On the Calculation of the $l_2 \to l_1$ Induced Matrix Norm

Konstantinos Drakakis

UCD CASL, University College Dublin
Belfield, Dublin 4, Co. Dublin, Ireland
konstantinos.drakakis@ucd.ie

Barak A. Pearlmutter

Hamilton Institute, NUI Maynooth, Co. Kildare, Ireland
barak@cs.nuim.ie

Abstract

We show that the $l_2 \to l_1$ induced matrix norm, namely the norm induced by the $l_2$ and $l_1$ vector norms in the domain and range space, respectively, can be calculated as the maximal element of a finite set involving discrete additive combinations of the rows of the involved matrix with weights of $\pm 1$; the number of elements this set contains is exponential in the number of rows involved. A geometric interpretation of the result allows us to extend the result to some other induced norms. Finally, we generalize the findings to bounded linear operators on separable Banach spaces that can be obtained as strong limits of sequences of finite-dimensional linear operators.

Mathematics Subject Classification: 15A60, 47A30, 52A20

Keywords: $l_2 \to l_1$ induced matrix norm, induced norms on Banach spaces

1 Introduction

Induced matrix norms are routinely used in both pure and applied mathematics. In many applications only an estimate or bound on norm is necessary, and extensive results have been obtained on the estimation or bounding of various

\footnote{The author is also affiliated with the School of Electrical, Electronic, and Mechanical Engineering, University College Dublin.}

We derive a simple expression for the $l_2 \to l_1$ induced matrix norm, and generalize it to infinite-dimensional matrices that play the role of bounded linear transformations between separable Banach spaces. We also derive a simple expression for the $l_p \to l_q$ induced matrix norm when $0 < p \leq 1 \leq q$.

The usual $l_p$ norms\(^2\) over vectors, $\|x\|_p \equiv \left(\sum_{i=1}^{n}|x_i|^p\right)^{1/p}$ for $0 < p < \infty$ and $\|x\|_\infty \equiv \max_{i \in \{1,\ldots,n\}}|x_i|$, can be used to induce a norm on mappings between two normed spaces: the maximum ratio between the norm of a vector in the domain and the norm of the corresponding vector in the domain [Horn and Johnson(1985), pp. 257–335]. We consider a matrix as a linear mapping from one normed linear space to another, so the $\|\cdot\|_{pq}$-norm of a matrix $A : \mathbb{R}^{m \times n}$ is defined to be the maximal ratio of the $l_q$-norm of the image of a vector in the range space over the $l_p$-norm of that vector in the domain space. Because $\|cx\|_r = c\|x\|_r$ for $c > 0$, we can constrain the pre- or post-image vector to unit length without loss of generality.

$$\|A\|_{pq} \equiv \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|_q}{\|x\|_p} = \sup_{\{x \in \mathbb{R}^n : \|x\|_p = 1\}} \|Ax\|_q = \left(\min_{\{x \in \mathbb{R}^n : \|Ax\|_q = 1\}} \|x\|_p\right)^{-1} \quad (1)$$

The supremum here is a maximum, as the spaces considered are finite-dimensional and the unit $l_p$-sphere is bounded and closed. As shorthand, we define $\|A\|_p \equiv \|A\|_{pp}$.

Closed-form solutions are known for some induced norms, in particular cases involving the $l_1$, $l_2$, and $l_\infty$-norms [Higham(1992)]. We now list some $l_p \to l_q$ induced matrix norms and their closed-form exact solutions. These are listed in lexicographic order by $p$ and $q$. Novel expressions, derived below, are starred.

$$\|A\|_{pq} = \max_{j=1,\ldots,n} \|a^j\|_q \quad (p \leq 1 \leq q) \quad (2a^*)$$

$$\|A\|_1 = \max_{j=1,\ldots,n} \|a^j\|_1 = \|A^T\|_\infty \quad (2b)$$

$$\|A\|_{12} = \max_{j=1,\ldots,n} \|a^j\|_2 = \|A^T\|_\infty \quad (2c)$$

\(^2\)Although when $p < 1$ the triangle inequality does not hold, this does not affect us below. For clarity we write “$l_p$ norm” instead of “$l_p$ seminorm” or “$l_p$ norm-or-possibly-seminorm”, and similarly we write “induced norm” instead of “induced norm-or-possibly-seminorm.”
Calculation of the $l_2 \to l_1$ induced matrix norm

\[
\|A\|_{21} = \max_{s \in \{-1,+1\}^m} \|s^T A\|_2
\]  
\[
\|A\|_2 = \max_{\lambda \in \text{eig}(A^T A)} \sqrt{|\lambda|} = \|A^T\|_2
\]  
\[
\|A\|_{1\infty} = \max_{s \in \{-1,+1\}^m} \|A s\|_1
\]  
\[
\|A\|_{1\infty} = \max_{(i,j) \in \{1,\ldots,m\} \times \{1,\ldots,n\}} |a_{ij}|
\]  
\[
\|A\|_{\infty 2} = \max_{i=1,\ldots,m} \|a_i\|_2 = \|A^T\|_{12}
\]  
\[
\|A\|_{\infty} = \max_{i=1,\ldots,m} \|a_i\|_1 = \|A^T\|_1
\]  

where $\text{eig} X$ is the set of eigenvalues of the array $X$, the vector $a^j$ is the $j$-th column of $A$, and the vector $a_i$ is the $i$-th row of $A$.

Out of these, (2b), (2e), and (2i) are well known and can be found in most textbooks of matrix theory or numerical analysis; (2f) and (2g) appear in [Rohn(2000)]; (2c) and (2h) appear in [Higham(1992)] (the former is a special case of the more general novel expression (2a*)). We now proceed to prove the novel cases, (2a*) in Section 5, and (2d*) in Section 2, using a combination of algebraic and geometric arguments. Section 6 generalizes some of the results to linear operators on Banach spaces.

2 $\|\cdot\|_{21}$

We first show that the $\|A\|_{21}$ induced matrix norm can be computed by considering a maximum over a finite set of vectors.

**Theorem 1.** Let $A : \mathbb{R}^{n \times m}$, and let $a_i, i = 1, \ldots, m$ be the row vectors of $A$; then,

\[
\|A\|_{21} = \max_{s \in \{-1,+1\}^m} \|s^T A\|_2 = \max_{s \in \{-1,+1\}^m} \left\| \sum_{i=1}^{n} s_i a_i \right\|_2
\]  

**Proof.** We need the maximum of $\|A x\|_1$ subject to the constraint that $\|x\|_2 = 1$. Observe that the constraint is continuously differentiable, while the function to be maximized is continuous but only piecewise differentiable. Within a differentiability region, we may use Lagrange multipliers to maximize; over regions of non-differentiability, however, we need to find the maximum manually. Fortunately, we can prove that the global maximum has to lie within a differentiability region (see Figure 1), so this last step is not necessary.

First we compute the local maxima within each differentiability region using Lagrange multipliers. Consider the function

\[
f(x, \lambda) \equiv \|A x\|_1 + \lambda(\|x\|_2 - 1) = \sum_{i=1}^{m} \left\| \sum_{j=1}^{n} a_{ij} x_j \right\| + \lambda \left( \left( \sum_{i=1}^{n} x_i^2 \right)^{1/2} - 1 \right).
\]
We need to find the partial derivatives of $f$ with respect to $x_i$ and $\lambda$ and set them to zero, but the absolute values in the formula make this difficult. We can circumvent this by setting $s_i = \text{sign}(A x)_i$.

$$f(x, \lambda) = \sum_{i=1}^{m} s_i \sum_{j=1}^{n} a_{ij} x_j + \lambda \left( \left( \sum_{i=1}^{n} x_i^2 \right)^{1/2} - 1 \right)$$

This function is now clearly differentiable in the set

$$W = \mathbb{R}^n - \bigcup_{i=1}^{m} \{ x \in \mathbb{R}^n : a_i x = 0 \},$$

which is a collection of connected regions, separated by hyperplanes, within each of which $f$ is differentiable and the components of $s$ are constant. It follows that, in $W$,

$$\frac{\partial f}{\partial x_k} = \sum_{i=1}^{m} s_i a_{ik} + \lambda \frac{x_k}{\|x\|_2} = \sum_{i=1}^{m} s_i a_{ik} + \lambda x_k = 0,$$

so $x = -\frac{1}{\lambda} A^T s$, and therefore $|\lambda| = \| A^T s \|_2$ and

$$f(x, \lambda) = -\frac{1}{\lambda} \sum_{i=1}^{m} s_i \sum_{j=1}^{n} a_{ik} \sum_{j=1}^{m} s_j a_{jk} = -\frac{1}{\lambda} \sum_{k=1}^{n} \left( \sum_{i=1}^{m} s_i a_{ik} \right) \left( \sum_{j=1}^{m} s_j a_{jk} \right) = \frac{\lambda^2}{\lambda} = \lambda.$$

Observe further that $\sum_{i=1}^{m} s_i a_i = s^T A$ so $\left\| \sum_{i=1}^{m} s_i a_i \right\|_2 = \sqrt{s^T AA^T s}$.

What about the values of $f$ outside $W$? Can perhaps the maximum be found there? We argue that this is not possible. Let $A : \mathbb{R}^{m \times n}$ be an array for which the vector $x$ yielding the maximum satisfies, without loss of generality, the $l \leq m$ conditions $a_i x = 0$, $i = 1, \ldots, l$. Now form the vector

$$y = \frac{x + \epsilon a_1}{\sqrt{1 + \epsilon^2 |a_1|^2}},$$

and assume that $\epsilon$ is small enough so that $s_i = \text{sign}(a_i y)$, $i = l + 1, \ldots, m$. Then,

$$\sum_{i=1}^{m} a_i y = \frac{1}{\sqrt{1 + \epsilon^2 |a_1|^2}} \left[ \sum_{i=l+1}^{m} |a_i x| + |\epsilon| \sum_{i=1}^{l} |a_i a_1| + \epsilon \sum_{i=l+1}^{m} s_i a_i a_1 \right]$$

$$= \sum_{i=l+1}^{m} |a_i x| + |\epsilon| \sum_{i=1}^{l} |a_i a_1| + \epsilon \sum_{i=l+1}^{m} s_i a_i a_1 + O(\epsilon^2)$$

$$= \sum_{i=l+1}^{m} |a_i x| + C_1 |\epsilon| + C_2 \epsilon + O(\epsilon^2)$$

where $C_1$ and $C_2$ are constants.
Assuming $a_1 \neq 0$, we get $C_1 > 0$, while regardless of the sign of $C_2$ we can choose the sign of $\epsilon$ appropriately so that $C_2 \epsilon \geq 0$. This makes $C_1 |\epsilon| + C_2 \epsilon > 0$, and

$$\sum_{i=1}^{m} |a_i y| > \sum_{i=l+1}^{m} |a_i x|$$

contradicting the assumption that the maximum occurs at $x$.

The requirement that $a_1 \neq 0$ can be eased by the continuity of vector norms, which allows us to approximate $A$ by a sequence of arrays $\{A_k\}_{k \in \mathbb{N}}$ with non-zero rows. This completes the proof.

As a final observation, we note that the maxima occur in pairs: if $s$ is a maximizing vector then $-s$ is as well.

3 Geometric intuition

Let us now develop a geometric intuition for the $l_2 \to l_1$ induced norm of $A$. We wish to find the smallest $t > 0$ such that

$$\exists x : \|x\|_2 = t \text{ given that } \|Ax\|_1 = 1.$$
Imagine we begin with a very small \( t \), and that we increase it gradually. Considered geometrically, this corresponds to expanding the sphere \( \{ x | \| x \|_2 = t \} \), as well as its image under \( A \), which is an \( m \)-dimensional ellipsoid, until this image, which lies fully within the \( m \)-dimensional octahedron \( \{ y : \| y \|_1 = 1 \} \) with vertices at \( y_i = \pm 1, \ i = 1, \ldots, m \), eventually touches the hyper-octahedron. The point of this first contact cannot be a vertex of the octahedron, because the ellipsoid is smooth and convex, so it must lie in the interior of a facet. Moreover, points of contact will occur in pairs on opposite and symmetric sides with respect to the origin. These two observations correspond to the results shown above that maximization solutions occur in pairs and that they occur within \( W \).

The same intuition with the shape of the balls in the domain and range spaces switched yields (2c), as a vertex of the transformed octahedron from the domain space will be the first point of contact.

### 4 Computational complexity of \( \| \cdot \|_{21} \)

The calculation of \( \| A \|_{21} \) when \( A : \mathbb{R}^{n \times m} \), by (2d*), requires a combinatorial search, i.e., the calculation of \( \| A^T s \|_2 \) for each \( s \in \{-1, +1\}^m \). This requires \( O(nm2^m) \) operations.

A simple argument shows that a maximizing \( s \) will be in \( \{-1, +1\}^m \). Let \( s' = s - s_i e_i \) and \( a_i \) be the \( i \)-th row of \( A \). Expanding, we obtain:

\[
\| A^T s \|_2 = s^T A A^T s = (s')^T A A^T s' + s_i^2 \| a_i \|_2^2 + 2 s_i (s')^T A A^T e_i
\]

This expression makes it clear that, keeping all elements of \( s \) fixed except \( s_i \), \( \| A^T s \|_2 \) is maximized when \( s_i \) becomes +1 or −1, according to whether \( (s')^T A A^T e_i \) is positive or negative, respectively; if it is 0 it obviously makes no difference. Repeating this argument for \( i = 1, \ldots, m \) proves the claim.

### 5 \( \| \cdot \|_{pq} \) for \( p \leq 1 \leq q \)

We can generalize (2c), the well known result for the \( l_1 \to l_2 \) induced matrix norm, to the \( l_p \to l_q \) induced matrix norm where \( p \leq 1 \leq q \).

**Theorem 2.**

\[
\| A \|_{pq} = \max_{j=1,\ldots,n} \| a_j \|_q \quad (p \leq 1 \leq q)
\]

**Proof.** We use a simple geometric argument in which the roles of the balls in Figure 1 are reversed, along with considerations of convexity. We wish to find the smallest \( t > 0 \) such that \( \exists x : \| x \|_p = t \) and \( \| Ax \|_q = 1 \). Imagine once more we begin with a very small \( t \), and that we increase it gradually.
Considered geometrically, this corresponds to expanding the closed surface 
\( P(t) = \{ x : \| x \|_p = t \} \), and therefore \( AP(t) \), the image of \( P(t) \) under \( A \), until this image, which lies fully within the closed surface 
\( Q = \{ y : \| y \|_q = 1 \} \), touches \( Q \).

If \( Q \) is convex, then the first point of contact between \( AP(t) \) and \( Q \) will be on a point of \( AP(t) \) that is on the convex hull of \( AP(t) \). When \( p < 1 \), the points of \( AP(t) \) on the convex hull of \( AP(t) \) are in the finite set \( \{ \pm t a_j : j = 1, \ldots, n \} \). When \( q \geq 1 \), the set \( Q \) is the boundary of a convex set. Together, these imply that we can find the minimal \( t \) separately for each \( \{ t a_j \} \), and take the smallest of them. This establishes the result for \( p < 1 \). A limit argument based on the continuity of the norm shows that this continues to hold when \( p = 1 \).

\[ \square \]

6 Induced norms on Banach spaces

How do our results need to be modified if we consider infinite-dimensional spaces instead? Since we are considering normed vector spaces, let us focus on Banach spaces, because they have the additional desirable feature of completeness. To avoid unnecessarily complicated cases, let us narrow our scope even further and exclusively consider separable Banach spaces, namely Banach spaces with a countable basis; after all, the most common ones are of this type.

Let then \( A : B_1 \to B_2 \) be a bounded linear operator from a separable (possibly complex) Banach space \( B_1 \) into another \( B_2 \), both equipped with (possibly different) \( l_p \) norms. Since the elements of a separable Banach space can be represented as vectors of countably many coordinates, \( A \) can be considered to be an array with countably many rows and columns (which we will also denote by \( A \)). As \( \| A \| < \infty \) by assumption, then \( \| A^T \| = \| A^* \| = \| A \| < \infty \).

Let now \( A_{m,n} : \mathbb{R}^{n \times m} \) be the constraint of \( A \) on the first \( n \) basis vectors of \( B_1 \) followed by the projection on the first \( m \) basis vectors of \( B_2 \). We want to investigate conditions under which there exists a sequence \( A_k = A_{m(k),n(k)} \), \( k \in \mathbb{N}^* \) such that \( \lim \| A_k - A \| = 0 \).

**Theorem 3.** Let \( A : B_1 \to B_2 \) be a bounded linear operator from a separable Banach space \( B_1 \) into another \( B_2 \), equipped with the norms \( \| \cdot \|_{p_1} \) and \( \| \cdot \|_{p_2} \), respectively, where \( 1 \leq p_1, p_2 < \infty \), let \( a_k \), \( k \in \mathbb{N}^* \) denote the rows of \( A \), let \( A_{m,n} : \mathbb{R}^{n \times m} \) be the constraint of \( A \) on the first \( n \) basis vectors of \( B_1 \) followed by the projection on the first \( m \) basis vectors of \( B_2 \), and let \( p_1' \) be such that

\[
\frac{1}{p_1'} + \frac{1}{p_1} = 1.
\]

If

\[
\sum_{k=1}^{\infty} \| a_k \|_{p_1'}^{p_2} < \infty \quad \text{[Row Norm Summability (RNS)]}
\]

then

\[
\lim_{k \to \infty} \| A_{m(k),n(k)} - A \| = 0.
\]

\[ \square \]
then there exists a sequence \( A_k = A_{m(k), n(k)}, \ k \in \mathbb{N}^* \), such that \( \lim \| A_k - A \| = 0 \).

**Proof.** The first step is to observe that the vector subspace of \( B_1 \) consisting of the union of the finite-dimensional subspaces of \( B_1 \) consisting of the first \( n \) basis vectors for all \( n \in \mathbb{N}^* \) is dense in \( B_1 \). Similarly, the image of this subspace under \( A \) is dense in \( AB_1 \subset B_2 \).

The second step is to write down explicitly what \( \|(A - A_k)x\|_{p_2} \) is when \( \|x\|_{p_1} = 1 \):

\[
\|(A - A_k)x\|_{p_2} = \sum_{l=1}^{m(k)} \left| a_{li}x_i \right|^{p_2} + \sum_{l=m(k)+1}^{\infty} \left| a_k \cdot x \right|^{p_2}.
\]

However, by Hölder’s inequality,

\[
|a_k \cdot x| \leq \|a_k\|_{p_1'} \|x\|_{p_1} = \|a_k\|_{p_1'}
\]

whence

\[
\|(A - A_k)x\|_{p_2} \leq \sum_{l=1}^{m(k)} \left| \sum_{i=n(k)+1}^{\infty} \left| a_{li} \right|^{p_2} \right|^{p_1' / p_2} + \sum_{l=m(k)+1}^{\infty} \left| a_k \right|^{p_2}.
\]

Let now

\[
m(k) = \min \left\{ m \in \mathbb{N} : \sum_{l=m+1}^{\infty} \left| a_k \right|_{p_1'}^{p_2} < \frac{1}{2k} \right\}
\]

and, having chosen \( m(k) \), let

\[
n(k) = \min \left\{ n \in \mathbb{N} : \max_{l=1, \ldots, m(k)} \sum_{i=n+1}^{\infty} |a_{li}|^{p_2} < \left( \frac{1}{2km(k)} \right)^{p_1' / p_2} \right\}.
\]

It follows that

\[
\|(A - A_k)x\|_{p_2} \leq \frac{1}{k} \tag{4}
\]

which implies that

\[
\sup_{\{x \in B_1 : \|x\|_{p_1} = 1\}} \|(A - A_k)x\|_{p_2} = \|A - A_k\|_{p_2} \leq \frac{1}{k} \tag{5}
\]

which finally results in

\[
\lim_{k \to \infty} \|A - A_k\|_{p_2} = 0. \tag{6}
\]

A “3-\( \epsilon \)” argument, to include vectors with infinitely many nonzero coefficients over the chosen bases, finishes the proof. \( \square \)
This theorem proves that the results proved in Theorems 1 and 2 remain valid for arrays of infinite dimension satisfying the RNS condition. Specifically in relation to Theorem 1, we make the following observation: Theorems 1 and 3 combined yield that

$$\|A\|_{21} = \lim_{k \to \infty} \max_{s \in \{-1,1\}^n(k)} \sqrt{s^T A_k A_k^T s} = \max_{s \in \{-1,1\}^\infty} \sqrt{s^T A A^T s}$$

despite the fact that the maximizing vector $s$ where $s_i = \pm 1$, $i \in \mathbb{N}$, is no longer a vector of finite norm. Therefore, the supremum involved in the computation of $\|\cdot\|_{21}$ in this case is a genuine supremum, not a maximum, in the sense that there exists no vector $x$ of finite norm, hence within the vector space, for which $\|A\|_{21} = x^T A A^T x$. Instead, if $A$ is a continuous linear operator between the spaces defined in the theorem, there exists a sequence of vectors $x_n$, $n \in \mathbb{N}$, such that $\lim \|A x_n\|_{1} = \|A\|_{21}$ and $\|x_n\|_2 = 1$.

7 Summary and conclusion

Using a combination of analytical and geometrical arguments, we have shown that:

- The induced norm $\|A\|_{21}$ can be computed as the maximal value of a quadratic form over a finite, albeit exponentially sized, set of vectors (Theorem 1).

- The induced norm $\|A\|_{pq}$, $p \leq 1 \leq q$ is the maximal $l_q$ norm of the columns of $A$ (Theorem 2).

- Analogous results hold for bounded linear operators on separable Banach spaces that satisfy RNS (Theorem 3).

Acknowledgements

This material is based upon works supported by the Science Foundation Ireland under Grant No. 06/MI/006 (Claude Shannon Institute), and 08/RFP/MTH1164.

References


Received: October, 2008