SUPPLEMENTARY MATERIALS: LEAST-SQUARES SPECTRAL METHODS FOR ODE EIGENVALUE PROBLEMS*

BEHNAM HASHEMI† AND YUJI NAKATSUKASA‡

Here we first briefly discuss pseudospectra for quasimatrices and then present an optimality result on the extension of the Ito-Murota approach to quasimatrix pairs; see subsection 5.2.

Let A and B be $\infty \times n$ quasimatrices. We define the spectrum of the pair (A, B) as follows:

$$\Lambda(\mathsf{A}, \mathsf{B}) := \{\lambda \in \mathbb{C} : \exists 0 \neq \mathbf{v} \in \mathbb{C}^n,\ \forall x \in [-1, 1],\ \mathsf{A}(x)\mathbf{v} = \lambda \mathsf{B}(x)\mathbf{v}\}. \tag{SM0.1}$$

We mentioned previously that rectangular matrix pencils usually do not have eigenvalues. Analogously, the spectrum of the quasimatrix pair (A, B) is often empty. We nonetheless wish to discuss points that are nearly in the spectrum.

A nice tool for analyzing the behavior of the spectrum of *matrices* and *operators* [SM3] under small perturbations is the pseudospectrum; see [SM11] and [SM13] for a detailed discussion of the various aspects of pseudospectra in those situations. This notion can be extended readily to the case of quasimatrices. Let $\epsilon > 0$ be arbitrary and $\sigma_{\min}$ denote the smallest singular value. We call

$$\Lambda_\epsilon(\mathsf{A}, \mathsf{B}) = \left\{ z \in \mathbb{C} : \frac{\sigma_{\min}(z\mathsf{B} - \mathsf{A})}{\sqrt{1 + |z|^2}} < \epsilon \right\}, \tag{SM0.2}$$

the ϵ-pseudospectrum of the quasimatrix pair (A, B), see [SM1, SM2]. Obviously $\Lambda(\mathsf{A}, \mathsf{B}) \subseteq \Lambda_\epsilon(\mathsf{A}, \mathsf{B})$, for every $\epsilon > 0$. More generally, $\Lambda_{\hat\epsilon}(\mathsf{A}, \mathsf{B}) \subseteq \Lambda_\epsilon(\mathsf{A}, \mathsf{B})$ if $\hat\epsilon < \epsilon$. By minimizing the perturbation so that solutions exist as in (5.2), the Ito-Murota algorithm can be seen as a method for finding a set of points in the ϵ-pseudospectrum of (A, B), for small values of ϵ.

Example 1. Let $\mathsf{A}(x) = [T_0(x), T_1(x), \cdots, T_5(x)]$ and $\mathsf{B}(x) = [P_0(x), P_1(x), \cdots, P_5(x)]$ defined on $[-1, 1]$, where P_i denotes the i-th Legendre polynomial[1]. It can be verified that $\Lambda(\mathsf{A}, \mathsf{B}) = \{1, 1, \frac{4}{3}, \frac{8}{5}, \frac{64}{35}, \frac{128}{63}\}$ and that for example $\mathbf{v}_3 = [\frac{-1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 0, 0, 0]^T$ is the eigenvector corresponding to $\lambda_3 = \frac{4}{3}$. Figure SM1 illustrates $\Lambda_\epsilon(\mathsf{A}, \mathsf{B})$ on a grid of z-values.

As explained before, a major difficulty with the generalized rectangular eigenvalue problem even in the discrete case is that the eigenpairs may fail to exist under perturbations. Motivated by the work [SM1] of Boutry, Elad, Golub, and Milanfar, we focus on the following reformulation of (SM0.1) that searches for the minimal perturbation to the quasimatrix-matrix pencil (A, B) such that the perturbed pencil $(\hat{\mathsf{A}}, \hat{\mathsf{B}})$ has n

*Supplementary material for SISC MS#M144593.
https://doi.org/10.1137/21M1445934

†Mathematics Program, Department of Mathematics, Statistics and Physics, College of Arts and Sciences, Qatar University, 2713, Doha, Qatar, and School of Mathematics, Institute for Research in Fundamental Sciences (IPM), P.O. Box 19395-5746, Tehran, Iran (hoseynhashemi@gmail.com).

‡Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK (nakatsukasa@maths.ox.ac.uk).

[1]In Chebfun [SM4] these $\infty \times 6$ quasimatrices can be readily constructed with the commands `A = chebpoly(0:5); B = legpoly(0:5);`

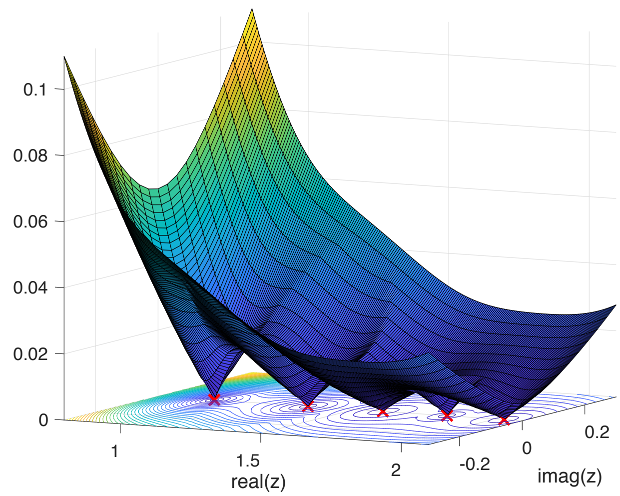

FIG. SM1. *Graphical inspection of the spectrum and the pseudospectra of the $\infty \times 6$ quasimatrix pencil of Example 1. Red crosses denote the eigenvalues.*

linearly independent eigenvectors:

$$\text{(SM0.3)} \qquad \begin{cases} \text{minimize} & \|[\hat{\mathsf{A}} - \mathsf{A} \quad \hat{\mathsf{B}} - \mathsf{B}]\|_F^2; \\ \text{subject to} & \hat{\mathsf{A}}, \hat{\mathsf{B}} \in \mathbb{C}^{(\infty+d)\times n},\ \{(\lambda_k, \mathbf{v}_k)\}_{k=1}^n \subseteq \mathbb{C} \times \mathbb{C}^n, \\ & \hat{\mathsf{A}}\mathbf{v}_k = \lambda_k \hat{\mathsf{B}}\mathbf{v}_k, \quad k = 1, 2, \dots, n, \\ & \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\} : \text{ linearly independent.} \end{cases}$$

The next proposition gives a sufficient condition for the existence and uniqueness of the optimal solution to (SM0.3) and hints an algorithm for its solution in terms of the following SVD[2].

$$\text{(SM0.4)} \qquad [\mathsf{B} \quad \mathsf{A}] = \mathsf{U}\Sigma V^* = [\mathsf{U}_1 \ \mathsf{U}_2] \begin{bmatrix} \Sigma_{1:n} & \\ & \Sigma_{n+1:2n} \end{bmatrix} \begin{bmatrix} V_{11}^* & V_{21}^* \\ V_{12}^* & V_{22}^* \end{bmatrix},$$

It is a direct extension of the results from the discrete case [SM5, SM7] to quasimatrix-matrix objects; see also [SM12, p. 51]. To keep the paper self-contained we give a proof of the proposition which requires a few definitions and the following three lemmas. For a quasimatrix-matrix $\mathsf{A} \in \mathbb{C}^{(\infty+d)\times n}$ we define

$$\text{(SM0.5)} \qquad \|\mathsf{A}\|_2 := \max_{0 \neq \mathbf{x} \in \mathbb{C}^n} \frac{\|\mathsf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2},$$

and

$$\|\mathsf{A}\|_F := \Big(\text{trace}(\mathsf{A}^*\mathsf{A})\Big)^{1/2} = \Big(\sum_{i=1}^n \sigma_i^2(\mathsf{A})\Big)^{1/2}.$$

In addition, A is called unitary if $\mathsf{A}^*\mathsf{A} = I_n$.

LEMMA SM0.1. For any $W \in \mathbb{C}^{n\times k}$ and unitary quasimatrix-matrix $\mathsf{U} \in \mathbb{C}^{(\infty+d)\times n}$, we have

$$\|\mathsf{U}W\|_2 = \|W\|_2.$$

[2]See (4.2) for details but notice that it was for the case of $[\mathsf{A} \ \mathsf{B}]$.

Proof. For every $\mathbf{x} \in \mathbb{C}^n$ we have $\|\mathsf{U}\mathbf{x}\|_2 = \|(\mathsf{U}\mathbf{x})^*(\mathsf{U}\mathbf{x})\|_2 = \|\mathbf{x}^* I_n \mathbf{x}\|_2 = \|\mathbf{x}\|_2$. The invariance of the 2-norm follows from (SM0.5). ∎

The following continuous analogue of the well-known Eckart-Young-Mirsky theorem will be used. It is stated in terms of the Frobenius norm of a quasimatrix-matrix.

LEMMA SM0.2. *The first k-terms in the SVD of a quasimatrix-matrix $\mathsf{A} \in \mathbb{C}^{(\infty+d)\times n}$ form its best rank-k approximation in the Frobenius norm.*

Proof. The reasoning is analogous to that of quasimatrices as in [SM9, p. 62] and [SM10]. ∎

The next lemma is a continuous analogue of the the result in [SM6, p. 321] for the discrete case.

LEMMA SM0.3. Let $\mathsf{A}, \mathsf{B} \in \mathbb{C}^{(\infty+d)\times n}$ and consider the SVD (SM0.4) of $[\mathsf{B} \ \ \mathsf{A}]$. If $\sigma_n(\mathsf{B}) > \sigma_{n+1}([\mathsf{B} \ \ \mathsf{A}])$, then V_{11} and V_{22} are nonsingular and $\sigma_n([\mathsf{B} \ \ \mathsf{A}]) > \sigma_{n+1}([\mathsf{B} \ \ \mathsf{A}])$.

Proof. We first use proof by contradiction to show that V_{22} is nonsingular. Assume that there exists a vector $\mathbf{x}$ with unit 2-norm such that $V_{22}\mathbf{x} = 0$. The second equation in $[\mathsf{B} \ \ \mathsf{A}]V = [\mathsf{U}_1 \ \ \mathsf{U}_2]\Sigma$ reads as $\mathsf{B}V_{12} + \mathsf{A}V_{22} = \mathsf{U}_2\Sigma_{n+1:2n}$ meaning that

$$\|\mathsf{B}V_{12}\mathbf{x}\|_2 = \|\mathsf{U}_2\Sigma_{n+1:2n}\mathbf{x}\|_2. \tag{SM0.6}$$

As we saw in Lemma SM0.1, the 2-norm is invariant under multiplication by a quasimatrix-matrix like U_2 whose columns are orthonormal function-vectors. Therefore, using (SM0.6) we have

$$\sigma_{n+1}([\mathsf{B} \ \ \mathsf{A}]) = \|\Sigma_{n+1:2n}\|_2 = \|\mathsf{U}_2\Sigma_{n+1:2n}\|_2 \geq \|\mathsf{U}_2\Sigma_{n+1:2n}\mathbf{x}\|_2 = \|\mathsf{B}V_{12}\mathbf{x}\|_2 \geq \sigma_{\min}(\mathsf{B}),$$

which is a contradiction.

The second part follows if we prove that $\sigma_n([\mathsf{B} \ \ \mathsf{A}]) \geq \sigma_n(\mathsf{B})$. This is an interlacing property for singular values of a quasimatrix-matrix which is valid because the singular values of any quasimatrix-matrix are just the singular values of the R factor of its QR factorization (see (4.3) and (4.4)) and the R factor is always a discrete matrix for which the interlacing property of singular values is a basic fact [SM8]. ∎

PROPOSITION SM0.4. *Let $\mathsf{A}, \mathsf{B} \in \mathbb{C}^{(\infty+d)\times n}$ and consider the SVD (SM0.4) of $[\mathsf{B} \ \ \mathsf{A}]$. If $\sigma_n(\mathsf{B}) > \sigma_{n+1}([\mathsf{B} \ \ \mathsf{A}])$, then there exists a unique optimal solution to (SM0.3) attained for*

$$\hat{\mathsf{A}} = \mathsf{A} - \mathsf{U}_2\Sigma_{n+1:2n}V_{22}^*, \qquad \textit{and} \qquad \hat{\mathsf{B}} = \mathsf{B} - \mathsf{U}_2\Sigma_{n+1:2n}V_{12}^*,$$

if and only if $V_{12}V_{22}^{-1}$ is diagonalizable.

Proof. We prove the result in two steps. In the first step we just extend the argument by Ito and Murota [SM7, Thm. 2, part i)] to the case of quasimatrix-matrix objects by showing that (SM0.3) is equivalent to the following continuous-discrete total least-squares problem

$$\begin{cases} \text{minimize} & \|[\hat{\mathsf{A}} - \mathsf{A} \quad \hat{\mathsf{B}} - \mathsf{B}]\|_F^2; \\ \text{subject to} & \hat{\mathsf{A}}, \hat{\mathsf{B}} \in \mathbb{C}^{(\infty+d)\times n} \\ & \ \operatorname{range}(\hat{\mathsf{A}}) \subseteq \operatorname{range}(\hat{\mathsf{B}}). \end{cases} \tag{SM0.7}$$

Let P_1 denote the set of all feasible solutions to (SM0.3) and assume that $\hat{\mathsf{A}}, \hat{\mathsf{B}}$ and $\{(\lambda_k, \mathbf{v}_k)\}_{k=1}^n$ are one of those feasible solutions. Assuming $V := [\mathbf{v}_1, \ \mathbf{v}_2, \ \ldots, \mathbf{v}_n]$ and

$\Lambda := \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, the constraints $\hat{\mathsf{A}}\mathbf{v}_k = \lambda_k \hat{\mathsf{B}}\mathbf{v}_k$ for $k = 1, 2, \ldots, n$ means that $\hat{\mathsf{A}}V = \hat{\mathsf{B}}V\Lambda$ and since columns of V are linearly independent, (SM0.3) is equivalent to $\hat{\mathsf{A}} = \hat{\mathsf{B}}V\Lambda V^{-1}$. This representation shows that P_1 is the same as the set of all $(\tilde{\mathsf{A}}, \tilde{\mathsf{B}}) \in \mathbb{C}^{(\infty+d)\times n} \times \mathbb{C}^{(\infty+d)\times n}$ satisfying $\tilde{\mathsf{A}} = \tilde{\mathsf{B}}Z$ where $Z \in \mathbb{C}^{n\times n}$ is diagonalizable.

Let P_2 denote the set of all $(\tilde{\mathsf{A}}, \tilde{\mathsf{B}}) \in \mathbb{C}^{(\infty+d)\times n} \times \mathbb{C}^{(\infty+d)\times n}$ satisfying $\tilde{\mathsf{A}} = \tilde{\mathsf{B}}Z$ where $Z \in \mathbb{C}^{n\times n}$ is *not* necessarily diagonalizable. This means that P_2 is the set of all $(\tilde{\mathsf{A}}, \tilde{\mathsf{B}})$ such that $\mathrm{range}(\tilde{\mathsf{A}}) \subseteq \mathrm{range}(\tilde{\mathsf{B}})$. Obviously $P_1 \subseteq P_2$. Since there exists a diagonalizable matrix in an arbitrarily close neighborhood of any square matrix Z, we have $P_2 \subseteq \overline{P_1}$ where $\overline{P_1}$ denotes the closure of P_1. In addition, $\|[\tilde{\mathsf{A}} - \mathsf{A} \quad \tilde{\mathsf{B}} - \mathsf{B}]\|_F^2$ is a continuous function of $(\tilde{\mathsf{A}}, \tilde{\mathsf{B}})$. Therefore,

$$\inf_{(\tilde{\mathsf{A}},\tilde{\mathsf{B}})\in P_1} \{\|[\tilde{\mathsf{A}} - \mathsf{A} \quad \tilde{\mathsf{B}} - \mathsf{B}]\|_F^2\} = \inf_{(\tilde{\mathsf{A}},\tilde{\mathsf{B}})\in P_2} \{\|[\tilde{\mathsf{A}} - \mathsf{A} \quad \tilde{\mathsf{B}} - \mathsf{B}]\|_F^2\},$$

which means that optimal solutions to (SM0.3) and (SM0.7) are the same.

Now in the second step we first derive explicit formulas for the unique optimal solution to (SM0.7) (and according to the first step an optimal solution to (SM0.3) as well.) The assumption $\sigma_n(\mathsf{B}) > \sigma_{n+1}([\mathsf{B}, \mathsf{A}])$ together with Lemma SM0.3 implies that V_{11} is nonsingular and that $\sigma_n([\mathsf{B} \quad \mathsf{A}]) > \sigma_{n+1}([\mathsf{B}, \mathsf{A}])$. We rewrite the formula $\hat{\mathsf{A}} = \hat{\mathsf{B}}Z$ as

$$[\hat{\mathsf{B}} \quad \hat{\mathsf{A}}] \begin{bmatrix} Z \\ -I \end{bmatrix} = 0,$$

implying that the rank of the augmented quasimatrix-matrix $[\hat{\mathsf{B}} \quad \hat{\mathsf{A}}]$ is at most n. Therefore, we can view solving (SM0.7) as finding the minimal (in the Frobenius norm) rank-n perturbation $[\hat{\mathsf{B}} \quad \hat{\mathsf{A}}]$ to $[\mathsf{B} \quad \mathsf{A}]$. According to Lemma SM0.2, the latter problem can be solved by the rank-n truncation of the SVD (SM0.4), i.e.,

$$[\hat{\mathsf{B}} \quad \hat{\mathsf{A}}] = \mathsf{U}_1 \Sigma_{1:n} [V_{11}^* \quad V_{21}^*],$$

which is unique as $\sigma_n([\mathsf{B} \quad \mathsf{A}]) > \sigma_{n+1}([\mathsf{B} \quad \mathsf{A}])$. To find the corresponding solution $\hat{Z}$ to $\hat{\mathsf{B}}Z = \hat{\mathsf{A}}$ we therefore put

$$\mathsf{U}_1 \Sigma_{1:n} V_{11}^* Z = \mathsf{U}_1 \Sigma_{1:n} V_{21}^*.$$

Since V_{11} is nonsingular, $\hat{Z} = (V_{21}V_{11}^{-1})^*$ solves $\hat{\mathsf{B}}Z = \hat{\mathsf{A}}$ in (SM0.7). From the orthogonality of the partitioned matrix V it follows that $(V_{21}V_{11}^{-1})^* = -V_{12}V_{22}^{-1}$.

On the other hand if $\hat{Z}$ corresponding with the optimal solution $[\hat{\mathsf{B}} \quad \hat{\mathsf{A}}]$ to (SM0.7) is diagonalizable, then its eigenpairs $\{(\lambda_k, v_k)\}_{k=1}^n$ satisfy a representation of the form $\hat{Z} = V\Lambda V^{-1}$ which by $\hat{\mathsf{B}}\hat{Z} = \hat{\mathsf{A}}$ means that $(\hat{\mathsf{B}}, \hat{\mathsf{A}})$ is an optimal solution to (SM0.3). Conversely, if $(\hat{\mathsf{B}}, \hat{\mathsf{A}})$ is an optimal solution to (SM0.3), then $\hat{\mathsf{A}} = \hat{\mathsf{B}}V\Lambda V^{-1}$ which means that $\hat{Z} = V\Lambda V^{-1}$ is diagonalizable. ∎

REFERENCES

[SM1] G. Boutry, M. Elad, G. H. Golub, and P. Milanfar, *The generalized eigenvalue problem for nonsquare pencils using a minimal perturbation approach*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 582–601, http://dx.doi.org/10.1137/S0895479803428795.

[SM2] S. Das and A. Neumaier, *Solving overdetermined eigenvalue problems*, SIAM J. Sci. Comput., 35 (2013), pp. A541–A560, https://doi.org/10.1137/110828514.

[SM3] E. B. Davies, *Pseudospectra of differential operators*, Journal of Operator Theory, (2000), pp. 243–262.

[SM4] T. A. Driscoll, N. Hale, and L. N. Trefethen, *Chebfun Guide*, Pafnuty Publications, Oxford, 2014.

[SM5] G. H. Golub and C. F. Van Loan, *An analysis of the total least squares problem*, SIAM J. Numer. Anal., 17 (1980), pp. 883–893, https://doi.org/10.1137/0717073.

[SM6] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, fourth ed., 2013.

[SM7] S. Ito and K. Murota, *An algorithm for the generalized eigenvalue problem for nonsquare matrix pencils by minimal perturbation approach*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 409–419, https://doi.org/10.1137/14099231X.

[SM8] R. C. Thompson, *Principal submatrices. IX. Interlacing inequalities for singular values of submatrices*, Linear Algebra Appl., 5 (1972), pp. 1–12, https://doi.org/10.1016/0024-3795(72)90013-4.

[SM9] A. Townsend, *Computing with Functions in Two Dimensions*, PhD thesis, University of Oxford, 2014.

[SM10] A. Townsend and L. N. Trefethen, *Continuous analogues of matrix factorizations*, Proc. A., 471 (2015), pp. 20140585, 21, https://doi.org/10.1098/rspa.2014.0585.

[SM11] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra*, Princeton University Press, Princeton, NJ, 2005.

[SM12] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem*, SIAM, Philadelphia, PA, 1991, https://doi.org/10.1137/1.9781611971002.

[SM13] T. G. Wright and L. N. Trefethen, *Pseudospectra of rectangular matrices*, IMA J. Numer. Anal., 22 (2002), pp. 501–519, http://dx.doi.org/10.1093/imanum/22.4.501.