Equivalences among the chi measure, Hoffman constant, and Renegar's distance to ill-posedness

Javier F. Peña* Juan C. Vera† Luis F. Zuluaga‡

May 15, 2019

Abstract

We show the equivalence among the following three condition measures of a full column rank matrix A: the chi measure, the signed Hoffman constant, and the signed distance to ill-posedness. The latter two measures are constructed via suitable collections of matrices obtained by flipping the signs of some rows of A. Our results provide a procedure to estimate $\chi(A)$ thereby opening an avenue to identify classes of linear programs solvable in polynomial time in the real model of computation.

1 Introduction

We establish new equivalences among three types of condition measures of a matrix that play central roles in numerical linear algebra and in convex optimization: the chi measure [3, 7, 9, 31, 32], the Hoffman constant [15, 17, 19, 37], and Renegar's distance to ill-posedness [29, 30]. We recall the definitions of these quantities in Section 2 below.

Let $A \in \mathbb{R}^{m\times n}$ be a full column rank matrix. The chi measure $\chi(A)$ arises naturally in weighted least-squares problems of the form $\min \|D^{1/2}(Ax-b)\|^2$, see, e.g., [4, 9, 10, 18]. The chi measure $\chi(A)$ is also a key component in the analysis of Vavasis and Ye's interior-point algorithm for linear programming [23, 36]. A remarkable feature of Vavasis and Ye's algorithm is its sole dependence on the matrix A defining the primal and dual constraints. The Hoffman constant $H(A)$ is associated to Hoffman's Lemma [15, 17], a fundamental *error bound* for systems of linear constraints of the form $Ax \le b$. The Hoffman constant and other similar error bounds are used to establish the convergence rate of a wide variety of optimization algorithms [2, 14, 16, 20–22, 24–26, 37, 37]. Renegar's distance to ill-posedness $\mathcal{R}(A)$ is a pillar for the concept of *condition number* in optimization introduced by Renegar in the seminal articles [29, 30] and subsequently extended in a number of articles [1, 5, 8, 11–13].

Our work is inspired by several relationships among $\chi(\cdot)$, $H(\cdot)$, and $\mathcal{R}(\cdot)$ previously established in [6, 8, 27, 34, 35, 39]. In particular, it is known that if $A \in \mathbb{R}^{m\times n}$ is full column

*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

rank, then $\chi(A) \geq H(A)$ and if $Ax < 0$ is feasible then $H(A) = 1/\mathcal{R}(A)$. However, $\chi(A)$ can be arbitrarily larger than $H(A)$ (see, e.g., [27]). Also, the equivalence between $\chi(A)$ and $1/\mathcal{R}(A)$ breaks down when $Ax < 0$ is infeasible. Our main result (Theorem 1) shows that the lack of equivalence among these quantities can be rectified by considering *signed* versions of $H(\cdot)$ and $\mathcal{R}(\cdot)$. In hindsight our equivalences are somewhat natural because $\chi(A)$ does not change when the signs of some rows of A are flipped whereas both $H(A)$ and $\mathcal{R}(A)$ evidently do. We show that $\chi(A)$ is exactly the largest $H(\hat{A})$ over the collection of matrices $\hat{A}$ obtained by flipping the signs of some rows of A. We also show that when all rows of A are non-zero, $1/\chi(A)$ is the same as the smallest $\mathcal{R}(\hat{A})$ over the collection of all matrices $\hat{A}$ obtained by flipping the signs of some rows of A so that $\hat{A}x < 0$ is feasible. Furthermore, we show that $\chi(A)$ is the same as $H(\mathbf{A})$ for the matrix $\mathbf{A}$ obtained by stacking the rows of A and $-A$. The latter equivalence together with the algorithmic machinery recently developed in [27] provides a procedure to compute or estimate $\chi(A)$. That computational ability in turn offers the potential to identify classes of linear programs that are solvable in polynomial time in the real model of computation via Vavasis-Ye's interior-point algorithm [23, 36], since the number of arithmetic operations of Vavasis-Ye's algorithm is polynomial on the dimensions of A and on $\log(\bar{\chi}(A))$ for a variant $\bar{\chi}(A)$ of $\chi(A)$.

Some of our equivalences are reminiscent of results previously developed by Tunçel [34] and by Todd, Tunçel, and Ye [33] to compare a variant $\bar{\chi}(A)$ of $\chi(A)$ and Ye's condition measure [38] for polyhedra of the form $\{A^{\mathsf{T}}y : y \geq 0, \|y\|_1 = 1\}$.

2 Definition of $\chi(\cdot)$, $H(\cdot)$, and $\mathcal{R}(\cdot)$

Let $A \in \mathbb{R}^{m\times n}$ have full column rank. The *chi measure* of A is defined as

$$\chi(A) = \sup\{\|(A^{\mathsf{T}}\operatorname{Diag}(d)A)^{-1}A^{\mathsf{T}}\operatorname{Diag}(d)\| : d \in \mathbb{R}^m_{++}\}.$$

In this expression and throughout the paper, $\operatorname{Diag}(d) \in \mathbb{R}^{m\times m}$ denotes the diagonal matrix whose vector of diagonal entries is $d \in \mathbb{R}^m$. Also, we write $\|\cdot\|$ to denote the canonical Euclidean norms in $\mathbb{R}^m$ and $\mathbb{R}^n$, and the corresponding induced operator norm (or equivalently the spectral norm) in $\mathbb{R}^{m\times n}$. The underlying space will always be clear from the context. Several authors [3, 7, 31, 32] independently showed that $\chi(A)$ is finite as long as A is full column rank. See [9] for a detailed discussion.

Let $A \in \mathbb{R}^{m\times n}$. The *Hoffman constant* $H(A)$ of A is defined as

$$H(A) = \sup\left\{\frac{\operatorname{dist}(u, P_A(b))}{\|(Au-b)_+\|} : b \in A(\mathbb{R}^n) + \mathbb{R}^m_+ \text{ and } u \notin P_A(b)\right\}$$

where $P_A(b) := \{x \in \mathbb{R}^n : Ax \leq b\}$ and $\operatorname{dist}(u, P_A(b)) = \min\{\|u-x\| : x \in P_A(b)\}$. Hoffman [17] showed that $H(A)$ is always finite. Other proofs of this fundamental result can be found in [15, 27, 37].

Let $A \in \mathbb{R}^{m\times n}$ be such that $Ax < 0$ is feasible. *Renegar's distance to ill-posedness* of A is defined as

$$\mathcal{R}(A) := \inf\{\|\Delta A\| : (A + \Delta A)x < 0 \text{ is infeasible}\}.$$

Renegar introduced the distance to ill-posedness as a main building block to develop the concept of *condition number* for optimization problems [29, 30].

The following proposition, which recalls properties previously established in [19,27,28,39], is our starting point.

Proposition 1. *Let $A \in \mathbb{R}^{m\times n}$. If A has full column rank then*

$$\chi(A) \geq H(A). \tag{1}$$

On the other hand, if $Ax < 0$ is feasible then

$$H(A) = \frac{1}{\mathcal{R}(A)}. \tag{2}$$

3 Equivalences among $\chi(\cdot)$, $H(\cdot)$, and $\mathcal{R}(\cdot)$

Let $A \in \mathbb{R}^{m\times n}$. The following two collections $\mathbb{S}(A)$ and $\mathbb{D}(A)$ of signed matrices associated to A play a central role in our main developments. Let

$$\mathbb{S}(A) := \{\mathrm{Diag}(d)A : d \in \{-1, 1\}^m\},$$

and

$$\mathbb{D}(A) := \{\hat{A} \in \mathbb{S}(A) : \hat{A}x < 0 \text{ is feasible}\}.$$

We are now ready to state our main result.

Theorem 1. *Let $A \in \mathbb{R}^{m\times n}$ have full column rank. Then*

$$\chi(A) = \max_{\hat{A}\in\mathbb{S}(A)} H(\hat{A}) = H(\mathbf{A}), \tag{3}$$

where $\mathbf{A} \in \mathbb{R}^{2m\times n}$ is the matrix obtained by stacking A and $-A$, that is, $\mathbf{A} = \begin{bmatrix} A \\ -A \end{bmatrix}$.

If in addition all rows of A are nonzero then

$$\chi(A) = \max_{\hat{A}\in\mathbb{D}(A)} H(\hat{A}) = \max_{\hat{A}\in\mathbb{D}(A)} \frac{1}{\mathcal{R}(\hat{A})}. \tag{4}$$

The identity (4) in Theorem 1 has the following natural extension when some rows of A are zero. Given $A \in \mathbb{R}^{m\times n}$, let $\tilde{A} \in \mathbb{R}^{\tilde{m}\times n}$ denote the submatrix of A obtained by dropping the zero rows from A. If $A \in \mathbb{R}^{m\times n}$ has full column rank then so does $\tilde{A}$ and Theorem 1 implies that

$$\chi(A) = \chi(\tilde{A}) = \max_{\hat{A}\in\mathbb{D}(\tilde{A})} H(\hat{A}) = \max_{\hat{A}\in\mathbb{D}(\tilde{A})} \frac{1}{\mathcal{R}(\hat{A})}. \tag{5}$$

The identity (5) in turn suggests an extension of $\chi(\cdot)$ to general (not necessarily full rank) matrices and general (not necessarily Euclidean) norms since both $H(\cdot)$ and $\mathcal{R}(\cdot)$ are defined in full generality and satisfy (2).

The proof of Theorem 1 relies on the two key building blocks stated as Proposition 2 and Proposition 3 below. We will use the following convenient notation. For a positive integer

m, let $[m]$ denote $\{1, \dots, m\}$. For $A \in \mathbb{R}^{m\times n}$ and $J \subseteq [m]$, we let $A_J \in \mathbb{R}^{J\times n}$ denote the submatrix of A defined by the rows indexed by J.

The first key building block for the proof of Theorem 1 is the following characterization of $\chi(\cdot)$ from [9]. The same characterization is also stated and proved in [39] by adapting a technique from [33].

Proposition 2. *Let $A \in \mathbb{R}^{m\times n}$ have full column rank. Then*

$$\chi(A) = \max_{\substack{J\subseteq[m], |J|=n \\ A_J \text{ non-singular}}} \max_{\substack{v\in\mathbb{R}^J \\ \|A_J^\mathsf{T} v\|=1}} \|v\|.$$

The second building block for the proof of Theorem 1 is the following characterization of $H(\cdot)$ discussed in [27] but that can be traced back to [19, 37, 39].

Proposition 3. *Let $A \in \mathbb{R}^{m\times n}$. Then*

$$H(A) = \max_{\substack{J\subseteq[m], |J|=n \\ A_J \text{ non-singular}}} \max_{\substack{v\in\mathbb{R}^J_+ \\ \|A_J^\mathsf{T} v\|=1}} \|v\| = \max_{J\in\overline{\mathcal{J}}(A)} \max_{\substack{v\in\mathbb{R}^J_+ \\ \|A_J^\mathsf{T} v\|=1}} \|v\|,$$

where $\mathcal{J}(A) = \{J \subseteq [m] : A_J x < 0 \text{ is feasible}\}$ and $\overline{\mathcal{J}}(A) \subseteq \mathcal{J}(A)$ is the collection of maximal sets in $\mathcal{J}(A)$.

Proof of Theorem 1. Let J and v be optimal for the characterization of $\chi(A)$ in Proposition 2. Then for $d = \operatorname{sign}(v) \in \{-1, 1\}^m$ and $\hat{A} := \operatorname{Diag}(d)A \in \mathbb{S}(A)$ Proposition 3 implies that

$$H(\hat{A}) \geq \|v\| = \chi(A).$$

On the other hand, the construction of $\chi(A)$ and Proposition 1 imply that for all $\hat{A} \in \mathbb{S}(A)$

$$\chi(A) = \chi(\hat{A}) \geq H(\hat{A}).$$

Thus the first identity in (3) follows. To prove the second identity in (3), notice that $J \subseteq [2m]$ is such that $|J| = n$ and $\mathbf{A}_J$ non-singular if and only if there exists $I \subseteq [m]$ such that $|I| = n$, A_I is non-singular, and $J = I_+ \cup (m + I_-)$ for some partition $I = I_+ \cup I_-$ of I. If $d \in \{-1, 1\}^m$ satisfies $d_i = 1$, $i \in I_+$ and $d_i = -1$, $i \in I_-$ then

$$\max_{\substack{v\in\mathbb{R}^J_+ \\ \|\mathbf{A}_J^\mathsf{T} v\|=1}} \|v\| = \max_{\substack{v\in\mathbb{R}^I_+ \\ \|(\operatorname{Diag}(d)A)_I^\mathsf{T} v\|=1}} \|v\|.$$

Hence Proposition 3 implies that

$$H(\mathbf{A}) = \max_{\substack{J\subseteq[2m], |J|=n \\ \mathbf{A}_J \text{ non-singular}}} \max_{\substack{v\in\mathbb{R}^J_+ \\ \|\mathbf{A}_J^\mathsf{T} v\|=1}} \|v\| = \max_{\hat{A}\in\mathbb{S}(A)} \max_{\substack{I\subseteq[m], |I|=n \\ \hat{A}_I \text{ non-singular}}} \max_{\substack{v\in\mathbb{R}^I_+ \\ \|\hat{A}_I^\mathsf{T} v\|=1}} \|v\| = \max_{\hat{A}\in\mathbb{S}(A)} H(\hat{A}).$$

The second identity in (3) thus follows.

The crux of the proof of (4) is the following one-to-one correspondence between $\overline{\mathcal{J}}(\mathbf{A})$ and $\mathbb{D}(A)$.

Claim. Suppose all rows of A are nonzero. Then $J \in \overline{\mathcal{J}}(\mathbf{A})$ if and only if $J = ([m] \setminus I) \cup (m + I)$ for some $I \subseteq [m]$ such that $\hat{A} \in \mathbb{D}(A)$ where $\hat{A}$ is the matrix obtained by flipping the signs of the rows of A indexed by I.

This claim, Proposition 3, and Proposition 1 imply that

$$H(\mathbf{A}) = \max_{J\in\overline{\mathcal{J}}(\mathbf{A})} \max_{\substack{v\in\mathbb{R}^J_+ \\ \|\mathbf{A}_J^\mathsf{T} v\|=1}} \|v\| = \max_{\hat{A}\in\mathbb{D}(A)} \max_{\substack{v\in\mathbb{R}^m_+ \\ \|\hat{A}^\mathsf{T} v\|=1}} \|v\| = \max_{\hat{A}\in\mathbb{D}(A)} H(\hat{A}) = \max_{\hat{A}\in\mathbb{D}(A)} \frac{1}{\mathcal{R}(\hat{A})}. \tag{6}$$

The third step follows from Proposition 3 and the fact that $\overline{\mathcal{J}}(\hat{A}) = \{[m]\}$ if $\hat{A}x < 0$ is feasible. Identity (4) follows from (6) and (3).

To finish, here is a proof of the above claim. For $u \in \mathbb{R}^n$ let $J_u := \{j : \mathbf{A}_j u < 0\}$. Observe that $J \in \mathcal{J}(\mathbf{A})$ if and only if $J \subseteq J_u$ for some $u \in \mathbb{R}^n$. Since all rows of A are nonzero, it follows that $J \in \overline{\mathcal{J}}(\mathbf{A})$ if and only if $J = J_u$ for some $u \in \mathbb{R}^n$ such that all entries of $\mathbf{A}u$ are non-zero. When the latter holds, we have $J_u = ([m] \setminus I_u) \cup (m + I_u)$ for $I_u = \{i : A_i u > 0\}$ and $A_{[m]\setminus I_u} u < 0,\ A_{I_u} u > 0$ which is equivalent to $\hat{A} \in \mathbb{D}(A)$ where $\hat{A}$ is the matrix obtained by flipping the signs of the rows of A indexed by I_u.

□

4 Conclusion

We showed that if $A \in \mathbb{R}^{m\times n}$ has full column rank and nonzero rows then

$$\chi(A) = \max_{\hat{A}\in\mathbb{S}(A)} H(\hat{A}) = \max_{\hat{A}\in\mathbb{D}(A)} H(\hat{A}) = \max_{\hat{A}\in\mathbb{D}(A)} \frac{1}{\mathcal{R}(\hat{A})} = H(\mathbf{A}), \tag{7}$$

where $\mathbf{A} \in \mathbb{R}^{2m\times n}$ is the matrix obtained by stacking the rows of A and $-A$. The first expression in (7) takes the maximum over the collection of matrices $\mathbb{S}(A)$ which has exponential size in m. The second and third expressions in in (7) take the maximum over the smaller but harder to describe collection of matrices $\mathbb{D}(A)$. By contrast, the last expression in (7) is the Hoffman constant of the single matrix $\mathbf{A} \in \mathbb{R}^{2m\times n}$. The identity $\chi(A) = H(\mathbf{A})$ and the machinery developed in [27] provide a novel algorithmic procedure to compute or estimate $\chi(A)$. This computational capability in turn creates an avenue to identify families of linear programs that are solvable in polynomial time in the real model of computation via Vavasis-Ye's interior-point algorithm [23, 36].

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

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

[4] E. Bobrovnikova and S. Vavasis. Accurate solution of weighted least squares by iterative methods. *SIAM Journal on Matrix Analysis and Applications*, 22(4):1153–1174, 2001.

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

[6] D. Cheung, F. Cucker, and J. Peña. Unifying condition numbers for linear programming. *Math. Oper. Res*, 28(4):609–624, 2003.

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

[8] 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.

[9] A. Forsgren. On linear least-squares problems with diagonally dominant weight matrices. *SIAM Journal on Matrix Analysis and Applications*, 17(4):763–788, 1996.

[10] A. Forsgren and G. Sporre. On weighted linear least-squares problems related to interior methods for convex quadratic programming. *SIAM Journal on Matrix Analysis and Applications*, 23(1):42–56, 2001.

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

[12] 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.

[13] 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.

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

[15] 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.

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

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

[18] P. Hough and S. Vavasis. Complete orthogonal decomposition for weighted least squares. *SIAM Journal on Matrix Analysis and Applications*, 18(2):369–392, 1997.

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

[20] 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.

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

[22] 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.

[23] R. Monteiro and T. Tsuchiya. A variant of the Vavasis-Ye layered-step interior-point algorithm for linear programming. *SIAM Journal on Optimization*, 13(4):1054–1079, 2003.

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

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

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

[27] J. Peña, J. Vera, and L. Zuluaga. New characterizations of Hoffman constants for systems of linear constraints. *arXiv preprint arXiv:1905.02894*, 2019.

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

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

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

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

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

[33] M. Todd, L. Tunçel, and Y. Ye. Characterizations, bounds, and probabilistic analysis of two complexity measures for linear programming problems. *Mathematical Programming*, 90(1):59–69, 2001.

[34] L. Tunçel. On the condition numbers for polyhedra in Karmarkar's form. *Operations Research Letters*, 24(4):149–155, 1999.

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

[36] 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.

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

[38] Y. Ye. Toward probabilistic analysis of interior-point algorithms for linear programming. *Math. of Oper. Res.*, 19:38–52, 1994.

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