A NOTE ON OISHI'S LOWER BOUND FOR THE SMALLEST SINGULAR VALUE OF LINEARIZED GALERKIN EQUATIONS

SIEGFRIED M. RUMP * AND SHIN'ICHI OISHI †

Abstract. Recently Oishi published a paper allowing lower bounds for the minimum singular value of coefficient matrices of linearized Galerkin equations, which in turn arise in the computation of periodic solutions of nonlinear delay differential equations with some smooth nonlinearity. The coefficient matrix of linearized Galerkin equations may be large, so the computation of a valid lower bound of the smallest singular value may be costly. Oishi's method is based on the inverse of a small upper left principal submatrix, and subsequent computations use a Schur complement with small computational cost. In this note some assumptions are removed and the bounds improved. Furthermore a technique is derived to reduce the total computationally cost significantly allowing to treat infinite dimensional matrices.

Key words. Bound for the norm of the inverse of a matrix, minimum singular value, Galerkin's equation, nonlinear delay differential equation, Schur complement.

AMS subject classifications. 65F45

1. Main result.

Certain periodic solutions of nonlinear delay differential equations with some smooth nonlinearity can be calculated by some Galerkin equations. Oishi's paper [3] discusses how to obtain a lower bound for the smallest singular value of the coefficient matrix G of the linearized Galerkin equation.

Throughout this note $\|\cdot\|$ denotes the ℓ_2-norm, and we use the convention that matrices are supposed to be nonsingular when using their inverse. Dividing the matrix G into blocks

$$G := \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{1}$$

with a square upper left block A of small size, often $\|G^{-1}\|$ is in practical applications not too far from $\|A^{-1}\|$. In other words, the main information on $\|G^{-1}\|$ sits in A. Moreover, the matrix G shares some kind of diagonal dominance so that $\|D^{-1}\|$ can be expected to be not too far from $\|D_d^{-1}\|$ for the splitting $D = D_d + D_f$ into diagonal and off-diagonal part.

Based on that Oishi discusses in [3] how to obtain an upper bound on $\|G^{-1}\| = \sigma_{\min}(G)^{-1}$ by estimating the influence of the remaining blocks B, C, D. Oishi proves the following theorem.

THEOREM 1.1. *Let n be a positive integer, and m be a non-negative integer satisfying $m \le n$. Let $G \in M_n(\mathbb{R})$ be as in* (1) *with $A \in M_m(\mathbb{R})$, $B \in M_{m,n-m}(\mathbb{R})$, $C \in M_{n-m,m}(\mathbb{R})$ and $D \in M_{n-m}(\mathbb{R})$. Let D_d and D_f be the diagonal part and the off-diagonal part of D, respectively. If*

$$\|A^{-1}B\| < 1, \quad \|CA^{-1}\| < 1 \quad \textit{and} \quad \|D_d^{-1}(D_f - CA^{-1}B)\| < 1, \tag{2}$$

*Institute for Reliable Computing, Hamburg University of Technology, Am Schwarzenberg-Campus 3, Hamburg 21073, Germany, and Visiting Professor at Waseda University, Faculty of Science and Engineering, 3–4–1 Okubo, Shinjuku-ku, Tokyo 169–8555, Japan (`rump@tuhh.de`).

†Department of Applied Mathematics, Faculty of Science and Engineering, Waseda University, 3–4–1 Okubo, Shinjuku-ku, Tokyo 169–8555, Japan (`oishi@waseda.jp`).

then G is invertible and

$$\|G^{-1}\| \le \frac{\max\left\{\|A^{-1}\|, \dfrac{\|D_d^{-1}\|}{1-\|D_d^{-1}(D_f - CA^{-1}B)\|}\right\}}{(1-\|A^{-1}B\|)(1-\|CA^{-1}\|)} \,. \tag{3}$$

In a practical application the matrix G may be large. So the advantage of Oishi's bound is that with small computational effort a reasonable upper bound for $\|G^{-1}\|$, and thus a lower bound for $\sigma_{\min}(G)^{-1}$, is derived. In particular the clever use of Schur's complement requires only a small upper left part A of G to be inverted, whereas the inverse of the large lower right part D is estimated using a Neumann expansion.

For $p := \min(m, n)$, denote by $I_{m,n}$ the matrix with the $p \times p$ identity matrix in the upper left corner. If $m = n$, we briefly write I_m. Oishi's estimate is based on the factorization [2] of G using its Schur complement

$$G = \begin{pmatrix} I_m & 0 \\ CA^{-1} & I_{n-m} \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{pmatrix} \begin{pmatrix} I_m & A^{-1}B \\ 0 & I_{n-m} \end{pmatrix}.$$

It follows

$$\|G^{-1}\| \le \varphi(CA^{-1}) \max\left\{\|A^{-1}\|, \|(D - CA^{-1}B)^{-1}\|\right\} \varphi(A^{-1}B) \tag{4}$$

where for $N \in M_{m,n-m}(\mathbb{R})$

$$\varphi(N) := \left\| \begin{pmatrix} I_m & -N \\ 0 & I_{n-m} \end{pmatrix} \right\| = \left\| \begin{pmatrix} I_m & N \\ 0 & I_{n-m} \end{pmatrix}^{-1} \right\| = \left\| \begin{pmatrix} I_m & 0 \\ N^T & I_{n-m} \end{pmatrix}^{-1} \right\| = \left\| \begin{pmatrix} I_m & 0 \\ -N^T & I_{n-m} \end{pmatrix} \right\|.$$

To bound $\|(D - CA^{-1}B)^{-1}\|$ the standard estimate

$$\|(D - CA^{-1}B)^{-1}\| = \| \left(I + D_d^{-1}(D_f - CA^{-1}B)\right)^{-1} D_d^{-1}\| \le \frac{\|D_d^{-1}\|}{1 - \|D_d^{-1}(D_f - CA^{-1}B)\|}$$

based on a Neumann expansion is used. In practical applications as those in [3] this imposes only a small overestimation, and often the maximum in the middle of (4) is equal to $\|A^{-1}\|$.

Thus it remains to bound $\varphi(N)$ for $N \in \{CA^{-1}, A^{-1}B\}$ which is done in [3] by $\varphi(N) \le \dfrac{1}{1-\|N\|}$ requiring $\|N\| < 1$. The following lemma removes that restriction and gives an explicit formula for $\varphi(N)$.

Lemma 1.2. *Let $H \in M_n(\mathbb{R})$ have the block representation*

$$H = \begin{pmatrix} I_k & -N \\ 0 & I_{n-k} \end{pmatrix}$$

for $1 \le k < n$ and $N \in M_{k,n-k}(\mathbb{R})$ with $\|N\| \ne 0$. Abbreviate $\mu := \|N\|$ and define

$$\alpha := \left\{ \frac{1}{2} \left(1 + \frac{1}{\sqrt{1 + \frac{4}{\mu^2}}} \right) \right\}^{1/2} \in (0, 1).$$

Then

$$\psi(N) := \|H\| = \|H^{-1}\| = \sqrt{1 + 2\alpha\mu\sqrt{1-\alpha^2} + \alpha^2\mu^2} \,. \tag{5}$$

Proof. Let $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^{n-k}$. Then

$$\left\| H \begin{pmatrix} x \\ y \end{pmatrix} \right\|^2 = \left\| \begin{pmatrix} x + Ny \\ y \end{pmatrix} \right\|^2 \le (\|x\| + \|N\|\|y\|)^2 + \|y\|^2$$

with equality if, and only if, $\|Ny\| = \|N\|\|y\|$ and x is a positive multiple of Ny. Let $u \in \mathbb{R}^k$ and $v \in \mathbb{R}^{n-k}$ with $\|u\| = \|v\| = 1$ and $Nv = \|N\|u$ and define $y := \alpha v$ and $x := \sqrt{1-\alpha^2}\, u$. Then $\left\| \begin{pmatrix} x \\ y \end{pmatrix} \right\| = 1$ and

$$\nu(\alpha) := \left\| H \begin{pmatrix} x \\ y \end{pmatrix} \right\|^2 = \left(\sqrt{1-\alpha^2} + \alpha\mu \right)^2 + \alpha^2 = 1 + 2\alpha\mu\sqrt{1-\alpha^2} + \alpha^2\mu^2, \tag{6}$$

so that the maximum of $\nu(\alpha)$ over $\alpha \in [0,1]$ is equal to $\|H\|^2$. A computation yields

$$\frac{d\nu(\alpha)}{d\alpha} = \frac{2\mu}{\sqrt{1-\alpha^2}} \left(1 - 2\alpha^2 + \alpha\mu\sqrt{1-\alpha^2} \right).$$

Setting the derivative to zero and abbreviating $\beta := \alpha^2$ gives

$$\beta^2 - \beta + \frac{1}{4+\mu^2} = 0$$

with positive solution

$$\beta = \frac{1}{2} \left(1 + \frac{1}{\sqrt{1 + \frac{4}{\mu^2}}} \right) \in (0,1)$$

for nontrivial N. That proves the lemma. □

The lemma removes the first two assumptions in (2). In practical applications those are usually satisfied, so the advantage is that they do not have to be verified. That proves the following.

THEOREM 1.3. *Let n, m and $G \in M_n(\mathbb{R})$ with the splitting as in* (1) *be given. With the splitting $D = D_d + D_f$ into diagonal and off-diagonal part assume*

$$\|D_d^{-1}(D_f - CA^{-1}B)\| < 1. \tag{7}$$

Then G is invertible and

$$\|G^{-1}\| \le \max\left\{ \|A^{-1}\|, \frac{\|D_d^{-1}\|}{1 - \|D_d^{-1}(D_f - CA^{-1}B)\|} \right\} \psi(A^{-1}B)\, \psi(CA^{-1}) \tag{8}$$

using $\psi(\cdot)$ as in (5) *in Lemma 1.2.*

For a matrix G arising in linearized Galerkin equations the norm $\|(D - CA^{-1}B)^{-1}\|$ of the inverse of the Schur complement is often dominated by $\|A^{-1}\|$. That is because of the increasing diagonal elements D_d of the lower right part of G, implying that $\|D_d^{-1}\|$ is equal to the reciprocal of the left upper element of D_d and $\|D_d^{-1}(D_f - CA^{-1}B)^{-1}\|$ to become small. In that case the bound (8) reduces to $\|G^{-1}\| \le \|A^{-1}\|\, \psi(A^{-1}B)\, \psi(CA^{-1})$. It can be adapted to other norms following the lines of the proof of Lemma 1.2.

2. Computational improvement and an infinite dimensional example. For both the original bound (3) and the new bound (8) the main computational effort is to estimate $\|D_d^{-1}(D_f - CA^{-1}B)\|$ requiring $\mathcal{O}((n-m)^2 m)$ operations. It often suffices to use $\|N\| \le \sqrt{\|N\|_1 \|N\|_\infty} =: \pi(N)$ and the crude estimate

$$\|D_d^{-1}(D_f - CA^{-1}B)\| \le \pi(D_d^{-1}D_f) + \pi(D_d^{-1}C)\pi(A^{-1}B) =: \gamma. \tag{9}$$

That avoids the product of large matrices and to compute the ℓ_2-norm of large matrices, thus reducing the total computational cost to $\mathcal{O}((n-m)m^2)$ operations. For instance, in Example 3 in [4] with $n = 10,000$ and $k = 0.9$ the bound (9) is successful in the sense that $\max\left(\|A^{-1}\|, \frac{\|D_d^{-1}\|}{1-\gamma}\right) = \|A^{-1}\|$ for dimension $m \ge 13$ of the upper left block A. The additional computational cost for (9) is marginal, so in any case it is worth to try. If successful, then the computational cost for dimension $n = 10,000$ reduces from 8 minutes to less than 2 seconds on a standard laptop.

In order to compute rigorous bounds, an upper bound for the ℓ_2 norm of a matrix $C \in M_n(\mathbb{R})$ is necessary. An obvious possibility is to use $\|C\|_2 \le \sqrt{\|C\|_1 \|C\|_\infty}$. Denoting the spectral radius of a matrix by $\varrho(\cdot)$, better bounds are obtained by $\|C\|_2^2 \le \||C|\|_2^2 = \varrho(|C|^T|C|)$ and Collatz' bound [1]

$$\varrho(|C|^T|C|) \le \max_i \frac{(|C|^T(|C|x))_i}{x_i} \ ,$$

which is true for any positive vector x. Then a few power iterations on x yield an accurate upper bound for $\||C|\|_2$ in some $\mathcal{O}(n^2)$ operations. More sophisticated methods for bounding $\|C\|_2$ can be found in [6].

A practical consideration, in particular for so-called verification methods [4, 5], is that the true value $\mu = \|N\|$ is usually not available. However, $\nu(N)$ in (6) is increasing with μ and implies

$$\|N\| \le \mu \quad \Rightarrow \quad \|H\| = \|H^{-1}\| \le \sqrt{1 + 2\alpha\mu\sqrt{1-\alpha^2} + \alpha^2\mu^2} \ . \tag{10}$$

For example, $\pi(N)$ can be used for an upper bound μ of $\|N\|$. Applying this method to (8) to bound $\psi(N)$ for $N \in \{A^{-1}B, CA^{-1}\}$ using $\|N\| \le \pi(N)$ entirely avoids the computation of the ℓ_2-norm of large matrices.

To that end we show how to compute an upper bound of $\|G^{-1}\|$ for the third example in [4] for infinite dimension. The matrix G is parameterized by some $k \in (0,5,1)$ with elements $G_{ij} = k^{|i-j|}$ for $i \ne j$ and $(1,\ldots,n)$ on the diagonal. As before m denotes the size of the upper left block A.

We first show how to estimate γ in (9) for arbitrarily large n. Define

$$P = D_d^{-1}D_f \quad \Rightarrow \quad P_{ij} := \begin{cases} 0 & \text{if } i = j \\ k^{|i-j|}/(i+m) & \text{otherwise} \end{cases} \ .$$

Then for every fixed $s \in \{1,\ldots,n-m\}$

$$\sum_{j=1}^{n-m} P_{sj} = \sum_{j=1}^{s-1} k^{s-j}/(s+m) + \sum_{j=s+1}^{n-m} k^{j-s}/(s+m) < 2\sum_{j=1}^{\infty} k^j/(1+m) = \frac{2k}{(1-k)(m+1)}$$

and

$$\sum_{i=1}^{n-m} P_{is} = \sum_{i=1}^{s-1} k^{s-i}/(i+m) + \sum_{i=s+1}^{n-m} k^{i-s}/(i+m) < 2\sum_{i=1}^{\infty} k^i/(1+m) = \frac{2k}{(1-k)(m+1)}$$

proves

$$\|D_d^{-1}D_f\| \le \sqrt{\|D_d^{-1}D_f\|_1 \|D_d^{-1}D_f\|_\infty} < \frac{2k}{(1-k)(m+1)} \ . \tag{11}$$

That estimate is valid for any dimension $n > m$. For $Q = D_d^{-1}C$ we have $Q_{ij} := k^{|m+i-j|}/(m+i)$, so that for every fixed $s \in \{1, \ldots, n-m\}$

$$\sum_{j=1}^{m} Q_{sj} \leq \sum_{j=1}^{s-1} k^{m+1-j}/(m+1) < \frac{k}{(1-k)(m+1)} \ ,$$

and for every fixed $s \in \{1, \ldots, m\}$

$$\sum_{i=1}^{n-m} Q_{is} = \sum_{i=1}^{n-m} k^{m+i-s}/(m+i) < \frac{k}{(1-k)(m+1)} \ .$$

It follows

$$\|D_d^{-1}C\| < \frac{k}{(1-k)(m+1)}. \tag{12}$$

In order to estimate $\|A^{-1}B\|$ we split $B = [S \ \ T]$ into blocks $S \in M_{m,q}$ and $T \in M_{m,n-m-q}$ and $1 \leq q < n-m$. Then $B_{ij} := k^{m-i+j}$ implies $T_{ij} := k^{q+m-i+j}$ and

$$\|T\| \leq \sqrt{\left(\sum_{i=1}^{m} k^{q+m-i+1}\right)\left(\sum_{j=1}^{n-m-q} k^{q+m-1+j}\right)} < \frac{k^{q+1}}{1-k} \ ,$$

so that

$$\|A^{-1}B\| \leq \|A^{-1}S\| + \|A^{-1}T\| < \|A^{-1}S\| + \frac{k^{q+1}\|A^{-1}\|}{1-k} \ . \tag{13}$$

Summarizing and using (11), (12), and (13), (9) becomes

$$\|D_d^{-1}(D_f - CA^{-1}B)\| < \frac{2k}{(1-k)(m+1)} + \frac{k}{(1-k)(m+1)}\left(\|A^{-1}S\| + \frac{k^{q+1}\|A^{-1}\|}{1-k}\right) =: \delta \ .$$

Note that only "small" matrices are involved in the computation of δ. For given k, we choose m and q large enough to ensure $\max\left(\|A^{-1}\|, \frac{\|D_d^{-1}\|}{1-\delta}\right) = \|A^{-1}\|$. Then it is true for the infinite dimensional matrix G as well. The computational effort is $\mathcal{O}(\max(m^2q, m^3))$

Note that the bound is very crude with much room for improvement. Nevertheless, $k = 0.9$, $m = 25$ and $q = 35$ ensures $\max\left(\|A^{-1}\|, \frac{\|D_d^{-1}\|}{1-\delta}\right) = \|A^{-1}\|$ for infinite n with marginal computational effort.

In our example, symmetry implies $\|A^{-1}B\| = \|CA^{-1}\|$, so that inserting the above quantities proves the bounds displayed in Table 2.1 for arbitrarily large dimension n.

TABLE 2.1
Bounds of $\|G^{-1}\|$ infinite dimensional G with $k = 0.9$ and choices of m and q.

m / q	30/50	30/100	50/100	100/100	200/100
$\|G^{-1}\| \leq$	3.07	2.75	2.61	2.50	2.45

Numerical evidence suggests that the true norm does not exceed 2.39. The total computing time for infinite dimensional G with $m = 200$ and $q = 100$ is less than 0.09 seconds on a laptop.

3. Comparison of the original and new bound. In [3] the bounds on $\|G^{-1}\|$ are tested by means of three typical application examples. The first example is a tridiagonal matrix G with all elements equal to 2 on the first subdiagonal, all elements equal to 3 on the first superdiagonal, and $(1, \ldots, n)$ on the diagonal. Therefore, with increasing dimension $n \geq 21$, neither $\varphi(CA^{-1})$ nor $\varphi(A^{-1}B)$ change. Since the maximum in (4) is equal to $\|A^{-1}\|$, the original bound and the new bound using (4) and (5) do not change for any $n \geq 21$, i.e., they are valid for the infinite dimensional case. Computed bounds for different size m of the left upper block are displayed in Table 3.1. Computational evidence suggests that $\|G^{-1}\|$ does not exceed 3.12 for infinite dimension.

Table 3.1

Bounds of $\|G^{-1}\|$ infinite dimensional G in Example 1 [4] and choices of m.

m	20	50	100	200
original bound	3.84	3.36	3.25	3.19
new bound	3.45	3.25	3.18	3.15

In the second example in [3] the upper left 5×5 matrix in G is replaced by some random positive matrix. Since the lower $(n-5) \times (n-5)$ block in G is still tridiagonal, again the original and new bound do not change for any dimension $n \geq 21$. An example for the original, the new bound and for $\|G^{-1}\|$ is $1.909, 1.648, 1.449$.

The third example in [3] was already treated in the previous section. For k close to 1 the bounds (3) and (8) are getting weak. For different values of k and dimension n the results are displayed in Table 3.2. For the true value of $\|G^{-1}\|$ verified bounds were computed using verification methods, so that the displayed digits in the last column are correct.

Table 3.2

Bounds of $\|G^{-1}\|$ using dimension $m = 20$ of the upper left block A for different parameters k and dimensions n of G.

k	n	original bound	new bound	$\|G^{-1}\|$
0.9	50	3.7548	2.9231	2.3882
	100	3.7566	2.9237	2.3882
	1000	3.7566	2.9237	2.3882
0.95	50	11.5604	4.8093	2.9483
	100	12.1180	4.8644	2.9483
	1000	12.1214	4.8647	2.9483
0.96	50	32.2895	6.1486	3.1352
	100	39.9103	6.3328	3.1352
	1000	40.0604	6.3359	3.1352
0.97	50	∞	9.0043	3.3798
	100	∞	9.7451	3.3803
	1000	∞	9.7816	3.3803

We see not much dependency on the dimension of the original and the new bound, and practically no dependency of $\|G^{-1}\|$. With k getting closer to the upper limit 1 there is more improvement by the new bound, but naturally the bounds become also weaker compared to $\|G^{-1}\|$. For $k \geq 0.97$ the original bound fails because the assumptions $\|CA^{-1}\| < 1$ and $\|A^{-1}B\| < 1$ are not satisfied.

4. Acknowledgement. Our dearest thanks to the anonymous referee for many helpful and constructive remarks.

REFERENCES

[1] L. Collatz. Einschließungssatz für die charakteristischen Zahlen von Matrizen. *Math. Z.*, 48:221–226, 1942.

[2] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, second edition, 2017.

[3] S. Oishi. Lower bounds for the smallest singular values of generalized asymptotic diagonal dominant matrices. *Japan Journal of Industrial and Applied Mathematics*, 2023. https://doi.org/10.1007/s13160-023-00596-5.

[4] S. Oishi, K. Ichihara, M. Kashiwagi, K. Kimura, X. Liu, H. Masai, Y. Morikura, T. Ogita, K. Ozaki, S.M. Rump, K. Sekine, A. Takayasu, and N. Yamanaka. *Principle of verified numerical computations.* Corona publisher, Tokyo, Japan, 2018 [in Japanese].

[5] S.M. Rump. Verification methods: Rigorous results using floating-point arithmetic. *Acta Numerica*, 19:287–449, 2010.

[6] S.M. Rump. Verified bounds for singular values, in particular for the spectral norm of a matrix and its inverse. *BIT Numerical Mathematics*, 51(2):367–384, 2011.