N &= D_N^2 \left[ c_N - A_N^T \left( A_d^+ \right) c \right] \\ &= D_N \left[ D_N \left( c_N - A_N^T \left( A_d^+ \right) c \right) \right]. \end{aligned}$$

$y(d) := (A_d^+)c$ is bounded; because, by Lemma 2.1 it is a convex combination of the dual feasible solutions $(A_J)^{-T} c_J$. Furthermore, since $D_N \left[ c_N - A_N^T (A_d^+) c \right] = (P_{AD} Dc)_N$ and $D_N$ converges to zero, we have the second claim. To show $(i)$, note that by (14) we have

$$[u(d)]_B = u(d_B) + \sum_{J \in \bar{\mathcal{B}}(A_B)} \sum_{i \in N} \lambda_J(d_B) \, [u(d)]_i \left[ w^{(J, -A_i)} \right]_B$$

where $\left[ w^{(J,-A_i)} \right]_J = - (A_J)^{-1} A_i$ (and the other entries of $w^{(J,-A_i)}$ are set to zero). By part $(ii)$ $[u(d)] \to 0$ and since $\lambda_J(d_B)$ is bounded we have that the second part of the equation above tends to zero and we have our claim. □

# 4 Minimal linear dependencies

The number $\tau_{A,c}$ will be estimated in terms of the minimal linear dependencies among the columns of the matrix $A$ and the angles such minimal linear dependencies make with the cost vector $c$.

As in [13], we shall focus on those characterizations of complexity measures involving the sign pattern of vectors in certain orthogonal linear subspaces. For $x \in \mathbb{R}^n$, $\text{sign}(x) \in$

$\{-,0,+\}^n$ encodes the signs of the entries of $x$. Let $S \subseteq \mathbb{R}^n$ be a linear subspace. We denote by $\text{sign}(S) \subseteq \{-,0,+\}^n$ the set of sign vectors of the elements of $S$.

Note that if $A \in \mathbb{R}^{m\times n}$ such that $\mathcal{N}(A) = S$ then every nonzero vector in $S$ represents a linear dependence among the columns of $A$. Minimal linear dependencies play a particularly important role in our work.

We denote the set of sign patterns of those minimal elements in $S$ by $\underline{\text{sign}}(S)$. That is, $\underline{\text{sign}}(S) \subset \text{sign}(S)$ denotes those nonzero sign patterns in $\text{sign}(S)$ such that setting any number of $+'s$ and $-'s$ to zero (without changing the others) does not give another nonzero element of $\text{sign}(S)$. Then, $\bar{x} \in S\backslash\{0\}$ is minimal if for all $\hat{x} \in S\backslash\{0\}$ satisfying $J_-(\hat{x}) \subseteq J_-(\bar{x})$, $J_+(\hat{x}) \subseteq J_+(\bar{x})$, $J_0(\hat{x}) \supseteq J_0(\bar{x})$ we have $\text{sign}(\hat{x}) = \text{sign}(\bar{x})$. So, $\bar{x} \in S$ is minimal if and only if $\text{sign}(\bar{x}) \in \underline{\text{sign}}(S)$.

Denote by $W_P$ the set of minimal vectors $w$ in $\mathcal{N}(A)$ such that $\|w\| = 1$ and $c^T w \geq 0$. Also denote by $\hat{W}_P$ the subset of $W_P$ whose elements satisfy $c^T w > 0$. Analogously, we define the dual sets $W_D$ and $\hat{W}_D$ just with $\mathcal{R}(A^T)$ in the role of $\mathcal{N}(A)$, and some feasible $\bar{x}$ in the role of $c$. $W_P$ is the set of minimal vectors in $\mathcal{N}(A)$ which make an acute angle with the cost vector $c$. If $\dim(\mathcal{N}(A)) \geq 1$ then $W_P$ is nonempty.

For each $J \in \mathcal{B}(A)$ and $i \notin J$ we define $w^{(J,i)}$ by $\left[w^{(J,i)}\right]_J := -\left(A_J\right)^{-1} A_i$, $w_i^{(J,i)} := 1$ and zero elsewhere. It is easy to prove that $\text{sign}\left(w^{(J,i)}\right) \in \underline{\text{sign}}(\mathcal{N}(A))$. Thus, $w^{(J,i)}/\|w^{(J,i)}\| \in W_P$, if $c^T w^{(J,i)} \geq 0$.

Let $I$ be a subset of the index set $\{1,2,\ldots,n\}$. We denote by $(W_I)_P$ the set defined in the same way as $W_P$ but with the data instance given by $A_I, c_I$. We define $\left(\hat{W}_I\right)_P$ analogously. The following lemma presents some properties of the sets $W_P$, $W_D$, $\hat{W}_P$ and $\hat{W}_D$:

**Lemma 4.1** *Consider $A$, $b$, $c$, $W_P, \hat{W}_P, W_D$ and $\hat{W}_D$ as defined above ($b \neq 0$ and $c \notin \mathcal{R}\left(A^T\right)$). Assume $n \geq m+1 \geq 2$. Then the following statements hold:*

*(a) The elements of $W_P$ span $\mathcal{N}(A)$ and the elements of $W_D$ span $\mathcal{R}(A^T)$.*

*(b) $\hat{W}_P \neq \emptyset$ and $\hat{W}_D \neq \emptyset$.*

*(c) For each $w \in \hat{W}_P$ there exists $\{d^k\}$ in $\mathbb{R}^n_{++}$ such that $\frac{u(d^k)}{\|u(d^k)\|} \to w$.*

*d) Consider $I \subset \{1,2,\ldots,n\}$. If $t \in (W_I)_P$ then the vector $\hat{t} \in \mathbb{R}^n$ defined as $\hat{t}_I := t, \hat{t}_{I^c} := 0$ belongs to $W_P$. Similarly if $t \in (W_I)_D$ then the vector $\hat{t} \in \mathbb{R}^n$ defined as $\hat{t}_I := t, \hat{t}_{I^c} := 0$ belongs to $W_D$.*

**Proof.** For part $(a)$: since $A$ has full row rank and $m < n$, we can choose $m$ linearly independent columns of $A$ which we index by $J$. So $\dim(\mathcal{N}(A_J)) = 0$ and the columns of $A_J$ generate $\mathcal{R}(A)$. We denote by $J^c$ the remaining $(n-m)$ columns of $A$. Each column $A_j$ $(j \in J^c)$ can be written as $A_J y^j_J = A_j$ with $y^j_J \neq 0$. We define for each $j \in J^c$ the vector $w^j \in \mathbb{R}^n$ as $w^j_J = \beta_j y^j_J$, $w^j_j = \beta_j$ and $w^j_{J^c \setminus \{j\}} = 0$, where $\beta_j$ is chosen in such way that $\|w^j\| = 1$ and $c^T w^j \geq 0$. By the construction $w^j \in W_P$ for all $j \in J^c$, and since $\dim(\mathcal{N}(A)) = (n-m)$ with $|J^c| = (n-m)$ and the vectors $w^j, j \in J^c$ are linearly independent, we conclude that the set $\{w^1, \ldots, w^{n-m}\}$ generates $\mathcal{N}(A)$. The second statement of part $(a)$ is similar.

To prove part $(b)$, first note that $\dim(\mathcal{N}(A)) = m - n \geq 1$. So, $W_P \neq \emptyset$. Next, assume for a contradiction that $\hat{W}_P = \emptyset$. That is, for all $w \in W_P$, we have $c^T w = 0$. This means that $c \perp \mathcal{N}(A)$ (by part $(a)$, the vectors in $W_P$ span $\mathcal{N}(A)$). Thus, $c \in \mathcal{R}(A^T)$, which is a contradiction. The dual part is analogous.

Now, consider $w \in \hat{W}_P$. Then by definition, there exists a $J \in \mathcal{B}(A)$ which gives rise to $w$. Take the sequence $\{d^k\}$ in $\mathbb{R}^n_{++}$ defined as $d^k_J = e_J$ and $d^k_{J^c} = \lambda_k e_{J^c}$, where $\{\lambda_k\}$ is a sequence in $\mathbb{R}_+$ converging to zero. Thus, $d^k \to \bar{d}$ where $\bar{d}_J = e_J$ and $\bar{d}_{J^c} = 0$. By Theorem 3.1 the sequence $u(d^k)$ converges to $\bar{u}$, where $\bar{u}_J = P_{A_J} c_J$ and $\bar{u}_{J^c} = 0$. We claim that $\bar{u}_J \neq 0$. In fact, if $\bar{u}_J = 0$ then $c_J \in \mathcal{R}(A^T)$ (because $\bar{u}_J \in \mathcal{N}(A)$). This leads to $c^T w = c^T_J w_J = 0$ which contradicts $c^T w > 0$. This means that $\frac{u(d^k)}{\|u(d^k)\|} \to \frac{\bar{u}}{\|\bar{u}\|}$. Since $c^T \bar{u} > 0$, $\bar{u}_J \in \mathcal{N}(A_J)$, $\bar{u}_{J^c} = 0$ and $\dim[\mathcal{N}(A_J)] = 1$ we conclude that $\bar{u}/\|\bar{u}\| = w$. This proves part $(c)$.

To show $(d)$, consider $t \in (W_I)_P$, and $J = J_+(t) \cup J_-(t) \subset I$. Now, we construct $\hat{t} \in \mathbb{R}^n$ by $\hat{t}_I := t$ and $\hat{t}_{I^c} := 0$. $\hat{t}$ satisfies $\hat{t}_J \in \mathcal{R}(A_J)$, $\hat{t}_{J^c} = 0$, $\|\hat{t}\| = 1$ and $c^T \hat{t} \geq 0$. This means that $\hat{t} \in W_P$. The proof for $(W_D)$ is similar. □

The following geometric result is the main tool in this part. We denote the cone generated by a set $S \subset \mathbb{R}^n$ by $\text{cone}(S)$. We shall establish that all primal affine-scaling directions are positive combinations of the elements of $\hat{W}_P$:

**Theorem 4.1** *Consider $d \in \mathbb{R}^n_{++}$. Then $u(d) \in \text{cone}(\hat{W}_P)$, and $v(d) \in \text{cone}(\hat{W}_D)$.*

**Proof.** We shall prove the primal statement. The dual one is analogous. Take a fixed $d \in \mathbb{R}^n_{++}$. By (6), we have $u(d) = D^2\left[c - A^T y(d)\right]$, where $y(d) = A^+_d c$, and by Theorem 2.1,

$$y(d) = \sum_{J \in \mathcal{B}(A)} \frac{\det(D^2_J)\det(A_J)^2}{\sum_{K \in \mathcal{B}} \det(D^2_K)\det(A_K)^2} (A_J)^{-T} c_J.$$

So,

$$\begin{aligned} u(d) &= D^2 \left( c - A^T \sum_{J\in\mathcal{B}(A)} \frac{\det(D_J^2)\det(A_J)^2}{\sum_{K\in\mathcal{B}(A)}\det(D_K^2)\det(A_K)^2} \left(A_J\right)^{-T} c_J \right) \\ &= D^2 \sum_{J\in\mathcal{B}(A)} \frac{\det(D_J^2)\det(A_J)^2}{\sum_{K\in\mathcal{B}(A)}\det(D_K^2)\det(A_K)^2} \left( c - A^T \left(A_J\right)^{-T} c_J \right). \end{aligned}$$

Each component $[u(d)]_i$ can be written as

$$[u(d)]_i = \sum_{J\in\mathcal{B}(A)} \frac{d_i^2 \det(D_J^2)\det(A_J)^2}{\sum_{K\in\mathcal{B}(A)}\det(D_K^2)\det(A_K)^2} \left( c_i - A_i^T \left(A_J\right)^{-T} c_J \right).$$

The form given by the coefficients $d_i^2 \det(D_J^2) = \Pi_{j\in J\cup\{i\}} d_j^2$ allows us to regroup the sum defining $u(d)$ : Consider a fixed $J \in \mathcal{B}(A)$ and $i \in \{1, 2, \ldots, n\}$. If $i \in J$ then $c_i - A_i^T A_J^{-T} c_J = c_i - c_i = 0$. This means that in the sum above only combinations of different $J$ and $i$ where $i \notin J$ are allowed, so we can assume $i \notin J$. Now fixing $i$ and varying $J \in \mathcal{B}(A)$ we define index sets $\bar{J} := J \cup \{i\}$ and consider the coefficients $\Pi_{j\in\bar{J}} d_j^2$. Define by $\mathcal{J}$ the set of all index sets constructed in the way above but varying also $i \in \{1, 2, \ldots, n\}$. Note that the same $\bar{J} \in \mathcal{J}$ can be built by using different combinations of basis $J$ and components $i$. Now focusing on coefficients of the form $\frac{\det(D_{\bar{J}}^2)}{\sum_{K\in\mathcal{B}(A)}\det(D_K^2)\det(A_K)^2}$ we can express $u(d)$ as

$$u(d) = \sum_{\bar{J}\in\mathcal{J}} \frac{\det(D_{\bar{J}}^2)\mu_{\bar{J}}}{\sum_{J\in\mathcal{B}}\det(D_K^2)\det(A_K)^2} w^{\bar{J}}, \tag{15}$$

where $w_i^{\bar{J}} := \mu_{\bar{J}}^{-1} \det(A_{\bar{J}\setminus\{i\}})^2 \left[ c_i - A_i^T \left(A_{\bar{J}\setminus\{i\}}\right)^{-1} c_{\bar{J}\setminus\{i\}} \right]$ ($\mu_J$ such that $\|w^{\bar{J}}\| = 1$), if $i \in \bar{J}$ and $w_i^{\bar{J}} := 0$ elsewhere. We claim that the vectors $w^{\bar{J}} \in \hat{W}$, for all $\bar{J} \in \mathcal{J}$. To show this, we prove that for fixed $\bar{J}$: $w^{\bar{J}} \in \mathcal{N}(A)$, $\mathrm{sign}(w^{\bar{J}}) \in \underline{\mathrm{sign}}(\mathcal{N}(A))$ and that $c^T w^{\bar{J}} > 0$. In fact: Firstly consider a sequence $\{d^k\}$ in $\mathbb{R}^n_{++}$ defined by $d_i^k = 1$ if $i \in \bar{J}$ and $d_i^k := \lambda_k$ if $i \notin \bar{J}$. Now suppose $0 < \lambda_k \downarrow 0$. It is easy to see from (15) that $\lim_{k\to\infty} \frac{u(d^k)}{\|u(d^k)\|} = w^{\bar{J}}$. Since $u(d^k) \in \mathcal{N}(A)$ for all $d^k$, and $\mathcal{N}(A)$ is a closed set we conclude that $w^{\bar{J}} \in \mathcal{N}(A)$. Secondly, since $\bar{J} = J \cup \{i\}$ for some $J \in \mathcal{B}(A)$ and $i \notin J$ we conclude that $\mathrm{sign}(w^{\bar{J}}) \in \underline{\mathrm{sign}}(\mathcal{N}(A))$. Finally, since $c^T u(d) > 0$ for all $d \in \mathbb{R}^n_{++}$, we have $\lim_{k\to\infty} c^T u(d^k) = c^T w^{\bar{J}} \geq 0$. So far, we have shown that $u(d) \in \mathrm{cone}(W_P)$. It remains to prove that only those vectors $w^{\bar{J}}$, that satisfy $c^T w^{\bar{J}} > 0$, participate in the sum (15). Suppose for a contradiction that there exists $\hat{J} \in \mathcal{J}$ such that $c^T w^{\hat{J}} = 0$. Then $c_{\hat{J}} \in \mathcal{R}(A_{\hat{J}}^T)$. Since $w_{\hat{J}}^{\hat{J}} \in \mathcal{N}(A_{\hat{J}})$ and $\dim(\mathcal{N}(A_{\hat{J}})) = 1$ we conclude that

$$w_{\hat{J}}^{\hat{J}} = \gamma P_{A_{\hat{J}}} c_{\hat{J}}$$

for some $\gamma \in \mathbb{R}$. But $c_{\hat{J}} \in \mathcal{R}(A_{\hat{J}}^T)$ implies that $P_{A_{\hat{J}}} c_{\hat{J}} = 0$. This means that in the sum (15) we consider only vectors $w^{\bar{J}}$ which satisfy $c^T w^{\bar{J}} > 0$. Therefore, $w^{\bar{J}} \in \hat{W}_P$ for all $\bar{J} \in \mathcal{J}$. □

As a consequence of this result, we have the following theorem:

**Theorem 4.2** *With the above definitions, we have*

$$\tau_{A,c} = \min\{c^T w \ : \ w \in \hat{W}_P\}.$$

*Moreover, for every $\bar{x} \in \mathbb{R}^n$ such that $A\bar{x} = b$,*

$$\tau_{A,b} = \min\{\bar{x}^T v \ : \ v \in \hat{W}_D\}.$$

**Proof.** We shall prove part $(a)$. Part $(b)$ is analogous. Since $u(d)/\|u(d)\| \in \text{cone}(\hat{W}_P)$ for all $d \in \mathbb{R}^n_{++}$, we have

$$\tau_{A,c} = \inf_{d>0} \left\{ \frac{c^T u(d)}{\|u(d)\|} \right\} \geq \min\{c^T w \ : \ w \in \hat{W}_P\}.$$

Now, since each $w \in \hat{W}_P$ is limit of affine-scaling directions (by Lemma 4.1) we conclude that

$$c^T w \geq \inf_{d>0} \left\{ \frac{c^T u(d)}{\|u(d)\|} \right\} = \tau_{A,c}$$

for each $w \in \hat{W}_P$. □

# 5 The smallest large variable complexity measure

In this section, we first quote the complexity measure first studied by Ye [29] and meant as the smallest large variable on the optimal set. Consider the optimal partition $[B, N]$. The smallest large variable complexity measure $\sigma$ is defined as the minimum of

$$\sigma_{A,b} = \min_{j\in B}\{\max x_j \ : \ Ax = b, x_B \geq 0, x_N = 0\}$$

and

$$\sigma_{A,c} = \min_{i\in N}\{\max s_i \ : \ A_B^T y = c_B, A_N^T y + s_N = c_N, s_N \geq 0\}.$$

That is,

$$\sigma_{A,b,c} := \min\left\{\sigma_{A,b}, \sigma_{A,c}\right\}.$$

The main result in [29] is establishing that the sequences generated by many of the interior-point path-following algorithms can be terminated in $O(\sqrt{n}(|\log \sigma_{A,b,c}| + n))$ iterations.

The primal smallest large variable measure has been related to the symmetry measure of the primal feasible set in case of homogoneous systems in Karmarkar's form (see Epelman and Freund [6]). Various results relating this complexity measure to others are given in [10, 13, 21, 27].

## 5.1 A characterization of $\sigma_{A,b,c}$

In this subsection, we give a characterization of $\sigma_{A,b,c}$ in terms of the elements of $\hat{W}_P$ and $\hat{W}_D$ that are feasible directions from the optimal primal and dual faces.

Given the optimal partition $[B, N]$, a feasible direction from the primal optimal set is any direction $v \in \mathcal{N}(A)$ such that $v_N \geq 0$. We denote by $F_P$ the set of feasible directions from the primal optimal set. Similarly, we define $F_D$ as the set of dual feasible directions from the dual optimal set (vectors $v$ in $\mathcal{R}(A^T)$ such that $v_B \geq 0$). For any $x \in \mathbb{R}^n$, $x^+$ stands for the maximum nonnegative component of $x$. Let us denote by $\eta$ the minimum of

$$\eta_P := \min\{c^T w / w_N^+ \ : \ w \in \hat{W}_P \cap F_P\}$$

and

$$\eta_D := \min\{\bar{x}^T v / v_B^+ \ : \ v \in \hat{W}_D \cap F_D\}$$

where $\bar{x}$ satisfies $A\bar{x} = b$.

In [12], Lara and Gonzaga proved that $\sigma_{A,c} \leq \min\{c^T w \ : \ w \in \hat{W}_P \cap F_P\}$ and $\sigma_{A,b} \leq \min\{\bar{x}^T v \ : \ v \in \hat{W}_D \cap F_D\}$. Here, we prove a tight characterization.

First we state an auxiliary problem which defines the maximum value of a large variable on the optimal dual set. For fixed $j \in N$ we define $\sigma_{D_j}$ as the optimal value for

$$(P_j) \quad \begin{array}{ll} \text{Maximize} & e_j^T s_N \\ \text{Subject to} & A_B^T y = c_B \\ & A_N^T y + s_N = c_N \\ & s_N \geq 0, \end{array}$$

where $e_j$ denotes the $j-th$ column of the identity matrix $I$. The dual problem associated with $(P_j)$ is

$$(D_j) \quad \begin{array}{ll} \text{Minimize} & c^T w \\ \text{Subject to} & Aw = 0 \\ & w_N \geq 0 \\ & w_j \geq 1. \end{array}$$

The problem $(P_j)$ has a positive optimal value, because the components indexed by $N$ are positive in the relative interior of the dual optimal set. It follows from the duality theorem that $(D_j)$ also has a positive optimal value. Let $w$ be an optimal solution for $(D_j)$, then $w_j = 1$, because otherwise $w_j > 1$ and $\tilde{w} := w/w_j$ would also be feasible with $c^T\tilde{w} < c^T w$, contradicting the optimality of $w$.

**Lemma 5.1** *Consider a fixed $j \in N$, and $\sigma_{D_j}$ as defined above. Then there exists $w^{(j)} \in \hat{W}_P \cap F_P$ such that $\sigma_{D_j} = c^T w^{(j)}/w_j^{(j)}$.*

**Proof.** By definition, $\sigma_{D_j}$ is the optimal value of the problem $(P_j)$. Consider the dual problem $(D_j)$. Since $w_j = 1$, we can write $(D_j)$ as

$$
\begin{array}{rll}
 & \text{Minimize} & c^T w \\
(D_j) & \text{Subject to} & A_B w_B + A_{N_j} w_{N_j} = -A_j \\
 & & w_{N_j} \geq 0,
\end{array}
$$

where $N_j = N \backslash \{j\}$. Among the optimal solutions for $(D_j)$, let us choose an optimal solution $w$ of this problem such that the number of zero components of $w$ is maximum. So, $\text{sign}(w) \in \underline{\text{sign}}(\mathcal{N}(A))$. Since $w_N \geq 0$ we conclude that $w \in \hat{W}_P \cap F_P$. We have shown that $w \in \hat{W}_P \cap F_P$ with $w_j = 1$ and $\sigma_{D_j} = c^T w = c^T w / w_j$. □

The same result can be established for $\sigma_{P_i} := \max\{x_i \ : \ A_B x_B = b, x_B \geq 0\}$, that is $\sigma_{P_i} = \bar{x}^T v^{(i)}/v_i^{(i)}$ for some primal feasible $\bar{x}$, and some $i \in B$.

In the sequel, we state the main result of this subsection:

**Theorem 5.1** *With the above definitions, we have*

$$\sigma_{A,c} = \eta_P; \sigma_{A,b} = \eta_D \ \textit{and thus} \ \sigma_{A,b,c} = \eta.$$

**Proof.** Let $s^*$ be a dual optimal solution such that for some $l$, $s_l^* = \sigma_{A,c}$. We want to show that there exists $w \in \hat{W}_P \cap F_P$ such that $s_l^* = c^T w / w_N^+$. In fact, $s_l^*$ is the optimal value for $(P_l)$. By the lemma above there exists $w^{(l)} \in \hat{W}_P \cap F_P$ satisfying $\sigma_{A,c} = \sigma_{D_l} = s_l^* = c^T w^{(l)}/w_l^{(l)}$. We claim that $w_l^{(l)} = \max_{i \in N}\{w_i^{(l)}\} = (w_N^{(l)})^+$. To show that, suppose for a contradiction that there exists $i \in N\backslash\{l\}$ such that $w_i^{(l)} > w_l^{(l)}$. Then $\tilde{w} := w^{(l)}/w_i^{(l)}$ is a feasible solution for $(D_i)$. Furthermore, $0 < c^T\tilde{w} < c^T w^{(l)}$. By the duality theory of linear programming, $c^T\tilde{w}$ is an upper bound for the optimal value of $(P_i)$. Take $\bar{s}$ any optimal solution for $(P_i)$. Then we have $\bar{s}_i > 0$, because $i \in N$, and $\bar{s}_i = \max\{s_i \ : \ s_N = c_N - A_N^T y \geq 0, c_B = A_B^T y, y \in \mathbb{R}^m\}$. So, we have

$$\bar{s}_i \leq c^T \tilde{w} < c^T w^{(l)} = \sigma_{A,c},$$

which contradicts the minimality of $\sigma_{A,c}$. This means that $w_l^{(l)} = (w_N^{(l)})^+$, and therefore, $\sigma_{A,c} = c^T w^{(l)}/(w_N^{(l)})^+ \geq \eta_P$.

Now, consider $w^* \in \hat{W}_P \cap F_P$ such that $\eta_P = c^T w^*/(w_N^*)^+$, and take $l$ as the index in $N$ which defines the maximum $(w_N^*)^+$. $w^*/(w_N^*)^+$ is feasible for $(D_l)$. Take an optimal solution $\bar{s}$ of $(P_l)$. We obtain,

$$\eta_P = c^T w^*/(w_N^*)^+ \geq \bar{s}_l = \sigma_{D_l} \geq \sigma_{A,c}.$$

□

## 5.2 Bounds on $\tau$

Now, we give some bounds on $\tau_{A,b,c}$ in terms of some complexity measures. Note that $\sigma_{A,c}$ is attained at an extreme point of the optimal set (similarly $\sigma_{A,b}$), so there exists an optimal basis $A_J$ and an index $l$ such that $\sigma_{A,c} = c_l - c_J^T (A_J)^{-1} A_l$. In the sequel, we define quantities associated with different basic solutions:

$$\xi_{A,b} := \min\{x_i \ : \ x_i > 0, x = (A_J)^{-1} b \geq 0, J \in \mathcal{B}(A)\},$$

$$\hat{\xi}_{A,b} := \min\{x_i \ : \ x_i > 0, x = (A_J)^{-1} b, J \in \mathcal{B}(A)\},$$

and

$$\underline{\xi}_{A,b} := \min\{|x_i| \ : \ x_i \neq 0, x = (A_J)^{-1} b, J \in \mathcal{B}(A)\}.$$

The following relations among these quantities are straightforward to establish:

$$\xi_{A,b} \geq \hat{\xi}_{A,b} \geq \underline{\xi}_{A,b}.$$

It can be shown by examples where these inequalities are not tight.

We also define the analogous dual quantities:

$$\xi_{A,c} := \{s_i \ : \ s_i > 0, s = c - A^T (A_J)^{-T} c_J \geq 0, J \in \mathcal{B}(A)\},$$

$$\hat{\xi}_{A,c} := \{s_i \ : \ s_i > 0, s = c - A^T (A_J)^{-T} c_J, J \in \mathcal{B}(A)\},$$

and

$$\underline{\xi}_{A,c} := \{|s_i| \ : \ s_i \neq 0, s = c - A^T (A_J)^{-T} c_J, J \in \mathcal{B}(A)\}.$$

A version of this complexity measure called $\xi(A)$ was studied in [13] and further used in the complexity analyses [14]. Now, we establish the following result which links $\tau_{A,c}$ to $\xi_{A,c}$ and $\tau_{A,b}$ to $\xi_{A,b}$:

**Theorem 5.2** *Consider* $A, b, c, \hat{\xi}_{A,c}, \hat{\xi}_{A,b}, \tau_{A,c}$ *and* $\tau_{A,b}$ *as defined above. Then*

*(a)* $\tau_{A,c} \geq \frac{\hat{\xi}_{A,c}}{\sqrt{\bar{\chi}(A)^2+1}}$.

*(b)* $\tau_{A,b} \geq \frac{\hat{\xi}_{A,b}}{\sqrt{\bar{\chi}(A)^2+1}}$.

**Proof.** We shall prove $(a)$. The proof for $(b)$ is analogous. By Theorem 4.2 we have

$$\tau_{A,c} = \min\left\{ \frac{c^T w}{\|w\|} : w \in \hat{W}_P \right\}.$$

Consider $w^* \in \hat{W}_P$ such that the minimum is attained. For $w^*$ there is a $J^* \in \mathcal{B}(A)$ and $i^* \notin J^*$ such that $w^*_{J^*} = -\left(A_J\right)^{-1} A_{i^*}, w^*_{i^*} = 1$, and $w^*_j$ is zero elsewhere. So

$$\begin{aligned} \tau_{A,c} &= \frac{c^T w^*}{\|w^*\|} \\ &= \frac{c_{i^*} - c^T_{J^*}\left(A_{J^*}\right)^{-1} A_{i^*}}{\sqrt{\|(A_{J^*})^{-1} A_{i^*}\|^2+1}}. \end{aligned}$$

Since $w^* \in \hat{W}_P$ we conclude that the numerator is positive. By the definition of $\hat{\xi}_{A,c}$, we have $\hat{\xi}_{A,c} \leq c_{i^*} - c^T_{J^*}\left(A_{J^*}\right)^{-1} A_{i^*}$. On the other hand, since $\|\left(A_J\right)^{-1} A_i\| \leq \|\left(A_J\right)^{-1} A\| \leq \bar{\chi}(A)$ (by Lemma 2.2) then the denominator in the last relation is at most $\sqrt{\bar{\chi}(A)^2+1}$. □

Bounds on $\tau$, provided by Theorem 5.2, are independent of the optimal partition and therefore, can be too rough if we want to measure how small the angle between $c$ and the affine-scaling direction is, when approaching optimal solutions. We can obtain better bounds on $\tau$, if we focus on the affine-scaling directions near the optimal face.

Now, we apply the results of the above subsection to study the limits of the affine-scaling directions when approaching the optimal face. First, we consider the quantity $\min\{w^+_N \ : \ w \in \hat{W}_P \cap F_P\}$ and a sequence $\{d^k\}$ in $\mathbb{R}^n_{++}$ with a limit identifying the optimal face, that is, $d^k_B \to d_B > 0, d^k_N \to 0$.

**Theorem 5.3** *Consider* $A, b, c$ *as above, the optimal partition* $[B, N]$ *and the sequence* $\{d^k\}$ *satisfying* $d^k_B \to \bar{d}_B > 0$, $d^k_N \to 0$. *Then*

$$\lim_{k\to\infty} \left\{ \frac{c^T u(d^k)}{\|u(d^k)\|} \right\} \geq \frac{\sigma_{A,c} \min\{w^+_N \ : \ w \in \hat{W}_P \cap F_P\}}{\sqrt{\bar{\chi}(A)^2+1}}.$$

**Proof.** Consider $\{d^k\}$ as in the hypothesis. For a fixed $k$, we have

$$u(d^k) = D_k^2(c - A^T y^k),$$

where

$$y^k := y(d^k) = \sum_{J \in \mathcal{B}(A)} \lambda_J(d^k)\,(A_J)^{-1}\,c_J \tag{16}$$

and $\lambda_J(d^k) = \det({D_k}_J^2)\det(A_J)^2 / \sum_{K \in \mathcal{B}(A)} \det((D_k)_K^2)\det(A_K)^2$. Since $y^k$ is a convex combination of some dual basic solutions $(A_J)^{-1}\,c_J$ with $J \in \mathcal{B}(A)$, we conclude that $\{y^k\}$ is bounded and so we can assume $\{y^k\}$ converges to, say, $\bar{y} \in \mathbb{R}^m$. By using the optimal partition $[B, N]$ we can split $u(d^k)$ as $\left[u(d^k)\right]_B$ and $\left[u(d^k)\right]_N$. $c_B \in \mathcal{R}(A_B^T)$, because, in optimal solutions we have $s_B = c_B - A_B^T y = 0$ for some $y \in \mathbb{R}^m$. Since $c \notin \mathcal{R}(A)$ we conclude that $c_N \notin \mathcal{R}(A_N^T)$. This means that $c_N - A_N^T \bar{y} \neq 0$. Let us denote $\mu_k := \|\left[u(d^k)\right]_N\|$. Note that $\mu_k \to 0$ by Theorem 3.1. We know $\left[u(d^k)\right]_N = (D_K^2)_N(c_N - A_N^T y^k)$. We can rearrange $\left[u(d^k)\right]_N$ as $\mu_k t_N^k$ where $t_N^k := \frac{\left[u(d^k)\right]_N}{\|\left[u(d^k)\right]\|}$. Clearly $t_N^k$ is bounded and we can assume that it converges (to $\bar{t}_N \neq 0$).

Now, let us focus on $u(d^k)$. We know that affine-scaling directions when approaching the optimal set, are feasible directions from the optimal face (i.e., $\left[u(d^k)\right]_N > 0$). So $u(d^k) \in F_P$ and therefore $\left[u(d^k)\right]_N \geq 0$ for all $k \geq \bar{k}$, for some $\bar{k}$. This means that $t_N^k > 0$ and so $\bar{t}_N \geq 0$.

By (14),

$$\begin{aligned}\left[u(d^k)\right]_B &= u(d_B) - ({A_B}_{d_B}^{+})^T A_N \left[u(d^k)\right]_N \\ &= u(d_B) + \textstyle\sum_{J \in \bar{\mathcal{B}}(A_B)} \sum_{i \in N} \lambda_J(d_B^k)\left[u(d^k)\right]_i w^{(J,-A_i)}.\end{aligned}$$

with $(w^{(J,-A_i)})_J := -(A_J)^{-1}A_i$ and zero in the remainder of the components. Since $c_B \in \mathcal{R}(A_B^T)$ we have $u_B(d_B) = 0$. Merging the expression for $\left[u(d^k)\right]_i$ $(i \in N)$ we obtain

$$\left[u(d^k)\right]_B = \mu_k^2 \textstyle\sum_{J \in \bar{\mathcal{B}}(A_B)} \sum_{i \in N} \lambda_J(d^k) t_i^k w_B^{(J,-A_i)}.$$

Consequently, by putting $u_B(d^k)$ and $u_N(d^k)$ together we have

$$u(d^k) = \mu_k^2 \sum_{J \in \bar{\mathcal{J}}(A_B)} \sum_{i \in N} \lambda_J(d_B^k) t_i^k w^{(J,i)},$$

with $w_J^{J,i} := -(A_J)^{-1}A_i$, $w_i^{J,i} := 1$ and zero elsewhere. By the construction, $w^{(J,i)} \in \hat{W}_P$ for all $J$. Since $\left[u(d^k)\right]_N \geq 0$, each $w_N^{(J,i)} \geq 0$ and the coefficients $\mu_k$, $\lambda_J(d_B^k)$ and $\bar{t}_i^k$ are nonnegative, we conclude $w^{(J,i)} \in F_P$. Therefore, $w^{(J,i)} \in \hat{W}_P \cap F_P$. This shows that $u(d^k)$ is a conic combination of the vectors in $\hat{W}_P \cap F_P$. Thus,

$$\begin{aligned} \frac{c^T u(d^k)}{\|u(d^k)\|} &\to \frac{c^T u(\bar{d})}{\|u(\bar{d})\|} \\ &\geq \min_{w \in \hat{W}_P \cap F_P} \frac{c^T w}{\|w\|} \frac{w_N^+}{w_N^+} \\ &\geq \frac{\sigma_D \min_{w \in \hat{W}_P \cap F_P} w_N^+}{\sqrt{\bar{\chi}(A)^2 + 1}}. \end{aligned}$$

The last inequality follows from Theorem 5.1. □

An analogue of this type of analysis would be very interesting for the primal-dual affine-scaling direction as well. This is left for future work.

# Appendix

In many proofs, we established the result for the primal form of the LP problem and omitted the proof for the dual-form LP. In all of these cases, the same proof also works for the dual-form LP, if we use the equivalent, subspace representation of $(D)$:

$$\begin{aligned} (D) \quad \min \quad & \bar{x}^T s \\ s &\in \left(S^\perp + \bar{s}\right), \\ s &\geq 0, \end{aligned}$$

where $S = \mathcal{N}(A)$.

# References

[1] I. Adler and R. D. C. Monteiro, Limiting behavior of the affine scaling continuous trajectories for linear programming problems, *Math. Prog. A* 50 (1991) 29–51.

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

[3] A. Ben–Tal and M. Teboulle, A geometric property of the least squares solution of linear equations, *Linear Algebra and Its Applications* 139 (1990) 165–170.

[4] I. I. Dikin. On the speed of an iterative process, *Upravlyaemye Sistemi* 12 (1974) 54–60.

[5] I. I. Dikin, Iterative solution of problems of linear and quadratic programming, *(Russian) Dokl. Akad. Nauk SSSR* 174 (1967) 747–748.

[6] M. Epelman and R. M. Freund, A new condition measure, pre-conditioners, and relations between different measures of conditioning for conic linear systems, *SIAM J. Optim.* 12 (2002) 627–655.

[7] C. C. Gonzaga and R. A. Tapia. On the convergence of the Mizuno-Todd-Ye algorithm to the analytic center of the solution set, *SIAM J. Optim.* 7 (1997) 47–65.

[8] C. C. Gonzaga and H. Lara. A note on properties of condition numbers, *Linear Algebra and Its Applications* 261 (1997) 269–273.

[9] L. A. Hall and R. J. Vanderbei, Two-thirds is sharp for affine scaling, *Oper. Res. Lett.* 13 (1993) 197–201.

[10] J. C. K. Ho and L. Tunçel, Reconciliation of various complexity and condition measures for linear programming problems and a generalization of Tardos' theorem, In: *Foundations of Computational Mathematics, Proceedings of the Smalefest 2000*, F. Cucker and J. M. Rojas (eds.) World Scientific, 2002, pp. 93–147.

[11] N. Karmarkar, A new polynomial-time algorithm for linear programming, *Combinatorica* 4 (1984) 373–395.

[12] H. J. Lara and C. C. Gonzaga. A characterization of the smallest large variable condition number. *Investigación Operativa* Volume 9, N 1,2 and 3, (2000).

[13] H. J. Lara and L. Tunçel, Condition and complexity measures for infeasibility certificates of systems of linear inequalities and their sensitivity analysis, *Research Report* 2002–10, Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada, May 2002 (revised: July 2002).

[14] H. J. Lara and L. Tunçel, On the computational complexity of solving systems of linear inequalities, *in preparation.*

[15] N. Megiddo and M. Shub, Boundary behavior of interior point algorithms in linear programming, *Math. Oper. Res.* 14 (1989) 97–146.

[16] R. D. C. Monteiro and T. Tsuchiya, A simplified global convergence proof of the affine scaling algorithm, *Annals of Operations Research* 47 (1993) 443–482.

[17] R. D. C. Monteiro and T. Tsuchiya, A new iteration-complexity bound for the MTY predictor-corrector algorithm, *manuscript*, October 2002.

[18] D. P. O'Leary. On bounds for scaled projections and pseudoinverses, *Linear Algebra and Its Applications* 132 (1990) 115–117.

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

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

[21] M. J. Todd, L. Tunçel and Y. Ye, Probabilistic analysis of two complexity measures for linear programming problems, *Math. Prog. A* 90 (2001) 59–69.

[22] P. Tseng and Z.-Q. Luo. On the convergence of the affine-scaling algorithm, *Math. Prog. A* 56 (1992) 301–319.

[23] T. Tsuchiya and M. Muramatsu, Global convergence of a long-step affine scaling algorithm for degenerate linear programming problems, *SIAM J. Optim.* 5 (1995) 525–551.

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

[25] S. A. Vavasis, Stable numerical algorithms for equilibrium systems, *SIAM J. Matrix Anal. Appl.* 15 (1994) 1108–1131.

[26] S. A. Vavasis and Y. Ye, A primal-dual interior point method whose running time depends only on the constraint matrix, *Math. Prog. A* 74 (1996) 79–120.

[27] S. A. Vavasis and Y. Ye, Condition numbers for polyhedra with real number data, *Operations Research Letters* 17 (1995) 209–214.

[28] Y. Ye, Improved complexity results on solving real-number linear feasibility problems, *manuscript*, May 2002.

[29] Y. Ye. On the finite convergence of interior-point algorithms for linear programming, *Mathematical Prog. A* 57 (1992) 325–335.