

Some preconditioners for systems of linear inequalities

Javier Peña* Vera Roshchina† Negar Soheili‡

June 20, 2013

Abstract

We show that a combination of two simple preprocessing steps would generally improve the conditioning of a homogeneous system of linear inequalities. Our approach is based on a comparison among three different but related notions of conditioning for linear inequalities.

1 Introduction

Condition numbers play an important role in numerical analysis. The condition number of a problem is a key parameter in the computational complexity of iterative algorithms as well as in issues of numerical stability. A related challenge of paramount importance is to *precondition* a given problem instance, that is, perform some kind of data preprocessing to transform a given problem instance into an equivalent one with that is better conditioned. Preconditioning has been extensively studied in numerical linear algebra [10, 15] and is an integral part of the computational implementation of numerous algorithms for solving linear systems of equations. In the more general optimization context, the task of designing preconditioners has been studied by Epelman and Freund [6] as well as by Belloni and Freund [3]. In a similar spirit, we propose two simple preconditioning procedures for homogeneous systems of linear inequalities. As we formalize in the sequel, these procedures lead to improvements in three types of condition

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

†Collaborative Research Network, University of Ballarat, AUSTRALIA, vroshchina@ballarat.edu.au

‡Tepper School of Business, Carnegie Mellon University, USA, nsoheili@andrew.cmu.edu

numbers for homogeneous systems of linear inequalities, namely Renegar’s [12], Goffin-Cucker-Cheung’s [5, 9], and the Grassmann condition number [1, 4]. Both Renegar’s and Goffin-Cheung-Chucker’s condition numbers are key parameters in the analyses of algorithmic schemes and other numerical properties of constraint systems [7, 8, 11–14]. The more recently developed Grassmann condition number is especially well-suited for probabilistic analysis [2].

We recall the definition of the above condition numbers in Section 2 below. These condition numbers quantify certain properties associated to the homogenous systems of inequalities (1) and (2) defined by a matrix $A \in \mathbb{R}^{m \times n}$, where $m \leq n$. Renegar’s condition number is defined in terms of the distance from A to a set of *ill-posed* instances in the space $\mathbb{R}^{m \times n}$. Goffin-Cucker-Cheung’s condition number is defined in terms of the *best conditioned* solution to the system defined by A . Alternatively, it can be seen as the reciprocal of a measure of thickness of the cone of feasible solutions to the system defined by A . Goffin-Cucker-Cheung’s condition number is intrinsic to certain geometry in the column space of A . The more recent Grassmann condition number, introduced in a special case by Belloni and Freund [4] and subsequently generalized by Amelunxen-Burgisser [1] is based on the projection distance from the linear subspace spanned by the rows of A to a set of *ill-posed* subspaces in \mathbb{R}^n . The Grassmann condition number is intrinsic to certain geometry in the row space of A .

The Goffin-Cucker-Cheung’s condition number and the Grassmann condition numbers are respectively invariant under column scaling and elementary row operations on the matrix A respectively. For suitable choices of norms, each of them is also always smaller than Renegar’s condition number (see (4) and (5) below). We observe that these properties have an interesting parallel and naturally suggest two preconditioning procedures, namely *column normalization* and *row balancing*. Our main results, presented in Section 3, discuss some interesting properties of these two preconditioning procedures. In particular, we show that a combination of them would improve the values of the three condition numbers.

2 Condition numbers and their relations

Given a matrix $A \in \mathbb{R}^{m \times n}$ with $m \leq n$, consider the homogeneous feasibility problem

$$Ax = 0, \quad x \geq 0, \quad x \neq 0, \quad (1)$$

and its alternative

$$A^T y \geq 0, \quad y \neq 0. \quad (2)$$

Let F_P and F_D denote the sets of matrices A such that (1) and (2) are respectively feasible. The sets F_P and F_D are closed and $F_P \cup F_D =$

$\mathbb{R}^{m \times n}$. The set $\Sigma := F_P \cap F_D$ is the set of *ill-posed* matrices. For a given $A \in \Sigma$, arbitrary small perturbations on A can lead to a change with respect to the feasibility of (1) and (2).

Renegar's condition number is defined as

$$\mathcal{C}_R(A) := \frac{\|A\|}{\min\{\|A - A'\| : A' \in \Sigma\}}.$$

Here $\|A\|$ denotes the operator norm of A induced by a given choice of norms in \mathbb{R}^n and \mathbb{R}^m . When the Euclidean norms are used in both \mathbb{R}^n and \mathbb{R}^m , we shall write $\mathcal{C}_R^{2,2}(A)$ for $\mathcal{C}_R(A)$. On the other hand, when the one-norm is used in \mathbb{R}^n and the Euclidean norm is used in \mathbb{R}^m we shall write $\mathcal{C}_R^{1,2}(A)$ for $\mathcal{C}_R(A)$. These are the two cases we will consider in the sequel.

Assume $A = [a_1 \ \cdots \ a_n] \in \mathbb{R}^{m \times n}$ with $a_i \neq 0$ for $i = 1, \dots, n$. Goffin-Cheung-Cucker's condition number is defined as

$$\mathcal{C}_{GCC}(A) := \min_{\|y\|_2=1} \max_{i \in \{1, \dots, n\}} \frac{\|a_i\|_2}{a_i^T y}.$$

Here $\|\cdot\|_2$ denotes the Euclidean norm in \mathbb{R}^m . The quantity $1/\mathcal{C}_{GCC}(A)$ is the Euclidean distance from the origin to the boundary of the convex hull of the set of normalized vectors $\left\{ \frac{a_i}{\|a_i\|_2}, i = 1, \dots, n \right\}$. Furthermore, when $A \in F_D$, this distance coincides with the thickness of the cone of feasible solutions to (2).

Let $\text{Gr}_{n,m}$ denote the set of m -dimensional linear subspaces in \mathbb{R}^n , that is, the m -dimensional Grassmann manifold in \mathbb{R}^n . Let

$$P_m := \{W \in \text{Gr}_{n,m} : W^\perp \cap \mathbb{R}_+^n \neq \{0\}\},$$

$$D_m := \{W \in \text{Gr}_{n,m} : W \cap \mathbb{R}_+^n \neq \{0\}\},$$

and

$$\Sigma_m = P_m \cap D_m = \{W \in \text{Gr}_{m,n} : W \cap \mathbb{R}_+^n \neq \{0\}, W \cap \text{int}(\mathbb{R}_+^n) = \emptyset\}.$$

The sets P_m, D_m , and Σ_m are analogous to the sets F_P, F_D , and Σ respectively: If $A \in \mathbb{R}^{m \times n}$ is full-rank then $A \in F_P \Leftrightarrow \text{span}(A^T) \in P_m$ and $A \in F_D \Leftrightarrow \text{span}(A^T) \in D_m$. In particular, $A \in \Sigma \Leftrightarrow \text{span}(A^T) \in \Sigma_m$. The Grassmann condition number is defined as [1]:

$$\mathcal{C}_{Gr}(A) := \frac{1}{\min\{d(\text{span}(A^T), W) : W \in \Sigma_m\}}.$$

Here $d(W_1, W_2) = \sigma_{\max}(\Pi_{W_1} - \Pi_{W_2})$, where Π_{W_i} denotes the orthogonal projection onto W_i for $i = 1, 2$. In the above expression and in the

sequel, $\sigma_{\max}(M)$ and $\sigma_{\min}(M)$ denote the largest and smallest singular values of the matrix M respectively.

We next recall some key properties of the condition numbers $\mathcal{C}_R^{1,2}$, $\mathcal{C}_R^{2,2}$, \mathcal{C}_{GCC} and \mathcal{C}_{Gr} . Throughout the remaining part of the paper assume $A = [a_1 \ \cdots \ a_n] \in \mathbb{R}^{m \times n}$ is full-rank and $a_i \neq 0$ for $i = 1, \dots, n$. From the relationship between the one-norm and the Euclidean norm, it readily follows that

$$\frac{\mathcal{C}_R^{2,2}(A)}{\sqrt{n}} \leq \mathcal{C}_R^{1,2}(A) \leq \sqrt{n} \mathcal{C}_R^{2,2}(A), \quad (3)$$

The following property is a consequence of [5, Theorem 1]:

$$\mathcal{C}_{GCC}(A) \leq \mathcal{C}_R^{1,2}(A) \leq \frac{\max_{i=1,\dots,n} \|a_i\|_2}{\min_{i=1,\dots,n} \|a_i\|_2} \mathcal{C}_{GCC}(A). \quad (4)$$

The following property was established in [1, Theorem 1.4]:

$$\mathcal{C}_{Gr}(A) \leq \mathcal{C}_R^{2,2}(A) \leq \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)} \mathcal{C}_{Gr}(A). \quad (5)$$

The inequalities (4) and (5) reveal an interesting parallel between the pairs $\mathcal{C}_{GCC}(A), \mathcal{C}_R^{1,2}(A)$ and $\mathcal{C}_{Gr}(A), \mathcal{C}_R^{2,2}(A)$. This parallel becomes especially striking by observing that the fractions in the right hand sides of (4) and (5) can be written respectively as

$$\frac{\max_{i=1,\dots,n} \|a_i\|_2}{\min_{i=1,\dots,n} \|a_i\|_2} = \frac{\max\{\|Ax\|_2 : \|x\|_1 = 1\}}{\min\{\|Ax\|_2 : \|x\|_1 = 1\}},$$

and

$$\frac{\sigma_{\max}(A)}{\sigma_{\min}(A)} = \frac{\max\{\|A^T y\|_2 : \|y\|_2 = 1\}}{\min\{\|A^T y\|_2 : \|y\|_2 = 1\}}.$$

3 Preconditioning via normalization and balancing

Inequalities (4) and (5) suggest two preconditioning procedures for improving Renegar's condition number $\mathcal{C}_R(A)$. The first one is to *normalize*, that is, scale the columns of A so that the preconditioned matrix \tilde{A} satisfies $\max_{i=1,\dots,n} \|\tilde{a}_i\| = \min_{i=1,\dots,n} \|\tilde{a}_i\|$. The second one is to *balance*, that is, apply an orthogonalization procedure to the rows of A such as Gram-Schmidt or QR-factorization so that the preconditioned matrix \tilde{A} satisfies $\sigma_{\max}(\tilde{A}) = \sigma_{\min}(\tilde{A})$. Observe that if the initial matrix A

is full rank and has non-zero columns, these properties are preserved by each of the above two preconditioning procedures. The following proposition formally states some properties of these two procedures.

Proposition 1 *Assume $A = [a_1 \ \cdots \ a_n] \in \mathbb{R}^{m \times n}$ is full-rank and $a_i \neq 0$ for $i = 1, \dots, n$.*

(a) *Let \tilde{A} be the matrix obtained after normalizing the columns of A , that is, $\tilde{a}_i = \frac{a_i}{\|a_i\|}$, $i = 1, \dots, n$. Then*

$$\mathcal{C}_R^{1,2}(\tilde{A}) = \mathcal{C}_{GCC}(\tilde{A}) = \mathcal{C}_{GCC}(A).$$

This transformation is optimal in the following sense

$$\mathcal{C}_R^{1,2}(\tilde{A}) = \min\{\mathcal{C}_R^{1,2}(AD) : D \text{ is diagonal positive definite}\}.$$

(b) *Let \tilde{A} be the matrix obtained after balancing the rows of A , that is, $\tilde{A} = Q^T$, where $QR = A^T$ with $Q \in \mathbb{R}^{n \times m}$, $R \in \mathbb{R}^{m \times m}$ is the QR-decomposition of A^T . Then*

$$\mathcal{C}_R^{2,2}(\tilde{A}) = \mathcal{C}_{Gr}(\tilde{A}) = \mathcal{C}_{Gr}(A).$$

This transformation is optimal in the following sense

$$\mathcal{C}_R^{2,2}(\tilde{A}) = \min\{\mathcal{C}_R^{2,2}(PA) : P \in \mathbb{R}^{m \times m} \text{ is non-singular}\}.$$

Proof: Parts (a) and (b) follow respectively from (4) and (5). ■

Observe that the normalization procedure transforms the solutions to the original problem (1) when this system is feasible. More precisely, observe that $Ax = 0$, $x \geq 0$ if and only if $\tilde{A}D^{-1}x = 0$, $D^{-1}x \geq 0$, where D is the diagonal matrix whose diagonal entries are the norms of the columns of A . Thus a solution to the original problem (1) can be readily obtained from a solution to the column-normalized preconditioned problem: premultiply by D . At the same time, observe that the solution to (2) does not change when the columns of A are normalized.

Similarly, the balancing procedure transforms the solutions to the original problem (2). In this case $A^T y \geq 0$, $y \neq 0$ if and only if $\tilde{A}^T R^{-1} y \geq 0$, $R^{-1} y \neq 0$, where $QR = A^T$ is the QR-decomposition of A^T . Hence a solution to the original problem (2) can be readily obtained from a solution to the row-balanced preconditioned problem: premultiply by R . At the same time, observe that the solution to (1) does not change when the rows of A are balanced.

We next present our main result concerning properties of the two combinations of normalization and balancing procedures.

Theorem 1 Assume $A = [a_1 \ \cdots \ a_n] \in \mathbb{R}^{m \times n}$ is full-rank and $a_i \neq 0$ for $i = 1, \dots, n$.

(a) Let \hat{A} be the matrix obtained after normalizing the columns of A and then balancing the rows of the resulting matrix. Then

$$\frac{1}{\sqrt{n}}\mathcal{C}_{GCC}(\hat{A}) \leq \mathcal{C}_R^{2,2}(\hat{A}) = \mathcal{C}_{Gr}(\hat{A}) \leq \sqrt{n}\mathcal{C}_{GCC}(A). \quad (6)$$

(b) Let \hat{A} be the matrix obtained after balancing the rows of A and then normalizing the columns of the resulting matrix. Then

$$\frac{1}{\sqrt{n}}\mathcal{C}_{Gr}(\hat{A}) \leq \mathcal{C}_R^{1,2}(\hat{A}) = \mathcal{C}_{GCC}(\hat{A}) \leq \sqrt{n}\mathcal{C}_{Gr}(A). \quad (7)$$

Proof:

(a) Let \tilde{A} be the matrix obtained by normalizing the columns of A . Proposition 1(a) yields

$$\mathcal{C}_R^{1,2}(\tilde{A}) = \mathcal{C}_{GCC}(\tilde{A}) = \mathcal{C}_{GCC}(A). \quad (8)$$

Since \hat{A} is obtained by balancing the rows of \tilde{A} , Proposition 1(b) yields

$$\mathcal{C}_R^{2,2}(\hat{A}) = \mathcal{C}_{Gr}(\hat{A}) = \mathcal{C}_{Gr}(\tilde{A}) \leq \mathcal{C}_R^{2,2}(\tilde{A}). \quad (9)$$

Inequality (6) now follows by combining (3), (4), (8), and (9).

(b) The proof this part is essentially identical to that of part (a) after using the parallel roles of the pairs $\mathcal{C}_{GCC}, \mathcal{C}_R^{1,2}$ and $\mathcal{C}_{Gr}, \mathcal{C}_R^{2,2}$ apparent from (4), (5), and Proposition 1. ■

Observe that both combined preconditioners in Theorem 1 transform A to a $\hat{A} = PAD$ where D is a diagonal matrix and P is a square non-singular matrix. The particular P and D depend on what procedure is applied first. Hence each of the combined preconditioners transforms both sets of solutions to the original pair of problems (1) and (2). In this case, solutions to the original problems can be readily obtained from solutions to the preconditioned problems via premultiplication by D for the solutions to (1) and by P^T for the solutions to (2).

As discussed in [1, Section 5], there is no general relationship between \mathcal{C}_{Gr} and \mathcal{C}_{GCC} , meaning that either one can be much larger than the other in a dimension-independent fashion. Theorem 1(a) shows that when $\mathcal{C}_{GCC}(A) \ll \mathcal{C}_{Gr}(A)$ column normalization followed by row balancing would reduce both \mathcal{C}_{Gr} and $\mathcal{C}_R^{1,2}$ while not increasing \mathcal{C}_{GCC} beyond a factor of \sqrt{n} . Likewise, Theorem 1(b) shows that when $\mathcal{C}_{Gr}(A) \ll \mathcal{C}_{GCC}(A)$ row balancing followed by column normalization

would reduce both \mathcal{C}_{GCC} and $\mathcal{C}_R^{1,2}$ while not increasing \mathcal{C}_{Gr} beyond a factor of \sqrt{n} . In other words, one of the two combinations of column normalization and by row balancing will always improve the values of all three condition numbers $\mathcal{C}_{GCC}(A)$, $\mathcal{C}_{Gr}(A)$ and $\mathcal{C}_R(A)$ modulo a factor of \sqrt{n} . The following two questions concerning a potential strengthening of Theorem 1 naturally arise:

- (i) Does the order of row balancing and column normalization matter in each part of Theorem 1?
- (ii) Are the leftmost and rightmost expressions in (6) and (7) within a small power of n ? In other words, do the combined preconditioners make all $\mathcal{C}_{GCC}(A)$, $\mathcal{C}_{Gr}(A)$ and $\mathcal{C}_R(A)$ the same modulo a small power of n ?

The examples below, adapted from [1, Section 5], show that the answers to these questions are ‘yes’ and ‘no’ respectively. Hence without further assumptions, the statement of Theorem 1 cannot be strengthened along these lines.

Example 1 Assume $\epsilon > 0$ and let $A = \begin{bmatrix} 2/\epsilon & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix}$. It is easy to see that $\mathcal{C}_{GCC}(A) = \sqrt{2}$. After balancing the rows of A and normalizing the columns of the resulting matrix we get

$$\hat{A} = \frac{1}{\sqrt{2(1+\epsilon^2)}} \begin{bmatrix} \sqrt{2(1+\epsilon^2)} & \epsilon & \epsilon \\ 0 & -\sqrt{2+\epsilon^2} & \sqrt{2+\epsilon^2} \end{bmatrix}.$$

It is easy to show that $\mathcal{C}_{GCC}(\hat{A}) = \frac{\sqrt{2(1+\epsilon^2)}}{\epsilon}$. Thus for $\epsilon > 0$ sufficiently small, $\mathcal{C}_{GCC}(\hat{A})$ can be arbitrarily larger than $\mathcal{C}_{GCC}(A)$. Therefore (6) does not hold for A, \hat{A} .

Example 2 Assume $0 < \epsilon < 1/2$ and let $A = \begin{bmatrix} -\epsilon & -1 & 1 \\ 0 & -1 & 1+\epsilon \end{bmatrix}$. Using [1, Proposition 1.6], it can be shown that $\mathcal{C}_{Gr}(A) = \sqrt{2 + (1+\epsilon)^2}$. After normalizing the columns of A and balancing the rows of the resulting matrix, we obtain

$$\hat{A} = \frac{1}{\sqrt{\delta}} \begin{bmatrix} -\theta\delta & -\sqrt{2}\theta\epsilon(1+\epsilon) & -\theta\epsilon\sqrt{1+(1+\epsilon)^2} \\ 0 & -\sqrt{1+(1+\epsilon)^2} & \sqrt{2}(1+\epsilon) \end{bmatrix}.$$

where $\delta = 1 + 3(1+\epsilon)^2$ and $\theta = \frac{1}{\sqrt{\delta+\epsilon^2}}$. Using [1, Proposition 1.6] again, it can be shown that $\mathcal{C}_{Gr}(\hat{A}) = \sqrt{1 + \frac{\delta}{\epsilon^2}}$. Therefore, $\mathcal{C}_{Gr}(\hat{A})$ can be arbitrarily larger than $\mathcal{C}_{Gr}(A)$. Hence (7) does not hold for A, \hat{A} .

Example 3 Assume $\epsilon > 0$ and let $A = \begin{bmatrix} 1 + \epsilon & 1 + \epsilon & -1 + \epsilon \\ -1 & -1 & 1 \end{bmatrix}$. In this case $\mathcal{C}_{GCC}(A) = \frac{\sqrt{2+2(1+\epsilon)^2}}{\epsilon}$. After normalizing the columns of A and balancing the rows of the resulting matrix, we get

$$\hat{A} = \frac{1}{\sqrt{2\gamma^2 + 2\beta^2}} \begin{bmatrix} \beta & \beta & \sqrt{2}\gamma \\ -\gamma & -\gamma & \sqrt{2}\beta \end{bmatrix},$$

where $\beta = \sqrt{\frac{1+(1+\epsilon)^2}{2}}$ and $\gamma = \sqrt{1 + (1 - \epsilon)^2}$. It is easy to see that $\mathcal{C}_{GCC}(\hat{A}) = \sqrt{2}$. Therefore, $\mathcal{C}_{GCC}(\hat{A})$ can be arbitrarily smaller than $\mathcal{C}_{GCC}(A)$ in (6).

Example 4 Assume $0 < \epsilon < 1/2$ and let $A = \begin{bmatrix} 2\epsilon & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix}$. In this case $\mathcal{C}_{Gr}(A) = \sqrt{1 + \frac{1}{2\epsilon^2}}$. After balancing the rows of A and normalizing the columns of the resulting matrix we get

$$\hat{A} = \frac{1}{\sqrt{2(1 + \epsilon^2)}} \begin{bmatrix} \sqrt{2(1 + \epsilon^2)} & 1 & 1 \\ 0 & -\sqrt{1 + 2\epsilon^2} & \sqrt{1 + 2\epsilon^2} \end{bmatrix}.$$

Using [1, Proposition 1.6], it can be shown that $\mathcal{C}_{Gr}(\hat{A}) = \sqrt{2 + \epsilon^2}$. Thus $\mathcal{C}_{Gr}(\hat{A})$ can be arbitrarily smaller than $\mathcal{C}_{Gr}(A)$ in (7).

References

- [1] D. Amelunxen and P. Bürgisser. A coordinate-free condition number for convex programming. *SIAM Journal on Optimization*, 22(3):1029–1041, 2012.
- [2] D. Amelunxen and P. Bürgisser. Probabilistic analysis of the Grassmann condition number. Technical report, University of Paderborn, Germany, 2013.
- [3] A. Belloni and R. Freund. A geometric analysis of renegars condition number, and its interplay with conic curvature. *Math. Program.*, 119:95–107, 2009.
- [4] A. Belloni and R. Freund. Projective re-normalization for improving the behavior of a homogeneous conic linear system. *Math. Program.*, 118:279–299, 2009.
- [5] D. Cheung and F. Cucker. A new condition number for linear programming. *Mathematical programming*, 91(1):163–174, 2001.
- [6] 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.

- [7] M. Epeleman and R. M. Freund. Condition number complexity of an elementary algorithm for computing a reliable solution of a conic linear system. *Math. Program.*, 88(3):451–485, 2000.
- [8] R. Freund and J. Vera. Condition-based complexity of convex optimization in conic linear form via the ellipsoid algorithm. *SIAM J. on Optim.*, 10:155–176, 1999.
- [9] J. L. Goffin. The relaxation method for solving systems of linear inequalities. *Mathematics of Operations Research*, pages 388–414, 1980.
- [10] G. Golub and C. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, third edition, 1996.
- [11] J. Peña and J. Renegar. Computing approximate solutions for convex conic systems of constraints. *Math Program.*, 87:351–383, 2000.
- [12] J. Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM Journal on Optimization*, 5(3):506–524, 1995.
- [13] J. Renegar. Linear programming, complexity theory and elementary functional analysis. *Mathematical Programming*, 70(1-3):279–351, 1995.
- [14] N. Soheili and J. Peña. A smooth perceptron algorithm. *SIAM Journal on Optimization*, 22(2):728–737, 2012.
- [15] L. Trefethen and D. Bau. *Numerical Linear Algebra*. SIAM, 1997.