Lawrence Berkeley National Laboratory report LBNL-50635

A Matrix Lower Bound

Joseph F. Grcar

Lawrence Berkeley National Laboratory
Mail Stop 50A-1148
One Cyclotron Road
Berkeley, CA 94720-8142 USA
e-mail: jfgrcar@lbl.gov

Abstract

A matrix lower bound is defined that generalizes ideas apparently due to S. Banach and J. von Neumann. The matrix lower bound has a natural interpretation in functional analysis, and it satisfies many of the properties that von Neumann stated for it in a restricted case.

Applications for the matrix lower bound are demonstrated in several areas. In linear algebra, the matrix lower bound of a full rank matrix equals the distance to the set of rank-deficient matrices. In numerical analysis, the ratio of the matrix norm to the matrix lower bound is a condition number for all consistent systems of linear equations. In optimization theory, the matrix lower bound suggests an identity for a class of min-max problems. In real analysis, a recursive construction that depends on the matrix lower bound shows that the level sets of continuously differentiable functions lie asymptotically near those of their tangents.

Keywords: condition number,
distance to rank deficiency,
functional analysis in matrix theory,
implicit function theorem,
level sets,
matrix inequalities,
matrix lower bound,
min-max problems,
triangle inequality

2000 MSC, primary:	15A60	(norms of matrices, applications of functional analysis to matrix theory);
secondary:	00A07	(problem books),
	15A45	(inequalities involving matrices),
application 1:	15A03	(linear dependence, rank),
application 2:	65F35	(matrix norms, conditioning),
application 3:	90C47	(minimax problems),
application 4:	26B10	(implicit function theorems)

Contents

List of Figures

This paper is dedicated to John von Neumann at the centennial of his birth, on 28th December 1903.

1 Introduction

John von Neumann and Herman Goldstine defined a "lower bound of a matrix" [16, p. 1042] and listed several of its properties, which they believed were "too well known to require much discussion." The value they arrived at for the lower bound of a square matrix is the smallest singular value. This value is not surprising, but its derivation and properties are. They reveal a symmetry with the matrix norm that seems to have gone unnoticed except by von Neumann and Goldstine.

Similar values for matrix lower bounds appeared after von Neumann and Goldstine's paper. D. K. Faddeev and V. N. Faddeeva [6, p. 109] noted the same value as a lower bound for the same class of matrices. A. S. Householder [10, p. 48] defined a lower bound for square matrices, which in the spectral case equals von Neumann and Goldstine's, and in the most general case is still zero for a singular matrix. None of these authors cite any literature about matrix lower bounds, but all seem to be aware of earlier sources. Whatever its origin, evidently the concept is a natural one that bears further investigation.

Here, a nonzero lower bound is proposed for *all* nonzero matrices of *any* shape. This new lower bound generalizes the earlier definitions for square matrices, when they are not zero, but it differs from them because it is never zero. Section 2 examines four places where matrix lower bounds arise, and proves that all the bounds are the same. Section 4 establishes properties for this matrix lower bound that von Neumann and Goldstine intuited in their special case. Section 5 demonstrates applications of the matrix lower bound in linear algebra, numerical analysis, optimization theory, and real analysis. Section 7 lists some open questions.

The matrix lower bound is a missing chapter in mathematics. Section 5 demonstrates that the lower bound finds applications in many different kinds of theorems and proofs. These uses benefit from recognizing that the same phenomenon appears in a variety of contexts, so it can be studied independently, which facilitates its fullest exploitation. Indeed, Sections 2 through 4 show that the matrix lower bound has a systematic collection of rules and properties which are analogous to but different from the matrix norm's. Finally, the lower bound has a tantalizing mixture of algebraic and analytic properties: it is the norm of a basic canonical mapping associated with all quotient spaces, and it is the distance to the set of rank deficient matrices. Thus, one suspects, the matrix lower bound has independent interest.

2 A Matrix Lower Bound

A matrix lower bound may be defined in terms of matrix analysis, convex sets, min-max optimization, and functional analysis. The first approach is the most easily motivated.

2.1 Definition

We will be dealing with: $\mathbb{R}^m$ and $\mathbb{R}^n$ as spaces of column vectors, with norms on these spaces, and with the associated norms for $m \times n$ and $n \times m$ matrices. The matrix norms are the ones that are variously called consistent, induced, operator, or subordinate. It will not be confusing to use the same notation for all the norms.

From the standpoint of establishing bounds and estimates, the principal use of the matrix norm,

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} \,,$$

is that for any $Ax = y$ there is an inequality of the form,

$$\|A\| \, \|x\| \geq \|y\| \,. \tag{1}$$

Von Neumann and Goldstine referred to $\|A\|$ as the "upper bound" of the matrix.

The idea for the lower bound is that inequality (1) might be reversed provided $\|A\|$ is replaced by some other number. Von Neumann and Goldstine restricted A to be square, so that for any nonsingular $Ax = y$ it is true that

$$\|A^{-1}\|^{-1} \, \|x\| \leq \|y\| \,.$$

Furthermore, they dealt with 2-norms so their $\|A^{-1}\|^{-1}$ is just the smallest singular value of A. Rather than choosing outright this numerical value for a lower bound, however, von Neumann and Goldstine derived the value and its properties. (The starting point for their derivations, $\min_{\|x\|=1} \|Ax\|$, will not be used here.) Among their findings is that their lower bound is the largest number that satisfies the reverse of equation (1) for square, nonsingular matrices. This suggests that, if more general lower bounds could be defined for a broader class of matrices, then the largest of them might satisfy some version of von Neumann and Goldstine's interrelated properties.

Stefan Banach supplied lower bounds ready for this purpose. Let $\mathcal{B}$ be the open ball of center 0 and radius 1 in whatever space is indicated. If $T : X \to Y$ is a continuous linear transformation from one Banach space onto another, then Banach proved [1, p. 38, chapter 3, equation (1)] that for every $\epsilon > 0$ there is a $\delta > 0$ so that $\delta \mathcal{B}_Y \subseteq T(\epsilon \mathcal{B}_X)$. Thus if $\|y\| = \delta$, then there is an x with $y = T(x)$ and $\|x\| \le \epsilon$, hence

$$\frac{\delta}{\epsilon}\,\|x\| \le \|y\|\,.$$

The restriction on $\|y\|$ can be discarded by jointly scaling x and y. In this way inequality (1) can be reversed in very general circumstances provided there is an *extra* qualification [1, p. 150, chapter 10, theorem 10]: for any y in the column space of A, there is *some* x with $Ax = y$ and

$$m\,\|x\| \le \|y\|\,,$$

where m is a number independent of x and y.

These developments may be summarized as follows. Banach showed that nonzero lower bounds always exist provided they are limited in scope to some x, while von Neumann and Goldstine found that the *greatest* lower bounds are most interesting. Combining these ideas and adopting von Neumann and Goldstine's notation and terminology leads to the following definition.

Definition 2.1 (Matrix Lower Bound) *(Compare [1, p. 150, chapter 10, theorem 10] and [16, p. 1042, equation (3.2.b)].) Let A be a nonzero matrix. The matrix lower bound, $\|A\|_\ell$, is the largest of the numbers, m, such that for every y in the column space of A, there is some x with $Ax = y$ and $m\,\|x\| \le \|y\|$.*

In establishing various alternative definitions for the matrix lower bound it will be convenient to have the set

$$\mathcal{M}_\ell(A) = \{m : \forall y \in \mathrm{col}(A),\, \exists x \text{ so } Ax = y \text{ and } m\|x\| \le \|y\|\}\,,$$

for which Definition 2.1 says $\|A\|_\ell = \max \mathcal{M}_\ell(A)$. It is straightforward to see that $\mathcal{M}_\ell(A)$ is nonempty and bounded above, but a little proof is needed to justify Definition 2.1's assertion that $\mathcal{M}_\ell(A)$ contains its supremum.

Lemma 2.2 (Existence and Bounds) *The matrix lower bound exists and is positive. In particular,*

$$\|B\|^{-1} \le \|A\|_\ell \le \|A\|$$

where B is any generalized inverse in the sense that $ABA = A$.

Proof. (Step 1.) It is always possible to find a matrix B that satisfies the condition $A = ABA$; the pseudoinverse of A is one. If $y \in \mathrm{col}(A)$, then let $x = By$ so $\|x\| \le \|B\|\,\|y\|$. The condition $ABA = A$ implies $Ax = ABy = y$. This is for any y, so $\|B\|^{-1} \in \mathcal{M}_\ell(A)$.

(Step 2.) Let $m \in \mathcal{M}_\ell(A)$ and choose some nonzero $y \in \mathrm{col}(A)$. For these choices there is some x with $Ax = y$ and $m\|x\| \le \|y\|$. Thus $m \le \|y\|/\|x\| = \|Ax\|/\|x\| \le \|A\|$.

(Step 3.) It remains to be shown that $\mathcal{M}_\ell(\mathcal{A})$ contains its positive cluster points and hence its supremum. Suppose $m_* > 0$ is the limit of a sequence $\{m_n\} \subseteq \mathcal{M}_\ell(\mathcal{A})$. If $y \in \mathrm{col}(A)$, then for each m_n there is a x_n with $Ax_n = y$ and $m_n\|x_n\| \le \|y\|$. These inequalities and the fact that $\{m_n\}$ converges to a positive limit imply that the sequence $\{x_n\}$ is bounded. Therefore $\{x_n\}$ has a convergent subsequence with limit x_*. Passing to the limit shows that $Ax_* = y$ and $m_*\|x_*\| \le \|y\|$. Since this is for any y, so $m_* \in \mathcal{M}_\ell(A)$. ∎

2.2 Geometric Formulation

The manner in which Banach derived lower bounds, explained in Section 2.1, suggests that the matrix lower bound has a geometric interpretation. Let A be an $m \times n$ matrix, and let $\mathcal{B}_m$ and $\mathcal{B}_n$ be the open unit balls centered at the origin in $\mathbb{R}^m$ and $\mathbb{R}^n$. Matrix-vector multiplication carries $\mathcal{B}_n$ to a convex set $A\mathcal{B}_n$ around the origin in $\mathbb{R}^m$. Banach and geometric intuition say this convex set contains balls; and Lemma 2.3 shows $\|A\|_\ell$ is the radius of the largest. However, these are balls only relative to the subspace $\mathrm{col}(A)$ and not with respect to all of $\mathbb{R}^m$ (unless $\mathrm{col}(A) = \mathbb{R}^m$). In convex analysis this concept is known as the relative interior. This is in contrast to $\|A\|$, which is the radius of the smallest enclosing ball with respect to either space $\mathrm{col}(A)$ or $\mathbb{R}^m$.

Lemma 2.3 (Geometric Characterization) *(Compare [1, p. 38, chapter 3, equation (1)] and [10, p. 48, equation (6)].) Let A be a nonzero $m \times n$ matrix, and let $\mathcal{B}_m$ and $\mathcal{B}_n$ be the unit balls in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. The matrix lower bound $\|A\|_\ell$ is the radius of the largest ball — with respect to the subspace* $\mathrm{col}(A)$ *— that is centered at the origin and contained in $A\mathcal{B}_n$,*

$$\|A\|_\ell = \max\{\, r : \mathrm{col}(A) \cap r\mathcal{B}_m \subseteq A\mathcal{B}_n \}\ .$$

Proof. Let $\mathcal{R}$ be the set in the Lemma. If $y \in \mathrm{col}(A) \cap \|A\|_\ell\mathcal{B}_m$, then by Definition 2.1 there is an x with $Ax = y$ and $\|A\|_\ell\,\|x\| \le \|y\| < \|A\|_\ell$. Thus $\|x\| < 1$ so $y \in A\mathcal{B}_n$ hence $\mathrm{col}(A) \cap \|A\|_\ell\mathcal{B}_m \subseteq A\mathcal{B}_n$ and then $\|A\|_\ell \in \mathcal{R}$.

Suppose $r \in \mathcal{R}$ and $r > 0$. It is always necessary to deal separately with the 0 case: if $y = 0$, then $A0 = y$ and $r0 \le \|y\|$. In the main case, if $y \in \mathrm{col}(A)$ and $y \ne 0$, then $yr/\|y\| \in \mathrm{col}(A) \cap \mathrm{cl}\,(r\mathcal{B}_m)$. The containment $\mathrm{col}(A) \cap r\mathcal{B}_m \subseteq A\mathcal{B}_n$ applies as well to the closures of these sets; thus there is an $x \in \mathrm{cl}\,(\mathcal{B}_n)$ with $Ax = yr/\|y\|$. Therefore $A(x\|y\|/r) = y$ where $r\|(x\|y\|/r)\| = \|x\|\,\|y\| \le \|y\|$. This is for any y, so $r \in \mathcal{M}_\ell(A)$, hence $r \le \|A\|_\ell$. In summary, $\|A\|_\ell$ is the largest member of $\mathcal{R}$. ∎

2.3 Min-max Formulation

The use of the matrix lower bound lies in the application of the property stated in Definition 2.1, but the bound can be characterized in a more formulaic way that is sometimes useful.

Lemma 2.4 (Min-max Characterization) *(Compare [16, p. 1042, equation (3.1.b)].) If A is a nonzero matrix, then*

$$\begin{aligned} \|A\|_\ell &= \min_{y \,\in\, \mathrm{col}(A)\backslash\{0\}} \; \max_{\{x \,:\, Ax = y\}} \frac{\|y\|}{\|x\|} \\ &= \min_{\{x \,:\, Ax \neq 0\}} \; \max_{\{z \,:\, Az = 0\}} \frac{\|Ax\|}{\|x+z\|} \,. \end{aligned} \tag{2}$$

Proof. The second formula needs no proof as it is a simple restatement of the first. (Step 1.) If $y \neq 0$ and $y \in \mathrm{col}(A)$, then $\{x : Ax = y\}$ is not empty. Since this set is closed, it must contain a point nearest the origin. At that point the maximum

$$M(y) = \max_{\{x \,:\, Ax = y\}} \frac{\|y\|}{\|x\|} \tag{3}$$

is attained. This function $M(y)$ is used throughout the proof.

(Step 2.) Let

$$m = \inf_{y \,\in\, \mathrm{col}(A)\backslash\{0\}} M(y)\,. \tag{4}$$

That $m = \|A\|_\ell$ is straightforward: for every y, Definition 2.1 says the ratio at which $M(y)$ is attained must be bounded below by $\|A\|_\ell$. Therefore the infemum of the $M(y)$ is also bounded below by $\|A\|_\ell$. To prove the reverse inequality, suppose $y \in \mathrm{col}(A)$. If $y = 0$, then $A0 = y$ and $m\|0\| \leq \|y\|$. If $y \neq 0$, then again consider some x at which $M(y)$ is attained. From this,

$$m \leq M(y) = \frac{\|y\|}{\|x\|}\,,$$

follows $m\|x\| \leq \|y\|$, which is for any y, so $m \in \mathcal{M}_\ell(A)$.

(Step 3.) It remains to be shown that the infemum, m, is attained. This is a surprisingly delicate argument. It is always possible to choose a sequence (a) $\{M(y_n)\}$ that converges to m. All that is known of the y_n is that $y_n \in \mathrm{col}(A)$ and $y_n \neq 0$.

Equation (3)'s objective function, $\|y_n\|/\|x\|$, and constraints, $Ax = y_n$, are unchanged by any joint scaling of y_n and the x, so without loss of generality it is possible to assume that $\|y_n\| = 1$. The sequence $\{y_n\}$ is then bounded, and hence has a convergent subsequence (b) $\{y_{n'}\}$ with limit y_*. The conditions

$y_{n'} \in \mathrm{col}(A)$ and $\|y_{n'}\| = 1$ are inherited by the limit, so that $M(y_*)$ is well-defined. Let (c) $M(y_*)$ be attained at x_*.

Since $y_{n'} - y_* \in \mathrm{col}(A)$, there is an $x_{n'}$ with $Ax_{n'} = y_{n'}$ and

$$\|A\|_\ell \, \|x_{n'} - x_*\| \leq \|y_{n'} - y_*\| \, . \tag{5}$$

This inequality is gotten by applying Definition 2.1 to $y = y_{n'} - y_*$ to find an x with $Ax = y$ and $\|A\|_\ell \, \|x\| \leq \|y\|$; then $x_{n'}$ is chosen so $x = x_{n'} - x_*$. In any case, equation (5) implies that (d) $\{x_{n'}\}$ converges to x_*. Moreover, by the definition of $M(y_{n'})$,

$$\frac{\|y_{n'}\|}{\|x_{n'}\|} \leq \max_{\{x \,:\, Ax \,=\, y_{n'}\}} \frac{\|y_{n'}\|}{\|x\|} = M(y_{n'}) \, . \tag{6}$$

Altogether now — from (b and d), equation (6), (a), equation (4), and (c),

$$\frac{\|y_*\|}{\|x_*\|} = \lim_{n' \to \infty} \frac{\|y_{n'}\|}{\|x_{n'}\|} \leq \lim_{n' \to \infty} M(y_{n'}) = m \leq M(y_*) = \frac{\|y_*\|}{\|x_*\|} \, ,$$

which proves that $m = M(y_*)$. ∎

Corollary 2.5 (Attainment) *There is a nonzero x with $\|A\|_\ell \, \|x\| = \|Ax\|$.*

Proof. In Lemma 2.4's equation (2), first choose y that attains the minimum, and then choose x that attains the maximum for this y. ∎

2.4 Functional Analysis

There is a simple, abstract interpretation of the matrix lower bound, but some background is needed to get at it.

If U is a Banach space and S is a subspace, then the quotient space U/S consists of the equivalence classes defined by $u \equiv u' \Leftrightarrow u - u' \in S$. The notation $[u] = S + u$ is used for the class that contains u. If S is closed, then U/S is a Banach space under the quotient norm,

$$\|[u]\| = \inf_{u' \,\in\, [u]} \|u'\| \, .$$

The canonical function $\phi : U \to U/S$ by $\phi(u) = [u]$ is continuous, linear, and has operator norm $\|\phi\| = 1$.

If $T : U \to V$ is a continuous linear transformation among Banach spaces, then $\ker(T)$ is closed, so $U/\ker(T)$ is a Banach space, and there is a further canonical function $\psi : U/\ker(T) \to \mathrm{im}\,(T)$ by $\psi([u]) = T(u)$. This function is continuous, linear and invertible. Moreover, if $\mathrm{im}\,(T)$ is closed, then the inverse map ψ^{-1} is continuous [1, p. 41, chapter 3, theorem 5], so it has an operator norm, see Figure 1.

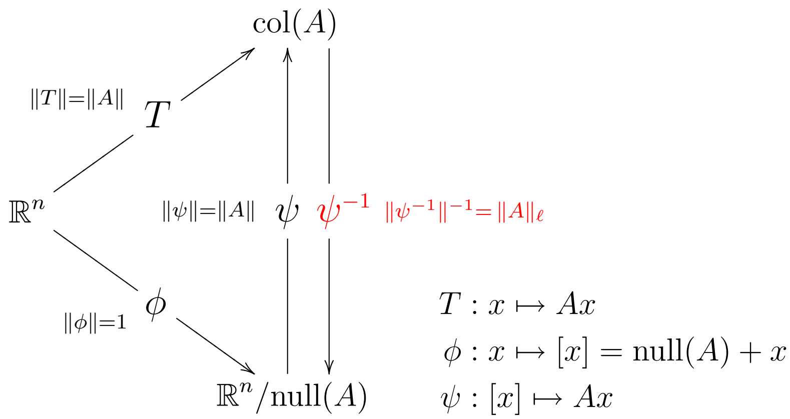

Figure 1: *The commutative diagram illustrating Lemma 2.6.*

Lemma 2.6 (Functional Analysis Characterization) *Let A be a nonzero $m \times n$ matrix, and let $\psi : \mathbb{R}^n/\text{null}(A) \to \text{col}(A)$ be the canonical map defined by $\psi([x]) = Ax$. Then*

$$\|\psi^{-1}\|^{-1} = \|A\|_\ell$$

where the norm of ψ^{-1} is the operator norm induced from the norms on col(A) *and* $\mathbb{R}^n/\text{null}(A)$.

Proof. All the spaces have finite dimension so the discussion preceding the Lemma applies. Only the evaluation of the norm remains,

$$\|\psi^{-1}\| = \max_{y \,\in\, \text{col}(A)\setminus\{0\}} \frac{\|\psi^{-1}(y)\|}{\|y\|} = \max_{y \,\in\, \text{col}(A)\setminus\{0\}} \; \min_{\{x \,:\, Ax = y\}} \frac{\|x\|}{\|y\|} \,.$$

Reciprocating this expression replaces the max-min by a min-max and inverts the ratio. Hence $\|\psi^{-1}\|^{-1} = \|A\|_\ell$ by Lemma 2.4. ∎

3 Examples

There are explicit formulas for the matrix lower bounds of invertible matrices, of rank 1 matrices, and with respect to spectral norms. For illustration, each of these formulas is established by a different method of proof. The spectral lower bound of a symmetric, positive definite matrix also has an interesting upper bound that is due directly to von Neumann and Goldstine.

Lemma 3.1 (Invertible Matrices) *(Compare [16, p. 1043, equations (3.5.d–e)].) If A is invertible, then $\|A\|_\ell = \|A^{-1}\|^{-1}$.*

Proof. Choose $y \neq 0$ at which $\|A^{-1}\|$ is attained, $\|A^{-1}\| = \|A^{-1}y\|/\|y\|$. Since there is exactly one x with $Ax = y$, it must be that $x = A^{-1}y$ satisfies Definition 2.1. From

$$\|A\|_\ell \, \|A^{-1}y\| \leq \|y\| = \|A^{-1}y\|/\|A^{-1}\|$$

follows $\|A\|_\ell \leq \|A^{-1}\|^{-1}$. Lemma 2.2 gives the reverse inequality. ∎

Lemma 3.2 (Rank 1 Matrices) *If A is a rank 1 matrix, where $A = uv^t$ for some $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, then $\|A\|_\ell = \|u\| \, \|v\|^*$, where $\|\cdot\|^*$ is the dual of the vector norm for $\mathbb{R}^n$. (The dual norm is discussed below in Section 4.1.)*

Proof. Let $\mathcal{B}_n$ be the open unit ball in $\mathbb{R}^n$. Since A has rank 1, therefore $A\mathcal{B}_n$ is an interval of radius $\|A\|$, which is a 1-dimensional ball. Thus $\|A\|_\ell = \|A\|$ by Lemma 2.3. It is well known that,

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{x \neq 0} \frac{\|uv^t x\|}{\|x\|} = \|u\| \max_{x \neq 0} \frac{|v^t x|}{\|x\|} = \|u\| \, \|v\|^* \, .$$

∎

Lemma 3.3 (Spectral Norms) *(Compare [16, p. 1046, equation (3.22.b)].) If A is not zero, and if the vector norms are 2-norms, then $\|A\|_\ell = \sigma_{\min}$, where $\sigma_{\min}$ is the smallest nonzero singular value of A.*

Proof. Let $A = U\Sigma V^t$ be the singular value decomposition of A: U and V are orthogonal matrices, and Σ is a diagonal matrix whose diagonal entries are the singular values, $\Sigma_{i,i} = \sigma_i$. If $y \in \mathrm{col}(A) \backslash \{0\}$, then $y = Ax$ for an x of the form $x = \sum_{i'} c_{i'} v_{i'}$, where i' indexes only the nonzero singular values, and where not all the coefficients $c_{i'}$ vanish. Altering the coefficients would change y; adding to x any vector orthogonal to the $v_{i'}$ would increase $\|x\|$; therefore for this y it must be that,

$$\max_{\{x \, : \, Ax = y\}} \frac{\|y\|}{\|x\|} = \left(\frac{\sum_{i'} \sigma_{i'}^2 \, c_{i'}^2}{\sum_{i'} c_{i'}^2} \right)^{1/2} .$$

This ratio is minimized by choosing $c_{i'} \neq 0$ exactly for the smallest $\sigma_{i'}$. The smallest singular value is then the minimum over all y. This is $\|A\|_\ell$ by Lemma 2.4's equation (2). ∎

Corollary 3.4 (Separation) *(Compare [16, p. 1045, equation (3.17.b); and p. 1073, equation (6.47)].) If A is symmetric and positive definite, and if the vector norms are 2-norms, then*

$$\|A\|_\ell \leq \max_i A_{i,i} \leq \|A\| \, .$$

Proof. From Lemma 3.3, $\|A\|_\ell$ is the smallest singular value of A. Because A is symmetric and positive definite, the singular values are the eigenvalues, $\lambda_n \geq \cdots \geq \lambda_2 \geq \lambda_1 = \|A\|_\ell > 0$. Von Neumann and Goldstine calculated

$$n\|A\|_\ell \leq \sum_{i=1}^{n} \lambda_i = \mathrm{trace}(A) = \sum_{i=1}^{n} A_{i,i} \leq n \max_i A_{i,i}\,.$$

The lower bound on $\|A\|$ comes from the fact that $\|A\|$ maximizes $u^t A v$ for all unit vectors u and v [16, p. 1043, equation (3.10)]. ∎

4 Properties

Von Neumann and Goldstine [16, pp. 1042–5, section 3.2] list several properties of the matrix lower bound. These exhibit a symmetry with the matrix norm whose aesthetic is all the more evident because they are not interrupted with proofs. The proofs are straightforward for the case that von Neumann and Goldstine considered; their contribution is recognizing and stating the properties.

With some effort the properties also can be established for Definition 2.1's lower bound. The first property, the transpose identity, has a proof that uses Lemma 2.6's characterization of the lower bound in terms of functional analysis. The next two groups of properties are inequalities for addition and multiplication. Many of these inequalities are not universally true, so some ingenuity is needed to find the most general hypotheses. The final group of properties are inner product identities.

In establishing these properties it sometimes will be assumed that a given matrix has full row rank. This is used in the following way. For an $m \times n$ matrix A to have full row rank means that the rows of A are linearly independent, so the matrix has rank m. The column space of A then has dimension m hence it must be all of $\mathbb{R}^m$. Thus, "A has full row rank" means "$Ax = y$ is consistent for every y."

4.1 Transpose Identity

To be clear about the meaning of some notation used here, it is necessary to digress into a review of some aspects of dual spaces.

If V is a real vector space, then $V^* = \mathrm{Hom}(V, \mathbb{R})$ is called the algebraic dual space of V. If V has a norm, $\|\cdot\|$, then the members $f \in V^*$ for which the following construction is finite,

$$\sup_{u \neq 0} \frac{f(u)}{\|u\|} < \infty\,,$$

are called bounded. These f form a space for which the construction serves as a norm. If V has finite dimension, then all members of V^* are bounded.

For $\mathbb{R}^n$ and other Hilbert spaces it is possible to define an isomorphism $V \to V^*$ by $v \mapsto \langle v, \cdot \rangle$. This is also an isometry between the inner product's norm on V and the norm constructed for V^*. Some parts of mathematics such as convex analysis find it convenient to "identify" $(\mathbb{R}^n)^*$ with $\mathbb{R}^n$ through this mapping. It seems unwise to make this identification when other norms are of interest, however, because the isomorphism is not an isometry for them.

For the present purposes it is better to recognize that each norm $\| \cdot \|$ on $\mathbb{R}^n$ has a dual norm $\| \cdot \|^*$, also on $\mathbb{R}^n$, that is defined by

$$\|v\|^* = \max_{u \neq 0} \frac{v^t u}{\|u\|} \, .$$

The value for this dual norm of $v \in \mathbb{R}^n$ equals the norm for the member of $(\mathbb{R}^n)^*$ that acts by $u \mapsto v^t u$. Similarly, if some column vector norms are used to define a matrix norm and a lower bound, $\| \cdot \|$ and $\| \cdot \|_\ell$, then the corresponding things defined by the dual column vector norms are indicated by $\| \cdot \|^*$ and $\| \cdot \|_\ell^*$. The norms for all other spaces and their duals continue to have the anonymous notation $\| \cdot \|$ whose identity is determined by the object to which it is applied.

In functional analysis it is well known that the norm of a bounded linear transformation equals the norm of its adjoint, $\|T\| = \|T^*\|$. In the notation just introduced for matrices this becomes $\|A\| = \|A^t\|^*$. Von Neumann and Goldstine noted a similar identity for matrix lower bounds.

Lemma 4.1 (Transpose Equality) *(Compare [16, p. 1043, equation (3.11.b)].) If A is not zero, then*

$$\|A\|_\ell = \|A^t\|_\ell^* \, .$$

Proof. (Part 1.) Figure 1's commutative diagram appears at the right side of Figure 2 where it has been rotated and restated in terms of the image and kernel spaces of T. Recall from Lemma 2.6 that

$$\|A\|_\ell = \|\psi_T^{-1}\|^{-1} \, . \tag{7}$$

(Part 2.) As mentioned in the text preceding the Lemma, operator norms are invariant with respect to adjoint, thus

$$\|\psi_T^{-1}\| = \|(\psi_T^{-1})^*\| \, . \tag{8}$$

(Part 3.) It is easy to see that adjoint commutes with inverse, since for $y \in \mathrm{im}\,(T)$ and $f \in \mathrm{im}\,(T)^*$,

$$f(y) = f(\psi_T(\psi_T^{-1}(y))) = (\psi_T^*(f))(\psi_T^{-1}(y)) = ((\psi_T^{-1})^*(\psi_T^*(f)))(y)$$

$$\begin{tikzcd}
& \mathrm{im}\,(T^*) & (\mathbb{R}^n/\ker(T))^* \arrow[l, "\phi_1"'] \\
(\mathbb{R}^m)^* \arrow[ur, dashed, "T^*"] \arrow[dr, "\phi_{T^*}"'] & & \\
& (\mathbb{R}^m)^*/\ker(T^*) \arrow[uu, dashed, "\psi_{T^*}"] \arrow[r, "\phi_2"] & \mathrm{im}\,(T)^* \arrow[uu, dashed, "\psi_T^*"]
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbb{R}^n/\ker(T) \arrow[dd, dashed, "\psi_T"] & \\
& \mathbb{R}^n \arrow[ul, "\phi_T"'] \arrow[dl, dashed, "T"] \\
\mathrm{im}\,(T) &
\end{tikzcd}$$

$$\begin{array}{lll}
T^* : f \mapsto f \circ T & \phi_1 : f \mapsto f \circ \phi_T & T : x \mapsto Ax \\
\phi_{T^*} : f \mapsto [f] = \ker(T^*) + f & \phi_2 : [f] \mapsto f|_{\mathrm{im}\,(T)} & \phi_T : x \mapsto [x] = \ker(T) + x \\
\psi_{T^*} : [f] \mapsto f \circ T & & \psi_T : [x] \mapsto T(x)
\end{array}$$

Figure 2: *The commutative diagram illustrating Lemma 4.1. The Lemma's proof associates $\|A\|_\ell$ with each of the dashed arrows, working from right to left, in five steps.*

so $(\psi_T^{-1})^* \circ \psi_T^* = 1$; similarly $\psi_T^* \circ (\psi_T^{-1})^* = 1$. Hence ψ_T^* is an isomorphism whose inverse $(\psi_T^*)^{-1}$ is given by $(\psi_T^{-1})^*$, therefore

$$\|(\psi_T^{-1})^*\| = \|(\psi_T^*)^{-1}\| \,. \tag{9}$$

(Part 4.) The next step requires that the square diagram in the center of Figure 2 commute. Suppose $f \in (\mathbb{R}^m)^*$ so that $[f]$ is an arbitrary member of $(\mathbb{R}^m)^*/\ker(T^*)$. With this choice, on the one hand $\psi_{T^*}([f]) = f \circ T : \mathbb{R}^n \to \mathbb{R}$; and on the other hand the image of $[f]$ under the result of

$$[f] \xmapsto{\phi_2} f|_{\mathrm{im}\,(T)} \xmapsto{\psi_T^*} f|_{\mathrm{im}\,(T)} \circ \psi_T \xmapsto{\phi_1} f|_{\mathrm{im}\,(T)} \circ \psi_T \circ \phi_T$$

acts on $x \in \mathbb{R}^n$ by

$$x \xmapsto{\phi_T} [x] \xmapsto{\psi_T} Ax \xmapsto{f} f(Ax) = (f \circ T)(x) \,,$$

so $\psi_{T^*} = \phi_1 \circ \psi_T^* \circ \phi_2$ as claimed. It is well known that the following are isometric isomorphisms,

$$\begin{array}{cc}
(\mathbb{R}^n/\ker(T))^* \to \ker(T)^\perp & (\mathbb{R}^m)^*/\mathrm{im}\,(T)^\perp \to \mathrm{im}\,(T)^* \\
\text{acting by } f \mapsto f \circ \phi_T & \text{acting by } [f] \mapsto f|_{\mathrm{im}\,(T)}
\end{array} \tag{10}$$

and since it is further well known that

$$\begin{aligned}
\ker(T)^\perp &= \mathrm{im}\,(T^*) \,, \\
\mathrm{im}\,(T)^\perp &= \ker(T^*) \,,
\end{aligned}$$

so the isometric isomorphisms in equation (10) are just the ϕ_1 and ϕ_2 in Figure 2. Altogether, from $\psi_{T^*} = \phi_1 \circ \psi_T^* \circ \phi_2$ follows $\psi_{T^*}^{-1} = \phi_2^{-1} \circ (\psi_T^*)^{-1} \circ \phi_1^{-1}$; so that

$$\|\psi_{T^*}^{-1}\| = \|(\psi_T^*)^{-1}\| \tag{11}$$

because ϕ_1^{-1} and ϕ_2^{-1} are isometric.

(Part 5.) The triangle at the left side of Figure 2 is identical to Figure 1's except that it has been drawn with respect to the adjoint transformation T^*. It is well known that if T represents the action of matrix-vector multiplication by A, then T^* does the same for A^t. Hence by Lemma 2.6,

$$\|A^t\|_\ell^* = \|\psi_{T^*}^{-1}\|^{-1} \tag{12}$$

(Step 6.) Equations (7 – 9, 11, 12) combine to give the desired result,

$$\frac{1}{\|A\|_\ell} = \|\psi_T^{-1}\| = \|(\psi_T^{-1})^*\| = \|(\psi_T^*)^{-1}\| = \|\psi_{T^*}^{-1}\| = \frac{1}{\|A^t\|_\ell^*}\,.$$

∎

4.2 Triangle Inequalities

The triangle inequalities for the matrix norm,

$$\Big|\|A\| - \|B\|\Big| \le \|A+B\| \le \|A\| + \|B\|\,,$$

find parallels (pun) in von Neumann and Goldstine's properties for the lower bound.

$$\|A\|_\ell - \|B\| \le \|A+B\|_\ell \le \|A\|_\ell + \|B\|\,.$$

The main implication of these inequalities is that the matrix lower bound is continuous in some circumstances.

These inequalities are not universal. For example, if

$$A = \begin{bmatrix} 1 & \\ & \epsilon \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & \\ & -\epsilon \end{bmatrix}\,,$$

then by Lemma 3.3 for spectral norms, $\|A+B\|_\ell = 1$ but $\|A\|_\ell = \|B\| = \epsilon$, so $\|A+B\|_\ell \not\le \|A\|_\ell + \|B\|$. The simplest case in which both triangle inequalities are true appears to be that A and $A+B$ have the same column or row spaces; each inequality then follows from containment in a different direction.

Lemma 4.2 (Triangle Inequalities) *(Compare [16, p. 1043, equation (3.8.b)].) If A and $A+B$ are not zero, then*

$$\|A\|_\ell - \|B\| \underset{1}{\le} \|A+B\|_\ell \underset{2}{\le} \|A\|_\ell + \|B\|\,,$$

provided the following conditions are satisfied, respectively.

1. $\mathrm{col}(A) \subseteq \mathrm{col}(A+B)$ *or* $\mathrm{row}(A) \subseteq \mathrm{row}(A+B)$
2. $\mathrm{col}(A+B) \subseteq \mathrm{col}(A)$ *or* $\mathrm{row}(A+B) \subseteq \mathrm{row}(A)$

Proof. (Part 1.) Reduction to column hypotheses: suppose A and B satisfy either of the row hypotheses, so A^t and B^t satisfy the corresponding column hypothesis. If the Lemma is valid for the column hypotheses, then it may be applied — to A^t, B^t and the dual vector norms — to give the respective inequality,

$$\|A^t\|_\ell^* - \|B^t\|^* \underset{1}{\leq} \|A^t + B^t\|_\ell^* \underset{2}{\leq} \|A^t\|_\ell^* + \|B^t\|^* \,.$$

The transposes can be removed to revert to the original matrix norm, and by Lemma 4.1, to the original matrix lower bound.

(Part 2.) Reduction to inequality (2): suppose A and B satisfy either of the hypotheses for inequality (1). Let $A' = A + B$ and $B' = -B$, so A and $A+B$ equal $A' + B'$ and A', respectively; therefore A' and B' satisfy the corresponding hypothesis for inequality (2). If the Lemma is valid for inequality (2), then it may be applied — to A' and B' — to obtain

$$\|A\|_\ell = \|A' + B'\|_\ell \leq \|A'\|_\ell + \|B'\| = \|A+B\|_\ell + \|B\| \,,$$

which is inequality (1).

(Part 3.) Thus it suffices to consider just the column hypothesis case of inequality (2). Suppose $(A+B)x = y \neq 0$. If (by 2) $\mathrm{col}(A+B) \subseteq \mathrm{col}(A)$, then by Definition 2.1 there is some δx with $A\,\delta x = Bx$ and $\|A\|_\ell \, \|\delta x\| \leq \|Bx\|$. Let $x' = x + \delta x$ so that both $Ax' = A(x + \delta x) = (A+B)x = y$ and

$$\|x'\| = \|x + \delta x\| \leq \|x\| + \|\delta x\| \leq \|x\| + \frac{\|Bx\|}{\|A\|_\ell} \leq \left(1 + \frac{\|B\|}{\|A\|_\ell}\right) \|x\| \,,$$

which rearranges to

$$\frac{\|y\|}{\|x\|} \leq \left(1 + \frac{\|B\|}{\|A\|_\ell}\right) \frac{\|y\|}{\|x'\|} \,.$$

This is for any x and y with $(A+B)x = y \neq 0$. Thus,

$$\max_{\{x \,:\, (A+B)x = y\}} \frac{\|y\|}{\|x\|} \;\leq\; \left(1 + \frac{\|B\|}{\|A\|_\ell}\right) \max_{\{x' \,:\, Ax' = y\}} \frac{\|y\|}{\|x'\|} \,,$$

so by Lemma 2.4,

$$\|A+B\|_\ell \leq \left(1 + \frac{\|B\|}{\|A\|_\ell}\right) \|A\|_\ell = \|A\|_\ell + \|B\| \,,$$

which is the desired inequality. ∎

Corollary 4.3 (Continuity) *The matrix lower bound is a continuous function on the open set of matrices with full rank.*

Proof. A has full row rank if and only if $\det(AA^t) \neq 0$, so the set of such matrices is open from the continuity of the determinant. Given such an A, for every sufficiently small $\epsilon > 0$, if $\|B\| \leq \epsilon$ then $A + B$ has full rank. Hence by Lemma 4.2,

$$\Big| \|A+B\|_\ell - \|A\|_\ell \Big| \leq \|B\| \leq \epsilon \,,$$

which is the condition that $\|A\|_\ell$ be a continuous function of A. The proof for matrices of full column rank is identical ∎

4.3 Multiplicative Inequalities

The multiplicative property of the matrix norm,

$$\|AB\| \leq \|A\| \, \|B\| \,,$$

is much expanded by von Neumann and Goldstine's properties for the lower bound. The first inequality is a direct analogue of the norm's inequality.

Lemma 4.4 (Product Rule) *(Compare [16, p. 1043, equation (3.9)].) If AB is not zero, then*

$$\|A\|_\ell \, \|B\|_\ell \leq \|AB\|_\ell \,.$$

Proof. Suppose that A and B are $m \times n$ and $n \times p$ matrices, respectively. Let $\mathcal{B}_k$ be the unit ball in $\mathbb{R}^k$, and let $r_X = \|X\|_\ell$. Since $\mathrm{col}(AB) \subseteq \mathrm{col}(A)$ so $\mathrm{col}(AB) = \mathrm{col}(AB) \cap \mathrm{col}(A)$, hence

$$\mathrm{col}(AB) \cap r_A r_B \mathcal{B}_m \;=\; \mathrm{col}(AB) \cap \mathrm{col}(A) \cap r_A r_B \mathcal{B}_m \,.$$

By Lemma 2.3, $\mathrm{col}(A) \cap r_A \mathcal{B}_m \subseteq A\mathcal{B}_n$ which after multiplication by r_B becomes $\mathrm{col}(A) \cap r_A r_B \mathcal{B}_m \subseteq r_B A \mathcal{B}_n$, hence

$$\mathrm{col}(AB) \cap \mathrm{col}(A) \cap r_A r_B \mathcal{B}_m \;\subseteq\; \mathrm{col}(AB) \cap r_B A \mathcal{B}_n \,.$$

Again by Lemma 2.3, $\mathrm{col}(B) \cap r_B \mathcal{B}_n \subseteq B\mathcal{B}_p$ so multiplying this by A gives

$$\mathrm{col}(AB) \cap r_B A \mathcal{B}_n \;\subseteq\; AB\mathcal{B}_p \,.$$

Altogether $\mathrm{col}(AB) \cap r_A r_B \mathcal{B}_m \subseteq AB\mathcal{B}_p$. Invoking Lemma 2.3 one last time,

$$r_{AB} \;=\; \max \{ \, r : \mathrm{col}(AB) \cap r\mathcal{B}_m \subseteq AB\mathcal{B}_p \} \,.$$

Since it has been shown that $r_A r_B$ lies in the set, so $r_A r_B \leq r_{AB}$. ∎

The other multiplicative inequalities mix the matrix norm and the matrix lower bound. They can be reduced to fewer cases, as for Lemma 4.2, but here the proof that treats each case separately is shorter.

Lemma 4.5 (Mixed Product Rules) *(Compare [16, p. 1043, equation (3.9)].) If AB is not zero, then*

$$\|AB\|_\ell \left\{ \begin{array}{c} \underset{1}{\leq} \|A\|\,\|B\|_\ell \underset{2'}{\leq} \\ \underset{2}{\leq} \|A\|_\ell\,\|B\| \underset{1'}{\leq} \end{array} \right\} \|AB\|$$

provided the following conditions are satisfied, respectively.

1. $\mathrm{row}(B) = \mathrm{row}(AB)$
   or $\mathrm{rank}(B) = \mathrm{rank}(AB)$
   or $\mathrm{null}(B) = \mathrm{null}(AB)$
   or $\mathrm{null}(A) \cap \mathrm{col}(B) = \{0\}$

$2'$. *B has full row rank*

2. $\mathrm{col}(A) = \mathrm{col}(AB)$
   or $\mathrm{rank}(A) = \mathrm{rank}(AB)$

$1'$. *A has full column rank*

Note that hypothesis $1' \Rightarrow 1$, $2' \Rightarrow 2$, and the multiple hypotheses for each of 1 and 2 are equivalent.

Proof. (Part 1.) If $\mathrm{row}(B) = \mathrm{row}(AB)$, then the row ranks are equal, so $\mathrm{rank}(B) = \mathrm{rank}(AB)$. This subtracted from the quantity of columns of B and AB gives the nullities, respectively, which therefore are equal. Since it is always true that $\mathrm{null}(B) \subseteq \mathrm{null}(AB)$, if these have the same dimension then they are equal. This means $Bw \neq 0$ implies $ABw \neq 0$ lest the null space grow, hence $\mathrm{null}(A) \cap \mathrm{col}(B) = \{0\}$. From this last hypothesis the others can be derived in reverse order using the much same reasoning, so all are equivalent.

Suppose $x \in \mathrm{col}(B)$. By Definition 2.1 there is a w with $Bw = x$ and $\|B\|_\ell\,\|w\| \leq \|x\|$, and there is a w' with $ABw' = ABw$ and $\|AB\|_\ell\,\|w'\| \leq \|ABw\|$. Altogether

$$\frac{\|AB\|_\ell}{\|A\|}\,\|w'\| \leq \frac{\|ABw\|}{\|A\|} = \frac{\|Ax\|}{\|A\|} \leq \frac{\|A\|\,\|x\|}{\|A\|} = \|x\|\,.$$

Since $ABw' = ABw$ and (by 1) $\mathrm{null}(A) \cap \mathrm{col}(B) = \{0\}$, therefore $Bw' = Bw = x$. This is for any $x \in \mathrm{col}(B)$, so $\|AB\|_\ell / \|A\| \leq \|B\|_\ell$.

(Part 2.) This inequality has the same number of equivalent hypotheses as part (1), but some of them are inconvenient to state because $\mathrm{null}(\,\cdot\,)$ is specific to right-side, matrix-vector multiplication.

If $y \in \mathrm{col}(A)$, then (by 2) $y \in \mathrm{col}(AB)$, so by Definition 2.1 there is a w with $ABw = y$ and $\|AB\|_\ell\,\|w\| \leq \|y\|$. Let $x = Bw$ so $\|x\| \leq \|B\|\,\|w\|$, which implies

$$\frac{\|AB\|_\ell}{\|B\|}\,\|x\| \leq \|AB\|_\ell\,\|w\| \leq \|y\|\,,$$

with $Ax = y$. This is for any $y \in \operatorname{col}(A)$, hence $\|AB\|_\ell / \|B\| \le \|A\|_\ell$.

(Part 2′.) Suppose $x \neq 0$. Since (by 2′) B has full row rank, by Definition 2.1 there is a w with $Bw = x$ and $\|B\|_\ell \, \|w\| \le \|x\|$. Therefore,

$$\frac{\|Ax\|}{\|x\|} \le \frac{\|ABw\|}{\|B\|_\ell \, \|w\|} \le \frac{\|AB\| \, \|w\|}{\|B\|_\ell \, \|w\|} = \frac{\|AB\|}{\|B\|_\ell} \,.$$

This is for any $x \neq 0$, so

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} \le \frac{\|AB\|}{\|B\|_\ell} \,.$$

(Part 1′.) Choose $w \neq 0$ that attains $\|B\|$, that is, $\|Bw\| = \|B\| \, \|w\|$. By Definition 2.1 there is an x with $Ax = ABw$ and $\|A\|_\ell \, \|x\| \le \|ABw\|$. Since (by 1′) A has full column rank, hence $x = Bw$. Altogether,

$$\|A\|_\ell \, \|B\| \, \|w\| = \|A\|_\ell \, \|Bw\| = \|A\|_\ell \, \|x\| \le \|ABw\| \le \|AB\| \, \|w\| \,.$$

∎

The following matrices provide counterexamples to Lemma 4.5's inequalities. Let

$$A = \begin{bmatrix} c & & \\ & 1 & \\ & & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & & \\ & 1 & \\ & & c \end{bmatrix},$$

where c is to be determined. One one hand, if $0 < c < 1$, then by Lemma 3.3 for spectral norms, $\|A\| = \|B\| = 1$, $\|A\|_\ell = \|B\|_\ell = c$, and $\|AB\|_\ell = 1$. Thus $\|AB\|_\ell \not\le \|A\| \, \|B\|_\ell$ and $\|AB\|_\ell \not\le \|A\|_\ell \, \|B\|$. On the other hand, if $1 < c$, then $\|A\| = \|B\| = c$, $\|A\|_\ell = \|B\|_\ell = 1$, and $\|AB\| = 1$. In this case $\|A\| \, \|B\|_\ell \not\le \|AB\|$ and $\|A\|_\ell \, \|B\| \not\le \|AB\|$.

4.4 Inner Product Identities

Identities of this kind,

$$\|A\| = \max_{\|x\| = 1} \|Ax\| = \max_{\|x\| = 1} \; \max_{\|y\|^* = 1} \; y^t A x \,,$$

are a natural starting point for investigations in matrix analysis. One of the corresponding minimizations,

$$\min_{\|x\| = 1} \|Ax\| \,,$$

serves as von Neumann and Goldstine's definition for their matrix lower bound [16, p. 1042, equation (3.1.b)]. Here it is found that such expressions give a nonzero value only for full rank matrices. The breakdown of these identities in the rank deficient case may be a reason why a comprehensive, nonzero matrix lower bound has not been defined previously.

Lemma 4.6 (Inner Product Formulas) *(Compare [16, p. 1042, equation (3.1.b)].) If A is not zero, then*

$$\left.\begin{array}{ll} 1. & \displaystyle\min_{\|x\|\,=\,1} \|Ax\| \\ 2. & \displaystyle\min_{\|x\|\,=\,1}\ \max_{\|y\|^*\,=\,1}\ y^t A x \\ 3. & \displaystyle -\max_{\|x\|\,=\,1}\ \min_{\|y\|^*\,=\,1}\ y^t A x \end{array}\right\} = \begin{cases} \|A\|_\ell & \textit{if } A \textit{ has full column rank,} \\ 0 & \textit{else,} \end{cases}$$

$$\left.\begin{array}{ll} 4. & \displaystyle\min_{\|y\|^*\,=\,1} \|A^t y\|^* \\ 5. & \displaystyle\min_{\|y\|^*\,=\,1}\ \max_{\|x\|\,=\,1}\ y^t A x \\ 6. & \displaystyle -\max_{\|y\|^*\,=\,1}\ \min_{\|x\|\,=\,1}\ y^t A x \end{array}\right\} = \begin{cases} \|A\|_\ell & \textit{if } A \textit{ has full row rank,} \\ 0 & \textit{else.} \end{cases}$$

Proof. Suppose that A is an $m \times n$ matrix. (Identity 1.) If A has full column rank, then $\mathrm{null}(A) = \{0\}$ so by Lemma 2.4,

$$\|A\|_\ell = \min_{\{x\,:\,Ax \neq 0\}}\ \max_{\{z\,:\,Az = 0\}} \frac{\|Ax\|}{\|x+z\|} = \min_{\{x\,:\,Ax\neq 0\}} \frac{\|Ax\|}{\|x\|} = \min_{\|x\|\,=\,1} \|Ax\| \,.$$

If A does not have full column rank, then $Ax = 0$ for some $x \neq 0$ so the minimum attains 0.

(Identity 2.) For any vector $Ax \in \mathbb{R}^m$,

$$\|Ax\| = \|(Ax)^{**}\| = \max_{\substack{f\,\in\,(\mathbb{R}^m)^* \\ \|f\|\,=\,1}} (Ax)^{**}(f) = \max_{\substack{f\,\in\,(\mathbb{R}^m)^* \\ \|f\|\,=\,1}} f(Ax) = \max_{\|y\|^*\,=\,1} y^t A x \,,$$

so identities (1) and (2) are equivalent.

(Identity 3.) This identity is the double negative of identity (2),

$$\begin{aligned} \min_{\|x\|\,=\,1}\ \max_{\|y\|^*\,=\,1}\ y^t A x \quad &= \quad -\,-\min_{\|x\|\,=\,1}\ \max_{\|y\|^*\,=\,1}\ y^t A x \\ y^t A x \qquad & \\ &= \quad -\max_{\|x\|\,=\,1}\ \min_{\|y\|^*\,=\,1}\ -y^t A x \\ &= \quad -\max_{\|x\|\,=\,1}\ \min_{\|y\|^*\,=\,1}\ y^t A x \,. \end{aligned}$$

(Identity 4.) If A has full row rank, then A^t has full column rank, so by Lemma 4.1 and identity (1),

$$\|A\|_\ell = \|A^t\|_\ell^* \ = \min_{\|y\|^*\,=\,1} \|A^t y\|^* \,.$$

If A^t does not have full column rank, then there is a nonzero $y \in \mathbb{R}^m$ with $A^t y = 0$, so minimization attains the value 0.

(Identities 5, 6.) These follow from identity (4) as (2, 3) follow from (1), or alternatively from (2, 3), respectively, as (4) is derived from (1). Either approach introduces members of $(\mathbb{R}^n)^{**}$ that must be removed by using the natural isometric isomorphism between $\mathbb{R}^n$ and $(\mathbb{R}^n)^{**}$. ∎

5 Applications

Like the matrix norm, the matrix lower bound supplies bounds and estimates for matrices and vectors. Each use perhaps could be handled by reasoning specific to the problem at hand, but a systematized collection of properties undeniably simplifies the task. The properties of the matrix norm are well known; the following examples suggest that some familiarity with the matrix lower bound might be useful as well.

Here is an overview of the examples. The first two deal with well-known formulas for distance-to-singularity, §5.1, and condition numbers, §5.2. It is found that the formulas can be meaningfully generalized simply by replacing the $\|A^{-1}\|^{-1}$ in them with $\|A\|_\ell$. Example three, §5.3, interprets Section 4.4's inner product identities in terms of convex analysis. Finally, both the first and fourth examples, §5.1 and §5.4, illustrate the use of matrix lower bounds to construct Cauchy sequences in existence proofs.

5.1 Linear Algebra: Rank Deficiency

C. Eckart and G. Young [5] apparently were the first to determine the distance from a given matrix to the set of matrices with lower rank. (Their priority is endorsed by Blum et al. [4].) Eckart and Young's interest in this problem was motivated by the question of how best to approximate one matrix by another. They considered the Frobenius norm and $m \times n$ matrices; the basis of their analysis was the singular value decomposition. W. Kahan [11, p. 775] treated essentially arbitrary operator norms and nonsingular matrices. The following Theorem is a generalization of Kahan's case to matrices of full rank.

Theorem 5.1 (Rank Deficiency) *If A has full rank, then $\|A\|_\ell$ is the distance from A to the set of a rank deficient matrices,*

$$\|A\|_\ell = \min \{\|E\| : \operatorname{rank}(A+E) < \operatorname{rank}(A)\} \ .$$

Proof. Since both $\| \cdot^t \| = \| \cdot \|^*$ and by Lemma 4.1 $\| \cdot^t \|_\ell = \| \cdot \|_\ell^*$, it suffices to consider the case that A has full row rank.

(Part 1.) Suppose $\|E\| < \|A\|_\ell$. Since A has full row rank, for any y there is some x_1 with $Ax_1 = y$. There also is some x_2 with $A(x_2 - x_1) = -Ex_1$.

Finally invoking Definition 2.1, for $n \geq 2$ there is an x_{n+1} with $A(x_{n+1} - x_n) = -E(x_n - x_{n-1})$ and $\|A\|_\ell \, \|x_{n+1} - x_n\| \leq \|E(x_n - x_{n-1})\| \leq \|E\| \, \|x_n - x_{n-1}\|$. Combining the inequalities shows, for $n \geq 2$,

$$\|x_{n+1} - x_n\| \leq \left(\frac{\|E\|}{\|A\|_\ell} \right)^{n-2} \|x_2 - x_1\| \, ,$$

so $\{x_n\}$ is a Cauchy sequence because $\|E\|/\|A\|_\ell < 1$. Let x_* be the limit of the sequence. Passing to the limit in

$$Ax_{n+1} = Ax_1 + \sum_{i=1}^{n} A(x_{i+1} - x_i) = y - Ex_1 - \sum_{j=1}^{n-1} E(x_{j+1} - x_j) = y - Ex_n$$

proves $Ax_* = y - Ex_*$, hence $y \in \mathrm{col}(A + E)$. This is for any y, therefore $A + E$ has full row rank.

(Part 2.) Let ϕ and ψ be the functions shown in Figure 1, and let ψ^{-1} attain its norm at $y_0 \in \mathrm{col}(A)$. Therefore from Lemma 2.6,

$$\|A\|_\ell^{-1} = \|\psi^{-1}\| = \frac{\|[x_0]\|}{\|y_0\|} \, ,$$

where $\psi^{-1}(y_0) = [x_0] \in \mathbb{R}^n/\mathrm{null}(A)$. The Hahn-Banach theorem says there is a functional $g : \mathbb{R}^n/\mathrm{null}(A) \to \mathbb{R}$ with $g([x_0]) = 1$ and $\|g\| = 1/\|[x_0]\|$. Define $e : \mathbb{R}^n/\mathrm{null}(A) \to \mathrm{col}(A)$ by $e([x]) = g([x])y_0$. Thus $e([x_0]) = y_0$ and $\|e\| = \|y_0\|/\|[x_0]\| = \|A\|_\ell$. The map $\psi - e$ has a nontrivial kernel because both ψ and e transform $[x_0]$ to y_0. Since

$$\dim(\ker(\psi - e)) + \dim(\mathrm{im}\,(\psi - e)) = \dim(\mathbb{R}^n/\mathrm{null}(A)) = \dim(\mathrm{col}(A)) \, ,$$

therefore $\dim(\mathrm{im}\,(\psi - e)) < \dim(\mathrm{col}(A))$. The composite map $e \circ \phi : \mathbb{R}^n \to \mathbb{R}^m$ is represented by an $m \times n$ matrix, E. Now $A - E$ represents $\psi \circ \phi - e \circ \phi = (\psi - e) \circ \phi$, so $\dim(\mathrm{col}(A - E)) \leq \dim(\mathrm{im}\,(\psi - e)) < \dim(\mathrm{col}(A))$, hence $A - E$ is rank deficient. Finally, $\|E\| = \|e \circ \phi\| \leq \|e\| \, \|\phi\| = \|e\| = \|A\|_\ell$ because $\|\phi\| = 1$, but $\|E\| \geq \|A\|_\ell$ by Part (1), so $\|E\| = \|A\|_\ell$.

(Alternate Part 2.) It is easier to find a perturbation for A^t. By Corollary 2.5 there are an x and y so $A^t y = x$ and $\|A^t\|_\ell^* \, \|y\|^* = \|x\|^*$. Let u be a unit vector at which $\|y\|^*$ is attained, that is, $\|u\| = 1$ and $y^t u = \|y\|^*$. Let $E^t = xu^t/\|y\|^*$ so by Lemma 3.2 $\|E^t\|^* = \|x\|^* \, \|u\|^{**}/\|y\|^* = \|x\|^*/\|y\|^* = \|A^t\|_\ell^*$. Since A^t has full column rank, but $A^t - E^t$ has a null space because $(A^t - E^t)y = 0$, thus $A^t - E^t$ and $A - E$ are rank deficient. Finally, the perturbation E has the desired size because $\|E\| = \|E^t\|^* = \|A^t\|_\ell^* = \|A\|_\ell$ by Lemma 4.1. This approach is shorter than the first Part (2) but it is less direct if the need to prove Lemma 4.1 is taken into account. ∎

5.2 Numerical Analysis: Condition Numbers

Von Neumann and Goldstine's use [16, pp. 1073–5, section 6.6] of their matrix lower bound occurs in a long proof that is beyond the scope of this paper. Their subject was the effect of rounding errors on numerical calculations, and the following example is in the spirit of their discussion.

Von Neumann and Goldstine suggested [16, p. 1092] that it would be useful to interpret numerical algorithms as though they proceed without rounding error, provided the calculations begin at some perturbed form of the initial data. The idea is that the algorithm exactly applied to the perturbed data should give the same result as the algorithm applied with rounded arithmetic to the unperturbed data. This concept became the central tenet of backward rounding error analysis. Von Neumann and Goldstine left their readers to work out, "in several different ways" [16, p. 1093], what effect the perturbed data would have on the solution of a problem.

To that end, suppose that the equations $Ax = b$ are consistent, and suppose an $\bar{x}$ has been given that solves a related problem $(A + E)\bar{x} = b$ where E is von Neumann and Goldstine's data perturbation. Since $E\bar{x} = b - A\bar{x}$ lies in the column space of A, Definition 2.1 says there is a δx with $A\,\delta x = E\bar{x}$ and $\|A\|_\ell\,\|\delta x\| \le \|E\bar{x}\|$. Let $x' = \bar{x} + \delta x$. This x' solves the original equations,

$$Ax' = A(\bar{x} + \delta x) = b - E\bar{x} + E\bar{x} = b\,.$$

It is also close to the erroneous solution,

$$\frac{\|x' - \bar{x}\|}{\|\bar{x}\|} = \frac{\|\delta x\|}{\|\bar{x}\|} \le \frac{\|E\bar{x}\|}{\|\bar{x}\|\,\|A\|_\ell} \le \frac{\|E\|}{\|A\|_\ell}\,,$$

provided only that the data perturbation, E, is small compared to $\|A\|_\ell$.

A calculation that produces an $\bar{x}$ that can be accounted for by an E that is small relative to the data, A, is called numerically stable. In this case, $\bar{x}$ is relatively close to an exact solution,

$$\frac{\|x' - \bar{x}\|}{\|\bar{x}\|} \le \frac{\|E\|}{\|A\|_\ell} = \frac{\|A\|}{\|A\|_\ell}\,\frac{\|E\|}{\|A\|}\,,$$

provided only that the ratio $\|A\|/\|A\|_\ell$ is small. When A is square and invertible, then this quantity specializes to the well-known value $\kappa(A) = \|A\|\,\|A^{-1}\|$ by Lemma 3.1.

In general, a condition number for a problem is a bound for the ratio of some measure of the solution error to some measure of the data perturbations [21, p. 29]. The matrix lower bound makes it possible to say, in von Neumann and Goldstine's terminology, that the condition number of any consistent system of linear equations is the ratio of the matrix upper and lower bounds.

Theorem 5.2 (Condition Number) *If $Ax = b \neq 0$ is a consistent system of linear equations, then a condition number for these equations is*

$$\kappa(A) = \frac{\|A\|}{\|A\|_\ell}\,.$$

This is a condition number in the sense that for any proposed solution $\bar{x}$ and any von Neumann perturbation E, with $(A+E)\bar{x} = b$, there is some x' with $Ax' = b$ and

$$\frac{\|x' - \bar{x}\|}{\|\bar{x}\|} \leq \kappa(A)\,\frac{\|E\|}{\|A\|}\,.$$

In formulas of this kind it is customary to replace the denominator's $\bar{x}$ by the exact solution, here x', at the cost of a marginally weaker bound.

Corollary 5.3 *Continuing Theorem 5.2, if*

$$\kappa(A)\,\frac{\|E\|}{\|A\|} \;<\; 1\,,$$

then

$$\frac{\|x' - \bar{x}\|}{\|x'\|} \leq \kappa(A)\,\frac{\|E\|}{\|A\|}\left(1 - \kappa(A)\,\frac{\|E\|}{\|A\|}\right)^{-1}. \tag{13}$$

Proof. Let $\delta = \kappa(A)\,\|E\|/\|A\|$, so from the triangle inequalities and by Theorem 5.2,

$$\frac{\|\bar{x}\|}{\|x'\|} - 1 = \frac{\|\bar{x}\| - \|x'\|}{\|x'\|} \leq \frac{|\,\|\bar{x}\| - \|x'\|\,|}{\|x'\|} \leq \frac{\|x' - \bar{x}\|}{\|x'\|} \leq \delta\,\frac{\|\bar{x}\|}{\|x'\|}\,.$$

This rearranges to

$$\frac{\|\bar{x}\|}{\|x'\|} \leq \frac{1}{1-\delta}$$

which, when multiplied by the Theorem's inequality, gives the Corollary's result. ∎

Equation (13) is new in that it has been derived previously only when A is square and invertible. H. Wittmeyer [22] apparently was the first to find a bound of this form. (His priority is endorsed by Stewart and Sun [19].) Wittmeyer dealt with nonsingular matrices and the 2-norm. The bound in exactly the form of equation (13) seems to be due to J. Wilkinson [21, p. 93, equation (12.15)] for the condition number $\kappa(A) = \|A\|\,\|A^{-1}\|$. Wilkinson treated nonsingular matrices and essentially arbitrary operator norms. Here, Wittmeyer and Wilkinson's results have been generalized to any consistent system of equations for any $m \times n$ matrix.

5.3 Optimization Theory: Min-max Identity

John von Neumann [14] proved the original minimax theorem for a problem whose optimal value is the expected outcome of certain parlor games, and thereby initiated game theory (at age twenty-two). Many people including von Neumann himself have since generalized the minimax theorem [18]. A familiar result similar to one that von Neumann actually proved is,

$$\min_{x \,\in\, K} \; \max_{y \,\in\, L} \; y^t A x \;=\; \max_{y \,\in\, L} \; \min_{x \,\in\, K} \; y^t A x \,,$$

where A is any matrix, and K and $L \subseteq \mathbb{R}^n$ are compact and convex.

Von Neumann's example of the minimax theorem suggests that it may be fruitful to look for identities among optimization problems that involve two convex sets. A likely candidate for abstraction is Lemma 4.6's inner product identity,

$$\min_{\|x\| \,=\, 1} \; \max_{\|y\|^* \,=\, 1} \; y^t A x \;=\; \|A\|_\ell \;=\; \min_{\|y\|^* \,=\, 1} \; \max_{\|x\| \,=\, 1} \; y^t A x \,,$$

where A is nonsingular. The two convex sets in this formula are the unit balls around the origin for the norms $\| \cdot \|$ and $\| \cdot \|^*$. It is interesting that this apparently algebraic identity involving the matrix lower bound can be derived from the following relationship that depends only on geometric properties of convex sets.

Theorem 5.4 (Min-max Identity [9]) *If $K \subseteq \mathbb{R}^n$ and $L^* \subseteq (\mathbb{R}^n)^*$ are compact, convex, and contain the origin in their interiors, then*

$$\min_{x \,\in\, \mathrm{bd}\,(K)} \; \max_{f \,\in\, \mathrm{bd}\,(L^*)} \; f(x) \;=\; \min_{f \,\in\, \mathrm{bd}\,(L^*)} \; \max_{x \,\in\, \mathrm{bd}\,(K)} \; f(x) \,. \tag{14}$$

Note that equation (14) resembles the minimax theorem in that the sets are exchanged, but unlike von Neumann's case the min and the max are not reversed.

The proof of Theorem 5.4 involves ideas from convex analysis that while elementary are beyond the scope of this paper. The theorem is a true duality result in that its proof relies on the polar duality relationship among convex sets that contain the origin. Nevertheless equation (14) cannot be characterized as a problem in convex optimization because the feasible sets are not convex. Moreover, the set of optimal pairs (x, f) that jointly attain the min-max of both sides in the equation needn't be convex although it has in interesting geometric description. For further discussion see [9].

5.4 Real Analysis: Collocated Level Sets

Suppose $\mathcal{D}$ is an open set in $\mathbb{R}^m$, and suppose $f : \mathcal{D} \to \mathbb{R}^n$ is continuously differentiable. By analogy with real-valued functions, the set

$$f^{-1}(y) = \{x : f(x) = y\}$$

may be called a level set of f.

At every $x_0 \in \mathcal{D}$ where the Jacobian matrix $Df(x_0)$ has full row rank, the implicit function theorem says that $f^{-1}(f(x_0))$ contains a smooth curve that is parameterized by an implicitly defined function. Usually the implicit function is emphasized, but the theorem also can be interpreted as describing the level set [3, p. 384, theorem 41.9 part (b)]: there is neighborhood N_{x_0} of x_0 where the implicit function's graph is all of $N_{x_0} \cap f^{-1}(f(x_0))$.

Here, a geometric comparison is made between all the level sets of f and those of its tangent function at x_0. Near x_0, the corresponding level sets are always present and they are asymptotically identical. The proof of this is a modification of a construction apparently due to L. M. Graves [8], see also [3, p. 378, theorem 41.6]. The matrix lower bound supplies a critical estimate in this construction.

The proof also depends on an inequality form of the mean value theorem that is valid for higher dimensions. If f is continuously differentiable, then for every x_0 and every $\rho > 0$ there is a neighborhood $N_{x_0}^{(15)}(\rho)$ of x_0 where

$$x, x' \in N_{x_0}^{(15)}(\rho) \quad \Rightarrow \quad \|f(x) - f(x') - Df(x_0)(x - x')\| \leq \epsilon \, \|x - x'\| \, . \qquad (15)$$

This inequality actually is equivalent to its hypothesis that the derivative be continuous [17, p. 72, lemma 3.2.10]. It has been discussed many times, see [3, p. 377, lemma 41.4] and [12, p. 212, notes for §7.1–4].

Theorem 5.5 (Collocated Level Sets) *Suppose $\mathcal{D} \subseteq \mathbb{R}^m$ is a neighborhood of x_0, and suppose $f : \mathcal{D} \to \mathbb{R}^n$ is continuously differentiable. Let $T(x)$ be the linear function that is tangent to $f(x)$ at x_0,*

$$T(x) = f(x_0) + Df(x_0)(x - x_0) \, .$$

Suppose the matrix $Df(x_0)$ has full row rank. For every $\epsilon > 0$ there is a neighborhood $N_{x_0}(\epsilon)$ of x_0 where $x \in N_{x_0}(\epsilon)$ implies

1. *$\exists \; x_T$ with $\|x_T - x\| \leq \epsilon \, \|x - x_0\|$ and $T(x_T) = f(x)$, and*
2. *$\exists \; x_f \in \mathcal{D}$ with $\|x_f - x\| \leq \epsilon \, \|x - x_0\|$ and $T(x) = f(x_f)$.*

Proof. The proof is based on the neighborhood $N_{x_0}^{(15)}(\rho)$ in equation (15) for a ρ determined from ϵ as follows. Let $\delta = \epsilon/(1+\epsilon) < 1$. Let $\ell = \|Df(x_0)\|_\ell$. Use the notation $B_c(r)$ for the open ball of center c and radius r. It is always possible to find an $r > 0$ so that $\mathrm{cl}\,(B_{x_0}(r)) \subseteq N_{x_0}^{(15)}(\delta\ell)$. The neighborhood in the statement of the theorem is then

$$N_{x_0}(\epsilon) = B_{x_0}(r(1-\delta)) \subseteq \mathrm{cl}\,(B_{x_0}(r)) \subseteq N_{x_0}^{(15)}(\delta\ell) \subseteq \mathcal{D} \, .$$

In both parts (1) and (2), suppose $x \in N_{x_0}(\epsilon)$.

(Part 1.) Since $Df(x_0)$ has full row rank, Definition 2.1 says there is an x_T with

$$Df(x_0)(x_T - x) = f(x) - T(x)\,,$$
$$\|Df(x_0)\|_\ell\,\|x_T - x\| \le \|f(x) - T(x)\|\,.$$

The equality and some algebra imply $T(x_T) = f(x)$, while the inequality and equation (15) for $\rho = \delta\ell$ and $x' = x_0$ imply

$$\|x_T - x\| \le \frac{\|f(x) - T(x)\|}{\|Df(x_0)\|_\ell} \le \frac{\delta\ell\|x - x_0\|}{\ell} = \delta\|x - x_0\| < \epsilon\|x - x_0\|\,. \quad (16)$$

(Part 2.) Let $x_1 = x$. With this notation equation (16) shows that the following conditions are satisfied for $j = 0$.

$$(1_j) \quad \|x_{j+1} - x_j\| \le \delta^j\,\|x - x_0\|$$
$$(2_j) \quad \|f(x_{j+1}) - T(x)\| \le \delta\ell\,\|x_{j+1} - x_j\|$$

Notice that summing (1_j) for $0 \le j \le k$ gives

$$\|x_{k+1} - x_0\| \le \sum_{j=0}^{k} \|x_{j+1} - x_j\| \le \frac{1 - \delta^{k+1}}{1 - \delta}\,\|x - x_0\|\,.$$

This combines with the choice $x \in B_{x_0}(r(1-\delta))$ to place $x_{k+1} \in \mathrm{cl}(B_{x_0}(r)) \subseteq \mathcal{D}$. Therefore the evaluation of $f(x_{k+1})$ in condition (2_k) is always well-defined provided that (1_j) holds for $0 \le j \le k$.

Suppose x_0, x_1, …, x_n have been constructed to satisfy (1_j) and (2_j) for $0 \le j \le n-1$. As in Part (1) — but note the change of sign — Definition 2.1 says there is an x_{n+1} with

$$Df(x_0)(x_{n+1} - x_n) = -\,[f(x_n) - T(x)]\,,$$
$$\|Df(x_0)\|_\ell\,\|x_{n+1} - x_n\| \le \|f(x_n) - T(x)\|\,.$$

The inequality and conditions (2_{n-1}) and (1_{n-1}) imply condition (1_n),

$$\|x_{n+1} - x_n\| \le \frac{\|f(x_n) - T(x)\|}{\|Df(x_0)\|_\ell} \le \frac{\delta\ell\|x_n - x_{n-1}\|}{\ell} \le \frac{\delta\ell\delta^{n-1}\|x - x_0\|}{\ell}\,.$$

It is therefore possible to evaluate $f(x_{n+1})$. Condition (2_n) now holds since

$$\begin{aligned}\|f(x_{n+1}) - T(x)\| &= \|f(x_{n+1}) - f(x_n) - Df(x_0)(x_{n+1} - x_n)\| \\ &\le \delta\ell\|x_{n+1} - x_n\|\,.\end{aligned}$$

The equality is from the choice of x_{n+1}, while the inequality is from equation

(15), which is applicable because x_n, $x_{n+1} \in \mathrm{cl}\,(B_{x_0}(r)) \subseteq N^{(15)}_{x_0}(\delta\ell)$.

In this way a sequence $\{x_n\} \subseteq \mathrm{cl}\,(B_{x_0}(r))$ is constructed that satisfies conditions (1_n) and (2_n) for all n. This is a Cauchy sequence by (1_n), so it has a limit $x_f \in \mathrm{cl}\,(B_{x_0}(r)) \subseteq \mathcal{D}$. Passing to the limit in (2_n) shows $f(x_f) = T(x)$. Summing (1_j), now for $1 \le j \le n$, gives

$$\|x_{n+1} - x\| \le \sum_{j=1}^{n} \|x_{j+1} - x_j\| \le \delta\, \frac{1-\delta^n}{1-\delta}\, \|x - x_0\|\,,$$

which in the limit becomes $\|x_f - x\| \le \delta(1-\delta)^{-1}\|x - x_0\| = \epsilon\|x - x_0\|$. ∎

6 Summary

6.1 Definitions

Definition 2.1 (Matrix Lower Bound). Let A be a nonzero matrix. The matrix lower bound, $\|A\|_\ell$, is the largest of the numbers, m, such that for every y in the column space of A, there is some x with $Ax = y$ and $m\,\|x\| \le \|y\|$.

Lemma 2.2 (Existence and Bounds). The matrix lower bound exists and is positive. In particular,

$$\|B\|^{-1} \le \|A\|_\ell \le \|A\|$$

where B is any matrix that satisfies the Penrose condition $ABA = A$.

Lemma 2.3 (Geometric Characterization). Let A be a nonzero $m \times n$ matrix, and let $\mathcal{B}_m$ and $\mathcal{B}_n$ be the unit balls in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. The matrix lower bound $\|A\|_\ell$ is the radius of the largest ball — with respect to the subspace $\mathrm{col}(A)$ — that is centered at the origin and contained in $A\mathcal{B}_n$,

$$\|A\|_\ell = \max\,\{\, r : \mathrm{col}(A) \cap r\mathcal{B}_m \subseteq A\mathcal{B}_n \}\ .$$

Lemma 2.4 (Min-max Characterization). If A is not zero, then

$$\begin{aligned} \|A\|_\ell &= \min_{y \,\in\, \mathrm{col}(A)\backslash\{0\}} \ \max_{\{x\,:\,Ax = y\}} \frac{\|y\|}{\|x\|} \\ &= \min_{\{x\,:\,Ax \ne 0\}} \ \max_{\{z\,:\,Az = 0\}} \frac{\|Ax\|}{\|x+z\|}\,. \end{aligned}$$

Corollary 2.5 (Attainment). There is a nonzero x with $\|A\|_\ell\,\|x\| = \|Ax\|$.

Lemma 2.6 (Functional Analysis Characterization). Let A be a nonzero $m \times n$ matrix, and let $\psi : \mathbb{R}^n/\text{null}(A) \to \text{col}(A)$ be the canonical map defined by $\psi([x]) = Ax$. Then

$$\|\psi^{-1}\|^{-1} = \|A\|_\ell$$

where the norm of ψ^{-1} is the operator norm induced from the norms on $\text{col}(A)$ and $\mathbb{R}^n/\text{null}(A)$.

6.2 Examples

Lemma 3.1 (Invertible Matrices). If A is invertible, then $\|A\|_\ell = \|A^{-1}\|^{-1}$.

Lemma 3.2 (Rank 1 Matrices) If A is a rank 1 matrix, where $A = uv^t$ for some $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, then $\|A\|_\ell = \|u\| \, \|v\|^*$, where $\| \cdot \|^*$ is the dual of the vector norm for $\mathbb{R}^n$.

Lemma 3.3 (Spectral Norms). If A is not zero, and if the vector norms are 2-norms, then $\|A\|_\ell = \sigma_{\text{min}}$, where σ_{min} is the smallest nonzero singular value of A.

Corollary 3.4 (Separation) If A is symmetric and positive definite, and if the vector norms are 2-norms, then

$$\|A\|_\ell \le \max_i A_{i,i} \le \|A\| \, .$$

6.3 Properties

Lemma 4.1 (Transpose Equality). If A is not zero, then

$$\|A\|_\ell = \|A^t\|_\ell^* \, .$$

Lemma 4.2 (Triangle Inequalities). If A and $A + B$ are not zero, then

$$\|A\|_\ell - \|B\| \underset{1}{\le} \|A + B\|_\ell \underset{2}{\le} \|A\|_\ell + \|B\| \, ,$$

provided the following conditions are satisfied, respectively.

1. $\text{col}(A) \subseteq \text{col}(A + B)$
   or $\text{row}(A) \subseteq \text{row}(A + B)$
2. $\text{col}(A + B) \subseteq \text{col}(A)$
   or $\text{row}(A + B) \subseteq \text{row}(A)$

Corollary 4.3 (Continuity). The matrix lower bound is a continuous function on the open set of matrices with full rank.

Lemma 4.4 (Product Rule). If AB is not zero, then

$$\|A\|_\ell \, \|B\|_\ell \le \|AB\|_\ell \,.$$

Lemma 4.5 (Mixed Product Rules). If AB is not zero, then

$$\|AB\|_\ell \left\{ \begin{array}{c} \underset{1}{\le} \|A\| \, \|B\|_\ell \underset{2'}{\le} \\[2ex] \underset{2}{\le} \|A\|_\ell \, \|B\| \underset{1'}{\le} \end{array} \right\} \|AB\|$$

provided the following conditions are satisfied, respectively.

1. $\operatorname{row}(B) = \operatorname{row}(AB)$ or $\operatorname{rank}(B) = \operatorname{rank}(AB)$ or $\operatorname{null}(B) = \operatorname{null}(AB)$ or $\operatorname{null}(A) \cap \operatorname{col}(B) = \{0\}$

2′. B has full row rank

2. $\operatorname{col}(A) = \operatorname{col}(AB)$ or $\operatorname{rank}(A) = \operatorname{rank}(AB)$

1′. A has full column rank

Note that hypothesis $1' \Rightarrow 1$, $2' \Rightarrow 2$, and the multiple hypotheses for each of 1 and 2 are equivalent.

Lemma 4.6 (Inner Product Formulas). If A is not zero, then

$$\left. \begin{array}{ll} 1. & \displaystyle \min_{\|x\| = 1} \|Ax\| \\[2ex] 2. & \displaystyle \min_{\|x\| = 1} \max_{\|y\|^* = 1} y^t A x \\[2ex] 3. & \displaystyle -\max_{\|x\| = 1} \min_{\|y\|^* = 1} y^t A x \end{array} \right\} = \left\{ \begin{array}{ll} \|A\|_\ell & \text{if } A \text{ has full column rank,} \\ 0 & \text{else,} \end{array} \right.$$

$$\left. \begin{array}{ll} 4. & \displaystyle \min_{\|y\|^* = 1} \|A^t y\|^* \\[2ex] 5. & \displaystyle \min_{\|y\|^* = 1} \max_{\|x\| = 1} y^t A x \\[2ex] 6. & \displaystyle -\max_{\|y\|^* = 1} \min_{\|x\| = 1} y^t A x \end{array} \right\} = \left\{ \begin{array}{ll} \|A\|_\ell & \text{if } A \text{ has full row rank,} \\ 0 & \text{else.} \end{array} \right.$$

6.4 Applications

Theorem 5.1 (Rank Deficiency). If A has full rank, then $\|A\|_\ell$ is the distance from A to the set of a rank deficient matrices,

$$\|A\|_\ell = \min \{ \|E\| : \operatorname{rank}(A + E) < \operatorname{rank}(A) \} \,.$$

Theorem 5.2 (Condition Number). If $Ax = b \neq 0$ is a consistent system of linear equations, then a condition number for these equations is

$$\kappa(A) = \frac{\|A\|}{\|A\|_\ell}\,.$$

This is a condition number in the sense that for any proposed solution $\bar{x}$ and any von Neumann perturbation E, with $(A+E)\bar{x} = b$, there is some x' with $Ax' = b$ and

$$\frac{\|x' - \bar{x}\|}{\|\bar{x}\|} \leq \kappa(A)\,\frac{\|E\|}{\|A\|}\,.$$

Corollary 5.3 Continuing Theorem 5.2, if

$$\kappa(A)\,\frac{\|E\|}{\|A\|} \;<\; 1\,,$$

then

$$\frac{\|x' - \bar{x}\|}{\|x'\|} \leq \kappa(A)\,\frac{\|E\|}{\|A\|}\left(1 - \kappa(A)\,\frac{\|E\|}{\|A\|}\right)^{-1}\,.$$

Theorem 5.4 (Min-max Identity) If $K \subseteq \mathbb{R}^n$ and $L^* \subseteq (\mathbb{R}^n)^*$ are compact, convex, and contain the origin in their interiors, then

$$\min_{x \,\in\, \mathrm{bd}(K)}\ \max_{f \,\in\, \mathrm{bd}(L^*)}\ f(x) \;=\; \min_{f \,\in\, \mathrm{bd}(L^*)}\ \max_{x \,\in\, \mathrm{bd}(K)}\ f(x)\,.$$

Theorem 5.5 (Collocated Level Sets). Suppose $\mathcal{D} \subseteq \mathbb{R}^m$ is a neighborhood of x_0, and suppose $f : \mathcal{D} \to \mathbb{R}^n$ is continuously differentiable. Let $T(x)$ be the linear function that is tangent to $f(x)$ at x_0,

$$T(x) = f(x_0) + Df(x_0)(x - x_0)\,.$$

Suppose the matrix $Df(x_0)$ has full row rank. For every $\epsilon > 0$ there is a neighborhood $N_{x_0}(\epsilon)$ of x_0 where $x \in N_{x_0}(\epsilon)$ implies

1. $\exists\, x_T$ with $\|x_T - x\| \leq \epsilon\,\|x - x_0\|$ and $T(x_T) = f(x)$, and
2. $\exists\, x_f \in \mathcal{D}$ with $\|x_f - x\| \leq \epsilon\,\|x - x_0\|$ and $T(x) = f(x_f)$.

7 Problems

In large part this paper was inspired by John von Neumann. He is often described, in lay terms, as someone to whom we are indebted for inventing many useful branches of science and mathematics [13]. A modest conclusion to draw from this is that von Neumann had a special talent to inspire further research.

Thus the best ideas for matrix lower bounds probably are to be found in von Neumann and Goldstine's paper. Although they do not consider matrix lower bounds in the generality discussed here, the associations made in their paper may indicate additional topics for study. Since their discussion of lower bounds is juxtaposed with a treatment of matrix 2-norms, von Neumann and Goldstine appear to suggest that best approximation problems may be solved in other norms just as pseudoinverses solve them for the 2-norm. This is the probably too-hopeful intent of the first two questions below.

1. What is the maximum of the $\|B\|^{-1}$ with $ABA = A$ in Lemma 2.2? Note that some B has a smallest norm so the maximum is attained.
2. When is there a B with $ABA = A$ and $\|B\|^{-1} = \|A\|_\ell$ in Lemma 2.2?
3. What can be said about the function $x = f(y)$ that attains the maximum in Lemma 2.4's equation (2)?

For the next three questions, let $\|A\|_{\ell,p}$ be the matrix lower bound defined with respect to vector spaces equipped with Hölder p-norms.

4. Does $\|A\|_{\ell,p}$ have bounds analogous to Corollary 3.4's bound for $\|A\|_{\ell,2}$?
5. Are there formulas for $\|A\|_{\ell,1}$ and $\|A\|_{\ell,\infty}$ analogous to the well-known formulas for $\|A\|_1$ and $\|A\|_\infty$?
6. Using the well-known inequalities

$$1 \le \frac{\|v\|_p}{\|v\|_{p'}} \le \frac{n^{1/p}}{n^{1/p'}}$$

   where $p \le p'$ and n is the dimension of the vector v, what inequalities can be established between $\|A\|_{\ell,p}$ and $\|A\|_{\ell,p'}$?
7. Can the lower bound in Lemma 4.2 be replaced by its absolute value? Note that von Neumann and Goldstine did not say it could.
8. Is there a commutative diagram that proves Lemma 4.4's product rule?
9. Can any of the hypotheses for Lemma 4.5's inequalities be weakened?
10. Can Theorem 5.1 be generalized to say that $\|A\|_\ell$ is the distance to the nearest matrix of lower rank even when A is already rank deficient?
11. As in Theorem 5.1, what is the distance to the nearest matrix whose rank differs by a given amount?

12. Let $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots$ be the singular values of a matrix. It is well known that [7, p. 428, cor. 8.3.2]

$$\sigma_k(A) - \sigma_1(E) \leq \sigma_k(A+E) \leq \sigma_k(A) + \sigma_1(E)\,. \tag{17}$$

If A and $A+E$ have full rank, then the choice $k = \operatorname{rank}(A)$ gives the following special case of equation (17),

$$\sigma_{\min}(A) - \sigma_1(E) \leq \sigma_{\min}(A+E) \leq \sigma_{\min}(A) + \sigma_1(E)\,. \tag{18}$$

For the 2-norm $\sigma_1(\,\cdot\,) = \|\cdot\|$, and by Lemma 3.3 $\sigma_{\min}(\,\cdot\,) = \|\cdot\|_\ell$. This shows that equation (18) also is a special case of Lemma 4.2's triangle-like inequality,

$$\|A\|_\ell - \|E\| \leq \|A+E\|_\ell \leq \|A\|_\ell + \|E\|\,. \tag{19}$$

Finally, by Theorem 5.1 $\|\cdot\|_\ell$ is the distance (as measured by a general operator norm) of the enclosed matrix to the nearest rank deficient matrix. Can Lemma 4.2 be extended to replace the matrix lower bound in equation (19) by the distance to the set of matrices of a given lower rank?

13. Many celebrated theorems either interpret the singular values of general matrices (or the eigenvalues of symmetric matrices) in terms of extremal formulas, or bound them in terms of perturbational inequalities. Examples are the Courant-Fischer-Weyl theorem, the Wielandt-Hoffman theorem, and the Cauchy interlace theorem. Are there similar results that replace the singular values or the eigenvalues by the distances to the nearest matrices having given changes of rank as measured by a general operator norm?

14. The ϵ-and-δ estimates in Theorem 5.5 depend on the Jacobian matrix, which is assumed to be continuous, and on its lower bound, which is continuous by Corollary 4.3. Can this continuity be used to extend the Theorem's conclusions?

15. Do the constructions $x \mapsto x_T$ or $x \mapsto x_f$ in the proof of Theorem 5.5 lead to immediate proofs of the implicit or inverse function theorems?

Some of the proofs in this paper implicitly use the Heine-Borel theorem. That theorem is why a matrix attains its norm while a linear transformation among Banach spaces in general does not. Most of the statements of the theorems in this paper, however, are oblivious to dimension. Banach himself showed that a linear transformation from one Banach space onto another has lower bounds. Those lower bounds have a supremum which is a lower bound itself from the considerations underlying Lemma 2.3. Thus it should be possible to obtain many of the conclusions of this paper by replacing min's and max's by inf's and sup's where necessary.

15. Restate Definition 2.1 and all the results in this paper for linear transformations among abstract spaces. To the extent possible, remove the assumption that the underlying spaces have finite dimension.

References

[1] S. Banach. *Théorie des Opérations linéaires*. Subwencji Funduszu Kultury Narodowej, Warsaw, 1932. Translated to English in [2].

[2] S. Banach. *Theory of Linear Operators*. North-Holland, Amsterdam, 1987. Translated by F. Jellet from a 1979 reprinting of [1].

[3] R. G. Bartle. *The Elements of Real Analysis*. John Wiley & Sons, New York, second edition, 1976.

[4] L. Blum, F. Cucker, M. Shub, and S. Smale. *Complexity and Real Computation*. Springer-Verlag, New York, 1997.

[5] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, September 1936.

[6] D. K. Faddeev and V. N. Faddeeva. *Computational Methods of Linear Algebra*. W. H. Freeman and Company, San Francisco, 1963. Translated from the Russian edition of 1960 by R. C. Williams.

[7] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, second edition, 1989.

[8] L. M. Graves. Some mapping theorems. *Duke Mathematical Journal*, 17:111–114, 1950.

[9] J. F. Grcar. Min-max identities on boundaries of convex sets around the origin. Submitted for publication, 2002.

[10] A. S. Householder. *The Theory of Matrices in Numerical Analysis*. Blaisdell, New York, 1964. Reprinted by Dover, New York, 1975.

[11] W. Kahan. Numerical linear algebra. *Canadian Mathematical Bulletin*, 9(6):757–801, 1966.

[12] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, New York, 1969.

[13] N. Macrae. *John von Neumann*. Pantheon Books, New York, 1992.

[14] J. von Neumann. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100:295–320, 1928. Presented to the Göttingen Mathematical Society on 7th December 1926. Translated to English in [15]. Reprinted in [20, v. 6, pp. 1–26].

[15] J. von Neumann. On the theory of games of strategy. In A. W. Tucker and R. D. Luce, editors, *Contributions to the Theory of Games IV*, number 40 in Annals of Mathematical Studies, pages 13–42. Princeton University Press, Princeton, 1959. Translated from [14] by S. Bargmann.

[16] J. von Neumann and H. H. Goldstine. Numerical inverting of matrices of high order. *Bulletin of the American Mathematical Society*, 53(11):1021–1099, November 1947. Reprinted in [20, v. 5, pp. 479–557].

[17] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York, 1970.

[18] S. Simons. Minimax theorems and their proofs. In D.-Z. Du and P. M. Pardalos, editors, *Minimax and Applications*, pages 1–24. Kluwer Academic Publishers, Dordrecht, 1995.

[19] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory.* Academic Press, San Diego, 1990.

[20] A. H. Taub, editor. *John von Neumann Collected Works.* Macmillan, New York, 1963.

[21] J. H. Wilkinson. *Rounding Errors in Algebraic Processes.* Prentice Hall, Englewood Cliffs, New Jersey, 1963.

[22] H. Wittmeyer. Einfluß der Änderung einer Matrix auf der Lösung des zugehörigen Gleichungssystems, sowie auf die charakteristischen Zahlen und die Eigenvektoren. *Zeitschrift für angewandte Mathematik und Mechanik*, 16:287–300, 1936.