# New characterizations of Hoffman constants for systems of linear constraints

JAVIER F. PENA[1], JUAN C. VERA[2], AND LUIS F. ZULUAGA[3]

[1]Carnegie Mellon University

[2] Tilburg University

[3]Lehigh University

**LEHIGH**
U N I V E R S I T Y.

# New characterizations of Hoffman constants for systems of linear constraints

Javier Peña[*]        Juan Vera[†]        Luis F. Zuluaga[‡]

May 6, 2019

## Abstract

We give a characterization of the Hoffman constant of a system of linear constraints in $\mathbb{R}^n$ *relative* to a *reference polyhedron* $R \subseteq \mathbb{R}^n$. The reference polyhedron $R$ represents constraints that are easy to satisfy such as box constraints. In the special case $R = \mathbb{R}^n$, we obtain a novel characterization of the classical Hoffman constant.

More precisely, given a reference polyhedron $R \subseteq \mathbb{R}^n$ and $A \in \mathbb{R}^{m \times n}$, we characterize the sharpest constant $H(A|R)$ such that for all $b \in A(R) + \mathbb{R}^m_+$ and $u \in R$

$$\text{dist}(u, P_A(b) \cap R) \leq H(A|R) \cdot \|(Au - b)_+\|,$$

where $P_A(b) = \{x \in \mathbb{R}^n : Ax \leq b\}$. Our characterization is stated in terms of the largest of a canonical collection of easily computable Hoffman constants. Our characterization in turn suggests new algorithmic procedures to compute Hoffman constants.

## 1 Introduction

A classical result of Hoffman [15] shows that the distance from a point $u \in \mathbb{R}^n$ to a non-empty polyhedron $P_A(b) := \{x \in \mathbb{R}^n : Ax \leq b\}$ can be bounded above in terms of the size of the *residual* vector $(Au - b)_+ := \max(0, Au - b)$. More precisely, for $A \in \mathbb{R}^{m \times n}$ there exists a *Hoffman constant* $H(A)$ that depends only on $A$ such that for all $b \in A(\mathbb{R}^n) + \mathbb{R}^m_+$ and all $u \in \mathbb{R}^n$,

$$\text{dist}(u, P_A(b)) \leq H(A) \cdot \|(Au - b)_+\|. \tag{1}$$

Here $\text{dist}(u, P_A(b)) := \min\{\|u - x\| : x \in P_A(b)\}$. For convenience, we will make the following slight abuse of notation throughout the paper. We will write $\|\cdot\|$ to denote both the norm in $\mathbb{R}^n$ and the norm in $\mathbb{R}^m$. The specific norm will always be evident from the context. The bound (1) is a type of *error bound* for the system of inequalities $Ax \leq b$, that is, an inequality bounding the distance from a point $u \in \mathbb{R}^n$ to a nonempty *solution set* in terms of a measure of the *error* or *residual* of the point $u$ for the constraints defining the solution set.

---

[*]Tepper School of Business, Carnegie Mellon University, USA, `jfp@andrew.cmu.edu`

[†]Department of Econometrics and Operations Research, Tilburg University, The Netherlands, `j.c.veralizcano@uvt.nl`

[‡]Department of Industrial and Systems Engineering, Lehigh University, USA, `luis.zuluaga@lehigh.edu`

We consider the following more general *relative* version of (1). Suppose $R \subseteq \mathbb{R}^n$ is a nonempty *reference polyhedron* and $A \in \mathbb{R}^{m \times n}$. The reference polyhedron $R$ represents some constraints that are easy to satisfy such as box constraints. We give a characterization of the sharpest *relative Hoffman constant* $H(A|R)$ that depends only on $(A, R)$ such that for all $b \in A(R) + \mathbb{R}_+^m$ and all $u \in R$,

$$\operatorname{dist}(u, P_A(b) \cap R) \leq H(A|R) \cdot \|(Au - b)_+\|. \tag{2}$$

In the special case when $R = \mathbb{R}^n$ we have $H(A|R) = H(A)$ and obtain the following novel characterization of $H(A)$:

$$H(A) = \max_{J \in \mathcal{S}(A)} \frac{1}{\min\limits_{v \in \mathbb{R}_+^J, \|v\|^* = 1} \|A_J^\mathsf{T} v\|^*}, \tag{3}$$

where $\mathcal{S}(A)$ is the collection of subsets $J \subseteq \{1, \ldots, m\}$ such that $A_J \mathbb{R}^n + \mathbb{R}_+^J = \mathbb{R}^J$, $A_J$ is the submatrix of $A$ defined by the rows indexed by $J$, and $\| \cdot \|^*$ denotes the dual norm of $\| \cdot \|$. Observe that $A_J \mathbb{R}^n + \mathbb{R}_+^J = \mathbb{R}^J$ if and only if $A_J x < 0$ is feasible.

Hoffman bounds of the classical form (1), the relative form (2), and more general error bounds play a fundamental role in mathematical programming [29, 30, 51]. In particular, these kinds of Hoffman bounds as well as other related error bounds play a central role in establishing convergence properties of a variety of modern convex optimization algorithms [4, 11, 14, 19, 20, 26, 28, 33, 47]. Hoffman bounds are also used to measure the optimality and feasibility of a point generated by rounding an optimal point of the continuous relaxation of a mixed-integer linear or quadratic optimization problem [12, 40]. Furthermore, Hoffman bounds are used in sensitivity analysis [16], and to design solution methods for non-convex quadratic programs [48].

The relative format (2) that includes a reference polyhedron arises naturally in various contexts. For instance, it usually occurs when there are *box constraints* of the form $\ell \leq x \leq u$ as these constraints are generally easy to satisfy. Our interest in characterizing the Hoffman constant in the more general relative case that includes a reference polyhedron is motivated by the recent articles [4, 11, 14, 19, 33, 48]. In each of these articles, relative Hoffman constants for systems of linear constraints for suitable reference polyhedra play a central role in establishing key properties of modern optimization algorithms. In particular, the *facial distance* or *pyramidal width* introduced in [19, 33] is precisely a relative Hoffman constant with the standard simplex as reference polyhedron.

The paper makes the following main contributions. First, we develop novel characterizations of Hoffman constants for systems of linear inequalities. The characterization is stated as the largest of a canonical collection of easily computable Hoffman constants. Our characterization applies to the general case that includes both linear inequalities and linear equations, and a reference polyhedron representing constraints that are easy to satisfy. As a special case we obtain the new characterization (3) for the classical Hoffman constant $H(A)$. The Hoffman constant and our characterization of it also extend to the broader context of polyhedral set-valued mappings.

Second, throughout the paper we highlight the interesting and natural but mostly overlooked connection between the Hoffman constant and Renegar's distance to ill-posedness [36,

37], which is a cornerstone of condition measures in continuous optimization. Our connection is along the lines of some developments by Ramdas and Peña [35]. In particular, we detail the tight connection between the reciprocal of the Hoffman constant $1/H(A)$ and Renegar's distance to ill-posedness for the system of linear inequalities $Ax < 0$. We also discuss similar interesting connections for other Hoffman constants.

Third, we leverage our characterizations of the relative Hoffman constants $H(A)$ and $H(A|R)$ to develop novel algorithmic approaches to compute or estimate Hoffman constants. We should note that the exact or even approximate computation of the Hoffman constant is a notoriously difficult and largely unexplored computational challenge. The characterization (3) suggests that while any $J \in \mathcal{S}(A)$ yields a lower bound on $H(A)$, the typically more interesting task of computing a reasonable upper bound on $H(A)$ is far more challenging since the upper bound must hold for all $J \in \mathcal{S}(A)$. Aside from the algorithmic procedures proposed in this paper, there appears to be only one other documented method to compute the Hoffman constant, namely Algorithm `ALG 2` proposed by Klatte and Thiere in [18]. This algorithm is based on the following popular characterization of $H(A)$ from [13, 18, 47]

$$H(A) = \max_{\substack{J \subseteq \{1,\dots,m\} \\ A_J \text{ full row rank}}} \frac{1}{\min_{v \in \mathbb{R}_+^J, \|v\|^* = 1} \|A_J^\mathsf{T} v\|^*}. \tag{4}$$

The characterization (4) is often alluded to in the optimization literature as an expression for computing $H(A)$. Indeed, Algorithm `ALG 2` of Klatte and Thiere in [18] is based on this characterization. It performs an exhaustive search over all $J \subseteq \{1, \dots, m\}$ to evaluate (4) which evidently is viable only for very small values of $m$. A main limitation of (4) is that it does not take advantage of any structural properties of $A$. As we discuss in Section 3, it is possible to compute $H(A)$ via a variant of (3) that takes the maximum over a potentially much smaller collection of subsets $\mathcal{F} \subseteq \mathcal{S}(A)$. In the most favorable case when $A\mathbb{R}^n + \mathbb{R}_+^m = \mathbb{R}^m$ we can take $\mathcal{F} = \{\{1, \dots, m\}\}$ and thus

$$H(A) = \frac{1}{\min_{v \in \mathbb{R}_+^m, \|v\|^* = 1} \|A^\mathsf{T} v\|^*},$$

which can be computed via a single and fairly tractable convex optimization problem for suitable choices of norms.

The paper is entirely self-contained and relies only on standard convex optimization techniques. Our results are related to a number of previous developments in the rich literature on error bounds [2, 6, 13, 23, 27, 38, 45, 49] and on condition measures for continuous optimization [5, 7–10, 21, 31, 32, 36, 37]. In particular, the expressions for the Hoffman constants in Proposition 1 and Proposition 2 have appeared, albeit in slightly different form or under more restrictive conditions, in the work of Klatte and Thiere [18], Li [23], Robinson [38], and Wang and Lin [47]. More precisely, Klatte and Thiere [18] state and prove a version of Proposition 1 under the more restrictive assumption that $\mathbb{R}^n$ is endowed with the $\ell_2$ norm. Li [23], Robinson [38], and Wang and Lin [47] give characterizations of Hoffman constants that are equivalent to Proposition 1 and Proposition 2 but where the maximum is taken over a different, and typically much larger, collection of index sets. As we detail in Section 2, our

characterization for $H(A)$ in Proposition 1 and Proposition 2 implies (4) and can readily be seen to be at least as sharp as some bounds on $H(A)$ previously derived by Güler et al. [13], Burke and Tseng [6], and Zhang [50]. We also note that weaker versions of Theorem 1 can be obtained from results on error bounds in Asplund spaces as those developed in the article by Van Ngai and Théra [45]. Our goal to characterize a relative version of Hoffman constants that accounts for the presence of a reference polyhedron is partly inspired by the concepts of relative smoothness, relative strong convexity, and relative continuity recently developed and used by Bauschke et al [3], Lu [24], Lu et al [25], and Teboulle [43]. Our characterization of the Hoffman constants is in the spirit of and draws on the seminal work by Renegar [36, 37] as well as related work by Freund and Vera [9, 10], Peña [31, 32] and Lewis [21, 22].

The contents of the paper are organized as follows. Section 2 presents our main developments. We give a novel characterization of the classical Hoffman constant $H(A)$ (see Proposition 1, Proposition 2, and Corollary 1). We also extend this characterization to the more general case that includes linear inequalities, linear equations, and a reference polyhedron (see Proposition 5, Proposition 6, and Corollary 3). Section 3 describes several algorithmic procedures to compute Hoffman constants. Section 4, the most technical section of the paper, presents developments similar to those in Section 2 but in the broader context of polyhedral set-valued mappings (see Theorem 1, Theorem 2, and Corollary 4). Finally, Section 5 details how the results in Section 2 follow from those in Section 4.

Throughout the paper whenever we work with an Euclidean space $\mathbb{R}^d$, we will assume that it is endowed with a (non-necessarily Euclidean) norm $\| \cdot \|$ and inner product $\langle \cdot, \cdot \rangle$. We will often use the dual norm $\| \cdot \|^*$ of $\| \cdot \|$ defined as follows

$$\|u\|^* := \max_{\|x\| \leq 1} \langle u, x \rangle.$$

Unless we explicitly state otherwise, our results apply to arbitrary norms.

We will also rely on the following notation. Given a polyhedron $Q \subseteq \mathbb{R}^d$ let $\mathcal{T}(Q) := \{T_Q(u) : u \in Q\}$ where $T_Q(u)$ denotes the *tangent cone* to $Q$ at $u \in Q$, that is,

$$T_Q(u) = \{d \in \mathbb{R}^d : u + td \in Q \text{ for some } t > 0\}.$$

Observe that since $Q$ is assumed to be a polyhedron then the collection of tangent cones $\mathcal{T}(Q)$ is finite. Given a convex cone $K \subseteq \mathbb{R}^n$ we let $K^*$ denote its dual cone, that is,

$$K^* := \{u \in \mathbb{R}^n : \langle u, x \rangle \geq 0 \text{ for all } x \in K\}.$$

Throughout the paper we will write $[m]$ as shorthand for $\{1, \ldots, m\}$.

# 2 Hoffman constants for systems of linear constraints

This section describes a characterization for the Hoffman constant $H(A|R)$ in (2) for systems of linear inequalities

$$Ax \leq b$$
$$x \in R.$$

We subsequently consider analogous Hoffman constants for systems of linear inequalities and linear equations

$$Ax \leq b$$
$$Cx = d$$
$$x \in R.$$

Although the latter case with inequalities and equations subsumes the former case, for exposition purposes we discuss separately the case with inequalities only. Furthermore, we start with the special case $R = \mathbb{R}^n$. The notation and main ideas in this special case are simpler and easier to grasp. In particular, the crux of the characterization of $H(A)$ based on a canonical collection of submatrices of $A$ is more apparent.

We defer the proofs of the propositions in this section to Section 5, where we show that they follow from more general results for polyhedral set-valued mappings detailed in Section 4.

## 2.1 The case of inequalities only

Proposition 1 below gives a characterization of the *sharpest* Hoffman constant $H(A)$ such that (1) holds. The characterization is stated in terms of a canonical collection of submatrices of $A$.

Let $A \in \mathbb{R}^{m \times n}$. We next consider systems of linear inequalities of the form

$$Ax \leq b.$$

Let $\mathcal{S}(A) := \{J \subseteq [m] : A_J \mathbb{R}^n + \mathbb{R}_+^J = \mathbb{R}^J\}$ and

$$H(A) := \max_{J \in \mathcal{S}(A)} H_J(A) \tag{5}$$

where

$$H_J(A) := \max_{\substack{y \in \mathbb{R}^m \\ \|y\| \leq 1}} \min_{\substack{x \in \mathbb{R}^n \\ A_J x \leq y_J}} \|x\|$$

for each $J \in \mathcal{S}(A)$. By convention $H_J(A) = 0$ if $J = \emptyset$.

Throughout the sequel, for $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ we let $P_A(b)$ denote the polyhedron $\{x \in \mathbb{R}^n : Ax \leq b\}$.

**Proposition 1.** *Let $A \in \mathbb{R}^{m \times n}$. Then for all $b \in A(\mathbb{R}^n) + \mathbb{R}_+^m$ and all $u \in \mathbb{R}^n$*

$$\operatorname{dist}(u, P_A(b)) \leq H(A) \cdot \operatorname{dist}(b - Au, \mathbb{R}_+^m) \leq H(A) \cdot \|(Au - b)_+\|. \tag{6}$$

*Furthermore, the first bound in (6) is tight: If $H(A) > 0$ then there exist $b \in A\mathbb{R}^n + \mathbb{R}_+^m$ and $u \in \mathbb{R}^n$ such that*

$$\operatorname{dist}(u, P_A(b)) = H(A) \cdot \operatorname{dist}(b - Au, \mathbb{R}_+^m) > 0.$$

The following result complements Proposition 1 and yields an alternative expression for $H(A)$.

**Proposition 2.** *Let $A \in \mathbb{R}^{m \times n}$. Then for all $J \in \mathcal{S}(A)$*

$$H_J(A) = \max_{\substack{y \in \mathbb{R}^m \\ \|y\| \leq 1}} \min_{\substack{x \in \mathbb{R}^n \\ A_J x \leq y_J}} \|x\| = \frac{1}{\min_{v \in \mathbb{R}^J_+, \|v\|^* = 1} \|A_J^\mathsf{T} v\|^*}, \tag{7}$$

*with the convention that the last denominator is $+\infty$ and thus $H_J(A) = 0$ when $J = \emptyset$. In particular,*

$$H(A) = \max_{J \in \mathcal{S}(A)} \frac{1}{\min_{v \in \mathbb{R}^J_+, \|v\|^* = 1} \|A_J^\mathsf{T} v\|^*}. \tag{8}$$

Proposition 2 implies that

$$H(A) = \max_{J \in \mathcal{S}(A)} \max\{\|v\|^* : v \in \mathbb{R}^J_+, \|A_J^\mathsf{T} v\|^* \leq 1\}$$

$$= \max_{J \in \mathcal{S}(A)} \max\{\|\tilde{v}\|^* : \tilde{v} \in \text{ext}\{v \in \mathbb{R}^J_+, \|A_J^\mathsf{T} v\|^* \leq 1\}\},$$

where $\text{ext}(C)$ denotes the set of extreme points of a closed convex set $C$. Thus the following bound on $H(A)$ previously established in [6, 13] readily follows:

$$H(A) = \max_{J \in \mathcal{S}(A)} \max\{\|\tilde{v}\|^* : \tilde{v} \in \text{ext}\{v \in \mathbb{R}^J_+, \|A_J^\mathsf{T} v\|^* \leq 1\}\}$$

$$\leq \max\{\|\tilde{v}\|^* : \tilde{v} \in \text{ext}\{v \in \mathbb{R}^m_+, \|A^\mathsf{T} v\|^* \leq 1\}\}.$$

Furthermore, observe that if $J \in \mathcal{S}(A)$ and $\tilde{v} \in \text{ext}\{v \in \mathbb{R}^J_+, \|A_J^\mathsf{T} v\|^* \leq 1\}$ then $A_{J'}$ must have full row rank for $J' := \{i : \tilde{v}_i > 0\} \subseteq J$. Therefore Proposition 2 also implies that

$$H(A) = \max_{\substack{J \subseteq \{1,\dots,m\} \\ A_J \text{ full row rank}}} \max\{\|v\|^* : v \in \mathbb{R}^J_+, \|A_J^\mathsf{T} v\|^* \leq 1\} = \max_{\substack{J \subseteq \{1,\dots,m\} \\ A_J \text{ full row rank}}} \frac{1}{\min_{v \in \mathbb{R}^J_+, \|v\|^* = 1} \|A_J^\mathsf{T} v\|^*},$$

which is precisely the characterization (4) of $H(A)$. In addition, Proposition 2 implies the following bound on $H(A)$ in terms of the $\chi(A)$ condition measure [41, 44, 46] established in [50] for the special case when $A$ is full column rank and both $\mathbb{R}^m$ and $\mathbb{R}^n$ are endowed with Euclidean norms:

$$H(A) = \max_{\substack{J \subseteq \{1,\dots,m\} \\ A_J \text{ full row rank}}} \max\{\|v\| : v \in \mathbb{R}^J_+, \|A_J^\mathsf{T} v\| \leq 1\}$$

$$= \max_{\substack{J \subseteq [m], |J| = n \\ A_J \text{ non-singular}}} \max\{\|v\| : v \in \mathbb{R}^J_+, \|A_J^\mathsf{T} v\| \leq 1\}$$

$$\leq \max_{\substack{J \subseteq [m], |J| = n \\ A_J \text{non-singular}}} \max\{\|v\| : v \in \mathbb{R}^J, \|A_J^\mathsf{T} v\| \leq 1\}$$

$$= \max_{\substack{J \subseteq [m], |J| = n \\ A_J \text{non-singular}}} \|A_J^{-1}\|$$

$$= \chi(A).$$

The last step follows from [50, Prop. 3.7].

6

Observe that the above inequality could be fairly loose. For example if $A = \begin{bmatrix} 1 & -\epsilon \\ 1 & \epsilon \end{bmatrix}$ for some small $\epsilon > 0$, then $H(A) = 1 + O(\epsilon)$ whereas $\|A^{-1}\| = \Omega(1/\epsilon)$.

Proposition 2 also implies that the Hoffman constant $H(A)$ can be computed by maximizing over a potentially much smaller collection $\mathcal{F} \subseteq \mathcal{S}(A)$ as stated in equation (10) below. The expressions (8) and (10) for $H(A)$ are at the heart of one of the algorithmic procedures for computing $H(A)$ that we discuss in Section 3.

**Corollary 1.** *Let $A \in \mathbb{R}^{m \times n}$.*

**(a)** *If $A\mathbb{R}^n + \mathbb{R}^m_+ = \mathbb{R}^m$ then*

$$H(A) = \frac{1}{\min\limits_{v \in \mathbb{R}^m_+, \|v\|^* = 1} \|A^\mathsf{T} v\|^*}. \tag{9}$$

**(b)** *Suppose $\mathcal{F} \subseteq \mathcal{S}(A)$ and $\mathcal{I} \subseteq 2^{[m]} \setminus \mathcal{S}(A)$ are such that for all $J \subseteq [m]$ either $J \subseteq F$ for some $F \in \mathcal{F}$, or $I \subseteq J$ for some $I \in \mathcal{I}$. Then*

$$H(A) = \max_{J \in \mathcal{F}} H_J(A) = \max_{J \in \mathcal{F}} \frac{1}{\min\limits_{v \in \mathbb{R}^J_+, \|v\|^* = 1} \|A_J^\mathsf{T} v\|^*}. \tag{10}$$

*Proof.* **(a)** This follows from part (b) applied to $\mathcal{F} = \{[m]\}$ and $\mathcal{I} = \emptyset$.

**(b)** The conditions on $\mathcal{F}$ and $\mathcal{I}$ imply that for all $J \in \mathcal{S}(A)$ there exists $F \in \mathcal{F}$ such that $J \subseteq F$. The latter condition implies $H_J(A) \leq H_F(A)$. Therefore

$$H(A) = \max_{J \in \mathcal{S}(A)} H_J(A) = \max_{J \in \mathcal{F}} H_J(A) = \max_{J \in \mathcal{F}} \frac{1}{\min\limits_{v \in \mathbb{R}^J_+, \|v\|^* = 1} \|A_J^\mathsf{T} v\|^*}.$$

$\square$

The identity (9) in Corollary 1 has the following geometric interpretation. By Gordan's Theorem, $Ax < 0$ has a solution if and only if $A^\mathsf{T} v = 0, v \neq 0$ does not. Equivalently, $A\mathbb{R}^n + \mathbb{R}^m_+ = \mathbb{R}^m$ if and only if $0 \notin \{A^\mathsf{T} v : v \geq 0, \|v\|^* = 1\}$. When this is the case, the quantity $1/H(A)$ is precisely the distance (in the dual norm $\|\cdot\|^*$) from the origin to $\{A^\mathsf{T} v : v \geq 0, \|v\|^* = 1\}$. The latter quantity in turn equals the *distance to non-surjectivity* of the mapping $x \mapsto Ax + \mathbb{R}^m_+$, that is, the norm of the smallest perturbation matrix $\Delta A \in \mathbb{R}^{m \times n}$ such that $(A + \Delta A)\mathbb{R}^n + \mathbb{R}^m_+ \neq \mathbb{R}^m$ as it is detailed in [21]. This distance to non-surjectivity is the same as Renegar's *distance to ill-posedness* of the system of linear inequalities $Ax < 0$ defined by $A$. The more general identity (8) in Proposition 2 in turn can be interpreted as follows. The quantity $1/H(A)$ is the smallest distance to ill-posedness of the collection of the *feasible* systems of linear inequalities of the form $A_J x < 0$ for $J \subseteq [m]$.

The distance to ill-posedness provides the main building block for Renegar's concept of *condition number* for convex optimization introduced in the seminal papers [36, 37] that has been further extended in [1, 5, 7–10, 31, 32] among many other articles.

Proposition 1, Proposition 2, and Corollary 1 extend to the more general context when there is a reference polyhedron representing some constraints that are easy to satisfy. More precisely, let $R \subseteq \mathbb{R}^n$ be a reference polyhedron and $A \in \mathbb{R}^{m \times n}$. Consider systems of the following form

$$Ax \leq b$$
$$x \in R,$$

where $R$ represents a set of constraints that are easy to satisfy. It is natural to consider a refinement of the Hoffman constant $H(A)$ that reflects the presence of these easy-to-satisfy constraints. To that end, let $\mathcal{S}(A|R)$ and $H(A|R)$ be the extensions of $\mathcal{S}(A)$ and $H(A)$ defined as follows.

$$\mathcal{S}(A|R) := \{(J, K) : J \subseteq [m], K \in \mathcal{T}(R) \text{ and } A_J(K) + \mathbb{R}_+^J = \mathbb{R}^J\},$$

and

$$H(A|R) = \max_{(J,K) \in \mathcal{S}(A|R)} H_{J,K}(A),$$

where

$$H_{J,K}(A) := \max_{\substack{y \in \mathbb{R}^m \\ \|y\| \leq 1}} \min_{\substack{x \in K \\ A_J x \leq y_J}} \|x\|$$

Once again, by convention $H_{J,K}(A) = 0$ if $J = \emptyset$.

We have the following analogue of Proposition 1.

**Proposition 3.** *Let $A \in \mathbb{R}^{m \times n}$ and $R \subseteq \mathbb{R}^n$ be a reference polyhedron. Then for all $b \in A(R) + \mathbb{R}_+^m$ and $u \in R$*

$$\mathrm{dist}(u, P_A(b) \cap R) \leq H(A|R) \cdot \mathrm{dist}(b - Au, \mathbb{R}_+^m) \leq H(A|R) \cdot \|(Au - b)_+\|.$$

*Furthermore the first bound is tight: If $H(A|R) > 0$ then there exist $b \in A(R) + \mathbb{R}_+^m$ and $u \in R$ such that*

$$\mathrm{dist}(u, P_A(b) \cap R) = H(A|R) \cdot \mathrm{dist}(b - Au, \mathbb{R}_+^m) > 0.$$

We also have the following analogue of Proposition 2 that provides an alternative expression for $H(A|R)$.

**Proposition 4.** *Let $A \in \mathbb{R}^{m \times n}$ and $R \subseteq \mathbb{R}^n$ be a reference polyhedron. Then for all $(J, K) \in \mathcal{S}(A|R)$*

$$H_{J,K}(A) = \max_{\substack{y \in \mathbb{R}^m \\ \|y\| \leq 1}} \min_{\substack{x \in K \\ A_J x \leq y_J}} \|x\| = \frac{1}{\min_{\substack{v \in \mathbb{R}_+^J, \|v\|^* = 1 \\ A_J^\mathsf{T} v - u \in K^*}} \|u\|^*}. \tag{11}$$

*In particular,*

$$H(A|R) = \max_{(J,K) \in \mathcal{S}(A|R)} \frac{1}{\min_{\substack{v \in \mathbb{R}_+^J, \|v\|^* = 1 \\ A_J^\mathsf{T} v - u \in K^*}} \|u\|^*}. \tag{12}$$

We also have the following analogue of Corollary 1.

8

**Corollary 2.** *Let $A \in \mathbb{R}^{m \times n}$ and $R \subseteq \mathbb{R}^n$ be a reference polyhedron.*

**(a)** *If $R$ is a cone and $A(R) + \mathbb{R}^m_+ = \mathbb{R}^m$ then*

$$H(A|R) = \max_{\substack{y \in \mathbb{R}^m \\ \|y\| \leq 1}} \min_{\substack{x \in R \\ Ax \leq y}} \|x\| = \frac{1}{\min_{\substack{v \in \mathbb{R}^m_+, \|v\|^* = 1 \\ A^\mathsf{T} v - u \in R^*}} \|u\|^*}. \tag{13}$$

**(b)** *Suppose $\mathcal{F} \subseteq \mathcal{S}(A|R)$ and $\mathcal{I} \subseteq 2^{[m]} \times \mathcal{T}(R) \setminus \mathcal{S}(A|R)$ are such that for all $(J, K) \in 2^{[m]} \times \mathcal{T}(R)$ either $J \subseteq F$ and $K \subseteq T$ for some $(F, T) \in \mathcal{F}$, or $I \subseteq J$ and $U \subseteq K$ for some $(I, U) \in \mathcal{I}$. Then*

$$H(A|R) = \max_{(J,K) \in \mathcal{F}} \frac{1}{\min_{\substack{v \in \mathbb{R}^J_+, \|v\|^* = 1 \\ A^\mathsf{T} v - u \in K^*}} \|u\|^*}.$$

## 2.2 The case of inequalities and equations

Let $A \in \mathbb{R}^{m \times n}$, $C \in \mathbb{R}^{p \times n}$, and $R \subseteq \mathbb{R}^n$ be a reference polyhedron. Consider systems of the form

$$Ax \leq b$$
$$Cx = d$$
$$x \in R,$$

where $R$ represents some constraints that are easy to satisfy.

Proposition 5 below gives a bound analogous to (1) for the distance from a point $u \in R$ to a nonempty polyhedron of the form

$$\{x \in R : Ax \leq b, \, Cx = d\} = P_A(b) \cap C^{-1}(d) \cap R.$$

For $J \subseteq [m]$ let $[A, C; J] : \mathbb{R}^n \rightrightarrows \mathbb{R}^m \times \mathbb{R}^p$ be the set-valued mapping defined by

$$x \mapsto \{(Ax + s, Cx) : s \in \mathbb{R}^m, \, s_J \geq 0\}.$$

Let

$$\mathcal{S}(A, C|R) := \{(J, K) : J \subseteq [m], \, K \in \mathcal{T}(R), \, [A, C; J](K) \text{ is a linear subspace}\},$$

and

$$H(A, C|R) := \max_{(J,K) \in \mathcal{S}(A,C|R)} \max_{\substack{(y,w) \in \mathbb{R}^m \times C(K) \\ \|(y,w)\| \leq 1}} \min_{\substack{x \in K \\ A_J x \leq y_J, \, Cx = w}} \|x\|. \tag{14}$$

We have the following more general versions of Proposition 3, Proposition 4, and Corollary 2.

**Proposition 5.** *Let $R \subseteq \mathbb{R}^n$ be a reference polyhedron, $A \in \mathbb{R}^{m \times n}$, and $C \in \mathbb{R}^{p \times n}$. Then for all $(b, d) \in \{(Ax + s, Cx) : x \in R, s \in \mathbb{R}^m_+\}$ and $u \in R$*

$$\operatorname{dist}(u, P_A(b) \cap C^{-1}(d) \cap R) \leq H(A, C|R) \cdot \operatorname{dist}\left((b - Au, d - Cu), \mathbb{R}^m_+ \times \{0\}\right) \tag{15}$$

*and this bound is tight: If $H(A, C|R) > 0$ then there exist $(b, d) \in \{(Ax + s, Cx) : x \in R, s \in \mathbb{R}^m_+\}$ and $u \in R$ such that*

$$\operatorname{dist}(u, P_A(b) \cap C^{-1}(d) \cap R) = H(A, C|R) \cdot \operatorname{dist}\left((b - Au, d - Cu), \mathbb{R}^m_+ \times \{0\}\right) > 0.$$

**Proposition 6.** *Let $R \subseteq \mathbb{R}^n$ be a reference polyhedron, $A \in \mathbb{R}^{m \times n}$, and $C \in \mathbb{R}^{p \times n}$. Then for all $(J, K) \in \mathcal{S}(A, C | R)$*

$$\max_{\substack{(y,w) \in \mathbb{R}^m \times C(K) \\ \|(y,w)\| \leq 1}} \min_{\substack{x \in K \\ A_J x \leq y_J, Cx = w}} \|x\| = \frac{1}{\min_{\substack{v \in \mathbb{R}_+^J, z \in C(K) \\ \|(v,z)\|^* = 1, A_J^\mathsf{T} v + C^\mathsf{T} z - u \in K^*}} \|u\|^*}.$$

*In particular*

$$H(A, C | R) = \max_{(J,K) \in \mathcal{S}(A,C|R)} \frac{1}{\min_{\substack{v \in \mathbb{R}_+^J, z \in C(K) \\ \|(v,z)\|^* = 1, A_J^\mathsf{T} v + C^\mathsf{T} z - u \in K^*}} \|u\|^*}.$$

**Corollary 3.** *Let $R \subseteq \mathbb{R}^n$ be a reference polyhedron, $A \in \mathbb{R}^{m \times n}$, and $C \in \mathbb{R}^{p \times n}$.*

**(a)** *If $R$ is a cone and $\{(Ax + s, Cx) : x \in R, \ s \in \mathbb{R}_+^m\}$ is a linear subspace then*

$$H(A, C | R) = \frac{1}{\min_{\substack{v \in \mathbb{R}_+^m, z \in C(R) \\ \|(v,z)\|^* = 1, A^\mathsf{T} v + C^\mathsf{T} z - u \in R^*}} \|u\|^*}. \tag{16}$$

**(b)** *Suppose $\mathcal{F} \subseteq \mathcal{S}(A, C | R)$ and $\mathcal{I} \subseteq 2^{[m]} \times \mathcal{T}(R) \setminus \mathcal{S}(A, C | R)$ are such that for all $(J, K) \in 2^{[m]} \times \mathcal{T}(R)$ either $J \subseteq F$ and $K \subseteq T$ for some $(F, T) \in \mathcal{F}$, or $I \subseteq J$ and $U \subseteq K$ for some $(I, U) \in \mathcal{I}$. Then*

$$H(A, C | R) = \max_{(J,K) \in \mathcal{F}} \frac{1}{\min_{\substack{v \in \mathbb{R}_+^J, z \in C(K) \\ \|(v,z)\|^* = 1, A_J^\mathsf{T} v + C^\mathsf{T} z - u \in K^*}} \|u\|^*}.$$

The constant $H(A, C | R)$ generalizes the previous constants $H(A)$, $H(A|R)$. More precisely, by taking $p = 0$ and $C = [\,]$ the "empty" $0 \times n$ matrix, we get

$$H(A, [\,] | R) = H(A | R).$$

If $C = [\,]$ and $R = \mathbb{R}^n$ then we get $H(A, [\,] | \mathbb{R}^n) = H(A)$.

Therefore Proposition 1 and Proposition 3 are special cases of Proposition 5. Likewise, Proposition 2 and Proposition 4 are special cases of Proposition 6. We present the proofs of Proposition 5 and Proposition 6 in Section 5. They are immediate consequences of the more general Theorem 1 and Theorem 2 for polyhedral sublinear mappings.

Another special and particularly interesting case occurs when $m = 0$ and $A = [\,]$. This concerns systems of the form

$$Cx = d$$
$$x \in R,$$

where $R$ represents some polyhedral constraints that are easy to satisfy. In this case Proposition 5 implies that for all $d \in C(R)$ and $x \in R$

$$\mathrm{dist}(x, C^{-1}(d) \cap R) \leq \tilde{H}(C | R) \cdot \|d - Cx\|, \tag{17}$$

10

where

$$\tilde{H}(C \mid R) = H([\,], C \mid R) = \max_{K \in \tilde{\mathcal{S}}(C|R)} \max_{\substack{w \in C(K) \\ \|w\| \leq 1}} \min_{\substack{x \in K \\ Cx = w}} \|x\|,$$

and

$$\tilde{\mathcal{S}}(C \mid R) = \mathcal{S}([\,], C \mid R) = \{K \in \mathcal{T}(R) : C(K) \text{ is a linear subspace}\}.$$

Proposition 5 also implies that the bound (17) is tight. Furthermore, Proposition 6 yields

$$\tilde{H}(C|R) = \max_{K \in \tilde{\mathcal{S}}(C|R)} \frac{1}{\min\limits_{\substack{z \in C(K), \|z\|^* = 1 \\ C^\mathsf{T} z - u \in K^*}} \|u\|^*}.$$

In addition, Corollary 3 implies that if $R$ is a cone and $C(R)$ is a linear subspace then

$$\tilde{H}(C \mid R) = \max_{\substack{w \in C(R) \\ \|z\| \leq 1}} \min_{\substack{x \in R \\ Cx = w}} \|x\| = \frac{1}{\min\limits_{\substack{z \in C(R), \|z\|^* = 1 \\ C^\mathsf{T} z - u \in R^*}} \|u\|^*}. \tag{18}$$

In a nice analogy to (9) in Corollary 1, the identity (18) has the following geometric interpretation. If $R$ is a cone then $L := C(R)$ is a linear subspace if and only if $Cx \in \mathrm{relint}(R)$ is feasible and when this is the case (18) implies that

$$\frac{1}{\tilde{H}(C \mid R)} = \max\{r : w \in L, \|w\| \leq r \Rightarrow y \in C(\mathbb{B} \cap R)\},$$

where $\mathbb{B} := \{x \in \mathbb{R}^n : \|x\| \leq 1\}$. In other words, $1/\tilde{H}(C \mid R)$ is the radius of the largest ball in $L$ centered at the origin and contained in $L \cap C(\mathbb{B} \cap R)$. This radius can be seen as a generalization of the smallest singular value of $C$. Indeed, observe that when $R = \mathbb{R}^n$ and both $\mathbb{R}^n$ and $\mathbb{R}^m$ are endowed with Euclidean norms, $1/\tilde{H}(C \mid \mathbb{R}^n)$ is the smallest positive singular value of $C$. In the special case when $R$ is a cone and $C(R) = \mathbb{R}^m$, the quantity $1/\tilde{H}(C \mid R)$ equals the *distance to non-surjectivity* of the mapping

$$x \mapsto \begin{cases} Cx & \text{if } x \in R \\ \emptyset & \text{otherwise} \end{cases}$$

as detailed in [21]. This distance to non-surjectivity is the same as Renegar's *distance to ill-posedness* of the system of constraints $Cx = 0$, $x \in \mathrm{relint}(R)$ defined by $C$.

## 3 Computing Hoffman constants

We next describe some algorithmic approaches to compute Hoffman constants. For ease of exposition, we focus on the computation of $H(A)$ but the approaches described below can be extended to compute or estimate more general relative Hoffman constants $H(A, C|R)$.

We describe two main approaches to compute $H(A)$. The first approach is based on a formulation of $H(A)$ as a mathematical program with linear complementarity constraints (MPLCC). The MPLCC formulation in turn can be rewritten as a mixed integer linear program or as a linear program with special order set constraints of type 1. The second

11

approach is based on identifying collections of sets that satisfy a certain *covering property* based on Corollary 1.

Throughout this section we assume that $A \neq 0$ as otherwise the computation of $H(A)$ is uninteresting.

## 3.1 MPLCC formulation of $H(A)$

The next proposition shows that $H(A)$ can be formulated as an MPLCC.

**Proposition 7.** *Let* $A \in \mathbb{R}^{m \times n} \setminus \{0\}$. *Then*

$$
\begin{aligned}
\frac{1}{H(A)} = \min_{x,s,v} \ & \|A^{\mathsf{T}} v\|^* \\
\text{s.t. } & Ax - s \leq -\mathbf{1} \\
& \|v\|^* = 1 \\
& s_i v_i = 0, \ i = 1, \ldots, m \\
& s \geq 0, \ v \geq 0.
\end{aligned}
\tag{19}
$$

*Proof.* This is essentially a restatement of (8). Indeed, (8) can be rewritten as

$$
\frac{1}{H(A)} = \min_{J \in \mathcal{S}(A)} \min_{u \in \mathbb{R}^J_+, \|u\|^*=1} \|A_J^{\mathsf{T}} u\|^*
\tag{20}
$$

To establish the equivalence between (19) and (20), observe that the set of feasible points $(x, s, v)$ for (19) is in one-to-one correspondence with the set of pairs $(J, u)$ such that $J \in \mathcal{S}(A)$ and $u \in \mathbb{R}^J_+, \|u\|^* = 1$ via

$$
J = \{i \in [m] : s_i = 0\} \quad \text{and} \quad u = v_J.
$$

Under this correspondence we have $\|A^{\mathsf{T}} v\|^* = \|A_J^{\mathsf{T}} u\|^*$ and thus (19) is equivalent to (20). $\square$

For a suitable choice of norms, problem (19) can be cast as a linear program with linear complementarity constraints (LPLCC). Indeed, suppose $\mathbb{R}^n$ and $\mathbb{R}^m$ are endowed with the $\ell_1$-norm and $\ell_\infty$-norm respectively. Then $\|A^{\mathsf{T}} v\|^* = \|A^{\mathsf{T}} v\|_\infty$ and $\|v\|^* = \|v\|_1$ for all $v \in \mathbb{R}^m$ and hence for this choice of norms (19) is equivalent to

$$
\begin{aligned}
\frac{1}{H(A)} = \min_{x,v,z,t} \ & t \\
\text{s.t. } & -t\mathbf{1} \leq A^{\mathsf{T}} v \leq t\mathbf{1} \\
& Ax - s \leq -\mathbf{1} \\
& \mathbf{1}^{\mathsf{T}} v = 1 \\
& s_i v_i = 0, \ i = 1, \ldots, m \\
& s \geq 0, \ v \geq 0
\end{aligned}
\tag{21}
$$

LPLCC is a large and important class of problems that subsumes linear bilevel optimization and non-convex quadratic programming among others. There a variety of solution methods

12

for LPLCCs, many of them based on enumerative schemes. For a detailed review on this subject, see [17]. We next describe how (21) can also be formulated as a mixed integer linear program and as a linear program with special order set constraints of type 1.

By using big-M constraints, we can reformulate (21) as the following mixed integer linear program (MILP):

$$
\begin{aligned}
\min_{x,v,z,t} \ & t \\
\text{s.t. } & A^{\mathsf{T}}v \leq \mathbf{1}t \\
& A^{\mathsf{T}}v \geq -\mathbf{1}t \\
& Ax \leq -z + M(\mathbf{1} - z) \\
& \mathbf{1}^{\mathsf{T}}v = 1 \\
& 0 \leq v \leq z \\
& z_j \in \{0,1\}, \quad j = 1, \ldots, m.
\end{aligned}
\tag{22}
$$

A potential limitation of (22) is the need for an appropriate and valid estimate for the value of $M$. Modern MILP solvers provide some ways to overcome this limitation. First, state-of-the-art MILP solvers enable the alternate reformulation of the third and sixth constraints in (22) as the following set of *indicator* constraints (see [42, Chapter 26]):

$$
\begin{aligned}
z_j = 1 &\Rightarrow A_j x \leq -1, \quad j = 1, \ldots, m. \\
z_j &\in \{0,1\}, \quad\quad\quad\quad j = 1, \ldots, m.
\end{aligned}
\tag{23}
$$

Another alternative to big-M constraints is to use *special order set constraints of type 1* (SOS1) as discussed in [34, 39]. An SOS1 constraint is a set of variables in which at most one member can be strictly positive. Problem (21) can be reformulated as a the following linear program with SOS1 constraints:

$$
\begin{aligned}
\min_{x,v,t,s} \ & t \\
\text{s.t. } & -t\mathbf{1} \leq A^{\mathsf{T}}v \leq t\mathbf{1} \\
& Ax - s \leq -\mathbf{1} \\
& \mathbf{1}^{\mathsf{T}}v = 1 \\
& s \geq 0, \ v \geq 0 \\
& \{v_j, s_j\} \in \text{SOS1}, \quad j = 1, \ldots, m.
\end{aligned}
\tag{24}
$$

## 3.2 Computation of $H(A)$ via the *covering property*

Let $A \in \mathbb{R}^{m \times n} \setminus \{0\}$. Corollary 1 suggests the following algorithmic approach to compute $H(A)$. Find $\mathcal{F} \subseteq \mathcal{S}(A)$ and $\mathcal{I} \subseteq 2^{[m]} \setminus \mathcal{S}(A)$ that satisfy the following *covering property:*

For all $J \in 2^{[m]}$ either $J \subseteq F$ for some $F \in \mathcal{F}$ or $I \subseteq J$ for some $I \in \mathcal{I}$.

Then compute

$$
H(A) := \max_{J \in \mathcal{F}} \frac{1}{\min\{\|A_J^{\mathsf{T}}v\|^* : v \in \mathbb{R}_+^J, \|v\|^* = 1\}}.
\tag{25}
$$

We choose the term *covering property* since the above condition can be alternatively stated as follows: every element $J \in 2^{[m]}$ of the the ground set $2^{[m]}$ is either "covered" by some set in $\mathcal{F}$ or its complement $[m] \setminus J$ is "covered" by the complement of some set in $\mathcal{I}$. Observe that $(\mathcal{F}, \mathcal{I}) = (\overline{\mathcal{S}}(A), \underline{\mathcal{S}}(A))$ satisfies the covering property for the collection $\overline{\mathcal{S}}(A) \subseteq \mathcal{S}(A)$ of maximal (inclusion-wise) sets in $\mathcal{S}(A)$ and the collection $\underline{\mathcal{S}}(A) \subseteq 2^{[m]} \setminus \mathcal{S}(A)$ of minimal (inclusion-wise) sets in $2^{[m]} \setminus \mathcal{S}(A)$. Furthermore, it is easy to see that if $(\mathcal{F}, \mathcal{I})$ satisfies the covering property then $\overline{\mathcal{S}}(A) \subseteq \mathcal{F}$ and $\underline{\mathcal{S}}(A) \subseteq \mathcal{I}$. In other words, $(\overline{\mathcal{S}}(A), \underline{\mathcal{S}}(A))$ is the minimal pair of collections that satisfies the covering property.

The main challenge in computing $H(A)$ via (25) is the identification of suitable collections $\mathcal{F}$ and $\mathcal{I}$ that satisfy the covering property. In some special cases, it is possible to find $\overline{\mathcal{S}}(A)$ analytically and thus compute $H(A)$ via (25) with $\mathcal{F} = \overline{\mathcal{S}}(A)$. We illustrate this approach via some examples in Section 3.2.1. In Section 3.2.2 Algorithm 1 describes a general procedure that gradually constructs $\overline{\mathcal{S}}(A)$ and $\underline{\mathcal{S}}(A)$. We should note that Algorithm 1 is only viable when the collections $\overline{\mathcal{S}}(A)$ and $\underline{\mathcal{S}}(A)$ are of reasonable size since the algorithm constructs both of these collections explicitly.

The following observation facilitates the computation of $H(A)$. If $\mathbb{R}^n$ and $\mathbb{R}^m$ are endowed with the $\ell_1$-norm and $\ell_\infty$-norm respectively then

$$\min\{\|A_J^\mathsf{T} v\|^* : v \in \mathbb{R}_+^J, \|v\|^* = 1\} = \min\{\|A_J^\mathsf{T} v\|_\infty : v \in \mathbb{R}_+^J, \mathbf{1}^\mathsf{T} v = 1\}. \qquad (26)$$

The latter expression is computable via linear programming.

It is worthwhile noting that although the computation of $H(A)$ depends on the choices of norms, the covering property does not. In particular, if the pair $(\mathcal{F}, \mathcal{I})$ satisfies the covering property, then $H(A)$ can be computed or estimated for any choice of norms via (25) provided that $\min\{\|A_J^\mathsf{T} v\|^* : v \in \mathbb{R}_+^J, \|v\|^* = 1\}$ can be computed or estimated. We also note that if the pair $(\mathcal{F}, \mathcal{I})$ satisfies the covering property then it provides a *certificate of optimality* for $H(A)$ since it certifies that $H_J(A) \leq H(A)$ for any $J \in \mathcal{S}(A)$. By contrast, the MPLCC and MILP approaches described in Section 3.1 do not readily provide such a certificate of optimality for $H(A)$.

### 3.2.1 Some examples

As the following examples illustrate, when the matrix $A$ is highly structured it may be possible to construct $\mathcal{F}$ and $\mathcal{I}$ directly so that the covering property holds. In the examples below we actually identify the collection $\overline{\mathcal{S}}(A)$ of maximal sets in $\mathcal{S}(A)$. Once $\overline{\mathcal{S}}(A)$ is identified, we can compute $H(A)$ via (25) with $\mathcal{F} = \overline{\mathcal{S}}(A)$. To facilitate the latter computation, in the next three examples we assume that $\mathbb{R}^n$ and $\mathbb{R}^m$ are endowed with the $\ell_1$-norm and $\ell_\infty$-norm respectively so that (26) holds.

**Example 1** (box). *Let $n \geq 1$ and consider the matrix*

$$A = \begin{bmatrix} \mathsf{I}_n \\ -\mathsf{I}_n \end{bmatrix} \in \mathbb{R}^{2n \times n}$$

*where $\mathsf{I}_n \in \mathbb{R}^{n \times n}$ denotes the $n \times n$ identity matrix. In this case the collection $\overline{\mathcal{S}}(A)$ consists of the $2^n$ sets of the form $\{i_1, i_2, \ldots, i_n\}$ where $i_k \in \{k, k+n\}$, $k = 1, \ldots, n$. A straightforward*

*calculation then shows that*

$$H(A) = \max_{J \in \overline{\mathcal{S}}(A)} \frac{1}{\min\limits_{v \in \mathbb{R}_+^J, \mathbf{1}^\mathsf{T} v = 1} \|A_J^\mathsf{T} v\|_\infty} = \frac{1}{\min\limits_{v \in \mathbb{R}_+^n, \mathbf{1}^\mathsf{T} v = 1} \|v\|_\infty} = n.$$

**Example 2** (simplex). *Let $n \geq 1$ and consider the matrix*

$$A = \begin{bmatrix} \mathsf{I}_n \\ -\mathbf{1}^\mathsf{T} \end{bmatrix} \in \mathbb{R}^{(n+1) \times n}.$$

*In this case the collection $\overline{\mathcal{S}}(A)$ consists of the $n+1$ sets $\{1, \ldots, n\}$ and $\{1, \ldots, n, n+1\} \setminus \{k\}$, $k = 1, \ldots, n$. A straightforward calculation then shows that*

$$\begin{aligned} H(A) &= \max \left( \frac{1}{\min\limits_{v \in \mathbb{R}_+^n, \mathbf{1}^\mathsf{T} v = 1} \|v\|_\infty}, \frac{1}{\min\limits_{(v,v_n) \in \mathbb{R}_+^n, \mathbf{1}^\mathsf{T} v + v_n = 1} \|(v - v_n \mathbf{1}, -v_n)\|_\infty} \right) \\ &= \max(n, 2n-1) \\ &= 2n-1. \end{aligned}$$

**Example 3** ($\ell_1$-unit ball). *Let $n \geq 1$ and consider the matrix $A \in \mathbb{R}^{2^n \times n}$ whose rows are the $2^n$ vectors in $\{-1, 1\}^n$ ordered lexicographically, that is,*

$$A = \begin{bmatrix} -1 & -1 & \cdots & -1 & -1 \\ -1 & -1 & \cdots & -1 & 1 \\ -1 & -1 & \cdots & 1 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & -1 \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix}.$$

*In this case the collection $\overline{\mathcal{S}}(A)$ consists of the collection of all sets of the form*

$$J_u = \{i : A_i u < 0\}$$

*for $u \in \mathbb{R}^n$ such that all components of $Au$ are non-zero. The symmetry of $A$ implies that*

$$H(A) = \max_{J \in \overline{\mathcal{S}}(A)} \frac{1}{\min\limits_{v \in \mathbb{R}_+^J, \mathbf{1}^\mathsf{T} v = 1} \|A_J^\mathsf{T} v\|_\infty} = \max_{J \in \mathcal{S}_0} \frac{1}{\min\limits_{v \in \mathbb{R}_+^J, \mathbf{1}^\mathsf{T} v = 1} \|A_J^\mathsf{T} v\|_\infty}, \tag{27}$$

*where $\mathcal{S}_0 \subseteq \mathcal{F}$ is the smaller collection of sets of the form $J_u$ where $u \in \mathbb{R}^n$ has non-increasing entries and all entries of $Au$ are non-zero. For small values of $n$, both $\mathcal{S}_0$ and $H(A)$ can be computed explicitly. The following table displays the values of $H(A)$ and the optimal index set $J \in \mathcal{S}_0$ where (27) attains its maximum for $n = 1, 2, 3, 4, 5$. To ease notation, the $J$ entry in each column only displays the new indices that need to be added to the $J$ entry in the previous column which is denoted by '$\cdots$':*

| $n$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $H(A)$ | 1 | 1 | 3 | 5 | 9 |
| $J$ | $\{1\}$ | $\cdots \cup \{2\}$ | $\cdots \cup \{3,5\}$ | $\cdots \cup \{4,6,7,9\}$ | $\cdots \cup \{8,10,11,12,13,17,18,19\}$ |

15

The values of the Hoffman constant $H(A)$ in Example 1, Example 2, and Example 3 can also be obtained via the MPLCC formulations discussed in Section 3.1 above or via Algorithm 1 below.

### 3.2.2  An algorithm that constructs $\mathcal{F}$ and $\mathcal{I}$ gradually

Algorithm 1 below formalizes the following simple iterative procedure to construct a pair $(\mathcal{F}, \mathcal{I})$ that satisfies the covering property: Start with $\mathcal{F} = \mathcal{I} = \emptyset$. At each subsequent iteration check whether $(\mathcal{F}, \mathcal{I})$ covers $2^{[m]}$. It is does, then we are done. Otherwise, find $J \in 2^{[m]}$ that is not covered by $(\mathcal{F}, \mathcal{I})$ and check whether $J \in \mathcal{S}(A)$. If indeed $J \in \mathcal{S}(A)$ then add $J$ to $\mathcal{F}$. Otherwise, find $I \subseteq J$ such that $I \notin \mathcal{S}(A)$ and add $I$ to $\mathcal{I}$. This procedure must eventually terminate since each iteration adds a new element to $\mathcal{F}$ or to $\mathcal{I}$. Furthermore, as Proposition 8 below shows, if this procedure is properly executed, it constructs the minimal collections $\mathcal{F} = \overline{\mathcal{S}}(A)$ and $\mathcal{I} = \underline{\mathcal{S}}(A)$ that satisfy the covering property.

The algorithm needs to perform two main steps. First, given $\mathcal{F} \subseteq \mathcal{S}(A)$ and $\mathcal{I} \subseteq 2^{[m]} \setminus \mathcal{S}(A)$, find $J \in 2^{[m]}$ not covered by $(\mathcal{F}, \mathcal{I})$ or verify that no such $J$ exists. Second, given $J \in 2^{[m]}$, either certify that $J \in \mathcal{S}(A)$ or else find $I \subseteq J$ with $I \in 2^{[m]} \setminus \mathcal{S}(A)$.

The first step can be accomplished by solving the following combinatorial optimization problem

$$
\begin{aligned}
\max_{J} \quad & |J| \\
\text{s.t.} \quad & |J^c \cap I| \geq 1, \ I \in \mathcal{I} \\
& |J \cap F^c| \geq 1, \ F \in \mathcal{F} \\
& J \subseteq [m].
\end{aligned}
\tag{28}
$$

Observe that $(\mathcal{F}, \mathcal{I})$ satisfy the covering property if and only if (28) is infeasible. Otherwise the optimal solution of (28) yields $J \in 2^{\{1,\dots,m\}}$ of maximal size that is not covered by $(\mathcal{F}, \mathcal{I})$.

The second step can be accomplished by solving the following optimization problem

$$
\min\{\|A_J^\mathsf{T} v\|^* : v \in \mathbb{R}_+^J, \|v\|^* = 1\}.
\tag{29}
$$

Observe that $J \in \mathcal{S}(A)$ if and only if the optimal value of (29) is positive. Otherwise, when the optimal value of (29) is zero, its optimal solution satisfies $v \in \mathbb{R}_+^J \setminus \{0\}$ and $A_J^\mathsf{T} v = 0$. In this case $I(v) := \{i \in J : v_i > 0\} \in 2^{[m]} \setminus \mathcal{S}(A)$ and $I(v) \subseteq J$. Furthermore, for additional efficiency we will assume that in the latter case $v$ is chosen so that $I(v)$ is of minimal size. This condition is easily enforceable by applying a straightforward post-processing procedure whenever the optimal value of (29) is zero. As we noted in (26), when $\mathbb{R}^n$ and $\mathbb{R}^m$ are endowed with the $\ell_1$-norm and $\ell_\infty$-norm respectively problem (29) can be rewritten as a linear program.

It is easy to see that if $[m] \in \mathcal{S}(A)$ then Algorithm 1 terminates after one iteration: the first iteration places $[m]$ in $\mathcal{F}$ and leaves $\mathcal{I}$ empty. At that point the collections $\mathcal{F} = \overline{\mathcal{S}}(A) = \{[m]\}$ and $\mathcal{I} = \underline{\mathcal{S}}(A) = \emptyset$ satisfy the covering property. Similarly when $[m]$ is the only set missing from $\mathcal{S}(A)$ Algorithm 1 terminates after $1 + m$ iterations: the first iteration places the set $[m]$ in $\mathcal{I}$ and the next $m$ iterations place the sets $[m] \setminus \{j\}$ for $j = 1, \dots, m$ in $\mathcal{F}$. At that point the collections $\mathcal{F} = \overline{\mathcal{S}}(A) = \{[m] \setminus \{j\} : j = 1, \dots, m\}$ and $\mathcal{I} = \underline{\mathcal{S}}(A) = \{[m]\}$ satisfy the covering property.

---
**Algorithm 1** Computation of a pair $(\mathcal{F}, \mathcal{I})$ satisfying the covering property and $H(A)$
---
1: **input** $A \in \mathbb{R}^{m \times n} \setminus \{0\}$
2: Let $\mathcal{F} := \emptyset$, $\mathcal{I} := \emptyset$, $H(A) := 0$
3: **while** $(\mathcal{F}, \mathcal{I})$ does not satisfy the covering property **do**
4:      Solve (28) to pick $J \in 2^{[m]}$ not covered by $(\mathcal{F}, \mathcal{I})$
5:      Let $v$ solve (29) to detect whether $J \in \mathcal{S}(A)$
6:      **if** $\|A_J^\mathsf{T} v\|^* > 0$ **then**
7:          $\mathcal{F} := \mathcal{F} \cup \{J\}$ and $H(A) := \max\left\{H(A), \frac{1}{\|A_J^\mathsf{T} v\|^*}\right\}$
8:      **else**
9:          Let $\mathcal{I} := \mathcal{I} \cup \{I(v)\}$
10:     **end if**
11: **end while**
12: **return** $\mathcal{F}, \mathcal{I}, H(A)$
---

For general $A \in \mathbb{R}^{m \times n} \setminus \{0\}$, Proposition 8 extends the above two special cases. We should note that, as the above two cases illustrate, the expression $|\overline{\mathcal{S}}(A)| + |\underline{\mathcal{S}}(A)|$ in (30) below is always smaller, and in some cases vastly smaller, than $2^m$ which is the number of steps that a brute force scan of the subsets of $[m]$ would require.

**Proposition 8.** *Let $A \in \mathbb{R}^{m \times n} \setminus \{0\}$ and let $\overline{\mathcal{S}}(A) \subseteq \mathcal{S}(A)$ denote the collection of maximal (inclusion-wise) sets in $\mathcal{S}(A)$ and let $\underline{\mathcal{S}}(A) \subseteq 2^{[m]} \setminus \mathcal{S}(A)$ denote the collection of minimal (inclusion-wise) sets in $2^{[m]} \setminus \mathcal{S}(A)$. Then Algorithm 1 terminates after*

$$|\overline{\mathcal{S}}(A)| + |\underline{\mathcal{S}}(A)| \tag{30}$$

*iterations. Furthermore, upon termination Algorithm 1 returns $\mathcal{F} = \overline{\mathcal{S}}(A)$ and $\mathcal{I} = \underline{\mathcal{S}}(A)$.*

*Proof.* Each iteration of Algorithm 1 either adds one element $J \in \overline{\mathcal{S}}(A)$ to $\mathcal{F}$ because the set $J$ selected in Step 4 is of maximal size, or adds one element from $I(v) \in \underline{\mathcal{S}}(A)$ to $\mathcal{I}$ because the set $I(v)$ selected in Step 9 is of minimal size. To finish, observe that Algorithm 1 terminates as soon as $\mathcal{F} = \overline{\mathcal{S}}(A)$ and $\mathcal{I} = \underline{\mathcal{S}}(A)$.

                □

The expression (30) can be interpreted as follows. Suppose $\overline{\mathcal{S}}(A) = \{J_1, \ldots, J_k\}$. Then $J \in 2^{[m]} \setminus \mathcal{S}(A)$ if and only if $J \setminus J_i \neq \emptyset$ for $i = 1, \ldots, k$. Therefore $|\underline{\mathcal{S}}(A)|$ is the number of minimal (componentwise) solutions $(x_1, \ldots, x_m) \in \{0, 1\}^m$ to the following system of set-covering constraints

$$\sum_{i \notin J_\ell} x_i \geq 1 \quad \text{for } \ell = 1, \ldots, k$$
$$x_i \in \{0, 1\} \quad \text{for } i = 1, \ldots, m.$$

Hence $|\overline{\mathcal{S}}(A)| + |\underline{\mathcal{S}}(A)|$ will not be too large if the sets in $\overline{\mathcal{S}}(A)$ are few and large. The most favorable case $|\overline{\mathcal{S}}(A)| + |\underline{\mathcal{S}}(A)| = 1$ occurs when $\overline{\mathcal{S}}(A) = \{[m]\}$ or equivalently when $Ax < 0$ is feasible. The next most favorable case $|\overline{\mathcal{S}}(A)| + |\underline{\mathcal{S}}(A)| = 1 + m$ occurs when $\underline{\mathcal{S}}(A) = \{[m]\}$ or equivalently when $Ax < 0$ is infeasible but it becomes feasible as soon as we drop one inequality.

## 3.3   Numerical experiments

Table 1 and Table 2 summarize two sets of numerical experiments that illustrate and compare three different methods to compute $H(A)$: the SOS1 formulation (24), Algorithm 1 based on the covering property, and an enumeration scheme based on (4) via a scan of the subsets $J \subseteq [m]$ such that $A_J$ is full row rank. We will refer to these three methods as SOS1, COVER, and ENUM respectively. To make the implementation of ENUM more efficient, we used a variant of Algorithm 1 that scans only for maximal subsets $J \subseteq [m]$ such that $A_J$ has full row rank. In particular, the number of sets that ENUM needs to scan is bounded above by $\binom{m}{n} + \binom{m}{n-1} + \cdots + \binom{m}{1}$.. All experiments were carried out in an iMac with a 3.5GHz Intel Core i7 and 32 GB 1600 MHz DDR3 RAM, using `Matlab R2017b` and `Gurobi 8.1.1`.

Table 1 reports results on a set of instances $A \in \mathbb{R}^{m \times n}$ with $n \leq m \leq 100$ drawn from `LPnetlib, CUTEr,` and `Globallib` test sets. For each instance we display the value $H(A)$ and the following number of "main steps" performed by each method. For SOS1, we display the number of nodes (including the root node) that `Gurobi` generated to solve (24). For COVER, we display the number of main iterations (iters.). For ENUM we display the number of sets $J \subseteq [m]$ that had to be scanned. The main task required at each main step in each method is the solution of a linear program. Thus the number of main steps give a rough comparison among the three methods. An entry '∗' in Table 1 indicates that the method did not solve the corresponding instances within 10,000 main steps.

Although we ran over 150 instances, we only display results for the more challenging instances where either SOS1 or COVER require more than one main step. The entries in Table 1 (which is ordered by the number of steps taken by the SOS1 method) reveal that many of the instances from the `LPnetlib, CUTEr,` and `Globallib` are easily solved by SOS1 and COVER whereas only the very smallest ones are solved by ENUM. The easiest instances (about 120 total, not resported in the Table) are those with $\underline{\mathcal{S}}(A) = \{[m]\}$ which COVER solves in one step and SOS1 generally solves in one or two steps. By contrast, ENUM cannot solve most of these instances within 10,000 steps. The first, fourth, and fifth instances in Table 1 illustrate this behavior. For other more challenging instances, COVER usually finds $H(A)$ after fewer main steps than SOS1. However, the SOS1 "steps" (nodes) are much faster (up to an order of magnitude) than the COVER iterations. This is not surprising since we used `Gurobi`, a state-of-the-art mathematical programming solver, in our SOS1 implementation. By contrast, we used fairly straightforward and simple `MATLAB` implementations for both COVER and ENUM.

Table 2 reports results on instances of $A$ generated as in Example 1 and Example 3. For each of these two examples we included two instances. The first one corresponds to the largest $n$ such that the instance was solvable by all three methods within 10,000 steps. The second one corresponds to the very next value of $n$. The general conclusions of these four experiments is that ENUM appears to be the least effective method, while the relationship between the efficiency of SOS1 and COVER is not always the same. It is noteworthy that Example 3 poses a significant challenge even for SOS1 when $n = 5$ thereby highlighting the formidable challenge that computing $H(A)$ entails. We also tested the methods for instances of $A$ generated as in Example 2 for $n$ up to 1000. Not surprisingly, all three methods solve those instances quite easily: SOS1 usually generates only one node beyond the root node, COVER terminates after $n + 2$ iterations, and ENUM terminates after scanning $n + 1$ sets.

| Instance | $m$ | $n$ | $H(A)$ | SOS1 | COVER | ENUM |
|---|---|---|---|---|---|---|
| qp50_25_3_3.mat | 50 | 25 | 2.7203 | 2 | 1 | * |
| st_bpv2.mat | 5 | 4 | 2.0000 | 2 | 3 | 3 |
| st_iqpbk1.mat | 8 | 7 | 28.0000 | 2 | 9 | 8 |
| st_fp5 | 11 | 10 | 3.2286 | 6 | 3 | 3 |
| qp50_25_1_3 | 50 | 25 | 3.5817 | 8 | 1 | * |
| qp40_20_4_2 | 40 | 20 | 10.4870 | 10 | 1 | * |
| st_e22 | 5 | 2 | 11.0000 | 26 | 10 | 10 |
| st_glmp_kk92 | 6 | 4 | 4.0000 | 30 | 11 | 11 |
| st_glmp_fp2 | 7 | 4 | 134.3334 | 38 | 18 | 22 |
| st_e25 | 8 | 4 | 173.2708 | 39 | 13 | 70 |
| qp20_10_4_2 | 20 | 10 | 36.2748 | 40 | 6 | * |
| qp30_15_1_1 | 30 | 15 | 3411.7169 | 40 | 8 | * |
| st_glmp_ss1 | 8 | 5 | 9.0000 | 58 | 20 | 20 |
| st_glmp_kky | 8 | 7 | 3.1429 | 64 | 20 | 20 |
| qp20_10_3_1 | 20 | 10 | 49.7414 | 88 | 15 | * |
| st_ph20 | 9 | 3 | 28.0000 | 108 | 33 | 84 |
| st_ph13 | 10 | 3 | 6.0000 | 118 | 15 | 99 |
| st_qpk2 | 12 | 6 | 2.0000 | 248 | 70 | 70 |
| biggsc4 | 13 | 4 | 5.0000 | 420 | 61 | 385 |
| qp1 | 50 | 1 | 483.3333 | 1026 | 501 | 50 |
| qp3 | 100 | 1 | 483.3333 | 1026 | 551 | 100 |
| st_qpk3 | 22 | 11 | 1.8196 | 7046 | 2059 | 2059 |
| lpi_woodinfe | 89 | 35 | 57.0006 | 8005 | * | * |
| qp30_15_3_1 | 30 | 15 | 104528.1035 | * | 371 | * |

∗: Algorithm reached limit on the number of steps.

Table 1: Performance of SOS1, COVER, and ENUM on test instances from `LPnetlib`, `CUTEr`, `Globallib`, with a limit of 10,000 main steps.

| Instance | $m$ | $n$ | $H(A)$ | SOS1 | COVER | ENUM |
|---|---|---|---|---|---|---|
| Example 1 | 26 | 13 | 13.0000 | 24 | 8205 | 8205 |
| Example 1 | 28 | 14 | 14.0000 | 24 | * | * |
| Example 3 | 16 | 4 | 5.0000 | 1552 | 152 | 1128 |
| Example 3 | 32 | 5 | 9.0000 | * | 4594 | * |

Table 2: Performance of SOS1, COVER, and ENUM on instances of $A$ generated as in Example 1 and Example 3.

We reiterate that the above results are based on entirely straightforward and simple implementations of SOS1, COVER, and ENUM. In particular, our basic implementations do not use warm-starts and do not make any attempt to exploit the symmetry structure that is

evident in Example 1, Example 2, and Example 3. The development of more sophisticated implementations of the above algorithmic schemes will be an interesting topic for future work.

# 4 A Hoffman constant for polyhedral set-valued mappings

We next present a characterization of the Hoffman constant for polyhedral set-valued mappings. Recall that a set-valued mapping $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ assigns a set $\Phi(x) \subseteq \mathbb{R}^m$ to each $x \in \mathbb{R}^n$. Let $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping. The inverse $\Phi^{-1} : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ of $\Phi$ is the set-valued mapping defined in the following natural way

$$x \in \Phi^{-1}(y) \text{ if and only if } y \in \Phi(x).$$

We will say that $\Phi$ is *polyhedral* if

$$\text{graph}(\Phi) := \{(x, y) : y \in \Phi(x)\} \subseteq \mathbb{R}^n \times \mathbb{R}^m$$

is a polyhedron. We will say that $\Phi$ is *sublinear* if graph($\Phi$) is a convex cone, and we will say that $\Phi$ is *closed* if graph($\Phi$) is closed. In particular, we will say that $\Phi$ is a *polyhedral sublinear mapping* if graph($\Phi$) is a polyhedral convex cone.

Let $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping. The domain and image of $\Phi$ are defined as follows:

$$\text{dom}(\Phi) = \{x \in \mathbb{R}^n : (x, y) \in \text{graph}(\Phi) \text{ for some } y \in \mathbb{R}^m\},$$
$$\text{Im}(\Phi) = \{y \in \mathbb{R}^m : (x, y) \in \text{graph}(\Phi) \text{ for some } x \in \mathbb{R}^n\}.$$

When $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a sublinear mapping, the norm of $\Phi$ is defined as follows

$$\|\Phi\| = \max_{\substack{x \in \text{dom}(\Phi) \\ \|x\| \le 1}} \min_{y \in \Phi(x)} \|y\|.$$

In particular, if $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a sublinear mapping then the norm of $\Phi^{-1} : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is

$$\|\Phi^{-1}\| = \max_{\substack{y \in \text{dom}(\Phi^{-1}) \\ \|y\| \le 1}} \min_{x \in \Phi^{-1}(y)} \|x\| = \max_{\substack{y \in \text{Im}(\Phi) \\ \|y\| \le 1}} \min_{x \in \Phi^{-1}(y)} \|x\|.$$

It is easy to see that both $\|\Phi\|$ and $\|\Phi^{-1}\|$ are finite if $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a polyhedral sublinear mapping.

Suppose $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is sublinear. Then the *upper adjoint* $\Phi^* : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is defined as follows

$$u \in \Phi^*(v) \Leftrightarrow \langle u, x \rangle \le \langle v, y \rangle \text{ for all } (x, y) \in \text{graph}(\Phi).$$

We will rely on the following correspondence between polyhedral sublinear mappings and polyhedral cones. By definition, $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a polyhedral sublinear mapping if

$\mathrm{graph}(\Phi) \subseteq \mathbb{R}^n \times \mathbb{R}^m$ is a polyhedral cone. Conversely, if $K \subseteq \mathbb{R}^n \times \mathbb{R}^m$ is a polyhedral cone, then the set-valued mapping $\Phi_K : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ defined via

$$y \in \Phi_K(x) \Leftrightarrow (x, y) \in K$$

is a polyhedral sublinear mapping with $\mathrm{graph}(\Phi_K) = K$.

A polyhedral set-valued mapping $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *surjective* if $\mathrm{Im}(\Phi) = \mathbb{R}^m$. More generally, we will say that $\Phi$ is *relatively surjective* if $\mathrm{Im}(\Phi)$ is a linear subspace. Suppose $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a polyhedral set-valued mapping. Let

$$\mathfrak{S}(\mathrm{graph}(\Phi)) := \{T \in \mathcal{T}(\mathrm{graph}(\Phi)) : \Phi_T \text{ is relatively surjective}\},$$

and

$$\mathcal{H}(\Phi) := \max_{T \in \mathfrak{S}(\mathrm{graph}(\Phi))} \|\Phi_T^{-1}\|.$$

We have the following general versions of Proposition 1, Proposition 2, and Corollary 1 for polyhedral set-valued mappings.

**Theorem 1.** *Let* $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *be a polyhedral set-valued mapping. Then for all* $b \in \mathrm{Im}(\Phi)$ *and* $u \in \mathrm{dom}(\Phi)$

$$\mathrm{dist}(u, \Phi^{-1}(b)) \leq \mathcal{H}(\Phi) \cdot \mathrm{dist}(b, \Phi(u)). \tag{31}$$

*Furthermore, this bound is tight: If* $\mathcal{H}(\Phi) > 0$ *then there exist* $b \in \mathrm{Im}(\Phi)$ *and* $u \in \mathrm{dom}(\Phi)$ *such that*

$$\mathrm{dist}(u, \Phi^{-1}(b)) = \mathcal{H}(\Phi) \cdot \mathrm{dist}(b, \Phi(u)) > 0.$$

**Theorem 2.** *Let* $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *be a polyhedral set-valued mapping. Then for all* $T \in \mathfrak{S}(\mathrm{graph}(\Phi))$

$$\|\Phi_T^{-1}\| = \max_{\substack{u \in \Phi_T^*(v) \\ \|u\|^* \leq 1}} \|\Pi_{\mathrm{Im}(\Phi_T)}(v)\|^* = \frac{1}{\displaystyle\min_{\substack{u \in \Phi_T^*(v) \\ \|\Pi_{\mathrm{Im}(\Phi_T)}(v)\|^*=1}} \|u\|^*},$$

*where* $\Pi_{\mathrm{Im}(\Phi_T)} : \mathbb{R}^m \to \mathrm{Im}(\Phi_T)$ *denotes the orthogonal projection on the subspace* $\mathrm{Im}(\Phi_T)$. *In particular,*

$$\mathcal{H}(\Phi) = \max_{T \in \mathfrak{S}(\mathrm{graph}(\Phi))} \frac{1}{\displaystyle\min_{\substack{u \in \Phi_T^*(v) \\ \|\Pi_{\mathrm{Im}(\Phi_T)}(v)\|^*=1}} \|u\|^*}.$$

**Corollary 4.** *Let* $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *be a polyhedral set-valued mapping.*

**(a)** *If* $\Phi$ *is sublinear and relatively surjective then*

$$\mathcal{H}(\Phi) = \frac{1}{\displaystyle\min_{\substack{u \in \Phi^*(v) \\ \|\Pi_{\mathrm{Im}(\Phi)}(v)\|^*=1}} \|u\|^*}.$$

**(b)** *Suppose* $\mathfrak{F} \subseteq \mathfrak{S}(\Phi)$ *and* $\mathfrak{I} \subseteq \mathcal{T}(\mathrm{graph}(\Phi)) \setminus \mathfrak{S}(\Phi)$ *are such that for all* $T \in \mathcal{T}(\mathrm{graph}(\Phi))$ *either* $T \subseteq F$ *for some* $F \in \mathfrak{F}$, *or* $I \subseteq T$ *for some* $I \in \mathfrak{I}$. *Then*

$$\mathcal{H}(\Phi) = \max_{T \in \mathfrak{F}} \|\Phi_T^{-1}\| = \max_{T \in \mathfrak{F}} \frac{1}{\displaystyle\min_{\substack{u \in \Phi_T^*(v) \\ \|\Pi_{\mathrm{Im}(\Phi_T)}(v)\|^*=1}} \|u\|^*}.$$

The proof of Theorem 1 relies on the following technical lemma.

**Lemma 1.** *Let* $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *be a polyhedral set-valued mapping. Then*

$$\mathcal{H}(\Phi) = \max_{T \in \mathfrak{S}(\mathrm{graph}(\Phi))} \|\Phi_T^{-1}\| = \max_{T \in \mathcal{T}(\mathrm{graph}(\Phi))} \|\Phi_T^{-1}\|.$$

*Proof of Theorem 1.* To ease notation, throughout this proof let $\mathcal{G} := \mathrm{graph}(\Phi)$. We will prove the following equivalent statement to (31): For all $b \in \mathrm{Im}(\Phi)$ and $(u, v) \in \mathcal{G}$

$$\mathrm{dist}(u, \Phi^{-1}(b)) \leq \mathcal{H}(\Phi) \cdot \|b - v\|. \tag{32}$$

Assume $b - v \neq 0$ as otherwise there is nothing to show. We proceed by contradiction. Suppose $b \in \mathrm{Im}(\Phi)$ and $(u, v) \in \mathcal{G}$ are such that $b - v \neq 0$ and

$$\|x - u\| > H \cdot \|b - v\| \tag{33}$$

for all $x$ such that $(x, b) \in \mathcal{G}$. Let $d := \frac{b-v}{\|b-v\|}$ and consider the optimization problem

$$
\begin{aligned}
\max_{w,t} \quad & t \\
\text{s.t.} \quad & (u + w, v + td) \in \mathcal{G}, \\
& \|w\| \leq \mathcal{H}(\Phi) \cdot t.
\end{aligned}
\tag{34}
$$

Since $b \in \mathrm{Im}(\Phi)$ it follows that $d = (b - v)/\|b - v\| \in \mathrm{Im}(\Phi_T)$ for $T := T_{\mathcal{G}}(u, v) \in \mathcal{T}(\mathcal{G})$. Since $\|d\| = 1$, the definition of $\|\Phi_T^{-1}\|$ and Lemma 1 imply that there exists $(z, d) \in T$ with $\|z\| \leq \|\Phi_T^{-1}\| \leq \mathcal{H}(\Phi)$. By the construction of $T = T_{\mathcal{G}}(u, v)$ it follows that for $t > 0$ sufficiently small $(u + tz, v + td) \in \mathcal{G}$ and so $(w, t) := (tz, t)$ is feasible for problem (34). Let

$$S := \{(w, t) \in \mathbb{R}^n \times \mathbb{R}_+ : (w, t) \text{ is feasible for } (34)\}.$$

Assumption (33) implies that $t < \|b - v\|$ for all $(w, t) \in S$. In addition, since $\mathcal{G}$ is polyhedral, it follows that $S$ is compact. Therefore (34) has an optimal solution $(\bar{w}, \bar{t})$ with $0 < \bar{t} < \|b - v\|$.

Let $(u', v') := (u + \bar{w}, v + \bar{t}d) \in \mathcal{G}$. Consider the modification of (34) obtained by replacing $(u, v)$ with $(u', v')$, namely

$$
\begin{aligned}
\max_{w',t'} \quad & t' \\
\text{s.t.} \quad & (u' + w', v' + t'd) \in \mathcal{G}, \\
& \|w'\| \leq \mathcal{H}(\Phi) \cdot t'.
\end{aligned}
\tag{35}
$$

Observe that $b - v' = b - v - \bar{t}d = (\|b - v\| - \bar{t})d \neq 0$. Again since $b \in \mathrm{Im}(\Phi)$ it follows that $d = \frac{b-v'}{\|b-v'\|} \in \mathrm{Im}(\Phi_{T'})$ for $T' := T_{\mathcal{G}}(u', v')$. Hence there exists $(z', d) \in T'$ such that $\|z'\| \leq \|\Phi_{T'}^{-1}\| \leq \mathcal{H}(\Phi)$. Therefore, (35) has a feasible point $(w', t') = (t'z', t')$ with $t' > 0$. In particular $(u' + w', v' + t'd) = (u + \bar{w} + w', v + (\bar{t} + t')d) \in \mathcal{G}$ with $\|\bar{w} + w'\| \leq \|\bar{w}\| + \|w'\| \leq \mathcal{H}(\Phi) \cdot (\bar{t} + t')$ and $\bar{t} + t' > \bar{t}$. This contradicts the optimality of $(\bar{w}, \bar{t})$ for (34).

To show that the bound is tight, suppose $\mathcal{H}(\Phi) = \|\Phi_T^{-1}\| > 0$ for some $T \in \mathfrak{S}(\Phi) \subseteq \mathcal{T}(\mathcal{G})$. The construction of $\|\Phi_T^{-1}\|$ implies that there exists $d \in \mathbb{R}^m$ with $\|d\| = 1$ such that the problem

$$
\begin{aligned}
\min_{z} \quad & \|z\| \\
\text{s.t.} \quad & (z, d) \in T
\end{aligned}
\tag{36}
$$

is feasible and has an optimal solution $\bar{z}$ with $\|\bar{z}\| = \|\Phi_T^{-1}\| = \mathcal{H}(\Phi) > 0$. Let $(u, v) \in \mathcal{G}$ be such that $T = T_{\mathcal{G}}(u, v)$. Let $b := v + td$ where $t > 0$ is small enough so that $(u, v) + t(\bar{z}, d) \in \mathcal{G}$. Observe that $b \in \mathrm{Im}(\Phi)$ and $b - v = td \neq 0$. To finish, notice that if $x \in \Phi^{-1}(b)$ then $(x - u, b - v) = (x - u, td) \in T_{\mathcal{G}}(u, v) = T$. The optimality of $\bar{z}$ then implies that

$$\|x - u\| \geq \mathcal{H}(\Phi) \cdot t = \mathcal{H}(\Phi) \cdot \|b - v\|.$$

Since this holds for all $x \in \Phi^{-1}(b)$ and $b - v \neq 0$, it follows that $\mathrm{dist}(u, \Phi^{-1}(b)) \geq \mathcal{H}(\Phi) \cdot \|b - v\| \geq \mathcal{H}(\Phi) \cdot \mathrm{dist}(b, \Phi(u)) > 0$. $\qquad \square$

The proofs of Theorem 2 and Lemma 1 rely on the following convex duality construction. Observe that for a polyhedral convex cone $T \subseteq \mathbb{R}^n \times \mathbb{R}^m$

$$
\begin{aligned}
\|\Phi_T^{-1}\| = \max_{y} \quad & \|\Phi_T^{-1}(y)\| \\
\mathrm{s.\,t.} \quad & y \in \mathrm{Im}(\Phi_T) \\
& \|y\| \leq 1,
\end{aligned}
$$

where

$$
\begin{aligned}
\|\Phi_T^{-1}(y)\| := \min_{x} \quad & \|x\| \\
\mathrm{s.\,t.} \quad & (x, y) \in T.
\end{aligned}
\tag{37}
$$

By convex duality it follows that

$$
\begin{aligned}
\|\Phi_T^{-1}(y)\| = \max_{u,v} \quad & -\langle v, y \rangle \\
\mathrm{s.\,t.} \quad & \|u\|^* \leq 1 \\
& (u, v) \in T^*.
\end{aligned}
\tag{38}
$$

Therefore when $T \subseteq \mathbb{R}^n \times \mathbb{R}^m$ is a polyhedral cone we have

$$
\begin{aligned}
\|\Phi_T^{-1}\| = \max_{u,v,y} \quad & -\langle v, y \rangle \\
\mathrm{s.\,t.} \quad & y \in \mathrm{Im}(\Phi_T) \\
& \|y\| \leq 1 \\
& \|u\|^* \leq 1 \\
& (u, v) \in T^*.
\end{aligned}
\tag{39}
$$

*Proof of Theorem 2.* Let $T \in \mathfrak{S}(\mathrm{graph}(\Phi))$. Since $\Phi_T$ is relatively surjective, from (39) it follows that

$$
\begin{aligned}
\|\Phi_T^{-1}\| = \max_{u,v} \quad & \|\Pi_{\mathrm{Im}(\Phi_T)}(v)\|^* \\
\mathrm{s.\,t.} \quad & \|u\|^* \leq 1 \\
& u \in \Phi_T^*(v).
\end{aligned}
$$

The latter quantity is evidently the same as

$$
\frac{1}{\displaystyle\min_{\substack{u \in \Phi_T^*(v) \\ \|\Pi_{\mathrm{Im}(\Phi_T)}(v)\|^* = 1}} \|u\|^*}.
$$

$\qquad \square$

Our proof of Lemma 1 relies on the following equivalence between *surjectivity* and *non-singularity* of sublinear mappings. A standard convex separation argument shows that a closed sublinear mapping $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is surjective if and only if

$$(0, v) \in \text{graph}(\Phi)^* \Rightarrow v = 0. \tag{40}$$

Condition (40) is a kind of *non-singularity* of $\Phi^*$ as it can be rephrased as $0 \in \Phi^*(v) \Rightarrow v = 0$.

*Proof of Lemma 1.* Without loss of generality assume affine$(\text{Im}(\Phi)) = \mathbb{R}^m$ as otherwise we can work instead with the modified mapping $\Phi_0 : \mathbb{R}^n \rightrightarrows L$ defined via $\Phi_0(x) := \Phi(x) - y_0$, where $y_0 \in \text{Im}(\Phi)$ and $L$ is the lineality space of affine$(\text{Im}(\Phi))$, that is, $L = \text{affine}(\text{Im}(\Phi)) - y_0$.

To ease notation let $\mathcal{G} := \text{graph}(\Phi)$. We need to show that

$$\max_{T \in \mathcal{T}(\mathcal{G})} \|\Phi_T^{-1}\| = \max_{T \in \mathcal{S}(\mathcal{G})} \|\Phi_T^{-1}\|.$$

By construction, it is immediate that

$$\max_{T \in \mathcal{T}(\mathcal{G})} \|\Phi_T^{-1}\| \geq \max_{T \in \mathcal{S}(\mathcal{G})} \|\Phi_T^{-1}\|.$$

To prove the reverse inequality let $T \in \mathcal{T}(\mathcal{G})$ be fixed and let $(\bar{u}, \bar{v}, \bar{y})$ attain the optimal value $\|\Phi_T^{-1}\|$ in (39). Let $\bar{F}$ be the minimal face of $T^*$ containing $(\bar{u}, \bar{v})$ and $\bar{T} := \bar{F}^* \in \mathcal{T}(T) \subseteq \mathcal{T}(\mathcal{G})$. As we detail below, $(\bar{u}, \bar{v}, \bar{y})$ can be chosen so that $\Phi_{\bar{T}}$ is surjective. Since $\|\bar{y}\| \leq 1$ we have

$$\|\Phi_T^{-1}\| = -\langle \bar{v}, \bar{y} \rangle \leq \|\bar{v}\|^*.$$

Furthermore, $(\bar{u}, \bar{v}) \in \bar{F} \subseteq T^* = \text{graph}(\Phi_{\bar{T}})^*$ and $\|\bar{u}\|^* \leq 1$, thus Theorem 2 yields

$$\|\Phi_T^{-1}\| \leq \|\bar{v}\|^* \leq \|\Phi_{\bar{T}}^{-1}\|.$$

Since this holds for any fixed $T \in \mathcal{T}(\mathcal{G})$, it follows that

$$\max_{T \in \mathcal{T}(\mathcal{G})} \|\Phi_T^{-1}\| \leq \max_{\bar{T} \in \mathcal{S}(\mathcal{G})} \|\Phi_{\bar{T}}^{-1}\|.$$

It remains to show that $(\bar{u}, \bar{v}, \bar{y})$ can be chosen so that $\Phi_{\bar{T}}$ is surjective, where $\bar{T} = \bar{F}^*$ and $\bar{F}$ is the minimal face of $T^*$ containing $(\bar{u}, \bar{v})$. To that end, pick a solution $(\bar{u}, \bar{v}, \bar{y})$ to (39) and consider the set

$$V := \{v \in \mathbb{R}^m : \langle v, \bar{y} \rangle = \langle \bar{v}, \bar{y} \rangle, (\bar{u}, v) \in T^*\}.$$

In other words, $V$ is the projection of the set of optimal solutions to (39) of the form $(\bar{u}, v, \bar{y})$. Since $T$ is polyhedral, so is $T^*$ and thus $V$ is a polyhedron. Furthermore, $V$ must have at least one extreme point. Otherwise there exist $\hat{v} \in V$ and a nonzero $\tilde{v} \in \mathbb{R}^m$ such that $\hat{v} + t\tilde{v} \in V$ for all $t \in \mathbb{R}$. In particular, $(\bar{u}, \hat{v} + t\tilde{v}) \in T^*$ for all $t \in \mathbb{R}$ and thus $(0, t\tilde{v}) \in T^*$ for all $t \in \mathbb{R}$. The latter in turn implies $\text{Im}(\Phi_T) = \{y \in \mathbb{R}^m : (x, y) \in T \text{ for some } x \in \mathbb{R}^n\} \subseteq \{y \in \mathbb{R}^m : \langle \tilde{v}, y \rangle = 0\}$. Since $T \in \mathcal{T}(\mathcal{G})$, if follows that $\text{Im}(\Phi)) \subseteq \{y \in \mathbb{R}^m : \langle \tilde{v}, y \rangle = 0\}$ and thus affine$(\text{Im}(\Phi)) \subseteq \{y \in \mathbb{R}^m : \langle \tilde{v}, y \rangle = 0\}$ thereby contradicting the assumption

24

affine$(\mathrm{Im}(\Phi)) = \mathbb{R}^m$. By replacing $\bar{v}$ if necessary, we can assume that $\bar{v}$ is an extreme point of $V$. We claim that the minimal face $\bar{F}$ of $K^*$ containing $(\bar{u}, \bar{v})$ satisfies

$$(0, v') \in \bar{F} = \bar{T}^* \Rightarrow v' = 0$$

thereby establishing the surjectivity of $\Phi_{\bar{T}}$ (cf., (40)). To prove this claim, proceed by contradiction. Assume $(0, v') \in \bar{F}$ for some nonzero $v' \in \mathbb{R}^m$. The choice of $\bar{F}$ ensures that $(\bar{u}, \bar{v})$ lies in the relative interior of $\bar{F}$ and thus for $t > 0$ sufficiently small both $(\bar{u}, \bar{v} + tv') \in \bar{F} \subseteq T^*$ and $(\bar{u}, \bar{v} - tv') \in \bar{F} \subseteq T^*$. The optimality of $(\bar{u}, \bar{v}, \bar{y})$ implies that both $\langle \bar{v} + tv', \bar{y} \rangle \geq \langle \bar{v}, \bar{y} \rangle$ and $\langle \bar{v} - tv', \bar{y} \rangle \geq \langle \bar{v}, \bar{y} \rangle$ and so $\langle v', \bar{y} \rangle = 0$. Thus both $\bar{v} + tv' \in V$ and $\bar{v} - tv' \in V$ with $tv' \neq 0$ thereby contradicting the assumption that $\bar{v}$ is an extreme point of $V$. $\qquad \square$

To conclude this section, we briefly describe how the approach to compute Hoffman constants via the covering property in Section 3.2 extends to the general context of polyhedral set-valued mappings. Suppose $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a polyhedral set-valued mapping. Corollary 4 suggests the following algorithmic approach to compute $\mathcal{H}(\Phi)$: Find $\mathfrak{F} \subseteq \mathfrak{S}(\Phi)$ and $\mathfrak{I} \subseteq \mathcal{T}(\mathrm{graph}(\Phi)) \setminus \mathfrak{S}(\Phi)$ that satisfy the following covering property:

For all $T \in \mathcal{T}(\mathrm{graph}(\Phi))$ either $T \subseteq F$ for some $F \in \mathfrak{F}$, or $I \subseteq T$ for some $I \in \mathfrak{I}$.

Then compute

$$\mathcal{H}(\Phi) = \max_{T \in \mathfrak{F}} \|\Phi_T^{-1}\| = \max_{T \in \mathfrak{F}} \frac{1}{\displaystyle\min_{\substack{u \in \Phi_T^*(v) \\ \|\Pi_{\mathrm{Im}(\Phi_T)}(v)\|^* = 1}} \|u\|^*}.$$

# 5 Proofs of propositions in Section 2

We next present the proofs of Proposition 5 and Proposition 6. As noted before, the other propositions in Section 2 follow as special cases of these two results.

Let $R \subseteq \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, and $C \in \mathbb{R}^{p \times n}$. Construct $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^{m+p}$ as follows

$$\Phi(x) = \begin{cases} \{(Ax + s, Cx) : s \geq 0\} & \text{if } x \in R \\ \emptyset & \text{otherwise.} \end{cases} \tag{41}$$

Observe that $\Phi$ is polyhedral since $R$ is a polyhedron.

*Proof of Proposition 5.* For $\Phi$ as in (41) we have

$$\mathfrak{S}(\mathrm{graph}(\Phi)) = \{T_{J,K} : (J, K) \in \mathcal{S}(A, C|R)\}$$

where $T_{J,K} = \{(x, Ax + s, Cx) : x \in K, s_J \geq 0\}$. Next, observe that for $(J, K) \in \mathcal{S}(A, C|R)$ we have $(y, w) \in \Phi_{T_{J,K}}(x) \Leftrightarrow x \in K$, $A_J x \leq y_J$, and $Cx = w$. Therefore

$$\mathcal{H}(\Phi) = \max_{J \in \mathcal{S}(A)} \|\Phi_{T_{J,K}}^{-1}\| = \max_{(J,K) \in \mathcal{S}(A,C|R)} \max_{\substack{(y,w) \in \mathbb{R}^m \times C(K) \\ \|(y,w)\| \leq 1}} \min_{\substack{x \in L \\ A_J x \leq y_J, Cx = w}} \|x\| = H(A, C|R).$$

Furthermore, $\mathrm{dom}(\Phi) = R$ and $\mathrm{Im}(\Phi) = \{(Ax+s, Cx) : x \in R,\, s \geq 0\}$. Therefore Theorem 1 implies that for all $(b, d) \in \mathrm{Im}(\Phi) = \{(Ax + s, Cx) : x \in R,\, s \geq 0\}$ and $u \in \mathrm{dom}(\Phi) = R$

$$
\begin{aligned}
\mathrm{dist}(u, P_A(b) \cap C^{-1}(d) \cap R) &= \mathrm{dist}(u, \Phi^{-1}(b, d)) \\
&\leq \mathcal{H}(\Phi) \cdot \mathrm{dist}((b, d), \Phi(u)) \\
&= H(A, C|R) \cdot \mathrm{dist}((b - Au, d - Cu), \mathbb{R}_+^m \times \{0\}).
\end{aligned}
$$

Theorem 1 also implies that this bound is tight.

$\square$

*Proof of Proposition 6.* Observe that for $(J, K) \in \mathcal{S}(A, C|R))$ and $T := T_{J,K}$ we have $\mathrm{Im}(\Phi_T) = \mathbb{R}^m \times C(K)$ and $u \in \Phi_T^*(v, z)$ if and only if $A^\mathsf{T} v + C^\mathsf{T} z - u \in K^*$, $v_J \geq 0$, and $v_{J^c} = 0$. Hence

$$
\min_{\substack{u \in \Phi_T^*(v,z) \\ \|\Pi_{\mathrm{Im}(\Phi_T)}(v,z)\|^* = 1}} \|u\|^* = \min_{\substack{v \in \mathbb{R}_+^J,\, z \in C(K) \\ \|(v,z)\|^* = 1,\, A_J^\mathsf{T} v + C^\mathsf{T} z - u \in L^*}} \|u\|^*.
$$

To finish, apply Theorem 2 and the facts $\mathfrak{S}(\mathrm{graph}(\Phi)) = \{T_{J,K} : (J, K) \in \mathcal{S}(A, C|R)\}$ and $\mathcal{H}(\Phi) = H(A, C|R)$ established in the previous proof.

$\square$

# 6 Conclusions

We provide a characterization of the Hoffman constant for a system of linear inequalities and equations relative to a reference polyhedron (Proposition 5). Our characterization is stated as the largest of a finite collection of easily computable Hoffman constants (Proposition 6).

We describe how our characterization can be leveraged to design two classes of algorithmic procedures to compute Hoffman constants. One of them is based on a formulation of the Hoffman constant as a mathematical program with linear complementarity constraints (Proposition 7). The other one is based on a certain type of covering property (Algorithm 1).

We also develop the concept of Hoffman constant generalized our characterization and covering property to compute it in the more general context of polyhedral set-valued mappings (Theorem 1 and Theorem 2).

# Acknowledgements

# References

[1] D. Amelunxen and P. Bürgisser. A coordinate-free condition number for convex programming. *SIAM J. on Optim.*, 22(3):1029–1041, 2012.

[2] D. Azé and J. Corvellec. On the sensitivity analysis of Hoffman constants for systems of linear inequalities. *SIAM Journal on Optimization*, 12(4):913–927, 2002.

[3] H. Bauschke, J. Bolte, and M. Teboulle. A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications. *Mathematics of Operations Research*, 42(2):330–348, 2016.

[4] A. Beck and S. Shtern. Linearly convergent away-step conditional gradient for non-strongly convex functions. *Mathematical Programming*, 164:1–27, 2017.

[5] P. Bürgisser and F. Cucker. *Condition*. Springer Berlin Heidelberg, 2013.

[6] J. Burke and P. Tseng. A unified analysis of Hoffman's bound via Fenchel duality. *SIAM Journal on Optimization*, 6(2):265–282, 1996.

[7] M. Epelman and R. Freund. A new condition measure, preconditioners, and relations between different measures of conditioning for conic linear systems. *SIAM J. on Optim.*, 12:627–655, 2002.

[8] R. Freund. Complexity of convex optimization using geometry-based measures and a reference point. *Math Program.*, 99:197–221, 2004.

[9] R. Freund and J. Vera. Some characterizations and properties of the "distance to ill-posedness" and the condition measure of a conic linear system. *Math Program.*, 86:225–260, 1999.

[10] R. Freund and J. Vera. On the complexity of computing estimates of condition measures of a conic linear system. *Mathematics of Operations Research*, 28(4):625–648, 2003.

[11] D. Garber. Fast rates for online gradient descent without strong convexity via Hoffman's bound. *arXiv preprint arXiv:1802.04623*, 2018.

[12] F. Granot and J. Skorin-Kapov. Some proximity and sensitivity results in quadratic integer programming. *Mathematical Programming*, 47(1-3):259–268, 1990.

[13] O. Güler, A. Hoffman, and U. Rothblum. Approximations to solutions to systems of linear inequalities. *SIAM Journal on Matrix Analysis and Applications*, 16(2):688–696, 1995.

[14] D. Gutman and J. Peña. The condition number of a function relative to a set. *arXiv preprint arXiv:1901.08359*, 2019.

[15] A. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, 49(4):263–265, 1952.

[16] A. Jourani. Hoffman's error bound, local controllability, and sensitivity analysis. *SIAM Journal on Control and Optimization*, 38(3):947–970, 2000.

[17] J. Júdice. Algorithms for linear programming with linear complementarity constraints. *Top*, 20(1):4–25, 2012.

[18] D. Klatte and G. Thiere. Error bounds for solutions of linear equations and inequalities. *Zeitschrift für Operations Research*, 41(2):191–214, 1995.

[19] S. Lacoste-Julien and M. Jaggi. On the global linear convergence of Frank-Wolfe optimization variants. In *Advances in Neural Information Processing Systems (NIPS)*, 2015.

[20] D. Leventhal and A. Lewis. Randomized methods for linear constraints: Convergence rates and conditioning. *Math. Oper. Res.*, 35:641–654, 2010.

[21] A. Lewis. Ill-conditioned convex processes and linear inequalities. *Math. Oper. Res.*, 24:829–834, 1999.

[22] A. Lewis. The structured distance to ill-posedness for conic systems. *Math. Oper. Res.*, 29:776–785, 2005.

[23] W. Li. The sharp Lipschitz constants for feasible and optimal solutions of a perturbed linear program. *Linear Algebra and its Applications*, 187:15–40, 1993.

[24] H. Lu. "Relative-continuity" for non-Lipschitz non-smooth convex optimization using stochastic (or deterministic) mirror descent. *arXiv preprint arXiv:1710.04718*, 2017.

[25] H. Lu, R. Freund, and Y. Nesterov. Relatively smooth convex optimization by first-order methods, and applications. *SIAM Journal on Optimization*, 28(1):333–354, 2018.

[26] Z. Luo and P. Tseng. Error bounds and convergence analysis of feasible descent methods: a general approach. *Annals of Operations Research*, 46(1):157–178, 1993.

[27] O. Mangasarian and T-H Shiau. Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems. *SIAM Journal on Control and Optimization*, 25(3):583–595, 1987.

[28] I. Necoara, Y. Nesterov, and F. Glineur. Linear convergence of first order methods for non-strongly convex optimization. *To Appear in Mathematical Programming*, 2018.

[29] T. Nguyen. A stroll in the jungle of error bounds. *arXiv preprint arXiv:1704.06938*, 2017.

[30] J. S. Pang. Error bounds in mathematical programming. *Math. Program.*, 79:299–332, 1997.

[31] J. Peña. Understanding the geometry on infeasible perturbations of a conic linear system. *SIAM J. on Optim.*, 10:534–550, 2000.

[32] J. Peña. A characterization of the distance to infeasibility under block-structured perturbations. *Linear algebra and its applications*, 370:193–216, 2003.

[33] J. Peña and D. Rodríguez. Polytope conditioning and linear convergence of the Frank-Wolfe algorithm. *To Appear in Mathematics of Operations Research*, 2018.

[34] S Pineda, H Bylling, and JM Morales. Efficiently solving linear bilevel programming problems using off-the-shelf optimization software. *Optimization and Engineering*, 19(1):187–211, 2018.

[35] A. Ramdas and J. Peña. Towards a deeper geometric, analytic and algorithmic understanding of margins. *Optimization Methods and Software*, 31(2):377–391, 2016.

[36] J. Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM J. on Optim.*, 5:506–524, 1995.

[37] J. Renegar. Linear programming, complexity theory and elementary functional analysis. *Math. Program.*, 70:279–351, 1995.

[38] S. Robinson. Bounds for error in the solution set of a perturbed linear program. *Linear Algebra and its applications*, 6:69–81, 1973.

[39] S. Siddiqui and S. Gabriel. An SOS1-based approach for solving mpecs with a natural gas market application. *Networks and Spatial Economics*, 13(2):205–227, 2013.

[40] O. Stein. Error bounds for mixed integer linear optimization problems. *Mathematical Programming*, 156(1-2):101–123, 2016.

[41] G. Stewart. On scaled projections and pseudoinverses. *Linear Algebra and its Applications*, 112:189–193, 1989.

[42] IBM ILOG CPLEX Optimization Studio-CPLEX. Users manual-version 12 release 6, 2013.

[43] M. Teboulle. A simplified view of first order methods for optimization. *Mathematical Programming*, pages 1–30, 2018.

[44] M. Todd. A Dantzig-Wolfe-like variant of Karmarkar's interior-point linear programming algorithm. *Operations Research*, 38(6):1006–1018, 1990.

[45] H. Van Ngai and M. Théra. Error bounds for systems of lower semicontinuous functions in Asplund spaces. *Mathematical Programming*, 116(1-2):397–427, 2009.

[46] S. Vavasis and Y. Ye. A primal-dual interior point method whose running time depends only on the constraint matrix. *Mathematical Programming*, 74(1):79–120, 1996.

[47] P. Wang and C. Lin. Iteration complexity of feasible descent methods for convex optimization. *Journal of Machine Learning Research*, 15(1):1523–1548, 2014.

[48] W. Xia, J. Vera, and L. F. Zuluaga. Globally solving non-convex quadratic programs via linear integer programming techniques. *arXiv preprint arXiv:1511.02423*, 2015.

[49] C. Zalinescu. Sharp estimates for Hoffman's constant for systems of linear inequalities and equalities. *SIAM Journal on Optimization*, 14(2):517–533, 2003.

[50] S. Zhang. Global error bounds for convex conic problems. *SIAM Journal on Optimization*, 10(3):836–851, 2000.

[51] Z. Zhou and A. So. A unified approach to error bounds for structured convex optimization problems. *Mathematical Programming*, 165(2):689–728, 2017.