

# Strict Feasibility and Degeneracy in Linear Programming

Jiyoung Im\*      Henry Wolkowicz†

March 5, 2022

## Abstract

Currently, the simplex method and the interior point method are indisputably the most popular algorithms for solving linear programs. Unlike general conic programs, linear programs with a finite optimal value do not require strict feasibility in order to establish strong duality. Hence strict feasibility is often less emphasized. In this note we discuss that the lack of strict feasibility necessarily causes difficulties in both simplex and interior point methods. In particular, the lack of strict feasibility implies that every basic feasible solution is degenerate. We achieve this using facial reduction and simple means of linear algebra. Furthermore, we emphasize that facial reduction involves two steps where the first guarantees strict feasibility, and the second recovers full-row rankness of the constraint matrix.

**Keywords:** facial reduction, preprocessing, linear programming, degeneracy.

**AMS Classification:** 90C05, 90C49.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Preliminaries</b>	<b>3</b>
2.1	Background and Notation . . . . .	3
2.2	Facial Reduction . . . . .	3
<b>3</b>	<b>Main Result</b>	<b>5</b>
3.1	Lack of Strict Feasibility and Relations to Degeneracy . . . . .	5
3.1.1	An Algebraic Proof of Theorem 3.1 via the Definition of Basic Feasible Solution	5
3.1.2	A Geometric Proof Using Extreme Points . . . . .	6
3.2	Discussions . . . . .	7
3.2.1	Distance to Infeasibility . . . . .	8
3.2.2	Applications to Known Characterizations for Strict Feasibility . . . . .	10
3.2.3	Lack of Strict Feasibility in the Dual . . . . .	10
3.2.4	Lack of Strict Feasibility and Interior Point Methods . . . . .	11

---

\*Department of Combinatorics and Optimization Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1;

†Department of Combinatorics and Optimization Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1; Research supported by The Natural Sciences and Engineering Research Council of Canada; [www.math.uwaterloo.ca/~hwolkowi](http://www.math.uwaterloo.ca/~hwolkowi).

<b>4 Numerics</b>	<b>12</b>
4.1 Numerical Experiments with Interior Point Methods . . . . .	12
4.1.1 Generating <b>LPs</b> without Strict Feasibility . . . . .	12
4.1.2 Empirics for Condition Numbers . . . . .	13
4.1.3 Empirics on Stopping Criteria . . . . .	14
4.1.4 Empirics on Distance to Infeasibility . . . . .	14
4.2 Numerical Experiments with Simplex Method . . . . .	15
4.2.1 Generating Dual <b>LPs</b> without Strict Feasibility . . . . .	16
4.2.2 Empirics on the Number of Degenerate Iterations . . . . .	16
<b>5 Conclusion</b>	<b>17</b>
<b>Index</b>	<b>19</b>
<b>References</b>	<b>20</b>

## 1 Introduction

The *Slater condition* (strict feasibility) is a useful property that a model can possess. Unlike general conic programs, linear programs (**LPs**) do not require strict feasibility as a constraint qualification that guarantees strong duality, and therefore, it is often not discussed. The Goldman-Tucker theorem [16] is related in that it guarantees a primal-dual optimal solution satisfying strict complementarity  $x^* + z^* > 0$ . However, it does not guarantee the existence of a strictly feasible primal solution  $\hat{x} > 0$ . The lack of strict feasibility for an **LP** does not seem to cause problems at first glance especially when the simplex method is used. In this manuscript, we show that the failure of strict feasibility causes degeneracy problems when the simplex-type method is used. More specifically, the lack of strict feasibility inevitably renders **LPs** degenerate, i.e., *every basic feasible solution is degenerate*.

The simplex method [9] is one of the most popular and successful algorithms for solving linear programs. Degeneracy that could result in cycling and nonconvergence is one of the early difficulties that arose. There are many anti-cycling rules, e.g., [3, 10, 29], that are developed in order to avoid these issues. However, techniques for the resolution of degeneracy often result in *stalling* [2, 6, 26], i.e., result in a large number of iterations to leave a degenerate point. Degeneracies are known to cause numerical issues when interior point methods are used, e.g., [20]. For example, degeneracy can result in singularity of the Jacobian and thus in ill-posedness and loss of accuracy [17].

Our main results on the degeneracy are shown using the process called *facial reduction*, **FR**. Facial reduction is an effective preprocessing tool to use in the absence of strict feasibility. Given a problem with lack of strict feasibility, facial reduction strives to formulate an equivalent problem so that the reformulation has a Slater point. By examining the facially reduced system, we obtain two results. First, we show that every basic feasible solution is degenerate when strict feasibility fails. Second, we understand a source of instability arising in problems that fail strict feasibility.

The manuscript is organized as follows. In Section 2 we present the background and notations. We then describe what facial reduction tries to achieve, and present related needed properties. In Section 3 we present our main result and immediate corollaries. We then relate our main result

to known results in the literature, such as distance to infeasibility. In section Section 4 we present various algorithmic performances of the interior point methods and the simplex method under the lack of strict feasibility. Finally we present our conclusions in Section 5.

## 2 Preliminaries

### 2.1 Background and Notation

We let  $\mathbb{R}^n, \mathbb{R}^{m \times n}$  be the standard real vector spaces of  $n$ -coordinates and  $m$ -by- $n$  matrices, respectively. We use  $\mathbb{R}_+^n$  ( $\mathbb{R}_{++}^n$ , resp) to denote the  $n$ -tuple with nonnegative (positive) entries. Given a matrix  $A \in \mathbb{R}^{m \times n}$ , we adopt the MATLAB notation to denote a submatrix of  $A$ . Given a subset  $\mathcal{I}$  of column indices,  $A(:, \mathcal{I}) \in \mathbb{R}^{m \times |\mathcal{I}|}$  is the submatrix of  $A$  that contains the columns of  $A$  in  $\mathcal{I}$ . Given a convex set  $\mathcal{C}$ ,  $\text{relint}(\mathcal{C})$  denotes the relative interior of the set  $\mathcal{C}$ .

Throughout this manuscript, we work with feasible **LPs** in standard form with finite optimal value:

$$(\mathcal{P}) \quad p^* = \min_x \{c^T x : Ax = b, x \geq 0\},$$

where  $p^* \in \mathbb{R}, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$  and  $c \in \mathbb{R}^n$ . We assume that  $\text{rank}(A) = m$ , i.e., there is no redundant constraint. We use  $\mathcal{F}$  to denote the feasible region of  $(\mathcal{P})$

$$\mathcal{F} = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}. \quad (2.1)$$

Given an index set  $\mathcal{B} \subset \{1, \dots, n\}, |\mathcal{B}| = m$ , a point  $x \in \mathcal{F}$  is called a *basic feasible solution* if  $A(:, \mathcal{B})$  is nonsingular and  $x_i = 0, \forall i \in \{1, \dots, n\} \setminus \mathcal{B}$ . It is well-known that the simplex method iterates from a basic feasible solution to a basic feasible solution. A basic feasible solution  $x \in \mathcal{F}$  is *nondegenerate* if  $x_i > 0, \forall i \in \mathcal{B}$ . A basic feasible solution  $x \in \mathcal{F}$  is *degenerate* if  $x_i = 0$ , for some  $i \in \mathcal{B}$ . It is clear, from the definition, that every basic feasible solution has at most  $m$  positive entries.<sup>1</sup>

### 2.2 Facial Reduction

In this section we describe the concept of facial reduction and present the properties that are used to establish the main result. We emphasize that facial reduction for  $(\mathcal{P})$  involves *two* steps: first, obtain an equivalent problem with strict feasibility; second, recover full-row rankness of the constraint matrix.

Let  $K \subset \mathbb{R}^n$  be a convex set. A convex set  $F \subseteq K$  is called a *face* of  $K$ , denoted  $F \trianglelefteq K$ , if for all  $y, z \in K$  with  $x = \frac{1}{2}(y + z) \in F$ , we have  $y, z \in F$ . Given a convex set  $\mathcal{C} \subseteq K$ , the *minimal face* for  $\mathcal{C}$  is the intersection of all faces containing the set  $\mathcal{C}$ .

**Proposition 2.1.** [12, Theorem 3.1.3](theorem of the alternative) *For the feasible system of (2.1), exactly one of the following statements holds:*

1. *There exists  $x \in \mathbb{R}_{++}^n$  with  $Ax = b$ , i.e., strict feasibility holds;*
2. *There exists  $y \in \mathbb{R}^m$  such that*

$$0 \neq z := A^T y \in \mathbb{R}_+^m, \quad \text{and} \quad \langle b, y \rangle = 0. \quad (2.2)$$

---

<sup>1</sup>We only consider primal degeneracy here, though everything follows through for dual degeneracy.

Proposition 2.1 gives rise to a process called *facial reduction*. The *facial reduction*, **FR**, for an **LP** is a process of identifying the minimal face of  $\mathbb{R}_+^n$  containing the feasible set  $\mathcal{F} = \{x \in \mathbb{R}_+^n : Ax = b\}$ . By finding the minimal face, we can work with a problem that lies in a smaller dimensional space and that satisfies strict feasibility. The **FR** process, i.e., finding the minimal face, is usually done by solving a sequence of auxiliary systems (2.2). More details on **FR** on general conic problems can be found in [4, 5, 12, 23, 27].

We now describe how the set  $\mathcal{F}$  (see (2.1)) is represented after **FR**. Suppose that strict feasibility fails. Then Proposition 2.1 implies that there must exist a nonzero  $y \in \mathbb{R}^m$  satisfying

$$\langle x, A^T y \rangle = \langle Ax, y \rangle = \langle b, y \rangle = 0, \quad \forall x \in \mathcal{F}. \quad (2.3)$$

Hence, every  $x \in \mathcal{F}$  is perpendicular to the nonnegative vector  $z = A^T y$ . We call this vector  $z = A^T y$  an *exposing vector* for  $\mathcal{F}$ , and let the cardinality of its support be  $s_z = |\{i : z_i > 0\}|$ .

Then  $z = \sum_{j=1}^{s_z} z_{t_j} e_{t_j}$ , where  $t_j$  is in nondecreasing order. We now have

$$0 = \langle z, x \rangle \quad \text{and} \quad x, z \in \mathbb{R}_+^n \implies x_i z_i = 0, \quad \forall i,$$

i.e., the positive elements in  $z$  fix the corresponding elements in  $x$  to zero. Then  $x = \sum_{j=1}^{n-s_z} x_{s_j} e_{s_j}$ , where  $s_j$  is in a nondecreasing order. We define the matrix with unit vectors for columns

$$V = [e_{s_1} \quad e_{s_2} \quad \dots \quad e_{s_{n-s_z}}] \in \mathbb{R}^{n \times (n-s_z)}.$$

Then we have

$$\mathcal{F} = \{x \in \mathbb{R}_+^n : Ax = b\} = \{x = Vv \in \mathbb{R}^n : AVv = b, v \in \mathbb{R}_+^{n-s_z}\}. \quad (2.4)$$

We call this matrix  $V \in \mathbb{R}^{n \times (n-s_z)}$  a *facial range vector*. The facial range vector confines the range that every feasible  $x$  can have. We use the identification (2.4) throughout this manuscript. This concludes the first step of **FR**, i.e., guaranteeing the strict feasibility.

It is known that every facial reduction yields at least one constraint being redundant, see e.g., [5], [21, Lemma 2.7], and [27, Section 3.5]. For completeness we now include a short proof tailored to **LP**, see Lemma 2.2.

**Lemma 2.2.** *Consider the facially reduced feasible set*

$$\mathcal{F}_r = \{v : AVv = b, v \in \mathbb{R}_+^{n-s_z}\}. \quad (2.5)$$

*Then at least one linear constraint of the **LP** is redundant.*

*Proof.* Let  $z = A^T y$  be the exposing vector satisfying the auxiliary system (2.2). And let  $V$  be a facial range vector induced by  $z$ . Then

$$0 = V^T z = V^T A^T y = (AV)^T y = \sum_{i=1}^m y_i ((AV)^T)_i.$$

Since  $y \in \mathbb{R}^m$  is a nonzero vector, the rows of  $AV$  are linearly dependent. □

We now see the result of the full two step facial reduction process, i.e., we get the constraint

matrix of the full-row rank:

$$\mathcal{F} = \{x \in \mathbb{R}_+^n : Ax = b\} = \{x = Vv \in \mathbb{R}^n : P_{\bar{m}}AVv = P_{\bar{m}}b, v \in \mathbb{R}_+^{n-s_z}\}, \quad (2.6)$$

where  $P_{\bar{m}} : \mathbb{R}^m \rightarrow \mathbb{R}^{\bar{m}}$ ,  $\bar{m} = \text{rank}(AV)$ , is the simple projection that chooses the linearly independent rows of  $AV$ . We emphasize the importance of this projection by relating to the so-called *distance to infeasibility* in Section 3.2.1 below. This concludes the second step of **FR**, i.e., guaranteeing the full rank.

For a general conic problem, such as semidefinite programs (**SDP**), the facial reduction iterations do not necessarily end in one iteration; see [8, 27, 28]. And there is a special name for the minimum length of **FR** iterations. Given a spectrahedron  $\mathcal{S}$ , the *singularity degree* of  $\mathcal{S}$ , denoted by  $\text{sd}(\mathcal{S})$ , is the smallest number of facial reduction iterations for finding  $\text{face}(\mathcal{S})$ . However, for **LPs**, it is known that **FR** can be done in *one* iteration, i.e.,  $\text{sd}(\mathcal{F}) \leq 1$ ; see [12, Theorem 4.4.1]. Unlike the **FR** performed on the class of **SDPs**, the **FR** performed on the **LPs** does not alter the sparsity pattern of the data matrix  $A$ . We emphasize that the **FR** on the set  $\mathcal{F}$  only involves the discarding the rows and columns of  $A$ ; the sparsity pattern of  $A$  does not change after these operations.

### 3 Main Result

In this section we present our main result Theorem 3.1. We provide two proofs: one takes an algebraic approach by using the definition of the basic feasible solution; and the other takes a geometric approach by using extreme points. Both proofs rely heavily on Lemma 2.2. In Section 3.2 we include immediate corollaries of the main result and interesting discussions.

#### 3.1 Lack of Strict Feasibility and Relations to Degeneracy

**Theorem 3.1.** *Suppose that strict feasibility of  $\mathcal{F}$  fails. Then every basic feasible solution to  $\mathcal{F}$  is degenerate.*

##### 3.1.1 An Algebraic Proof of Theorem 3.1 via the Definition of Basic Feasible Solution

*Proof.* Since there is no strictly feasible point in  $\mathcal{F}$ , there exists a facial range vector  $V$ , and as in (2.4) we have

$$\mathcal{F} = \{x \in \mathbb{R}^n : AVv = b, v \in \mathbb{R}_+^{n-s_z}\}.$$

By Lemma 2.2,  $AV$  has at least one redundant row. By permuting the columns of  $A$ , we may assume that the matrix  $V$  is of the form

$$V = \begin{bmatrix} I_r \\ 0 \end{bmatrix} \text{ and } r = n - s_z.$$

We partition the index set  $\{1, \dots, n\}$  as

$$\{1, \dots, n\} = \mathcal{I}_+ \cup \mathcal{I}_0, \text{ where } \mathcal{I}_+ = \{1, \dots, r\} \text{ and } \mathcal{I}_0 = \{r+1, \dots, n\}.$$

Let  $\bar{x} \in \mathcal{F}$  be a basic feasible solution with basic indices

$$\mathcal{B} \subset \{1, \dots, n\}, |\mathcal{B}| = m, \det(A(:, \mathcal{B})) \neq 0, \text{ and } A(:, \mathcal{B})\bar{x}(\mathcal{B}) = b.$$

Suppose  $\mathcal{B} \subseteq \mathcal{I}_+$ . We note, by Lemma 2.2 again, that  $A(:, \mathcal{I}_+) = AV$  has redundant rows, i.e.,  $\text{rank}(A(:, \mathcal{I}_+)) < m$ . Hence  $\bar{x}$  must include a basic variable in  $\mathcal{I}_0$  and this concludes that every basic feasible solution is degenerate.  $\square$

### 3.1.2 A Geometric Proof Using Extreme Points

We now give the second proof of our main result. We first employ the statement presented in [22]. In Proposition 3.2 below,  $\mathbb{S}_+^n$  denotes the set of  $n$ -by- $n$  positive semidefinite matrices.

**Proposition 3.2.** [22, Theorem 2.1] *Suppose that  $X \in F$ , where  $F$  is a face of the set  $\{X \in \mathbb{S}_+^n : \text{trace}(A_i X) = b_i, \forall i = 1, \dots, m\}$ . Let  $d = \dim F$ ,  $r = \text{rank}(X)$ . Then  $\frac{r(r+1)}{2} \leq m + d$ .*

The set in Proposition 3.2 is called a spectrahedron. Feasible sets of standard semidefinite programs are represented as spectrahedra. A sepectrahedron is a generalization of the polyhedral set  $\mathcal{F}$  and the proof from [22, Theorem 2.1] can be altered to work with  $\mathcal{F}$ . We include the proof for completeness.

**Corollary 3.3.** *Suppose that  $x \in F$ , where  $F$  is a face of the set  $\mathcal{F}$ . Let  $r$  be the number of nonzeros in  $x$  and  $d = \dim F$ . Then the number of nonzero entries of  $x \in F$  is at most  $m + d$ .*

*Proof.* Let  $x \in F$  and let  $r$  be the number positive entries in  $x$ . Let  $\bar{x} \in \mathbb{R}^r$  be the vector obtained by discarding the 0 entries in  $x$ . This is readily given by the following matrix-vector multiplication  $\bar{x} = I(\text{supp}(x), :)x$ , where  $\text{supp}(x)$  is the support of  $x$ , the set of indices  $\{i : x_i > 0\}$ . Let  $\bar{A} \in \mathbb{R}^{m \times r}$  be the matrix after removing the columns of  $A$  that are not in the support of  $x$ , i.e.,  $\bar{A} = A(:, \text{supp}(x))$ . We note that  $\bar{x}$  is a particular solution to the system  $\bar{A}z = b$  and  $\bar{x} > 0$ .

Suppose to the contrary that  $r > m + d$ . Since  $r - m > d$ , there exists at least  $d + 1$  linearly independent vectors, say  $v_1, \dots, v_{d+1} \in \mathbb{R}^r$ , satisfying  $\bar{A}v_i = 0$ ,  $\forall i = 1, \dots, d + 1$ . For each  $i \in \{1, \dots, d + 1\}$  and for  $\epsilon \in \mathbb{R}$ , we define

$$\begin{aligned} v_{i,+} &:= \bar{x} + \epsilon v_i, & v_{i,-} &:= \bar{x} - \epsilon v_i, \\ x_{i,+} &:= I(:, \text{supp}(x))(\bar{x} + \epsilon v_i), & x_{i,-} &:= I(:, \text{supp}(x))(\bar{x} - \epsilon v_i). \end{aligned}$$

For a sufficiently small  $\epsilon$ , we have  $x_{i,+}, x_{i,-} \in \mathcal{F}$ . We note that  $x = \frac{1}{2}(x_{i,+} + x_{i,-})$ ,  $\forall i$ . Hence, by the definition of face,  $x_{i,+} \in F$ ,  $\forall i$ . Therefore,  $F$  contains vectors  $\{x_{i,+}\}_{i=1, \dots, d+1} \cup \{x\}$  that are affinely independent and hence  $\dim(F) \geq d + 1$ .  $\square$

A point  $x$  in a convex set  $\mathcal{C}$  is called an *extreme point* if, for all  $y, z \in \mathcal{C}$ ,  $x = \frac{1}{2}(y + z)$  implies  $x = y = z$ . An extreme point is itself a face and the dimension of this face is 0. Hence, we obtain Corollary 3.4 by writing Corollary 3.3 through the lens of extreme points.

**Corollary 3.4.** *Every extreme point  $x \in \mathcal{F}$  has at most  $m$  positive entries.*

We now restate the main result of this paper Theorem 3.1 in the language of extreme points and number of rows of  $A$ .

**Theorem 3.5.** *Suppose that strict feasibility of  $\mathcal{F}$  fails. Then every extreme point  $x \in \mathcal{F}$  has at most  $m - 1$  positive entries.*

*Proof.* Since strict feasibility fails for  $\mathcal{F}$ , we have  $\mathcal{F} = \{x = Vv \in \mathbb{R}^n : AVv = b, v \in \mathbb{R}_+^{n-s_z}\}$ ; see (2.4). From Lemma 2.2, we note that at least one equality in  $AVv = b$  is redundant. Let  $P_{\bar{m}}AVv = P_{\bar{m}}b$  be the system obtained after removing redundant rows of  $AV$ ; see (2.6). Then, by Corollary 3.4, every extreme point of the set  $\{v \in \mathbb{R}_+^{n-s_z} : P_{\bar{m}}AVv = P_{\bar{m}}b\}$  has at most  $m - 1$  nonzero entries. Hence, the statement follows.  $\square$

**Remark 3.6.** *The idea used in the proof of Theorem 3.5 is the same as the one presented in [21] for spectrahedron. In [21], the authors use Proposition 3.2 to strengthen the bound called the Barvinok-Pataki bound. The bound is strengthened by the means of singularity degree that stems from the facial reduction algorithm [12, 23, 27]. The number of nonzeros in  $x$  in Theorem 3.5 plays the role of  $\text{rank}(X)$  in Proposition 3.2. Facial reduction applied to spectrahedra also yields redundant constraints and hence a similar result follows for spectrahedra.*

## 3.2 Discussions

In this section we discuss the main result in Section 3.1 and make connections to known results in the literature.

Theorem 3.1 and Theorem 3.5 are equivalent owing to the well-known characterization:

$$x \in \mathcal{F} \text{ is a basic feasible solution} \iff x \in \mathcal{F} \text{ is an extreme point.}$$

We highlight that Theorem 3.1 and Theorem 3.5 do not merely state the existence of a *single* degenerate basic feasible solution. It states that *every* basic feasible solution is degenerate. Developing a pivot rule that prevents the simplex method from visiting degenerate points is not possible as it can never stay away from the degeneracies when strict feasibility fails. We provide an example for an illustration.

**Example 3.7.** *Consider  $\mathcal{F}$  with the data*

$$A = \begin{bmatrix} 1 & 1 & 3 & 5 & 2 \\ 0 & 1 & 2 & -2 & 2 \end{bmatrix} \text{ and } b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

*Consider the vector  $y = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ . Then*

$$A^T y = (1 \ 0 \ 1 \ 7 \ 0)^T \text{ and } b^T y = 0.$$

*Hence, Proposition 2.1 certifies that  $\mathcal{F}$  does not contain a strictly feasible point. There are exactly six feasible bases in  $\mathcal{F}$ ; the basic feasible solution associated with  $\mathcal{B} = \{\{1, 2\}, \{2, 3\}, \{2, 4\}\}$  is  $x = (0 \ 1 \ 0 \ 0 \ 0)^T$  and the basic feasible solution associated with  $\mathcal{B} \in \{\{1, 5\}, \{3, 5\}, \{4, 5\}\}$  is  $x = (0 \ 0 \ 0 \ 0 \ \frac{1}{2})^T$ . Clearly, all basic feasible solutions are degenerate.*

Below, we obtain an interesting statement by writing the contrapositive of Theorem 3.1. Similarly, we provide Example 3.9 below to illustrate Corollary 3.8.

**Corollary 3.8.** *If there exists a nondegenerate basic feasible solution, then there exists a strictly feasible point in  $\mathcal{F}$ .*

**Example 3.9.** *Consider  $\mathcal{F}$  with the data*

$$A = \begin{bmatrix} 1 & 0 & -2 & 3 & -4 \\ 0 & -1 & -2 & 3 & 1 \end{bmatrix} \text{ and } b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

*The system  $\mathcal{F}$  has exactly four feasible bases; the basic feasible solution associated with  $\mathcal{B} \in \{\{1, 4\}, \{2, 4\}, \{4, 5\}\}$  is  $x = (0 \ 0 \ 0 \ 1/3 \ 0)^T$  and the basic feasible solution associated with  $\mathcal{B} = \{1, 5\}$  is  $x = (5 \ 0 \ 0 \ 0 \ 1)^T$ . We note that the basic feasible solution associated with  $\mathcal{B} = \{1, 5\}$  is nondegenerate. As Corollary 3.8 states, the system  $\mathcal{F}$  has a strictly feasible point, and it is verified by the point  $\frac{1}{10}(4 \ 1 \ 1 \ 4 \ 1)^T$ .*

We exhibit Example 3.10 below to show that the converse of Theorem 3.1 and Theorem 3.5 is not true.

**Example 3.10.** Consider  $\mathcal{F}$  with the data

$$A = \begin{bmatrix} 1 & 0 & 2 & 0 & -2 \\ 1 & -3 & 2 & 1 & -2 \end{bmatrix} \text{ and } b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

$\mathcal{F}$  has exactly four feasible bases and all of them are degenerate; the basic feasible solution associated with  $\mathcal{B} \in \{\{1, 2\}, \{1, 4\}\}$  is  $x = (1 \ 0 \ 0 \ 0 \ 0)^T$  and the basic feasible solution associated with  $\mathcal{B} \in \{\{2, 3\}, \{3, 4\}\}$  is  $x = (0 \ 0 \ 1/2 \ 0 \ 0)^T$ . However,  $\mathcal{F}$  contains a strictly feasible point  $\frac{1}{10}(1 \ 1 \ 5.5 \ 3 \ 1)^T$ .

Given a basic feasible solution  $\bar{x} \in \mathcal{F}$ , the *degree of degeneracy* of  $\bar{x}$  is the number of 0's in its basic variables. By exploiting the facially reduced model we can check how degenerate the basic feasible solutions of  $\mathcal{F}$  are.

**Corollary 3.11.** Suppose that strict feasibility fails for  $\mathcal{F}$  and let  $\mathcal{F}$  have the representation (2.4). Let  $\mathcal{I}_0$  be the set of indices that holds  $\mathcal{I}_0 = \{i \in \{1, \dots, n\} : x_i = 0, \forall x \in \mathcal{F}\}$ . Let  $\bar{x} \in \mathcal{F}$  be a basic feasible solution. Then, the followings hold.

1. The degree of degeneracy of  $\bar{x}$  is at least  $m - \text{rank}(AV)$ .
2. At least  $m - \text{rank}(AV)$  number of basic indices of  $\bar{x}$  are contained in  $\mathcal{I}_0$ .

*Proof.* Let  $\bar{x} \in \mathcal{F}$  be a basic feasible solution and let  $\mathcal{B}$  be a basis for  $\bar{x}$ . We note that  $A(:, \mathcal{B})$  contains linearly independent columns. Then  $A(:, \mathcal{B})$  can contain at most  $\text{rank}(AV)$  number of columns from  $AV$ . Thus,  $\bar{x}(\mathcal{B})$  must contain at least  $m - \text{rank}(AV)$  number of zeros. Item 2 is a direct consequence of Item 1.  $\square$

### 3.2.1 Distance to Infeasibility

The *distance to infeasibility* is a measure of the smallest perturbations of the data of a problem that renders the problem infeasible. In our setting, we can use the following simplification of the distance to infeasibility from [24] by restricting the perturbations to  $b$ , i.e., we can force infeasibility using only perturbations in  $b$ ;

$$\text{dist}(b, \mathcal{F} = \emptyset) := \inf \left\{ \|b - \tilde{b}\| : \{x \in \mathbb{R}^n : Ax = \tilde{b}, x \geq 0\} = \emptyset \right\}.$$

Many interesting bounds, condition numbers, are shown in [24] under the assumption that the distance to infeasibility is positive and known. It is known that a positive distance to infeasibility of  $\mathcal{F}$  implies that strict feasibility holds for  $\mathcal{F}$ ; see e.g., [13, 14]. The contrapositive of this statement is that, if strict feasibility fails for  $\mathcal{F}$ , then the distance to infeasibility is 0. We revisit this statement with the facially reduced system (2.4). We provide an elementary proof that there is an arbitrarily small perturbation for the data vector  $b$  of  $\mathcal{F}$  that yields the set  $\mathcal{F}$  infeasible, i.e.,  $\text{dist}(b, \mathcal{F} = \emptyset) = 0$ . Furthermore, we provide an explicit perturbation that renders the set  $\mathcal{F}$  empty.

Suppose that  $\mathcal{F}$  fails strict feasibility. Recall the representation (2.4) for  $\mathcal{F}$ . Let  $AV = QR$  be a qr decomposition of  $AV$ , where  $Q \in \mathbb{R}^{m \times m}$  orthogonal,  $R \in \mathbb{R}^{m \times (n-s_z)}$  upper triangular. We write  $Q = [Q_1 \ Q_2]$  so that  $\text{range}(Q_1) = \text{range}(AV)$ . Then, by the orthogonality of  $Q$ , we have

$$Ax = AVv = b \iff Q^T Ax = Rv = Q^T b. \tag{3.1}$$



Since  $AV$  is a rank deficient matrix (see Lemma 2.2), the upper triangular matrix  $R$  is of the form

$$R = \begin{bmatrix} \bar{R} \\ 0 \end{bmatrix} \in \mathbb{R}^{m \times (n-s_z)} \text{ and } \bar{R} \in \mathbb{R}^{\text{rank}(AV) \times (n-s_z)} \text{ with nonzero diagonal.} \quad (3.2)$$

If we perturb the vector  $b \in \mathbb{R}^m$  so that at least one of the last  $m - \text{rank}(AV)$  entries of  $Q^T b$  becomes nonzero, then the system (3.1) becomes infeasible. Such perturbations can be found using linear combinations of the columns of  $Q_2$ ; see Proposition 3.12 below.

**Proposition 3.12.** *Suppose that strict feasibility fails for  $\mathcal{F}$  and let  $\mathcal{F}$  have the representation (2.4). Then the following hold.*

1. *For all  $\Delta b \in \text{range}(AV)$  with sufficiently small norm, the set  $\{x \in \mathbb{R}_+^n : Ax = b + \Delta b\}$  is feasible.*
2. *The distance to infeasibility of  $\mathcal{F}$  is 0, i.e.,  $\text{dist}(b, \mathcal{F} = \emptyset) = 0$ .*

*Proof.* Let  $\Delta b$  be any perturbation in  $\text{range}(AV)$ . Let  $QR = AV$  be a qr decomposition of  $AV$ . In particular, let  $R$  have the form (3.2) and  $Q = [Q_1 \ Q_2]$  so that  $\text{range}(Q_1) = \text{range}(AV)$ . Then

$$Ax = AVv = b + \epsilon \Delta b \iff Rv = Q^T b + \epsilon Q^T \Delta b \iff \bar{R}v = Q_1^T b + \epsilon Q_1^T \Delta b. \quad (3.3)$$

The last equivalence holds since  $Ax = b$  and  $\Delta b \in \text{range}(AV) = \text{range}(Q_1)$ . Since the system  $\bar{R}v = Q_1^T b$  satisfies strict feasibility, the distance to infeasibility of this system is positive. Thus, the perturbed system  $\bar{R}v = Q_1^T b + \epsilon Q_1^T \Delta b$  remains feasible. Therefore, by (3.3), perturbing  $\mathcal{F}$  along the direction  $\Delta b \in \text{range}(AV)$  maintains the feasibility and this concludes the proof for Item 1.

For Item 2 we present a perturbation  $\Delta b$  to  $b$  that renders  $\mathcal{F}$  infeasible. By Proposition 2.1, we have a nonzero vector  $\bar{y} \in \mathbb{R}^m$  that satisfies (2.2). Then we have

$$\bar{y} \in \text{range}(AV)^\perp \implies \bar{y} = Q_2 \bar{u} \text{ for some nonzero } \bar{u}.$$

We recall Farkas' lemma:

$$\{y \in \mathbb{R}^m : A^T y \geq 0, \langle b, y \rangle < 0\} \neq \emptyset \implies \mathcal{F} = \emptyset.$$

Now, for any  $\epsilon > 0$ , setting  $\Delta b_\epsilon = -\epsilon \bar{y}$  yields

$$A^T \bar{y} \geq 0, \langle b, \bar{y} \rangle = 0 \implies A^T \bar{y} \geq 0, \langle b + \Delta b_\epsilon, \bar{y} \rangle < 0. \quad (3.4)$$

Hence, by letting  $\epsilon \rightarrow 0^+$ , we see that the distance to infeasibility,  $\text{dist}(b, \mathcal{F} = \emptyset)$ , is equal to 0.  $\square$

We emphasize that the result

$$\mathcal{F} \text{ fails strictly feasibility} \implies \text{dist}(b, \mathcal{F} = \emptyset) = 0$$

gives rise to the second step (2.6) of **FR** discussed in Section 2.2. We note that the instability discussed in this section essentially originates from the observation made in Lemma 2.2, i.e., redundant equalities arise in the facially reduced system. Facially reduced system allows us to exploit the root of potential instability when the right-hand-side vector  $b$  is perturbed. Although the distance to infeasibility is 0 in the absence of strict feasibility, Proposition 3.12 suggests that a carefully chosen perturbation of  $b$  does not have an impact on the feasibility of  $\mathcal{F}$ . We provide a related numerical experiment in Section 4.1.4 below.

The distance to infeasibility directly impacts the measure of well-posedness of the problem, [13, 14, 25]. Given the pair  $d = (A, b)$  of the data for an instance  $(\mathcal{P})$ , the *condition measure* of  $(\mathcal{P})$  is defined by

$$\mathcal{C}(d) := \frac{\|d\|}{\inf\{\|\Delta d\| : d + \Delta d \text{ yields } (\mathcal{P}) \text{ infeasible}\}}.$$

The value  $\mathcal{C}(d)$  is a measure of well-posedness of the problem  $(\mathcal{P})$ . Since  $\text{dist}(b, \mathcal{F} = \emptyset) = 0$ , we have  $\mathcal{C}(d) = \infty$ . Namely, when strict feasibility fails for  $(\mathcal{P})$ , the problem is ill-posed.

### 3.2.2 Applications to Known Characterizations for Strict Feasibility

There are some known characterizations for strict feasibility of  $\mathcal{F}$ . Using these characterizations we can obtain extensions of Theorem 3.1, Theorem 3.5, and Corollary 3.8.

The dual  $(\mathcal{D})$  of  $(\mathcal{P})$  is

$$(\mathcal{D}) \quad \max_{y,s} \{b^T y : A^T y + s = c, s \geq 0\}. \quad (3.5)$$

It is known that strict feasibility fails for  $\mathcal{F}$  if, and only if, the set of optimal solutions for the dual  $(\mathcal{D})$  is unbounded; see e.g., [30, Theorem 2.3] and [15]. Then Corollary 3.13 follows.

**Corollary 3.13.** *1. Suppose that the set of optimal solutions for the dual  $(\mathcal{D})$  is unbounded. Then every basic feasible solution to  $\mathcal{F}$  is degenerate.*

*2. Suppose that there exists a nondegenerate basic feasible solution to  $\mathcal{F}$ . Then the set of optimal solutions for the dual  $(\mathcal{D})$  is bounded.*

It is known that strict feasibility holds for  $\mathcal{F}$  if, and only if,  $b \in \text{relint}(A(\mathbb{R}_+^n))$ , where  $\text{relint}$  denote the relative interior; see e.g., [12, Proposition 4.4.1]. Then if one finds a set of indices  $\mathcal{I} \subset \{1, \dots, n\}$  such that  $A(:, \mathcal{I})$  is nonsingular and  $A(:, \mathcal{I})z = b$  has a solution  $z$  with positive entries, then  $b \in \text{relint}(A(\mathbb{R}_+^n))$ .

### 3.2.3 Lack of Strict Feasibility in the Dual

In this section we consider the facial reduction for the dual  $(\mathcal{D})$ ; see (3.5). We denote the feasible set of the dual  $(\mathcal{D})$  by

$$\mathcal{G} := \{(y, s) \in \mathbb{R}^m \oplus \mathbb{R}_+^n : A^T y + s = c\}. \quad (3.6)$$

The facial reduction arguments applied to the dual are parallel to the ones given in Section 2.2. Hence, we provide a short derivation for the facially reduced system for  $\mathcal{G}$ . We also conclude that the absence of strict feasibility for  $\mathcal{G}$  implies the dual degeneracy at all basic feasible solutions.

The following lemma is the theorem of the alternative applied to the set  $\mathcal{G}$ .

**Lemma 3.14.** [7, Theorem 3.3.10] *(theorem of the alternative in dual form) Let  $\mathcal{G}$  in (3.6) be feasible. Then, exactly one of the following statements holds:*

1. *There exists  $(y, s) \in \mathbb{R}^m \oplus \mathbb{R}_+^n$  with  $A^T y + s = c$ , i.e., strict feasibility holds for  $\mathcal{G}$ ;*
2. *There exists  $w \in \mathbb{R}^n$  such that*

$$0 \neq w \in \mathbb{R}_+^n, Aw = 0 \text{ and } \langle c, w \rangle = 0. \quad (3.7)$$

We recall that the vector  $A^T y$  in (2.3) is an exposing vector to the set  $\mathcal{F}$ . Let  $w$  be a solution to the auxiliary system (3.7). Similarly, the vector  $w$  plays the role of an exposing vector for  $\mathcal{G}$ :

$$\forall (y, s) \in \mathcal{G}, \text{ it holds } \langle w, s \rangle = \langle w, c - A^T y \rangle = \langle c, w \rangle - \langle Aw, y \rangle = 0 - \langle 0, y \rangle = 0. \quad (3.8)$$

We let

$$\mathcal{I}_w = \{1, \dots, n\} \setminus \text{supp}(w), \quad U = I(:, \mathcal{I}_w) \quad \text{and} \quad s_w = |\text{supp}(w)|.$$

Then, the facially reduced system of  $\mathcal{G}$  appears

$$\mathcal{G} = \left\{ (y, s) \in \mathbb{R}^m \oplus \mathbb{R}_+^n : \begin{bmatrix} A^T & I \end{bmatrix} \begin{pmatrix} y \\ s \end{pmatrix} = c \right\} = \left\{ (y, u) \in \mathbb{R}^m \oplus \mathbb{R}_+^{n-s_w} : \begin{bmatrix} A^T & U \end{bmatrix} \begin{pmatrix} y \\ u \end{pmatrix} = c \right\}. \quad (3.9)$$

The notion of degeneracy in Section 2.1 naturally extends to an arbitrary polyhedron, e.g., see [1, Section 2]. For a general polyhedron  $P \subseteq \mathbb{R}^n$ , a point  $p$  in  $P$  is called a *basic solution* if there are  $n$  linearly independent active constraints at  $p$ . In addition, if there are more than  $n$  active constraints at the point  $p \in P$ , then the point  $p$  is called *degenerate*. Using this definition of the degeneracy, we now show that the absence of strict feasibility for  $\mathcal{G}$  implies that every basic solution of  $\mathcal{G}$  is degenerate.

We show that the facially reduced system in (3.9) contains a redundant constraint. Let  $w$  be a solution to the system (3.7), i.e.,  $w$  is an exposing vector for  $\mathcal{G}$ . Then we have

$$\begin{bmatrix} A \\ U^T \end{bmatrix} w = \begin{bmatrix} Aw \\ U^T w \end{bmatrix} = \begin{bmatrix} 0_m \\ 0_{n-s_w} \end{bmatrix}.$$

In other words, there is a nontrivial row combination of  $\begin{bmatrix} A^T & U \end{bmatrix}$  that yields the 0 vector, i.e., there exists a redundant row in  $\begin{bmatrix} A^T & U \end{bmatrix}$ . Hence, the facially reduced system contains a redundant constraint. The redundancy immediately implies the dual degeneracy; for any basic solution of  $\mathcal{G}$ , there always exists a redundant equality in  $\begin{bmatrix} A^T & I \end{bmatrix} \begin{pmatrix} y \\ s \end{pmatrix} = c$ .

### 3.2.4 Lack of Strict Feasibility and Interior Point Methods

Many algorithm constructions for interior point methods stem from the optimality condition (KKT conditions) of the primal ( $\mathcal{P}$ ) and the dual ( $\mathcal{D}$ ). And many practical interior point methods find the search direction  $d$  by solving the so-called normal equation, a square system  $ADA^T d = r$ , where  $D$  is a diagonal matrix with positive diagonal and  $r$  is some residual; see e.g., [30, Chapter 11]. The diagonal of  $D$  consists of some element-wise product of the primal variable  $x$  and the dual slack variable  $s$ . We have shown that the lack of strict feasibility for  $\mathcal{F}$  makes all vertices of  $\mathcal{F}$  degenerate. Thus, all vertices that form the optimal face of ( $\mathcal{P}$ ) are also degenerate. Hence, the primal degeneracy can cause the diagonal of  $D$  to have a deficient number of nice positives near the optimum. Thus the normal equation may be ill-conditioned sooner as it get near the optimum, i.e., numerical stability could be hard to achieve. We present related numerics in Section 4.1.

There is a comprehensive survey [20] that concerns problems caused by degeneracies when an interior point method is chosen for LPs. The survey [20] addresses the effect of degeneracy on the convergence of interior point methods and numerical performance, etc.

## 4 Numerics

We now provide empirical evidence that **FR** indeed is a useful preprocessing tool in reducing the size of the problem, and in particular improving the condition number of the problem. We do this first for interior point methods and then for simplex methods.

### 4.1 Numerical Experiments with Interior Point Methods

In this section we compare the behaviour for finding near-optimal points with instances that have strictly feasible points and instances that do not. More specifically, given a near optimal primal-dual point  $(x^*, s^*) \in \mathbb{R}_{++}^n \oplus \mathbb{R}_{++}^n$  from interior point method solvers, we observe the condition number, i.e., the ratio of largest to smallest eigenvalues of the normal matrix at  $(x^*, s^*)$ :

$$\kappa(AD^*A^T), \quad \text{where } D^* = \text{Diag}(x^*) \text{Diag}(s^*)^{-1}. \quad (4.1)$$

We show that instances that do not have strictly feasible points tend to have significantly larger condition numbers of the normal equation near the optimum. We also present a numerical experiment on the perturbation imposed on the right-hand-side vector  $b$ .

#### 4.1.1 Generating LPs without Strict Feasibility

We first show how to generate an instance for  $\mathcal{F}$  that fails strict feasibility. More specifically, given  $m, n, r \in \mathbb{N}$ , we construct the data  $A \in \mathbb{R}^{m \times n}$  and  $b \in \mathbb{R}^m$  to satisfy (2.2) with  $r$  as the dimension of the relative interior of  $\mathcal{F}$ ,  $\text{relint}(\mathcal{F})$ .

1. Pick any  $0 \neq y \in \mathbb{R}^m$ . Let

$$\{y\}^\perp = \text{span}\{a_i\}_{i=1}^{m-1} \quad (= \text{null}(y^T)).$$

We let  $R \in \mathbb{R}^{(m-1) \times r}$  be a random matrix, and get

$$A_1 := [a_1 \ \dots \ a_{m-1}]R \in \mathbb{R}^{m \times r}, \quad A_1^T y = 0 \in \mathbb{R}^r.$$

2. Pick any  $\hat{v} \in \mathbb{R}_{++}^r$  and set  $b = A_1 \hat{v}$ . We note that  $y^T A_1 = 0$  and  $\langle b, y \rangle = 0$ .
3. Pick any matrix  $A_2 \in \mathbb{R}^{m \times (n-r)}$  satisfying  $(y^T A_2)_i \neq 0, \forall i$ . If there exists  $i$  such that  $(y^T A_2)_i < 0$ , then change the sign of  $i$ -th column of  $A_2$  so that we conclude

$$(A_2^T y) \in \mathbb{R}_{++}^{n-r}.$$

4. We define the matrix  $A = [A_1, A_2] \in \mathbb{R}^{m \times n}$ . Then  $\{x \in \mathbb{R}_+^n : Ax = b\}$  is a polyhedron with a feasible point  $\hat{x} = [\hat{v}; 0]$  having  $r$  number of positives. The vector  $y$  is a solution for the system (2.2):

$$0 \preceq z = A^T y = \begin{pmatrix} A_1^T y = 0 \\ A_2^T y > 0 \end{pmatrix}, \quad b^T y = 0.$$

We then randomly permute the columns of  $A$  to avoid the zeros always being at the bottom of the feasible variables  $x$ .

For the empirics, we construct the objective function  $c^T x$  of  $(\mathcal{P})$  as follows. We choose any  $\bar{s} \in \mathbb{R}_{++}^n, \bar{y} \in \mathbb{R}^m$  and set  $c = A^T \bar{y} + \bar{s}$ . Then we have the data for the primal-dual pair of **LPs** and

the primal *fails* strict feasibility:

$$(\mathcal{P}_{(A,b,c)}) \quad \min\{c^T x : Ax = b, x \geq 0\} \quad \text{and} \quad (\mathcal{D}_{(A,b,c)}) \quad \min\{b^T y : A^T y + s = c, s \geq 0\}.$$

We note that by choosing  $\bar{s} \in \mathbb{R}_{++}^n$ , the dual problem  $(\mathcal{D}_{(A,b,c)})$  has a strictly feasible point. In order to generate instances with strictly feasible points, we maintain the same data  $A, c$  used for the pair  $(\mathcal{P}_{(A,b,c)})$  and  $(\mathcal{D}_{(A,b,c)})$ . We only redefine the right-hand-side vector by  $\bar{b} = Ax^\circ$ , where  $x^\circ \in \mathbb{R}_{++}^n$ :

$$(\bar{\mathcal{P}}_{(A,\bar{b},c)}) \quad \min\{c^T x : Ax = \bar{b}, x \geq 0\} \quad \text{and} \quad (\bar{\mathcal{D}}_{(A,\bar{b},c)}) \quad \min\{\bar{b}^T y : A^T y + s = c, s \geq 0\}.$$

The facially reduced instances of  $(\mathcal{P}_{(A,b,c)})$  are denoted by  $(\mathcal{P}_{(A_{FR},b_{FR},c_{FR})})$ . They are obtained by discarding the variables that are identically 0 in the feasible set  $\mathcal{F}$  and the redundant constraints. In other words, the affine constraints of  $(\mathcal{P}_{(A_{FR},b_{FR},c_{FR})})$  are of the form (2.6).

### 4.1.2 Empirics for Condition Numbers

We use three different solvers in our test; linprog from MATLAB<sup>2</sup>, SDPT3<sup>3</sup> and MOSEK<sup>4</sup>. MATLAB version 2021a is used to access the solvers for the test, and we use the default settings for stopping criteria for all solvers. In order to illustrate the differences in condition numbers of the normal matrices, we solve the three families of instances listed below:

1.  $(\mathcal{P}_{(A,b,c)})$  : a family of instances that do not have strictly feasible points;
2.  $(\bar{\mathcal{P}}_{(A,\bar{b},c)})$  : a family of instances that have strictly feasible points;
3.  $(\mathcal{P}_{(A_{FR},b_{FR},c_{FR})})$  : a family of facially reduced instances of  $(\mathcal{P}_{(A,b,c)})$ .

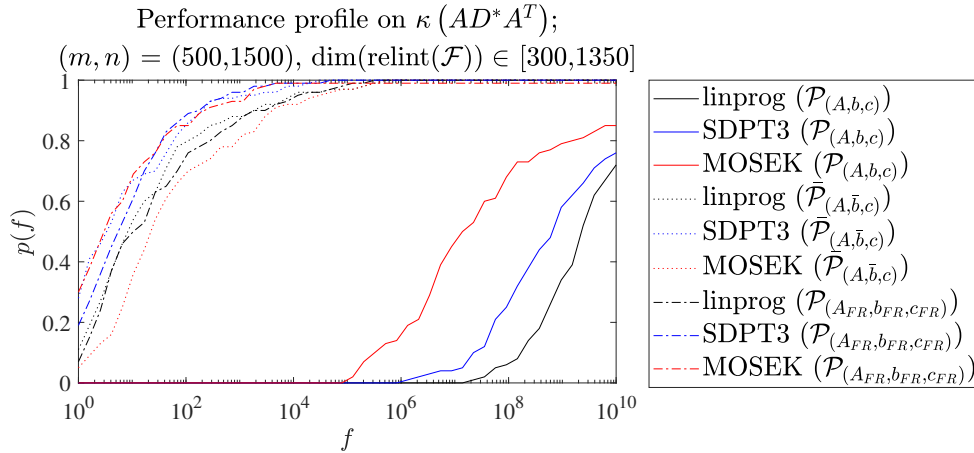


Figure 4.1: Performance profile on  $\kappa(ADA^T)$  with(out) strict feasibility near optimum; various solvers

We use the *performance profile* [11, 18] to observe the overall behaviours on different families of instances using the three solvers. The performance profile provides a useful graphical comparison for solver performances. Figure 4.1 displays the performance profile on the condition numbers of

<sup>2</sup><https://www.mathworks.com/>. Version 9.10.0.1669831 (R2021a) Update 2.

<sup>3</sup><https://www.math.cmu.edu/~reha/sdpt3.html>. Version SDPT3 4.0.

<sup>4</sup><https://www.mosek.com/>. Version 8.0.0.60.

the normal matrix  $AD^*A^T$  near optimal points from different solvers. We generate 100 instances for each family that have  $\dim(\text{relint}(\mathcal{F})) \in [300, 1350]$ . The instance sizes are fixed with  $(m, n) = (500, 1500)$ . The vertical axis in Figure 4.1 represents the statistics of the performance ratio on  $\kappa(AD^*A^T)$ , the condition number of normal matrix near optimum  $(x^*, s^*)$ ; see (4.1). The solid lines in Figure 4.1 represent the performance of the instances  $(\mathcal{P}_{(A,b,c)})$  that fail strict feasibility. They show that the condition numbers of the normal matrices near optima are significantly higher when strict feasibility fails. That is, when strict feasibility fails for  $\mathcal{F}$ , the matrix  $AD^*A^T$  is more ill-conditioned and it is difficult to obtain search directions of high accuracy. We also observe that facially reduced instances yield smaller condition numbers near optima. We note that the instances  $(\mathcal{P}_{(A,b,c)})$  and  $(\mathcal{P}_{(A_{FR}, b_{FR}, c_{FR})})$  are equivalent.

### 4.1.3 Empirics on Stopping Criteria

We now use the three solvers to observe the accuracy of the first-order optimality conditions (KKT conditions), and the running time for the instances  $(\mathcal{P}_{(A,b,c)})$  and  $(\mathcal{P}_{(A_{FR}, b_{FR}, c_{FR})})$ . Table 4.1 exhibits the numerics on these instances. We test the average performance of 10 instances of the size  $(n, m, r) = (3000, 500, 2000)$ . The headers used in Table 4.1 provide the following. Given solver outputs  $(x^*, y^*, s^*)$ , the header ‘KKT’ exhibits the average of the triple consisting of the primal feasibility, dual feasibility and complementarity;

$$\text{KKT} = \left( \frac{\|Ax^* - b\|}{1 + \|b\|}, \frac{\|A^T y^* + s^* - c\|}{1 + \|c\|}, \frac{\langle x^*, s^* \rangle}{n} \right).$$

The headers ‘iter’ and ‘time’ in Table 4.1 refer to the average of the number of iterations and the running time in seconds, respectively.

		Non-Facially Reduced System	Facially Reduced System
linprog	KKT	(9.58e-16, 1.80e-12, 5.17e-09)	(5.78e-16, 1.51e-15, 5.57e-08)
	iter	23.30	17.60
	time	1.10	0.76
SDPT3	KKT	(1.51e-10, 1.49e-12, 4.67e-03)	(8.54e-12, 3.75e-16, 4.19e-06)
	iter	25.40	19.80
	time	0.82	0.53
MOSEK	KKT	(8.40e-09, 7.54e-16, -5.16e-06)	(5.16e-09, 3.81e-16, -2.03e-08)
	iter	35.90	10.10
	time	0.58	0.31

Table 4.1: Average of KKT conditions, iterations and time of (non)-facially reduced problems

From Table 4.1 we observe that facially reduced instances provide significant improvement in first order optimality conditions, the number of iterations and the running times for all solvers. We note that the instances  $(\mathcal{P}_{(A,b,c)})$  and  $(\mathcal{P}_{(A_{FR}, b_{FR}, c_{FR})})$  are equivalent. Hence, our empirics show that performing facial reduction as a preprocessing step not only improves the solver running time but also the *quality* of solutions.

### 4.1.4 Empirics on Distance to Infeasibility

In this section we present a numerical experiment that illustrates the affect of the perturbation imposed on the right-hand-side vector of the system  $\mathcal{F}$  when strict feasibility fails. We recall, from Proposition 3.12, that there exists an arbitrarily small perturbation of the right-hand-side vector

$b$  of  $\mathcal{F}$  that renders the set  $\mathcal{F}$  infeasible, i.e.,  $\text{dist}(b, \mathcal{F} = \emptyset) = 0$ . Moreover, the vector  $\Delta b = y$  that satisfies the auxiliary system (2.2) is a perturbation that makes the set  $\mathcal{F}$  empty; see (3.4).

We follow the steps in Section 4.1.1 to generate instances of the order  $(n, m) = (1000, 200)$  and  $r = \text{relint}(\mathcal{F}) = 900$ . The objective function  $c^T x$  is chosen as presented in Section 4.1.1. For the fixed  $(n, m, r)$ , we generate 10 instances and observe the average performance of these instances as we gradually increase the magnitude of the perturbation. We recall the matrix  $AV$  from (2.4). We use two types of perturbations for  $b$ ;

$$\Delta b, \text{ where } \Delta b \in \text{range}(AV)^\perp, \quad \Delta \bar{b}, \text{ where } \Delta \bar{b} \in \text{range}(AV).$$

We choose  $\Delta b$  to be the vector  $y$  that satisfies (2.2). For  $\Delta \bar{b}$ , we choose  $AVd$ , where  $d \in \mathbb{R}^r$  is a randomly chosen vector. As we increase  $\epsilon > 0$ , we observe the performance of the two families of the systems

$$Ax = b_\epsilon := b - \epsilon \Delta b \quad \text{and} \quad Ax = \bar{b}_\epsilon := b - \epsilon \Delta \bar{b}.$$

We use the interior point method from MATLAB's linprog for the test. Figure 4.2 contains the average of the first-order optimality conditions evaluated at the solver outputs  $(x^*, y^*, s^*)$  of these instances; primal feasibility, dual feasibility and the complementarity.

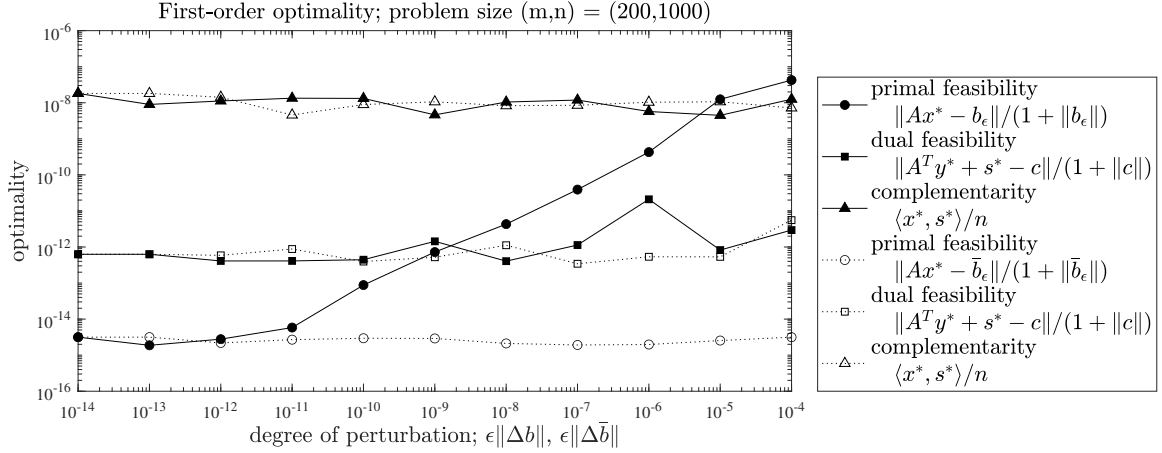


Figure 4.2: Changes in the first-order optimality condition as the perturbation of  $b$  increases

The horizontal axis of Figure 4.2 indicates the degree of the perturbation imposed on the right-hand-side vector  $b$ ,  $\epsilon \|\Delta b\|$  and  $\epsilon \|\Delta \bar{b}\|$ . The vertical axis indicates the individual component of the first-order optimality. From Figure 4.2, we observe that the KKT conditions with the perturbation  $\Delta \bar{b}$  display a steady performance regardless of the perturbation degree; see the markers  $\circ, \square, \triangle$  with the dotted lines. In contrast, the markers  $\bullet, \blacksquare, \blacktriangle$  in Figure 4.2 exhibit the performance of the instances that are perturbed with  $\Delta b$  and they display a different performance. In particular, we see that the relative primal feasibility  $\|Ax^* - b_\epsilon\|/(1 + \|b_\epsilon\|)$ , marked with  $\bullet$ , consistently increases as the perturbation magnitude  $\epsilon \|\Delta b\|$  increases when strict feasibility fails for  $\mathcal{F}$ .

## 4.2 Numerical Experiments with Simplex Method

In this section we compare the behaviour of the dual simplex method with instances that have strictly feasible points and instances that do not.



### 4.2.1 Generating Dual LPs without Strict Feasibility

We first show how to generate an instance for the dual feasible set  $\mathcal{G}$  that fails strict feasibility. The construction is similar to the one in Section 4.1.1. We generate a degenerate problem by finding a feasible auxiliary system (3.7). Given  $m, n, r \in \mathbb{N}$ , we construct  $A \in \mathbb{R}^{m \times n}$  and  $c \in \mathbb{R}^n$  that satisfy (3.7) with  $\dim(\text{relint}(\mathcal{G})) = m + r$ .

1. Pick any  $0 \neq w \in \mathbb{R}_+^n$  with  $|\text{supp}(w)| = n - r$ . Let

$$\{w\}^\perp = \text{span}\{a_i\}_{i=1}^{n-1} \quad (= \text{null}(w^T)).$$

We let the rows of the matrix  $A \in \mathbb{R}^{m \times n}$  consist of a random linear combination of the row vectors in the set  $\{a_i^T\}_{i=1}^{n-1}$ . We note that  $Aw = 0$ .

2. Pick  $s \in \mathbb{R}_+^n$  so that

$$s_i = \begin{cases} 0 & \text{if } i \in \text{supp}(w) \\ \text{positive} & \text{if } i \notin \text{supp}(w). \end{cases}$$

We note that  $\langle w, s \rangle = 0$  holds.

3. Pick  $y \in \mathbb{R}^m$  and set  $c = A^T y + s$ . We note that  $\langle c, w \rangle = 0$  holds.

For the empirics, we construct the objective function  $b^T y$  of  $(\mathcal{D})$  by choosing a vector  $\hat{x} \in \mathbb{R}_{++}^n$  and setting  $b = A\hat{x}$ .

### 4.2.2 Empirics on the Number of Degenerate Iterations

In this section we test how the lack of strict feasibility affects the performance of the dual simplex method. We choose MOSEK for our tests since MOSEK reports the percentage of degenerate iterations as a part of the solver report. MOSEK reports the quantity ‘DEGITER(%)’, the ratio of degenerate iterations.

Given a set  $\mathcal{G}$  and a point  $(y, s) \in \text{relint}(\mathcal{G}) \subseteq \mathbb{R}^m \oplus \mathbb{R}_+^n$ , let  $r$  be the number of positive entries of  $s$ , i.e.,  $r = |\text{supp}(s)|$ . In our tests, we gradually increase  $r$  for fixed  $n, m$  and generate instances for  $\mathcal{G}$  as described in Section 4.2.1. We then observe the behaviour of the dual simplex method. Table 4.2 contains the results. In Table 4.2, a smaller value for the header  $(r/n)\%$  means that there are more entries of  $s$  that are identically 0 in the set  $\mathcal{G}$ ; and the value 100% means that strict feasibility holds. For each triple  $(n, m, r)$ , we generated 10 instances and we report the average of ‘DEGITER(%)’ of these instances.

		$(r/n)\%$				
		60	70	80	90	100
$(n, m)$	(1000, 250)	36.62	10.18	0.01	0.02	0.00
	(2000, 500)	39.72	18.28	0.07	0.15	0.01
	(3000, 750)	25.99	10.66	0.32	0.75	0.02
	(4000, 1000)	29.78	18.25	0.25	0.53	0.02

Table 4.2: Average of the ratio of degenerate iterations

We recall Theorem 3.1: lack of strict feasibility implies that all basic feasible solutions are degenerate. However, we observe more, i.e., from Table 4.2, the frequency of degenerate iterations increases as  $r$  decreases. In other words, higher degeneracy of the set  $\mathcal{G}$  yields more degenerate iterations when the dual simplex method is used.



## 5 Conclusion

In this manuscript we addressed the importance of the strict feasibility constraint qualification for linear programming for both theoretical and numerical computations. For our numerics we illustrated this using the two most popular classes of algorithms, simplex and interior point methods. For the theory, we proved, using the two-step facial reduction, that if strict feasibility fails for a linear program, then every basic feasible solution is degenerate.

Strict feasibility is assumed for many algorithms for both the theory and implementation, in particular for interior point methods that claim polynomial time convergence. Strict feasibility is a good measure of the quality of a model. For example, for **SDP**, the absence of strict feasibility may lead to stability problems even for instances of small dimension, see e.g., [28]. Hence designing a model that has strict feasibility is important. However in cases where the absence of strict feasibility is inevitable (e.g., **SDP** relaxations of discrete optimization problems), facial reduction can be performed to regularize the model, e.g., see [19]. Typically, strict feasibility for **LPs** is less emphasized and many algorithms show strong numerical performances without this assumption. In this paper we showed that even for **LPs**, strict feasibility is a valuable property to guarantee.

An essential step for almost all algorithms for linear programming is preprocessing. This often transforms the right-hand-side vector  $b$  and we lose control over round-off errors introduced by finite-precision computing. As we have shown in Section 4.1.4, carefully chosen perturbations do not necessarily aggravate the primal feasibility; see Figure 4.2. However, it is difficult to confine the round-off errors accumulated to  $b$  to the range of  $AV$ ; see Proposition 3.12. As illustrated in Figure 4.2 and the nearness to infeasibility relationships to condition numbers Section 3.2.1, the primal feasibility worsens as the perturbations become larger. This further emphasizes that ensuring strict feasibility should be part of preprocessing for linear programming.

# Index

- $(\mathcal{P})$ , 3
- $A(:, \mathcal{I})$ , submatrix of  $A$  with columns in  $\mathcal{I}$ , 3
- $I$ , the identity matrix, 4
- $P_{\bar{m}} : \mathbb{R}^m \rightarrow \mathbb{R}^{\bar{m}}$ , 5
- $\bar{m} = \text{rank}(AV)$ , 5
- $\text{dist}(b, \mathcal{F} = \emptyset)$ , distance to infeasibility, 8
- $\mathbb{R}^{m \times n}$ , real vector space of  $m$ -by- $n$  matrices, 3
- $\mathbb{R}_+^n$ , nonnegative orthant, 3
- $\mathbb{R}_{++}^n$ , positive orthant, 3
- relint, relative interior, 3, 10
- supp, support, 6
- $p^*$ , 3
- $s_w$ , support of exposing vector for  $\mathcal{G}$ , 11
- $s_z$ , support of exposing vector for  $\mathcal{F}$ , 4
- $\mathcal{C}(\cdot)$ , condition measure, 9
- $\mathcal{F}$ , feasible region, 3
- $\mathcal{G}$ , dual feasible set, 10
- $(\mathcal{D})$ , dual of  $(\mathcal{P})$ , 10
- FR**, facial reduction, 2
- LP**, linear program, 2
  
- basic feasible solution, 3
- basic solution, 11
  
- condition measure,  $\mathcal{C}(\cdot)$ , 9
  
- degenerate, 3, 11
- degree of degeneracy, 8
- distance to infeasibility, 5, 8
- distance to infeasibility,  $\text{dist}(b, \mathcal{F} = \emptyset)$ , 8
- dual feasible set,  $\mathcal{G}$ , 10
- dual of  $(\mathcal{P})$ ,  $(\mathcal{D})$ , 10
  
- exposing vector, 4
- extreme point, 6
  
- face, 3, 6
- facial range vector, 4
- facial reduction, **FR**, 2, 4
- feasible region,  $\mathcal{F}$ , 3
  
- linear program, **LP**, 2
  
- minimal face, 3
  
- nondegenerate, 3
- nonnegative orthant,  $\mathbb{R}_+^n$ , 3
  
- performance profile, 13
- positive orthant,  $\mathbb{R}_{++}^n$ , 3
  
- real vector space of  $m$ -by- $n$  matrices,  $\mathbb{R}^{m \times n}$ , 3
- relative interior, relint, 3, 10
  
- singularity degree, 5
- Slater condition, 2
- stalling, 2
- support of exposing vector for  $\mathcal{F}$ ,  $s_z$ , 4
- support of exposing vector for  $\mathcal{G}$ ,  $s_w$ , 11
- support, supp, 6

## References

- [1] D. Bertsimas and J. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, Belmont, MA, 1997. [11](#)
- [2] R.E. Bixby. Solving real-world linear programs: a decade and more of progress. *Oper. Res.*, 50(1):3–15, 2002. 50th anniversary issue of Operations Research. [2](#)
- [3] R.G. Bland. New finite pivoting rules for the simplex method. *Math. Oper. Res.*, 2(2):103–107, 1977. [2](#)
- [4] J.M. Borwein and H. Wolkowicz. Facial reduction for a cone-convex programming problem. *J. Austral. Math. Soc. Ser. A*, 30(3):369–380, 1980/81. [4](#)
- [5] J.M. Borwein and H. Wolkowicz. Regularizing the abstract convex program. *J. Math. Anal. Appl.*, 83(2):495–530, 1981. [4](#)
- [6] A. Charnes. Optimality and degeneracy in linear programming. *Econometrica*, 20:160–170, 1952. [2](#)
- [7] Y.-L. Cheung. *Preprocessing and Reduction for Semidefinite Programming via Facial Reduction: Theory and Practice*. PhD thesis, University of Waterloo, 2013. [10](#)
- [8] Y-L. Cheung, S. Schurr, and H. Wolkowicz. Preprocessing and regularization for degenerate semidefinite programs. In D.H. Bailey, H.H. Bauschke, P. Borwein, F. Garvan, M. Thera, J. Vanderwerff, and H. Wolkowicz, editors, *Computational and Analytical Mathematics, In Honor of Jonathan Borwein’s 60th Birthday*, volume 50 of *Springer Proceedings in Mathematics & Statistics*, pages 225–276. Springer, 2013. [5](#)
- [9] G.B. Dantzig. *Linear Programming and Extensions*. Princeton University Press, Princeton, New Jersey, 1963. [2](#)
- [10] G.B. Dantzig, A. Orden, and P. Wolfe. The generalized simplex method for minimizing a linear form under linear inequality restraints. *Pacific J. Math.*, 5:183–195, 1955. [2](#)
- [11] E.D. Dolan and J.J. Moré. Benchmarking optimization software with performance profiles. *Math. Program.*, 91(2, Ser. A):201–213, 2002. [13](#)
- [12] D. Drusvyatskiy and H. Wolkowicz. The many faces of degeneracy in conic optimization. *Foundations and Trends<sup>®</sup> in Optimization*, 3(2):77–170, 2017. [3](#), [4](#), [5](#), [7](#), [10](#)
- [13] R.M. Freund and F. Ordonez. On an extension of condition number theory to nonconic convex optimization. *Mathematics of operations research*, 30(1):173–194, 2005. [8](#), [10](#)
- [14] R.M. Freund and J.R. Vera. Some characterizations and properties of the “distance to ill-posedness” and the condition measure of a conic linear system. Technical report, MIT, Cambridge, MA, 1997. [8](#), [10](#)
- [15] J. Gauvin. A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. *Mathematical programming*, 12(1):136–138, 1977. [10](#)
- [16] A.J. Goldman and A.W. Tucker. Theory of linear programming. In *Linear inequalities and related systems*, pages 53–97. Princeton University Press, Princeton, N.J., 1956. *Annals of Mathematics Studies*, no. 38. [2](#)

- [17] M. Gonzalez-Lima, H. Wei, and H. Wolkowicz. A stable primal-dual approach for linear programming under nondegeneracy assumptions. *Comput. Optim. Appl.*, 44(2):213–247, 2009. [2](#)
- [18] Nicholas Gould and Jennifer Scott. A note on performance profiles for benchmarking software. *ACM transactions on mathematical software*, 43(2):1–5, 2016. [13](#)
- [19] N. Graham, H. Hu, J. Im, X. Li, and H. Wolkowicz. A restricted dual peaceman-rachford splitting method for a strengthened dnn relaxation for qap. *INFORMS J. Comput.*, 2021. 29 pages, accepted Sept. 2021. [17](#)
- [20] O. Güler, D. Den Hertog, C. Roos, T. Terlaky, and T. Tsuchiya. Degeneracy in interior point methods for linear programming: a survey. *Ann. Oper. Res.*, 46/47(1-4):107–138, 1993. Degeneracy in optimization problems. [2](#), [11](#)
- [21] J. Im and H. Wolkowicz. A strengthened Barvinok-Pataki bound on SDP rank. *Oper. Res. Lett.*, 49(6):837–841, 2021. 11 pages, accepted Aug. 2021. [4](#), [7](#)
- [22] G. Pataki. On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Math. Oper. Res.*, 23(2):339–358, 1998. [6](#)
- [23] F.N. Permenter. *Reduction methods in semidefinite and conic optimization*. PhD thesis, Massachusetts Institute of Technology, 2017. [4](#), [7](#)
- [24] J. Renegar. Some perturbation theory for linear programming. *Math. Programming*, 65(1, Ser. A):73–91, 1994. [8](#)
- [25] J. Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM J. Optim.*, 5(3):506–524, 1995. [10](#)
- [26] D. M. Ryan and M. R. Osborne. On the solution of highly degenerate linear programmes. *Math. Program.*, 41:385–392, 1988. [2](#)
- [27] S. Sremac. *Error bounds and singularity degree in semidefinite programming*. PhD thesis, University of Waterloo, 2019. [4](#), [5](#), [7](#)
- [28] S. Sremac, H.J. Woerdeman, and H. Wolkowicz. Error bounds and singularity degree in semidefinite programming. *SIAM J. Optim.*, 31(1):812–836, 2021. [5](#), [17](#)
- [29] T. Terlaky and S. Zhang. Pivot rules for linear programming: A survey on recent theoretical developments. *Annals of operations research*, 46-47(1):203–233, 1993. [2](#)
- [30] S. Wright. *Primal-Dual Interior-Point Methods*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, 1996. [10](#), [11](#)