A top nine list: Most popular induced matrix norms

Andrew D. Lewis*

2010/03/20

Abstract

Explicit formulae are given for the nine possible induced matrix norms corresponding to the 1-, 2-, and ∞-norms for Euclidean space. The complexity of computing these norms is investigated.

Keywords. Induced norm.

AMS Subject Classifications (2020). 15A60

1. Introduction

Arguably the most commonly used norms for real Euclidean space $\mathbb{R}^n$ are the norms $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ defined by

$$\|\boldsymbol{x}\|_1 = \sum_{j=1}^n |x_j|, \quad \|\boldsymbol{x}\|_2 = \Big(\sum_{j=1}^n |x_j|^2\Big)^{1/2}, \quad \|\boldsymbol{x}\|_\infty = \max\{|x_1|, \dots, |x_n|\},$$

respectively, for $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$. Let $\mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$ be the set of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$, which we identify with the set of $m \times n$ matrices in the usual way. If $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$ and if $p, q \in \{1, 2, \infty\}$ then the norm of $\boldsymbol{A}$ induced by the p-norm on $\mathbb{R}^n$ and the q-norm on $\mathbb{R}^m$ is

$$\|\boldsymbol{A}\|_{p,q} = \sup\{\|\boldsymbol{A}(\boldsymbol{x})\|_q \mid \|\boldsymbol{x}\|_p = 1\}.$$

This is well-known to define a norm on $\mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$. There are other equivalent characterisations of the induced norm, but the one given above is the only one we will need. We refer to [Horn and Johnson 2013] for a general discussion of induced matrix norms.

For certain combinations of (p, q), explicit expressions for $\|\cdot\|_{p,q}$ are known. For example, in [Horn and Johnson 2013] expressions are given in the cases $(1, 1)$ (in §5.6.4), $(2, 2)$ (§5.6.6), and (∞, ∞) (§5.6.5). In [Rohn 2000] the case $(\infty, 1)$ is studied, and its computation is shown to be NP-hard. The case $(2, 1)$ is given by Drakakis and Pearlmutter [2009], although the details of the degenerate case given there are a little sketchy. Drakakis and Pearlmutter also list all of the other combinations except $(2, \infty)$, for which no expression seems to be available, and which we give here, apparently for the first time. The formula given by Drakakis and Pearlmutter for $(\infty, 2)$ is presented without reference or proof, and is incorrect, probably a typographical error.

Here we present the correct formulae for all nine of the induced norms. Although most of these formulae are known in the literature, we give proofs in all nine cases so that, for the

*Professor, Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada
Email: `andrew.lewis@queensu.ca`, URL: `http://www.mast.queensu.ca/~andrew/`

first time, all proofs for all cases are given in one place. We also analyse the computational complexity of computing these various norms.

Here is the notation we use. By $\{\boldsymbol{e}_1,\dots,\boldsymbol{e}_n\}$ we denote the standard basis for $\mathbb{R}^n$. For a matrix $\boldsymbol{A}\in\mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$, $\boldsymbol{r}(\boldsymbol{A},a)\in\mathbb{R}^n$ denotes the ath row and $\boldsymbol{c}(\boldsymbol{A},j)\in\mathbb{R}^m$ denotes the jth column. The components of $\boldsymbol{A}$ are denoted by A_{aj}, $a\in\{1,\dots,m\}$, $j\in\{1,\dots,n\}$. The transpose of $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^T$. The Euclidean inner product is denoted by $\langle\cdot,\cdot\rangle$. For a differentiable map $\boldsymbol{f}\colon\mathbb{R}^n\to\mathbb{R}^m$, $\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x})\in\mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$ denotes the derivative of $\boldsymbol{f}$ at $\boldsymbol{x}$. For a set X, $\mathbf{2}^X$ denotes the power set of X.

2. Formulae for induced norms

1 Theorem: *Let $p,q\in\{1,2,\infty\}$ and let $\boldsymbol{A}\in\mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$. The induced norm $\|\cdot\|_{p,q}$ satisfies the following formulae:*

(i) $\|\boldsymbol{A}\|_{1,1}=\max\{\|\boldsymbol{c}(\boldsymbol{A},j)\|_1\mid j\in\{1,\dots,n\}\}$;

(ii) $\|\boldsymbol{A}\|_{1,2}=\max\{\|\boldsymbol{c}(\boldsymbol{A},j)\|_2\mid j\in\{1,\dots,n\}\}$;

(iii) $$\begin{aligned}\|\boldsymbol{A}\|_{1,\infty}&=\max\{|A_{aj}|\mid a\in\{1,\dots,m\},\ j\in\{1,\dots,n\}\};\\&=\max\{\|\boldsymbol{c}(\boldsymbol{A},j)\|_\infty\mid j\in\{1,\dots,n\}\}\\&=\max\{\|\boldsymbol{r}(\boldsymbol{A},a)\|_\infty\mid a\in\{1,\dots,m\}\}\end{aligned}$$

(iv) $\|\boldsymbol{A}\|_{2,1}=\max\{\|\boldsymbol{A}^T(\boldsymbol{u})\|_2\mid \boldsymbol{u}\in\{-1,1\}^m\}$;

(v) $\|\boldsymbol{A}\|_{2,2}=\max\{\sqrt{\lambda}\mid \lambda$ *is an eigenvalue for* $\boldsymbol{A}^T\boldsymbol{A}\}$;

(vi) $\|\boldsymbol{A}\|_{2,\infty}=\max\{\|\boldsymbol{r}(\boldsymbol{A},a)\|_2\mid a\in\{1,\dots,m\}\}$;

(vii) $\|\boldsymbol{A}\|_{\infty,1}=\max\{\|\boldsymbol{A}(\boldsymbol{u})\|_1\mid \boldsymbol{u}\in\{-1,1\}^n\}$;

(viii) $\|\boldsymbol{A}\|_{\infty,2}=\max\{\|\boldsymbol{A}(\boldsymbol{u})\|_2\mid \boldsymbol{u}\in\{-1,1\}^n\}$;

(ix) $\|\boldsymbol{A}\|_{\infty,\infty}=\max\{\|\boldsymbol{r}(\boldsymbol{A},a)\|_1\mid a\in\{1,\dots,m\}\}$.

Proof: (i) We compute

$$\begin{aligned}\|\boldsymbol{A}\|_{1,1}&=\sup\{\|\boldsymbol{A}(\boldsymbol{x})\|_1\mid \|\boldsymbol{x}\|_1=1\}\\&=\sup\Big\{\sum_{a=1}^m|\langle\boldsymbol{r}(\boldsymbol{A}(\boldsymbol{x})),\boldsymbol{x}\rangle|\ \Big|\ \|\boldsymbol{x}\|_1=1\Big\}\\&\le\sup\Big\{\sum_{a=1}^m\sum_{j=1}^n|A_{aj}||x_j|\ \Big|\ \|\boldsymbol{x}\|_1=1\Big\}\\&=\sup\Big\{\sum_{j=1}^n|x_j|\Big(\sum_{a=1}^m|A_{aj}|\Big)\ \Big|\ \|\boldsymbol{x}\|_1=1\Big\}\\&\le\max\Big\{\sum_{a=1}^m|A_{aj}|\ \Big|\ j\in\{1,\dots,n\}\Big\}\\&=\max\{\|\boldsymbol{c}(\boldsymbol{A},j)\|_1\mid j\in\{1,\dots,n\}\}.\end{aligned}$$

To establish the opposite inequality, suppose that $k\in\{1,\dots,n\}$ is such that

$$\|\boldsymbol{c}(\boldsymbol{A},k)\|_1=\max\{\|\boldsymbol{c}(\boldsymbol{A},j)\|_1\mid j\in\{1,\dots,n\}\}.$$

Then,

$$\|\boldsymbol{A}(\boldsymbol{e}_k)\|_1 = \sum_{a=1}^{m}\Big|\Big(\sum_{j=1}^{n} A_{aj}\boldsymbol{e}_{k,j}\Big)\Big| = \sum_{a=1}^{m}|A_{ak}| = \|\boldsymbol{c}(\boldsymbol{A},k)\|_1.$$

Thus

$$\|\boldsymbol{A}\|_{1,1} \geq \max\{\|\boldsymbol{c}(\boldsymbol{A},j)\|_1 \mid j \in \{1,\dots,n\}\},$$

since $\|\boldsymbol{e}_k\|_1 = 1$.

(ii) We compute

$$\begin{aligned}
\|\boldsymbol{A}\|_{1,2} &= \sup\{\|\boldsymbol{A}(\boldsymbol{x})\|_2 \mid \|\boldsymbol{x}\|_1 = 1\} \\
&= \sup\Big\{\Big(\sum_{a=1}^{m}\langle \boldsymbol{r}(\boldsymbol{A},a),\boldsymbol{x}\rangle^2\Big)^{1/2} \;\Big|\; \|\boldsymbol{x}\|_1 = 1\Big\} \\
&\leq \sup\Big\{\Big(\sum_{a=1}^{m}\Big(\sum_{j=1}^{n}|A_{aj}x_j|\Big)^2\Big)^{1/2} \;\Big|\; \|\boldsymbol{x}\|_1 = 1\Big\} \\
&\leq \sup\Big\{\Big(\sum_{a=1}^{m}(\max\{|A_{aj}| \mid j \in \{1,\dots,n\}\})^2\Big(\sum_{j=1}^{n}|x_j|\Big)^2\Big)^{1/2} \;\Big|\; \|\boldsymbol{x}\|_1 = 1\Big\} \\
&= \Big(\sum_{a=1}^{m}(\max\{|A_{aj}| \mid j \in \{1,\dots,n\}\})^2\Big)^{1/2} \\
&= \Big(\max\Big\{\sum_{a=1}^{m}A_{aj}^2 \;\Big|\; j \in \{1,\dots,n\}\Big\}\Big)^{1/2} = \max\{\|\boldsymbol{c}(\boldsymbol{A},j)\|_2 \mid j \in \{1,\dots,n\}\},
\end{aligned}$$

using the fact that

$$\sup\{\|\boldsymbol{x}\|_2 \mid \|\boldsymbol{x}\|_1 = 1\} = 1.$$

To establish the other inequality, note that if we take $k \in \{1,\dots,n\}$ such that

$$\|\boldsymbol{c}(\boldsymbol{A},k)\|_2 = \max\{\|\boldsymbol{c}(\boldsymbol{A},j)\|_2 \mid j \in \{1,\dots,n\}\},$$

then we have

$$\|\boldsymbol{A}(\boldsymbol{e}_k)\|_2 = \Big(\sum_{a=1}^{m}\Big(\sum_{j=1}^{n}A_{aj}\boldsymbol{e}_{k,j}\Big)^2\Big)^{1/2} = \Big(\sum_{a=1}^{m}A_{ak}^2\Big)^{1/2} = \|\boldsymbol{c}(\boldsymbol{A},k)\|_2.$$

Thus

$$\|\boldsymbol{A}\|_{1,2} \geq \max\{\|\boldsymbol{c}(\boldsymbol{A},j)\|_2 \mid j \in \{1,\dots,n\}\},$$

since $\|\boldsymbol{e}_k\|_1 = 1$.

(iii) Here we compute

$$\begin{aligned}
\|\boldsymbol{A}\|_{1,\infty} &= \sup\{\|\boldsymbol{A}(\boldsymbol{x})\|_\infty \mid \|\boldsymbol{x}\|_1 = 1\} \\
&= \sup\Big\{\max\Big\{\Big|\sum_{j=1}^{n}A_{aj}x_j\Big| \;\Big|\; a \in \{1,\dots,m\}\Big\} \;\Big|\; \|\boldsymbol{x}\|_1 = 1\Big\} \\
&\leq \sup\Big\{\max\Big\{|A_{aj}| \;\Big|\; j \in \{1,\dots,n\},\ a \in \{1,\dots,m\}\Big\}\Big(\sum_{j=1}^{n}|x_j|\Big) \;\Big|\; \|\boldsymbol{x}\|_1 = 1\Big\} \\
&= \max\{|A_{aj}| \mid j \in \{1,\dots,n\},\ a \in \{1,\dots,m\}\}.
\end{aligned}$$

For the converse inequality, let $k \in \{1, \dots, n\}$ be such that

$$\max\{|A_{ak}| \mid a \in \{1, \dots, m\}\} = \max\{|A_{aj}| \mid j \in \{1, \dots, n\},\ a \in \{1, \dots, m\}\}.$$

Then

$$\begin{aligned}\|\boldsymbol{A}(\boldsymbol{e}_k)\|_\infty &= \max\Big\{\Big|\sum_{j=1}^n A_{aj}\boldsymbol{e}_{k,j}\Big| \ \Big|\ a \in \{1, \dots, m\}\Big\} \\ &= \max\{|A_{ak}| \mid a \in \{1, \dots, m\}\}.\end{aligned}$$

Thus

$$\|\boldsymbol{A}\|_{1,\infty} \geq \max\{|A_{aj}| \mid j \in \{1, \dots, n\},\ a \in \{1, \dots, m\}\},$$

since $\|\boldsymbol{e}_k\|_1 = 1$.

(iv) In this case we maximise the function $\boldsymbol{x} \mapsto \|\boldsymbol{A}(\boldsymbol{x})\|_1$ subject to the constraint that $\|\boldsymbol{x}\|_2^2 = 1$. We shall do this using the Lagrange Multiplier Theorem [e.g., Edwards 1973, §II.5], defining

$$f(\boldsymbol{x}) = \|\boldsymbol{A}(\boldsymbol{x})\|_1, \quad g(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2 - 1.$$

Let us first assume that none of the rows of $\boldsymbol{A}$ are zero. We must exercise some care because f is not differentiable on $\mathbb{R}^n$. However, f is differentiable at points off the set

$$B_{\boldsymbol{A}} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \text{there exists } a \in \{1, \dots, m\} \text{ such that } \langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{x}\rangle = 0\}.$$

To facilitate computations, let us define $\boldsymbol{u}_{\boldsymbol{A}} \colon \mathbb{R}^n \to \mathbb{R}^m$ by asking that

$$u_{\boldsymbol{A},a}(\boldsymbol{x}) = \operatorname{sign}(\langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{x}\rangle).$$

Note that $B_{\boldsymbol{A}} = \boldsymbol{u}_{\boldsymbol{A}}^{-1}(\boldsymbol{0})$ and that on $\mathbb{R}^n \setminus B_{\boldsymbol{A}}$ the function $\boldsymbol{u}_{\boldsymbol{A}}$ is locally constant. Moreover, it is clear that

$$f(\boldsymbol{x}) = \langle \boldsymbol{u}_{\boldsymbol{A}}(\boldsymbol{x}), \boldsymbol{A}(\boldsymbol{x})\rangle.$$

Now let $\boldsymbol{x}_0 \in \mathbb{R}^n \setminus B_{\boldsymbol{A}}$ be a maximum of f subject to the constraint that $g(\boldsymbol{x}) = 0$. One easily verifies that $\boldsymbol{D}g$ has rank 1 at points that satisfy the constraint. Thus, by the Lagrange Multiplier Theorem, there exists $\lambda \in \mathbb{R}$ such that

$$\boldsymbol{D}(f - \lambda g)(\boldsymbol{x}_0) = \boldsymbol{0}.$$

We compute

$$\boldsymbol{D}f(\boldsymbol{x}_0) \cdot \boldsymbol{v} = \langle \boldsymbol{u}_{\boldsymbol{A}}(\boldsymbol{x}_0), \boldsymbol{A}(\boldsymbol{v})\rangle, \quad \boldsymbol{D}g(\boldsymbol{x}) \cdot \boldsymbol{v} = 2\langle \boldsymbol{x}, \boldsymbol{v}\rangle.$$

Thus $\boldsymbol{D}(f - \lambda g)(\boldsymbol{x}_0) = \boldsymbol{0}$ if and only if

$$\boldsymbol{A}^T(\boldsymbol{u}_{\boldsymbol{A}}(\boldsymbol{x}_0)) = 2\lambda \boldsymbol{x}_0 \quad \Longrightarrow \quad |\lambda| = \frac{1}{2}\|\boldsymbol{A}^T(\boldsymbol{u}_{\boldsymbol{A}}(\boldsymbol{x}_0))\|_2,$$

since $\|\boldsymbol{x}_0\|_2 = 1$. Thus $\lambda = 0$ if and only if $\boldsymbol{A}^T(\boldsymbol{u}_{\boldsymbol{A}}(\boldsymbol{x}_0)) = \boldsymbol{0}$. Therefore, if $\lambda = 0$, then $f(\boldsymbol{x}_0) = 0$. We can disregard this possibility since f cannot have a maximum of zero as we are assuming that $\boldsymbol{A}$ has no zero rows. As $\lambda \neq 0$ we have

$$f(\boldsymbol{x}_0) = \langle \boldsymbol{A}^T(\boldsymbol{u}_{\boldsymbol{A}}(\boldsymbol{x}_0)), \boldsymbol{x}_0\rangle = \frac{1}{2\lambda}\|\boldsymbol{A}^T(\boldsymbol{u}_{\boldsymbol{A}}(\boldsymbol{x}_0))\|_2^2 = 2\lambda.$$

We conclude that, at solutions of the constrained maximisation problem, we must have

$$f(\boldsymbol{x}_0) = \|\boldsymbol{A}^T(\boldsymbol{u})\|_2,$$

where $\boldsymbol{u}$ varies over the nonzero points in the image of $\boldsymbol{u}_{\boldsymbol{A}}$, i.e., over points from $\{-1,1\}^m$.

This would conclude the proof of this part of the theorem in the case that $\boldsymbol{A}$ has no zero rows, but for the fact that it is possible that f attains its maximum on $B_{\boldsymbol{A}}$. We now show that this does not happen. Let $\boldsymbol{x}_0 \in B_{\boldsymbol{A}}$ satisfy $\|\boldsymbol{x}_0\|_2 = 1$ and denote

$$A_0 = \{a \in \{1,\dots,m\} \mid u_{\boldsymbol{A},a}(\boldsymbol{x}_0) = 0\}.$$

Let $A_1 = \{1,\dots,m\} \setminus A_0$. Let $a_0 \in A_0$. For $\epsilon \in \mathbb{R}$ define

$$\boldsymbol{x}_\epsilon = \frac{\boldsymbol{x}_0 + \epsilon \boldsymbol{r}(\boldsymbol{A}, a_0)}{\sqrt{1 + \epsilon^2 \|\boldsymbol{r}(\boldsymbol{A}, a_0)\|_2^2}}.$$

Note that $\boldsymbol{x}_\epsilon$ satisfies the constraint $\|\boldsymbol{x}_\epsilon\|_2^2 = 1$. Now let $\epsilon_0 \in \mathbb{R}_{>0}$ be sufficiently small that

$$\langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{x}_\epsilon \rangle \neq 0$$

for all $a \in A_1$ and $\epsilon \in [-\epsilon_0, \epsilon_0]$. Then we compute

$$\begin{aligned}
\|\boldsymbol{A}(\boldsymbol{x}_\epsilon)\|_1 &= \sum_{a=1}^{m} |\langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{x}_0 \rangle + \epsilon \langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{r}(\boldsymbol{A}, a_0) \rangle| + O(\epsilon^2) \\
&= \sum_{a \in A_0} |\epsilon| |\langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{r}(\boldsymbol{A}, a_0) \rangle| \\
&\quad + \sum_{a \in A_1} |\langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{x}_0 \rangle + \epsilon \langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{r}(\boldsymbol{A}, a_0) \rangle| + O(\epsilon^2). \qquad (1)
\end{aligned}$$

Since we are assuming that none of the rows of $\boldsymbol{A}$ are zero,

$$\sum_{a \in A_0} |\epsilon| |\langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{r}(\boldsymbol{A}, a_0) \rangle| > 0 \qquad (2)$$

for $\epsilon \in [-\epsilon_0, \epsilon_0]$, as long as ϵ_0 is sufficiently small. Now take $a \in A_1$. If ϵ is sufficiently small we can write

$$|\langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{x}_0 \rangle + \epsilon \langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{r}(\boldsymbol{A}, a_0) \rangle| = |\langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{x}_0 \rangle| + \epsilon C_a$$

for some $C_a \in \mathbb{R}$. As a result, and using (1), we have

$$\|\boldsymbol{A}(\boldsymbol{x}_\epsilon)\|_1 = \|\boldsymbol{A}(\boldsymbol{x}_0)\|_1 + \sum_{a \in A_0} (|\epsilon| |\langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{r}(\boldsymbol{A}, a_0) \rangle| + \epsilon \sum_{a \in A_1} C_a + O(\epsilon^2).$$

It therefore follows, by choosing ϵ_0 to be sufficiently small, that we have

$$\|\boldsymbol{A}(\boldsymbol{x}_\epsilon)\|_1 > \|\boldsymbol{A}(\boldsymbol{x}_0)\|_1$$

either for all $\epsilon \in [-\epsilon_0, 0)$ or for all $\epsilon \in (0, \epsilon_0]$, taking (2) into account. Thus if $\boldsymbol{x}_0 \in B_{\boldsymbol{A}}$ then $\boldsymbol{x}_0$ is not a local maximum for f subject to the constraint $g^{-1}(0)$.

Finally, suppose that $\boldsymbol{A}$ has some rows that are zero. Let

$$A_0 = \{a \in \{1, \dots, m\} \mid \boldsymbol{r}(\boldsymbol{A}, a) = \boldsymbol{0}\}$$

and let $A_1 = \{1, \dots, m\} \setminus A_0$. Let $A_1 = \{a_1, \dots, a_k\}$ with $a_1 < \cdots < a_k$, and define $\hat{\boldsymbol{A}} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^k)$ by

$$\hat{\boldsymbol{A}}(\boldsymbol{x}) = \sum_{r=1}^{k} \langle \boldsymbol{r}(\boldsymbol{A}, a_r), \boldsymbol{x} \rangle \boldsymbol{e}_r,$$

and note that $\|\boldsymbol{A}(\boldsymbol{x})\|_1 = \|\hat{\boldsymbol{A}}(\boldsymbol{x})\|_1$ for every $\boldsymbol{x} \in \mathbb{R}^n$. If $\boldsymbol{y} \in \mathbb{R}^m$ define $\hat{\boldsymbol{y}} \in \mathbb{R}^k$ by removing from $\boldsymbol{y}$ the elements corresponding to the zero rows of $\boldsymbol{A}$:

$$\hat{\boldsymbol{y}} = (y_{a_1}, \dots, y_{a_k}).$$

Then we easily determine that $\boldsymbol{A}^T(\boldsymbol{y}) = \hat{\boldsymbol{A}}^T(\hat{\boldsymbol{y}})$. Therefore,

$$\begin{aligned}
\|\boldsymbol{A}\|_{2,1} &= \sup\{\|\boldsymbol{A}(\boldsymbol{x})\|_1 \mid \|\boldsymbol{x}\|_2 = 1\} \\
&= \sup\{\|\hat{\boldsymbol{A}}(\boldsymbol{x})\|_1 \mid \|\boldsymbol{x}\|_2 = 1\} = \|\hat{\boldsymbol{A}}\|_{2,1} \\
&= \max\{\|\hat{\boldsymbol{A}}^T(\hat{\boldsymbol{u}})\|_2 \mid \hat{\boldsymbol{u}} \in \{-1, 1\}^k\} \\
&= \max\{\|\boldsymbol{A}^T(\boldsymbol{u})\|_2 \mid \boldsymbol{u} \in \{-1, 1\}^m\},
\end{aligned}$$

and this finally gives the result.

(v) Note that, in this case, we wish to maximise the function $\boldsymbol{x} \mapsto \|\boldsymbol{A}(\boldsymbol{x})\|_2^2$ subject to the constraint that $\|\boldsymbol{x}\|_2^2 = 1$. In this case, the function we are maximising and the function defining the constraint are infinitely differentiable. Therefore, we can use the Lagrange Multiplier Theorem to determine the character of the maxima. Thus we define

$$f(\boldsymbol{x}) = \|\boldsymbol{A}(\boldsymbol{x})\|_2^2, \quad g(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2 - 1.$$

As $\boldsymbol{D}g$ has rank 1 at points satisfying the constraint, if a point $\boldsymbol{x}_0 \in \mathbb{R}^n$ solves the constrained maximisation problem, then there exists $\lambda \in \mathbb{R}$ such that

$$\boldsymbol{D}(f - \lambda g)(\boldsymbol{x}_0) = 0.$$

Since $f(\boldsymbol{x}) = \langle \boldsymbol{A}^T \circ \boldsymbol{A}(\boldsymbol{x}), \boldsymbol{x} \rangle$, we compute

$$\boldsymbol{D}f(\boldsymbol{x}) \cdot \boldsymbol{v} = 2\langle \boldsymbol{A}^T \circ \boldsymbol{A}(\boldsymbol{x}), \boldsymbol{v} \rangle.$$

As above, $\boldsymbol{D}g(\boldsymbol{x}) \cdot \boldsymbol{v} = 2\langle \boldsymbol{x}, \boldsymbol{v} \rangle$. Thus $\boldsymbol{D}(f - \lambda g)(\boldsymbol{x}_0) = 0$ implies that

$$\boldsymbol{A}^T \circ \boldsymbol{A}(\boldsymbol{x}_0) = \lambda \boldsymbol{x}_0.$$

Thus it must be the case that λ is an eigenvalue for $\boldsymbol{A}^T \circ \boldsymbol{A}$ with eigenvector $\boldsymbol{x}_0$. Since $\boldsymbol{A}^T \circ \boldsymbol{A}$ is symmetric and positive-semidefinite, all eigenvalues are real and nonnegative. Thus there exist $\lambda_1, \dots, \lambda_n \in \mathbb{R}_{\geq 0}$ and vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ such that

$$\lambda_1 \leq \cdots \leq \lambda_n,$$

such that $\boldsymbol{A}^T \circ \boldsymbol{A}(\boldsymbol{x}_j) = \lambda_j \boldsymbol{x}_j$, $j \in \{1, \dots, n\}$, and such that a solution to the problem of maximising f with the constraint $g^{-1}(0)$ is obtained by evaluating f at one of the points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$. Thus the problem can be solved by evaluating f at this finite collection of points, and determining at which of these f has its largest value. A computation gives $f(\boldsymbol{x}_j) = \lambda_j$, and this part of the result follows.

(vi) First of all, we note that this part of the theorem certainly holds when $\boldsymbol{A} = \boldsymbol{0}$. Thus we shall freely assume that $\boldsymbol{A}$ is nonzero when convenient. We maximise the function $\boldsymbol{x} \mapsto \|\boldsymbol{A}(\boldsymbol{x})\|_\infty$ subject to the constraint that $\|\boldsymbol{x}\|_2^2 = 1$. We shall again use the Lagrange Multiplier Theorem, defining

$$f(\boldsymbol{x}) = \|\boldsymbol{A}(\boldsymbol{x})\|_\infty, \quad g(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2 - 1.$$

Note that $\boldsymbol{A}$ is not differentiable on $\mathbb{R}^n$, so we first restrict to a subset where f is differentiable. Let us define

$$\begin{aligned} S_{\boldsymbol{A}} \colon \mathbb{R}^n &\to \boldsymbol{2}^{\{1,\dots,m\}} \\ \boldsymbol{x} &\mapsto \{a \in \{1, \dots, m\} \mid \langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{x} \rangle = \|\boldsymbol{A}(\boldsymbol{x})\|_\infty\}. \end{aligned}$$

Then denote

$$B_{\boldsymbol{A}} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \operatorname{card}(S_{\boldsymbol{A}}(\boldsymbol{x})) > 1\}.$$

We easily see that f is differentiable at points that are not in the set $B_{\boldsymbol{A}}$.

Let us first suppose that $\boldsymbol{x}_0 \in \mathbb{R}^n \setminus B_{\boldsymbol{A}}$ is a maximum of f subject to the constraint that $g(\boldsymbol{x}) = 0$. Then there exists a unique $a_0 \in \{1, \dots, m\}$ such that $f(\boldsymbol{x}_0) = \langle \boldsymbol{r}(\boldsymbol{A}, a_0), \boldsymbol{x}_0 \rangle$. Since we are assuming that $\boldsymbol{A}$ is nonzero, it must be that $\boldsymbol{r}(\boldsymbol{A}, a_0)$ is nonzero. Moreover, there exists a neighbourhood U of $\boldsymbol{x}_0$ such that

$$\operatorname{sign}(\langle \boldsymbol{r}(\boldsymbol{A}, a_0), \boldsymbol{x} \rangle) = \operatorname{sign}(\langle \boldsymbol{r}(\boldsymbol{A}, a_0), \boldsymbol{x}_0 \rangle)$$

and $f(\boldsymbol{x}) = \langle \boldsymbol{r}(\boldsymbol{A}, a_0), \boldsymbol{x} \rangle$ for each $\boldsymbol{x} \in U$. Abbreviating

$$u_{\boldsymbol{A}, a_0}(\boldsymbol{x}) = \operatorname{sign}(\langle \boldsymbol{r}(\boldsymbol{A}, a_0), \boldsymbol{x} \rangle),$$

we have

$$f(\boldsymbol{x}) = u_{\boldsymbol{A}, j}(\boldsymbol{x}_0) \langle \boldsymbol{r}(\boldsymbol{A}, a_0), \boldsymbol{x} \rangle$$

for every $\boldsymbol{x} \in U$. Note that, as in the proofs of parts (iv) and (v) above, $\boldsymbol{D}g(\boldsymbol{x})$ has rank 1 for $\boldsymbol{x} \neq 0$. Therefore there exists $\lambda \in \mathbb{R}$ such that

$$\boldsymbol{D}(f - \lambda g)(\boldsymbol{x}_0) = \boldsymbol{0}.$$

We compute

$$\boldsymbol{D}(f - \lambda g)(\boldsymbol{x}_0) \cdot \boldsymbol{v} = u_{\boldsymbol{A}, j}(\boldsymbol{x}_0) \langle \boldsymbol{r}(\boldsymbol{A}, a_0), \boldsymbol{v} \rangle - 2\lambda \langle \boldsymbol{x}_0, \boldsymbol{v} \rangle$$

for every $\boldsymbol{v} \in \mathbb{R}^n$. Thus we must have

$$2\lambda \boldsymbol{x}_0 = u_{\boldsymbol{A}, a_0}(\boldsymbol{x}_0) \boldsymbol{r}(\boldsymbol{A}, a_0).$$

This implies that $\boldsymbol{x}_0$ and $\boldsymbol{r}(\boldsymbol{A}, a_0)$ are collinear and that

$$|\lambda| = \frac{1}{2} \|\boldsymbol{r}(\boldsymbol{A}, a_0)\|_2$$

since $\|\boldsymbol{x}_0\|_2 = 1$. Therefore,

$$f(\boldsymbol{x}_0) = u_{\boldsymbol{A},a_0}(\boldsymbol{x}_0)\langle \boldsymbol{r}(\boldsymbol{A},a_0), \tfrac{1}{2\lambda}u_{\boldsymbol{A},a_0}(\boldsymbol{x}_0)\boldsymbol{r}(\boldsymbol{A},a_0)\rangle = 2\lambda.$$

Since $|\lambda| = \frac{1}{2}\|\boldsymbol{r}(\boldsymbol{A},a_0)\|_2$ it follows that

$$f(\boldsymbol{x}_0) = \|\boldsymbol{r}(\boldsymbol{A},a_0)\|_2.$$

This completes the proof, but for the fact that maxima of f may occur at points in $B_{\boldsymbol{A}}$. Thus let $\boldsymbol{x}_0 \in B_{\boldsymbol{A}}$ be such that $\|\boldsymbol{x}_0\|_2 = 1$. For $a \in S_{\boldsymbol{A}}(\boldsymbol{x}_0)$ let us write

$$\boldsymbol{r}(\boldsymbol{A},a) = \rho_a \boldsymbol{x}_0 + \boldsymbol{y}_a,$$

where $\langle \boldsymbol{x}_0, \boldsymbol{y}_a\rangle = 0$. Therefore, $\langle \boldsymbol{r}(\boldsymbol{A},a), \boldsymbol{x}_0\rangle = \rho_a$. We claim that if there exists $a_0 \in S_{\boldsymbol{A}}(\boldsymbol{x}_0)$ for which $\boldsymbol{y}_{a_0} \neq \boldsymbol{0}$, then $\boldsymbol{x}_0$ cannot be a maximum of f subject to the constraint $g^{-1}(0)$. Indeed, if $\boldsymbol{y}_{a_0} \neq 0$ then define

$$\boldsymbol{x}_\epsilon = \frac{\boldsymbol{x}_0 + \epsilon \boldsymbol{y}_{a_0}}{\sqrt{1 + \epsilon^2\|\boldsymbol{y}_{a_0}\|_2^2}}.$$

As in the proof of part (iv) above, one shows that $\boldsymbol{x}_\epsilon$ satisfies the constraint for every $\epsilon \in \mathbb{R}$. Also as in the proof of part (iv), we have

$$\boldsymbol{x}_\epsilon = \boldsymbol{x}_0 + \epsilon \boldsymbol{y}_0 + O(\epsilon^2).$$

Thus, for ϵ sufficiently small,

$$|\langle \boldsymbol{r}(\boldsymbol{A},a_0), \boldsymbol{x}_\epsilon\rangle| = |\langle \boldsymbol{r}(\boldsymbol{A},a_0), \boldsymbol{x}_0\rangle| + \epsilon C_{a_0} + O(\epsilon^2)$$

where C_{a_0} is nonzero. Therefore, there exists $\epsilon_0 \in \mathbb{R}_{>0}$ such that

$$|\langle \boldsymbol{r}(\boldsymbol{A},a_0), \boldsymbol{x}_\epsilon\rangle| > |\langle \boldsymbol{r}(\boldsymbol{A},a_0), \boldsymbol{x}_0\rangle|$$

either for all $\epsilon \in [-\epsilon_0, 0)$ or for all $\epsilon \in (0, \epsilon_0]$. In either case, $\boldsymbol{x}_0$ cannot be a maximum for f subject to the constraint $g^{-1}(0)$.

Finally, suppose that $\boldsymbol{x}_0 \in B_{\boldsymbol{A}}$ is a maximum for f subject to the constraint $g^{-1}(0)$. Then, as we saw in the preceding paragraph, for each $a \in S_{\boldsymbol{A}}(\boldsymbol{x}_0)$, we must have

$$\boldsymbol{r}(\boldsymbol{A},a) = \langle \boldsymbol{r}(\boldsymbol{A},a), \boldsymbol{x}_0\rangle \boldsymbol{x}_0.$$

It follows that $\|\boldsymbol{r}(\boldsymbol{A},a)\|_2^2 = \langle \boldsymbol{r}(\boldsymbol{A},a), \boldsymbol{x}_0\rangle^2$. Moreover, by definition of $S_{\boldsymbol{A}}(\boldsymbol{x}_0)$ and since we are supposing that $\boldsymbol{x}_0$ is a maximum for f subject to the constraint $g^{-1}(0)$, we have

$$\|\boldsymbol{r}(\boldsymbol{A},a)\|_2 = \|\boldsymbol{A}\|_{2,\infty}. \tag{3}$$

Now, if $a \in \{1,\dots,m\}$, we claim that

$$\|\boldsymbol{r}(\boldsymbol{A},a)\|_2 \leq \|\boldsymbol{A}\|_{2,\infty}. \tag{4}$$

Indeed suppose that $a \in \{1,\dots,m\}$ satisfies

$$\|\boldsymbol{r}(\boldsymbol{A},a)\|_2 > \|\boldsymbol{A}\|_{2,\infty}.$$

Define $\boldsymbol{x} = \frac{\boldsymbol{r}(\boldsymbol{A},a)}{\|\boldsymbol{r}(\boldsymbol{A},a)\|_2}$ so that $\boldsymbol{x}$ satisfies the constraint $g(\boldsymbol{x}) = 0$. Moreover,

$$f(\boldsymbol{x}) \geq \langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{x}\rangle = \|\boldsymbol{r}(\boldsymbol{A}, a)\|_2 > \|\boldsymbol{A}\|_{2,\infty},$$

contradicting the assumption that $\boldsymbol{x}_0$ is a maximum for f. Thus, given that (3) holds for every $a \in S_{\boldsymbol{A}}(\boldsymbol{x}_0)$ and (4) holds for every $a \in \{1, \dots, m\}$, we have

$$\|\boldsymbol{A}\|_{2,\infty} = \max\{\|\boldsymbol{r}(\boldsymbol{A}, a)\|_2 \mid a \in \{1, \dots, m\}\},$$

as desired.

For the last three parts of the theorem, the following result is useful.

Lemma: *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ and let $|||\cdot|||_\infty$ be the norm induced on $\mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$ by the norm $\|\cdot\|_\infty$ on $\mathbb{R}^n$ and the norm $\|\cdot\|$ on $\mathbb{R}^m$. Then*

$$|||\boldsymbol{A}|||_\infty = \max\{\|\boldsymbol{A}(\boldsymbol{u})\| \mid \boldsymbol{u} \in \{-1, 1\}^n\}.$$

Proof: Note that the set

$$\{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_\infty \leq 1\}$$

is a convex polytope. Therefore, this set is the convex hull of the vertices $\{-1, 1\}^n$; see [Webster 1994, Theorem 2.6.16]. Thus, if $\|\boldsymbol{x}\|_\infty = 1$ we can write

$$\boldsymbol{x} = \sum_{\boldsymbol{u} \in \{-1,1\}^n} \lambda_{\boldsymbol{u}} \boldsymbol{u}$$

where $\lambda_{\boldsymbol{u}} \in [0, 1]$ for each $\boldsymbol{u} \in \{-1, 1\}^n$ and

$$\sum_{\boldsymbol{u} \in \{-1,1\}^n} \lambda_{\boldsymbol{u}} = 1.$$

Therefore,

$$\begin{aligned}
\|\boldsymbol{A}(\boldsymbol{x})\| &= \Big\| \sum_{\boldsymbol{u} \in \{-1,1\}^n} \lambda_{\boldsymbol{u}} \boldsymbol{A}(\boldsymbol{u}) \Big\| \leq \sum_{\boldsymbol{u} \in \{-1,1\}^n} \lambda_{\boldsymbol{u}} \|\boldsymbol{A}(\boldsymbol{u})\| \\
&\leq \Big(\sum_{\boldsymbol{u} \in \{-1,1\}^n} \lambda_{\boldsymbol{u}} \Big) \max\{\|\boldsymbol{A}(\boldsymbol{u})\| \mid \boldsymbol{u} \in \{-1, 1\}^n\} \\
&= \max\{\|\boldsymbol{A}(\boldsymbol{u})\| \mid \boldsymbol{u} \in \{-1, 1\}^n\}.
\end{aligned}$$

Therefore,

$$\sup\{\|\boldsymbol{A}(\boldsymbol{x})\| \mid \|\boldsymbol{x}\|_\infty = 1\} \leq \max\{\|\boldsymbol{A}(\boldsymbol{u})\| \mid \boldsymbol{u} \in \{-1, 1\}^n\} \leq \sup\{\|\boldsymbol{A}(\boldsymbol{x})\| \mid \|\boldsymbol{x}\|_\infty = 1\},$$

the last inequality holding since if $\boldsymbol{u} \in \{-1, 1\}^n$ then $\|\boldsymbol{u}\|_\infty = 1$. The result follows since the previous inequalities must be equalities. ▼

(vii) This follows immediately from the preceding lemma.

(viii) This too follows immediately from the preceding lemma.

(ix) Note that for $\boldsymbol{u} \in \{-1,1\}^n$ we have

$$|\langle \boldsymbol{r}(\boldsymbol{A},a),\boldsymbol{u}\rangle| = \Big|\sum_{j=1}^n A_{aj}u_j\Big| \le \sum_{j=1}^n |A_{aj}| = \|\boldsymbol{r}(\boldsymbol{A},a)\|_1.$$

Therefore, using the previous lemma,

$$\begin{aligned}\|\boldsymbol{A}\|_{\infty,\infty} &= \max\{\|\boldsymbol{A}(\boldsymbol{u})\|_\infty \mid \boldsymbol{u} \in \{-1,1\}^n\}\\ &= \max\{\max\{|\langle \boldsymbol{r}(\boldsymbol{A},a),\boldsymbol{u}\rangle| \mid a \in \{1,\dots,m\}\} \mid \boldsymbol{u} \in \{-1,1\}^n\}\\ &\le \max\{\|\boldsymbol{r}(\boldsymbol{A},a)\|_1 \mid a \in \{1,\dots,m\}\}.\end{aligned}$$

To establish the other inequality, for $a \in \{1,\dots,m\}$ define $\boldsymbol{u}_a \in \{-1,1\}^n$ by

$$u_{a,j} = \begin{cases} 1, & A_{aj} \ge 0,\\ -1, & A_{aj} < 0\end{cases}$$

and note that a direct computation gives the ath component of $\boldsymbol{A}(\boldsymbol{u}_a)$ as $\|\boldsymbol{r}(\boldsymbol{A},a)\|_1$. Therefore,

$$\begin{aligned}\max\{\|\boldsymbol{r}(\boldsymbol{A},a)\|_1 \mid a \in \{1,\dots,m\}\} &= \max\{|\boldsymbol{A}(\boldsymbol{u}_a)_a| \mid a \in \{1,\dots,m\}\}\\ &\le \max\{\|\boldsymbol{A}(\boldsymbol{u})\|_\infty \mid \boldsymbol{u} \in \{-1,1\}^n\} = \|\boldsymbol{A}\|_{\infty,\infty},\end{aligned}$$

giving this part of the theorem. ■

3. Complexity of induced norm computations

Let us consider a comparison of the nine induced matrix norms in terms of the computational effort required. One would like to know how many operations are required to compute any of the norms. We shall do this making the following assumptions on our computational model.

> Floating point operations are carried out to an accuracy of $\epsilon = 2^{-N}$ for some fixed $N \in \mathbb{Z}_{>0}$. By $M(N)$ we denote the number of operations required to multiply integers j_1 and j_2 satisfying $0 \le j_1, j_2 \le 2^N$. We assume that addition and multiplication of floating point numbers can be performed with a relative error of $O(2^{-N})$ using $O(M(N))$ operations.

With this assumption, we can deduce the computational complexity of the basic operations we will need.

1. Computing a square root takes $O(M(N))$ operations; see [Brent 1976].
2. Computing the absolute value of a number is 1 operation (a bit flip).
3. Comparing two numbers takes $O(N)$ operations.
4. Finding the maximum number in a list of k numbers takes $O(kN)$ operations; see [Blum, Floyd, Pratt, Rivest, and Tarjan 1973].

5. If $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$ and $\boldsymbol{B} \in \mathrm{L}(\mathbb{R}^m; \mathbb{R}^p)$ then the matrix multiplication $\boldsymbol{BA}$ takes $O(mnpM(n))$ operations. Faster matrix multiplication algorithms are possible than the direct one whose complexity we describe here, [e.g., Coppersmith and Winograd 1990], but we are mainly interested in the fact that matrix multiplication has polynomial complexity in the size of the matrices.
6. Computation of the QR-decomposition of an $k \times k$ matrix $\boldsymbol{A}$ has computational complexity $O(k^3 M(N))$. Note that the QR-decomposition can be used to determine the Gram–Schmidt orthogonalisation of a finite number of vectors. We refer to [Golub and Van Loan 1996, §5.2] for details.
7. Let us describe deterministic bounds for the operations needed to compute the eigenvalues and eigenvectors of a $k \times k$ matrix $\boldsymbol{A}$, following Pan and Chen [1999]. Let us fix some norm $|||\cdot|||$ on $\mathrm{L}(\mathbb{R}^k; \mathbb{R}^k)$. Given $\epsilon = 2^{-N}$ as above, let $\beta \in \mathbb{R}_{>0}$ be such that $2^{-\beta}|||\boldsymbol{A}||| \leq \epsilon$. Then Pan and Chen show that the eigenvalues and eigenvectors of $\boldsymbol{A}$ can be computed for $\boldsymbol{A}$ using an algorithm of complexity
$$O(k^3 M(N)) + O((k \log^2 k)(\log \beta + \log^2 k) M(N)). \tag{5}$$
There are stochastic, iterative, or gradient flow algorithms that will generically perform computations with fewer operations than predicted by this bound. However, the complexity of such algorithms is difficult to understand, or they require unbounded numbers of operations in the worst case. In any event, here we only care that the complexity of the eigenproblem is polynomial.
8. The previous two computational complexity results can be combined to show that finding the square root of a symmetric positive-definite matrix has computational complexity given by (5). This is no doubt known, but let us see how this works since it is simple. First compute the eigenvalues and eigenvectors of $\boldsymbol{A}$ using an algorithm with complexity given by (5). The eigenvectors can be made into an orthonormal basis of eigenvectors using the Gram–Schmidt procedure. This decomposition can be performed using an algorithm of complexity $O(n^3 M(N))$. Assembling the orthogonal eigenvectors into the columns of a matrix gives an orthogonal matrix $\boldsymbol{U} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ and a diagonal matrix $\boldsymbol{D} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ with positive diagonals such that $\boldsymbol{A} = \boldsymbol{UDU}^T$. Then the matrix $\boldsymbol{D}^{1/2}$ with diagonal entries equal to the square roots of the diagonal of $\boldsymbol{D}$ can be constructed with complexity $O(nM(n))$. Finally, $\boldsymbol{A}^{1/2} = \boldsymbol{UD}^{1/2}\boldsymbol{U}$ is computed using matrix multiplication with complexity of $O(n^3 M(n))$.

Using these known computational complexity results, it is relatively straightforward to assess the complexity of the computations of the various norms in Theorem 1. In Table 1 we display this data, recording only the dependency of the computations on the number of

Table 1: Complexity of computing the norms $\|\cdot\|_{p,q}$

p \ q	1	2	∞
1	$O(mn)$	$O(mn)$	$O(mn)$
2	$O(mn2^m)$	$O(n^3)$	$O(mn)$
∞	$O(mn2^n)$	$O(mn2^n)$	$O(mn)$

rows m and columns n of the matrix. Note that the cases of $(p,q) \in \{(2,1),(\infty,1),(\infty,2)\}$ are exceptional in that the required operations grow exponentially with the size of $\boldsymbol{A}$. One must exercise some care in drawing conclusions here. For example, as we show in the proof of Theorem 1,

$$\|\boldsymbol{A}\|_{\infty,\infty} = \max\{\|\boldsymbol{A}(\boldsymbol{u})\|_\infty \mid \boldsymbol{u} \in \{-1,1\}^n\}, \tag{6}$$

and this computation has complexity $O(mn2^n)$. However, it turns out that the norm can be determined with a formula that is actually less complex. Indeed, our proof of the formula for $\|\cdot\|_{\infty,\infty}$—which is not the usual proof—starts with the formula (6) and produces a result with complexity $O(mn)$ as stated in Table 1.

One is then led to ask, are there similar simplifications of the norms corresponding to the cases $(p,q) \in \{(2,1),(\infty,1),(\infty,2)\}$? Rohn [2000] shows that the computation of $\|\cdot\|_{\infty,1}$ is NP-hard. We shall show here, using his ideas, that the computation of the norms $\|\cdot\|_{2,1}$ and $\|\cdot\|_{\infty,2}$ are likewise difficult, perhaps impossible, to reduce to algorithms with polynomial complexity.

2 Theorem: *If there exists an algorithm to compute $\|\boldsymbol{A}\|_{2,1}$ or $\|\boldsymbol{A}\|_{\infty,2}$ whose computational complexity is polynomial in the number of rows and the number of columns of $\boldsymbol{A}$, then P=NP.*

Proof: First note that $\|\boldsymbol{A}\|_{2,1} = \|\boldsymbol{A}^T\|_{\infty,2}$, so it suffices to prove the theorem only for $(p,q) = (\infty,2)$.

Following Rohn [2000] we introduce the notion of an ***MC-matrix*** ("MC" stands for "max-cut" since these matrices are related to the "max-cut problem" in graph theory) as a symmetric matrix $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^n;\mathbb{R}^n)$ with the property that the diagonal elements are equal to n and the off-diagonal elements are either 0 or -1. [Rohn 1994] shows that MC-matrices are positive-definite. Paljak and Rohn [1993] also prove the following.

> *The following decision problem is NP-complete:*
> *Given an $n \times n$ MC-matrix $\boldsymbol{A}$ and $M \in \mathbb{Z}_{>0}$, is $\langle \boldsymbol{A}(\boldsymbol{u}), \boldsymbol{u}\rangle \geq M$ for some $\boldsymbol{u} \in \{-1,1\}^n$?*

We will use this fact crucially in our proof.

Let us call a symmetric matrix $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^n;\mathbb{R}^n)$ a ***$\sqrt{MC}$-matrix*** if $\boldsymbol{A} \circ \boldsymbol{A}$ is an MC-matrix. Note that the map $\boldsymbol{A} \mapsto \boldsymbol{A} \circ \boldsymbol{A}$ from the set of $\sqrt{MC}$-matrices to the set of MC-matrices is surjective since MC-matrices have symmetric positive-definite square roots by virtue of their being themselves symmetric and positive-definite.

Now suppose that there exists an algorithm for determining the $(\infty,2)$-norm of a matrix, the computational complexity of which is of polynomial order in the number of rows and columns of the matrix. Let $\boldsymbol{A}$ be an $n \times n$ MC-matrix and let $M \in \mathbb{Z}_{>0}$. As we pointed out prior to stating the theorem, one can determine the $\sqrt{MC}$-matrix $\boldsymbol{A}^{1/2}$ using an algorithm with computational complexity that is polynomial in n. Then, by assumption, we can compute

$$\begin{aligned}\|\boldsymbol{A}^{1/2}\|_{\infty,2}^2 &= \max\{\|\boldsymbol{A}^{1/2}(\boldsymbol{u})\|_2^2 \mid \boldsymbol{u} \in \{-1,1\}^n\}\\ &= \max\{\langle \boldsymbol{A}(\boldsymbol{u}), \boldsymbol{u}\rangle \mid \boldsymbol{u} \in \{-1,1\}^n\}\end{aligned}$$

in polynomial time. In particular, we can determine whether $\langle \boldsymbol{A}(\boldsymbol{u}), \boldsymbol{u}\rangle \geq M$ in polynomial time. As we stated above, this latter decision problem is NP-complete, and so we must have P=NP. ■

References

Blum, M., Floyd, R. W., Pratt, V., Rivest, R. L., and Tarjan, R. E. [1973] *Time bounds for selection*, Journal of Computer and System Sciences, **7**(4), pages 448–461, ISSN: 0022-0000, DOI: 10.1016/S0022-0000(73)80033-9.

Brent, R. P. [1976] *Fast multiple-precision evaluation of elementary functions*, Journal of the Association for Computing Machinery, **23**(2), pages 242–251, ISSN: 0004-5411, DOI: 10.1145/321941.321944.

Coppersmith, D. and Winograd, S. [1990] *Matrix multiplication via arithmetic progressions*, Journal of Symbolic Computation, **9**(3), pages 251–280, ISSN: 0747-7171, DOI: 10.1016/S0747-7171(08)80013-2.

Drakakis, K. and Pearlmutter, B. A. [2009] *On the calculation of the $\ell_2 \to \ell_1$ induced matrix norm*, International Journal of Algebra, **3**(5), pages 231–240, ISSN: 1312-8868, URL: http://www.m-hikari.com/ija/ija-password-2009/ija-password5-8-2009/drakakisIJA5-8-2009.pdf.

Edwards, C. H. [1973] *Advanced Calculus of Several Variables*, Harcourt Brace & Company: New York, NY, Reprint: [Edwards 1995].

— [1995] *Advanced Calculus of Several Variables*, Dover Publications, Inc.: New York, NY, ISBN: 978-0-486-68336-2, Original: [Edwards 1973].

Golub, G. H. and Van Loan, C. F. [1996] *Matrix Computations*, 3rd edition, Johns Hopkins Studies in the Mathematical Sciences, The Johns Hopkins University Press: Baltimore, MD, ISBN: 978-0-8018-5414-9.

Horn, R. A. and Johnson, C. R. [2013] *Matrix Analysis*, 2nd edition, Cambridge University Press: New York/Port Chester/Melbourne/Sydney, ISBN: 978-0-521-83940-2.

Paljak, S. and Rohn, J. [1993] *Checking robust nonsingularity is NP-hard*, Mathematics of Control, Signals, and Systems, **6**, pages 1–9, ISSN: 0932-4194, DOI: 10.1007/BF01213466.

Pan, V. Y. and Chen, Z. Q. [1999] *The complexity of the matrix eigenproblem*, in *Conference Record of 31st Annual ACM Symposium on Theory of Computing*, ACM Symposium on Theory of Computing, (Atlanta, GA, May 1999), Association for Computing Machinery, pages 507–516.

Rohn, J. [1994] *Checking positive-definiteness or stability of symmetric interval matrices is NP-hard*, Commentationes Mathematicae Universitatis Carolinae, **35**(4), pages 795–797, ISSN: 0010-2628, URL: http://hdl.handle.net/10338.dmlcz/118721.

— [2000] *Computing the norm $\|A\|_{\infty,1}$ is NP-hard*, Linear and Multilinear Algebra, **47**(3), ISSN: 0308-1087, DOI: 10.1080/03081080008818644.

Webster, R. J. [1994] *Convexity*, Oxford Science Publications, Oxford University Press: Oxford, ISBN: 978-0-19-853147-0.