Estimating the Condition Number of the Frechet Derivative of a Matrix Function

Higham, Nicholas J. and Relton, Samuel D.

2014

MIMS EPrint: **2013.84**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: http://eprints.maths.manchester.ac.uk/
And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

SIAM J. SCI. COMPUT.
Vol. 36, No. 6, pp. C617–C634

ESTIMATING THE CONDITION NUMBER OF THE FRÉCHET DERIVATIVE OF A MATRIX FUNCTION*

NICHOLAS J. HIGHAM† AND SAMUEL D. RELTON†

Abstract. The Fréchet derivative L_f of a matrix function $f : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ is used in a variety of applications and several algorithms are available for computing it. We define a condition number for the Fréchet derivative and derive upper and lower bounds for it that differ by at most a factor 2. For a wide class of functions we derive an algorithm for estimating the 1-norm condition number that requires $O(n^3)$ flops given $O(n^3)$ flops algorithms for evaluating f and L_f; in practice it produces estimates correct to within a factor $6n$. Numerical experiments show the new algorithm to be much more reliable than a previous heuristic estimate of conditioning.

Key words. matrix function, condition number, Fréchet derivative, Kronecker form, matrix exponential, matrix logarithm, matrix powers, matrix pth root, MATLAB, expm, logm, sqrtm

AMS subject classifications. 65F30, 65F60

DOI. 10.1137/130950082

1. Introduction.

Condition numbers are widely used in numerical analysis for measuring the sensitivity of the solution to a problem when the input data is subject to small perturbations such as rounding or measurement errors [14]. For matrix functions the condition number is intimately related to the Fréchet derivative. The Fréchet derivative of $f\colon \mathbb{C}^{n\times n} \mapsto \mathbb{C}^{n\times n}$ is a mapping $L_f(A,\cdot)\colon \mathbb{C}^{n\times n} \mapsto \mathbb{C}^{n\times n}$ that is linear in its second argument and for any $E \in \mathbb{C}^{n\times n}$ satisfies

$$f(A+E) - f(A) - L_f(A,E) = o(\|E\|). \tag{1.1}$$

The Fréchet derivative of a matrix function also arises as an object of interest in its own right in a variety of applications, of which some recent examples are the computation of correlated choice probabilities [1], computing linearized backward errors for matrix functions [6], analysis of complex networks [7], [8], Markov models of cancer [9], computing matrix geometric means [21], nonlinear optimization for model reduction [25], [26], and tensor-based morphometry [30]. Software for computing Fréchet derivatives of matrix functions is available in a variety of languages [16].

The aims of this work are to define the condition number of the Fréchet derivative, obtain bounds for it, and construct an efficient algorithm to estimate it. We first recall the definition of the condition number of a matrix function f. The absolute condition number of f is defined by

$$\operatorname{cond}_{\mathrm{abs}}(f,A) := \lim_{\epsilon\to 0} \sup_{\|E\|\le\epsilon} \frac{\|f(A+E)-f(A)\|}{\epsilon}$$

and is characterized as the norm of the Fréchet derivative [15, Thm. 3.1], [27]:

$$\operatorname{cond}_{\mathrm{abs}}(f,A) = \max_{\|E\|=1} \|L_f(A,E)\|. \tag{1.2}$$

*Submitted to the journal's Software and High-Performance Computing section December 20, 2013; accepted for publication (in revised form) September 2, 2014; published electronically November 25, 2014. This work was supported by European Research Council Advanced grant MATFUN (267526) and Engineering and Physical Sciences Research Council grant EP/I03112X/1.
http://www.siam.org/journals/sisc/36-6/95008.html

†School of Mathematics, The University of Manchester, Manchester, M13 9PL, UK (nick.higham@manchester.ac.uk, http://www.maths.manchester.ac.uk/~higham, samuel.relton@manchester.ac.uk, http://www.maths.manchester.ac.uk/~srelton).

In practice the relative condition number

$$\text{cond}_{\text{rel}}(f,A) := \lim_{\epsilon\to 0}\sup_{\|E\|\le\epsilon\|A\|}\frac{\|f(A+E)-f(A)\|}{\epsilon\|f(A)\|} = \max_{\|E\|=1}\frac{\|L_f(A,E)\|\|A\|}{\|f(A)\|} \tag{1.3}$$

is more relevant. The two condition numbers are closely related, since

$$\text{cond}_{\text{rel}}(f,A) = \text{cond}_{\text{abs}}(f,A)\frac{\|A\|}{\|f(A)\|}. \tag{1.4}$$

Associated with the Fréchet derivative is its Kronecker matrix form [15, eq. (3.17)], which is the unique matrix $K_f(A)\in\mathbb{C}^{n^2\times n^2}$ such that for any $E\in\mathbb{C}^{n\times n}$,

$$\text{vec}(L_f(A,E)) = K_f(A)\,\text{vec}(E) = (\text{vec}(E)^T\otimes I_{n^2})\,\text{vec}(K_f(A)), \tag{1.5}$$

where vec is the operator which stacks the columns of a matrix vertically from first to last and $\otimes$ is the Kronecker product. The second equality is obtained by using the formula

$$\text{vec}(YAX) = (X^T\otimes Y)\,\text{vec}(A)$$

in the special case, with $x\in\mathbb{C}^n$,

$$Ax = \text{vec}(Ax) = (x^T\otimes I_n)\,\text{vec}(A). \tag{1.6}$$

The principal use of the Kronecker form is that its norm, estimated via the 1-norm or 2-norm power methods, for example, gives an estimate of $\text{cond}_{\text{abs}}(f,A)$ [15, sect. 3.4], [22].

To investigate the condition number of the Fréchet derivative we will need higher order Fréchet derivatives of matrix functions, which were recently investigated by Higham and Relton [19]. We now summarize the key results that we will need from that work.

The second Fréchet derivative $L_f^{(2)}(A,E_1,E_2)\in\mathbb{C}^{n\times n}$ is linear in both E_1 and E_2 and satisfies

$$L_f(A+E_2,E_1) - L_f(A,E_1) - L_f^{(2)}(A,E_1,E_2) = o(\|E_2\|). \tag{1.7}$$

Higham and Relton show that a sufficient condition for the second Fréchet derivative $L_f^{(2)}(A,\cdot,\cdot)$ to exist is that f is $4p-1$ times continuously differentiable on an open set containing the eigenvalues of A [19, Thm. 3.5], where p is the size of the largest Jordan block of A. This condition is certainly satisfied if f is $4n-1$ times continuously differentiable on a suitable open set. In this work we assume without further comment that this latter condition holds: it clearly does for common matrix functions such as the exponential, logarithm, and matrix powers A^t with $t\in\mathbb{R}$ or indeed for any analytic function. Under this condition it follows that $L_f^{(2)}(A,E_1,E_2)$ is independent of the order of E_1 and E_2 (which is analogous to the equality of mixed second order partial derivatives for scalar functions) [19, sect. 2].

One method for computing Fréchet derivatives is to apply f to a $2n\times 2n$ matrix and read off the top right-hand block [15, eq. (3.16)], [24]:

$$f\left(\begin{bmatrix} A & E\\ 0 & A\end{bmatrix}\right) = \begin{bmatrix} f(A) & L_f(A,E)\\ 0 & f(A)\end{bmatrix}. \tag{1.8}$$

Higham and Relton [19, Thm. 3.5] show that the second Fréchet derivative can be calculated in a similar way, as the top right-hand block of a $4n \times 4n$ matrix:

$$L_f^{(2)}(A, E_1, E_2) = f\left(\begin{bmatrix} A & E_1 & E_2 & 0 \\ 0 & A & 0 & E_2 \\ 0 & 0 & A & E_1 \\ 0 & 0 & 0 & A \end{bmatrix}\right)(1\colon n, 3n+1\colon 4n). \tag{1.9}$$

There is also a second order Kronecker matrix form [19, sect. 4], analogous to (1.5) and denoted by $K_f^{(2)}(A) \in \mathbb{C}^{n^4 \times n^2}$, such that for any E_1 and E_2,

$$\operatorname{vec}(L_f^{(2)}(A, E_1, E_2)) = (\operatorname{vec}(E_1)^T \otimes I_{n^2}) K_f^{(2)}(A) \operatorname{vec}(E_2) \tag{1.10}$$

$$= (\operatorname{vec}(E_2)^T \otimes \operatorname{vec}(E_1)^T \otimes I_{n^2}) \operatorname{vec}(K_f^{(2)}(A)). \tag{1.11}$$

Note that $K_f^{(2)}(A)$ encodes information about $L_f^{(2)}(A)$—the Fréchet derivative of $L_f(A)$—in n^6 numbers. Our challenge is to estimate the condition number of L_f in just $O(n^3)$ flops.

This paper is organized as follows. In section 2 we define the absolute and relative condition numbers of a Fréchet derivative, relate the two, and bound them above and below in terms of the second Fréchet derivative and the condition number of f. The upper and lower bounds that we obtain differ by at most a factor 2. Section 3 relates the bounds to the Kronecker matrix $K_f^{(2)}(A)$ at the cost, for the 1-norm, of introducing a further factor n of uncertainty, and this leads to an $O(n^3)$ flops algorithm given in section 4 for estimating the 1-norm condition number of the Fréchet derivative. We test the accuracy and robustness of our algorithm via numerical experiments in section 5. Concluding remarks are given in section 6.

2. The condition number of the Fréchet derivative. We begin by proposing a natural definition for the absolute and relative condition numbers of a Fréchet derivative and showing that the two are closely related. We define the absolute condition number of a Fréchet derivative $L_f(A, E)$ by

$$\operatorname{cond}_{\mathrm{abs}}(L_f, A, E) = \lim_{\epsilon \to 0} \sup_{\substack{\|\Delta A\| \le \epsilon \\ \|\Delta E\| \le \epsilon}} \frac{\|L_f(A + \Delta A, E + \Delta E) - L_f(A, E)\|}{\epsilon}, \tag{2.1}$$

which measures the maximal effect that small perturbations in the data A and E can have on the Fréchet derivative. Similarly, we define the relative condition number by

$$\operatorname{cond}_{\mathrm{rel}}(L_f, A, E) = \lim_{\epsilon \to 0} \sup_{\substack{\|\Delta A\| \le \epsilon \|A\| \\ \|\Delta E\| \le \epsilon \|E\|}} \frac{\|L_f(A + \Delta A, E + \Delta E) - L_f(A, E)\|}{\epsilon \|L_f(A, E)\|}, \tag{2.2}$$

where the changes are now measured in a relative sense. By taking ΔA and ΔE sufficiently small we can rearrange (2.2) to obtain the approximate upper bound

$$\frac{\|L_f(A + \Delta A, E + \Delta E) - L_f(A, E)\|}{\|L_f(A, E)\|} \lesssim \max\left(\frac{\|\Delta A\|}{\|A\|}, \frac{\|\Delta E\|}{\|E\|}\right) \operatorname{cond}_{\mathrm{rel}}(L_f, A, E). \tag{2.3}$$

A useful property of the relative condition number is its lack of dependence on

the norm of E: for any positive scalar $s \in \mathbb{R}$,

$$\begin{aligned}\operatorname{cond}_{\mathrm{rel}}(L_f, A, sE) &= \lim_{\epsilon\to 0} \sup_{\substack{\|\Delta A\|\le\epsilon\|A\| \\ \|\Delta E\|\le s\epsilon\|E\|}} \frac{\|L_f(A+\Delta A, sE+\Delta E) - L_f(A, sE)\|}{\epsilon\|L_f(A, sE)\|} \\ &= \lim_{\epsilon\to 0} \sup_{\substack{\|\Delta A\|\le\epsilon\|A\| \\ \|\Delta E/s\|\le\epsilon\|E\|}} \frac{\|L_f(A+\Delta A, E+\Delta E/s) - L_f(A, E)\|}{\epsilon\|L_f(A, E)\|} \\ &= \operatorname{cond}_{\mathrm{rel}}(L_f, A, E). \qquad (2.4)\end{aligned}$$

Furthermore we can obtain a similar relationship to (1.4) relating the absolute and relative condition numbers. This is useful since it allows us to state results and algorithms using the absolute condition number before reinterpreting them in terms of the relative condition number.

LEMMA 2.1. *The absolute and relative condition numbers of the Fréchet derivative L_f are related by*

$$\operatorname{cond}_{\mathrm{rel}}(L_f, A, E) = \frac{\operatorname{cond}_{\mathrm{abs}}(L_f, A, sE)\|E\|}{\|L_f(A, E)\|}, \qquad s = \frac{\|A\|}{\|E\|}.$$

Proof. Using (2.4) and setting $s\|E\| = \|A\|$ and $\delta = \epsilon\|A\|$ we have

$$\begin{aligned}\operatorname{cond}_{\mathrm{rel}}(L_f, A, E) &= \operatorname{cond}_{\mathrm{rel}}(L_f, A, sE) \\ &= \lim_{\epsilon\to 0} \sup_{\substack{\|\Delta A\|\le\epsilon\|A\| \\ \|\Delta E\|\le\epsilon s\|E\|}} \frac{\|L_f(A+\Delta A, sE+\Delta E) - L_f(A, sE)\|}{\epsilon\|L_f(A, sE)\|} \\ &= \frac{\|A\|}{\|L_f(A, sE)\|} \lim_{\delta\to 0} \sup_{\substack{\|\Delta A\|\le\delta \\ \|\Delta E\|\le\delta}} \frac{\|L_f(A+\Delta A, sE+\Delta E) - L_f(A, sE)\|}{\delta} \\ &= \frac{\operatorname{cond}_{\mathrm{abs}}(L_f, A, sE)\|A\|}{\|L_f(A, sE)\|} = \frac{\operatorname{cond}_{\mathrm{abs}}(L_f, A, sE)\|E\|}{\|L_f(A, E)\|}. \qquad \square\end{aligned}$$

In order to bound the relative condition number we will derive computable bounds on the absolute condition number and use the relationship in Lemma 2.1 to translate them into bounds on the relative condition number. We first obtain lower bounds.

LEMMA 2.2. *The absolute condition number of the Fréchet derivative satisfies both of the lower bounds*

$$\begin{aligned}&\operatorname{cond}_{\mathrm{abs}}(L_f, A, E) \ge \operatorname{cond}_{\mathrm{abs}}(f, A), \\ &\operatorname{cond}_{\mathrm{abs}}(L_f, A, E) \ge \max_{\|\Delta A\|=1} \|L_f^{(2)}(A, E, \Delta A)\|.\end{aligned}$$

Proof. For the first bound we set $\Delta A = 0$ in (2.1) and use the linearity of the derivative:

$$\begin{aligned}\operatorname{cond}_{\mathrm{abs}}(L_f, A, E) &\ge \lim_{\epsilon\to 0} \sup_{\|\Delta E\|\le\epsilon} \frac{\|L_f(A, E+\Delta E) - L_f(A, E)\|}{\epsilon} \\ &= \lim_{\epsilon\to 0} \sup_{\|\Delta E\|\le\epsilon} \frac{\|L_f(A, \Delta E)\|}{\epsilon} \\ &= \operatorname{cond}_{\mathrm{abs}}(f, A).\end{aligned}$$

Similarly, for the second bound we set $\Delta E = 0$ and obtain, using (1.7),

$$\begin{aligned}\operatorname{cond}_{\mathrm{abs}}(L_f, A, E) &\geq \lim_{\epsilon\to 0} \sup_{\|\Delta A\|\leq\epsilon} \frac{\|L_f(A+\Delta A, E) - L_f(A,E)\|}{\epsilon} \\ &= \lim_{\epsilon\to 0} \sup_{\|\Delta A\|\leq\epsilon} \frac{\|L_f^{(2)}(A,E,\Delta A) + o(\|\Delta A\|)\|}{\epsilon} \\ &= \lim_{\epsilon\to 0} \sup_{\|\Delta A\|\leq\epsilon} \|L_f^{(2)}(A,E,\Delta A/\epsilon)\| \\ &= \max_{\|\Delta A\|=1} \|L_f^{(2)}(A,E,\Delta A)\|. \qquad \square \end{aligned} \tag{2.5}$$

Next, we derive an upper bound.

LEMMA 2.3. *The absolute condition number of the Fréchet derivative satisfies*

$$\operatorname{cond}_{\mathrm{abs}}(L_f, A, E) \leq \max_{\|\Delta A\|=1} \|L_f^{(2)}(A,E,\Delta A)\| + \operatorname{cond}_{\mathrm{abs}}(f, A).$$

Proof. Notice that by linearity of the second argument of L_f,

$$\begin{aligned}\operatorname{cond}_{\mathrm{abs}}(L_f, A, E) &= \lim_{\epsilon\to 0} \sup_{\substack{\|\Delta A\|\leq\epsilon \\ \|\Delta E\|\leq\epsilon}} \frac{\|L_f(A+\Delta A, E+\Delta E) - L_f(A,E)\|}{\epsilon} \\ &\leq \lim_{\epsilon\to 0} \sup_{\substack{\|\Delta A\|\leq\epsilon \\ \|\Delta E\|\leq\epsilon}} \left(\frac{\|L_f(A+\Delta A, E) - L_f(A,E)\|}{\epsilon} + \frac{\|L_f(A+\Delta A, \Delta E)\|}{\epsilon} \right) \\ &\leq \lim_{\epsilon\to 0} \sup_{\|\Delta A\|\leq\epsilon} \frac{\|L_f(A+\Delta A, E) - L_f(A,E)\|}{\epsilon} \\ &\qquad + \lim_{\epsilon\to 0} \sup_{\substack{\|\Delta A\|\leq\epsilon \\ \|\Delta E\|\leq\epsilon}} \|L_f(A+\Delta A, \Delta E/\epsilon)\|. \end{aligned} \tag{2.6}$$

The first term on the right-hand side of (2.6) is equal to $\max_{\|\Delta A\|=1} \|L_f^{(2)}(A,E,\Delta A)\|$ by (2.5). For the second half of the bound (2.6) we have, using (1.7) and the fact that $L_f^{(2)}(A, E_1, E_2)$ is linear in E_2,

$$\begin{aligned}\lim_{\epsilon\to 0} \sup_{\substack{\|\Delta A\|\leq\epsilon \\ \|\Delta E\|\leq\epsilon}} \|L_f(A+\Delta A, \Delta E/\epsilon)\| &= \lim_{\epsilon\to 0} \sup_{\substack{\|\Delta A\|\leq\epsilon \\ \|\Delta E\|\leq\epsilon}} \big\| L_f(A, \Delta E/\epsilon) + L_f^{(2)}(A, \Delta E/\epsilon, \Delta A) \\ &\qquad + o(\|\Delta A\|)\big\| \\ &= \lim_{\epsilon\to 0} \sup_{\substack{\|\Delta A\|\leq\epsilon \\ \|\Delta E\|\leq\epsilon}} \|L_f(A, \Delta E/\epsilon) + O(\epsilon)\| \\ &= \lim_{\epsilon\to 0} \sup_{\|\Delta E\|\leq\epsilon} \|L_f(A, \Delta E/\epsilon)\| = \operatorname{cond}_{\mathrm{abs}}(f, A). \end{aligned}$$

Combining the two halves of the bound gives the result. $\square$

We now give the corresponding bounds for the relative condition number.

THEOREM 2.4. *The relative condition number of the Fréchet derivative L_f satisfies* $\operatorname{cond}_{\mathrm{rel}}(L_f, A, E) \geq 1$ *and*

$$\max(\operatorname{cond}_{\mathrm{abs}}(f, A), sM)r \leq \operatorname{cond}_{\mathrm{rel}}(L_f, A, E) \leq (\operatorname{cond}_{\mathrm{abs}}(f, A) + sM)r,$$

where $s = \|A\|/\|E\|$, $r = \|E\|/\|L_f(A,E)\|$, *and* $M = \max_{\|\Delta A\|=1} \|L_f^{(2)}(A,E,\Delta A)\|$.

Proof. To show $\mathrm{cond}_{\mathrm{rel}}(L_f,A,E) \geq 1$ we can use Lemmas 2.1 and 2.2, along with the linearity of $L_f(A,E)$ in E, as follows:

$$\begin{aligned}
\mathrm{cond}_{\mathrm{rel}}(L_f,A,E) &= \frac{\mathrm{cond}_{\mathrm{abs}}(L_f,A,sE)\|E\|}{\|L_f(A,E)\|} \\
&\geq \frac{\mathrm{cond}_{\mathrm{abs}}(f,A)\|E\|}{\|L_f(A,E)\|} \\
&= \frac{\max_{\|Z\|=1}\|L_f(A,Z)\|\|E\|}{\|L_f(A,E)\|} \\
&= \frac{\max_{\|Z\|=1}\|L_f(A,Z)\|}{\|L_f(A,E/\|E\|)\|} \geq 1.
\end{aligned}$$

For the other inequalities apply Lemma 2.1 to Lemmas 2.2 and 2.3 similarly. ∎

Theorem 2.4 gives upper and lower bounds for $\mathrm{cond}_{\mathrm{rel}}(L_f,A,E)$ that differ by at most a factor 2. During our numerical experiments in section 5 we found that typically $\mathrm{cond}_{\mathrm{abs}}(f,A)$ and sM were of comparable size, though on occasion they differed by many orders of magnitude. Finding sufficient conditions for these two quantities to differ significantly remains an open question which will depend on the complex interaction between f, A, and E.

There are already efficient algorithms for estimating $\mathrm{cond}_{\mathrm{abs}}(f,A)$ based on matrix norm estimation [15, sect. 3.4] in conjunction with methods for evaluating the Fréchet derivative [2], [3], [4], [17]. The key question is therefore how to estimate the maximum of $\|L_f^{(2)}(A,E,\Delta A)\|$ over all ΔA with $\|\Delta A\| = 1$. This is the subject of the next section.

3. Maximizing the second Fréchet derivative. Our techniques for estimating the required maximum norm of the second Fréchet derivative are analogous to those for estimating $\mathrm{cond}_{\mathrm{abs}}(f,A)$, so we first recall the latter.

We begin by considering the Frobenius norm, because the condition number of a matrix function can be computed as [15, eq. (3.20)]

$$\mathrm{cond}_{\mathrm{abs}}(f,A) = \max_{\|E\|_F=1} \|L_f(X,E)\|_F = \|K_f(X)\|_2. \tag{3.1}$$

In practice we usually estimate $\mathrm{cond}_{\mathrm{abs}}(f,A)$ in the 1-norm by $\|K_f(A)\|_1$, which is justified by the inequalities [15, Lem. 3.18]

$$\frac{\mathrm{cond}_{\mathrm{abs}}(f,A)}{n} \leq \|K_f(A)\|_1 \leq n\,\mathrm{cond}_{\mathrm{abs}}(f,A). \tag{3.2}$$

To estimate $\|K_f(A)\|_1$ the block 1-norm power method of Higham and Tisseur [20] is used [15, Alg. 3.22]. This approach requires around $4t$ matrix–vector products in total (using both $K_f(A)$ and $K_f(A)^*$) and produces estimates rarely more than a factor 3 from the true norm. The parameter t is usually set to 2, but can be increased for greater accuracy at the cost of extra flops.

Using (1.10) we obtain a result similar to (3.1) for maximizing the norm of the

second Fréchet derivative:

$$\begin{aligned}\max_{\|\Delta A\|_F=1} \|L_f^{(2)}(A,E,\Delta A)\|_F &= \sup_{\|\operatorname{vec}(\Delta A)\|_2\le 1} \|\operatorname{vec}(L_f^{(2)}(A,E,\Delta A))\|_2 \\ &= \sup_{\|\operatorname{vec}(\Delta A)\|_2\le 1} \|\big(\operatorname{vec}(E)^T\otimes I_{n^2}\big)K_f^{(2)}(A)\operatorname{vec}(\Delta A)\|_2 \\ &= \|\big(\operatorname{vec}(E)^T\otimes I_{n^2}\big)K_f^{(2)}(A)\|_2. \qquad (3.3)\end{aligned}$$

The next result shows that using the 1-norm instead gives the same accuracy guarantees as (3.2).

THEOREM 3.1. *With* $M = \max_{\|\Delta A\|_1\le 1} \|L_f^{(2)}(A,E,\Delta A)\|_1$, *we have*

$$\frac{1}{n}M \le \|(\operatorname{vec}(E)^T\otimes I_{n^2})K_f^{(2)}(A)\|_1 \le nM.$$

Proof. Making use of (1.10), for the lower bound we obtain

$$\begin{aligned}\max_{\|\Delta A\|_1\le 1} \|L_f^{(2)}(A,E,\Delta A)\|_1 &\le \sup_{\|\Delta A\|_1\le 1} \|\operatorname{vec}(L_f^{(2)}(A,E,\Delta A))\|_1 \\ &= \sup_{\|\Delta A\|_1\le 1} \|(\operatorname{vec}(E)^T\otimes I_{n^2})K_f^{(2)}(A)\operatorname{vec}(\Delta A)\|_1 \\ &\le \sup_{\|\operatorname{vec}(\Delta A)\|_1\le n} \|(\operatorname{vec}(E)^T\otimes I_{n^2})K_f^{(2)}(A)\operatorname{vec}(\Delta A)\|_1 \\ &= n \sup_{\|\operatorname{vec}(\Delta A)\|_1\le 1} \|(\operatorname{vec}(E)^T\otimes I_{n^2})K_f^{(2)}(A)\operatorname{vec}(\Delta A)\|_1 \\ &= n\|(\operatorname{vec}(E)^T\otimes I_{n^2})K_f^{(2)}(A)\|_1.\end{aligned}$$

For the upper bound, using (1.10) again,

$$\begin{aligned}\max_{\|\Delta A\|_1\le 1} \|L_f^{(2)}(A,E,\Delta A)\|_1 &\ge \frac{1}{n}\sup_{\|\Delta A\|_1\le 1} \|\operatorname{vec}(L_f^{(2)}(A,E,\Delta A))\|_1 \\ &= \frac{1}{n}\sup_{\|\Delta A\|_1\le 1} \|(\operatorname{vec}(E)^T\otimes I_{n^2})K_f^{(2)}(A)\operatorname{vec}(\Delta A)\|_1 \\ &\ge \frac{1}{n}\sup_{\|\operatorname{vec}(\Delta A)\|_1\le 1} \|(\operatorname{vec}(E)^T\otimes I_{n^2})K_f^{(2)}(A)\operatorname{vec}(\Delta A)\|_1 \\ &= \frac{1}{n}\|(\operatorname{vec}(E)^T\otimes I_{n^2})K_f^{(2)}(A)\|_1. \qquad \square\end{aligned}$$

Explicitly computing matrix–vector products with $(\operatorname{vec}(E)^T\otimes I_{n^2})K_f^{(2)}(A)$ and its conjugate transpose is not feasible, as computing $K_f^{(2)}(A)$ costs $O(n^7)$ flops [19, Alg. 4.2]. Fortunately we can compute the matrix–vector products implicitly since, by (1.10),

$$(\operatorname{vec}(E)^T\otimes I_{n^2})K_f^{(2)}(A)\operatorname{vec}(V) = \operatorname{vec}(L_f^{(2)}(A,E,V)),$$

where the evaluation of the right-hand side costs only $O(n^3)$ flops using (1.9). This is analogous to the relation $K_f(A)\operatorname{vec}(V) = \operatorname{vec}(L_f(A,V))$ used in the estimation of $K_f(A)$ in the 1-norm [15, Alg 3.22].

Similarly, we would like to implicitly compute products with the conjugate transpose $\left[(\mathrm{vec}(E)^T \otimes I_{n^2})K_f^{(2)}(A)\right]^*$ so that the entire 1-norm estimation can be done in $O(n^3)$ flops. To do so we need the following result.

THEOREM 3.2. *Let f be analytic on an open subset $\mathcal{D}$ of $\mathbb{C}$ for which each connected component is closed under conjugation and let f satisfy $f(z) = \overline{f(\overline{z})}$ for all $z \in \mathcal{D}$. Then for all $k \le m$ and A with spectrum in $\mathcal{D}$,*

$$L_f^{(k)}(A, E_1, \ldots, E_k)^* = L_f^{(k)}(A^*, E_1^*, \ldots, E_k^*).$$

Proof. Our proof is by induction on k, where the base case $k = 1$ is established by Higham and Lin [17, Lem. 6.2]. Assume that the result holds for the kth Fréchet derivative, which exists under the given assumptions. Then, since the Fréchet derivative is equal to the Gâteaux derivative [19],

$$L_f^{(k+1)}(A, E_1, \ldots, E_{k+1})^* = \left.\frac{d}{dt}\right|_{t=0} L_f^{(k)}(A + tE_{k+1}, E_1, \ldots, E_k)^*.$$

Using the inductive hypothesis the right-hand side becomes

$$\left.\frac{d}{dt}\right|_{t=0} L_f^{(k)}(A^* + tE_{k+1}^*, E_1^*, \ldots, E_k^*) = L_f^{(k+1)}(A^*, E_1^*, \ldots, E_{k+1}^*). \qquad \square$$

The conditions of Theorem 3.2 are not very restrictive; they are satisfied by the exponential, the logarithm, real powers A^t ($t \in \mathbb{R}$), the matrix sign function, and trigonometric functions, for example. The condition $f(z) = \overline{f(\overline{z})}$ is, in fact, equivalent to $f(A)^* = f(A^*)$ for all A with spectrum in $\mathcal{D}$ [18, Thm. 3.2 and its proof]. Under the conditions of the theorem it can be shown that

$$K_f(A)^* = K_f(A^*), \tag{3.4}$$

which is implicit in [15, pp. 66–67] and [17], albeit not explicitly stated there (and this equality will be needed in the appendix). Matrix–vector products with $K_f(A)^*$ can therefore be computed efficiently since

$$K_f(A)^* \,\mathrm{vec}(V) = K_f(A^*)\,\mathrm{vec}(V) = \mathrm{vec}(L_f(A^*, V)) = \mathrm{vec}(L_f(A, V^*)^*), \tag{3.5}$$

using Theorem 3.2 for the last equality. The next result gives an analog of (3.4) for $\left[(\mathrm{vec}(E)^T \otimes I_{n^2})K_f^{(2)}(A)\right]^*$.

THEOREM 3.3. *Under the conditions of Theorem 3.2, for $A \in \mathbb{C}^{n\times n}$ with spectrum in $\mathcal{D}$,*

$$\left[(\mathrm{vec}(E)^T \otimes I_{n^2})K_f^{(2)}(A)\right]^* = (\mathrm{vec}(E^*)^T \otimes I_{n^2})K_f^{(2)}(A^*).$$

Proof. We will need to use the Kronecker product property

$$(A \otimes B)(C \otimes D) = AC \otimes BD. \tag{3.6}$$

We also need the commutation (or vec-permutation) matrix $C_n \in \mathbb{C}^{n^2\times n^2}$, which is a permutation matrix defined by the property that for $A \in \mathbb{C}^{n\times n}$, $\mathrm{vec}(A^T) = C_n\,\mathrm{vec}(A)$. It is symmetric and satisfies, for $A, B \in \mathbb{C}^{n\times n}$ and $x, y \in \mathbb{C}^n$ [10], [23, Thm. 3.1],

$$(A \otimes B)C_n = C_n(B \otimes A), \tag{3.7}$$

$$(x^T \otimes y^T)C_n = y^T \otimes x^T. \tag{3.8}$$

We will prove that the two matrices in the theorem statement are equal by showing that they take the same value when multiplied by the arbitrary vector $v = \operatorname{vec}(V)$, where $V \in \mathbb{C}^{n\times n}$. Multiplying both sides by v and taking vec of the right-hand side we find that we need to show

$$\left[(\operatorname{vec}(E)^T \otimes I_{n^2})K_f^{(2)}(A)\right]^* v = (v^T \otimes \operatorname{vec}(E^*)^T \otimes I_{n^2}) \operatorname{vec}(K_f^{(2)}(A^*)).$$

Manipulating the left-hand side we have

$$\begin{aligned}
&\left[(\operatorname{vec}(E)^T \otimes I_{n^2})K_f^{(2)}(A)\right]^* v = K_f^{(2)}(A)^* \left(\operatorname{vec}(\overline{E}) \otimes I_{n^2}\right) v \\
&\quad = K_f^{(2)}(A)^* \left(\operatorname{vec}(\overline{E}) \otimes v\right) \quad \text{using } v = 1 \otimes v \text{ and (3.6)} \\
&\quad = \left[\left(\operatorname{vec}(\overline{E}) \otimes v\right)^T \otimes I_{n^2}\right] \operatorname{vec}(K_f^{(2)}(A)^*) \quad \text{by (1.6)} \\
&\quad = \left[((C_n \otimes I_{n^2})(\operatorname{vec}(E^*) \otimes v))^T \otimes I_{n^2}\right] \operatorname{vec}(K_f^{(2)}(A)^*) \\
&\quad = \left[\left((\operatorname{vec}(E^*)^T \otimes v^T)(C_n \otimes I_{n^2})\right) \otimes I_{n^2}\right] \operatorname{vec}(K_f^{(2)}(A)^*) \quad \text{using } C_n = C_n^T \\
&\quad = (\operatorname{vec}(E^*)^T \otimes v^T \otimes I_{n^2})(C_n \otimes I_{n^4}) \operatorname{vec}(K_f^{(2)}(A)^*) \quad \text{by (3.6) and } I_{n^2} \otimes I_{n^2} = I_{n^4} \\
&\quad = (v^T \otimes \operatorname{vec}(E^*)^T \otimes I_{n^2})(C_{n^2} \otimes I_{n^2})(C_n \otimes I_{n^4}) \operatorname{vec}(K_f^{(2)}(A)^*),
\end{aligned}$$

using (3.8) for the last equality. Therefore it remains to show that

$$(C_{n^2} \otimes I_{n^2})(C_n \otimes I_{n^4}) \operatorname{vec}(K_f^{(2)}(A)^*) = \operatorname{vec}(K_f^{(2)}(A^*)),$$

the proof of which can be found in the appendix. ∎

Theorem 3.3 shows that we can compute matrix–vector products with the conjugate transpose as

$$\begin{aligned}
\left[(\operatorname{vec}(E)^T \otimes I_{n^2})K_f^{(2)}(A)\right]^* \operatorname{vec}(V) &= (\operatorname{vec}(E^*)^T \otimes I_{n^2})K_f^{(2)}(A^*) \operatorname{vec}(V) \\
&= \operatorname{vec}(L_f^{(2)}(A^*, E^*, V)) \quad \text{by (1.10)} \\
&= \operatorname{vec}(L_f^{(2)}(A, E, V^*)^*), \qquad (3.9)
\end{aligned}$$

where the final equality is from Theorem 3.2. Therefore the block 1-norm estimator can be used to estimate efficiently $\|(\operatorname{vec}(E)^T \otimes I_{n^2})K_f^{(2)}(A)\|_1$ in Theorem 3.1.

4. An algorithm for estimating the relative condition number. We are now ready to state our complete algorithm for estimating the relative condition number of a Fréchet derivative in the 1-norm.

In the following algorithm we use the unvec operator, which for a vector $v \in \mathbb{C}^{n^2}$ returns the unique matrix in $\mathbb{C}^{n\times n}$ such that $\operatorname{vec}(\operatorname{unvec}(v)) = v$.

ALGORITHM 4.1. *Given $A \in \mathbb{C}^{n\times n}$, $E \in \mathbb{C}^{n\times n}$, and f satisfying the conditions of Theorem* 3.2 *this algorithm produces an estimate γ of the relative condition number* $\operatorname{cond}_{\mathrm{rel}}(L_f, A, E)$. *It uses the block* 1*-norm estimation algorithm of* [20] *with $t = 2$, which we denote by normest* (*an implementation is* [12, `funm_condest1`]).

1 Compute $f(A)$ and $L_f(A, E)$ via specialized algorithms such as those in [2], [4], or [17] if possible. Alternatively, compute $L_f(A, E)$ by finite differences, the complex step method [3], or (1.8).

2 Compute an estimate c of $\mathrm{cond}_{\mathrm{rel}}(f, A)$ using [15, Alg. 3.22] and normest.
3 $c \leftarrow c\|f(A)\|_1/\|A\|_1$ % Now $c = \mathrm{cond}_{\mathrm{abs}}(f, A)$.
4 $s = \|A\|_1/\|E\|_1$
5 Estimate $\mu = \| \left(\mathrm{vec}(E)^T \otimes I_{n^2}\right) K_f^{(2)}(A)\|_1$ using normest with lines 7–14.
6 $\gamma = (c + s\mu)\|E\|_1/\|L_f(A, E)\|_1$
7 ... To compute $(\mathrm{vec}(E)^T \otimes I_{n^2})K_f^{(2)}(A)v$ for a given v:
8 $V = \mathrm{unvec}(v)$
9 Calculate $W = L_f^{(2)}(A, E, V)$ using (1.9) for example.
10 Return $\mathrm{vec}(W)$ to the norm estimator.
11 ... To compute $\left[(\mathrm{vec}(E)^T \otimes I_{n^2})K_f^{(2)}(A)\right]^* v$ for a given v:
12 $V = \mathrm{unvec}(v)$
13 Calculate $W = L_f^{(2)}(A, E, V^*)$ using (1.9) for example.
14 Return $\mathrm{vec}(W^*)$ to the norm estimator.

Cost: Around 9 Fréchet derivative evaluations for $L_f(A, E)$ and $\mathrm{cond}_{\mathrm{rel}}(f, A)$, plus about 8 second Fréchet derivative evaluations. The cost depends on which particular methods are chosen to compute the Fréchet derivatives required in lines 1 and 2 and $L_f^{(2)}(A, E, V)$, but the total cost is $O(n^3)$ flops.

The quality of the estimate returned by Algorithm 4.1 depends on the quality of the underlying bounds and the quality of the computed norm estimate. The estimate has a factor 2 uncertainty from Theorem 2.4 and another factor n uncertainty from (3.2) and Theorem 3.1. The norm estimates are usually correct to within a factor 3, so overall we can expect the estimate from Algorithm 4.1 to differ from $\mathrm{cond}_{\mathrm{rel}}(f, A)$ by at most a factor $6n$.

Even though the Fréchet derivative $L_f^{(2)}(A, E_1, E_2)$ is linear in E_1 and E_2, the scaling of E_1 and E_2 may affect the accuracy of the computation. Heuristically we might expect that scaling E_1 and E_2 so that $\|A\|_1 \approx \|E_1\|_1 \approx \|E_2\|_1$ would give good accuracy. When implementing Algorithm 4.1 we scale E_1 and E_2 in this way before taking the derivatives and rescaling the result.

5. Numerical experiments. Our experiments are all performed in MATLAB R2013a. We examine the performance of Algorithm 4.1 for the matrix logarithm and matrix powers A^t with $t \in \mathbb{R}$ using the Fréchet derivative evaluation algorithms from [4] and [17], respectively. Throughout this section $u = 2^{-53}$ denotes the unit roundoff. Since the Fréchet derivative algorithms in question have been shown to perform in a forward stable manner in [4] and [17] (assessed therein using the `Kronecker` condition number estimator that we will show tends to underestimate the true condition number) we expect their relative errors to be bounded by the condition number times the unit roundoff.

We will compare Algorithm 4.1, denoted in this section by `condest_FD`, with three other methods in terms of the accuracy and reliability of using the estimated value of $\mathrm{cond}_{\mathrm{rel}}(L_f, A, E)u$ as a bound on the relative error

$$\frac{\|\widehat{L}_f(A, E) - L_f(A, E)\|_1}{\|L_f(A, E)\|_1},$$

where $\widehat{L}_f(A, E)$ is the computed Fréchet derivative. Unfortunately, we cannot directly assess the quality of our condition number estimates as we have no way to compute the exact condition number $\mathrm{cond}_{\mathrm{rel}}(L_f, A, E)$.

For our tests we need to choose the matrices A and E at which to evaluate the Fréchet derivative and its condition number. For A we use the same test matrices as in [4] and [17]. These (mostly 10×10) matrices are from the Matrix Computation Toolbox [11], the MATLAB `gallery` function, and the literature. Ideally we would choose the direction E as a direction that maximizes the relative error above; however, it is unclear how to do so without resorting to expensive optimization procedures. Instead we choose the direction E to be a matrix with normal $(0, 1)$ distributed elements, but we give a specific example of a worst case direction for the matrix logarithm in section 5.3.

To compute an accurate value of $L_f(A, E)$, used solely to calculate the relative errors mentioned above, we evaluate (1.8) in 250 digit precision by performing the diagonalization $VDV^{-1} = \left[\begin{smallmatrix} X & E \\ 0 & X \end{smallmatrix}\right]$, applying f to the diagonal matrix D, and returning the $(1, 2)$ block. If the matrix $\left[\begin{smallmatrix} X & E \\ 0 & X \end{smallmatrix}\right]$ is not diagonalizable we add a random perturbation of norm 10^{-125} to make the eigenvalues distinct. This idea was introduced by Davis [5] and has been used in [4] and [17]. These high precision calculations are performed in the Symbolic Math Toolbox.

We compare our algorithm against three approximations. The first is

$$\mathrm{cond}_{\mathrm{rel}}(L_f, A, E) \approx \frac{\|L_f(A + \Delta A, E + \Delta E) - L_f(A, E)\|_1}{\epsilon \|L_f(A, E)\|_1},$$

where ΔA and ΔE are chosen to have normal $(0, 1)$ distributed elements and then are scaled so that $\|\Delta A\|_1/\|A\|_1 = \|\Delta E\|_1/\|E\|_1 = \epsilon = 10^{-8}$ (cf. (2.2)). We would expect this method to generally underestimate the condition number since ΔA and ΔE are unlikely to point in the directions of greatest sensitivity. This estimate will be referred to as the `random` method throughout this section. Since this method requires only two Fréchet derivative evaluations (as opposed to around 17 for Algorithm 4.1) one possible extension of this method would be to run it k times and take the mean as an estimate of the condition number. Further experiments, not reported here, took $k = 5$, 10, and 20 without seeing any significant change in the results.

Our next alternative approximation is

$$\mathrm{cond}_{\mathrm{rel}}(L_f, A, E) \approx \frac{\|K_f(A + \Delta A) - K_f(A)\|_1}{\epsilon \|K_f(A)\|_1},$$

where $K_f(A)$ is the Kronecker form of the Fréchet derivative in (1.5), and ΔA is generated with normal $(0, 1)$ distributed elements and then scaled so that $\|\Delta A\|_1/\|A\|_1 = \epsilon = 10^{-8}$. This heuristic approximation has been used in [2], [4], but has two drawbacks. First, the dependence on E is ignored, which (see Lemmas 2.2 and 2.3) essentially corresponds to neglecting an additive $\mathrm{cond}_{\mathrm{abs}}(f, A)$ term and so could lead to underestimating the condition number. Second, the random direction ΔA will generally not point in the direction in which $K_f(A)$ is most sensitive, again leading to underestimation. We refer to this as the `Kronecker` method in our experiments. This method costs $O(n^5)$ flops and is therefore the most expensive. We might also try running this method k times and taking the mean of the results, in an attempt to better estimate the condition number. Further experiments averaging $k = 5$, 10, and 20 runs of this algorithm made no significant difference to the results.

The final approximation method for comparison is a modification of Algorithm 4.1 that estimates the second Fréchet derivative by the finite difference approximation $L_f^{(2)}(A, E, V) \approx t^{-1}\big(L_f(A + tV, E) - L_f(A, E)\big)$ for a small t instead of using (1.9). This is done by invoking `funm_condest1` from the Matrix Function Toolbox [11] on the

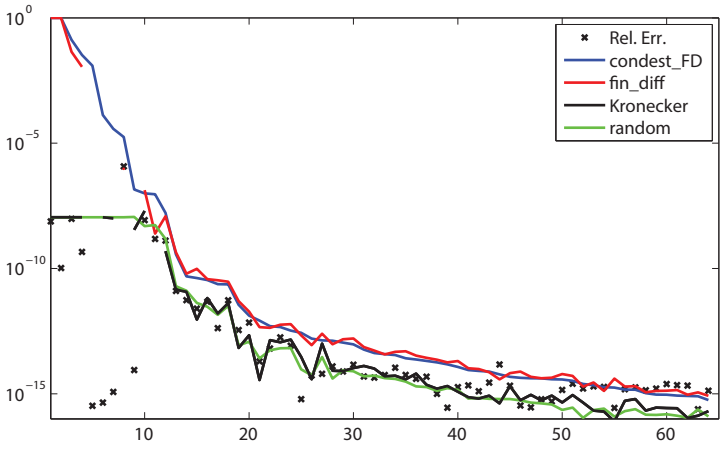

FIG. 1. *Relative errors of computed* $L_{\log}(A,E)$ *and estimates of* $\mathrm{cond}_{\mathrm{rel}}(L_{\log},A,E)u$ *for* 66 *test matrices sorted by decreasing value of* `condest_FD`.

function $g(A) = L_f(A,E)$ with the option to use finite differences selected, with the default value $t = 10^{-8}$. We will refer to this method as `fin_diff` in our experiments. This method has essentially identical cost to Algorithm 4.1, the only difference being the computation of the second Fréchet derivatives.

5.1. Condition number of Fréchet derivative of matrix logarithm. In our first experiment we compute the Fréchet derivative of the logarithm of 66 test matrices using the algorithm of Al-Mohy, Higham, and Relton [4]. Figure 1 shows the normwise relative errors and the estimates of $\mathrm{cond}_{\mathrm{rel}}(L_{\log},A,E)u$.

We see that `fin_diff` and `condest_FD` give similar output in most cases, as do `Kronecker` and `random`, though neither of these latter two seems able to yield values higher than 10^{-8} (the length of the finite difference step used in the algorithm). All four methods agree on which problems are well conditioned. On the right-hand side of the figure we see that some relative errors are slightly above the estimates. However all are within a factor 2.7 of the estimate from `condest_FD`, which is much less than the factor $6n$ we can expect in the worst case, as explained at the end of section 4.

For the ill conditioned problems both `Kronecker` and `fin_diff` fail to return condition number estimates for some of the test matrices, as indicated by the broken lines at the left end of Figure 1. This is due to a perturbed matrix $A+V$ having negative eigenvalues during the computation of the Fréchet derivatives using finite differences, which raises an error since the principal matrix logarithm and its Fréchet derivative are not defined for such matrices. In principle this same problem could happen when using the `random` method. Since `condest_FD` computes bounds on the second Fréchet derivative without perturbing A it does not encounter this problem. In section 5.3 we analyze the second test matrix in more detail and find that, despite the error bound being pessimistic, the condition number truly is as large as estimated by `fin_diff` and `condest_FD`.

5.2. Condition number of Fréchet derivative of matrix power. Our second experiment compares the algorithms on the function A^t with $t = 1/15$ over 60 test matrices from the previous set, where the Fréchet derivative is computed using the algorithm of Higham and Lin [17]. Figure 2 shows the normwise relative errors and

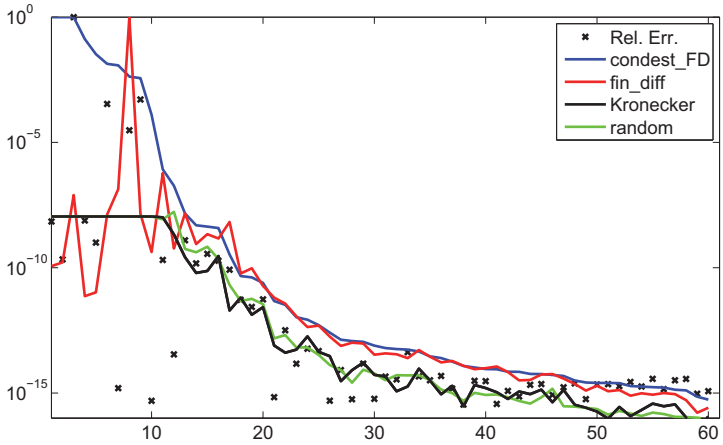

FIG. 2. *Relative errors of computed* $L_{x^t}(A,E)$ *and estimates of* $\mathrm{cond}_{\mathrm{rel}}(L_{x^t},A,E)u$, *with* $t = 1/15$, *for* 60 *test matrices sorted by decreasing value of* `condest_FD`.

the estimated quantities $\mathrm{cond}(L_{x^t},A,E)u$, sorted by decreasing `condest_FD`. Again we see that the condition number estimates from `Kronecker` and `random` are bounded above by about 10^{-8}, though the actual relative errors are sometimes much higher.

The methods return similar condition number estimates for the well conditioned problems but give very different results on the ill conditioned problems in the first 10 test cases. In particular `fin_diff`, `Kronecker`, and `random` do not provide reliable error bounds for the badly conditioned cases, their bounds being several orders of magnitude lower than the observed relative errors for test matrices 6 and 9. There is also some significant overestimation by `fin_diff` on test matrix 8. In contrast, `condest_FD` provides reliable error bounds for all the ill conditioned problems.

Similar experiments with the matrix exponential, not reported here, show analogous results: both `condest_FD` and `fin_diff` give good bounds on the relative errors while `Kronecker` and `random` generally underestimate them. The only difference is that `fin_diff` also gives good bounds for the ill-conditioned problems, instead of failing or giving spurious results as above.

5.3. An ill conditioned Fréchet derivative. In this section we give a more detailed analysis of the Fréchet derivative of the logarithm on test problem 2 of Figure 1. The matrices A and E are

$$A = \begin{bmatrix} e^{(\pi-10^{-7}i)} & 1000 \\ 0 & e^{(\pi+10^{-7}i)} \end{bmatrix}, \quad E = \begin{bmatrix} 0.3 & 0.012 \\ -0.76 & -0.49 \end{bmatrix}.$$

This example is particularly interesting because the condition number estimated by Algorithm 4.1 is large, $\mathrm{cond}_{\mathrm{rel}}(L_{\log},A,E) \approx 1.5 \times 10^{20}$, but we observed a relative error of around 10^{-10} when computing the Fréchet derivative in our experiments. We will show that a tiny perturbation to A that greatly changes $L_{\log}(A,E)$ exists.

What we need to do is find a matrix V with $\|V\|_1 = 1$ such that $\|L^{(2)}_{\log}(A,E,V)\|_1$ is large, since by Theorem 2.4 this will imply that $\mathrm{cond}_{\mathrm{rel}}(L_{\log},A,E)$ is large. Such a V can be obtained as output from the 1-norm estimator. However, we will obtain it from first principles by applying direct search optimization [13], with the code `mdsmax` from [11] that implements the algorithm from [28], [29]. Direct search yields

the putative optimal point

$$V = \begin{bmatrix} 0.1535 + 0.1535i & 0.1535 + 0.1535i \\ 0.1535 + 0.7677i & 0.1535 + 0.1535i \end{bmatrix},$$

shown to four significant figures, for which $\|L_{\log}^{(2)}(A, E, V)\|_1 = 1.4 \times 10^{44}$. Calculating the Fréchet derivatives $L_{\log}(A, E)$ and $L_{\log}(A + uV, E)$ in 250 digit arithmetic—using the procedure outlined at the beginning of this section—leads to a relative difference of

$$\frac{\|L_{\log}(A + uV, E) - L_{\log}(A, E)\|_1}{\|L_{\log}(A, E)\|_1} = 1.0318,$$

showing that the Fréchet derivative evaluation is extremely sensitive to perturbations in the direction V. We were fortunate not to experience this sensitivity during the evaluation of $L_{\log}(A, E)$. This computation confirms that, as `condest_FD` suggests, a relative perturbation of order u to A can produce a change of order 1 in the Fréchet derivative. But as we saw in the experiments, ill conditioning is not identified consistently by the approximations from `fin_diff`, `Kronecker`, or `random`.

6. Conclusion. We have defined, for the first time, the condition number of the Fréchet derivative of a matrix function and derived an algorithm for estimating it (Algorithm 4.1) that applies to a wide class of functions that includes the exponential, the logarithm, and real matrix powers. In practice, the algorithm produces estimates within a factor $6n$ of the true 1-norm condition number at a cost of $O(n^3)$ flops, given $O(n^3)$ flops algorithms for computing the function and its Fréchet derivative. The norms being estimated by the algorithm involve $n^4 \times n^2$ matrices, so structure is being exploited. An interesting open question is whether the highly structured nature of the second Fréchet derivative and its Kronecker form can be exploited to gain further theoretical insight into the conditioning of the Fréchet derivative.

The new algorithm is particularly useful for testing the forward stability of algorithms for computing Fréchet derivatives, and for this purpose our experiments show it to be much more reliable than a heuristic estimate used previously.

Appendix. Continued proof of Theorem 3.3. This section completes the proof of Theorem 3.3. We need to show that

$$(C_{n^2} \otimes I_{n^2})(C_n \otimes I_{n^4}) \operatorname{vec}(K_f^{(2)}(A)^*) = \operatorname{vec}(K_f^{(2)}(A^*)).$$

We will begin by showing that $\operatorname{vec}(K_f^{(2)}(A)^*) = \operatorname{vec}(K_f^{(2)}(A^*)C_n)$ which (after some manipulation) reduces the problem to one showing that

$$(C_{n^2} \otimes I_{n^2}) \operatorname{vec}(K_f^{(2)}(A^*)) = \operatorname{vec}(K_f^{(2)}(A^*)). \tag{A.1}$$

Before proceeding we recall that C_n is a permutation matrix corresponding to some permutation σ on the integers from 1 to n^2. This permutation can be defined by the property that when $\operatorname{vec}(E_i) = e_i$ is the ith standard basis vector then

$$E_{\sigma(i)} = E_i^T, \tag{A.2}$$

which follows from the observation that $C_n \operatorname{vec}(E_i) = C_n e_i = e_{\sigma(i)} = \operatorname{vec}(E_{\sigma(i)})$ along with $C_n \operatorname{vec}(E_i) = \operatorname{vec}(E_i^T)$.

Expanding [19, Alg. 4.2] for the case $k = 2$, (or from [19, eq. (4.4)]), we see that $K_f^{(2)}(X) \in \mathbb{C}^{n^4 \times n^2}$ is made from $n^2 \times 1$ blocks

$$\left[K_f^{(2)}(X)\right]_{ij} = \operatorname{vec}(L_f^{(2)}(X, E_j, E_i)), \qquad i, j = 1 : n^2,$$

so that applying C_n to the right of $K_f^{(2)}(A^*)$ permutes its columns and

$$\begin{aligned}\left[K_f^{(2)}(A^*)C_n\right]_{ij} &= \operatorname{vec}(L_f^{(2)}(A^*, E_{\sigma(j)}, E_i)) \\ &= \operatorname{vec}(L_f^{(2)}(A^*, E_j^T, E_i)) \\ &= \operatorname{vec}(L_f^{(2)}(A, E_j, E_i^T)^*),\end{aligned}$$

because $L_f^{(2)}(A^*, F, G) = L_f^{(2)}(A, F^*, G^*)^*$ by Theorem 3.2. Similarly $K_f^{(2)}(A)^*$ is made from $1 \times n^2$ blocks

$$\left[K_f^{(2)}(A)^*\right]_{ij} = \operatorname{vec}(L_f^{(2)}(A, E_i, E_j))^*, \qquad i, j = 1 : n^2.$$

To continue, note that $K_f^{(2)}(A^*)C_n$ and $K_f^{(2)}(A)^*$ are of sizes $n^4 \times n^2$ and $n^2 \times n^4$, respectively, and so cannot be equal, though we need only prove that their vectorizations are equal. We need to show that each $n^2 \times n^2$ block column of $K_f^{(2)}(A)^*$ is equal to the "unvec" of the corresponding $n^4 \times 1$ column of $K_f^{(2)}(A^*)C_n$. That is, for $j = 1 : n^2$ we want to show that

(A.3)
$$\begin{bmatrix} \operatorname{vec}(L_f^{(2)}(A, E_1, E_j))^* \\ \vdots \\ \operatorname{vec}(L_f^{(2)}(A, E_{n^2}, E_j))^* \end{bmatrix} = \begin{bmatrix} \operatorname{vec}(L_f^{(2)}(A, E_j, E_1^T)^*) & \cdots & \operatorname{vec}(L_f^{(2)}(A, E_j, E_{n^2}^T)^*) \end{bmatrix}.$$

To do so, we will expand the rows and columns then show they are equal elementwise. Since [19]

$$L_f^{(2)}(A, E_k, E_j) = \left.\frac{d}{dt}\right|_{t=0} L_f(A(t), E_k), \qquad A(t) = A + tE_j,$$

the left-hand side of (A.3) can be written as

$$\begin{bmatrix} \operatorname{vec}(L_f^{(2)}(A, E_1, E_j))^* \\ \vdots \\ \operatorname{vec}(L_f^{(2)}(A, E_{n^2}, E_j))^* \end{bmatrix} = \left.\frac{d}{dt}\right|_{t=0} \begin{bmatrix} e_1^T \overline{\operatorname{vec}(L_f(A(t), E_1))} & \cdots & e_{n^2}^T \overline{\operatorname{vec}(L_f(A(t), E_1))} \\ \vdots & \ddots & \vdots \\ e_1^T \overline{\operatorname{vec}(L_f(A(t), E_{n^2}))} & \cdots & e_{n^2}^T \overline{\operatorname{vec}(L_f(A(t), E_{n^2}))} \end{bmatrix}.$$

Similarly, using (A.2) on the right-hand side of (A.3) we have

$$\operatorname{vec}(L_f^{(2)}(A, E_j, E_i^T)^*) = \left.\frac{d}{dt}\right|_{t=0} C_n \overline{\operatorname{vec}(L_f(A(t), E_{\sigma(i)}))},$$

and therefore the right-hand side of (A.3) can be written as

$$\begin{aligned}&\begin{bmatrix} \operatorname{vec}(L_f^{(2)}(A, E_j, E_1^T)^*) & \cdots & \operatorname{vec}(L_f^{(2)}(A, E_j, E_{n^2}^T)^*) \end{bmatrix} \\ &\quad = \left.\frac{d}{dt}\right|_{t=0} \begin{bmatrix} e_{\sigma(1)}^T \overline{\operatorname{vec}(L_f(A(t), E_{\sigma(1)}))} & \cdots & e_{\sigma(1)}^T \overline{\operatorname{vec}(L_f(A(t), E_{\sigma(n^2)}))} \\ \vdots & \ddots & \vdots \\ e_{\sigma(n^2)}^T \overline{\operatorname{vec}(L_f(A(t), E_{\sigma(1)}))} & \cdots & e_{\sigma(n^2)}^T \overline{\operatorname{vec}(L_f(A(t), E_{\sigma(n^2)}))} \end{bmatrix}.\end{aligned}$$

Suppressing the dependence on t, we need to prove that

$$e_j^T \operatorname{vec}(L_f(A, E_i)) = e_{\sigma(i)}^T \operatorname{vec}(L_f(A, E_{\sigma(j)})),$$

since these are the (i,j) elements of the left- and right-hand side of (A.3), respectively (with the complex conjugation removed from both sides). Beginning from the right-hand side we have

$$\begin{aligned}
e_{\sigma(i)}^T \operatorname{vec}(L_f(A, E_{\sigma(j)})) &= e_i^T C_n \operatorname{vec}(L_f(A, E_{\sigma(j)})) \\
&= e_i^T \overline{\operatorname{vec}(L_f(A^*, E_j))} \quad \text{by (A.2)} \\
&= e_i^T (e_j^T \otimes I_{n^2}) \overline{\operatorname{vec}(K_f(A^*))} \quad \text{by (1.5)} \\
&= e_i^T (e_j^T \otimes I_{n^2}) \overline{\operatorname{vec}(K_f(A)^*)} \quad \text{by (3.4)} \\
&= e_i^T (e_j^T \otimes I_{n^2}) C_n \operatorname{vec}(K_f(A)) \\
&= e_i^T (I_{n^2} \otimes e_j^T) \operatorname{vec}(K_f(A)) \quad \text{by (3.8)} \\
&= e_j^T (e_i^T \otimes I_{n^2}) \operatorname{vec}(K_f(A)) \\
&= e_j^T \operatorname{vec}(L_f(A, E_i)) \quad \text{by (1.5)},
\end{aligned}$$

as required, which completes the proof of

$$\operatorname{vec}(K_f^{(2)}(A)^*) = \operatorname{vec}(K_f^{(2)}(A^*) C_n).$$

To complete the result we need to prove (A.1). To make the notation slightly easier we will use $X = A^*$ from now on. By [23, Thm. 3.1 (i)] we can write

$$C_{n^2} = \sum_{j=1}^{n^2} e_j^T \otimes I_{n^2} \otimes e_j,$$

where $e_k \in \mathbb{C}^{n^2}$, and so the left-hand side of (A.1) becomes

$$\begin{aligned}
(C_{n^2} \otimes I_{n^2}) \operatorname{vec}(K_f^{(2)}(X)) &= \left(\sum_{j=1}^{n^2} e_j^T \otimes I_{n^2} \otimes e_j \otimes I_{n^2} \right) \operatorname{vec}(K_f^{(2)}(X)) \\
&= \sum_{j=1}^{n^2} \operatorname{vec}\left((I_{n^2} \otimes e_j \otimes I_{n^2}) K_f^{(2)}(X) e_j \right) \\
&= \sum_{j=1}^{n^2} \operatorname{vec}\left((I_{n^2} \otimes e_j \otimes I_{n^2}) \operatorname{vec}(K_f^{(1)}(X, E_j)) \right) \\
&= \sum_{j=1}^{n^2} \operatorname{vec}\left((e_j \otimes I_{n^2}) K_f^{(1)}(X, E_j) \right) \\
&= \sum_{j=1}^{n^2} \operatorname{vec}\left(e_j \otimes K_f^{(1)}(X, E_j) \right) \\
&= \operatorname{vec}\left(\begin{bmatrix} K_f^{(1)}(X, E_1) \\ \vdots \\ K_f^{(1)}(X, E_{n^2}) \end{bmatrix} \right),
\end{aligned}$$

where $K_f^{(1)}(X, E_i)$ is defined in [19, sect. 4]. To show that this is equal to $\mathrm{vec}(K_f^{(2)}(X))$ we can write the two vectors out elementwise. For $\mathrm{vec}(K_f^{(2)}(X))$ we know from [19, Alg. 4.2] that

$$\mathrm{vec}(K_f^{(2)}(X)) = \begin{bmatrix} \mathrm{vec}(L_f^{(2)}(X, E_1, E_1)) \\ \vdots \\ \mathrm{vec}(L_f^{(2)}(X, E_1, E_{n^2})) \\ \mathrm{vec}(L_f^{(2)}(X, E_2, E_1)) \\ \vdots \\ \mathrm{vec}(L_f^{(2)}(X, E_{n^2}, E_{n^2})) \end{bmatrix}, \tag{A.4}$$

whereas

$$\begin{aligned} \mathrm{vec}\left(\begin{bmatrix} K_f^{(1)}(X, E_1) \\ \vdots \\ K_f^{(1)}(X, E_{n^2}) \end{bmatrix}\right) &= \mathrm{vec}\left(\begin{bmatrix} K_f^{(1)}(X, E_1)e_1 \\ K_f^{(1)}(X, E_2)e_1 \\ \vdots \\ K_f^{(1)}(X, E_1)e_2 \\ \vdots \\ K_f^{(1)}(X, E_{n^2})e_{n^2} \end{bmatrix}\right) \\ &= \mathrm{vec}\left(\begin{bmatrix} \mathrm{vec}(L_f^{(2)}(X, E_1, E_1)) \\ \vdots \\ \mathrm{vec}(L_f^{(2)}(X, E_{n^2}, E_1)) \\ \mathrm{vec}(L_f^{(2)}(X, E_1, E_2)) \\ \vdots \\ \mathrm{vec}(L_f^{(2)}(X, E_{n^2}, E_{n^2})) \end{bmatrix}\right). \end{aligned}$$

This is equal to (A.4) since $L_f^{(2)}(X, F, G) = L_f^{(2)}(X, G, F)$, by the ordering independence noted in section 1.

REFERENCES

[1] S. D. Ahipasaoglu, X. Li, and K. Natarajan, *A Convex Optimization Approach for Computing Correlated Choice Probabilities With Many Alternatives*, Preprint 4034, Optimization Online, 2013.

[2] A. H. Al-Mohy and N. J. Higham, *A new scaling and squaring algorithm for the matrix exponential*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 970–989.

[3] A. H. Al-Mohy and N. J. Higham, *The complex step approximation to the Fréchet derivative of a matrix function*, Numer. Algorithms, 53 (2010), pp. 133–148.

[4] A. H. Al-Mohy, N. J. Higham, and S. D. Relton, *Computing the Fréchet derivative of the matrix logarithm and estimating the condition number*, SIAM J. Sci. Comput., 35 (2013), pp. C394–C410.

[5] E. B. Davies, *Approximate diagonalization*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 1051–1064.

[6] E. Deadman and N. J. Higham, *Testing Matrix Function Algorithms Using Identities*, MIMS EPrint 2014.13, Manchester Institute for Mathematical Sciences, The University of Manchester, Manchester, UK, 2014.

[7] E. Estrada and D. J. Higham, *Network properties revealed through matrix functions*, SIAM Rev., 52 (2010), pp. 696–714.
[8] E. Estrada, D. J. Higham, and N. Hatano, *Communicability betweenness in complex networks*, Phys. A, 388 (2009), pp. 764–774.
[9] B. García-Mora, C. Santamaría, G. Rubio, and J. L. Pontones, *Computing survival functions of the sum of two independent Markov processes: An application to bladder carcinoma treatment*, Int. J. Comput. Math., 91 (2014), pp. 209–220.
[10] H. V. Henderson and S. R. Searle, *The vec-permutation matrix, the vec operator and Kronecker products: A review*, Linear and Multilinear Algebra, 9 (1981), pp. 271–288.
[11] N. J. Higham, *The Matrix Computation Toolbox*, http://www.maths.manchester.ac.uk/∼higham/mctoolbox.
[12] N. J. Higham, *The Matrix Function Toolbox*, http://www.maths.manchester.ac.uk/∼higham/mftoolbox.
[13] N. J. Higham, *Optimization by direct search in matrix computations*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 317–333.
[14] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002.
[15] N. J. Higham, *Functions of Matrices: Theory and Computation*, SIAM, Philadelphia, 2008.
[16] N. J. Higham and E. Deadman, *A Catalogue of Software for Matrix Functions. Version* 1.0. MIMS EPrint 2014.8, Manchester Institute for Mathematical Sciences, The University of Manchester, Manchester, UK, 2014.
[17] N. J. Higham and L. Lin, *An improved Schur–Padé algorithm for fractional powers of a matrix and their Fréchet derivatives*, SIAM J. Matrix Anal. Appl., 34 (2013), 1341–1360.
[18] N. J. Higham, D. S. Mackey, N. Mackey, and F. Tisseur, *Functions preserving matrix groups and iterations for the matrix square root*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 849–877.
[19] N. J. Higham and S. D. Relton, *Higher order Fréchet derivatives of matrix functions and the level-*2 *condition number*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1019–1037.
[20] N. J. Higham and F. Tisseur, *A block algorithm for matrix* 1*-norm estimation, with an application to* 1*-norm pseudospectra*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1185–1201.
[21] B. Jeuris, R. Vandebril, and B. Vandereycken, *A survey and comparison of contemporary algorithms for computing the matrix geometric mean*, Electron. Trans. Numer. Anal., 39 (2012), pp. 379–402.
[22] C. S. Kenney and A. J. Laub, *Condition estimates for matrix functions*, SIAM J. Matrix Anal. Appl., 10 (1989), pp. 191–209.
[23] J. R. Magnus and H. Neudecker, *The commutation matrix: Some properties and applications*, Ann. Statist., 7 (1979), pp. 381–394.
[24] R. Mathias, *A chain rule for matrix functions and applications*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 610–620.
[25] D. Petersson, *A Nonlinear Optimization Approach to* $\mathcal{H}_2$*-Optimal Modeling and Control*, Dissertation No. 1528, Ph.D. thesis, Department of Electrical Engineering, Linköping University, Sweden, Linköping, Sweden, 2013.
[26] D. Petersson and J. Löfberg, *Model reduction using a frequency-limited* $\mathcal{H}_2$*-cost*, Technical report, Department of Electrical Engineering, Linköpings Universitet, Linköping, Sweden, 2012.
[27] J. R. Rice, *A theory of condition*, SIAM J. Numer. Anal., 3 (1966), pp. 287–310.
[28] V. Torczon, *Multi-Directional Search: A Direct Search Algorithm for Parallel Machines*, Ph.D. thesis, Rice University, Houston, TX, 1989.
[29] V. Torczon, *On the convergence of the multidirectional search algorithm*, SIAM J. Optim., 1 (1991), pp. 123–145.
[30] E. Zacur, M. Bossa, and S. Olmos, *Multivariate tensor-based morphometry with a right-invariant Riemannian distance on* $GL^+(n)$, J. Math. Imaging Vis., 50 (2014), pp. 18–31.