ROBUST PARAMETER IDENTIFIABILITY ANALYSIS VIA COLUMN SUBSET SELECTION*

KATHERINE J. PEARCE†, ILSE C.F. IPSEN‡, MANSOOR A. HAIDER§, ARVIND K. SAIBABA¶, AND RALPH C. SMITH‖

Abstract. We advocate a numerically reliable and accurate approach for practical parameter identifiability analysis: Applying column subset selection (CSS) to the sensitivity matrix, instead of computing an eigenvalue decomposition of the Fischer information matrix. Identifiability analysis via CSS has three advantages: (i) It quantifies reliability of the subsets of parameters selected as identifiable and unidentifiable. (ii) It establishes criteria for comparing the accuracy of different algorithms. (iii) The implementations are numerically more accurate and reliable than eigenvalue methods applied to the Fischer matrix, yet without an increase in computational cost. The effectiveness of the CSS methods is illustrated with extensive numerical experiments on sensitivity matrices from six physical models, as well as on adversarial synthetic matrices. Among the CSS methods, we recommend an implementation based on the strong rank-revealing QR algorithm because of its rigorous accuracy guarantees for both identifiable and non-identifiable parameters.

Key words. Sensitivity matrix, Fischer information matrix, systems of ordinary differential equations, dynamical systems, eigenvalue decomposition, singular value decomposition, pivoted QR decomposition, rank-revealing QR decomposition

AMS subject classifications. 65F25, 65F35, 65Z05, 15A12, 15A18, 15A23, 15A42, 37N25, 92C42

1. Introduction.

In data-driven mathematical modeling, the ability to reliably estimate model parameters depends on the set of available observations, the scope of system responses for which such observations are available, the inherent mathematical structure of the model, and the parameter estimation method. Identifiability analysis evaluates the ability to accurately estimate each parameter in a model and, in some cases, quantifies the extent to which this estimate is reliable. It has wide-ranging implications for a variety of applications, including analysis of disease and epidemiology models to guide treatment regimes, physiologically-based pharmacokinetic (PBPK) and quantitative system pharmacology (QSP) models for drug development, and coupled multi-physics models for next-generation nuclear power plant design. In particular, identifiability analysis can be more challenging, yet also have greater impact, in applications where the number of model variables and parameters is significantly greater than the number of responses with available data.

Practical identifiability analysis refers to the partitioning of parameters in a mathematical model into two groups: *identifiable* parameters that can be reliably estimated from data and those that cannot, termed *unidentifiable*. At the heart of many practical identifiability methods is the sensitivity matrix $\boldsymbol{S}$, whose columns represent model parameters and whose rows represent observations (data) for a quantity of interest. A

*Submitted to the editors May 2022.

Funding: This research was supported in part by grants NSF DMS-1745654 and DOE DE-SC0022085 (ICFI); NSF DMS-1845406 (AKS), NSF-DMS-1638521 (KJP, MAH), NSF-DMS-2053812 (MAH, RCS), AFOSR FA9550-18-1-0457 (RCS)

Address for all authors: Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205, USA

†kjpearce@ncsu.edu, https://kjpearce.github.io/.

‡ipsen@ncsu.edu, https://ipsen.math.ncsu.edu/.

§mahaider@ncsu.edu, https://haider.wordpress.ncsu.edu.

¶asaibab@ncsu.edu, https://asaibab.math.ncsu.edu/.

‖rsmith@ncsu.edu, https://https://rsmith.math.ncsu.edu/.

common approach extracts identifiable and unidentifiable parameters from eigenvalues and eigenvectors of the Fischer information matrix $\boldsymbol{S}^T\boldsymbol{S}$. However, the sensitivity matrix $\boldsymbol{S}$ is often ill-conditioned, that is, sensitive to small perturbations, so that the explicit formation of the cross product $\boldsymbol{S}^T\boldsymbol{S}$ can inflict a serious loss of accuracy.

We apply instead column subset selection (CSS) to the sensitivity matrix $\boldsymbol{S}$, which has the same computational complexity as eigenvalue methods on the Fischer matrix $\boldsymbol{S}^T\boldsymbol{S}$. We derive bounds that show the superior accuracy of CSS, and corroborate this with extensive numerical experiments on a variety of model-based and adversarial synthetic matrices. The higher accuracy of the CSS methods produces a more reliable distinction between identifiable and unidentifiable parameters, as illustrated by their highly consistent performance across across this suite of test matrices. This is especially critical when the identifiable parameters inform subsequent investigations [3, 9, 38].

1.1. Contributions. We advocate a numerically reliable and accurate approach for practical parameter identifiability analysis: Applying column subset selection to the sensitivity matrix, instead of computing an eigenvalue decomposition of the Fischer information matrix.

1. We interpret algorithms based on eigenvalue decompositions of the Fischer matrix [27] as known column subset selection (CSS) methods applied to the sensitivity matrix (section 3). This connection allows us to derive rigorous guarantees for the accuracy and reliability of the parameter identification that were previously lacking.
2. Identifiability analysis via CSS (section 4) has five advantages:
   (a) It broadens the applicability of parameter identifiability analysis by permitting the use of synthetic data generated from an additive observation model. This is crucial when experimental data are not available or optimization for determining nominal parameter values is not feasible.
   (b) It incorporates parameter correlation.
   (c) It quantifies reliability of the subsets of parameters selected as identifiable and unidentifiable.
   (d) It establishes criteria for comparing the accuracy of different algorithms.
   (e) The implementations are numerically more accurate and reliable than eigenvalue methods applied to the Fischer matrix, yet without an increase in computational cost.
3. We perform extensive numerical experiments (section 5.1) on sensitivity matrices from six physical models (section 2.2, Appendix B) to illustrate the accuracy and reliability of the CSS methods.
4. Among the four CSS methods (Algorithms 4.1–4.4), we recommend an implementation based on the strong rank-revealing QR algorithm (Algorithm 4.4) because of its rigorous accuracy guarantees for both, identifiable and unidentifiable parameters, through bounds that have only a polynomial dependence on the number of relevant parameters, rather than an exponential dependence as in Algorithms 4.1–4.3.
5. We construct an adversarial matrix, the SHIPS matrix (section 5.2) to amplify accuracy differences among the CSS methods. Although synthetic, the adversarial matrices (section 5.2) still admit an interpretation as sensitivity matrices for certain dynamical systems (Appendix C).

The CSS algorithms (section 4) are based on existing work and presented with a view towards understanding rather than efficiency. In the same vein, the correct-

ness proofs (section A) are geared towards exposition: self-contained, as simple as possible, and more general with slightly fewer assumptions. With a view towards reproducibility, our implementations are available on https://github.com/kjpearce/CSS-Algs-for-Sens-Identifiability.

2. Parameter sensitivity and identifiability.

We define the notion of parameter identifiability (section 2.1), and present real applications that require it (section 2.2).

2.1. Parameter identifiability.

We assume that a model's quantity of interest y, such as a state variable in a system of differential equations, can be expressed as a scalar-valued function of system inputs and parameters, $y = h(\mathbf{u}; \mathbf{q})$. Here the vector $\mathbf{u}$ represents system inputs, such as time, and the vector $\mathbf{q} \in \mathbb{R}^p$ the model parameters.

We denote the *sensitivity* of y with respect to the parameter q_j, evaluated at the ith observation and a specific point $\mathbf{q}^*$ in the admissible parameter space, by

$$s_{ij} = \frac{\partial h_i(\mathbf{u}; \mathbf{q})}{\partial q_j}\Big|_{\mathbf{q}=\mathbf{q}^*}, \qquad 1 \le i \le n, \quad 1 \le j \le p.$$

The sensitivity matrix is $\boldsymbol{S} = (s_{ij}) \in \mathbb{R}^{n\times p}$, and has more rows than columns, $n \ge p$. The parameters $\mathbf{q}$ are *sensitivity-identifiable* at $\mathbf{q}^*$ if $\boldsymbol{S}^T\boldsymbol{S}$ is invertible [9, 32, 42]. Our goal is to determine those columns of $\boldsymbol{S}$ that correspond to the most sensitivity-identifiable and the least sensitivity-identifiable parameters.

2.2. Practical applications with sensitivity matrices.

We describe an epidemiological compartment model in detail (section 2.2.1), and summarize five other mathematical models together with their quantities of interest (section 2.2.2).

2.2.1. SVIR Model.

The epidemiological SVIR compartment model in Figure B.1(c), models the spread of disease among susceptible S, vaccinated V, infectious I, and recovered R in a population of N individuals; and consists of a coupled system of four ordinary differential equations with specified initial conditions,

$$\begin{aligned}
\frac{dS}{dt} &= -\beta\frac{IS}{N}, & S(0) &= S_0\\
\frac{dV}{dt} &= \nu S - \alpha\beta\frac{IV}{N}, & V(0) &= V_0,\\
\frac{dI}{dt} &= \beta\frac{IS}{N} + \alpha\beta\frac{IV}{N} - \gamma I, & I(0) &= I_0,\\
\frac{dR}{dt} &= \gamma I, & R(0) &= R_0.
\end{aligned}$$

The epidemiological parameters $\mathbf{q} = \begin{bmatrix}\beta & \nu & \alpha & \gamma\end{bmatrix}^\top$ govern the system dynamics; the system input $\mathbf{u}$ is time t; and the quantity of interest is $y = h(t; \mathbf{q}) \equiv I(t; \mathbf{q})$ the number of infectious individuals at time t. Discretization with respect to time $t = t_i$, $1 \le i \le n$, produces a sensitivity matrix evaluated at a nominal point $\mathbf{q}^*$,

$$\boldsymbol{S} = \left[\frac{\partial h(t_i;\mathbf{q})}{\partial\beta} \quad \frac{\partial h(t_i;\mathbf{q})}{\partial\nu} \quad \frac{\partial h(t_i;\mathbf{q})}{\partial\alpha} \quad \frac{\partial h(t_i;\mathbf{q})}{\partial\gamma}\right]\Big|_{\mathbf{q}=\mathbf{q}^*} \in \mathbb{R}^{n\times 4}.$$

Nominal parameter values are often selected from the literature, as shown in Table B.1, or as solutions of inverse problems with available data. Numerical sensitivities in $\boldsymbol{S}$ are estimated from derivative approximations, such as finite difference or complex-step approximations [33, 34].

2.2.2. Six models from physical applications. We present numerical experiments (section 5) for the six models below, with quantities of interest in Table 2.1. More details can be found in section B.

- SVIR: See above.
- SEVIR [39]: This extension of SVIR model adds an additional compartment for individuals E who have been exposed but are not yet infectious.
- COVID [40]: This extension of SEVIR splits the infectious group into compartments for asymptomatic, symptomatic, and hospitalized individuals.
- HGO [19]: This model for the biomechanical deformation of the left pulmonary artery vessel wall is based on nonlinear hyperelastic structural relations, and calibrated to in vitro experiments on normal and hypertensive mice.
- Wound [38]: This model for in vitro fibrin matrix polymerization during hemostasis concerns clot formation during the first stage of wound healing, and is based on biochemical reaction kinetics.
- Neuro [20]: This model of the neurovascular coupling (NVC) response describes local changes in vascular resistance that result from neuronal activity, and is based on nonlinear ODEs.

Model	Type	p	Quantity of Interest
SVIR	Epidemiological	4	# Infectious individuals
SEVIR	Epidemiological	5	# Infectious individuals
COVID	Epidemiological	8	# Infectious (sympt., asymp., hospitalized)
HGO	Cardiovascular	8	Vessel lumen area and wall thickness
Wound	Wound Healing	11	Fibrin matrix (*in vitro* clot) concentration
Neuro	Neurological	175	Blood oxyhemoglobin concentration

TABLE 2.1
Number of parameters p and quantities of interest for the models in section 2.2

3. Background. We express sensitivity analysis on the eigenvectors of the Fischer matrix $\boldsymbol{F} = \boldsymbol{S}^T\boldsymbol{S}$ as column subset selection on the sensitivity matrix $\boldsymbol{S}$.

After briefly introducing notation (section 3.1), we review identfiability analysis based on eigenvectors of the Fischer matrix (section 3.2), the singular value decomposition of the sensitivity matrix (section 3.3), column subset selection on the sensitivity matrix (section 3.4), determination of the number k of identifiable parameters (section 3.5), and finally the implementation of column subset selection via QR decompositions (section 3.6).

3.1. Notation. We denote matrices by bold upper case letters. The identity matrix is

$$\boldsymbol{I}_p \equiv \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1 & \cdots & \boldsymbol{e}_p \end{bmatrix} \in \mathbb{R}^{p\times p}$$

with columns that are the canonical vectors $\boldsymbol{e}_j \in \mathbb{R}^p$.

We assume that the sensitivity matrix $\boldsymbol{S} \in \mathbb{R}^{n\times p}$ is tall and skinny, with at least as many rows as columns, $n \geq p$. The p columns of $\boldsymbol{S}$ represent parameters and its rows represent observations. The Fischer information matrix is the cross product matrix $\boldsymbol{F} \equiv \boldsymbol{S}^T\boldsymbol{S} \in \mathbb{R}^{p\times p}$, where the superscript T denotes the transpose.

3.2. Eigenvalue decomposition of the Fischer matrix. Existing methods [27, 35] select parameters by inspecting the eigenvectors of the Fischer matrix $\boldsymbol{F} = \boldsymbol{S}^T\boldsymbol{S} \in \mathbb{R}^{p\times p}$. Since it is real symmetric positive semi-definite, its eigenvalue decomposition has the form

$$\boldsymbol{F} = \boldsymbol{V}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_p\end{bmatrix}\boldsymbol{V}^T, \qquad \lambda_1 \geq \cdots \geq \lambda_p \geq 0, \tag{3.1}$$

where λ_j are the eigenvalues. The eigenvector matrix $\boldsymbol{V} \in \mathbb{R}^{p\times p}$ is an orthogonal matrix with $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_p = \boldsymbol{V}\boldsymbol{V}^T$. Its columns and elements are

$$\boldsymbol{V} = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_p\end{bmatrix} = \begin{bmatrix}v_{11} & \cdots & v_{1p} \\ \vdots & & \vdots \\ v_{p1} & \cdots & v_{pp}\end{bmatrix}$$

In particular, the trailing column $\mathbf{v}_p$ is an eigenvector associated with a smallest eigenvalue λ_p, so $\boldsymbol{F}\mathbf{v}_p = \lambda_p\mathbf{v}_p$. If $\lambda_p > 0$, then $\boldsymbol{F}$ is nonsingular.

The parameter with index j is represented by column j of $\boldsymbol{S}$. The corresponding column of the Fischer matrix is

$$\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{e}_j = \boldsymbol{F}\boldsymbol{e}_j = \boldsymbol{V}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_p\end{bmatrix}\boldsymbol{V}^T\boldsymbol{e}_j \quad \text{where} \quad \boldsymbol{V}^T\boldsymbol{e}_j = \begin{bmatrix}v_{j1} \\ \vdots \\ v_{jp}\end{bmatrix}, \qquad 1 \leq j \leq p.$$

Thus, column j of $\boldsymbol{S}$ depends on column j of $\boldsymbol{V}^T$ which, in turn, contains element j of each eigenvector.

> Selecting element j of any eigenvector of $\boldsymbol{F} = \boldsymbol{S}^T\boldsymbol{S}$ amounts to selecting the parameter with index j in $\boldsymbol{S}$.

CAUTION. *Explicit formation of the Fischer matrix $\boldsymbol{F} = \boldsymbol{S}^T\boldsymbol{S}$ can lead to significant loss of information, thus affecting subsequent practical identifiability analysis.*

For instance [16, Section 5.3.2], in customary double precision floating point arithmetic with unit roundoff $2^{-53} \approx 1.1 \cdot 10^{-16}$, the sensitivity matrix

$$\boldsymbol{S} = \begin{bmatrix}1 & 1 \\ 10^{-9} & 0 \\ 0 & 10^{-9}\end{bmatrix}$$

has linearly independent columns, and $\operatorname{rank}(\boldsymbol{S}) = 2$. *In contrast, the Fischer information matrix computed in double precision floating point arithmetic*

$$\operatorname{fl}(\boldsymbol{S}^T\boldsymbol{S}) = \begin{bmatrix}1 & 1 \\ 1 & 1\end{bmatrix}$$

is singular, because the diagonal elements computed in double precision are

$$\operatorname{fl}(1 + 10^{-9}\cdot 10^{-9}) = \operatorname{fl}(1 + 10^{-18}) = 1,$$

where the operator $\operatorname{fl}(\cdot)$ *represents the output of a computation in floating point arithmetic.*

3.3. Singular value decomposition of the sensitivity matrix. We avoid the explicit formation of the Fischer matrix $\boldsymbol{F} = \boldsymbol{S}^T\boldsymbol{S}$, and instead operate directly on the sensitivity matrix $\boldsymbol{S}$, without increasing the computation time.

This is done with the help of the (thin) singular value decomposition (SVD) [16, section 8.6]

$$\boldsymbol{S} = \boldsymbol{U} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_p \end{bmatrix} \boldsymbol{V}^T, \qquad \sigma_1 \geq \cdots \geq \sigma_p \geq 0, \tag{3.2}$$

where σ_j are the singular values of $\boldsymbol{S}$, the left singular vector matrix $\boldsymbol{U} \in \mathbb{R}^{n\times p}$ has orthonormal columns with $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_p$, and the right singular vector matrix $\boldsymbol{V}$is is identical to the orthogonal matrix in (3.1).

Substituting the SVD of $\boldsymbol{S}$ into $\boldsymbol{F}$ gives (3.1) with eigenvalues $\lambda_j = \sigma_j^2$, $1 \leq j \leq p$. Thus, squared singular values of $\boldsymbol{S}$ are the eigenvalues of $\boldsymbol{F}$, and the right singular vectors of $\boldsymbol{S}$ are eigenvectors of $\boldsymbol{F}$.

> Selecting element j of any right singular vector of $\boldsymbol{S}$ amounts to selecting the parameter with index j in column j of $\boldsymbol{S}$.

As a consequence, all information provided by the eigenvalue decomposition of the Fischer matrix $\boldsymbol{F} = \boldsymbol{S}^T\boldsymbol{S}$ is available from the SVD of the sensitivity matrix $\boldsymbol{S}$. Computation of the SVD is not more expensive; see Remark 3.1.

3.4. Column subset selection on the sensitivity matrix. We go a step further, and select the parameters directly from the sensitivity matrix $\boldsymbol{S}$, rather than detouring through an eigenvalue or singular value decomposition.

Specifically, we compute a permutation matrix $\boldsymbol{P} \in \mathbb{R}^{p\times p}$ that reorders the columns of the sensitivity matrix $\boldsymbol{S}$,

$$\boldsymbol{S}\boldsymbol{P} = \begin{bmatrix} \boldsymbol{S}_1 & \boldsymbol{S}_2 \end{bmatrix} \tag{3.3}$$

so that the, say k columns of $\boldsymbol{S}_1$ represent the identifiable parameters, and the $p-k$ columns of $\boldsymbol{S}_2$ the unidentifiable parameters.

In practice, one wants the columns of $\boldsymbol{S}_1$ to represent an approximate basis for range($\boldsymbol{S}$). A basis satisfies two criteria: Its vectors are linearly independent, and they span the host space.

1. Linear independence of the columns of $\boldsymbol{S}_1 \in \mathbb{R}^{n\times k}$ is quantified by the magnitude of its smallest singular value, which is bounded above by the kth largest singular value of the host matrix,

$$\sigma_k(\boldsymbol{S}_1) \leq \sigma_k. \tag{3.4}$$

   The larger $\sigma_k(\boldsymbol{S}_1)$, the more linearly independent the columns of $\boldsymbol{S}_1$. A more specific statement is presented in (3.9).
2. Spanning the host space range($\boldsymbol{S}$) is quantified by the accuracy of $\boldsymbol{S}_1$ as a low-rank approximation of the host matrix $\boldsymbol{S}$. One measure of accuracy is the residual norm, which is bounded below by the $(k+1)$st singular value of

the host matrix[1]

$$\|(\boldsymbol{I} - \boldsymbol{S}_1\boldsymbol{S}_1^\dagger)\boldsymbol{S}\|_2 = \|(\boldsymbol{I} - \boldsymbol{S}_1\boldsymbol{S}_1^\dagger)\boldsymbol{S}_2\|_2 \geq \sigma_{k+1}. \tag{3.5}$$

The smaller the residual, the better range($\boldsymbol{S}_1$) spans the host space. Criterion (3.5) is a special case of the subsequent (3.10).

Identifiable parameters are the 'most linearly independent' columns' of $\boldsymbol{S}$.
Unidentifiable parameters are the 'most linearly dependent' columns of $\boldsymbol{S}$.

Algorithms 4.1 and 4.3 select unidentifiable parameters, Algorithm 4.2 selects identifiable parameters, while Algorithm 4.4 selects both.

CAUTION. *The separation into linearly dependent and independent columns is highly non-unique. For instance, the matrix*

$$\boldsymbol{S} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

has $\operatorname{rank}(\boldsymbol{S}) = k = 2$ *with* $\sigma_1 = \sigma_2 = \sigma_k = \sqrt{2}$ *and* $\sigma_4 = \sigma_3 = \sigma_{k+1} = 0$. *Moving two linearly independent columns of* $\boldsymbol{S}$ *to the front can be accomplished by any of the following permutation matrices* $\boldsymbol{P}$,

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

to produce the same matrix

$$\boldsymbol{S}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

with residual (3.5) equal to $\sigma_3 = 0$.

One can require the criteria (3.4) or (3.5) to hold either for all $1 \leq k \leq p$ [21], or else for only one specific k [7, 18]. In the latter case, section 3.5 discusses approaches for selecting k.

3.5. Choosing the number k of identifiable parameters. If one knows a bound η on the error or noise in the elements of $\boldsymbol{S}$, one can use criterion (3.5) to designate as small all those singular values below η, in the absolute or the relative sense,

$$\sigma_{k+1} \leq \eta \qquad \text{or} \qquad \sigma_{k+1} \leq \eta\,\sigma_1.$$

[1] The superscript $\boldsymbol{S}_1^\dagger$ denotes the Moore-Penrose inverse, and the equalities follow from the Moore-Penrose property $\boldsymbol{S}_1\boldsymbol{S}_1^\dagger\boldsymbol{S}_1 = \boldsymbol{S}_1$ and the unitary invariance of the two-norm with regard to the permutation $\boldsymbol{P}$.

For instance, if η bounds the relative error in the elements of $\boldsymbol{S}$, then the value of k determined by $\sigma_{k+1} \leq \eta\,\sigma_1$ is called the *numerical rank* of $\boldsymbol{S}$ [15, Definition 2.1], [16, section 5.4.2]. If $\boldsymbol{S}$ is accurate to double precision unit roundoff, then $\eta \approx 1.1 \cdot 10^{-16}$.

Alternatively, one can use criterion (3.4) to designate as large all those singular values exceeding η, in the absolute or the relative sense,

$$\sigma_k > \eta \qquad \text{or} \qquad \sigma_k > \eta\,\sigma_1.$$

If the accuracy of the elements in $\boldsymbol{S}$ is unknown, but its singular values contain a prominent gap, then one can choose k to capture this gap,

$$\sigma_1 \geq \cdots \geq \sigma_k \;\gg\; \sigma_{k+1} \geq \cdots \geq \sigma_p$$

An upgrade [18, Algorithm 5] of Algorithm 4.4 looks for a large gap between adjacent singular values, in order to compute k automatically [18, Remark 1].

The number k of identifiable parameters
can be chosen as the numerical rank of $\boldsymbol{S}$,
or based on a large gap in the singular values.

3.6. Implementing column subset selection with QR decompositions. We show how to compute, by means of pivoted QR decompositions, permutation matrices $\boldsymbol{P}$ that try to optimize criteria (3.4) or (3.5). As a matter of exposition, we introduce plain QR decompositions (section 3.6.1), pivoted QR decompositions (section 3.6.2) and then rank revealing QR decompositions (section 3.6.3).

3.6.1. QR decompositions. Assume that the sensitivity matrix $\boldsymbol{S} \in \mathbb{R}^{n\times p}$ has full column rank with $\operatorname{rank}(\boldsymbol{S}) = p$. A 'thin QR decomposition' [16, section 5.2], [22, Chapter 19] is a basis transformation that transforms the basis for $\operatorname{range}(\boldsymbol{S})$ from linearly independent columns of $\boldsymbol{S}$ to orthonormal columns of $\boldsymbol{Q}$,

$$\boldsymbol{S} = \boldsymbol{Q}\boldsymbol{R}. \tag{3.6}$$

Here $\boldsymbol{Q} \in \mathbb{R}^{n\times p}$ has orthonormal columns with $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}_p$, and the nonsingular upper triangular matrix $\boldsymbol{R} \in \mathbb{R}^{p\times p}$ represents an easy relation between the two bases. Substituting (3.6) into $\boldsymbol{S}$ gives for Fischer information matrix

$$\boldsymbol{F} = \boldsymbol{S}^T\boldsymbol{S} = \boldsymbol{R}^T\boldsymbol{R}.$$

Thus the eigenvalues of $\boldsymbol{F}$ are equal to the squared singular values of the triangular matrix $\boldsymbol{R}$.

3.6.2. Pivoted QR decompositions. These decompositions have more flexibility because they can additionally permute (pivot) the columns of $\boldsymbol{S}$ to compute an orthonormal basis for $\operatorname{range}(\boldsymbol{S})$ [16, 5.4.2], [22, Chapter 19],

$$\boldsymbol{S}\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{R}, \tag{3.7}$$

where $\boldsymbol{P} \in \mathbb{R}^{p\times p}$ is the permutation matrix in (3.3); $\boldsymbol{Q} \in \mathbb{R}^{n\times p}$ has orthonormal columns with $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}_p$; and $\boldsymbol{R} \in \mathbb{R}^{p\times p}$ is upper triangular. Substituting the factorization (3.7) into the sensitivity matrix $\boldsymbol{S}$ gives for Fischer matrix

$$\boldsymbol{F} = \boldsymbol{S}^T\boldsymbol{S} = \boldsymbol{P}\,\boldsymbol{R}^T\boldsymbol{R}\,\boldsymbol{P}^T.$$

Since permutation matrices are orthogonal matrices, the eigenvalues of $\boldsymbol{F}$ are still equal to the squared singular values of $\boldsymbol{R}$, while each eigenvector of $\boldsymbol{R}^T\boldsymbol{R}$ is a permutation of the corresponding eigenvector of $\boldsymbol{F}$.

Algorithms 4.1–4.4 start with a preliminary QR decomposition to reduce the dimension of the matrix. The following remark shows that such a preliminary decomposition is also effective prior to an SVD computation, and the proofs in Appendix A exploit this.

REMARK 3.1. *A preliminary QR decomposition $\boldsymbol{SP} = \boldsymbol{QR}$ is an efficient way to compute the SVD of a dense matrix $\boldsymbol{S} \in \mathbb{R}^{n\times p}$ with $n \geq p$ [6], since it reduces the dimension for the SVD from that of a tall and skinny $n \times p$ matrix down to that of a small square $p \times p$ matrix with the same dimension as the Fischer matrix $\boldsymbol{F} = \boldsymbol{S}^T\boldsymbol{S}$.*

To see this, compute the pivoted QR factorization $\boldsymbol{SP} = \boldsymbol{QR}$, and let the upper triangular $\boldsymbol{R}$ have an SVD

$$\boldsymbol{R} = \boldsymbol{U}_r \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_p \end{bmatrix} \boldsymbol{V}^T,$$

where $\boldsymbol{U}_r$ and $\boldsymbol{V} \in \mathbb{R}^{p\times p}$ are orthogonal matrices. Then the SVD of the permuted sensitivity matrix $\boldsymbol{SP}$ is

$$\boldsymbol{SP} = (\boldsymbol{QU}_r) \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_p \end{bmatrix} \boldsymbol{V}^T,$$

where the left singular vector matrix $\boldsymbol{QU}_r \in \mathbb{R}^{n\times p}$ has orthonormal columns.

This approach retains the asymptotic complexity of an SVD of $\boldsymbol{S}$, but has the advantage of reducing the actual operation count, and, in particular, reducing the problem dimension to that of the Fischer matrix $\boldsymbol{F} = \boldsymbol{S}^T\boldsymbol{S}$.

3.6.3. Rank revealing QR decompositions. These pivoted QR decompositions are designed to 'reveal' the numerical rank of a matrix $\boldsymbol{S}$ that is rank deficient, or ill-conditioned with regard to left inversion [7, section 2], [16, 5.4.2]. [18, section 1.1]. Although there are numerous ways to compute such decompositions [7, 18], most share the same overall strategy.

Assume the sensitivity matrix has numerical rank$(\boldsymbol{S}) \approx k$, where $1 \leq k < p$. Partition the pivoted QR decomposition (3.7) commensurately with the column partitioning (3.3),

$$\underbrace{\begin{bmatrix}\boldsymbol{S}_1 & \boldsymbol{S}_2\end{bmatrix}}_{\boldsymbol{SP}} = \boldsymbol{S}\underbrace{\begin{bmatrix}\boldsymbol{P}_1 & \boldsymbol{P}_2\end{bmatrix}}_{\boldsymbol{P}} = \underbrace{\begin{bmatrix}\boldsymbol{Q}_1 & \boldsymbol{Q}_2\end{bmatrix}}_{\boldsymbol{Q}}\underbrace{\begin{bmatrix}\boldsymbol{R}_{11} & \boldsymbol{R}_{12}\\ \boldsymbol{0} & \boldsymbol{R}_{22}\end{bmatrix}}_{\boldsymbol{R}}, \tag{3.8}$$

with submatrices $\boldsymbol{P}_1 \in \mathbb{R}^{p\times k}$, $\boldsymbol{Q}_1 \in \mathbb{R}^{n\times k}$, $\boldsymbol{R}_{11} \in \mathbb{R}^{k\times k}$, and $\boldsymbol{R}_{22} \in \mathbb{R}^{(p-k)\times(p-k)}$.

Since $\boldsymbol{S}_1 = \boldsymbol{Q}_1\boldsymbol{R}_{11}$, the leading diagonal block $\boldsymbol{R}_{11}$ has the same singular values as the matrix $\boldsymbol{S}_1$ of identifiable parameters

$$\sigma_j(\boldsymbol{S}_1) = \sigma_j(\boldsymbol{R}_{11}), \qquad 1 \leq j \leq k. \tag{3.9}$$

Similarly, since

$$\sigma_j((\boldsymbol{I} - \boldsymbol{S}_1\boldsymbol{S}_1^\dagger)\boldsymbol{S}) = \sigma_j((\boldsymbol{I} - \boldsymbol{S}_1\boldsymbol{S}_1^\dagger)\boldsymbol{S}_2) = \sigma_j(\boldsymbol{R}_{22}), \qquad 1 \leq j \leq p-k, \tag{3.10}$$

the trailing diagonal block $\boldsymbol{R}_{22}$ has the same non-zero singular values as the residuals of the low-rank approximation of range($\boldsymbol{S}$) by range($\boldsymbol{S}_1$).

We call a QR decomposition (3.8) qualitatively 'rank-revealing' if it tries to optimize subselection criteria (3.4) or (3.5), that is,

$$\sigma_k(\boldsymbol{R}_{11}) \approx \sigma_k \qquad \text{or} \qquad \sigma_1(\boldsymbol{R}_{22}) = \|\boldsymbol{R}_{22}\|_2 \approx \sigma_{k+1}.$$

The first criterion tries to produce a well conditioned basis $\boldsymbol{S}_1 = \boldsymbol{Q}_1\boldsymbol{R}_{11}$, and its approximation $\boldsymbol{S}_1\boldsymbol{P}_1^T \approx \boldsymbol{S}$. The second criterion aligns with the popular and robust requirement $\|(\boldsymbol{I} - \boldsymbol{S}_1\boldsymbol{S}_1^\dagger)\boldsymbol{S}\|_2 \approx \sigma_{k+1}$ for low-rank approximations [12].

> Rank-revealing QR decompositions try to select as identifiable parameters those columns of $\boldsymbol{S}$ that are the most linearly independent or that approximate well the unidentifiable parameters.

Rigorous, stringent versions of the subselection criteria (3.4) and (3.5) are presented in [18, Section 1.2] and Theorem 4.4.

4. Identifiability as column subset selection. We express practical identifiability analysis [35, Definition 5.11], [41, page 4 of 21] as column subset selection, to quantify accuracy and to compare the accuracy of different algorithms. We start with Jollife's methods [27, 35]: PCA method B1 (section 4.1), PCA method B4 (section 4.2), and PCA method B3 (section 4.3), and then propose the strong rank-revealing QR factorization [18] as the most accurate option for practical identifiability analysis (section 4.4).

Algorithms 4.1–4.4 input a tall and skinny sensitivity matrix $\boldsymbol{S}$ and the number k of identifiable parameters, say from section 3.5; and output the factors of a pivoted QR decomposition $\boldsymbol{SP} = \boldsymbol{QR}$.

> We recommend Algorithm 4.4 in section 4.4.
> It has the most rigorous and realistic accuracy guarantees for both, identifiable and unidentifiable parameters.

The algorithms are formulated with a focus on understanding, rather than efficiency.

4.1. PCA method B1. This method [27, section 2.2], [35, (5.13)] selects unidentifiable parameters, by detecting large-magnitude components in the eigenvectors $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_p$ corresponding to the $p-k$ smallest eigenvalues of the Fischer matrix $\boldsymbol{F}$, starting from the smallest eigenvalue.

Method B1 starts with a unit-norm eigenvector $\mathbf{v}_p$ corresponding to λ_p, picks a magnitude largest element in $\mathbf{v}_p$,

$$|v_{m_1,p}| = \max_{1\le j\le p} |v_{jp}|,$$

and designates the parameter with index m_1 as unidentifiable. Method B1 repeats this on eigenvectors corresponding to eigenvalues $\lambda_{p-1} \le \cdots \le \lambda_{k+1}$ in that order, by selecting magnitude-largest elements that have not been selected previously,

$$|v_{m_\ell,\ell}| = \max_{\substack{1\le j\le p \\ j\neq m_1,\ldots,m_{p-\ell+1}}} |v_{j\ell}|, \qquad \ell = p-1, \ldots, k+1,$$

and declares the parameters with indices $m_1, \ldots m_{p-k}$ as unidentifiable.

Expressing PCA method B1 as column subset selection. PCA method B1 is almost identical to the subset selection algorithm in [5, Section 3], which is also [7, Algorithm Chan-II], and is related to the algorithms in [13, 17].

Algorithm 4.1, which represents [5, Algorithm RRQR(r)], selects $p-k$ unidentifiable parameters $\boldsymbol{S}_2$ to optimize subset selection criterion (3.5) and moves them to the back of the matrix. Once a column for $\boldsymbol{S}_2$ has been identified, Algorithm 4.1 ignores it from then on, and continues on a lower-dimensional submatrix.

Algorithm 4.1 Column subset selection version of PCA B1

Input: $\boldsymbol{S} \in \mathbb{R}^{n\times p}$ with $n \geq p$, $1 \leq k < p$

Set $\boldsymbol{P} = \boldsymbol{I}_p$

Compute decomposition (3.7): $\boldsymbol{SP} = \boldsymbol{QR}$ {Unpivoted QR of $\boldsymbol{S}$}

for $\ell = p : k+1$

{If $\ell = p$, then $\boldsymbol{R}_{11} = \boldsymbol{R}$ }

Partition $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22} \end{bmatrix}$ where $\boldsymbol{R}_{11} \in \mathbb{R}^{\ell\times\ell}$ {Focus on leading $\ell \times \ell$ block}

Compute right singular vector $\mathbf{v} \in \mathbb{R}^{\ell}$ of $\boldsymbol{R}_{11}$ corresponding to $\sigma_\ell(\boldsymbol{R}_{11})$

Compute permutation $\widetilde{\boldsymbol{P}} \in \mathbb{R}^{\ell\times\ell}$ so that $|(\widetilde{\boldsymbol{P}}^T\mathbf{v})_\ell| = \|\mathbf{v}\|_\infty$

{Move magnitude-largest element of $\mathbf{v}$ to bottom}

Compute QR decomposition (3.6): $\boldsymbol{R}_{11}\widetilde{\boldsymbol{P}} = \widetilde{\boldsymbol{Q}}\widetilde{\boldsymbol{R}}_{11}$ {Unpivoted QR of $\boldsymbol{R}_{11}\widetilde{\boldsymbol{P}}$}

Update $\boldsymbol{Q} := \boldsymbol{Q}\begin{bmatrix} \widetilde{\boldsymbol{Q}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{p-\ell} \end{bmatrix}$, $\boldsymbol{P} := \boldsymbol{P}\begin{bmatrix} \widetilde{\boldsymbol{P}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{p-\ell} \end{bmatrix}$, $\boldsymbol{R} := \begin{bmatrix} \widetilde{\boldsymbol{R}}_{11} & \widetilde{\boldsymbol{Q}}^T\boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22} \end{bmatrix}$

end for

return $\boldsymbol{P}$, $\boldsymbol{Q}$, $\boldsymbol{R}$

Theorem 4.1 shows that the unidentifiable parameters $\boldsymbol{S}_2$ from Algorithm 4.1 can be interpreted as column subsets satisfying criterion (3.5).

THEOREM 4.1. *Let $\boldsymbol{S} \in \mathbb{R}^{n\times p}$ with $n \geq p$ be the sensitivity matrix, and $1 \leq k < p$. Algorithm 4.1 computes a pivoted QR decomposition*

$$\boldsymbol{SP} = \begin{bmatrix} \boldsymbol{S}_1 & \boldsymbol{S}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{Q}_1 & \boldsymbol{Q}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22} \end{bmatrix}, \qquad \boldsymbol{R}_{22} \in \mathbb{R}^{(p-k)\times(p-k)},$$

where

$$\sigma_{k+1} \leq \|(\boldsymbol{I} - \boldsymbol{S}_1\boldsymbol{S}_1^\dagger)\boldsymbol{S}_2\|_2 = \|\boldsymbol{R}_{22}\|_2 \leq 2^{p-k-1}\sigma_{k+1}.$$

If the mumerical $\operatorname{rank}(\boldsymbol{S}) = k$*, then the columns of $\boldsymbol{S}_2$ represent the $p-k$ unidentifiable parameters.*

Proof. The equality follows from (3.10), while the lower bound follows from interlacing (A.1). The upper bound is derived in section A.1, and in particular in Lemma A.3. ∎

Theorem 4.1 bounds the residual in the low rank approximation $\boldsymbol{S}_1$ according to criterion (3.5). Like many subset selection bounds, the upper bound can be achieved by artificially contrived matrices [22, section 8.3], but tends to be quantitatively pessimistic in practice. Fortunately, it is informative from a qualitative perspective.

4.2. PCA method B4. This method [27, section 2.2], [35, (5.15)], [41, Appendix C] selects identifiable parameters, by detecting large-magnitude components in the eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ corresponding to the k largest eigenvalues of the Fischer matrix, starting from the largest eigenvalue. Our detailed interpretation of the algorithm follows that in [41, Appendix C, Third Criterion].

Method B4 starts with a unit-norm eigenvector $\mathbf{v}_1$ corresponding to λ_1, picks a magnitude largest element in $\mathbf{v}_1$,

$$|v_{m_1,1}| = \max_{1 \le j \le p} |v_{j1}|,$$

and declares the parameter with index m_1 as identifiable. Method B4 repeats this on eigenvectors corresponding to eigenvalues $\lambda_2 \ge \cdots \ge \lambda_k$ in that order, by selecting magnitude-largest elements that have not been selected previously,

$$|v_{m_\ell,\ell}| = \max_{\substack{1 \le j \le p \\ j \ne m_1,\ldots,m_\ell}} |v_{j\ell}|, \qquad \ell = 2, \ldots, k,$$

and declares the parameters with indices $m_1, \ldots, m_k$ as identifiable.

Expressing PCA method B4 as column subset selection. PCA method B4 is almost identical to the subset selection algorithm in [6, Section 3], which is also [7, Algorithm Chan-I].

Algorithm 4.2, which represents [6, Algorithm L-RRQR], selects k identifiable parameters $\boldsymbol{S}_1$ to optimize subset selection criterion (3.4) and moves them to the front of the matrix. Once a column for $\boldsymbol{S}_1$ has been identified, Algorithm 4.2 ignores it from then on, and continues on a lower-dimensional submatrix.

Algorithm 4.2 Column subset selection version of PCA B4

Input: $\boldsymbol{S} \in \mathbb{R}^{n \times p}$ with $n \ge p$, $1 \le k < p$

Set $\boldsymbol{P} = \boldsymbol{I}_p$

Compute decomposition (3.7): $\boldsymbol{S}\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{R}$ {Unpivoted QR of $\boldsymbol{S}$}

for $\ell = 1 : k$

{If $\ell = 1$, then $\boldsymbol{R}_{22} = \boldsymbol{R}$ }

Partition $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22} \end{bmatrix}$ where $\boldsymbol{R}_{22} \in \mathbb{R}^{(p-\ell+1)\times(p-\ell+1)}$

{Focus on trailing $(p - \ell + 1) \times (p - \ell + 1)$ block}

Compute right singular vector $\mathbf{v} \in \mathbb{R}^{p-\ell+1}$ of $\boldsymbol{R}_{22}$ corresponding to $\sigma_1(\boldsymbol{R}_{22})$

Compute permutation $\widetilde{\boldsymbol{P}} \in \mathbb{R}^{(p-\ell+1)\times(p-\ell+1)}$ so that $|(\widetilde{\boldsymbol{P}}^T \mathbf{v})_1| = \|\mathbf{v}\|_\infty$

{Move magnitude-largest element of $\mathbf{v}$ to top}

Compute QR decomposition (3.6): $\boldsymbol{R}_{22}\widetilde{\boldsymbol{P}} = \widetilde{\boldsymbol{Q}}\widetilde{\boldsymbol{R}}_{22}$ {Unpivoted QR of $\boldsymbol{R}_{22}\widetilde{\boldsymbol{P}}$}

Update $\boldsymbol{Q} := \boldsymbol{Q}\begin{bmatrix} \boldsymbol{I}_{\ell-1} & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{Q}} \end{bmatrix}$, $\boldsymbol{P} := \boldsymbol{P}\begin{bmatrix} \boldsymbol{I}_{\ell-1} & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{P}} \end{bmatrix}$, $\boldsymbol{R} := \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \widetilde{\boldsymbol{R}}_{22} \end{bmatrix}$

end for

return $\boldsymbol{P}$, $\boldsymbol{Q}$, $\boldsymbol{R}$

Theorem 4.2 shows that the identifiable parameters $\boldsymbol{S}_1$ from Algorithm 4.2 can be interpreted as parameters that satisfy criterion (3.4).

THEOREM 4.2. *Let $\boldsymbol{S} \in \mathbb{R}^{n\times p}$ with $n \geq p$ be the sensitivity matrix, and $1 \leq k < p$. Then Algorithm 4.2 computes a QR decomposition*

$$\boldsymbol{SP} = \begin{bmatrix}\boldsymbol{S}_1 & \boldsymbol{S}_2\end{bmatrix} = \begin{bmatrix}\boldsymbol{Q}_1 & \boldsymbol{Q}_2\end{bmatrix} \begin{bmatrix}\boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22}\end{bmatrix}, \qquad \boldsymbol{R}_{11} \in \mathbb{R}^{k\times k},$$

where

$$2^{-k+1}\sigma_k \leq \sigma_k(\boldsymbol{R}_{11}) = \sigma_k(\boldsymbol{S}_1) \leq \sigma_k.$$

If numerical $\operatorname{rank}(\boldsymbol{S}) = k$*, then the columns of $\boldsymbol{S}_1$ represent the k identifiable parameters.*

Proof. The equality follows from (3.9), while the upper bound follows from interlacing (A.1). The lower bound is derived in section A.2, and in particular in Lemma A.6. ∎

Theorem 4.2 bounds the linear independence of the columns in $\boldsymbol{S}_1$ according to criterion (3.4). As before, the lower bound in Theorem 4.2 can be quantitatively very pessimistic in practice, but tends to be qualitatively informative.

4.3. PCA method B3. This method [27, section 2.2], [35, (5.14)] selects unidentifiable parameters by detecting large squared row sums in the matrix $\boldsymbol{V}_{k+1:p} \equiv \begin{bmatrix}\mathbf{v}_{k+1} & \cdots & \mathbf{v}_p\end{bmatrix}$ of eigenvectors corresponding to the $p-k$ smallest eigenvalues of the Fischer matrix.

The squared row norms of $\boldsymbol{V}_{k+1:p}$,

$$\omega_j \equiv \left\| \begin{bmatrix} v_{j,k+1} & \cdots & v_{jp}\end{bmatrix} \right\|_2^2 = \sum_{\ell=k+1}^{p} v_{j\ell}^2, \qquad 1 \leq j \leq p,$$

are called 'leverage scores' in the statistics literature [8, 23, 50]. The largest leverage score

$$\max_{1\leq j\leq p} \omega_j = \max_{1\leq j\leq p} \sum_{\ell=k+1}^{p} v_{j\ell}^2$$

is called 'coherence' in the compressed sensing literature [11] and reflects the difficulty of sampling rows from $\boldsymbol{V}_{k+1:p}$.

Method B3 picks a largest leverage score from $\boldsymbol{V}_{k+1:p}$,

$$\omega_{m_1} = \max_{1\leq j\leq p} \omega_j = \max_{1\leq j\leq p} \sum_{\ell=k+1}^{p} v_{j\ell}^2$$

and declares the parameter with index m_1 as unidentifiable. Method B3 repeats this on the remaining rows of $\boldsymbol{V}_{k+1:p}$, by selecting parameters that have not been selected previously,

$$\omega_{m_\ell} = \max_{\substack{k+1\leq j\leq p \\ j\neq m_1,\ldots,m_{p-\ell+1}}} \omega_j, \qquad \ell = p-1,\ldots,k+1,$$

and declares the parameters with index $m_1,\ldots,m_{p-k}$ as unidentifiable.

Expressing PCA method B3 as column subset selection. PCA method B3 can be interpreted in two ways: Either as selecting parameters according to the largest leverage scores of the subdominant eigenvector matrix $\boldsymbol{V}_{k+1:p}$ of the Fischer matrix [8, 23, 50]; or else as selecting parameters based on column subset selection with [7, Algorithm GKS-II]. We choose the latter interpretation.

Algorithm 4.3 may look different from PCA method B3 but accomplishes the same thing in an easier manner (in exact arithmetic). The algorithm in [15, Section 6], which is also [7, Algorithm GKS-I], operates instead the dominant right singular vectors, and applies the column subset selection method [4, Section 4], [16, section 5.4.2], which is also [7, Algorithm Golub-I].

The idea is the following: partition the SVD of the triangular matrix $\boldsymbol{R} = \boldsymbol{U}_r \boldsymbol{\Sigma} \boldsymbol{V}^T$ in Remark 3.1,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix}, \qquad \boldsymbol{U}_r = \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix}, \qquad \boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_2 \end{bmatrix},$$

where $\boldsymbol{\Sigma}_1 = \operatorname{diag}\begin{pmatrix} \sigma_1 & \cdots & \sigma_k \end{pmatrix} \in \mathbb{R}^{k \times k}$ contains the k dominant singular values of $\boldsymbol{R}$, hence $\boldsymbol{S}$; and $\boldsymbol{U}_1 \in \mathbb{R}^{p \times k}$ and $\boldsymbol{V}_1 \in \mathbb{R}^{p \times k}$ are the k associated left and right singular vectors, respectively. Applying a permutation to $\boldsymbol{V}_1^T$ corresponds to applying a permutation to $\boldsymbol{R}$, hence $\boldsymbol{S}$. In Algorithm 4.3, column j of $\boldsymbol{W}$ is denoted by $\boldsymbol{W}\boldsymbol{e}_j$.

Theorem 4.3 quantifies how well the identifiable parameters $\boldsymbol{S}_1$ from Algorithm 4.3 satisfy criterion (3.4).

THEOREM 4.3. *Let $\boldsymbol{S} \in \mathbb{R}^{n \times p}$ with $n \geq p$ be the sensitivity matrix, and $1 \leq k < p$. Then Algorithm 4.3 computes a QR decomposition*

$$\boldsymbol{S}\boldsymbol{P} = \begin{bmatrix} \boldsymbol{S}_1 & \boldsymbol{S}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{Q}_1 & \boldsymbol{Q}_2 \end{bmatrix}}_{\boldsymbol{Q}} \underbrace{\begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22} \end{bmatrix}}_{\boldsymbol{R}}, \qquad \boldsymbol{R}_{11} \in \mathbb{R}^{k \times k},$$

where

$$\sigma_k / \|\boldsymbol{V}_{11}^{-1}\|_2 \leq \sigma_k(\boldsymbol{R}_{11}) \leq \sigma_k,$$
$$\sigma_{k+1} \leq \sigma_1(\boldsymbol{R}_{22}) \leq \|\boldsymbol{V}_{11}^{-1}\|_2 \, \sigma_{k+1}$$

and $\boldsymbol{V}_{11} \in \mathbb{R}^{k \times k}$ is the leading principal submatrix of $\boldsymbol{V}$ in the SVD $\boldsymbol{R} = \boldsymbol{U}_r \boldsymbol{\Sigma} \boldsymbol{V}^T$.

If Algorithm 4.3 applies Algorithm 4.2 to $\boldsymbol{V}_1^T$, then

$$\|\boldsymbol{V}_{11}^{-1}\|_2 \leq 2^{k-1}$$

If numerical $\operatorname{rank}(\boldsymbol{S}) = k$*, then the columns of $\boldsymbol{S}_1$ represent the k identifiable parameters.*

Proof. The upper bound for $\boldsymbol{R}_{11}$ and the lower bound for $\boldsymbol{R}_{22}$ follow from interlacing (A.1). The remaining two bounds are derived in section A.3.

The bound for $\|\boldsymbol{V}_{11}^{-1}\|_2 = 1/\sigma_k(\boldsymbol{V}_{11})$ follows by applying Theorem 4.2 to $\boldsymbol{V}_1^T$ and remembering that all singular values of $\boldsymbol{V}_1$ are equal to 1. ∎

4.4. Strong rank-revealing QR decompositions. The final method [18, section 4] selects identifiable parameters by trying to maximize the volume of $\boldsymbol{S}_1$ via pairwise column permutations.

Algorithm 4.3 Column subset selection version of PCA B3

Input: $\boldsymbol{S} \in \mathbb{R}^{n\times p}$, $n \geq p$, $1 \leq k < p$

Set $\boldsymbol{P} = \boldsymbol{I}_p$

Compute decomposition (3.7): $\boldsymbol{S} = \boldsymbol{Q}\boldsymbol{R}$ {Unpivoted QR of $\boldsymbol{S}$}

for $\ell = 1 : k$

{If $\ell = 1$, then $\boldsymbol{R}_{22} = \boldsymbol{R}$ }

Partition $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22} \end{bmatrix}$ where $\boldsymbol{R}_{22} \in \mathbb{R}^{(p-\ell+1)\times(p-\ell+1)}$

{Focus on trailing $(p-\ell+1)\times(p-\ell+1)$ block}

Compute $k-\ell+1$ right singular vectors $\boldsymbol{V}_1 \in \mathbb{R}^{(p-\ell+1)\times(k-\ell+1)}$ of $\boldsymbol{R}_{22}$ corresponding to $\sigma_1 \geq \cdots \geq \sigma_{k-\ell+1}$

Set $\boldsymbol{W} = \boldsymbol{V}_1^T \in \mathbb{R}^{(k-\ell+1)\times(p-\ell+1)}$

Compute permutation $\widetilde{\boldsymbol{P}} \in \mathbb{R}^{(p-\ell+1)\times(p-\ell+1)}$ so that

$$\|\boldsymbol{W}(\widetilde{\boldsymbol{P}}\boldsymbol{e}_1)\|_2 = \max_{1\leq j\leq p-\ell+1} \|\boldsymbol{W}\boldsymbol{e}_j\|_2$$

{Move column of $\boldsymbol{W}$ with largest norm to front}

Compute QR decomposition (3.6): $\boldsymbol{R}_{22}\widetilde{\boldsymbol{P}} = \widetilde{\boldsymbol{Q}}\widetilde{\boldsymbol{R}}_{22}$ {Unpivoted QR of $\boldsymbol{R}_{22}\widetilde{\boldsymbol{P}}$}

Update $\boldsymbol{Q} := \boldsymbol{Q}\begin{bmatrix} \boldsymbol{I}_{\ell-1} & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{Q}} \end{bmatrix}$, $\boldsymbol{P} := \boldsymbol{P}\begin{bmatrix} \boldsymbol{I}_{\ell-1} & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{P}} \end{bmatrix}$, $\boldsymbol{R} := \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \widetilde{\boldsymbol{R}}_{22} \end{bmatrix}$

end for

return $\boldsymbol{P}$, $\boldsymbol{Q}$, $\boldsymbol{R}$

A 'strong rank-revealing' QR decomposition tries to optimize both subset selection criteria (3.4) and (3.5) and bounds every element of $|\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}|$. The component-wise boundedness ensures that the columns of

$$\boldsymbol{P}\begin{bmatrix} -\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12} \\ \boldsymbol{I}_{p-k} \end{bmatrix}$$

represents an approximate basis for the null space of $\boldsymbol{S}$, provided $\boldsymbol{R}_{11}$ is not too ill-conditioned [18, section 1.2]. A rigorous definition of the strong rank-revealing QR decomposition is presented in [18, Section 1.2] and Theorem 4.4 below.

Algorithm 4.4, which represents [18, Algorithm 4], exchanges a column of $\boldsymbol{S}_1$ with a column of $\boldsymbol{S}_2$ until $\det(\boldsymbol{S}_1^T\boldsymbol{S}_1) = \det(\boldsymbol{R}_{11})^2$ stops increasing. More specifically [18, Lemma 3.1], after permuting columns i and $k+j$ of $\boldsymbol{R}$ with a permutation matrix $\boldsymbol{P}^{(ij)}$, and performing an unpivoted QR decomposition $\boldsymbol{S}\boldsymbol{P}^{(ij)} = \widetilde{\boldsymbol{Q}}\widetilde{\boldsymbol{R}}$, we compare the determinant of the leading principal submatrix $\widetilde{\boldsymbol{R}}_{11} \in \mathbb{R}^{k\times k}$ of $\widetilde{\boldsymbol{R}}$ with that of the original submatrix $\boldsymbol{R}_{11}$,

$$\rho_{ij} \equiv \frac{\det(\widetilde{\boldsymbol{R}}_{11})}{\det(\boldsymbol{R}_{11})} = \sqrt{(\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12})_{ij}^2 + \left(\|\boldsymbol{R}_{22}\boldsymbol{e}_j\|_2\, \|\boldsymbol{e}_i^T\boldsymbol{R}_{11}^{-1}\|_2\right)^2}. \tag{4.1}$$

Given a user-specified tolerance $f > 1$, Algorithm 4.4 iterates as long as it can find columns i and $j+k$ with $\rho_{ij} > f$ and, by permuting columns i and $j+k$. increase the determinant to $\det(\widetilde{\boldsymbol{R}}_{11}) \geq f\det(\boldsymbol{R}_{11})$. The correctness of Algorithm 4.4 follows from Lemma A.8.

Theorem 4.4 shows that the columns $\boldsymbol{S}_1$ from Algorithm 4.4 can be interpreted as identifiable parameters that satisfy even stronger conditions than criteria (3.4) and (3.5) combined.

THEOREM 4.4. *Let $\boldsymbol{S} \in \mathbb{R}^{n\times p}$ with $n \geq p$ be the sensitivity matrix and $1 \leq k < p$. Algorithm 4.4 with input $f \geq 1$ computes a QR decomposition*

$$\boldsymbol{S}\boldsymbol{P} = \begin{bmatrix}\boldsymbol{S}_1 & \boldsymbol{S}_2\end{bmatrix} = \begin{bmatrix}\boldsymbol{Q}_1 & \boldsymbol{Q}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{R}_{11} & \boldsymbol{R}_{12}\\ \boldsymbol{0} & \boldsymbol{R}_{22}\end{bmatrix},$$

where $\boldsymbol{R}_{11} \in \mathbb{R}^{k\times k}$ and $\boldsymbol{R}_{22} \in \mathbb{R}^{(p-k)\times(p-k)}$ satisfy

$$\sigma_i(\boldsymbol{R}_{11}) \geq \frac{\sigma_i}{\sqrt{1+f^2k(p-k)}}, \qquad 1 \leq i \leq k$$
$$\sigma_j(\boldsymbol{R}_{22}) \leq \sigma_{j+k}\sqrt{1+f^2k(p-k)}, \qquad 1 \leq j \leq p-k,$$

and

$$|\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}|_{ij} \leq f, \qquad 1 \leq i \leq k,\ 1 \leq j \leq p-k.$$

If numerical $\operatorname{rank}(\boldsymbol{S}) = k$*, then the columns of $\boldsymbol{S}_1$ represent the k identifiable parameters, and the columns of $\boldsymbol{S}_2$ the unidentifiable parameters.*

Proof. This follows from [18, Lemma 3.1 and Theorem 3.2]. See section A.4, and in particular in Lemma A.9. ∎

Algorithm 4.4 Column subset selection with strong rank-revealing QR (srrqr)

Input: Sensitivity matrix $\boldsymbol{S} \in \mathbb{R}^{n\times p}$, $n \geq p$, $1 \leq k < p$, $f \geq 1$
  Compute $\boldsymbol{S}\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{R}$ {Pivoted QR to make $\boldsymbol{R}_{11}$ nonsingular}
  Compute ρ_{ij} as defined in (4.1), $1 \leq i \leq k$, $1 \leq j \leq k-p$
  while $\max_{1\leq i\leq k,\ 1\leq j\leq p-k}\{\rho_{ij}\} > f$
    Find some $1 \leq i \leq k$ and $1 \leq j \leq k-p$ with $\rho_{ij} > f$
    Compute permutation $\boldsymbol{P}^{(ij)}$ to permute columns i and $j+k$
    Decomposition (3.6) $\boldsymbol{R}\boldsymbol{P}^{(ij)} = \widetilde{\boldsymbol{Q}}\widetilde{\boldsymbol{R}}$ {Unpivoted QR of $\boldsymbol{R}\boldsymbol{P}^{(ij)}$}
    Update $\boldsymbol{P} := \boldsymbol{P}\boldsymbol{P}^{(ij)}$, $\boldsymbol{Q} := \boldsymbol{Q}\widetilde{\boldsymbol{Q}}$, $\boldsymbol{R} := \widetilde{\boldsymbol{R}}$
    Update ρ_{ij}
  end while
  return $\boldsymbol{P}$, $\boldsymbol{Q}$, $\boldsymbol{R}$

5. Applications. We compare the accuracy of the four Algorithms 4.1–4.4 on the sensitivity matrices from physical applications (section 5.1) and on the synthetic matrices from classical column pivoting 'counterexamples' (section 5.2).

Numerical experiments were performed in MATLAB 2021b on a 16 GB MacBook Pro with an M1 chip. We compute relative versions of the subset selection criteria (3.4) and (3.5),

$$\gamma_1 \equiv \frac{\sigma_k(\boldsymbol{S}_1)}{\sigma_k(\boldsymbol{S})}, \tag{5.1}$$

and

$$\gamma_2 \equiv \frac{\|(\boldsymbol{I} - \boldsymbol{S}_1\boldsymbol{S}_1^{\dagger})\boldsymbol{S}_2\|_2}{\sigma_{k+1}(\boldsymbol{S})}. \tag{5.2}$$

The closer γ_1 and γ_2 are to 1, the more accurate the algorithm. We also compute the improvement in condition number of the selected columns,

$$\tau \equiv \frac{\text{cond}(\boldsymbol{S}_1)}{\text{cond}(\boldsymbol{S})} \tag{5.3}$$

The lower τ_1, the better the conditioning of the selected columns.

5.1. Sensitivity matrices from physical models. We apply Algorithms 4.1–4.4 to the sensitivity matrices from the mathematical models in sections 2.2 and B.

The sensitivity matrices $\boldsymbol{S}$ are evaluated at given nominal parameter values. For the epidemiological models (SVIR, SEVIR, COVID) in particular, $\boldsymbol{S}$ is evaluated at the nominal values in Table B.1, and additionally at 10,000 points sampled uniformly within 50% of the nominal value.

Table 5.1. For each model, Algorithms 4.1–4.4 produce the same identifiable parameters, that is, the same column subsets and the same identical values for the subset selection criteria γ_1 in (5.1) and γ_2 in (5.1). The consistent accuracy illustrates the robustness of column subset selection for identifiability analysis in applications, particularly since each sensitivity matrix originates from a different type of mechanistic model.

Model	n	p	k	τ	γ_1	γ_2
SVIR	31	4	3	1.6e-03	1.0	1.0
SEVIR	31	5	4	1.2e-02	1.0	1.0
COVID	31	8	5	1.5e-03	0.9	1.1
HGO	14	8	5	4.0e-04	1.0	1.0
Wound	46	11	6	2.2e-08	0.9	1.2
Neuro	200	175	14	9.8e-23	0.6	1.7

TABLE 5.1

Identical accuracy of Algorithms 4.1–4.4 on the models in section 2.2. Here p = number of parameters and number of columns of $\boldsymbol{S}$; n = number of observations and number of rows of $\boldsymbol{S}$; k= numerical rank of $\boldsymbol{S}$ and number of identifiable parameters; τ= ratio of condition numbers in (5.3); and γ_1 and γ_2 are the subset selection criteria in (5.1), and in (5.2), respectively.

When applied to the physical models,
Algorithms 4.1–4.4 exhibit similar accuracy and reliability.
We recommend Algorithm 4.4 because, in theory,
it has the most stringent accuracy guarantees.

5.2. Synthetic adversarial matrices. We apply Algorithms 4.1–4.4 to synthetic adversarial matrices designed to thwart the accuracy of subset selection algorithms. Although synthetic, these matrices still represent sensitivity matrices for specific dynamical systems (Appendix C). Each algorithm is applied to 10,000 realizations of each of the following matrices.

- *Kahan* [28]: $\boldsymbol{S} = \boldsymbol{D}_n \boldsymbol{K}_n \in \mathbb{R}^{n\times n}$, where

$$\boldsymbol{D}_n \equiv \text{diag}\begin{pmatrix} 1 & \zeta & \zeta^2 & \cdots & \zeta^{n-1} \end{pmatrix}, \qquad \boldsymbol{K}_n \equiv \begin{pmatrix} 1 & -\varphi & -\varphi & \cdots & -\varphi \\ & 1 & -\varphi & \cdots & -\varphi \\ & & \ddots & \ddots & \vdots \\ & & & 1 & -\varphi \\ & & & & 1 \end{pmatrix},$$

with $\zeta^2 + \varphi^2 = 1$ for $\zeta, \varphi > 0$, and $k = n - 1$.
We choose $n = 100$, and sample ζ uniformly from $[0.9, 0.99999]$. The average condition number over 10,000 realizations is $\text{cond}(\boldsymbol{S}) \approx 2.4 \cdot 10^{19}$.

- *Gu-Eisenstat* [18, Example 2]:

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{D}_{n-3}\boldsymbol{K}_{n-3} & \mathbf{0} & \mathbf{0} & -\varphi \boldsymbol{D}_{n-3}\mathbb{1}_{n-3} \\ & \mu & 0 & 0 \\ & & \mu & 0 \\ & & & \mu \end{pmatrix} \in \mathbb{R}^{n \times n},$$

where $k = n - 2$, and

$$\mu \equiv \frac{1}{\sqrt{k}} \min_{1 \le i \le n-3} \|\boldsymbol{e}_i^T (\boldsymbol{D}_{n-3}\boldsymbol{K}_{n-3})^{-1}\|_2^{-1}.$$

We choose $n = 100$, and sample ζ uniformly from $[0.9, 0.99999]$. The average condition number over 10,000 realizations is $\text{cond}(\boldsymbol{S}) \approx 2.0 \cdot 10^{34}$.

- *Jolliffe* [27, Appendix A1]: $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, where $\boldsymbol{U} \in \mathbb{R}^{n \times p}$ has orthonormal columns with Haar measure [47]; $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$ is diagonal; and $\boldsymbol{V}$ is the orthonormal factor from the QR factorization of

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_1 & & & \\ & \boldsymbol{\Lambda}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{\Lambda}_k \end{pmatrix}, \qquad \boldsymbol{\Lambda}_i = \begin{pmatrix} 1 & \rho_i & \cdots & \rho_i \\ \rho_i & 1 & \cdots & \rho_i \\ \vdots & \vdots & \ddots & \vdots \\ \rho_i & \rho_i & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{p_i \times p_i},$$

where $\rho_i \approx 1$ and $p = \sum_{i=1}^{k} p_i$.
We choose $n = 200$, $p = 100$, $p_i = 5$, and $k = 20$; and sample the leading k diagonal elements of $\boldsymbol{\Sigma}$ uniformly from $[10^2, 10^3]$, the $p-k$ trailing diagonal elements of $\boldsymbol{\Sigma}$ uniformly from $[10^{-10}, 10^{1.9}]$, and ρ_i uniformly from $[0.9, 0.99999]$. The average condition number over 10,000 realizations is $\text{cond}(\boldsymbol{S}) \approx 4.8 \cdot 10^{14}$.

- *Sorensen-Embree* [46]: $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, where $\boldsymbol{U} \in \mathbb{R}^{n \times p}$ has Haar measure with orthonormal columns; $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$ is diagonal; and $\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_k & \boldsymbol{V}_{p-k} \end{pmatrix} \in \mathbb{R}^{p \times p}$ is an orthogonal matrix, and $\boldsymbol{V}_k \in \mathbb{R}^{p \times k}$ is the orthonormal factor from the QR factorization of

$$\mathbf{L} = \begin{pmatrix} 1 & & & \\ -1 & 1 & & \\ \vdots & \ddots & \ddots & \\ -1 & \cdots & -1 & 1 \\ -1 & \cdots & -1 & -1 \\ \vdots & & \vdots & \vdots \\ -1 & \cdots & -1 & -1 \end{pmatrix} \in \mathbb{R}^{p \times k}.$$

We choose $n = 200$, $p = 100$, and $k = 20$; and sample the leading k diagonal elements of $\boldsymbol{\Sigma}$ uniformly from $[10^2, 10^3]$, the $p - k$ trailing ones uniformly from $[10^{-10}, 10^{1.9}]$. The average condition number over 10,000 realizations is $\text{cond}(\boldsymbol{S}) \approx 1.4 \cdot 10^{14}$.

- *SHIPS*: We constructed this matrix to amplify differences in the accuracy of Algorithms 4.1–4.4. Here $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, where $\boldsymbol{U}$ and $\boldsymbol{\Sigma}$ as for Joliffe, and $\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_k & \boldsymbol{V}_{p-k} \end{pmatrix} \in \mathbb{R}^{p\times p}$ is an orthogonal matrix with

$$\boldsymbol{V}_k = \begin{pmatrix} \boldsymbol{V}_{11} \\ \tilde{\boldsymbol{U}}(\boldsymbol{I} - \boldsymbol{V}_{11}\boldsymbol{V}_{11})^{1/2} \end{pmatrix} \in \mathbb{R}^{p\times k}$$

  where $\tilde{\boldsymbol{U}} \in \mathbb{R}^{(p-k)\times k}$ has orthonormal columns with Haar measure [47], and

$$\boldsymbol{V}_{11} = \frac{\mathbf{T}}{2\|\mathbf{T}\|_2} \in \mathbb{R}^{k\times k}, \qquad \boldsymbol{T} = \begin{pmatrix} 1 & -1 & \cdots & -1 \\ & 1 & \cdots & -1 \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in \mathbb{R}^{k\times k}.$$

  We choose $n = 200$, $p = 100$, and $k = 20$. The leading k diagonal elements of $\boldsymbol{\Sigma}$ are logarithmically spaced in $[10^2, 10^3]$, and the $p - k$ trailing ones logarithmically spaced in $[10^{-10}, 10^{1.9}]$. The average condition number over 10,000 realizations is $\text{cond}(\boldsymbol{S}) \approx 1.0 \cdot 10^{13}$.

In Algorithm 4.4, we set $f = \sqrt{2}$ for the *Gu-Eisenstat* matrix, and $f = 1$ for all other matrices.

Table 5.2. It displays the average of the condition number ratio (5.3), and subset selection criteria (5.1) and (5.2) for 10,000 realizations of each synthetic matrix.

Algorithm 4.2 produces the smallest values of τ and γ_1, that is, the worst conditioned columns $\boldsymbol{S}_1$, for the *Kahan* and *Gu-Eisenstat* matrices.

Algorithm 4.3 produces the smallest values of γ_2, that is, the best low-rank approximation $\boldsymbol{S}_1$. Algorithms 4.1 and 4.4 are close with only slightly larger γ_2 on all matrices except for the *Sorensen-Embree* matrix, where their γ_2 is more than 5 times larger than that of Algorithm 4.3.

Algorithms 4.1 and 4.4 produce better conditioned $\boldsymbol{S}_1$ than Algorithm 4.3, most notably for the *Sorensen-Embree* and *SHIPS* matrices.

The *Jolliffe* matrix was constructed to thwart Algorithm 4.3 [27, Appendix A1], and there is slight evidence of its loss of accuracy with these matrices. While all of the algorithms performed nearly identically, the absolute version of criterion (3.4) for Algorithm 4.3 (to more digits than could be represented in Table 5.2) is 1.8e-14, compared to 1.9e-14 for Algorithms 4.1, 4.2, and 4.4.

Figure 5.1. The box plots illustrate the accuracy of Algorithms 4.1–4.4 on 10,000 realizations of our *SHIPS* matrix. The top and bottom of each box represent the first and third quartiles, respectively, while the red line through the box itself is the average. Values below and above the short black horizontal lines are outliers, and the horizontal lines themselves show the minimum and maximum excluding the outliers.

We constructed the SHIPS matrix to force differences in the accuracy of Algorithms 4.1–4.4. It illustrates the superior accuracy of Algorithm 4.4 in the conditioning (5.3) of the selected columns $\boldsymbol{S}_1$, as well as subset selection criteria (5.1) and (5.2).

Figure 5.1(a). Algorithm 4.4 gives the best, that is smallest, ratio of condition numbers. In contrast, Algorithms 4.3 and 4.2 have a larger number of outliers above the maximum, illustrating more less reliable accuracy.

Figure 5.1(b). Algorithm 4.4 gives the best, that is closest to 1, values of γ_1. In contrast, Algorithm 4.3 has more outliers below its minimum, indicating less reliable accuracy.

Figure 5.1(c). Algorithm 4.4 has the most consistent values of γ_2, but they are slightly larger than those for Algorithms 4.3 and 4.2. Its maximum and the outliers above are comparable to those of 4.3. In contrast, Algorithm 4.1 is much less accurate.

While there are differences among Algorithms 4.1–4.4 they are relatively small, suggesting that all are effective in practice. However, we still recommend Algorithm 4.4 since it is numerically stable, computationally efficient, and is the only one whose bounds do not depend exponentially on p or k.

$\boldsymbol{S}$	Algorithms	τ	γ_1	γ_2
Kahan	4.1, 4.4	• 3.7e-03	• 1.0	1.8e03
	4.2	6.4e-01	1.6e-03	1.9e15
	4.3	• 3.7e-03	• 1.0	• 1.7e03
GuEis	4.1, 4.4	4.1e-03	0.6	0.9
	4.2	4.1e-03	0.6	5.2e11
	4.3	4.1e-03	0.6	• 1.0
Joll	4.1,4.2, 4.3, 4.4	1.6e-12	1.0	1.0
SorEm	4.1, 4.4	• 1.4e-12	• 0.9	5.4
	4.2	2.2e-12	0.5	1.1
	4.3	2.3e-12	0.5	• 1.0
SHIPS	4.1	1.9e-12	0.3	2.4
	4.2	2.9e-12	0.2	1.4
	4.3	2.0e-12	0.3	•1.4
	4.4	• 1.6e-12	• 0.4	1.9

TABLE 5.2

Accuracy of Algorithms 4.1–4.4 on the synthetic matrices. For each matrix $\boldsymbol{S}$, the average condition number ratio τ in (5.3), and the average subset selection criteria γ_1 in (5.1) and γ_2 in (5.2) over 10,000 realizations are displayed. A • denotes an optimal value for the corresponding criterion.

6. Conclusion. We have presented a numerically accurate and reliable approach for practical parameter identifiability analysis in the context of physical models.

Our recommendation is to perform column subset selection (CSS) directly on the sensitivity matrix $\boldsymbol{S}$, rather than detouring through the error-prone formation of the Fischer matrix $\boldsymbol{F} = \boldsymbol{S}^T\boldsymbol{S}$ followed by an eigenvalue decomposition.

We applied the four CSS Algorithms 4.1–4.4, to a large variety of practical and adversarial sensitivity matrices, and they produced almost identical sets of identifiable parameters $\boldsymbol{S}_1$ with vastly improved condition numbers compared to the condition number of the original matrix $\boldsymbol{S}$.

The superior accuracy of CSS is important when identifiability analysis is part of a larger application. In the context of inverse problems, for instance, parameters designated as unidentifiable may be fixed at a nominal value, for the purpose of dimension reduction. If this is an iterative process, reliable designation of unidentifiable parameters is important.

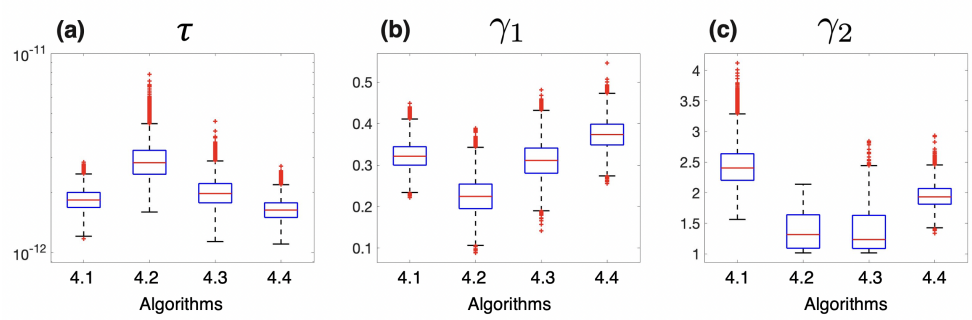

FIG. 5.1. *Application of Algorithms 4.1–4.4 to 10,000 realizations of the* `SHIPS` *matrix. Box plots show (a) the ratio of condition numbers τ in (5.3), and the subset selection criteria (b) γ_1 in (5.1), and (c) γ_2 in (5.2).*

Future research. We discuss several avenues for future research, many of which will necessitate challenging modifications to Algorithms 4.1–4.4.

1. Efficient implementation of Algorithms 4.1–4.4.
   This includes the choice of QR decompositions and data structure; as well as fast updates, searches for magnitude-largest elements, and computation of k.
2. Application of CSS methods to pharmacology.
   Physiologically-based pharmacokinetic (PBPK) and quantitative systems pharmacology (QSP) models exhibit moderate- to high-dimensional parameter spaces with highly nonlinear dependencies in their ODEs. For example, the minimal brain PBPK model in [2] has as many as 37 parameters in 16 coupled ODEs. This requires that unidentifiable parameters be determined and fixed at nominal values at the very start –prior to optimization, sensitivity analysis, Bayesian inference for computing parameter distributions, and uncertainty propagation for constructing prediction intervals for QoIs.
   Another difficulty is the optimization of criteria (5.1)–(5.3) for larger QSP and PBPK models, as they may depend strongly on the number n of observations, the number p of parameters, and the number k of identifiable parameters.
3. Global CSS algorithms.
   Algorithms 4.1–4.4 are local in the sense that they operate on a single set of nominal parameter values. However, there is significant motivation in the PBPK and QSP communities to identify parameter dependencies for a *range* of admissible parameter values. Although it might be tempting to simply average the sensitivity values, in the manner of active subspace analysis [10], the highly nonlinear nature of parameter dependencies tends to rule out this approach.
4. Mixed effects.
   Another challenge in PBPK and QSP models are the regimes that combine both, population and individual attributes. This necessitates mixed-effects models, which try to quantify the fixed-effects due to population parameters on the one hand; and the distributions for random effects associated with individuals on the other. A first step would be to incorporate CSS methods into the initial parameter subset selection algorithm for mixed-effects models in [45].

5. Virtual populations.
   A broad area of research in QSP models concerns the generation of virtual populations for the purpose of safe and efficient drug development [1]. This requires the perturbation of QSP models about nominal values and characterization of sensitivities and uncertainties associated with model parameters. We anticipate that the CSS algorithms will play an increasing role in this growing field of virtual population generation and selection.

Appendix A. Proofs. We present the proofs of Theorem 4.1 (section A.1), Theorem 4.2 (section A.2), Theorem 4.3 (section A.3), and Theorem 4.4 (section A.4).

Let $\boldsymbol{S} \in \mathbb{R}^{n\times p}$ be the sensitivity matrix with $n \geq p$, singular values $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$, and a pivoted QR decomposition, partitioned for some $1 \leq k < p$ so that

$$\boldsymbol{S}\boldsymbol{P} = \boldsymbol{Q}\begin{bmatrix}\boldsymbol{R}_{11} & \boldsymbol{R}_{12}\\ \boldsymbol{0} & \boldsymbol{R}_{22}\end{bmatrix}, \qquad \boldsymbol{R}_{11} \in \mathbb{R}^{k\times k}, \quad \boldsymbol{R}_{22} \in \mathbb{R}^{(p-k)\times(p-k)}.$$

Singular value interlacing [16, Corollary 8.6.3] implies that the singular values of $\boldsymbol{R}_{11}$ cannot exceed the corresponding dominant singular values of $\boldsymbol{S}$, while the singular values of $\boldsymbol{R}_{22}$ cannot be smaller than the corresponding subdominant singular values of $\boldsymbol{S}$, that is,

$$\begin{aligned}\sigma_j(\boldsymbol{R}_{11}) &\leq \sigma_j, \qquad 1 \leq j \leq k\\ \sigma_j(\boldsymbol{R}_{22}) &\geq \sigma_{k+j}, \qquad 1 \leq j \leq p-k.\end{aligned} \tag{A.1}$$

A.1. Proof of Theorem 4.1. We present an approximation for the smallest singular value (Lemma A.1), a correctness proof Algorithm 4.1 (Lemma A.2), and a proof of Theorem 4.1 (Lemma A.3).

In the subsequent proofs we combine different bits and pieces from [7, sections 7 and 8] and [5, section 3], and add more details for comprehension.

The key observation is that a judiciously chosen permutation can reveal a smallest singular value in a diagonal element of the triangular matrix in a QR decomposition. Below is a consequence of a more general statement in [5, Theorem 2.1].

LEMMA A.1 (Revealing a smallest singular value). *Let* $\mathbf{v}$ *with* $\|\mathbf{v}\|_2 = 1$ *be a right singular vector of* $\boldsymbol{B} \in \mathbb{R}^{m\times m}$ *associated with a smallest singular value* $\sigma_m(\boldsymbol{B})$*, so that* $\|\boldsymbol{B}\mathbf{v}\|_2 = \sigma_m(\boldsymbol{B})$*. Let* $\boldsymbol{P} \in \mathbb{R}^{m\times m}$ *be a permutation that moves a magnitude-largest element of* $\mathbf{v}$ *to the bottom,* $|(\boldsymbol{P}^T\mathbf{v})_m| = \|\mathbf{v}\|_\infty$*. If* $\boldsymbol{B}\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{R}$ *is an unpivoted QR decomposition (3.6) of* $\boldsymbol{B}\boldsymbol{P}$*, then the trailing diagonal element of the upper triangular matrix* $\boldsymbol{R}$ *satisfies*

$$\sigma_m(\boldsymbol{B}) \leq |r_{mm}| \leq \sqrt{m}\,\sigma_m(\boldsymbol{B}).$$

Proof. The lower bound follows from singular value interlacing (A.1). As for the upper bound, the relation between the right singular vector $\mathbf{v}$ and a corresponding left singular vector $\mathbf{u}$ with $\boldsymbol{B}\mathbf{v} = \sigma_m(\boldsymbol{B})\mathbf{u}$ and $\|\mathbf{u}\|_2 = 1$ implies

$$\sigma_m(\boldsymbol{B})\mathbf{u} = \boldsymbol{B}\mathbf{v} = (\boldsymbol{B}\boldsymbol{P})\,(\boldsymbol{P}^T\mathbf{v}) = \boldsymbol{Q}\boldsymbol{R}\,(\boldsymbol{P}^T\mathbf{v}) = \boldsymbol{Q}\boldsymbol{R}\begin{bmatrix}*\\ (\boldsymbol{P}^T\mathbf{v})_m\end{bmatrix}$$

From this, $\|\mathbf{u}\|_2 = 1$, the unitary invariance of the two-norm, and the upper triangular nature of $\boldsymbol{R}$ follows

$$\sigma_m(\boldsymbol{B}) = \|\sigma_m(\boldsymbol{B})\mathbf{u}\|_2 = \|\boldsymbol{R}(\boldsymbol{P}^T\mathbf{v})\|_2 \geq |r_{mm}(\boldsymbol{P}^T\mathbf{v})_m| = |r_{mm}|\,\|\mathbf{v}\|_\infty \geq |r_{mm}|/\sqrt{m}.$$

The last inequality follows from the fact that $\mathbf{v} \in \mathbb{R}^m$ has unit two-norm $\|\mathbf{v}\|_2 = 1$, so at least one of its m elements must be sufficiently large with $\|\mathbf{v}\|_\infty \geq 1/\sqrt{m}$. ∎

LEMMA A.2 (Correctness of Algorithm 4.1). *Let $\boldsymbol{S} \in \mathbb{R}^{n\times p}$ with $n \geq p$ have singular values $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$, and pick some $1 \leq k < p$. Then Algorithm 4.1 computes a QR decomposition $\boldsymbol{SP} = \boldsymbol{QR}$ where the $p-k$ trailing diagonal elements of $\boldsymbol{R}$ satisfy*

$$|\boldsymbol{R}_{\ell\ell}| \leq \sqrt{\ell}\,\sigma_\ell, \qquad k+1 \leq \ell \leq p.$$

Proof. This is an induction proof on the iterations i of Algorithm 4.1 with more discerning notation. The initial pivoted decomposition reduces the problem size

$$\boldsymbol{SP}^{(0)} = \boldsymbol{Q}^{(0)}\boldsymbol{R}^{(0)}, \tag{A.2}$$

where $\boldsymbol{P}^{(0)} \in \mathbb{R}^{p\times p}$ is a permutation, $\boldsymbol{Q}^{(0)} \in \mathbb{R}^{n\times p}$ has orthonormal columns, and $\boldsymbol{R}^{(0)} \in \mathbb{R}^{p\times p}$ is upper triangular.

Induction basis. Set $\boldsymbol{R}_{11}^{(1)} = \boldsymbol{R}^{(0)} \in \mathbb{R}^{p\times p}$, and let $\mathbf{v}^{(1)}, \mathbf{u}^{(1)} \in \mathbb{R}^p$ be right and left singular vectors associated with a smallest singular value,

$$\boldsymbol{R}_{11}^{(1)}\mathbf{v}^{(1)} = \sigma_p\mathbf{u}^{(1)}, \qquad \|\mathbf{v}^{(1)}\|_2 = \|\mathbf{u}^{(1)}\|_2 = 1.$$

Determine a permutation $\widetilde{\boldsymbol{P}}^{(1)}$ that moves a magnitude-largest element of $\mathbf{v}^{(1)}$ to the bottom,

$$|((\widetilde{\boldsymbol{P}}^{(1)})^T\mathbf{v}^{(1)})_p| = \|\mathbf{v}^{(1)}\|_\infty \geq 1/\sqrt{p}.$$

Compute an unpivoted QR decomposition $\boldsymbol{R}_{11}^{(1)}\widetilde{\boldsymbol{P}}^{(1)} = \widetilde{\boldsymbol{Q}}^{(1)}\widetilde{\boldsymbol{R}}_{11}^{(1)}$, where $\widetilde{\boldsymbol{Q}}^{(1)} \in \mathbb{R}^{p\times p}$ is an orthogonal matrix. Lemma A.1 implies that the trailing diagonal element of the triangular matrix reveals a smallest singular value, $|(\widetilde{\boldsymbol{R}}_{11}^{(1)})_{pp}| \leq \sqrt{p}\,\sigma_p$. Insert this into the initial decomposition (A.2)

$$\boldsymbol{SP}^{(0)} = \boldsymbol{Q}^{(0)}\boldsymbol{R}^{(0)} = \boldsymbol{Q}^{(0)}\widetilde{\boldsymbol{Q}}^{(1)}\widetilde{\boldsymbol{R}}_{11}^{(1)}(\widetilde{\boldsymbol{P}}^{(1)})^T.$$

Multiply by $\widetilde{\boldsymbol{P}}^{(1)}$ on the right,

$$\boldsymbol{S}\,\underbrace{\boldsymbol{P}^{(0)}\widetilde{\boldsymbol{P}}^{(1)}}_{\boldsymbol{P}^{(1)}} = \underbrace{\boldsymbol{Q}^{(0)}\widetilde{\boldsymbol{Q}}^{(1)}}_{\boldsymbol{Q}^{(1)}}\,\underbrace{\widetilde{\boldsymbol{R}}_{11}^{(1)}}_{\boldsymbol{R}^{(1)}} \qquad \text{where} \qquad |\boldsymbol{R}_{pp}^{(1)}| \leq \sqrt{p}\sigma_p.$$

Induction hypothesis. Assume that $\boldsymbol{SP}^{(i)} = \boldsymbol{Q}^{(i)}\boldsymbol{R}^{(i)}$ for $i = p-\ell$ and $\ell > k+1$ with

$$|\boldsymbol{R}_{jj}^{(i)}| \leq \sqrt{j}\,\sigma_j, \qquad \ell \leq j \leq p.$$

Induction step. Here $\ell = k+2$ is the dimension of the leading block, while $i \equiv p-\ell$ is the dimension of the trailing block. Partition

$$\boldsymbol{R}^{(i)} = \begin{bmatrix} \boldsymbol{R}_{11}^{(i)} & \boldsymbol{R}_{12}^{(i)} \\ \mathbf{0} & \boldsymbol{R}_{22}^{(i)} \end{bmatrix} \qquad \boldsymbol{R}_{11}^{(i)} \in \mathbb{R}^{\ell\times\ell}, \quad \boldsymbol{R}_{22}^{(i)} \in \mathbb{R}^{i\times i}. \tag{A.3}$$

Let $\mathbf{v}^{(i+1)}, \mathbf{u}^{(i+1)} \in \mathbb{R}^\ell$ be right and left singular vectors associated with a smallest singular value of $\boldsymbol{R}_{11}^{(i)}$,

$$\boldsymbol{R}_{11}^{(i)}\mathbf{v}^{(i+1)} = \sigma_\ell(\boldsymbol{R}_{11}^{(i)})\,\mathbf{u}^{(i+1)}, \qquad \|\mathbf{v}^{(i+1)}\|_2 = \|\mathbf{u}^{(i+1)}\|_2 = 1. \tag{A.4}$$

Determine a permutation $\widetilde{\boldsymbol{P}}^{(i+1)}$ that moves a magnitude-largest element of $\mathbf{v}^{(i+1)}$ to the bottom,

$$|((\widetilde{\boldsymbol{P}}^{(i+1)})^T \mathbf{v}^{(i+1)})_\ell| = \|\mathbf{v}^{(i+1)}\|_\infty \geq 1/\sqrt{\ell}.$$

Compute an unpivoted QR decomposition $\boldsymbol{R}_{11}^{(i)} \widetilde{\boldsymbol{P}}^{(i+1)} = \widetilde{\boldsymbol{Q}}^{(i+1)} \widetilde{\boldsymbol{R}}_{11}^{(i+1)}$, where $\widetilde{\boldsymbol{Q}}^{(i+1)} \in \mathbb{R}^{\ell \times \ell}$ is an orthogonal matrix. Lemma A.1 implies that the trailing diagonal element of the triangular matrix reveals a smallest singular value,

$$|(\widetilde{\boldsymbol{R}}_{11}^{(i+1)})_{\ell\ell}| \leq \sqrt{\ell}\, \sigma_\ell(\boldsymbol{R}_{11}^{(i)}). \tag{A.5}$$

Insert this into the decomposition $\boldsymbol{S}\boldsymbol{P}^{(i)} = \boldsymbol{Q}^{(i)}\boldsymbol{R}^{(i)}$ with partitioning (A.3), and exploit the fact that the inverse of the orthogonal matrix $\widetilde{\boldsymbol{Q}}^{(i+1)}$ is $(\widetilde{\boldsymbol{Q}}^{(i+1)})^T$,

$$\begin{aligned} \boldsymbol{S}\boldsymbol{P}^{(i)} = \boldsymbol{Q}^{(i)}\boldsymbol{R}^{(i)} &= \boldsymbol{Q}^{(i)} \begin{bmatrix} \widetilde{\boldsymbol{Q}}^{(i+1)} \widetilde{\boldsymbol{R}}_{11}^{(i+1)} (\widetilde{\boldsymbol{P}}^{(i+1)})^T & \boldsymbol{R}_{12}^{(i)} \\ \mathbf{0} & \boldsymbol{R}_{22}^{(i)} \end{bmatrix} \\ &= \boldsymbol{Q}^{(i)} \begin{bmatrix} \widetilde{\boldsymbol{Q}}^{(i+1)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_i \end{bmatrix} \begin{bmatrix} \widetilde{\boldsymbol{R}}_{11}^{(i+1)} & (\widetilde{\boldsymbol{Q}}^{(i+1)})^T \boldsymbol{R}_{12}^{(i)} \\ \mathbf{0} & \boldsymbol{R}_{22}^{(i)} \end{bmatrix} \begin{bmatrix} (\widetilde{\boldsymbol{P}}^{(i+1)})^T & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_i \end{bmatrix} \end{aligned}$$

Multiply by the permutation on the right,

$$\underbrace{\boldsymbol{S}\,\boldsymbol{P}^{(i)} \begin{bmatrix} \widetilde{\boldsymbol{P}}^{(i+1)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_i \end{bmatrix}}_{\boldsymbol{P}^{(i+1)}} = \underbrace{\boldsymbol{Q}^{(i)} \begin{bmatrix} \widetilde{\boldsymbol{Q}}^{(i+1)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_i \end{bmatrix}}_{\boldsymbol{Q}^{(i+1)}} \underbrace{\begin{bmatrix} \widetilde{\boldsymbol{R}}_{11}^{(i+1)} & (\widetilde{\boldsymbol{Q}}^{(i+1)})^T \boldsymbol{R}_{12}^{(i)} \\ \mathbf{0} & \boldsymbol{R}_{22}^{(i)} \end{bmatrix}}_{\boldsymbol{R}^{(i+1)}}.$$

From (A.5), interlacing (A.1), and the fact that $\boldsymbol{R}^{(i)}$ has the same singular values as $\boldsymbol{S}$ follows

$$|(\boldsymbol{R}^{(i+1)})_{\ell\ell}| = |(\widetilde{\boldsymbol{R}}_{11}^{(i+1)})_{\ell\ell}| \leq \sqrt{\ell}\, \sigma_\ell(\boldsymbol{R}_{11}^{(i)}) \leq \sqrt{\ell}\, \sigma_\ell(\boldsymbol{R}^{(i)}) = \sqrt{\ell}\, \sigma_\ell.$$

Together with the induction hypothesis, and $i = p - \ell = p - (k+2)$ this implies

$$|\boldsymbol{R}_{jj}^{(p-k+1)}| \leq \sqrt{j}\, \sigma_j, \qquad k+1 \leq j \leq p. \qquad \square$$

LEMMA A.3 (Proof of Theorem 4.1). *Let $\boldsymbol{S} \in \mathbb{R}^{n \times p}$ with $n \geq p$ have singular values $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$, and pick some $1 \leq k < p$. Then Algorithm 4.1 computes a QR decomposition*

$$\boldsymbol{S}\boldsymbol{P} = \begin{bmatrix} \boldsymbol{Q}_1 & \boldsymbol{Q}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \mathbf{0} & \boldsymbol{R}_{22} \end{bmatrix},$$

where the largest singular value of $\boldsymbol{R}_{22} \in \mathbb{R}^{(p-k)\times(p-k)}$ is bounded by

$$\|\boldsymbol{R}_{22}\|_2 \leq p\, \|\boldsymbol{W}^{-1}\|_2\, \sigma_{k+1}.$$

Here $\boldsymbol{W} \in \mathbb{R}^{(p-k)\times(p-k)}$ is a triangular matrix with diagonal elements $|w_{jj}| = 1$, $1 \leq j \leq p-k$; offdiagonal elements $|w_{ij}| \leq 1$ for $i \neq j$; and

$$\|\boldsymbol{W}^{-1}\|_2 \leq 2^{p-k-1}.$$

Proof. Let $\boldsymbol{S}\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{R}$ be computed by Algorithm 4.1 with input k. The proof is an extension of Lemma A.1. From the right singular vectors in Algorithm 4.1 we construct a matrix $\boldsymbol{Z}$, and then bound $\|\boldsymbol{R}\boldsymbol{Z}\|_2$ to derive an upper bound for $\|\boldsymbol{R}_{22}\|_2$.

Construction of $\boldsymbol{Z}$. The indexing of the partition is different than the one in (A.3),

$$\boldsymbol{R}^{(\ell)} = \begin{bmatrix} \boldsymbol{R}_{11}^{(\ell)} & \boldsymbol{R}_{12}^{(\ell)} \\ \mathbf{0} & \boldsymbol{R}_{22}^{(\ell)} \end{bmatrix} \qquad \boldsymbol{R}_{11}^{(\ell)} \in \mathbb{R}^{\ell\times\ell}, \quad \boldsymbol{R}_{22}^{(\ell)} \in \mathbb{R}^{(p-\ell)\times(p-\ell)}, \qquad k+1 \le \ell \le p.$$

In the statement of this lemma, the partitioning is $\ell = k$.

Let $\mathbf{v}^{(\ell)}, \mathbf{u}^{(\ell)} \in \mathbb{R}^{\ell}$ be right and left singular vectors associated with a smallest singular value of $\boldsymbol{R}_{11}^{(\ell)}$,

$$\boldsymbol{R}_{11}^{(\ell)}\mathbf{v}^{(\ell)} = \sigma_\ell(\boldsymbol{R}_{11}^{(\ell)})\,\mathbf{u}^{(\ell)}, \qquad \|\mathbf{v}^{(\ell)}\|_2 = \|\mathbf{u}^{(\ell)}\|_2 = 1, \qquad k+1 \le \ell \le p.$$

Algorithm 4.1 has permuted the right singular vectors so that a magnitude-largest element is at the bottom,

$$|\mathbf{v}_\ell^{(\ell)}| \ge 1/\sqrt{\ell} \quad \text{and} \quad |\mathbf{v}_j^{(\ell)}| \le |\mathbf{v}_\ell^{(\ell)}|, \qquad 1 \le j < \ell, \quad k+1 \le \ell \le p. \tag{A.6}$$

The trailing elements in singular vectors associated with larger-dimensional blocks are not affected by subsequent permutations, see (A.3), where permutations in the $(1,1)$ block do not affect the $(2,2)$ block and its placement of diagonal elements.

Construct an upper trapezoidal matrix $\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{z}_1 & \cdots & \boldsymbol{z}_{p-k} \end{bmatrix} \in \mathbb{R}^{p\times(p-k)}$, whose columns are the right singular vectors

$$\boldsymbol{z}_{\ell-k} = \begin{bmatrix} \mathbf{v}^{(\ell)} \\ \mathbf{0}_{p-\ell} \end{bmatrix}, \qquad k+1 \le \ell \le p.$$

Factor out the diagonal elements and focus on the trailing $(p-k)\times(p-k)$ submatrix

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 \\ \boldsymbol{W} \end{bmatrix} \boldsymbol{D}, \qquad \text{where} \qquad \boldsymbol{D} = \begin{bmatrix} v_{k+1}^{(k+1)} & & \\ & \ddots & \\ & & v_p^{(p)} \end{bmatrix} \in \mathbb{R}^{(p-k)\times(p-k)} \tag{A.7}$$

has diagonal elements $|d_{\ell\ell}| = |v_\ell^{(\ell)}| \ge 1/\sqrt{\ell}$, $k+1 \le \ell \le p$. From (A.6) follows that $\boldsymbol{W} \in \mathbb{R}^{(p-k)\times(p-k)}$ is a nonsingular upper triangular matrix with elements

$$|w_{\ell\ell}| = 1, \qquad |w_{\ell j}| \le 1, \qquad 1 \le \ell \le p-k, \quad j > \ell.$$

Bounds for $\|\boldsymbol{RZ}\|_2$. We derive an upper and a lower bound. Multiplying the QR decomposition $\boldsymbol{SP} = \boldsymbol{QR}$ by $\boldsymbol{Q}^T$ on the left and by $\boldsymbol{Z}$ on the right gives

$$\boldsymbol{Q}^T\boldsymbol{SPZ} = \boldsymbol{RZ} \in \mathbb{R}^{p-k}.$$

The columns of $\boldsymbol{RZ}$ are

$$\boldsymbol{R}\boldsymbol{z}_{\ell-k} = \begin{bmatrix} \boldsymbol{R}_{11}^{(\ell)}\mathbf{v}^{(\ell)} \\ \mathbf{0}_{p-\ell} \end{bmatrix} = \sigma_\ell(\boldsymbol{R}_{11}^{(\ell)}) \begin{bmatrix} \mathbf{u}^{(\ell)} \\ \mathbf{0}_{p-\ell} \end{bmatrix}, \qquad k+1 \le \ell \le p.$$

From $\|\mathbf{u}^{(\ell)}\|_2 = 1$ and interlacing (A.1) follows

$$\|\boldsymbol{R}\boldsymbol{z}_{\ell-k}\|_2 = \sigma_\ell(\boldsymbol{R}_{11}^{(\ell)}) \le \sigma_\ell, \qquad k+1 \le \ell \le p.$$

Bound the norm of $\boldsymbol{RZ} \in \mathbb{R}^{p\times(p-k)}$ in terms of its largest column norm [16, section 2.3.2] to obtain the upper bound

$$\|\boldsymbol{RZ}\|_2 \leq \sqrt{p-k}\, \max_{k+1\leq \ell\leq p} \|\boldsymbol{R}\boldsymbol{z}_{\ell-k}\|_2 \leq \sqrt{p-k}\, \max_{k+1\leq \ell\leq p} \sigma_\ell \leq \sqrt{p}\,\sigma_{k+1}. \tag{A.8}$$

As for the lower bound, use the partitioning in the statement of this lemma,

$$\boldsymbol{RZ} = \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{Z}_1\boldsymbol{D} \\ \boldsymbol{W}\boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}_{11}\boldsymbol{Z}_1\boldsymbol{D} + \boldsymbol{R}_{12}\boldsymbol{W}\boldsymbol{D} \\ \boldsymbol{R}_{22}\boldsymbol{W}\boldsymbol{D} \end{bmatrix},$$

and bound $\|\boldsymbol{RZ}\|_2$ in terms of the trailing component

$$\|\boldsymbol{RZ}\|_2 \geq \|\boldsymbol{R}_{22}\boldsymbol{W}\boldsymbol{D}\|_2 \geq \frac{\|\boldsymbol{R}_{22}\|_2}{\|\boldsymbol{W}^{-1}\|_2\|\boldsymbol{D}^{-1}\|_2} \geq \frac{\|\boldsymbol{R}_{22}\|_2}{\sqrt{p}\,\|\boldsymbol{W}^{-1}\|_2}.$$

At last combine the above upper bound with the lower bound (A.8),

$$\|\boldsymbol{R}_{22}\| \leq p\,\|\boldsymbol{W}^{-1}\|_2\,\sigma_{k+1}.$$

The bound for $\|\boldsymbol{W}^{-1}\|_2$ is derived in [22, Theorem 8.14]; and there are classes of matrices for which it can essentially be tight [22, section 8.3]. ∎

A.2. Proof of Theorem 4.2. We present an approximation for the largest singular value (Lemma A.4), a correctness proof Algorithm 4.2 (Lemma A.5), and a proof of Theorem 4.2 (Lemma A.6).

In the subsequent proofs, we present more general and simpler derivations than the ones in [7, section 7] and [6, sections 2 and 3], and add more details for comprehension.

The key observation is that a judiciously chosen permutation can reveal a largest singular value in a diagonal element of the triangular matrix in a QR decomposition. The next statement represents part of [6, Theorem 2.1], however with a simpler proof that does not require a pseudo inverse as in [6, Theorems 6.1 and 6.2].

LEMMA A.4 (Revealing a largest singular value). *Let* $\mathbf{v}$ *with* $\|\mathbf{v}\|_2 = 1$ *be a right singular vector of* $\boldsymbol{B} \in \mathbb{R}^{m\times m}$ *associated with a largest singular value* $\sigma_1(\boldsymbol{B})$*, so that* $\|\boldsymbol{B}\mathbf{v}\|_2 = \sigma_1(\boldsymbol{B})$*. Let* $\boldsymbol{P} \in \mathbb{R}^{m\times m}$ *be a permutation that moves a magnitude-largest element of* $\mathbf{v}$ *to the top,* $|(\boldsymbol{P}^T\mathbf{v})_1| = \|\mathbf{v}\|_\infty$*. If* $\boldsymbol{BP} = \boldsymbol{QR}$ *is an unpivoted QR decomposition (3.6) of* $\boldsymbol{BP}$*, then the leading diagonal element of the upper triangular matrix* $\boldsymbol{R}$ *satisfies*

$$\sigma_1(\boldsymbol{B})/\sqrt{m} \leq |r_{11}| \leq \sigma_1(\boldsymbol{B}).$$

Proof. The upper bound follows from singular value interlacing (A.1). As for the lower bound, the relation between the right singular vector $\mathbf{v}$ and a corresponding left singular vector $\mathbf{u}$ with $\boldsymbol{B}^T\mathbf{u} = \sigma_1(\boldsymbol{B})\mathbf{v}$ and $\|\mathbf{u}\|_2 = 1$ implies

$$\sigma_1(\boldsymbol{B})\,\boldsymbol{P}^T\mathbf{v} = \boldsymbol{P}^T\boldsymbol{B}\,\mathbf{u} = \boldsymbol{R}^T\boldsymbol{Q}^T\mathbf{u}.$$

From this, the lower triangular nature of $\boldsymbol{R}^T$, the Cauchy Schwartz inequality, and $\|\mathbf{u}\|_2 = 1$ follows for the leading element

$$\sigma_1(\boldsymbol{B})\|\mathbf{v}\|_\infty = |\sigma_1(\boldsymbol{B})(\boldsymbol{P}^T\mathbf{v})_1| = |\boldsymbol{e}_1^T\boldsymbol{R}^T(\boldsymbol{Q}^T\mathbf{u})| \leq \|\boldsymbol{R}\boldsymbol{e}_1\|_2\|\boldsymbol{Q}^T\mathbf{u}\|_2 = |r_{11}|.$$

Then $\|\mathbf{v}\|_\infty \geq 1/\sqrt{m}$ follows from the fact that $\mathbf{v} \in \mathbb{R}^m$ has unit two-norm $\|\mathbf{v}\|_2 = 1$, so at least one of its m elements must be sufficiently large. ∎

LEMMA A.5 (Correctness of Algorithm 4.2). *Let $\boldsymbol{S} \in \mathbb{R}^{n\times p}$ with $n \geq p$ have singular values $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$, and pick some $1 \leq k < p$. Then Algorithm 4.2 computes a QR decomposition $\boldsymbol{SP} = \boldsymbol{QR}$ where the k leading diagonal elements of $\boldsymbol{R}$ satisfy*

$$\sigma_\ell/\sqrt{p-\ell+1} \leq |\boldsymbol{R}_{\ell\ell}|, \qquad 1 \leq \ell \leq k.$$

Proof. This is an induction proof on the iterations ℓ of Algorithm 4.2 with more discerning notation. The initial pivoted decomposition reduces the problem size

$$\boldsymbol{SP}^{(0)} = \boldsymbol{Q}^{(0)}\boldsymbol{R}^{(0)}, \tag{A.9}$$

where $\boldsymbol{P}^{(0)} \in \mathbb{R}^{p\times p}$ is a permutation, $\boldsymbol{Q}^{(0)} \in \mathbb{R}^{n\times p}$ has orthonormal columns, and $\boldsymbol{R}^{(0)} \in \mathbb{R}^{p\times p}$ is upper triangular.

Induction basis. Set $\boldsymbol{R}_{22}^{(1)} = \boldsymbol{R}^{(0)} \in \mathbb{R}^{p\times p}$, and let $\mathbf{v}^{(1)}, \mathbf{u}^{(1)} \in \mathbb{R}^p$ be right and left singular vectors associated with a largest singular value,

$$\boldsymbol{R}_{22}^{(1)}\mathbf{v}^{(1)} = \sigma_1\mathbf{u}^{(1)}, \qquad \|\mathbf{v}^{(1)}\|_2 = \|\mathbf{u}^{(1)}\|_2 = 1.$$

Determine a permutation $\widetilde{\boldsymbol{P}}^{(1)}$ that moves a magnitude-largest element of $\mathbf{v}^{(1)}$ to the top,

$$|((\widetilde{\boldsymbol{P}}^{(1)})^T\mathbf{v}^{(1)})_1| = \|\mathbf{v}^{(1)}\|_\infty \geq 1/\sqrt{p}.$$

Compute an unpivoted QR decomposition $\boldsymbol{R}_{22}^{(1)}\widetilde{\boldsymbol{P}}^{(1)} = \widetilde{\boldsymbol{Q}}^{(1)}\widetilde{\boldsymbol{R}}_{22}^{(1)}$, where $\widetilde{\boldsymbol{Q}}^{(1)} \in \mathbb{R}^{p\times p}$ is an orthogonal matrix. Lemma A.4 implies that the leading diagonal element of the triangular matrix reveals a largest singular value, $|(\widetilde{\boldsymbol{R}}_{22}^{(1)})_{11}| \geq \sigma_1/\sqrt{p}$. Insert this into the initial decomposition (A.9)

$$\boldsymbol{SP}^{(0)} = \boldsymbol{Q}^{(0)}\boldsymbol{R}^{(0)} = \boldsymbol{Q}^{(0)}\widetilde{\boldsymbol{Q}}^{(1)}\widetilde{\boldsymbol{R}}_{22}^{(1)}(\widetilde{\boldsymbol{P}}^{(1)})^T.$$

Multiply by $\widetilde{\boldsymbol{P}}^{(1)}$ on the right,

$$\boldsymbol{S}\,\underbrace{\boldsymbol{P}^{(0)}\widetilde{\boldsymbol{P}}^{(1)}}_{\boldsymbol{P}^{(1)}} = \underbrace{\boldsymbol{Q}^{(0)}\widetilde{\boldsymbol{Q}}^{(1)}}_{\boldsymbol{Q}^{(1)}}\,\underbrace{\widetilde{\boldsymbol{R}}_{22}^{(1)}}_{\boldsymbol{R}^{(1)}} \qquad \text{where} \qquad |\boldsymbol{R}_{22}^{(1)}| \geq \sigma_1/\sqrt{p}.$$

Induction hypothesis. Assume that $\boldsymbol{SP}^{(\ell)} = \boldsymbol{Q}^{(\ell)}\boldsymbol{R}^{(\ell)}$ for $\ell < k$ with

$$|\boldsymbol{R}_{jj}^{(\ell)}| \geq \sigma_j/\sqrt{p-j+1}, \qquad 1 \leq j \leq \ell.$$

Induction step. Here $\ell = k-1$. The dimension of the leading block is $\ell-1$, while the dimension of the trailing block is $i \equiv p-(\ell-1)$. Partition

$$\boldsymbol{R}^{(\ell)} = \begin{bmatrix} \boldsymbol{R}_{11}^{(\ell)} & \boldsymbol{R}_{12}^{(\ell)} \\ \mathbf{0} & \boldsymbol{R}_{22}^{(\ell)} \end{bmatrix} \qquad \boldsymbol{R}_{11}^{(\ell)} \in \mathbb{R}^{(\ell-1)\times(\ell-1)}, \quad \boldsymbol{R}_{22}^{(\ell)} \in \mathbb{R}^{i\times i}. \tag{A.10}$$

Let $\mathbf{v}^{(\ell+1)}, \mathbf{u}^{(\ell+1)} \in \mathbb{R}^i$ be right and left singular vectors associated with a largest singular value of $\boldsymbol{R}_{22}^{(\ell)}$,

$$\boldsymbol{R}_{22}^{(\ell)}\mathbf{v}^{(\ell+1)} = \sigma_1(\boldsymbol{R}_{22}^{(\ell)})\,\mathbf{u}^{(\ell+1)}, \qquad \|\mathbf{v}^{(\ell+1)}\|_2 = \|\mathbf{u}^{(\ell+1)}\|_2 = 1. \tag{A.11}$$

Determine a permutation $\widetilde{\boldsymbol{P}}^{(\ell+1)} \in \mathbb{R}^{i\times i}$ that moves a magnitude-largest element of $\mathbf{v}^{(\ell+1)}$ to the top,

$$|((\widetilde{\boldsymbol{P}}^{(\ell+1)})^T \mathbf{v}^{(\ell+1)})_1| = \|\mathbf{v}^{(\ell+1)}\|_\infty \geq 1/\sqrt{i}.$$

Compute an unpivoted QR decomposition $\boldsymbol{R}_{22}^{(\ell)} \widetilde{\boldsymbol{P}}^{(\ell+1)} = \widetilde{\boldsymbol{Q}}^{(\ell+1)} \widetilde{\boldsymbol{R}}_{22}^{(\ell+1)}$, where $\widetilde{\boldsymbol{Q}}^{(\ell+1)} \in \mathbb{R}^{i\times i}$ is an orthogonal matrix. Lemma A.4 implies that the leading diagonal element of the triangular matrix reveals a largest singular value,

$$|(\widetilde{\boldsymbol{R}}_{22}^{(\ell+1)})_{11}| \geq \sigma_1(\boldsymbol{R}_{22}^{(\ell)})/\sqrt{i}. \tag{A.12}$$

Insert this into the decomposition $\boldsymbol{SP}^{(\ell)} = \boldsymbol{Q}^{(\ell)}\boldsymbol{R}^{(\ell)}$ with partitioning (A.10), and exploit the fact that the inverse of the orthogonal matrix $\widetilde{\boldsymbol{Q}}^{(\ell+1)}$ equals $(\widetilde{\boldsymbol{Q}}^{(\ell+1)})^T$,

$$\begin{aligned}\boldsymbol{SP}^{(\ell)} = \boldsymbol{Q}^{(\ell)}\boldsymbol{R}^{(\ell)} &= \boldsymbol{Q}^{(\ell)} \begin{bmatrix} \boldsymbol{R}_{11}^{(\ell)} & \boldsymbol{R}_{12}^{(\ell)} \\ \mathbf{0} & \widetilde{\boldsymbol{Q}}^{(\ell+1)} \widetilde{\boldsymbol{R}}_{22}^{(\ell+1)} (\widetilde{\boldsymbol{P}}^{(\ell+1)})^T \end{bmatrix} \\ &= \boldsymbol{Q}^{(\ell)} \begin{bmatrix} \boldsymbol{I}_{\ell-1} & \mathbf{0} \\ \mathbf{0} & \widetilde{\boldsymbol{Q}}^{(\ell+1)} \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_{11}^{(\ell)} & \boldsymbol{R}_{12}^{(\ell)} \\ \mathbf{0} & \widetilde{\boldsymbol{R}}_{22}^{(\ell+1)} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_{\ell-1} & \mathbf{0} \\ \mathbf{0} & (\widetilde{\boldsymbol{P}}^{(\ell+1)})^T \end{bmatrix}\end{aligned}$$

Multiply by the permutation on the right,

$$\underbrace{\boldsymbol{S}\,\boldsymbol{P}^{(\ell)} \begin{bmatrix} \boldsymbol{I}_{\ell-1} & \mathbf{0} \\ \mathbf{0} & \widetilde{\boldsymbol{P}}^{(\ell+1)} \end{bmatrix}}_{\boldsymbol{P}^{(\ell+1)}} = \underbrace{\boldsymbol{Q}^{(\ell)} \begin{bmatrix} \boldsymbol{I}_{\ell-1} & \mathbf{0} \\ \mathbf{0} & \widetilde{\boldsymbol{Q}}^{(\ell+1)} \end{bmatrix}}_{\boldsymbol{Q}^{(\ell+1)}} \underbrace{\begin{bmatrix} \boldsymbol{R}_{11}^{(\ell)} & \boldsymbol{R}_{12}^{(\ell)} \\ \mathbf{0} & \widetilde{\boldsymbol{R}}_{22}^{(\ell+1)} \end{bmatrix}}_{\boldsymbol{R}^{(\ell+1)}}.$$

From (A.12), interlacing (A.1), and the fact that $\boldsymbol{R}^{(\ell)}$ has the same singular values as $\boldsymbol{S}$ follows

$$|\boldsymbol{R}_{\ell\ell}^{(\ell+1)}| = |(\widetilde{\boldsymbol{R}}_{22}^{(\ell+1)})_{11}| \geq \sigma_1(\boldsymbol{R}_{22}^{(\ell)})/\sqrt{i} \geq \sigma_\ell(\boldsymbol{R}^{(\ell)})/\sqrt{i} = \sigma_\ell/\sqrt{i}.$$

Together with the induction hypothesis, and $\ell = k-1$ this implies

$$|\boldsymbol{R}_{jj}^{(k)}| \geq \sigma_j/\sqrt{p-j+1}, \qquad 1 \leq j \leq k. \qquad \square$$

LEMMA A.6 (Proof of Theorem 4.2). *Let $\boldsymbol{S} \in \mathbb{R}^{n\times p}$ with $n \geq p$ have singular values $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$, and pick some $1 \leq k < p$. Then Algorithm 4.2 computes a QR decomposition*

$$\boldsymbol{SP} = \begin{bmatrix} \boldsymbol{Q}_1 & \boldsymbol{Q}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \mathbf{0} & \boldsymbol{R}_{22} \end{bmatrix},$$

where the smallest singular value of $\boldsymbol{R}_{11} \in \mathbb{R}^{k\times k}$ is bounded by

$$\sigma_k(\boldsymbol{R}_{11}) \geq \frac{\sigma_k}{p\,\|\boldsymbol{W}^{-1}\|_2}.$$

Here $\boldsymbol{W} \in \mathbb{R}^{k\times k}$ is a triangular matrix with diagonal elements $|w_{jj}| = 1$, $1 \leq j \leq k$; offdiagonal elements $|w_{ij}| \leq 1$ for $i \neq j$; and

$$\|\boldsymbol{W}^{-1}\|_2 \leq 2^{k-1}.$$

Proof. Let $\boldsymbol{SP} = \boldsymbol{QR}$ be computed by Algorithm 4.2 with input k. The proof is an extension of Lemma A.4, and is more general than the one in [7, section 7] due to the absence of inverses and no need for the requirement $\sigma_k > 0$.

From the right singular vectors in Algorithm 4.2 we construct a matrix $\boldsymbol{Z}$, and also a matrix $\boldsymbol{Y}$ of left singular vectors. Then we bound the kth singular value of a top submatrix of $\boldsymbol{R}^T\boldsymbol{Y}$, to derive a lower bound for $\sigma_k(\boldsymbol{R}_{11})$.

Construction of $\boldsymbol{Z}$ and $\boldsymbol{Y}$. Consider the partitionings as in (A.10) with $i \equiv p - (\ell - 1)$

$$\boldsymbol{R}^{(\ell)} = \begin{bmatrix} \boldsymbol{R}_{11}^{(\ell)} & \boldsymbol{R}_{12}^{(\ell)} \\ \boldsymbol{0} & \boldsymbol{R}_{22}^{(\ell)} \end{bmatrix} \qquad \boldsymbol{R}_{11}^{(\ell)} \in \mathbb{R}^{(\ell-1)\times(\ell-1)}, \quad \boldsymbol{R}_{22}^{(\ell)} \in \mathbb{R}^{i\times i}, \quad 1 \le \ell \le k.$$

In the statement of this lemma, the partitioning is $\ell = k + 1$.

Let $\mathbf{v}^{(\ell)}, \mathbf{u}^{(\ell)} \in \mathbb{R}^i$ be right and left singular vectors associated with a largest singular value of $\boldsymbol{R}_{22}^{(\ell)}$,

$$\boldsymbol{R}_{22}^{(\ell)}\mathbf{v}^{(\ell)} = \sigma_1(\boldsymbol{R}_{22}^{(\ell)})\,\mathbf{u}^{(\ell)}, \qquad \|\mathbf{v}^{(\ell)}\|_2 = \|\mathbf{u}^{(\ell)}\|_2 = 1, \qquad 1 \le \ell \le k.$$

Algorithm 4.2 has permuted the right singular vectors so that a magnitude-largest element is at the top, for $1 \le \ell \le k$

$$|\mathbf{v}_1^{(\ell)}| \ge 1/\sqrt{i} \quad \text{and} \quad |\mathbf{v}_j^{(\ell)}| \le |\mathbf{v}_1^{(\ell)}|, \qquad 1 < j \le i. \tag{A.13}$$

The leading elements in singular vectors associated with larger-dimensional blocks are not affected by subsequent permutations, see (A.10), where permutations in the $(2,2)$ block do not affect the $(1,1)$ block and its placement of diagonal elements.

Construct a lower trapezoidal matrix $\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{z}_1 & \cdots & \boldsymbol{z}_k \end{bmatrix} \in \mathbb{R}^{p\times k}$, whose columns are the right singular vectors

$$\boldsymbol{z}_\ell = \begin{bmatrix} \mathbf{0}_{\ell-1} \\ \mathbf{v}^{(\ell)} \end{bmatrix}, \qquad 1 \le \ell \le k.$$

Factor out the diagonal elements and distinguish the leading $k \times k$ submatrix

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{W} \\ \boldsymbol{Z}_2 \end{bmatrix} \boldsymbol{D}, \qquad \text{where} \qquad \boldsymbol{D} = \begin{bmatrix} v_1^{(1)} & & \\ & \ddots & \\ & & v_1^{(k)} \end{bmatrix} \in \mathbb{R}^{k\times k} \tag{A.14}$$

has diagonal elements $|d_{\ell\ell}| = |v_1^{(\ell)}| \ge 1/\sqrt{p-\ell+1}$, $1 \le \ell \le k$. From (A.13) follows that $\boldsymbol{W} \in \mathbb{R}^{k\times k}$ is a nonsingular lower triangular matrix with elements

$$|w_{\ell\ell}| = 1, \qquad |w_{j\ell}| \le 1, \qquad 1 \le \ell \le k, \quad j > \ell.$$

Analogously, construct a second lower trapezoidal matrix $\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1 & \cdots & \boldsymbol{y}_k \end{bmatrix} \in \mathbb{R}^{p\times k}$, whose columns are the right left vectors

$$\boldsymbol{y}_\ell = \begin{bmatrix} \mathbf{0}_{\ell-1} \\ \mathbf{u}^{(\ell)} \end{bmatrix}, \qquad \|\boldsymbol{y}_\ell\|_2 = 1, \qquad 1 \le \ell \le k,$$

and distinguish the leading $k \times k$ submatrix

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{bmatrix}, \qquad \text{where} \qquad \boldsymbol{Y}_1 \in \mathbb{R}^{k\times k}, \qquad \|\boldsymbol{Y}_1\|_2 \le \sqrt{k}. \tag{A.15}$$

Bounds for $\sigma_k(\boldsymbol{R}_{11}^T \boldsymbol{Y}_1)$. We derive an upper and a lower bound.
The columns of $\boldsymbol{R}^T\boldsymbol{Y}$ are for $1 \le \ell \le k$,

$$\boldsymbol{R}^T \boldsymbol{y}_\ell = \begin{bmatrix} (\boldsymbol{R}_{11}^{(\ell)})^T & \mathbf{0} \\ (\boldsymbol{R}_{12}^{(\ell)})^T & (\boldsymbol{R}_{22}^{(\ell)})^T \end{bmatrix} \begin{bmatrix} \mathbf{0}_{\ell-1} \\ \mathbf{u}_\ell \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{\ell-1} \\ (\boldsymbol{R}_{22}^{(\ell)})^T \mathbf{u}_\ell \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{\ell_1} \\ \sigma_1(\boldsymbol{R}_{22}^{\ell}) \mathbf{v}_\ell \end{bmatrix} = \sigma_1(\boldsymbol{R}_{22}^{\ell}) \boldsymbol{z}_\ell.$$

Collecting all the columns gives

$$\boldsymbol{R}^T\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\Delta} \qquad \text{where} \quad \boldsymbol{\Delta} = \begin{bmatrix} \sigma_1(\boldsymbol{R}_{22}^{(1)}) & & \\ & \ddots & \\ & & \sigma_1(\boldsymbol{R}_{22}^{(k)}) \end{bmatrix} \in \mathbb{R}^{k \times k}.$$

With the partitioning of $\boldsymbol{R}$ as in the statement of this lemma, the top $k \times k$ submatrix of $\boldsymbol{R}^T\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\Delta}$ equals

$$\boldsymbol{R}_{11}^T \boldsymbol{Y}_1 = \boldsymbol{W}\boldsymbol{D}\boldsymbol{\Delta}.$$

First derive the lower bound from the right side. The Weyl product inequalities [26, 7.3.P16] imply

$$\sigma_k(\boldsymbol{R}_{11}^T \boldsymbol{Y}_1) = \sigma_k(\boldsymbol{W}\boldsymbol{D}\boldsymbol{\Delta}) \ge \sigma_k(\boldsymbol{W})\, \sigma_k(\boldsymbol{D})\, \sigma_k(\boldsymbol{\Delta}) \ge \frac{\sigma_k}{\sqrt{p-k+1}\, \|\boldsymbol{W}^{-1}\|_2} \tag{A.16}$$

where the last inequality follows from applying interlacing (A.1) to

$$\sigma_k(\boldsymbol{\Delta}) = \min_{1 \le \ell \le k} \sigma_1(\boldsymbol{R}_{22}^{(\ell)}) \ge \sigma_k,$$

and bounding the diagonal elements of $\boldsymbol{D}$ in (A.14) by

$$\sigma_k(\boldsymbol{D}) = \min_{1 \le \ell \le k} |v_1^{(\ell)}| \ge 1/\sqrt{p-k+1}.$$

Now derive the lower bound from the left side. The Weyl product inequalities [26, 7.3.P16] and (A.15) imply

$$\sigma_k(\boldsymbol{R}_{11}^T \boldsymbol{Y}_1) \le \sigma_k(\boldsymbol{R}_{11}) \|\boldsymbol{Y}_1\|_2 \le \sqrt{k}\, \sigma_k(\boldsymbol{R}_{11}).$$

At last, combine this with (A.16) to obtain

$$\sigma_k(\boldsymbol{R}_{11}) \ge \frac{\sigma_k}{\sqrt{k(p-k+1)}\, \|\boldsymbol{W}^{-1}\|_2} \ge \frac{\sigma_k}{p\, \|\boldsymbol{W}^{-1}\|_2}.$$

The bound for $\|\boldsymbol{W}^{-1}\|_2$ follows as in the proof of Lemma A.3. ∎

A.3. Proof of Theorem 4.3. The following is an extension of [16, Theorem 5.5.2].

LEMMA A.7 (Proof of Theorem 4.3). *Let* $\boldsymbol{S} \in \mathbb{R}^{n \times p}$ *with* $n \ge p$ *have singular values* $\sigma_1 \ge \cdots \ge \sigma_p \ge 0$, *and pick some* $1 \le k < p$. *If Algorithm 4.3 computes a QR decomposition*

$$\boldsymbol{S}\boldsymbol{P} = \begin{bmatrix} \boldsymbol{Q}_1 & \boldsymbol{Q}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \mathbf{0} & \boldsymbol{R}_{22} \end{bmatrix},$$

and chooses the permutation $\boldsymbol{P}$ *so that* $\boldsymbol{V}_{11} \in \mathbb{R}^{k\times k}$ *is nonsingular, then* $\boldsymbol{R}_{11} \in \mathbb{R}^{k\times k}$ *and* $\boldsymbol{R}_{22} \in \mathbb{R}^{(p-k)\times(p-k)}$ *satisfy*

$$\sigma_k/\|\boldsymbol{V}_{11}^{-1}\|_2 \le \sigma_k(\boldsymbol{R}_{11}) \le \sigma_k$$
$$\sigma_{k+1} \le \sigma_1(\boldsymbol{R}_{22}) \le \|\boldsymbol{V}_{11}^{-1}\|_2\, \sigma_{k+1}.$$

Proof. Let $\boldsymbol{S} = \boldsymbol{Q}\boldsymbol{R}$ be a preliminary unpivoted QR decomposition, where $\boldsymbol{Q} \in \mathbb{R}^{n\times p}$ has orthonormal columns, and $\boldsymbol{R} \in \mathbb{R}^{p\times p}$ is upper triangular. Then let $\boldsymbol{R} = \boldsymbol{U}_r\boldsymbol{\Sigma}\boldsymbol{V}^T$ be an SVD of the triangular matrix as in Remark 3.1. Distinguish the matrix of k largest singular values $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{k\times k}$ of $\boldsymbol{S}$, and the corresponding right singular vectors $\boldsymbol{V}_1 \in \mathbb{R}^{p\times k}$,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \in \mathbb{R}^{p\times p}, \qquad \boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_2 \end{bmatrix} \in \mathbb{R}^{p\times p}.$$

Main idea. Perform a QR decomposition with column pivoting on $\boldsymbol{V}_1^T$,

$$\boldsymbol{V}_1^T\boldsymbol{P} = \boldsymbol{Q}_1 \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \end{bmatrix},$$

where $\boldsymbol{P} \in \mathbb{R}^{p\times p}$ is a permutation matrix, $\boldsymbol{Q}_1 \in \mathbb{R}^{k\times k}$ is an orthogonal matrix, and $\boldsymbol{V}_{11} \in \mathbb{R}^{k\times k}$ is nonsingular upper triangular. Partition commensurately,

$$\boldsymbol{V}_2^T\boldsymbol{P} = \begin{bmatrix} \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{bmatrix},$$

where $\boldsymbol{V}_{22} \in \mathbb{R}^{(p-k)\times(p-k)}$. Express the permuted upper triangular matrix $\boldsymbol{R}\boldsymbol{P}$ in terms of these partitions,

$$\begin{aligned} \boldsymbol{R}\boldsymbol{P} = \boldsymbol{U}_r\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{P} &= \boldsymbol{U}_r \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{bmatrix} \\ &= \boldsymbol{U}_r \begin{bmatrix} \widehat{\boldsymbol{\Sigma}}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{bmatrix} \qquad \text{where} \quad \widehat{\boldsymbol{\Sigma}}_1 \equiv \boldsymbol{\Sigma}_1\boldsymbol{Q}_1. \end{aligned} \tag{A.17}$$

Because $\boldsymbol{Q}_1$ is an orthogonal matrix, $\widehat{\boldsymbol{\Sigma}}_1$ has the same singular values as $\boldsymbol{\Sigma}_1$, that is,

$$\sigma_j(\widehat{\boldsymbol{\Sigma}}_1) = \sigma_j(\boldsymbol{\Sigma}_1) = \sigma_j, \qquad 1 \le j \le k. \tag{A.18}$$

Re-triangularize by computing an unpivoted QR decomposition of $\boldsymbol{R}\boldsymbol{P}$,

$$\boldsymbol{R}\boldsymbol{P} = \boldsymbol{Q}_r \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22} \end{bmatrix} \tag{A.19}$$

where $\boldsymbol{R}_{11} \in \mathbb{R}^{k\times k}$ is upper triangular.

Inequality for $\boldsymbol{R}_{11}$. Equate (A.19) with (A.17) and move $\boldsymbol{U}_r$ to the left

$$\boldsymbol{U}_r^T\boldsymbol{Q}_r \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22} \end{bmatrix} = \boldsymbol{U}_r^T\boldsymbol{R}\boldsymbol{P} = \begin{bmatrix} \widehat{\boldsymbol{\Sigma}}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{bmatrix}.$$

The goal is to extract $\boldsymbol{R}_{11}$. To this end partition

$$\boldsymbol{U}_r^T\boldsymbol{Q}_r = \begin{bmatrix} \boldsymbol{U}_{11} & \boldsymbol{U}_{12} \\ \boldsymbol{U}_{21} & \boldsymbol{U}_{22} \end{bmatrix}$$

and substitute this into the above expression for $\boldsymbol{U}_r^T \boldsymbol{R}\boldsymbol{P}$,

$$\begin{bmatrix} \boldsymbol{U}_{11} & \boldsymbol{U}_{12} \\ \boldsymbol{U}_{21} & \boldsymbol{U}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22} \end{bmatrix} = \begin{bmatrix} \widehat{\boldsymbol{\Sigma}}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{bmatrix}.$$

Due to the triangular and diagonal matrices, the (1,1) block of this equation is

$$\boldsymbol{U}_{11}\boldsymbol{R}_{11} = \widehat{\boldsymbol{\Sigma}}_1 \boldsymbol{V}_{11}.$$

Apply the Weyl product inequalities for singular values [26, (7.3.14)] to the smallest singular value of the matrices on both sides and remember (A.18),

$$\frac{\sigma_k}{\|\boldsymbol{V}_{11}^{-1}\|_2} = \sigma_k(\widehat{\boldsymbol{\Sigma}}_1)\sigma_k(\boldsymbol{V}_{11}) \leq \sigma_k(\widehat{\boldsymbol{\Sigma}}_1 \boldsymbol{V}_{11}) = \sigma_k(\boldsymbol{U}_{11}\boldsymbol{R}_{11}).$$

Because the orthogonal matrix $\boldsymbol{U}$ has all singular values equal to one,

$$\sigma_k(\boldsymbol{U}_{11}\boldsymbol{R}_{11}) \leq \sigma_1(\boldsymbol{U}_{11})\sigma_k(\boldsymbol{R}_{11}) \leq \sigma_1(\boldsymbol{U})\sigma_k(\boldsymbol{R}_{11}) = \sigma_k(\boldsymbol{R}_{11}).$$

Combining the extreme ends of the sequence of inequalities gives $\sigma_k/\|\boldsymbol{V}_{11}^{-1}\|_2 \leq \sigma_k(\boldsymbol{R}_{11})$.

Inequality for $\boldsymbol{R}_{22}$. Again, equate (A.19) with (A.17) but now move the $\boldsymbol{V}$ matrix to the left,

$$\begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_{11}^T & \boldsymbol{V}_{21}^T \\ \boldsymbol{V}_{12}^T & \boldsymbol{V}_{22}^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}_{11}^T & \boldsymbol{U}_{21}^T \\ \boldsymbol{U}_{12}^T & \boldsymbol{U}_{22}^T \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\Sigma}}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix}.$$

As before, the triangular and diagonal matrices imply that the (2,2) block of this equation is

$$\boldsymbol{R}_{22}\boldsymbol{V}_{22}^T = \boldsymbol{U}_{22}^T \boldsymbol{\Sigma}_2.$$

Apply the Weyl product inequalities for singular values [26, (7.3.14)] to the largest singular value of the matrices on both sides,

$$\frac{\sigma_1(\boldsymbol{R}_{22})}{\|\boldsymbol{V}_{22}^{-1}\|_2} = \sigma_1(\boldsymbol{R}_{22})\sigma_{p-k}(\boldsymbol{V}_{22}) \leq \sigma_1(\boldsymbol{R}_{22}\boldsymbol{V}_{22}^T) = \sigma_1(\boldsymbol{U}_{22}^T\boldsymbol{\Sigma}_2).$$

Because the orthogonal matrix $\boldsymbol{U}$ has all singular values equal to one,

$$\sigma_1(\boldsymbol{U}_{22}^T\boldsymbol{\Sigma}_2) \leq \sigma_1(\boldsymbol{U}_{22})\sigma_1(\boldsymbol{\Sigma}_2) \leq \sigma_1(\boldsymbol{U})\sigma_{k+1} = \sigma_{k+1}.$$

Since $\boldsymbol{V}_{11}$ is nonsingular, the CS decomposition [16, Theorem 2.5.3] implies that $\|\boldsymbol{V}_{11}^{-1}\|_2 = \|\boldsymbol{V}_{22}^{-1}\|_2$. Combining the extreme ends of the sequence of the above inequalities gives $\sigma_1(\boldsymbol{R}_{22})/\|\boldsymbol{V}_{11}^{-1}\|_2 \leq \sigma_{k+1}$. ∎

A.4. Proof of Theorem 4.4. We prove the correctness of Algorithm 4.4 (Lemma A.8), and present a proof of Theorem 4.4 (Lemma A.9).

Our proofs follow those in [18] but without the full rank assumption on the sensitivity matrix and with more details. To keep the proofs simple, we assume that the QR decompositions are implemented so that the upper triangular matrices have non-negative diagonal elements [16, Theorem 5.2.3].

We prove the correctness of stopping criterion of Algorithm 4.4, which depends on the row norms of $\boldsymbol{R}_{11}^{-1}$ and the column norms of $\boldsymbol{R}_{22}$,

$$\omega_i(\boldsymbol{R}_{11}) \equiv 1/\|\boldsymbol{e}_i^T \boldsymbol{R}_{11}^{-1}\|_2, \qquad 1 \leq i \leq k$$
$$\gamma_j(\boldsymbol{R}_{22}) \equiv \|\boldsymbol{R}_{22}\boldsymbol{e}_j\|_2, \qquad 1 \leq j \leq p-k.$$

LEMMA A.8 (Correctness of Algorithm 4.4). *Let*

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22} \end{bmatrix} \in \mathbb{R}^{p\times p}$$

be upper triangular with non-negative diagonal elements, nonsingular $\boldsymbol{R}_{11} \in \mathbb{R}^{k\times k}$, *and* $\boldsymbol{R}_{22} \in \mathbb{R}^{(p-k)\times(p-k)}$. *Let* $\boldsymbol{P}$ *be a permutation that permutes columns* i *and* $k+j$ *of* $\boldsymbol{R}$ *for some* $1 \le i \le k$ *and some* $1 \le j \le p-k$, *and let* $\boldsymbol{RP} = \widetilde{\boldsymbol{Q}}\widetilde{\boldsymbol{R}}$ *be an unpivoted QR decomposition with*

$$\widetilde{\boldsymbol{R}} = \begin{bmatrix} \widetilde{\boldsymbol{R}}_{11} & \widetilde{\boldsymbol{R}}_{12} \\ \boldsymbol{0} & \widetilde{\boldsymbol{R}}_{22} \end{bmatrix}.$$

Then

$$\rho_{ij} \equiv \frac{\det(\widetilde{\boldsymbol{R}}_{11})}{\det(\boldsymbol{R}_{11})} = \sqrt{(\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12})_{i,j}^2 + (\gamma_j(\boldsymbol{R}_{22})/\omega_i(\boldsymbol{R}_{11}))^2}.$$

Proof. We give the proof for the special case $i = k$ and $j = 1$, and first argue that this represents no loss of generality. Note that column j of $\boldsymbol{R}_{22}$ corresponds to column $k+j$ of $\boldsymbol{R}$.

Reduction to the case $i = k$ *and* $j = 1$. Suppose that $i < k$ and $j > 1$. Let $\boldsymbol{P}_{i,k}$ be the permutation that permutes columns i and k of $\boldsymbol{R}$, and let $\boldsymbol{R}_{11}\boldsymbol{P}_{i,k} = \bar{\boldsymbol{Q}}_{11}\bar{\boldsymbol{R}}_{11}$ be the unpivoted QR decomposition. Similarly, let $\boldsymbol{P}_{1,j}$ be the permutation that permutes columns $k+j$ and $k+1$ of $\boldsymbol{R}$, and let $\boldsymbol{R}_{22}\boldsymbol{P}_{1,j} = \bar{\boldsymbol{Q}}_{22}\bar{\boldsymbol{R}}_{22}$ be the unpivoted QR decomposition. With

$$\bar{\boldsymbol{R}}_{12} \equiv \bar{\boldsymbol{Q}}_{11}^T\boldsymbol{R}_{12}\boldsymbol{P}_{1,j}, \qquad \bar{\boldsymbol{P}} \equiv \begin{bmatrix} \boldsymbol{P}_{i,k} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}_{1,j} \end{bmatrix},$$

the matrix

$$\boldsymbol{R}\bar{\boldsymbol{P}} = \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22} \end{bmatrix}\begin{bmatrix} \boldsymbol{P}_{i,k} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}_{1,j} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}_{11}\boldsymbol{P}_{i,k} & \boldsymbol{R}_{12}\boldsymbol{P}_{1,j} \\ \boldsymbol{0} & \boldsymbol{R}_{22}\boldsymbol{P}_{1,j} \end{bmatrix}$$

has the unpivoted QR decomposition

$$\boldsymbol{R}\bar{\boldsymbol{P}} = \begin{bmatrix} \bar{\boldsymbol{Q}}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \bar{\boldsymbol{Q}}_{22} \end{bmatrix}\begin{bmatrix} \bar{\boldsymbol{R}}_{11} & \bar{\boldsymbol{R}}_{12} \\ \boldsymbol{0} & \bar{\boldsymbol{R}}_{22} \end{bmatrix}.$$

The assumption of non-negative diagonal elements in the upper triangular matrices implies $\det(\boldsymbol{R}_{11}) = \det(\bar{\boldsymbol{R}}_{11})$. From

$$\bar{\boldsymbol{R}}_{11}^{-1}\bar{\boldsymbol{R}}_{12} = (\bar{\boldsymbol{Q}}_{11}^T\boldsymbol{R}_{11}\boldsymbol{P}_{i,k})^{-1}(\bar{\boldsymbol{Q}}_{11}^T\boldsymbol{R}_{12}\boldsymbol{P}_{1,j}) = \boldsymbol{P}_{i,k}^T\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}\boldsymbol{P}_{1,j},$$

the invariance of the two-norm under multiplication by orthogonal matrices, and the non-negativity of the diagonal elements follows

$$|\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}|_{i,j} = |\bar{\boldsymbol{R}}_{11}^{-1}\bar{\boldsymbol{R}}_{12}|_{k,1}, \qquad \omega_i(\boldsymbol{R}_{11}) = \omega_k(\bar{\boldsymbol{R}}_{11}), \qquad \gamma_j(\boldsymbol{R}_{22}) = \gamma_1(\bar{\boldsymbol{R}}_{22}).$$

Thus, the relevant quantities do not change under permutations and subsequent QR decompositions.

Relevant quantities induced by the partitioning of upper triangular matrices. With $i = k$ and $j = 1$, distinguish[2] rows and columns k and $k+1$,

$$\boldsymbol{R} = \left[\begin{array}{c|c} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \hline & \boldsymbol{R}_{22} \end{array}\right] = \left[\begin{array}{cc|cc} \hat{\boldsymbol{R}}_{11} & \mathbf{a} & \mathbf{b} & \hat{\boldsymbol{R}}_{12} \\ & \omega & \beta & \mathbf{c}^T \\ \hline & & \gamma & \mathbf{d}^T \\ & & & \hat{\boldsymbol{R}}_{22} \end{array}\right]$$

where $\hat{\boldsymbol{R}}_{11} \in \mathbb{R}^{(k-1)\times(k-1)}$, $\hat{\boldsymbol{R}}_{12} \in \mathbb{R}^{(k-1)\times(p-k-1)}$, $\omega > 0$ and $\gamma > 0$. Upper triangularity implies the determinant relation

$$\det(\boldsymbol{R}_{11}) = \omega \det(\hat{\boldsymbol{R}}_{11}). \tag{A.20}$$

Looking at the trailing row of $\boldsymbol{R}_{11}^{-1}$ and the leading column of $\boldsymbol{R}_{22}$,

$$\boldsymbol{R}_{11}^{-1} = \begin{bmatrix} \hat{\boldsymbol{R}}_{11}^{-1} & -\frac{1}{\omega}\hat{\boldsymbol{R}}_{11}^{-1}\mathbf{a} \\ \mathbf{0} & \frac{1}{\omega} \end{bmatrix}, \qquad \boldsymbol{R}_{22} = \begin{bmatrix} \gamma & \mathbf{d}^T \\ \mathbf{0} & \hat{\boldsymbol{R}}_{22} \end{bmatrix},$$

gives

$$1/\|\boldsymbol{e}_k^T \boldsymbol{R}_{11}^{-1}\|_2 = \omega_k(\boldsymbol{R}_{11}) = \omega, \qquad \|\boldsymbol{R}_{22}\boldsymbol{e}_1\|_2 = \gamma_1(\boldsymbol{R}_{22}) = \gamma.$$

Element $(k, 1)$ of $\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}$ equals

$$(\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12})_{k,1} = \boldsymbol{e}_k^T \begin{bmatrix} \hat{\boldsymbol{R}}_{11}^{-1} & -\frac{1}{\omega}\hat{\boldsymbol{R}}_{11}^{-1}\mathbf{a} \\ \mathbf{0} & \frac{1}{\omega} \end{bmatrix} \begin{bmatrix} \mathbf{b} & \hat{\boldsymbol{R}}_{12} \\ \beta & \mathbf{c}^T \end{bmatrix} \boldsymbol{e}_1 = \begin{bmatrix} \mathbf{0} & \frac{1}{\omega} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \beta \end{bmatrix} = \frac{\beta}{\omega},$$

The action. Let $\boldsymbol{P}$ be the permutation that permutes columns k and $k+1$ of $\boldsymbol{R}$,

$$\boldsymbol{R}\boldsymbol{P} = \left[\begin{array}{cc|cc} \hat{\boldsymbol{R}}_{11} & \mathbf{b} & \mathbf{a} & \hat{\boldsymbol{R}}_{12} \\ & \beta & \omega & \mathbf{c}^T \\ \hline & \gamma & 0 & \mathbf{d}^T \\ & & & \hat{\boldsymbol{R}}_{22} \end{array}\right]$$

To return to upper triangular form, perform an unpivoted QR decomposition $\boldsymbol{R}\boldsymbol{P} = \widetilde{\boldsymbol{Q}}\widetilde{\boldsymbol{R}}$ that zeros out γ by rotating rows k and $k+1$. The resulting triangular matrix $\widetilde{\boldsymbol{R}}$ has a leading principal submatrix

$$\widetilde{\boldsymbol{R}}_{11} = \begin{bmatrix} \hat{\boldsymbol{R}}_{11} & \mathbf{b} \\ \mathbf{0} & \sqrt{\beta^2 + \gamma^2} \end{bmatrix}$$

with the determinant relation

$$\det(\widetilde{\boldsymbol{R}}_{11}) = \sqrt{\beta^2 + \gamma^2}\det(\hat{\boldsymbol{R}}_{11}).$$

Combine this with the old determinant relation (A.20)

$$\begin{aligned} \frac{\det(\widetilde{\boldsymbol{R}}_{11})}{\det(\boldsymbol{R}_{11})} &= \frac{\sqrt{\beta^2 + \gamma^2}}{\omega} = \sqrt{\left(\frac{\beta}{\omega}\right)^2 + \left(\frac{\gamma}{\omega}\right)^2} \\ &= \sqrt{(\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12})_{k,1}^2 + (\gamma_1(\boldsymbol{R}_{22})/\omega_k(\boldsymbol{R}_{11}))^2}. \end{aligned}$$

∎

[2] To increase readability, we sometimes use blank spaces to represent 0 elements.

The following proof of Theorem 4.4 relies on results from [25, section 3.3] and [18, section 3] but without the assumption that $\boldsymbol{S}$ has full column rank.

LEMMA A.9 (Proof of Theorem 4.4). *Let $\boldsymbol{S} \in \mathbb{R}^{n\times p}$ with $n > p$ have singular values $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$; and QR decomposition $\boldsymbol{S} = \boldsymbol{QR}$. Let $1 \leq k < p$ so that the leading $k \times k$ principal submatrix of $\boldsymbol{R}$ is non-singular. Then Algorithm 4.4 with $f \geq 1$ computes a QR decomposition*

$$\boldsymbol{SP} = \begin{bmatrix}\boldsymbol{Q}_1 & \boldsymbol{Q}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{R}_{11} & \boldsymbol{R}_{12}\\ \boldsymbol{0} & \boldsymbol{R}_{22}\end{bmatrix},$$

with singular values

$$\sigma_i(\boldsymbol{R}_{11}) \geq \frac{\sigma_i}{\sqrt{1+f^2k(p-k)}}, \qquad 1 \leq i \leq k,$$
$$\sigma_j(\boldsymbol{R}_{22}) \leq \sigma_{j+k}\sqrt{1+f^2k(p-k)}, \qquad 1 \leq j \leq p-k.$$

Additionally, the elements of $\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}$ are bounded by

$$|\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}|_{i,j} \leq f, \qquad 1 \leq i \leq k \; 1 \leq j \leq p-k.$$

Proof. We prove the inequality in the reverse order.

Third inequality. It follows from the observation that Algorithm 4.4 terminates once

$$|\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}|_{i,j} \leq \sqrt{|\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}|_{i,j}^2 + (\gamma_j(\boldsymbol{R}_{22})/\omega_i(\boldsymbol{R}_{11}))^2} \leq f$$

holds for $1 \leq i \leq k$, $1 \leq j \leq p-k$.

Second inequality. We scale the leading diagonal block so that it contains the k dominant singular values by $\alpha \equiv \sigma_1(\boldsymbol{R}_{22})/\sigma_k(\boldsymbol{R}_{11})$. Extract a judiciously scaled block-diagonal matrix

$$\boldsymbol{R}_D \equiv \begin{bmatrix}\alpha\boldsymbol{R}_{11} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}_{22}\end{bmatrix} = \underbrace{\begin{bmatrix}\boldsymbol{R}_{11} & \boldsymbol{R}_{12}\\ \boldsymbol{0} & \boldsymbol{R}_{22}\end{bmatrix}}_{\boldsymbol{R}}\underbrace{\begin{bmatrix}\alpha\boldsymbol{I}_k & -\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}\\ \boldsymbol{0} & \boldsymbol{I}_{p-k}\end{bmatrix}}_{\boldsymbol{W}}.$$

where $\alpha\boldsymbol{R}_{11}$ contains the k dominant singular values, because

$$\sigma_k(\alpha\boldsymbol{R}_{11}) = \alpha\sigma_k(\boldsymbol{R}_{11}) = \sigma_1(\boldsymbol{R}_{22}).$$

This means the largest singular value of $\boldsymbol{R}_{22}$ is equal to the smallest singular value of $\alpha\boldsymbol{R}_{11}$, thus less than or equal to all other singular values of $\alpha\boldsymbol{R}_{11}$. Therefore, the trailing block $\boldsymbol{R}_{22}$ contains the $p-k$ smallest singular values of $\boldsymbol{R}_D$.

The Weyl product inequality [26, (7.3.13)] implies

$$\sigma_{j+k}(\boldsymbol{R}_D) = \sigma_j(\boldsymbol{R}_{22}) \leq \sigma_{j+k}(\boldsymbol{R})\|\boldsymbol{W}\|_2, \qquad 1 \leq j \leq p-k. \tag{A.21}$$

We bound $\|\boldsymbol{W}\|_2$, by bounding the two-norm in terms of the Frobenius norm and, in

turn, expressing this as a sum,

$$\begin{aligned}\|\boldsymbol{W}\|_2^2 &\le 1 + \|\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}\|_2^2 + \alpha^2 \\ &= 1 + \|\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}\|_2^2 + \|\boldsymbol{R}_{22}\|_2^2\|\boldsymbol{R}_{11}^{-1}\|_2^2 \\ &\le 1 + \|\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}\|_F^2 + \|\boldsymbol{R}_{22}\|_F^2\|\boldsymbol{R}_{11}^{-1}\|_F^2 \\ &= 1 + \sum_{i=1}^{k}\sum_{j=1}^{p-k}\left((\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12})_{i,j}^2 + (\gamma_j(\boldsymbol{R}_{22})/\omega_i(\boldsymbol{R}_{11}))^2\right) \\ &\le 1 + \sum_{i=1}^{k}\sum_{j=1}^{p-k} f^2 \;\; = 1 + f^2 k(p-k).\end{aligned}$$

Now substitute $\|\boldsymbol{W}\|_2 \le \sqrt{1 + f^2 k(p-k)}$ into (A.21).

First inequality. If $\sigma_1(\boldsymbol{R}_{22}) = 0$, then $\boldsymbol{R}_{22} = \boldsymbol{0}$ and the first inequality holds.

Thus assume that $\sigma_1(\boldsymbol{R}_{22}) > 0$ so that $\alpha \equiv \sigma_1(\boldsymbol{R}_{22})/\sigma_k(\boldsymbol{R}_{11}) > 0$. Deriving a lower bound for the large singular values requires a slightly different ansatz. We scale the trailing block by $1/\alpha$ so that it contains the $p-k$ smallest singular values. Extract a differently scaled block-diagonal matrix,

$$\boldsymbol{R} = \begin{bmatrix}\boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22}\end{bmatrix} = \underbrace{\begin{bmatrix}\boldsymbol{R}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_{22}/\alpha\end{bmatrix}}_{\hat{\boldsymbol{R}}_D}\underbrace{\begin{bmatrix}\boldsymbol{I}_k & \boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12} \\ \boldsymbol{0} & \alpha\boldsymbol{I}_{p-k}\end{bmatrix}}_{\hat{\boldsymbol{W}}}.$$

where $\boldsymbol{R}_{22}/\alpha$ contains the $p-k$ subdominant singular values, because

$$\sigma_1(\boldsymbol{R}_{22}/\alpha) = \sigma_1(\boldsymbol{R}_{22})/\alpha = \sigma_k(\boldsymbol{R}_{11}).$$

This means the smallest singular value of $\boldsymbol{R}_{11}$ is equal to the smallest singular value of $\boldsymbol{R}_{22}/\alpha$, thus larger or equal to all other singular values of $\boldsymbol{R}_{22}/\alpha$. Therefore, the leading block $\boldsymbol{R}_{11}$ contains the k largest singular values of $\boldsymbol{R}_D$.

An analogous argument as above shows

$$\begin{aligned}\sigma_i(\boldsymbol{R}) &\le \sigma_i(\hat{\boldsymbol{R}}_D)\|\hat{\boldsymbol{W}}\|_2 = \sigma_i(\boldsymbol{R}_{11})\|\hat{\boldsymbol{W}}\|_2 \\ &\le \sigma_i(\boldsymbol{R}_{11})\sqrt{1 + f^2 k(p-k)}, \qquad 1 \le i \le k.\end{aligned}$$

∎

Appendix B. Supplemental Material. We present more details for the models in section 2.2: Epidemiological (section B.1), cardiovascular tissue (section B.2), fibrin polymerization (section B.3), and neurological (section B.4). All models are represented as coupled systems of ODEs (ordinary differential equations), and parameter sensitivities are determined from their numerical solution via complex-step or finite differences.

B.1. Epidemiological Models. We implemented five (nested) epidemiological compartment models in section 2.2 that represent COVID-19 spread among the US population for identifiability anaylsis of the model parameters.

Figure B.1 displays the different compartments associated with the state variables in each model, and the possible transitions from one infection status to another within a population [49, 39]. The parameters above the arrows represent the transition rates. From these diagrams, nonlinear ordinary differential equations for each system can be derived by analogy with leading-order mass action reaction kinetics.

For the SIR, SEIR, SVIR, and SEVIR models in Figure B.1, the quantity of interest is the number $I(t)$ of infectious individuals at time t; and for the COVID model it is $(A + I + H)(t)$.

We calibrated the models to the spread of COVID-19 through the US based on CDC data and relevant studies. Table B.1 describes the physical interpretation of each parameter and the average nominal value for generating sensitivities.

As outlined in §5.1, letting q_j^* represent the nominal value of the jth parameter in Table B.1, the algorithms were tested on 10,000 matrices for each model, evaluated at parameter vectors for which the jth component is sampled uniformly from the interval $[0.5q_j^*, 1.5q_j^*]$.

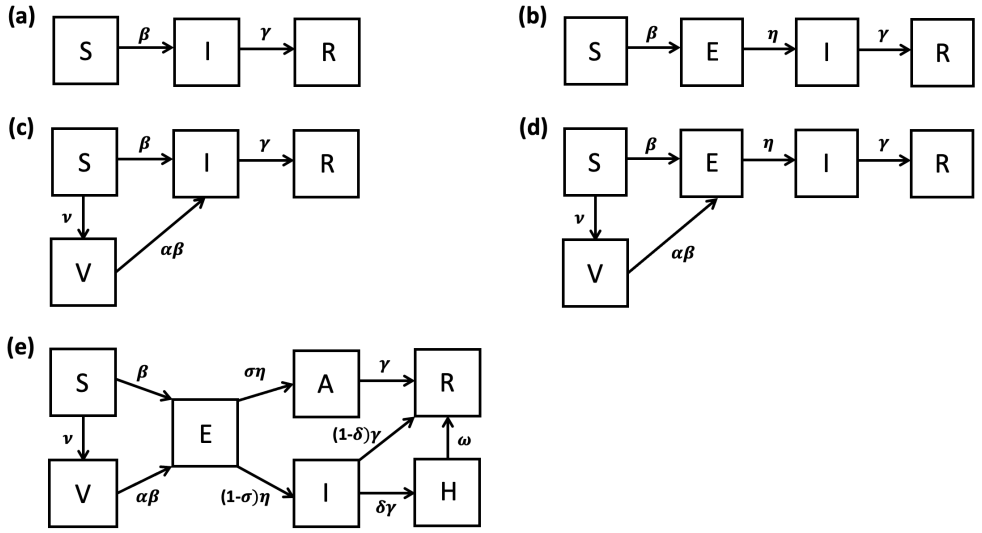

FIG. B.1. *Compartment diagrams for the epidemiological models in section 2.2 to illustrate the possible transitions from one infection statius to another within a population for (a) SIR, (b) SEIR, (c) SVIR, (d) SEVIR, and (e) COVID-19 models.*

Par	Mean	Description	Ref.
β	0.80	Transmission coefficient	[29]
η	0.33	Rate of progression to infectiousness (following exposure)	[30, 31]
γ	0.14	Rate of progression through infectious stage	[44, 37]
α	0.10	Probability of infection after vaccination	[39]
ν	0.004	Rate of vaccination	[48]
σ	0.35	Percentage of infected that are asymptomatic	[43]
δ	0.05	Rate of hospitalization for symptomatic infected	[14]
ω	0.82	Rate of recovery for hospitalized infections	[36]

TABLE B.1
Parameter values and physical interpretations in epidemiological models from §5.1

B.2. Cardiovascular Tissue Biomechanics (HGO) Model. The sensitivity matrix corresponding to this model arose from a nonlinear hyperelastic structural model of the vessel wall for a large pulmonary artery in the context of ex vivo

biomechanical experiments. A two-layer, anisotropic vessel wall model was developed, within the general framework of the Holzapfel-Gasser-Ogden (HGO) model [24], and systematically reduced with identifiabilty techniques rooted in the scaled sensitivity matrix. The quantity of interest was a hybrid normalized residual vector amalgamating data measuring lumen area and wall thickness changes with increasing fluid pressure.

This data set arises from ex vivo biomechanical testing of coupled flow and deformation for left pulmonary arteries excised from normal and hypertensive mice. This model contains 16 model parameters: 8 are fixed based on values in the literature or information from the experiments, while the remaining 8 are estimated via systematic model reduction in the context of an inverse problem [19].

Results of the systematic model reduction in [19] are consistent with the value $k = 5$ for HGO in Table 4.1.

B.3. Fibrin Polymerization Model for Wound Healing Applications. Motivated by a wound healing application, this is a biochemical reaction kinetics model for in vitro fibrin polymerization, mediated by the enzyme thrombin, and [38]. The 46×11 sensitivity matrix represents 46 time points for the concentration of fibrin matrix, i.e. in-vitro clots; and 11 parameters that represent reaction rates for the associated biochemical reaction species. The parameters are chosen from the last row of [38, Table 1] for a mathematical model of hemostasis, the first stage of wound healing during which fibrin (extracellular) matrix polymerization occurs.

The corresponding system of ODEs is based on first-order reaction kinetics, analogous to the mass-action assumptions for the epidemiological compartment models in section B.1. The systematic identifiability analysis and model reduction for the inverse problem in [38] are consistent with the value $k = 6$ under the "Wound" model in Table 5.1.

B.4. Neurological Model. This complex model consists of a system of nonlinear ODEs [20] that quantify the neurovascular coupling (NVC) response, and the local changes in vascular resistance due to neuronal activity [20]. The state variables represent different components of the human brain, while the parameters represent the ion channels and metabolic signalling among them.

The sensitivity matrix is the largest and most ill-conditioned sensitivity matrix in Table 5.1, along with the largest number of state variables, parameters $p = 175$, and observations $n = 200$.

Appendix C. Dynamical systems for adversarial CSS matrices.

Given an adversarial CSS matrix $\boldsymbol{S} \in \mathbb{R}^{n\times p}$ with $n \geq p$, we construct a dynamical system whose sensitivity matrix is identical to $\boldsymbol{S}$.

Let $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ have have a thin SVD as in (3.2) and distinguish the columns of the singular vector matrices,

$$\boldsymbol{U} = \begin{bmatrix}\mathbf{u}_1 & \dots & \mathbf{u}_p\end{bmatrix} \in \mathbb{R}^{n\times p}, \qquad \boldsymbol{V} = \begin{bmatrix}\mathbf{v}_1 & \dots & \mathbf{v}_p\end{bmatrix} \in \mathbb{R}^{p\times p}.$$

Pick some vector $\mathbf{q} \in \mathbb{R}^p$, we are going to construct a system of ODEs parameterized by $\mathbf{q}$.

To this end, let

$$\boldsymbol{\Lambda} \equiv \operatorname{diag}\begin{pmatrix}\lambda_1 & \cdots & \lambda_p\end{pmatrix} \in \mathbb{R}^{p\times p}$$

be a diagonal matrix yet to be specified. Denote by $\boldsymbol{x}(t) \in \mathbb{R}^p$ the state vector and

by $\boldsymbol{y}(t)$ the observation vector, and combine everything into the initial value problem

$$\begin{aligned} \frac{d\boldsymbol{x}}{dt} &= \boldsymbol{\Lambda}\boldsymbol{x}, \qquad \boldsymbol{x}(0) = \boldsymbol{V}^T\mathbf{q}, \\ \boldsymbol{y}(t) &= \boldsymbol{U}\boldsymbol{x}(t) \end{aligned} \tag{C.1}$$

with solution $\boldsymbol{x}(t) = \exp(t\boldsymbol{\Lambda})\boldsymbol{V}^T\mathbf{q}$. Since $\boldsymbol{\Lambda}$ is diagonal, the observation equals

$$\boldsymbol{y}(t) = \boldsymbol{U}\exp(t\boldsymbol{\Lambda})\boldsymbol{V}^T\mathbf{q} = \sum_{j=1}^{p} \mathbf{u}_j e^{t\lambda_j}\mathbf{v}_j^T\mathbf{q}.$$

As in section 2.2.1, differentiate $\boldsymbol{y}$ with respect to $\mathbf{q}$, and then evaluate at time $t = \tau > 0$. The rows of the resulting sensitivity matrix equal

$$\frac{\partial y_i}{\partial \mathbf{q}} = \sum_{j=1}^{p} u_{ij}\, e^{\tau\lambda_j}\, \mathbf{v}_j^T, \qquad 1 \le i \le n.$$

Set $\sigma_j = e^{\tau\lambda_j}$ so that $\lambda_j = \frac{1}{\tau}\ln\sigma_j$, $1 \le j \le p$. Then the sensitivity matrix at time $t = \tau$ is

$$\boldsymbol{S}(t;\mathbf{q}) = \sum_{j=1}^{p} \mathbf{u}_j\sigma_j\, \mathbf{v}_j^T = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T.$$

Therefore, the dynamical system (C.1) has the desired sensitivity matrix $\boldsymbol{S}$ at time τ.

REFERENCES

[1] R. Allen, T. Rieger, and C. Musante, *Efficient generation and selection of virtual populations in quantitative systems pharmacology models*, CPT Pharmacometrics Syst. Pharmacol., 5 (2016), pp. 140–146.

[2] P. Bloomingdale, S. Bakshi, C. Maass, C. Pichardo-Almarza, D. B. Yadav, P. van der Graaf, and N. Mehrotra, *Minimal brain PBPK model to support the preclinical and clinical development of antibody therapeutics for CNS diseases*, Journal of Pharmacokinetics and Pharmacodynamics, (2021).

[3] M. Burth, G. C. Verghese, and M. Vélez-Reyes, *Subset selection for improved parameter estimation in on-line identification of a synchronous generator*, IEEE Trans. Power Syst., 14 (1999), pp. 218–225.

[4] P. A. Businger and G. H. Golub, *Linear least squares solutions by Householder transformations*, Numer. Math., 7 (1965), pp. 269–76.

[5] T. F. Chan, *Rank revealing QR factorizations*, Linear Algebra Appl., 88/89 (1987), pp. 67–82.

[6] T. F. Chan and P. C. Hansen, *Computing truncated singular value decomposition least squares solutions by rank revealing QR-factorizations*, SIAM J. Sci. Statist. Comput., 11 (1990), pp. 519–530.

[7] S. Chandrasekaran and I. C. F. Ipsen, *On rank-revealing QR factorisations*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 592–622.

[8] S. Chatterjee and A. S. Hadi, *Influential observations, high leverage points, and outliers in linear regression*, Statist. Sci., 1 (1986), pp. 379–393.

[9] A. Cintrón-Arias, H. T. Banks, A. Capaldi, and A. L. Lloyd, *A sensitivity matrix based methodology for inverse problem formulation*, J. Inverse Ill-Posed Probl., 17 (2009), pp. 545–564.

[10] P. Constantine, *Active Subspaces: Emerging Ideas for Dimension Reduction in Parametric Studies*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2015.

[11] D. L. Donoho and X. Huo, *Uncertainty principles and ideal atomic decomposition*, IEEE Trans. Inform. Theory, 47 (2001), pp. 2845–2862.

[12] P. Drineas and I. C. F. Ipsen, *Low-rank approximations do not need a singular value gap*, SIAM J. Matrix Anal. Appl., 40 (2019), pp. 299–319.

[13] L. V. Foster, *Rank and null space calculations using matrix decomposition without column interchanges*, Linear Algebra Appl., 74 (1986), pp. 47–71.

[14] S. Garg, L. Kim, and M. Whitaker, *Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019,* COVID-NET*, 14 states, March 1–30*, MMWR Morb Mortal Wkly Rep, 69 (2020), pp. 458–464.
[15] G. H. Golub, V. Klema, and G. W. Stewart, *Rank degeneracy and least squares problems*, Tech. Rep. STAN-CS-76-559, Computer Science Department, Stanford University, August 1976.
[16] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, fourth ed., 2013.
[17] W. B. Gragg and G. W. Stewart, *A stable variant of the secant method for solving nonlinear equations*, SIAM J. Numer. Anal., 13 (1976), pp. 889–903.
[18] M. Gu and S. C. Eisenstat, *Efficient algorithms for computing a strong rank-revealing QR factorization*, SIAM J. Sci. Comput., 17 (1996), pp. 848–869.
[19] M. Haider, K. Pearce, N. Chesler, N. Hill, and M. Olufsen, *Application of the HGO model to capturing in vitro relationships between pressure, area, and wall thickness in murine left pulmonary arteries*, Biomechanics and Modeling in Mechanobiology (under review), (2022).
[20] J. L. Hart, P. A. Gremaud, and T. David, *Global sensitivity analysis of high dimensional neuroscience models: An example of neurovascular coupling*, Bull. Math. Biol., 81 (2019), pp. 1805–1828.
[21] N. Heavner, C. Chen, A. Gopal, and M. P.-G., *Efficient algorithms for computing rank-revealing factorizations on a GPU*, 2021.
[22] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, second ed., 2002.
[23] D. C. Hoaglin and R. E. Welsch, *The Hat matrix in regression and ANOVA*, Amer. Statist., 32 (1978), pp. 17–22.
[24] G. Holzapfel, T. Gasser, and R. Ogden, *A new constitutive framework for arterial wall mechanics and a comparative study of material models*, Journal of Elasticity, 61 (2000), pp. 1–48.
[25] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.
[26] ———, *Matrix Analysis*, Cambridge University Press, Cambridge, second ed., 2013.
[27] I. T. Jolliffe, *Discarding variables in a principal component analysis. I. Artificial data*, J. Roy. Statist. Soc. Ser. C, 21 (1972), pp. 160–173.
[28] W. Kahan, *Numerical linear algebra*, Canadian Math. Bull., 9 (1966), pp. 757–801.
[29] R. Ke, E. R. Severson, S. Sanche, and N. Hengartner, *Estimating the reproductive number* R_0 *of* SARS-CoV-2 in the United States and eight European countries *and implications for vaccination*, Journal of Theoretical Biology, 517 (2021), p. 110621.
[30] C. M. A. d. O. Lima, *Information about the new coronavirus disease* (COVID-19), 2020.
[31] Z. Liu, P. Magal, O. Seydi, and G. Webb, *A covid-19 epidemic model with latency period*, Infectious Disease Modelling, 5 (2020), pp. 323–337.
[32] D. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, Inc, 1969.
[33] J. Lyness, *Numerical algorithms based on the theory of complex variables*, In Proc. ACM 22nd Nat. Conf., 4 (1967), pp. 124–134.
[34] J. Lyness and C. Moler, *Numerical differentiation of analytic functions*, SIAM J. Numer. Anal., 4(2) (1967), pp. 202–210.
[35] H. Miao, X. Xia, A. S. Perelson, and H. Wu, *On identifiability of nonlinear ODE models and applications in viral dynamics*, SIAM Rev., 53 (2011), pp. 3–39.
[36] N. T. Nguyen, J. Chinn, J. Nahmias, S. Yuen, K. A. Kirby, S. Hohmann, and A. Amin, *Outcomes and mortality among adults hospitalized with COVID-19 at US medical centers*, JAMA Network Open, 4 (2021).
[37] M. Park, A. Cook, J. Lim, Y. Sun, and B. Dickens, *A systemic review of* COVID-19 *epidemiology based on current evidence*, Journal of Clinical Medicine, 9 (2020), p. 967.
[38] K. J. Pearce, K. Nellenbach, R. C. Smith, A. C. Brown, and M. A. Haider, *Modeling and parameter subset selection for fibrin polymerization kinetics with applications to wound healing*, Bull. Math. Biol., 83 (2021), pp. Paper No. 47, 22.
[39] T. Perkins and G. España, *Optimal control of the COVID-19 pandemic with non-pharmaceutical interventions*, Bull. Math. Biol., 82 (2020), p. 118.
[40] ———, *Optimal control of the covid-19 pandemic with non-pharmaceutical interventions*, Bull. Math. Biol., 82 (2020), p. 118.
[41] T. Quaiser and M. Mönnigmann, *Systematic identifiability testing for unambiguous mechanistic modeling: Application to JAK-STAT, MAP kinase, and NF-κB signaling pathway models*, BMC Syst. Biol., 3 (2009), pp. 50–71.

[42] J. G. Reid, *Structural identifiability in linear time-invariant systems*, IEEE Trans. Automat. Control, 22 (1977), pp. 242–246.
[43] P. Sah, M. C. Fitzpatrick, C. F. Zimmer, E. Abdollahi, L. Juden-Kelly, S. M. Moghadas, B. H. Singer, and A. P. Galvani, *Asymptomatic sars-cov-2 infection: A systematic review and meta-analysis*, Proceedings of the National Academy of Sciences, 118 (2021).
[44] S. Sanche, Y. Lin, C. Xu, E. Romero-Severson, N. Hengartner, and R. Ke, *High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2*, Emerging Infectious Diseases, 26 (2020), pp. 1470–1477.
[45] K. Schmidt and R. Smith, *A parameter subset selection algorithm for mixed-effect models*, Int. J. Uncertain. Quantif., 6 (2016), pp. 1–12.
[46] D. C. Sorensen and M. Embree, *A DEIM induced CUR factorization*, SIAM J. Sci. Comput., 38 (2016), pp. A1454–A1482.
[47] G. W. Stewart, *The efficient generation of random orthogonal matrices with an application to condition estimators*, SIAM J. Numer. Anal., 17 (1980), pp. 403–409.
[48] Center for Disease Control, *http://covid.cdc.gov/covid-data-tracker, accessed 2021-08-20*.
[49] N. Tuncer and T. Le, *Structural and practical identifiability analysis of outbreak models*, Math. Biosci., 299 (2018), pp. 1–18.
[50] P. F. Velleman and R. E. Welsch, *Efficient computing of regression diagnostics*, Amer. Statist., 35 (1981), pp. 234–242.