The Structured Distance to the Nearest System Without Property $\mathcal{P}$

Scott C. Johnson, *Student Member, IEEE*, Mark Wicks, *Senior Member, IEEE*, Miloš Žefran, *Senior Member, IEEE*, Raymond A. DeCarlo, *Fellow, IEEE*

***Abstract*—For a system matrix M, this paper explores the smallest (Frobenius) norm additive structured perturbation δM for which a system property $\mathcal{P}$ (e.g. controllability, observability, stability, etc.) fails to hold, i.e., δM is the structured perturbation with smallest Frobenius norm such that there exists a property matrix $R \in \mathcal{P}$ for which $M - \delta M - R$ drops rank. The Frobenius norm is used because of its direct dependence on the magnitude of each entry in the perturbation matrix. Necessary conditions on a locally minimum norm structured rank-reducing perturbation δM and associated property matrix R are set forth and proven. An iterative algorithm is also set forth that computes a locally minimum norm structured perturbation and associated property matrix satisfying the necessary conditions. Algorithm convergence is proven using a discrete Lyapunov function.**

***Index Terms*—Matrix perturbation theory, structured perturbation, robust control, linear systems, computational methods.**

I. Introduction

System properties, such as controllability and observability, are often characterized by binary labels, e.g., controllable or uncontrollable and stable or unstable. These binary labels fail to capture the robustness of these properties. For example, consider the LTI system

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. The pair (A, B) is controllable if and only if for each $\lambda \in \mathbb{C}$

$$\text{rank}\begin{bmatrix} A - \lambda I_n & B \end{bmatrix} = n, \tag{2}$$

where I_n denotes the $n \times n$ identity matrix [1]. The set of uncontrollable pairs $(A, B) \in \mathbb{R}^{n\times(n+m)}$, i.e., pairs failing to satisfy (2), is an algebraic variety of lower dimension and hence has measure zero in $\mathbb{R}^{n\times(n+m)}$. Since LTI models are only approximations of physical systems, it is also necessary to characterize the robustness of the controllability property, e.g., by determining the distance to the nearest uncontrollable system [2], [3]

$$\mu_{\mathbb{R}}(A, B) = \inf_{(\delta A,\delta B)\in\mathcal{C}} \|[\delta A, \delta B]\|_F, \tag{3}$$

where $\mathcal{C} = \{(\delta A, \delta B) : \exists\lambda \in \mathbb{C}, \text{rank}[A-\delta A-\lambda I_n, B-\delta B] < n\}$. Similar metrics can be constructed for system properties including, but not limited to reachability, stabilizability, observability, and detectability.

Computing metrics such as $\mu_{\mathbb{R}}(A, B)$ and the associated minimizing perturbations has been an active area of research over the past 40 years [2]–[16]. The norm used to measure robustness separates the robust system property literature. The Frobenius norm metric in (3) is based on the work of [2] and is used in [3]–[9]. The primary alternative to the Frobenius norm metric is the spectral norm, i.e., the largest singular value of the matrix $[\delta A, \delta B]$. The spectral norm metric, usually referred to by the names controllability radius or observability radius, is explored in several works including [11]–[16]. This paper utilizes a Frobenius norm metric because a perturbation on each entry of a system matrix affects the Frobenius norm in a strong and direct way.

The primary challenge to either robustness metric is developing an algorithm to compute the minimum distance and associated perturbation matrices. In [3], the algorithm for computing (3) for real but otherwise unstructured perturbations is based on computing a coordinate transformation into a "nearly" Kalman uncontrollable form. Another approach for computing (3) is considered in [8] wherein one constructs a large $n(n + 1) \times n(n + m)$ matrix X_{n-1} consisting of a structured arrangement of blocks of matrices

$$\begin{bmatrix} A & B \\ I & 0 \end{bmatrix}.$$

The "Structured Total Least Norm" algorithm then computes a low rank approximation to X_{n-1} where only the A and B matrices are potentially perturbed. The low rank approximation also provides the smallest perturbations $\delta A \in \mathbb{R}^{n\times n}$ and $\delta B \in \mathbb{R}^{n\times m}$ causing uncontrollability.

Reference [11] develops an algorithm for computing the controllability radius for real but otherwise unstructured perturbations utilizing a constrained optimization problem; the perturbations causing uncontrollability are constructed from singular vectors. In [13], a fast algorithm for computing the controllability radius is developed for unstructured complex perturbations. Extensions to higher-order LTI systems with affine perturbations are considered in [14]. Additional extensions including descriptor and time-delay LTI systems are

Manuscript received July 12, 2016; revised July 18, 2016, March 23, 2017, and September 11, 2017; accepted November 15, 2017. Recommended by Associate Editor W. Michiels. (*Corresponding author: Miloš Žefran*). The work of M. Žefran has been partially supported by NSF grants IIS-1705058 and CMMI-1762924.

S. Johnson and R. DeCarlo are with the Department of Electrical and Computer Engineering at Purdue University, West Lafayette, IN 47907, USA (email: johns924@purdue.edu, decarlo@ecn.purdue.edu).

M. Wicks is with Xometry, Bethesda, MD 20814, USA (email: mawicks@gmail.com).

M. Žefran is with the Department of Electrical and Computer Engineering at the University of Illinois at Chicago, Chicago, IL 60607, USA (email: mzefran@uic.edu).

considered in [16]. Finally, [15] develops an upper bound on the spectral distance to uncontrollability of a switched LTI system.

In [17], [18], a class of low rank approximation problems is studied. As a special case, distance to an uncontrollable space is treated in the context of input-output descriptions which are invariant with regard to a particular realization. While there is similarity with the work described here, our problem setup explicitly includes structure on the perturbation and structure on the property and is thus inherently different. On the other hand, [17] deals with problems of arbitrary rank while we only consider rank deficiency of 1.

In the context of distance to uncontrollability, [7] extends the formulas of [2] to account for structured real perturbations having a particular product structure (see Equation (2) in [7]). Alternatively, reference [19] and the work herein, recasts this problem and generalizes it to determine when a system subject to a structured perturbation looses a system property such as controllability, stability, gain margin, rise time levels, etc. Specifically, [19] formulates the problem of structured rank reducing perturbations, belonging to a subspace $\mathcal{S} \subset \mathbb{C}^{n\times m}$, on a rectangular matrix $M \in \mathbb{C}^{n\times m}$ which cause the failure of a system property, $\mathcal{P}$. This general $\mathcal{P}$–robustness framework encompasses many of the robustness problems previously addressed in the literature. No prior work has extended the $\mathcal{P}$–robustness framework, proven the necessary conditions for $\mathcal{P}$–robustness, or completed and proven convergence of the algorithm suggested in [19]; this list constitutes the main contributions of our work.

Section II introduces the $\mathcal{P}$–robustness framework. Section III establishes necessary conditions for solving the $\mathcal{P}$–robustness problem (Theorem 1). They motivate the solution algorithm in Section IV (Algorithm 1). Section V shows that the algorithm converges to a point satisfying the necessary conditions (Theorem 2). Numerical examples follow in Section VI.

In this paper, the following notation will be used:

dim($\mathcal{S}$)	Dimension of a linear space $\mathcal{S}$.
$\|\|M\|\|_F$	Frobenius norm of a matrix M.
$\|\|\nu\|\|_2$	Euclidean norm of a vector ν.
$M^\top$, M^H	Transpose and conj. transpose of M.
$\sigma_n(M)$	n^{th} singular value of matrix M.
I_m	$m \times m$ identity matrix.
diag(ν)	Diag. matrix with diag. entries given by ν.
vec(M)	Vectorizes M by stacking the columns.
$M \otimes N$	The kronecker product of M and N.
$M^\dagger$	The Moore-Penrose pseudoinverse.
col-sp(M)	Column space of M.
Im(M)	Imaginary component of M.
Re(M)	Real component of M.
$\sigma_i(M)$	i^{th} largest singular value of M.
cl(U)	The closure of the set U.
$\langle A_1, A_2\rangle$	Inner product of $A_1, A_2 \in \mathbb{C}^{n\times m}$ defined as $\langle A_1, A_2\rangle = \text{Re}(\text{vec}(A_1))^\top \text{Re}(\text{vec}(A_2)) + \text{Im}(\text{vec}(A_1))^\top \text{Im}(\text{vec}(A_2))$

II. $\mathcal{P}$–Robustness Problem

Observe that in the problem of controllability (and by duality observability) described above, there is a matrix $M = [A, B]$, a perturbation matrix $\delta M = [\delta A, \delta B]$ with a particular structure described by some basis in a space of possible perturbations $\mathcal{S}$, and a property matrix $P = [\lambda I, 0]$, a complex matrix living in a property space $\mathcal{P}$ also with an underlying basis. So the problem illustrated above is to find the smallest normed perturbation for which $[A - \delta A, B - \delta B]$ does not have the controllability property, i.e.,

$$\text{rank}([A - \delta A, B - \delta B] - [\lambda I, 0]) < n$$

over the space of purturbations $\mathcal{S}$ and the space of property matrices $\mathcal{P}$. In several other problems, a distance to a region $\mathcal{D}$ can be reduced to a distance to its boundary, $\partial\mathcal{D}$. Problems of this sort include:

- Asymptotic Stability: Find the smallest structured perturbation δA such that $\text{rank}([A-\delta A]-[j\omega I]) < n$ for some $\omega \in \mathbb{R}$.
- Damping Ratio greater than ζ: Find the smallest structured perturbation δA such that $\text{rank}([A-\delta A]-[\omega(-\zeta + j\sqrt{1-\zeta^2})I]) < n$ for some $\omega \in \mathbb{R}$.
- Prescribed Pole Locations: Given a polygonal region of the complex plane, $\mathcal{D}$, the system has all poles in the interior of $\mathcal{D}$ if and only if $\text{rank}[A - \lambda I] = n$ for all $\lambda \in \mathbb{C}$ but not in the interior of $\mathcal{D}$.

In each of the above problems, there is a matrix M, a perturbation matrix δM in a linear space $\mathcal{S}$ having a basis, and a property matrix P also in a linear space $\mathcal{P}$ with an underlying basis, for which one seeks the smallest norm matrix δM such that $M - \delta M$ does not have the pertinent property. This leads to the following definition:

Definition 1. *[19] Let $M \in \mathbb{C}^{n\times m}$ with $n \leq m$ (without loss of generality), and $\mathcal{P} \subset \mathbb{C}^{n\times m}$ and $\mathcal{S} \subset \mathbb{C}^{n\times m}$ be linear spaces over $\mathbb{R}$. The $\mathcal{P}$–robustness of M with respect to parameter variations in $\mathcal{S}$ is defined as*

$$r(M; \mathcal{S}, \mathcal{P}) = \inf_{\delta M \in \mathcal{T}} \|\delta M\|_F \tag{4}$$

where

$$\mathcal{T} = \{\delta M \in \mathcal{S} : \exists R \in \mathcal{P}, \text{rank}[M - \delta M - R] < n\}. \tag{5}$$

As mentioned, the Frobenius norm metric directly measures the magnitude of the parameter variations and thus appears to more accurately represent the robustness of the system property. This is in contrast to the controllability (and observability) radius which measures the largest singular value of the perturbation causing uncontrollability (unobservability), a metric that may not reflect some parameter variations: for a fixed largest singular value, changes in the smaller singular values due to parameter variations go unnoticed.

It is useful to consider bases for $\mathcal{S}$ and $\mathcal{P}$ (which are linear subspaces over the field $\mathbb{R}$). Let $\{S_1, S_2, \cdots, S_k\}$ be an orthonormal basis for $\mathcal{S}$ and $\{P_1, P_2, \cdots, P_r\}$ be an orthonormal basis for $\mathcal{P}$, where by orthonormal we mean that $\langle S_i, S_j\rangle$ is 0 if $i \neq j$ and 1 if $i = j$. Each perturbation $\delta M \in \mathcal{S}$ can be represented by an associated vector $\zeta \in \mathbb{R}^k$ in this basis $\{S_1, S_2, \cdots, S_k\}$, i.e., $\delta M = \sum_{i=1}^k \zeta_i S_i$. Similarly, each $R \in \mathcal{P}$ is represented by a vector $\rho \in \mathbb{R}^r$. Using the fixed bases for $\mathcal{S}$ and $\mathcal{P}$, we can reformulate the $\mathcal{P}$–robustness of M with respect to perturbations in $\mathcal{S}$.

Definition 2. *Let $M \in \mathbb{C}^{n\times m}$; let $\mathcal{P} \subset \mathbb{C}^{n\times m}$ and $\mathcal{S} \subset \mathbb{C}^{n\times m}$ be linear spaces over $\mathbb{R}$ with orthonormal bases $\{S_1, S_2, \cdots, S_k\}$ and $\{P_1, P_2, \cdots, P_r\}$, respectively. Given $\zeta \in \mathbb{R}^k$ and $\rho \in \mathbb{R}^r$, let*

$$M(\zeta, \rho) = M - \sum_{i=1}^{k} \zeta_i S_i - \sum_{i=1}^{r} \rho_i P_i. \tag{6}$$

The $\mathcal{P}$–robustness of M with respect to parameter variations in $\mathcal{S}$ is

$$r(M; \mathcal{S}, \mathcal{P}) = \inf_{\zeta\in\mathbb{R}^k,\ \rho\in\mathbb{R}^r} \|\zeta\|_2 \tag{7}$$

subject to:

$$0 = \sigma_n\left(M(\zeta, \rho)\right) \triangleq H(\zeta, \rho). \tag{8}$$

Note, Definition 2 is equivalent to Definition 1. Also, the $\mathcal{P}$–robustness of M with respect to parameter variations in $\mathcal{S}$ is independent of the choice of orthonormal basis, although of course, the minimizing pair (ζ_*, ρ_*) depends on the choice.

For use later in the paper, we define

$$f(\zeta, \rho) = 0.5\|\zeta\|_2^2. \tag{9}$$

If $\|\zeta\|_2$ is replaced by $f(\zeta, \rho)$ in (7), the minimizer (when it exists) does not change. The function $f(\zeta, \rho)$ is preferable because it simplifies proofs later in the paper.

Example 1. *Applying the $\mathcal{P}$–robustness formulation to controllability of an LTI state model (A, B, C), we set $M = [A, B]$, $R = [\lambda I, 0]$, and $\delta M = [\delta A, \delta B]$, where δM has a specific perturbation structure defined by a basis for $\mathcal{S}$.*

The original motivation for this work stems from the need for specific structured real perturbations for the state and mode sequence (SMS) observability problem of switched LTI (SLTI) systems with safety applications described in [9], [20].

Example 2. *For the problem of computing the distance to the nearest SMS unobservable SLTI system (which can model transitions from safe to unsafe operation), the $\mathcal{P}$–robustness framework can be applied to each pair of modes i (safe) and j (unsafe) by defining*

$$M_{ij} = \begin{bmatrix} A_i^\top & 0 & C_i^\top \\ 0 & A_j^\top & C_j^\top \end{bmatrix} \in \mathbb{R}^{2n\times(2n+p)}$$

$$\delta M_{ij} = \begin{bmatrix} \delta A_i^\top & 0 & \delta C_i^\top \\ 0 & \delta A_j^\top & \delta C_j^\top \end{bmatrix} \in \mathbb{R}^{2n\times(2n+p)}$$

$$R = \begin{bmatrix} \lambda I & 0 & 0 \\ 0 & \lambda I & 0 \end{bmatrix} \in \mathbb{C}^{2n\times(2n+p)}.$$

Clearly, the perturbation δM_{ij} has a specialized structure that is problematic for most existing approaches. The $\mathcal{P}$–robustness of M_{ij} with respect to parameter variations $\delta M_{ij} \in \mathcal{S}$ provides $r_{ij} \triangleq r(M_{ij}; \mathcal{S}, \mathcal{P})$, see (4). Then $\min_{i,j}\{r_{ij}\}$ is exactly the distance to the nearest SMS unobservable SLTI system.

The rank reduction in the $\mathcal{P}$–robustness problem is characterized by the n^{th} singular value of $M - \delta M - R$ becoming zero. We observe that the set of perturbation and property matrices δM and R for which $M - \delta M - R$ has the two smallest singular values equal to zero is an algebraic variety of lower dimension in $\mathcal{S} \times \mathcal{P}$. In other words, generically, when the n^{th} singular value of $M - \delta M - R$ is zero, the $(n-1)^{th}$ is different from zero. In this paper we thus focus on this most common problem structure. Also, see the discussion after Assumption 3 in Section V.

To analyze the n^{th} singular value, we define the following linear operator:

Definition 3. *Each pair $u \in \mathbb{C}^n$ and $V \in \mathbb{C}^{m\times(m-n+1)}$ induces a linear operator $L_{uV} : \mathbb{C}^{n\times m} \to \mathbb{C}^{1\times(m-n+1)}$ given by*

$$L_{uV}(N) = u^H N V. \tag{10}$$

Proposition 1. *Let $N = \widehat{U}\widehat{\Sigma}\widehat{V}^H$ be a singular value decomposition (SVD) of N. Define u to be the last column of $\widehat{U}$ and V to be the last $m - n + 1$ columns of $\widehat{V}$. Then*

$$L_{uV}(N) = \begin{bmatrix} \sigma_n(N) & 0 & \cdots & 0 \end{bmatrix}.$$

Consequently, $\|L_{uV}(N)\|_F = \sigma_n(N)$.

The linear operator L_{uV} is defined for any u and V, independent of the argument. For example, u and V can be related to the SVD of a matrix $M - \delta M - R$ and operate on any matrix $N \in \mathbb{C}^{n\times m}$. Since the perturbations and property matrices belong to lower dimensional subspaces $\mathcal{S}$ and $\mathcal{P}$, we define additional linear operators that have domains restricted to these subspaces.

Definition 4. *For $u \in \mathbb{C}^n$ and $V \in \mathbb{C}^{m\times(m-n+1)}$, the linear operators $L_{uV\mathcal{S}} : \mathcal{S} \to \mathbb{C}^{1\times(m-n+1)}$ and $L_{uV\mathcal{P}} : \mathcal{P} \to \mathbb{C}^{1\times(m-n+1)}$ are defined as*

$$L_{uV\mathcal{S}}(\delta M) \triangleq L_{uV}|_{\mathcal{S}}(\delta M) = u^H \delta M V \tag{11}$$

$$L_{uV\mathcal{P}}(R) \triangleq L_{uV}|_{\mathcal{P}}(R) = u^H R V. \tag{12}$$

On occasion, we will replace V with a vector $v \in$ col-sp$(V) \subset \mathbb{C}^m$. Specifically, we define $L_{uv\mathcal{S}} : \mathcal{S} \to \mathbb{C}$ and $L_{uv\mathcal{P}} : \mathcal{P} \to \mathbb{C}$ as $L_{uv\mathcal{S}}(\delta M) = u^H \delta M v$ and $L_{uv\mathcal{P}}(R) = u^H R v$.

The distinctions of the operator domains are pertinent when considering the pseudoinverses $L_{uV\mathcal{S}}^\dagger : \mathbb{C}^{1\times(m-n+1)} \to \mathcal{S}$ and $L_{uV\mathcal{P}}^\dagger : \mathbb{C}^{1\times(m-n+1)} \to \mathcal{P}$. The map $L_{uV\mathcal{S}}$ is surjective if for each $y \in \mathbb{C}^{1\times(m-n+1)}$ there exists $\delta M \in \mathcal{S}$ such that $L_{uV\mathcal{S}}(\delta M) = y$. When $L_{uV\mathcal{S}}$ is surjective, the pseudoinverse $L_{uV\mathcal{S}}^\dagger(y) = \delta M$ is the matrix $\delta M \in \mathcal{S}$ with the smallest Frobenius norm solving the equation $L_{uV}(\delta M) = y$.

Fundamental to the solution of the $\mathcal{P}$–robustness problem is the surjectivity of maps $L_{uV\mathcal{S}}$ as per the following assumption:

Assumption 1. *Let $\delta M \in \mathcal{S}$ and $R \in \mathcal{P}$. Let $M - \delta M - R$ have a SVD $\widehat{U}\widehat{\Sigma}\widehat{V}^H$. Define u to be the n^{th} column of $\widehat{U}$ and V to be the last $m-n+1$ columns of $\widehat{V}$ and let $L_{uV\mathcal{S}}$ be as given in Definition 4. While the SVD $\widehat{U}\widehat{\Sigma}\widehat{V}^H$, and therefore the operator $L_{uV\mathcal{S}}$, may not be unique, we assume that for each choice of $\delta M \in \mathcal{S}$ and $R \in \mathcal{P}$, there exists an appropriate SVD $\widehat{U}\widehat{\Sigma}\widehat{V}^H$ for which $L_{uV\mathcal{S}}$ is surjective on $\mathcal{S}$.*

Observe that we just use one column of $\widehat{U}$, even when the associated singular value is repeated. In any case, surjectivity of $L_{uV\mathcal{S}}$ ensures an improvement direction can be found for the $\mathcal{P}$–robustness problem when not at the minimum. Repeated singular values normally correspond to a point where the singular values cross. If the smallest singular values are repeated at one iteration of our proposed algorithm, in the subsequent iteration the minimum singular value will typically be simple (not repeated) unless the problem structure constrains the two minimum singular values to be equal throughout $\mathcal{P}$ and $\mathcal{S}$. When the minimum singular value is repeated, the additional freedom in choosing u and V provides more options for finding a pair that achieves surjectivity (see Example 4 in Appendix C).

The following proposition provides further insight into surjectivity of $L_{uV\mathcal{S}}$[1].

Proposition 2. *Let $u \in \mathbb{C}^n$, and $V \in \mathbb{C}^{m\times(m-n+1)}$. Let $\mathcal{S}$ have basis $\{S_1, S_2, \ldots, S_k\}$. $L_{uV\mathcal{S}}$ is surjective if and only if*

$$\text{rank}\left(\begin{bmatrix}\text{Re}[(V^\top \otimes u^H)B_{\mathcal{S}}]\\ \text{Im}[(V^\top \otimes u^H)B_{\mathcal{S}}]\end{bmatrix}\right) = 2(m-n+1) \tag{13}$$

where $B_{\mathcal{S}} = [\text{vec}(S_1), \text{vec}(S_2), \ldots, \text{vec}(S_k)]$.

Proof. See Appendix A. □

Clearly, $B_{\mathcal{S}}$ must have at least $2(m-n+1)$ columns for (13) to be satisfied, i.e., $\mathcal{S}$ as a vector space over $\mathbb{R}$ must have dimension no less than $2(m-n+1)$. Consequently, for the problem to be solvable, we require that the perturbation space $\mathcal{S}$ be sufficiently rich.

As described in [19], the surjectivity of $L_{uV\mathcal{S}}$ ensures a certain regularity condition on a rank reducing perturbation/property matrix pair $(\delta M, R) \in \mathcal{S}\times\mathcal{P}$. This regularity condition guarantees that there are neighboring perturbation/property matrix pairs $(\delta M', R')$ which are also rank reducing, i.e., δM is not an isolated rank reducing perturbation. It is worth pointing out that isolated rank reducing perturbations can be computed algebraically. Finally, we exclude $\mathcal{P}$–robustness problems where $\text{rank}(M-R) < n$ for some $R \in \mathcal{P}$ (i.e., M does not have property $\mathcal{P}$) since $r(M;\mathcal{S},\mathcal{P}) = 0$ in this case.

III. Necessary Conditions

This section states and proves the necessary conditions on a minimum norm rank-reducing perturbation $\delta M_* \in \mathcal{S}$ and the associated property matrix $R_* \in \mathcal{P}$ (when they exist), i.e., $\|\delta M_*\|_F = r(M;\mathcal{S},\mathcal{P})$ and $\text{rank}(M-\delta M_*-R_*) < n$. We next provide intuitive development of the necessary conditions. Suppose that at the solution, $\text{rank}(M-\delta M_*-R_*) = n-1$. See Example 4 and the discussion following Assumption 3 for the case when $\text{rank}(M-\delta M_*-R_*) < n-1$.

Let us first assume that the property matrix R_* is fixed. For δM_* to be the minimum norm rank-reducing perturbation for $M-R_*$, the tangent plane to the hypersurface $\Upsilon_1 = \{\delta M \in \mathcal{S} : \sigma_n(M-R_*-\delta M) = 0\}$ must be perpendicular to the line connecting $M-R_*$ and $M-R_*-\delta M_*$ (see Figure 1). For u_* the n^{th} left singular vector (lsv) and V_* having columns equal to the n^{th} through m^{th} right singular vectors (rsv) of $M-R_*-\delta M_*$, $\|L_{u_*V_*}(M-R_*-\delta M_*)\|_F = \sigma_n(M-R_*-\delta M_*) = 0$. As will be seen in the proof of Theorem 1, the hyperplane $\Upsilon_2 = \{\delta M \in \mathcal{S} : L_{u_*V_*}(M-R_*-\delta M) = 0\}$ is related to the tangent plane to Υ_1 at δM_*. Note, elements of Υ_2 are precisely the minima of $\|L_{u_*V_*\mathcal{S}}(\delta M) - L_{u_*V_*}(M-R_*)\|_F$ over $\delta M \in \mathcal{S}$. The minimum Frobenius norm rank reducing perturbation on $M-R_*$ is therefore the element of Υ_2 that has the least Frobenious norm. Since $L_{u_*V_*\mathcal{S}}$ is surjective by Assumption 1, this minimum is given by

$$\delta M_* = (L^\dagger_{u_*V_*\mathcal{S}} \circ L_{u_*V_*})(M-R_*), \tag{14}$$

the first necessary condition in Theorem 1.

The second necessary condition addresses the locally optimal property matrix R_*. As per the discussion above, the optimal rank reducing perturbation satisfies (14). Let $\Delta R \in \mathcal{P}$ be an alteration to R_*. Given any δM_0, let u_0 be the n^{th} lsv and V_0 have columns equal to the n^{th} through m^{th} rsv of $M-R_*-\Delta R-\delta M_0$. According to (14), given property matrix $R_*+\Delta R$, the corresponding minimum Frobenius norm perturbation satisfies

$$\delta M_0 = (L^\dagger_{u_0V_0\mathcal{S}} \circ L_{u_0V_0})(M-R_*-\Delta R). \tag{15}$$

It is difficult to directly minimize the norm of (15) with respect to ΔR because the matrices u_0 and V_0 change with ΔR. However, for sufficiently small ΔR, we can approximate δM_0 with

$$\delta M_0 \approx (L^\dagger_{u_*V_*\mathcal{S}} \circ L_{u_*V_*})(M-R_*-\Delta R), \tag{16}$$

where u_* and V_* are associated with $M-R_*-\delta M_*$. Thus in a sufficiently small neighborhood of R_*, minimizing the norm of (15) with respect to ΔR is equivalent to minimizing

$$\|(L^\dagger_{u_*V_*\mathcal{S}}\circ L_{u_*V_*\mathcal{P}})\Delta R - (L^\dagger_{u_*V_*\mathcal{S}}\circ L_{u_*V_*})(M-R_*)\|_F. \tag{17}$$

The minimizer of (17) is given by

$$\Delta R = (L^\dagger_{u_*V_*\mathcal{S}}\circ L_{u_*V_*\mathcal{P}})^\dagger (L^\dagger_{u_*V_*\mathcal{S}}\circ L_{u_*V_*})(M-R_*). \tag{18}$$

If R_* is the optimal property matrix, then (17) is minimized at $\Delta R = 0$. Hence, the right-hand side of (18) equals zero, the second necessary condition.

Theorem 1. *Suppose there exists $\delta M_* \in \mathcal{S}$ that is a local minimum norm element of the set $\mathcal{T} = \{\delta M \in \mathcal{S} : \exists R \in \mathcal{P}, \text{rank}[M-\delta M-R] < n\}$; choose $R_* \in \{R \in \mathcal{P} : \text{rank}[M-\delta M_*-R] < n\}$ and let u be a non-trivial element of $\ker[(M-\delta M_*-R_*)^H]$ and let V be a matrix whose columns span $\ker[M-\delta M_*-R_*]$. If $L_{uV\mathcal{S}}$ is surjective, then the following two necessary conditions both hold:*

1a) *$\delta M_* \in \mathcal{S}$ is a minimum norm matrix minimizing*

$$\|L_{uV\mathcal{S}}(\delta M_*) - L_{uV}(M-R_*)\|_F. \tag{19}$$

2a) *$0 = \Delta R_*$, where $\Delta R_* \in \mathcal{P}$ is the matrix minimizing*

$$\|(L^\dagger_{uV\mathcal{S}}\circ L_{uV\mathcal{P}})(\Delta R) - (L^\dagger_{uV\mathcal{S}}\circ L_{uV})(M-R_*)\|_F. \tag{20}$$

Equivalently,

1b) *$\delta M_* = (L^\dagger_{uV\mathcal{S}}\circ L_{uV})(M-R_*)$ and*

2b) *$0 = \Delta R_* = (L^\dagger_{uV\mathcal{S}}\circ L_{uV\mathcal{P}})^\dagger (L^\dagger_{uV\mathcal{S}}\circ L_{uV})(M-R_*)$.*

[1] More extensive investigation of surjectivity and additional results can be found in [21].

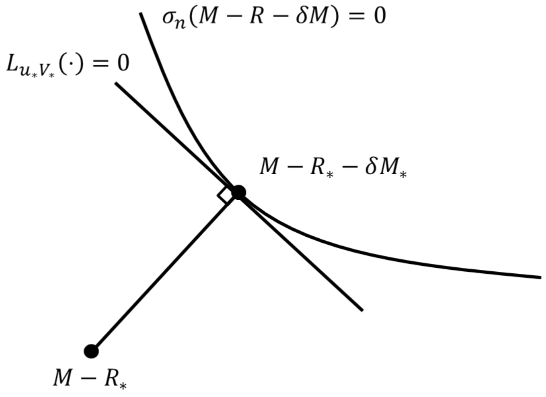

Fig. 1. This figure illustrates the first necessary condition. Given appropriate assumptions, the surface $L_{u_*V_*}(\cdot)=0$ is tangent to the curve $\sigma_n(M-R_*-\delta M)=0$ for $\delta M\in\mathcal{S}$ at $M-R_*-\delta M_*$, where u_* is the n^{th} lsv and V_* is the n^{th} through m^{th} rsv of $M-R_*-\delta M_*$. For δM_* to be a local minimum rank reducing perturbation for the property matrix R_* the line connecting $M-R_*$ to $M-R_*-\delta M_*$ must be perpendicular to the tangent surface $L_{u_*V_*}(\cdot)=0$.

Remark 1. *Conditions 1b*[2] *and 2b are Frobenius-norm specific. However, we conjecture that conditions 1a and 2a can be generalized to other norms, such as the spectral norm, as follows: in 1a, it should read "... minimum spectral norm minimizing", and in 2a, since* $\Delta R_*=0$*, the statement remains the same while in* (20) *the operator* $L_{uV\mathcal{S}}$ *needs to be modified to* $L_{\widehat{u}\widehat{V}\mathcal{S}}$ *so that* $L^{\dagger}_{\widehat{u}\widehat{V}\mathcal{S}}$ *provides the minimum spectral norm matrix. How these modifications might impact the proof of Theorem 1 is beyond the scope of this work.*

Note, condition 1a essentially requires δM_* to be the smallest matrix minimizing $\sigma_n(M-R_*-\delta M)$ for $\delta M\in\mathcal{S}$. So even when no rank reducing perturbation exists, condition 1a provides the "best" solution.

The proof of Theorem 1 requires some machinery and four technical lemmas. As will be seen, proving the necessary conditions in Theorem 1 requires the application of the inverse function theorem[3] which in turn requires Fréchet differentiability of the equality constraint $H(\zeta,\rho)=0$ in (8). Unfortunately, there are points at which H is only directionally differentiable. These non-Fréchet differentiable points are caused by two structural components of the SVD: i) the ordering of the singular values and ii) the requirement that the singular values be positive. We observe that, in general, perturbation and property matrices δM and R for which $M-\delta M-R$ has a repeated smallest singular value or a zero smallest singular value is an algebraic variety of lower dimension in $\mathcal{S}\times\mathcal{P}$. Consequently, the function $H(\zeta,\rho)$ is Fréchet differentiable almost everywhere. Since we are concerned with rank reducing perturbations, we need to resolve the non-Fréchet differentiability when $H(\zeta,\rho)=0$.

De Moor and Boyd in [22] suggest an alternative SVD that relaxes the reordering of the singular values/vectors and positivity of the singular values. The focus of [22] is computing analytic unsigned and unordered SVDs along an analytic path. These results on analytic paths are extended herein to an open set in $\mathbb{R}^k\times\mathbb{R}^r$; in this way, we can construct a Fréchet differentiable function $\widetilde{H}(\zeta,\rho)$ which is zero exactly when $H(\zeta,\rho)=0$.

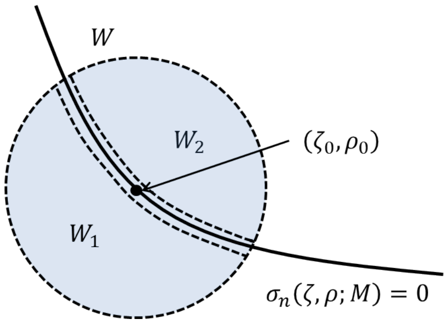

Fig. 2. This figure illustrates the simply-connected neighborhood W of (ζ_0,ρ_0) partitioned three simply connected regions: W_1 and W_2 disjoint open sets with $W\subset\operatorname{cl}(W_1\cup W_2)$ and the surface $W\cap\{\sigma_n(\cdot)=0\}$.

Let $(\delta M_0,R_0)\in\mathcal{S}\times\mathcal{P}$ be a pair of matrices which satisfy

$$\operatorname{rank}(M-\delta M_0-R_0)=n-1.$$

Let $(\zeta_0,\rho_0)\in\mathbb{R}^k\times\mathbb{R}^r$ represent $(\delta M_0,R_0)$ in the bases $\{S_1,\cdots,S_k\}$ and $\{P_1,\cdots,P_r\}$, respectively. To simplify the notation, define the map $\sigma_n:\mathbb{R}^k\times\mathbb{R}^r\times\mathbb{C}^{n\times m}\to\mathbb{R}$ given by

$$\sigma_n(\zeta,\rho;M)\triangleq\sigma_n\left(M(\zeta,\rho)\right), \tag{21}$$

where $M(\zeta,\rho)$ has been defined in (6). Since $\sigma_n(M-\delta M_0-R_0)$ is simple and $\sigma_n(\cdot)$ is continuous everywhere, there exists a simply-connected and open neighborhood $W\subset\mathbb{R}^k\times\mathbb{R}^r$ of (ζ_0,ρ_0) sufficiently small such that

1) for each $(\zeta,\rho)\in W$, $\sigma_n(\zeta,\rho;M)$ (the smallest singular value) is simple,
2) there exists simply-connected and open subsets $W_1,W_2\subset W$ such that
    a) $W\subset\operatorname{cl}(W_1\cup W_2)$,
    b) W_1 and W_2 are disjoint, and
    c) for each $(\zeta,\rho)\in W_1\cup W_2$, $\sigma_n(\zeta,\rho;M)>0$.

The simply-connected and open subsets W, W_1, and W_2 are illustrated in Figure 2. Note that the open set W includes the common boundary of the open sets W_1 and W_2.

Let $g:[0,1]\to W$ be an analytic function with $g(s)\in W_1$ for $s<0.5$ and $g(s)\in W_2$ for $s>0.5$. According to [22, Theorem 1], there exists an analytic function $f_g:[0,1]\to\mathbb{R}$, called the unsigned n^{th} singular value function, such that

$$|f_g(s)|=\sigma_n(g(s);M),\quad s\in[0,1].$$

The function f_g can only change sign as it transitions through the common boundary of W_1 and W_2, i.e., at $s=0.5$. By [22, Theorem 3], there exists analytic singular vector functions $u_g:[0,1]\to\mathbb{C}^n$ and $v_g:[0,1]\to\mathbb{C}^m$ such that for each $s\in[0,1]$, $u_g(s)$ and $v_g(s)$ are the unsigned n^{th} lsv and rsv associated with $f_g(s)$, i.e., $u_g(s)$ and $v_g(s)$ are unit vectors satisfying

$$u_g^H(s)\left(M-\sum_{i=1}^{k}\zeta_{gi}(s)S_i-\sum_{j=1}^{r}\rho_{gj}(s)P_j\right)=f_g(s)v_g^H(s),$$

[2] Note the similarity of 1b with (5) in [7].

[3] A simpler proof is outlined in [19], but it requires conditions stronger than surjectivity.

for each $s \in [0,1]$, where $\zeta_g : [0,1] \to \mathbb{R}^k$ and $\rho_g(s) \to \mathbb{R}^r$ are defined by the relation $g \triangleq (\zeta_g, \rho_g)$.

Let $\widetilde{H} : W \to \mathbb{R}$ be the extension of the unsigned singular value function f_g to the set W, i.e.,

$$\widetilde{H}(\zeta,\rho) = \begin{cases} \mathrm{sign}(f_g(0))\sigma_n(\zeta,\rho;M) & (\zeta,\rho) \in W_1, \\ \mathrm{sign}(f_g(1))\sigma_n(\zeta,\rho;M) & \text{otherwise.} \end{cases} \tag{22}$$

Note that, $\widetilde{H}(g(s)) = f_g(s)$ for each $s \in [0,1]$ since by construction of $g(s)$, f_g can change sign only at $s = 0.5$. In addition, the form of $\widetilde{H}$ implies that for each $(\zeta,\rho) \in W$, $|\widetilde{H}(\zeta,\rho)| = \sigma_n(\zeta,\rho;M)$. We will show that $\widetilde{H}$ is Fréchet differentiable on W, as per the following lemma.

Lemma 1. *Let $(\zeta_0,\rho_0) \in \mathbb{R}^k \times \mathbb{R}^r$ satisfy* rank$[M(\zeta_0,\rho_0)] = n-1$*. Let $W \subset \mathbb{R}^k \times \mathbb{R}^r$ be as defined above. Let $\widetilde{H} : W \to \mathbb{R}$ be as in* (22)*. Then $\widetilde{H}$ is Fréchet differentiable on W with partial derivatives given by*

$$\begin{aligned} \frac{\partial \widetilde{H}(\zeta,\rho)}{\partial \zeta_i} &= -\,\mathrm{Re}(u^H S_i v) = -\,\mathrm{Re}(L_{uv\mathcal{S}}(S_i)) \\ \frac{\partial \widetilde{H}(\zeta,\rho)}{\partial \rho_i} &= -\,\mathrm{Re}(u^H P_i v) = -\,\mathrm{Re}(L_{uv\mathcal{P}}(P_i)) \end{aligned} \tag{23}$$

where u and v are unsigned n^{th} lsv and rsv of $M(\zeta_0,\rho_0)$, respectively, i.e., $u^H M(\zeta_0,\rho_0) = \widetilde{H}(\zeta_0,\rho_0)v^H$ and $M(\zeta_0,\rho_0)v = \widetilde{H}(\zeta_0,\rho_0)u$. Equivalently, (23) *has the matrix form:*

$$\begin{aligned} &\widetilde{H}'(\zeta_0,\rho_0) = \left[\tfrac{\partial \widetilde{H}(\zeta,\rho_0)}{\partial \zeta}\big|_{\zeta=\zeta_0} \;\Big|\; \tfrac{\partial \widetilde{H}(\zeta_0,\rho)}{\partial \rho}\big|_{\rho=\rho_0} \right] = \\ &-\,\mathrm{Re}\left[u^H S_1 v, \cdots, u^H S_k v \mid u^H P_1 v, \cdots, u^H P_r v\right] \\ &= -\,\mathrm{Re}[L_{uv\mathcal{S}}(S_1), \cdots \mid L_{uv\mathcal{P}}(P_1), \cdots], \end{aligned} \tag{24}$$

Proof. See Appendix A. □

Thus Lemma 1 allows the replacement of the constraint $H(\zeta,\rho) = 0$ in (8) by the Fréchet differentiable constraint $\widetilde{H}(\zeta,\rho) = 0$ when restricted to the set W.

The next lemma proves that if the claimed necessary condition 1a (or equivalently 1b) of Theorem 1 is not satisfied, then there exists a direction $\Delta M \in \mathcal{S}$ to change the perturbation δM_* on the tangent plane $L_{uV}(\cdot) = 0$ (see Figure 1). This new perturbation $\delta M_* + \Delta \widetilde{M}$ may not be rank reducing, but will allow us to prove the existence of rank reducing perturbations with norms smaller than $\|\delta M_*\|_F$.

Lemma 2. *Let $M \in \mathbb{C}^{n\times m}$, $\delta M_0 \in \mathcal{S}$, $R_0 \in \mathcal{P}$ satisfy* rank$[M - \delta M_0 - R_0] < n$*. Let u be the n^{th} lsv and V have columns equal to the n^{th} through m^{th} rsv of $M - \delta M_0 - R_0$. Suppose $L_{uV\mathcal{S}}$ surjective and $\delta M_0 \neq (L^{\dagger}_{uV\mathcal{S}} \circ L_{uV})(M - R_0)$. Then there exists a matrix $\Delta M \in \mathcal{S}$ such that*

$$L_{uV\mathcal{S}}(\Delta M) = 0 \tag{25}$$

and

$$\langle \delta M_0, \Delta M \rangle < 0. \tag{26}$$

Proof. See Appendix A. □

Remark 2. *The condition in* (26) *is used to construct ζ_Δ satisfying the condition in* (32) *of Lemma 4. A similar condition in* (29) *is also needed for a similar reason. These conditions are necessary to set up the hypotheses of the inverse function theorem used in the proof of Theorem 1.*

Similar to Lemma 2, Lemma 3 proves that if the second necessary condition (2a or 2b) of Theorem 1 is not satisfied, then there exist directions ΔM and ΔR for changing the perturbation δM_* and the property matrix R_*, respectively, that reduce the norm of the perturbation on the tangent surface $L_{uV}(\cdot) = 0$. Again, the resulting perturbed matrices may not be rank reducing, but will allow us to prove the existence of a rank reducing perturbation with smaller norm and associated property matrix.

Lemma 3. *Let $M \in \mathbb{C}^{n\times m}$, $\delta M_0 \in \mathcal{S}$, $R_0 \in \mathcal{P}$ satisfy* rank$[M - \delta M_0 - R_0] < n$*. Let u be the n^{th} lsv and V have columns equal to the n^{th} through m^{th} rsv of $M - \delta M_0 - R_0$. Suppose $L_{uV\mathcal{S}}$ is surjective, $\delta M_0 = (L^{\dagger}_{uV\mathcal{S}} \circ L_{uV})(M - R_0)$, and*

$$0 \neq \Delta R \triangleq (L^{\dagger}_{uV\mathcal{S}} \circ L_{uV\mathcal{P}})^{\dagger}(L^{\dagger}_{uV\mathcal{S}} \circ L_{uV})(M - R_0). \tag{27}$$

Then there exist $\Delta M \in \mathcal{S}$ such that

$$L_{uV}(\Delta M + \Delta R) = 0 \tag{28}$$

and

$$\langle \delta M_0, \Delta M \rangle < 0. \tag{29}$$

Proof. See Appendix A. □

As a preliminary to the statement of Lemma 4, let $(\zeta_0,\rho_0) \in \mathbb{R}^k \times \mathbb{R}^r$ be a pair satisfying rank$(M(\zeta_0,\rho_0)) = n-1$. Let $W \subset \mathbb{R}^k \times \mathbb{R}^r$ be a simply-connected and open neighborhood of (ζ_0,ρ_0) where the n^{th} singular value is simple, as constructed in Lemma 1. If (ζ_0,ρ_0) does not satisfy both condition 1a and 2a (or equivalently 1b and 2b), Lemmas 2 and 3 can be used to prove the existence of smaller normed rank reducing perturbations. To this end, we combine the function $f(\zeta,\rho) = 0.5\|\zeta\|_F^2$ in (9) with the equality constraint $\widetilde{H}(\zeta,\rho) = 0$ to form a new function $T : W \to \mathbb{R}^2$ given by

$$T(\zeta,\rho) = \begin{bmatrix} f(\zeta,\rho) - f(\zeta_0,\rho_0) \\ \widetilde{H}(\zeta,\rho) \end{bmatrix}. \tag{30}$$

The verification of the necessary conditions of Theorem 1 follows by applying the inverse function theorem to the map T. The underlying hypotheses of the inverse function theorem require proving that T is Fréchet differentiable and that $T'(\zeta_0,\rho_0)$ is surjective, under appropriate conditions.

Lemma 4. *Let $(\zeta_0,\rho_0) \in \mathbb{R}^k \times \mathbb{R}^r$ satisfy* rank$(M(\zeta_0,\rho_0)) = n-1$ *and $W \subset \mathbb{R}^k \times \mathbb{R}^r$ be a neighborhood of (ζ_0,ρ_0) as in Lemma 1 where the n^{th} singular value is simple. If*

$$\frac{\partial}{\partial \zeta}\widetilde{H}(\zeta,\rho_0)|_{\zeta=\zeta_0} \tag{31}$$

is surjective and there exists $\rho_\Delta \in \mathbb{R}^r$ and $\zeta_\Delta \in \mathbb{R}^k$ such that

$$\zeta_0^{\top}\zeta_\Delta < 0 \tag{32}$$

and

$$\widetilde{H}'(\zeta_0,\rho_0)\begin{bmatrix} \zeta_\Delta \\ \rho_\Delta \end{bmatrix} = 0, \tag{33}$$

then T defined in (30) *is Fréchet differentiable and $T'(\zeta_0,\rho_0)$ is surjective.*

Proof. See Appendix A. □

We can now prove necessary conditions of Theorem 1.

Proof of Theorem 1.

Condition 1: For contradiction, assume that δM_0 is a minimum norm element in $\mathcal{T}$ (defined in (5)) with associated property matrix R_0, but condition 1b is not satisfied, i.e.,

$$\delta M_0 \neq (L_{uV\mathcal{S}}^{\dagger} \circ L_{uV})(M - R_0) \tag{34}$$

where $u \in \mathbb{C}^n$ and $V \in \mathbb{C}^{m\times(m-n+1)}$ are the n^{th} lsv and n^{th} through m^{th} rsv of $M - \delta M_0 - R_0$, respectively. By Lemma 2, there exists $\Delta M \in \mathcal{S}$ such that $L_{uV\mathcal{S}}(\Delta M) = 0$ and $\langle \delta M_0, \Delta M_0\rangle < 0$. Let ζ_0 and ζ_Δ in $\mathbb{R}^k$ represent δM_0 and ΔM in the orthonormal basis $\{S_1, \cdots, S_k\}$, respectively. Let $\rho_0 \in \mathbb{R}^r$ represent R_0 in the orthonormal basis $\{P_1, \cdots, P_r\}$. Since the basis $\{S_1, \cdots, S_k\}$ is orthonormal, $\zeta_0^\top \zeta_\Delta = \langle \delta M_0, \Delta M\rangle < 0$. By the form of $\widetilde{H}'(\zeta_0, \rho_0)$ in Lemma 1,

$$\widetilde{H}'(\zeta_0, \rho_0) \begin{bmatrix} \zeta_\Delta \\ 0 \end{bmatrix} = -\operatorname{Re}(L_{uv\mathcal{S}}(\Delta M)), \tag{35}$$

where v is the n^{th} unsigned rsv of $M - \delta M_0 - R_0$. But $v \in \text{col-sp}(V)$ so there exists a unique $w \in \mathbb{C}^{m-n+1}$ such that $v = Vw$. Since $L_{uV\mathcal{S}}(\Delta M) = 0$, $L_{uv\mathcal{S}}(\Delta M) = 0$. Thus (35) is zero.

Consider the function $\frac{\partial}{\partial\zeta}\widetilde{H}(\zeta, \rho_0)|_{\zeta=\zeta_0} : \mathbb{R}^k \to \mathbb{R}$. To show surjectivity of $\frac{\partial}{\partial\zeta}\widetilde{H}(\zeta, \rho_0)|_{\zeta=\zeta_0}$, for each $\alpha \in \mathbb{R}$ we need to show there exists $\zeta_\alpha \in \mathbb{R}^k$ such that $\frac{\partial}{\partial\zeta}\widetilde{H}(\zeta, \rho_0)|_{\zeta=\zeta_0}\zeta_\alpha = \alpha$. Let $w = [w_1, w_2, \cdots, w_{m-n+1}]^\top \in \mathbb{C}^{m-n+1}$ satisfy $v = Vw$. Without loss of generality, suppose $w_i \neq 0$ for some i.

Similar to (35), for each $\zeta \in \mathbb{R}^k$ define $\delta M_\zeta \triangleq \sum_i^k \zeta_i S_i$. From (24) and the definition of w,

$$\begin{aligned} \frac{\partial}{\partial\zeta}\widetilde{H}(\zeta, \rho_0)|_{\zeta=\zeta_0}\zeta &= -\operatorname{Re}(u^H \delta M_\zeta v) \\ &= -\operatorname{Re}(L_{uV\mathcal{S}}(\delta M_\zeta) w). \end{aligned} \tag{36}$$

To demonstrate surjectivity, suppose $\alpha \in \mathbb{R}$ is an arbitrary scalar. Since $L_{uV\mathcal{S}}$ is surjective, there exists $\delta M_\alpha = \sum_i^k \zeta_{\alpha i} S_i \in \mathcal{S}$ for $\zeta_\alpha \in \mathbb{R}^k$ such that

$$L_{uV\mathcal{S}}(\delta M_\alpha) = [\cdots, 0, -\alpha/w_i, 0, \cdots], \tag{37}$$

where only the i^{th} entry is nonzero. Then from (36) and (37), $\frac{\partial}{\partial\zeta}\widetilde{H}(\zeta, \rho_0)|_{\zeta=\zeta_0}\zeta_\alpha = \alpha$. Since α was arbitrary, $\frac{\partial}{\partial\zeta}\widetilde{H}(\zeta, \rho_0)|_{\zeta=\zeta_0}$ is surjective.

Hence by Lemma 4, $T(\zeta, \rho)$ given in (30) is Fréchet differentiable and $T'(\zeta_0, \rho_0)$ is surjective. Thus by the inverse function theorem [23], there exists an open set $W_0 \subset \mathbb{R}^2$ containing zero such that for all $y \in W_0$, there exists $\zeta_y \in \mathbb{R}^k$ and $\rho_y \in \mathbb{R}^r$ such that $T(\zeta_y, \rho_y) = y$. Hence for all sufficiently small neighborhoods of 0 in $\mathbb{R}^2$, there exists $\delta > 0$, $\zeta_* \in \mathbb{R}^k$, and $\rho_* \in \mathbb{R}^r$ such that $T(\zeta_*, \rho_*) = [-\delta, 0]^\top$. This implies that $\delta M_* = \sum_{i=1}^k \zeta_{*i} S_i \in \mathcal{T}$, i.e., it is a rank reducing perturbation, with associated property matrix $R_* = \sum_{j=1}^r \rho_{*j} R_j$. Since $f(\zeta_*, \rho_*) - f(\zeta_0, \rho_0) = -\delta < 0$, $\|\delta M_*\|_F < \|\delta M_0\|_F$ contradicting that δM_0 is a local minimum norm element in $\mathcal{T}$.

Condition 2: For contradiction, assume that δM_0 is a minimum norm element in $\mathcal{T}$ with associated property matrix R_0 and condition 1b is satisfied, but condition 2b is not, i.e.,

$$\delta M_0 = (L_{uV\mathcal{S}}^{\dagger} \circ L_{uV})(M - R_0), \text{ and} \tag{38}$$

$$0 \neq (L_{uV\mathcal{S}}^{\dagger} \circ L_{uV\mathcal{P}})^{\dagger}(L_{uV\mathcal{S}}^{\dagger} \circ L_{uV})(M - R_0) \tag{39}$$

where $u \in \mathbb{C}^n$ and $V \in \mathbb{C}^{m\times(m-n+1)}$ are the n^{th} lsv and n^{th} through m^{th} rsv of $M - \delta M_0 - R_0$, respectively. By Lemma 3, there exists $\Delta M \in \mathcal{S}$ and $\Delta R \in \mathcal{P}$ such that $L_{uV}(\Delta M + \Delta R) = 0$ and $\langle \delta M_0, \Delta M\rangle < 0$. Let ζ_0 and ζ_Δ in $\mathbb{R}^k$ specify δM_0 and ΔM in the orthonormal basis $\{S_1, \cdots, S_k\}$, respectively. Let ρ_0 and ρ_Δ specify R_0 and ΔR in the orthonormal basis $\{P_1, \cdots, P_r\}$, respectively. Since the basis $\{S_1, \cdots, S_k\}$ is orthonormal, $\zeta_0^\top \zeta_\Delta = \langle \delta M_0, \Delta M\rangle < 0$. By the form of $\widetilde{H}'(\zeta_0, \rho_0)$ in Lemma 1,

$$\widetilde{H}'(\zeta_0, \rho_0) \begin{bmatrix} \zeta_\Delta \\ \rho_\Delta \end{bmatrix} = -\operatorname{Re}(L_{uv\mathcal{S}}(\Delta M)).$$

Since $L_{uV}(\Delta M + \Delta R) = 0$, $L_{uv}(\Delta M + \Delta R) = 0$. This implies $\widetilde{H}'(\zeta_0, \rho_0)[\zeta_\Delta^\top, \rho_\Delta^\top]^\top = 0$. Since $L_{uV\mathcal{S}}$ is surjective, using arguments from the proof of Condition 1, $\frac{\partial}{\partial\zeta}\widetilde{H}(\zeta, \rho_0)|_{\zeta=\zeta_0}$ is surjective onto $\mathbb{R}$. Hence by Lemma 4, $T(\zeta, \rho)$ given in (30) is Fréchet differentiable and $T'(\zeta_0, \rho_0)$ is surjective. Using the same arguments as while proving condition 1, this implies that there exists $\delta M_* \in \mathcal{T}$ smaller than δM_0, contradicting that δM_0 is a local minimum element. □

The next section sets forth an algorithm which is proven to converge to a perturbation and property matrix pair $(\delta M_*, R_*)$ satisfying the necessary conditions of Theorem 1.

IV. $\mathcal{P}$–Robustness Algorithm

This section introduces the steps in the $\mathcal{P}$–Robustness Algorithm. The bold statements in Algorithm 1 below provide a high-level description of individual steps. The algorithm computes norm reducing and rank reducing directions of search in each iteration. It then proceeds along the direction of their vector sum with a step size α_k chosen to reduce a discrete step-dependent Lyapunov function. Each step of the algorithm, as well as the implementation, are discussed in detail after the algorithm is stated.

Algorithm 1. $\mathcal{P}$–Robustness.

```
1) $k = 0$
2) Initialize $\delta M_0 \in \mathcal{S}$ and $R_0 \in \mathcal{P}$. Set $g_0 = 1$.
3) REPEAT
4) Let $u$ and $V$ be the $n^{th}$ lsv and $n^{th}$ through $m^{th}$ rsv
   of $M - \delta M_k - R_k$, respectively[4], such that the operator
   $L_{uV\mathcal{S}}$ is surjective. Define $[\sigma_n]_k \triangleq \sigma_n(M - \delta M_k - R_k)$.
5) Norm reducing direction $(\delta\widetilde{M}_k, \Delta\widetilde{R}_k)$:

   Set $\widetilde{\phi}_k = \min_{\delta M\in\mathcal{S},\Delta R\in\mathcal{P}} \|L_{uV}(\delta M + \Delta R) - L_{uV\mathcal{S}}(\delta M_k)\|_F$ and let
      $\widetilde{Z} = \{(\delta M, \Delta R) :$
            $\|L_{uV}(\delta M + \Delta R - \delta M_k)\|_F = \widetilde{\phi}_k\}$        (40)
```

[4] We suppress the k-dependence of u and V to prevent overburdening the notation, i.e., the singular vectors change in each iteration.

Then,

$$\delta\widetilde{M}_k = \underset{\substack{\delta M' \in \{\delta M \in \mathcal{S} : \exists \Delta R \in \mathcal{P} \\ \text{s.t. } (\delta M, \Delta R) \in \widetilde{Z}\}}}{\operatorname{argmin}} \|\delta M'\|_F$$

and

$$\Delta\widetilde{R}_k = \underset{\Delta R' \in \{\Delta R \in \mathcal{P} : (\delta\widetilde{M}_k, \Delta R) \in \widetilde{Z}\}}{\operatorname{argmin}} \|\Delta R'\|_F.$$

Equivalently, since $L_{uV\mathcal{S}}$ is surjective, the actual computations are

$$\Delta\widetilde{R}_k = (L^\dagger_{uV\mathcal{S}} \circ L_{uV\mathcal{P}})^\dagger (L^\dagger_{uV\mathcal{S}} \circ L_{uV\mathcal{S}})(\delta M_k) \quad (41)$$

$$\delta\widetilde{M}_k = (L^\dagger_{uV\mathcal{S}} \circ L_{uV})(\delta M_k - \Delta\widetilde{R}_k). \quad (42)$$

6) **Rank reducing direction** $(\delta\overline{M}_k, \Delta\overline{R}_k)$**:**

Set $\overline{\phi}_k = \min_{\delta M \in \mathcal{S}, \Delta R \in \mathcal{P}} \|L_{uV}(\delta M + \Delta R) - L_{uV}(M - R_k - \delta M_k)\|_F$ and

$$\overline{Z} = \{(\delta M, \Delta R) : \quad (43)$$
$$\|L_{uV}(\delta M + \Delta R - (M - R_k - \delta M_k))\|_F = \overline{\phi}_k\}$$

Then,

$$\delta\overline{M}_k = \underset{\substack{\delta M' \in \{\delta M \in \mathcal{S} : \exists \Delta R \in \mathcal{P} \\ \text{s.t. } (\delta M, \Delta R) \in \overline{Z}\}}}{\operatorname{argmin}} \|\delta M'\|_F$$

and

$$\Delta\overline{R}_k = \underset{\Delta R' \in \{\Delta R \in \mathcal{P} : (\delta\overline{M}_k, \Delta R) \in \overline{Z}\}}{\operatorname{argmin}} \|\Delta R'\|_F.$$

Equivalently, since $L_{uV\mathcal{S}}$ is surjective, the actual computations are

$$\Delta\overline{R}_k = (L^\dagger_{uV\mathcal{S}} \circ L_{uV\mathcal{P}})^\dagger \quad (44)$$
$$(L^\dagger_{uV\mathcal{S}} \circ L_{uV})(M - R_k - \delta M_k)$$

$$\delta\overline{M}_k = (L^\dagger_{uV\mathcal{S}} \circ L_{uV})(M - R_k - \Delta\overline{R}_k - \delta M_k) \quad (45)$$

7) **Lyapunov function reducing direction:**

$$\Delta R_k = \Delta\widetilde{R}_k + \Delta\overline{R}_k \text{ and } \delta\widehat{M}_k = \delta\widetilde{M}_k + \delta\overline{M}_k$$

8) **Normalizing weights:**

$$g_k = \min\left(g_{k-1}, [\sigma_n]_k / (2\|\delta\overline{M}_k\|_F)\right),$$
$$b_k = 0 \text{ if } \delta M_k = 0, \text{ otherwise } b_k = \frac{1}{2}\|\delta M_k\|_F^{-1}$$

9) **Choosing a step size:** Define

$$f^{(k)}_{ub}(\alpha) = \frac{-[\sigma_n]_k}{2}\alpha + a_k\alpha^2 - g_k b_k \|\delta M_k\|_F^2$$
$$+ g_k b_k \|(1-\alpha)\delta M_k + \alpha\delta\widetilde{M}_k\|_F^2 \quad (46)$$

where

$$a_k = \|[u_n]_k^H (\delta\widehat{M}_k - \delta M_k - \Delta R_k)(I - V_k V_k^H) \quad (47)$$
$$* (M - \delta M_k - R_k)^\dagger (\delta\widehat{M}_k - \delta M_k + \Delta R_k)\|_2.$$

Compute

$$\alpha_k = \underset{\alpha \in [0,1]}{\operatorname{argmin}} f^{(k)}_{ub}(\alpha) \quad (48)$$

10) **Update estimates:** $R_{k+1} = R_k + \alpha_k \Delta R_k$ and $\delta M_{k+1} = (1-\alpha_k)\delta M_k + \alpha_k \delta\widehat{M}_k$.

11) $k \to k+1$

12) **UNTIL** $\|\Delta R_k\|_F < \epsilon$, $\|\delta M_k - \delta\widetilde{M}_k\|_F < \epsilon$, $[\sigma_n]_k < \epsilon$

Different choices of the initial guess for δM_0 and R_0 may lead to different local minima. In our work we have used random perturbations near zero. Alternatively, one can use perturbations that give upper or lower bounds, computed using e.g. [3], [9].

Algorithm 1 is designed to reduce a Lyapunov energy function of the form

$$E_k = [\sigma_n]_k + g_k \|\delta M_k\|_F, \quad (49)$$

where $[\sigma_n]_k = \sigma_n(M - R_k - \delta M_k)$ and g_k is a nonzero adaptive weight computed in step 8. A direction for reducing E_{k+1} is found by moving along the vector sum of the directions $(\delta\widetilde{M}_k, \Delta\widetilde{R}_k)$ (step 5) and $(\delta\overline{M}_k, \Delta\overline{R}_k)$ (step 6), which reduce $\|\delta M_{k+1}\|_F$ and $[\sigma_n]_{k+1}$, respectively.

To illustrate how these directions affect (49), consider step 5. To find $(\delta\widetilde{M}_k, \Delta\widetilde{R}_k)$ reducing $\|\delta M_{k+1}\|_F$, we search for the minimum norm pair that does not change σ_n by approximating the function $\sigma_n(\cdot)$ with L_{uV}. Specifically, $(\delta\widetilde{M}_k, \Delta\widetilde{R}_k)$ satisfies

$$L_{uV}(M - R_k - \Delta\widetilde{R}_k - \delta\widetilde{M}_k) =$$
$$= L_{uV}(M - R_k - \delta M_k) = \begin{bmatrix} [\sigma_n]_k & 0 & \cdots & 0 \end{bmatrix}. \quad (50)$$

Hence the pairs $(\delta\widetilde{M}_k, \Delta\widetilde{R}_k)$ satisfying (50) constitute the set $\widetilde{Z}$ in (40), i.e., if $L_{uV\mathcal{S}}$ is surjective then $\widetilde{\phi}_k = 0$, because for any $\Delta R \in \mathcal{P}$ setting

$$\delta M = (L^\dagger_{uV\mathcal{S}} \circ L_{uV})(\delta M_k - \Delta R) \quad (51)$$

results in

$$0 = \|L_{uV\mathcal{S}}(\delta M) - L_{uV}(\delta M_k - \Delta R)\|_F \geq \widetilde{\phi}_k \geq 0. \quad (52)$$

Moreover, δM defined by (51) is the matrix with the smallest Frobenius norm in $\mathcal{S}$ such that (52) is zero. Any pair $(\delta M, \Delta R) \in \widetilde{Z}$ for which $\|\delta M\|_F$ is minimized, satisfies (51), i.e., for a yet unspecified $\Delta\widetilde{R}_k$,

$$\delta\widetilde{M}_k = (L^\dagger_{uV\mathcal{S}} \circ L_{uV})(\delta M_k - \Delta\widetilde{R}_k). \quad (53)$$

Choosing $\Delta\widetilde{R}_k$ to minimize $\|\delta\widetilde{M}_k\|_F$ (for pairs in $\widetilde{Z}$) is then equivalent to minimizing the norm of the right hand side of (53), i.e.,

$$\Delta\widetilde{R}_k = \underset{\Delta R \in \mathcal{P}}{\operatorname{argmin}} \|\widetilde{\psi}(\Delta R)\|_F, \quad (54)$$

where

$$\widetilde{\psi}(\Delta R) \triangleq (L^\dagger_{uV\mathcal{S}} \circ L_{uV\mathcal{P}})(\Delta R) - (L^\dagger_{uV\mathcal{S}} \circ L_{uV\mathcal{S}})(\delta M_k).$$

The matrix $\Delta\widetilde{R}_k$ with smallest Frobenius norm minimizing (54) is

$$\Delta\widetilde{R}_k = (L^\dagger_{uV\mathcal{S}} \circ L_{uV\mathcal{P}})^\dagger (L^\dagger_{uV\mathcal{S}} \circ L_{uV\mathcal{S}})(\delta M_k). \quad (55)$$

Since δM_k is known from the previous step, when $L_{uV\mathcal{S}}$ is surjective, $\Delta\widetilde{R}_k$ can be computed first using (55) – which is identical to (41), prior to computing $\delta\widetilde{M}_k$ using (53) – which is identical to (42). This justifies the statements of step 5.

Step 6 computes a direction $(\delta\overline{M}_k, \Delta\overline{R}_k)$ for reducing $[\sigma_n]_{k+1}$, i.e., it finds $(\delta\overline{M}_k, \Delta\overline{R}_k)$ minimizing $\|\delta\overline{M}_k\|_F$ subject to

$$L_{uV}(\delta\overline{M}_k + \Delta\overline{R}_k) = L_{uV}(M - R_k - \delta M_k). \tag{56}$$

As in step 5, the linear operator L_{uV} approximates the smallest singular value function $\sigma_n(\cdot)$. Hence (56) is an approximation of the constraint $\sigma_n(M - R_k - \Delta\overline{R}_k - \delta M_k - \delta\overline{M}_k) = 0$. Using analogous arguments as for step 5, pairs $(\delta\overline{M}_k, \Delta\overline{R}_k)$ satisfying (56) are in $\overline{Z}$ if $L_{uV\mathcal{S}}$ is surjective, i.e., $\overline{\phi}_k = 0$ and the pair $(\delta\overline{M}_k, \Delta\overline{R}_k) \in \overline{Z}$ minimizing $\|\delta\overline{M}_k\|_F$ satisfies

$$\delta\overline{M}_k = (L^{\dagger}_{uV\mathcal{S}} \circ L_{uV})(M - R_k - \Delta\overline{R}_k - \delta M_k). \tag{57}$$

Here $\Delta\overline{R}_k$ is chosen to be the smallest norm matrix in $\mathcal{P}$ minimizing the norm of the right side of (57), i.e.,

$$\Delta\overline{R}_k = \underset{\Delta R \in \mathcal{P}}{\operatorname{argmin}} \|\overline{\psi}(\Delta R)\|_F, \tag{58}$$

where

$$\begin{aligned}\overline{\psi}(\Delta R) &= (L^{\dagger}_{uV\mathcal{S}} \circ L_{uV\mathcal{P}})(\Delta R) \\ &\quad - (L^{\dagger}_{uV\mathcal{S}} \circ L_{uV})(M - R_k - \delta M_k).\end{aligned} \tag{59}$$

The matrix $\Delta\overline{R}_k$ with smallest Frobenius norm minimizing (58) is given by

$$\Delta\overline{R}_k = (L^{\dagger}_{uV\mathcal{S}} \circ L_{uV\mathcal{P}})^{\dagger}(L^{\dagger}_{uV\mathcal{S}} \circ L_{uV})(M - R_k - \delta M_k), \tag{60}$$

completing the justification of step 6.

What remains is to specify the step size α_k. It is chosen to decrease E_{k+1} (see (49)) in the direction of ΔR_k and $\delta\widehat{M}_k$ in step 7. Since the singular value function $\sigma_n(\cdot)$ is not differentiable at 0, we instead minimize a surrogate quadratic function $f^{(k)}_{ub}(\alpha)$ in (46) that upper bounds (see proof of Theorem 2) the decrease $E_{k+1}(\alpha) - E_k \leq f^{(k)}_{ub}(\alpha)$. Choosing α_k to minimize $f^{(k)}_{ub}$ will cause the sequence $\{E_k\}$ to converge to a positive constant $d = g_* \|\delta M_*\|_F$. This suffices to guarantee the necessary conditions are met at the terminating values δM_* and R_*.

A. Algorithm 1 Implementation

To implement steps 5 and 6 in Algorithm 1, the pseudoinverse $L_{uV\mathcal{S}}$ is computed via Kronecker products and the vec operator [24], [25]:

$$\begin{aligned}\operatorname{vec}(L_{uV\mathcal{S}}(\delta M)) &= (V^{\top} \otimes u^H)\operatorname{vec}(\delta M) \\ &= (V^{\top} \otimes u^H) B_{\mathcal{S}} \zeta,\end{aligned}$$

where ζ respresents δM in the orthonormal basis $\{S_1, \cdots, S_k\}$ and $B_{\mathcal{S}} \triangleq [\operatorname{vec}(S_1), \cdots, \operatorname{vec}(S_k)]$. Taking the real and imaginary components we obtain

$$\begin{bmatrix}(\operatorname{Re}[L_{uV\mathcal{S}}(\delta M)])^{\top} \\ (\operatorname{Im}[L_{uV\mathcal{S}}(\delta M)])^{\top}\end{bmatrix} = \begin{bmatrix}\operatorname{Re}[(V^{\top} \otimes u^H) B_{\mathcal{S}}] \\ \operatorname{Im}[(V^{\top} \otimes u^H) B_{\mathcal{S}}]\end{bmatrix} \zeta.$$

Let $N_{\mathcal{S}} \in \mathbb{C}^{2(m-n+1)\times k}$ and $N_{\mathcal{P}} \in \mathbb{C}^{2(m-n+1)\times r}$ be

$$N_{\mathcal{S}} = \begin{bmatrix}\operatorname{Re}[(V_k^{\top} \otimes [u_n]_k^H) B_{\mathcal{S}}] \\ \operatorname{Im}[(V_k^{\top} \otimes [u_n]_k^H) B_{\mathcal{S}}]\end{bmatrix},$$

$$N_{\mathcal{P}} = \begin{bmatrix}\operatorname{Re}[(V_k^{\top} \otimes [u_n]_k^H) B_{\mathcal{P}}] \\ \operatorname{Im}[(V_k^{\top} \otimes [u_n]_k^H) B_{\mathcal{P}}]\end{bmatrix},$$

where $B_{\mathcal{P}} \triangleq [\operatorname{vec}(P_1), \cdots, \operatorname{vec}(P_r)]$. With this notation, $\Delta\widetilde{R}_k$ of (41) and $\delta\widetilde{M}_k$ of (42) satisfy

$$\operatorname{vec}(\Delta\widetilde{R}_k) = B_{\mathcal{P}}(N^{\dagger}_{\mathcal{S}} N_{\mathcal{P}})^{\dagger} N^{\dagger}_{\mathcal{S}} N_{\mathcal{S}} \zeta_k$$

$$\operatorname{vec}(\delta\widetilde{M}_k) = B_{\mathcal{S}} N^{\dagger}_{\mathcal{S}} \begin{bmatrix}\operatorname{Re}[(V_k^{\top} \otimes [u_n]_k^H)\operatorname{vec}(\delta M_k - \Delta\widetilde{R}_k)] \\ \operatorname{Im}[(V_k^{\top} \otimes [u_n]_k^H)\operatorname{vec}(\delta M_k - \Delta\widetilde{R}_k)]\end{bmatrix},$$

where ζ_k represents δM_k in $\{S_1, \cdots, S_k\}$. Similarly, $\Delta\overline{R}_k$ in (44) and $\delta\overline{M}_k$ in (45) satisfy

$$\begin{aligned}\operatorname{vec}(\Delta\overline{R}_k) &= B_{\mathcal{P}}(N^{\dagger}_{\mathcal{S}} N_{\mathcal{P}})^{\dagger} N^{\dagger}_{\mathcal{S}} * \\ &\qquad \begin{bmatrix}\operatorname{Re}[(V_k^{\top} \otimes [u_n]_k^H)\operatorname{vec}(M - R_k - \delta M_k)] \\ \operatorname{Im}[(V_k^{\top} \otimes [u_n]_k^H)\operatorname{vec}(M - R_k - \delta M_k)]\end{bmatrix} \\ \operatorname{vec}(\delta\widetilde{M}_k) &= B_{\mathcal{S}} N^{\dagger}_{\mathcal{S}} * \\ &\qquad \begin{bmatrix}\operatorname{Re}[(V_k^{\top} \otimes [u_n]_k^H)\operatorname{vec}(M - R_k - \Delta\overline{R}_k - \delta M_k)] \\ \operatorname{Im}[(V_k^{\top} \otimes [u_n]_k^H)\operatorname{vec}(M - R_k - \Delta\overline{R}_k - \delta M_k)]\end{bmatrix}\end{aligned}$$

Now we consider step 9 that requires the minimization of the function $f^{(k)}_{ub}$ in (46) with respect to the step size α_k. One possibility is a one-dimensional constrained line search for $\alpha_k \in [0, 1]$. Since a decrease in E_{k+1} is guaranteed for α sufficiently small, an appropriate initial guess for α_k is 0. Alternatively, since $f^{(k)}_{ub}(\alpha) = \alpha(c^{(k)}_1 + c^{(k)}_2 \alpha)$ where

$$\begin{aligned}c^{(k)}_1 &= -\left(\frac{[\sigma_n]_k}{2} + 2\operatorname{Re}[\langle \delta M_k, \delta\widetilde{M}_k - \delta M_k \rangle]\right) \\ c^{(k)}_2 &= a_k + \|\delta\widetilde{M}_k - \delta M_k\|^2_F,\end{aligned}$$

is a quadratic function of α, one can compute the minimizer in the interval $[0, 1]$ analytically: $\alpha_k = \min\left\{fracc^{(k)}_1 2c^{(k)}_2, 1\right\}$ if $c^{(k)}_2 \neq 0$, and 0 otherwise.

B. Algorithm 1 Complexity

The computational complexity of Algorithm 1 depends on four parameters which can vary independently: n, m, $\dim(\mathcal{S})$, and $\dim(\mathcal{P})$. The relationship of these parameters to the computational complexity is complex. However, we can derive an expression for a conservative upper bound. Let $r = m - n + 1$. If we assume that $\dim(\mathcal{S}) < m \cdot n$ (conservative for a structured problem), that $r < n$, and that $\dim(\mathcal{P})$ is a small constant (2 in the case of controllability), then a conservative upper bound for the cost per iteration is $O(r \cdot n^4)$ (corresponding to the computation of $(L^{\dagger}_{uV\mathcal{S}} \circ L_{uV\mathcal{S}})$ in (41)). A tight lower bound is $O(n^3)$ (corresponding to an SVD of a $n \times m$ matrix where $m < 2n$). The more structure imposed, the smaller $\dim(\mathcal{S})$. The smaller $\dim(\mathcal{S})$, the closer the complexity will be to the lower bound. The number of iterations required depends on the parameters of the problem: M, $\mathcal{R}$, and $\mathcal{S}$.

V. Convergence of Algorithm 1

The well-posedness of the $\mathcal{P}$–robustness problem and the convergence of Algorithm 1 to a finite property matrix R_* requires a structural assumption on the matrices of the property space $\mathcal{P}$. The problem is that if there exist a non-zero $R \in \mathcal{P}$ for which rank$(R) < n$, then $\sigma_n(M - \delta M - \eta R)$ may be finite (and possibly optimal) as $\eta \to \infty$.

Assumption 2. *Each nonzero property matrix $R \in \mathcal{P}$ is full rank, i.e.,* rank $R = n$.

In consequence, "inf" in (4) and (7) can be replaced with "min".

To simplify the convergence analysis we will make two additional assumptions. A convergent algorithm could be constructed without them, but in this way the discussion is much clearer.

Assumption 3. *The sequence $\{[\sigma_{n-1}]_k\}$ computed in step 4 is bounded away from zero.*

Assumption 3 requires that the $(n-1)^{th}$ singular value, $[\sigma_{n-1}]_k$, be nonzero. It is possible that the solution satisfies rank$(M - \delta M_* - R_*) < n - 1$. However, arbitrarily close to every M for which rank$(M - \delta M_* - R_*) < n - 1$ is a perturbed M for which the solution occurs at $\sigma_{n-1} = \epsilon$ (see Example 4 in Appendix C). In the presence of roundoff, such a perturbation occurs implicitly. Nonetheless, to ensure Assumption 3 an explicit perturbation can be added to M when $[\sigma_{n-1}]_k$ falls below some pre-determined threshold, and the algorithm will converge to a (local) solution to a nearby problem. The perturbation need not respect the structure of $\mathcal{S}$.

Assumption 4. *The sequence $\{g_k\}$ computed in step 8 is bounded away from zero.*

The reasoning for this assumption is motivated by Lemma 8 in Appendix B where it is shown that close to the solution, g_k is bounded away from 0 and that the lower bound essentially measures the surjectivity of $L_{uV\mathcal{S}}$.

The proof of convergence utilizes the three lemmas below. Observe that by Assumption 1, u and V as required in step 4 can always be found and satisfy $u^H u = u^\dagger u = 1$ and $V^H V = V^\dagger V = I$.

Lemma 5. *Let u and V be as in step 4, $[\sigma_n]_k \triangleq \sigma_n(M - \delta M_k - R_k)$ and a_k be as defined in* (47). *Then, since $L_{uV\mathcal{S}}$ is surjective, for all $\alpha \in (0, 1)$,*

$$\sigma_n\left(M - \delta M_k - R_k - \alpha(\delta\widehat{M}_k - \delta M_k + \Delta R_k)\right) \leq (1-\alpha)[\sigma_n]_k + \alpha^2 a_k.$$

Proof. See Appendix A. □

The next lemma constructs an upper bound on the norm $\|\delta M_{k+1}\|_F$ as a function of α. We will require the following linear orthogonal projection operators from the proofs of Lemmas 2 and 3:

$$Q_1 \triangleq (L^\dagger_{uV\mathcal{S}} \circ L_{uV\mathcal{S}}) \quad (61)$$

$$Q_2 \triangleq I - (L^\dagger_{uV\mathcal{S}} \circ L_{uV\mathcal{P}})(L^\dagger_{uV\mathcal{S}} \circ L_{uV\mathcal{P}})^\dagger. \quad (62)$$

Lemma 6. *Let u and V be as in step 4. Since $L_{uV\mathcal{S}}$ is surjective, for all $\alpha \in [0, 1]$*

$$\begin{aligned}\|(1-\alpha)\delta M_k + \alpha\delta\widehat{M}_k\|_F &\leq \|\delta M_k\|_F + \alpha\|\delta\overline{M}_k\|_F \quad (63)\\ &+ b_k\left(\|(1-\alpha)\delta M_k + \alpha(Q_2 \circ Q_1)(\delta M_k)\|_F^2 - \|\delta M_k\|_F^2\right),\end{aligned}$$

where $\delta\overline{M}_k$ and b_k are given in steps 6 and 8, respectively.

Proof. See Appendix A. □

Convergence will be shown by using the Lyapunov function $E_k = [\sigma_n]_k + g_k\|\delta M_k\|_F$. Given Assumptions 2 and 4, we will show that if the necessary conditions of Theorem 1 are not satisfied then i) $E_{k+1} - E_k \leq f_{ub}^{(k)}(\alpha_k) \leq 0$ and ii) $f_{ub}^{(k)}(\alpha_k) < 0$. Since E_k is nonnegative for each k, proving that $\{E_k\}$ is nonincreasing implies it is a bounded monotone function so the sequence converges, i.e., $E_{k+1} - E_k \to 0$. This guarantees that $f_{ub}^{(k)}(\alpha_k) \to 0$. The next lemma shows why this in turn implies convergence to a pair δM_* and R_* satisfying the necessary conditions in Theorem 1.

Lemma 7. *Let $R \in \mathcal{P}$ and $\delta M \in \mathcal{S}$ satisfy* rank$[M - R - \delta M] < n$, *i.e., they are a candidate solution. R and δM satisfy the necessary conditions in Theorem 1 if and only if*

$$(Q_2 \circ Q_1)(\delta M) = \delta M,$$

where Q_1 and Q_2 are given in (61) *and* (62)*, respectively.*

Proof. See Appendix A. □

The next Theorem proves that if Algorithm 1 is carried out to infinite precision, then the algorithm converges to a necessary condition for an optimal solution R_* and δM_*. The stopping conditions in Algorithm 1 guarantee that the algorithm terminates. The parameter ϵ determines how far the terminal points δM_k and R_k are from satisfying the necessary conditions.

Theorem 2. *If Assumptions 2-3 hold, then the sequence $\{E_k\}$ computed by Algorithm 1 converges, where*

$$E_k \triangleq [\sigma_n]_k + g_k\|\delta M_k\|_F. \quad (64)$$

Further, the sequences $\{\delta M_k\}$ and $\{R_k\}$ have limit points δM_* and R_* satisfying the necessary conditions of Theorem 1.

Proof. First we show that $E_{k+1} - E_k \leq f_{ub}^{(k)}(\alpha_k)$. Because g_k is nonincreasing, Lemma 5 and 6 imply that

$$\begin{aligned}&E_{k+1} - E_k\\ &= [\sigma_n]_{k+1} + g_{k+1}\|\delta M_{k+1}\|_F - ([\sigma_n]_k + g_k\|\delta M_k\|_F)\\ &\leq -\alpha_k[\sigma_n]_k + a_k\alpha_k^2 + \alpha_k g_k\|\delta\overline{M}_k\|_F - g_k b_k\|\delta M_k\|_F^2\\ &\quad + g_k b_k\|(1-\alpha_k)\delta M_k + \alpha_k(Q_2 \circ Q_1)(\delta M_k)\|_F^2\end{aligned}$$

Since $g_k \leq [\sigma_n]_k/(2\|\delta\overline{M}_k\|_F)$,

$$\begin{aligned}&E_{k+1} - E_k\\ &\leq \frac{-[\sigma_n]_k}{2}\alpha_k + a_k\alpha_k^2 - g_k b_k\|\delta M_k\|_F^2\\ &\quad + g_k b_k\|(1-\alpha_k)\delta M_k + \alpha_k(Q_2 \circ Q_1)(\delta M_k)\|_F^2\\ &\triangleq f_{ub}^{(k)}(\alpha_k).\end{aligned}$$

Note that $f_{ub}^{(k)}(\alpha)$ is a quadratic function of α and $f_{ub}^{(k)}(0) = 0$. Therefore, there exist constants $c_1^{(k)}$ and $c_2^{(k)}$ such that

$$f_{ub}^{(k)}(\alpha) = c_1^{(k)}\alpha + c_2^{(k)}\alpha^2.$$

Careful inspection of $f_{ub}^{(k)}(\alpha)$ shows that $c_2^{(k)} \geq 0$, i.e., $f_{ub}^{(k)}(\alpha)$ admits a global minimum. Since g_k, b_k, and $[\sigma_n]_k$ are all nonnegative, $c_1^{(k)} \leq 0$ if the coefficient of the linear term in the quadratic

$$\|(1-\alpha_k)\delta M_k + \alpha_k(Q_2 \circ Q_1)(\delta M_k)\|_F^2 - \|\delta M_k\|_F^2$$

is nonpositive. This is clearly the case since $\|(Q_2 \circ Q_1)(\delta M_k)\|_F \leq \|\delta M_k\|_F$. Further, $c_1^{(k)} = 0$ if and only if $[\sigma_n]_k = 0$ and $(Q_2 \circ Q_1)(\delta M_k) = \delta M_k$ since $g_k > 0$ by Assumption 4. Equivalently, Lemma 7 implies that $c_1^{(k)} = 0$ if and only if the necessary conditions are satisfied. Since α_k is chosen to minimize $f_{ub}^{(k)}$ over the interval $[0,1]$, $f_{ub}^{(k)}(\alpha_k) < 0$ so long as the necessary conditions are not satisfied.

Thus $\{E_k\}$ is nonnegative and decreasing since $E_{k+1} - E_k \leq f_{ub}^{(k)}(\alpha_k) \leq 0$. By the monotone convergence theorem, $\{E_k\}$ converges, i.e., $E_{k+1} - E_k \to 0$. To prove that we converge to a necessary condition, we will prove that the sequence $\{c_1^{(k)}\}$ converges to zero. Based on Lemma 7, this implies that a necessary condition is satisfied.

Since $E_{k+1} - E_k \leq f_{ub}^{(k)}(\alpha_k) \leq 0$ and $E_{k+1} - E_k \to 0$, the sequence $\{f_{ub}^{(k)}(\alpha_k)\} \to 0$. As long as $\{c_2^{(k)}\}$ is bounded, this implies that $\{c_1^{(k)}\} \to 0$ as desired. The sequence $\{c_2^{(k)}\}$ is unbounded only if the quadratic coefficient of the function

$$\begin{aligned} &a_k\alpha_k^2 - g_k b_k\|\delta M_k\|_F^2 \\ &\quad + g_k b_k\|(1-\alpha_k)\delta M_k + \alpha_k(Q_2 \circ Q_1)(\delta M_k)\|_F^2 \end{aligned} \tag{65}$$

goes unbounded. The quadratic term coefficient in (65) is given by $a_k + g_k b_k\|(I - Q_2 \circ Q_1)\delta M_k\|^2$ and by construction $b_k\|(I - Q_2 \circ Q_1)\delta M_k\|_F^2 \leq b_k\|\delta M_k\|_F^2 \leq \|\delta M_k\|_F$. Since $\{E_k\}$ converges and $\{g_k\} > 0$, $\|\delta M_k\|_F$ is bounded. By (47), $a_k \leq \|\widehat{\delta M_k} - \delta M_k - \Delta R_k\|_F^2/[\sigma_{n-1}]_k$, which by Assumption 3 is bounded if ΔR_k is bounded, or equivalently if R_k is bounded.

Assume for contradiction that $\{R_k\}$ is unbounded. Since $\{E_k\}$ converges, $[\sigma_n]_k$ is bounded. Since $\{g_k\} > 0$, $\{\delta M_k\}$ is bounded as well. Let R_{min} be the norm one property matrix minimizing σ_n, i.e., $R_{min} = \operatorname{argmin}_{R \in \mathcal{P}, \|R\|_F = 1} \sigma_n(R)$. By Assumption 2, $\sigma_n(R_{min}) > 0$ and for any $R \in \mathcal{P}$, $\sigma_n(R) \geq \|R\|_F \sigma_n(R_{min})$. Hence, letting u_k be the n^{th} lsv of $M - R_k - \delta M_k$, for sufficiently large k

$$\begin{aligned} [\sigma_n]_k &= \|u_k^H(M - R_k - \delta M_k)\|_2 \\ &\geq \|u_k^H R_k\|_2 - \|u_k^H(M - \delta M_k)\|_2 \\ &\geq \sigma_n(R_k) - \|u_k^H(M - \delta M_k)\|_2 \\ &\geq \|R_k\|_F \sigma_n(R_{min}) - \|u_k^H(M - \delta M_k)\|_2. \end{aligned}$$

Hence, if $\|R_k\|_F \to \infty$, then $[\sigma_n]_k \to \infty$ contradicting that $\{E_k\}$ converges. Hence $\{\Delta R_k\}$ is bounded and thus $\{a_k\}$ and $\{c_2^{(k)}\}$ are bounded. This implies that as $k \to \infty$, $(Q_2 \circ Q_1)(\delta M_k) \to \delta M_k$, i.e., the two necessary conditions in Theorem 1, become satisfied. Finally, since $\{g_k\} > 0$ and $\{E_k\}$ converges, the sequence of perturbations $\{\delta M_k\}$ has a bounded accumulation point δM_*. Since δM_* satisfies the two necessary conditions, the sequence $\{\Delta R_k\}$ has an accumulation point $\Delta R_* = 0$, i.e., $R_k \to R_*$, completing the proof. □

VI. Numerical Examples

A. Example 1 (continued)

Consider the third example in [3] (also appears in [11] and [8]), which in the $\mathcal{P}$–robustness framework has system matrix M, structured perturbations δM, and property matrices R given by

$$\begin{aligned} M &= \begin{bmatrix} A & B \end{bmatrix} \\ \delta M &= \begin{bmatrix} \delta A & \delta B \end{bmatrix} \in \mathbb{R}^{3\times 4} \\ R &= \lambda \begin{bmatrix} I & 0 \end{bmatrix} \in \mathbb{C}^{3\times 4}, \end{aligned}$$

where

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0.1 & 3 & 5 \\ 0 & -1 & -1 \end{bmatrix}, \; B = \begin{bmatrix} 1 \\ 0.1 \\ 0 \end{bmatrix}.$$

With initial guesses $\delta M_0 = 0$ and $R_0 = [jI, 0]$, and $\epsilon = 10^{-10}$, Algorithm 1 terminates in 9 iterations. The $\mathcal{P}$–robustness of M with respect to parameter variations in $\mathcal{S}$ is computed to be $r(M; \mathcal{S}, \mathcal{P}) = 0.057737$. The minimizing property and perturbation matrices are $R_* = (0.9824 + 0.9731j)[I, 0]$ and $\delta M_* = [\delta A_*, \delta B_*]$, respectively, where

$$\begin{aligned} \delta A_* &= 10^{-4}\begin{bmatrix} -5.8878 & -0.49659 & 0.29287 \\ 168.48 & 14.210 & -8.3803 \\ 167.31 & 14.111 & -8.3221 \end{bmatrix} \\ \delta B_*^\top &= 10^{-3}\begin{bmatrix} 1.1427 & 15.754 & 49.685 \end{bmatrix}. \end{aligned}$$

Upon termination, $\sigma_n(M - \delta M_* - R_*) = 2.8944 \times 10^{-16} \approx 0$. These results are consistent with [3] and [8]. As noted in [8], we cannot compare the results of this example to [11] due to the different norm used therein (largest singular value of δM versus the Frobenius norm).

B. Example 2 (continued)

Consider the example in [9], which in the $\mathcal{P}$–robustness framework has system matrix M, structured perturbations δM, and property matrices R given by

$$\begin{aligned} M &= \begin{bmatrix} A_0^\top & 0 & C_0^\top \\ 0 & A_1^\top & C_1^\top \end{bmatrix} \\ \delta M &= \begin{bmatrix} \delta A_0^\top & 0 & \delta C_0^\top \\ 0 & \delta A_1^\top & \delta C_1^\top \end{bmatrix} \in \mathbb{R}^{4\times 5} \\ R &= \lambda \begin{bmatrix} I & 0 \end{bmatrix} \in \mathbb{C}^{4\times 5}, \end{aligned}$$

where

$$\begin{aligned} A_0 &= \begin{bmatrix} -1 & 2 \\ 0 & -2 \end{bmatrix}, & C_0 &= \begin{bmatrix} 1 & 0 \end{bmatrix}, \\ A_1 &= \begin{bmatrix} -3 & 0.1 \\ 5 & -1 \end{bmatrix}, & C_1 &= \begin{bmatrix} 1 & 1 \end{bmatrix}. \end{aligned}$$

The perturbation space $\mathcal{S}$ is real and does not allow perturbations of the off-diagonal entries of M. In [9], the distance to the nearest SMS SLTI system is computed to satisfy

$$0.0506 \leq r(M; \mathcal{S}, \mathcal{P}) \leq 0.4570.$$

Setting the terminating condition $\epsilon = 10^{-15}$ and initial guesses $\delta M_0 = 0$ and $R_0 = 0$, the Algorithm 1 terminated in 13 iterations. For reference, this algorithm took 135 ms using an Intel Core i5-2410M processor. The computed distance is $r(M; \mathcal{S}, \mathcal{P}) = 0.071821$ where $R_* = -0.9065 \begin{bmatrix} I & 0 \end{bmatrix}$,

$$\delta M_* = 10^{-3} \begin{bmatrix} -21.5 & -39.3 & 0 & 0 & 0 \\ -0.8 & -1.4 & 0 & 0 & 0 \\ 0 & 0 & 1.2 & 0.5 & 0 \\ 0 & 0 & 51.8 & 21.7 & 0 \end{bmatrix},$$

and $\sigma_n(M - \delta M_* - R_*) = 5.256 \times 10^{-16} \approx 0$. Note that $r(M; \mathcal{S}, \mathcal{P}) = 0.071821$ is between 0.0506 and 0.4570.

C. Example 3

Consider the $\mathcal{P}$–robustness problem given by

$$M = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 1 & -1 & 1 & 0 & 1 \end{bmatrix}$$
$$\delta M = \begin{bmatrix} \delta m_{11} & 0 & 0 & 0 & 0 \\ \delta m_{21} & 0 & 0 & 0 & 0 \\ \delta m_{31} & 0 & 0 & \delta m_{34} & 0 \\ \delta m_{41} & \delta m_{42} & \delta m_{43} & \delta m_{44} & \delta m_{45} \end{bmatrix} \in \mathbb{R}^{4 \times 5}$$
$$R = \lambda \begin{bmatrix} 0 & I \end{bmatrix} \in \mathbb{C}^{4 \times 5},$$

Setting the terminating condition $\epsilon = 10^{-15}$ and initial guesses

$$\delta M_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and $R_0 = 2j[0, I]$, the Algorithm 1 terminated in 93 iterations. For reference, this algorithm took 1.36 s using an Intel Core i5-2410M processor. The distance $r(M; \mathcal{S}, \mathcal{P})$ is computed to be $r(M; \mathcal{S}, \mathcal{P}) = 0.66548$ where $R_* = 0.6988 \begin{bmatrix} 0 & I \end{bmatrix}$,

$$\delta M_* = \begin{bmatrix} -0.3245 & 0 & 0 & 0 & 0 \\ 0.4644 & 0 & 0 & 0 & 0 \\ 0.2376 & 0 & 0 & 0.2558 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and $\sigma_n(M - \delta M_* - R_*) = 1.1186 \times 10^{-16} \approx 0$.

VII. Conclusion

In this work, the $\mathcal{P}$–robustness framework developed in [19] is used to solve a family of robustness problems. Specifically, the Frobenius norm metric is used to measure the $\mathcal{P}$–robustness of M with respect to perturbations in $\mathcal{S}$. Necessary conditions for a minimal rank reducing perturbation are proven in Theorem 1. The necessary conditions motivate Algorithm 1 for computing the metric $r(M; \mathcal{S}, \mathcal{P})$ as well as the minimizing property matrix R_* and perturbation matrix δM_*. For $M \in \mathbb{C}^{n \times m}$, the computational complexity of each iteration of the algorithm is at most $O((m - n + 1) \cdot n^4)$.

In future work, we will modify Algorithm 1 to solve $\mathcal{P}$–robustness problems with singular property matrices, i.e., $\text{rank}(R) < n$. This modification will address the case where the norm of the optimal property matrix R_* is unbounded. In addition, we expect that Algorithm 1 can be modified to compute the $\mathcal{P}$–robustness of M using the spectral norm metric, i.e., minimizing $\sigma_1(\delta M_*)$. Although the Frobenius norm may be a more accurate measure of robustness, extending the work to the spectral norm metric would unify the robustness property literature.

Appendix

A. Additional Proofs

Proof of Proposition 2. Let $\zeta_0 \in \mathbb{R}^k$ satisfy $\text{vec}(\delta M) = B_{\mathcal{S}} \zeta_0$. Then,

$$\begin{aligned} \text{vec}(L_{uV\mathcal{S}}(\delta M)) &= (V^\top \otimes u^H)\, \text{vec}(\delta M) \\ &= (V^\top \otimes u^H) B_{\mathcal{S}} \zeta_0. \end{aligned} \tag{66}$$

Let $y \in \mathbb{C}^{m-n+1}$ be an arbitrary vector. Using (66), $L_{uV\mathcal{S}}(\delta M) = y^\top$ if and only if

$$\begin{bmatrix} \text{Re}(y) \\ \text{Im}(y) \end{bmatrix} = \begin{bmatrix} \text{Re}((V^\top \otimes u^H) B_{\mathcal{S}}) \\ \text{Im}((V^\top \otimes u^H) B_{\mathcal{S}}) \end{bmatrix} \zeta_0. \tag{67}$$

$L_{uV\mathcal{S}}$ is surjective if and only if for each $y \in \mathbb{C}^{m-n+1}$, there exists $\zeta_0 \in \mathbb{R}^k$ such that (67) holds. Hence, $L_{uV\mathcal{S}}$ is surjective if and only if

$$\begin{bmatrix} \text{Re}((V^\top \otimes u^H) B_{\mathcal{S}}) \\ \text{Im}((V^\top \otimes u^H) B_{\mathcal{S}}) \end{bmatrix}$$

is full row rank, i.e., (13) is satisfied. □

Proof of Lemma 1.

Step 1: First we show that every analytic path $g_a : [0, 1] \to W$ satisfying $g_a(s) \in W_1$ for $s < 0.5$ and $g_a(s) \in W_2$ for $s > 0.5$ has an analytic unsigned n^{th} singular value function $f_a : [0, 1] \to \mathbb{R}$ such that $f_a(s) = \widetilde{H}(g_a(s))$, i.e., $\widetilde{H}$ is an unsigned n^{th} singular value function for all such paths. Without loss of generality assume $f_a(0) = \widetilde{H}(g_a(0))$. Recall $\widetilde{H}$ was constructed to be consistent with f_g, the unsigned n^{th} singular value function associated with the curve g. Construct a continuous closed path Γ by connecting $g_a(0)$ with $g(0)$ in W_1 and $g_a(1)$ with $g(1)$ in W_2. Since (i) σ_n is continuous and (ii) σ_n is nonzero on $W_1 \cup W_2$, then σ_n is continuous on Γ and thus $\text{sign}(f_a(1)) = \text{sign}(f_g(1))$, i.e., $f_a(s) = \widetilde{H}(g_a(s))$ for all $s \in [0, 1]$ as desired. Consequently, $\widetilde{H}$ can be used for an unsigned n^{th} singular value for any analytic curve in W.

Step 2: Next we prove that $\widetilde{H}$ is Fréchet differentiable on $W_1 \cup W_2$ with partial derivatives given in (23). Let $(\tilde{\zeta}, \tilde{\rho}) \in W_1 \cup W_2$. Then the n^{th} singular value of $M(\tilde{\zeta}, \tilde{\rho})$ is simple and the n^{th} lsv and rsv are unique up to multiplication by unitary scalars. By [22, Theorem 3], there exists a simply connected neighborhood $W_0 \subset W_1 \cup W_2$ of $(\tilde{\zeta}, \tilde{\rho})$ and analytic (unsigned) singular vector functions $\tilde{u} : W_0 \to \mathbb{C}^n$ and $\tilde{v} : W_0 \to \mathbb{C}^m$ such that for all $(\zeta, \rho) \in W_0$,

$$\tilde{u}^H(\zeta, \rho) M(\zeta, \rho) \tilde{v}(\zeta, \rho) = \widetilde{H}(\zeta, \rho).$$

Since $\tilde{u}$, $\tilde{v}$, and $M(\zeta, \rho)$ are analytic, so is $\widetilde{H}$.

To take the derivative $\frac{\partial \widetilde{H}(\tilde{\zeta},\tilde{\rho})}{\partial \zeta_i}$, we consider replacing ζ with a complex argument z.[5] The complex partial derivative with respect to z_i at $z=\tilde{\zeta}$ is

$$\begin{aligned}\frac{\partial \widetilde{H}(z,\tilde{\rho})}{\partial z_i}\bigg|_{z=\tilde{\zeta}} &= \frac{\partial \tilde{u}^H}{\partial z_i} M(\tilde{\zeta},\tilde{\rho})\tilde{v} + \tilde{u}^H \frac{\partial M(z,\tilde{\rho})}{\partial z_i}\tilde{v} \\ &\quad + \tilde{u}^H M(\tilde{\zeta},\tilde{\rho})\frac{\partial \tilde{v}}{\partial z_i} \\ &= \widetilde{H}(\tilde{\zeta},\tilde{\rho})\left(\frac{\partial \tilde{u}^H}{\partial z_i}\tilde{u} + \tilde{v}^H\frac{\partial \tilde{v}}{\partial z_i}\right) - \tilde{u}^H S_i \tilde{v},\end{aligned}$$

since $\tilde{u}$ and $\tilde{v}$ are singular vector functions, i.e.,

$$M(\tilde{\zeta},\tilde{\rho})\tilde{v}(\tilde{\zeta},\tilde{\rho}) = \widetilde{H}(\tilde{\zeta},\tilde{\rho})\tilde{u}(\tilde{\zeta},\tilde{\rho})$$

and

$$\tilde{u}^H(\tilde{\zeta},\tilde{\rho})M(\tilde{\zeta},\tilde{\rho}) = \widetilde{H}(\tilde{\zeta},\tilde{\rho})\tilde{v}^H(\tilde{\zeta},\tilde{\rho}).$$

However, since $1=\tilde{u}^H\tilde{u}$, we obtain $0 = \frac{\partial \tilde{u}^H}{\partial z_i}\tilde{u} + \left(\frac{\partial \tilde{u}^H}{\partial z_i}\tilde{u}\right)^H$. Hence Re $\left(\frac{\partial \tilde{u}^H}{\partial z_i}\tilde{u}\right) = 0$. Similarly, Re $\left(\tilde{v}^H\frac{\partial \tilde{v}}{\partial z_i}\right) = 0$. The real derivative of $\widetilde{H}$ with respect to ζ_i is given by

$$\begin{aligned}\frac{\partial \widetilde{H}(\tilde{\zeta},\tilde{\rho})}{\partial \zeta_i} &= \text{Re}\left[\frac{\partial \widetilde{H}(z,\tilde{\rho})}{\partial z_i}\right] \\ &= -\,\text{Re}(\tilde{u}^H(\tilde{\zeta},\tilde{\rho})S_i\tilde{v}(\tilde{\zeta},\tilde{\rho})).\end{aligned}$$

The same argument holds for computing $\frac{\partial \widetilde{H}(\tilde{\zeta},\tilde{\rho})}{\partial \rho_i} = -\,\text{Re}(\tilde{u}^H(\tilde{\zeta},\tilde{\rho})P_i\tilde{v}(\tilde{\zeta},\tilde{\rho}))$.

Step 3: What remains is to show that $\widetilde{H}$ is Fréchet differentiable for $(\zeta_0,\rho_0) \in W \setminus W_1 \cup W_2$, i.e., points where $M(\zeta_0,\rho_0)$ drops rank. Although $\sigma_n(M(\zeta_0,\rho_0)) = 0$, the n^{th} lsv is unique (up to unitary scalar multiplication) since the last singular value is simple and $M(\zeta_0,\rho_0)$ has fewer rows than columns. However, the n^{th} (unsigned) rsv is not unique since $M(\zeta_0,\rho_0)$ has a right null space of dimension $m-n+1$. However, not all n^{th} (unsigned) rsv can be a continuation of an analytic n^{th} rsv function for analytic paths passing through (ζ_0,ρ_0). Let $g_b : [0,1] \to W$ be an analytic curve from W_1 to W_2 with $g_b(0.5) = (\zeta_0,\rho_0)$. By [22, Theorem 3], there exists analytic n^{th} (unsigned) lsv and rsv functions $u_b : [0,1] \to \mathbb{C}^n$ and $v_b : [0,1] \to \mathbb{C}^m$, respectively, associated with the unsigned n^{th} singular value function $\widetilde{H}(g_b(\cdot))$. Since for each $s \neq 0.5$, $M(g_b(s))$ is a fixed matrix with a nonzero n^{th} singular value, the product $u_b(s)v_b^H(s)$ is unique for all $s \in [0,1]$ (even though $u_b(s)$ and $v_b(s)$ are not unique). The same uniqueness result holds for analytic curves from W_2 to W_1 passing through (ζ_0,ρ_0). For analytic paths in $W \setminus W_1 \cup W_2$ passing through (ζ_0,ρ_0), the right null space of $M(\zeta,\rho)$ changes analytically and hence one can choose $u_b(0.5)$ and $v_b(0.5)$ as the n^{th} unsigned lsv and rsv along these paths as well. Because W, W_1, and W_2 are simply connected, there exists a neighborhood $W_0 \subset W$ of (ζ_0,ρ_0) and analytic singular vector functions $u_0 : W_0 \to \mathbb{C}^n$ and $v_0 : W_0 \to \mathbb{C}^m$ such that

$$u_0^H(\zeta,\rho)M(\zeta,\rho)v_0(\zeta,\rho) = \widetilde{H}(\zeta,\rho)$$

for all $(\zeta,\rho) \in W_0$. Using the same arguments in step 2, it follows that $\widetilde{H}$ is Fréchet differentiable in W, specifically at points $(\zeta_0,\rho_0) \in W \setminus W_1 \cup W_2$. Finally, the associated partial derivatives are given in (23). □

Proof of Lemma 2.

Let $\delta\widetilde{M} \triangleq (L_{uV\mathcal{S}}^{\dagger} \circ L_{uV})(M - R_0)$. Since $L_{uV\mathcal{S}}$ is surjective, $L_{uV\mathcal{S}}(\delta\widetilde{M}) = L_{uV}(M - R_0)$. Since $\sigma_n(M - \delta M_0 - R_0) = 0$, $L_{uV\mathcal{S}}(\delta M_0) = L_{uV}(M - R_0)$. Hence, defining $\Delta M \triangleq \delta\widetilde{M} - \delta M_0$, we obtain

$$L_{uV\mathcal{S}}(\Delta M) = L_{uV\mathcal{S}}(\delta\widetilde{M}) - L_{uV\mathcal{S}}(\delta M_0) = 0.$$

Observe that by definition $\Delta M = -(I - L_{uV\mathcal{S}}^{\dagger} \circ L_{uV\mathcal{S}})(\delta M_0) = -(I - Q_1)(\delta M_0)$, where the linear operator Q_1 is

$$Q_1 \triangleq (L_{uV\mathcal{S}}^{\dagger} \circ L_{uV\mathcal{S}}). \tag{68}$$

By the properties of the Moore-Penrose pseudoinverse, Q_1 is an orthogonal projection on $\mathcal{S}$ with respect to the inner product $\langle \cdot,\cdot \rangle$, i.e., Q_1 is a self-adjoint linear operator and $Q_1^2 = Q_1$. In addition, $I - Q_1$ is also an orthogonal projection on $\mathcal{S}$. Hence,

$$\begin{aligned}\langle \delta M_0, \Delta M \rangle &= -\langle \delta M_0, (I - Q_1)(\delta M_0) \rangle \\ &= -\langle (I - Q_1)(\delta M_0), (I - Q_1)(\delta M_0) \rangle \\ &= -\|(I - Q_1)(\delta M_0)\|_F^2 \\ &= -\|\Delta M\|_F^2\end{aligned}$$

Since $\Delta M \neq 0$ by the assumption that $\delta M_0 \neq \delta\widetilde{M}$, $\langle \delta M_0, \Delta M \rangle = -\|\Delta M\|_F^2 < 0$. □

Proof of Lemma 3.

Equation (27) states that ΔR is the matrix in $\mathcal{P}$ minimizing

$$\|(L_{uV\mathcal{S}}^{\dagger} \circ L_{uV\mathcal{P}})R - \delta M_0\|_F \tag{69}$$

for all $R \in \mathcal{P}$. Hence, if $(L_{uV\mathcal{S}}^{\dagger} \circ L_{uV\mathcal{P}})(\Delta R) = 0$ then $\Delta R' = 0$ also minimizes (69), contradicting (27). Consequently, $(L_{uV\mathcal{S}}^{\dagger} \circ L_{uV\mathcal{P}})(\Delta R) \neq 0$ by the properties of the Moore-Penrose pseudoinverse. Let $\Delta M \in \mathcal{S}$ be defined as $\Delta M \triangleq -(L_{uV\mathcal{S}}^{\dagger} \circ L_{uVP})(\Delta R) \neq 0$. Since $L_{uV\mathcal{S}}$ is surjective, $L_{uV\mathcal{S}}(\Delta M) = -L_{uV\mathcal{P}}(\Delta R)$. Hence, by linearity

$$L_{uV}(\Delta M + \Delta R) = L_{uV\mathcal{S}}(\Delta M) + L_{uV\mathcal{P}}(\Delta R) = 0.$$

Since $\delta M_0 = (L_{uV\mathcal{S}}^{\dagger} \circ L_{uV})(M - R_0)$,

$$\begin{aligned}\Delta M &= -(L_{uV\mathcal{S}}^{\dagger} \circ L_{uV\mathcal{P}})(\Delta R) \\ &= -(L_{uV\mathcal{S}}^{\dagger} \circ L_{uV\mathcal{P}})(L_{uV\mathcal{S}}^{\dagger} \circ L_{uV\mathcal{P}})^{\dagger}\delta M_0 \\ &\triangleq -(I - Q_2)(\delta M_0),\end{aligned}$$

where the linear operator Q_2 is

$$Q_2 \triangleq I - (L_{uV\mathcal{S}}^{\dagger} \circ L_{uV\mathcal{P}})(L_{uV\mathcal{S}}^{\dagger} \circ L_{uV\mathcal{P}})^{\dagger}. \tag{70}$$

By the properties of the Moore-Penrose pseudoinverse, Q_2 is an orthogonal projection on $\mathcal{S}$ with respect to the inner product $\langle \cdot,\cdot \rangle$, i.e., Q_2 is a self-adjoint linear operator and $Q_2^2 = Q_2$. In

[5] Rigorously, we should define a new function with a complex domain, but we have chosen to keep the presentation more direct.

addition, $I - Q_2$ is also an orthogonal projection on $\mathcal{S}$. This implies

$$\begin{aligned}\langle \delta M_0, \Delta M\rangle &= -\langle \delta M_0, (I - Q_2)(M_0)\rangle \\ &= -\langle (I - Q_2)(\delta M_0), (I - Q_2)(M_0)\rangle \\ &= -\|(I - Q_2)(\delta M_0)\|_F^2 \\ &= -\|\Delta M\|_F^2 < 0\end{aligned}$$

as was to be shown. □

Proof of Lemma 4.

By definition, f is differentiable with derivative $f'(\zeta_0, \rho_0) = [\zeta_0^\top, 0]$. Since the Fréchet differential $\widetilde{H}'(\zeta_0, \rho_0)$ exists by Lemma 1, T is Fréchet differentiable.

It remains to show that $T'(\zeta_0, \rho_0)$ is surjective. Let $y = [y_1, y_2]^\top \in \mathbb{R}^2$ be an arbitrary vector. Since $\frac{\partial}{\partial \zeta}\widetilde{H}(\zeta, \rho_0)|_{\zeta=\zeta_0}$ is surjective, there exists $\zeta_1 \in \mathbb{R}^k$ such that $\frac{\partial}{\partial \zeta}\widetilde{H}(\zeta, \rho_0)|_{\zeta=\zeta_0}\zeta_1 = y_2$. Thus since $f'(\zeta_0, \rho_0) = [\zeta_0^\top, 0]$,

$$T'(\zeta_0, \rho_0)\begin{bmatrix}\zeta_1 \\ 0\end{bmatrix} = \begin{bmatrix}\zeta_0^\top \zeta_1 \\ y_2\end{bmatrix}.$$

By assumption, there exists ζ_Δ and ρ_Δ such that $\zeta_0^\top \zeta_\Delta < 0$ and $\widetilde{H}'(\zeta_0, \rho_0)[\zeta_\Delta^\top, \rho_\Delta^\top]^\top = 0$. Define $\zeta_2 \in \mathbb{R}^k$ and $\rho_2 \in \mathbb{R}^r$ by

$$\begin{aligned}\zeta_2 &= \left(\frac{y_1 - \zeta_0^\top \zeta_1}{\zeta_0^\top \zeta_\Delta}\right)\zeta_\Delta \\ \rho_2 &= \left(\frac{y_1 - \zeta_0^\top \zeta_1}{\zeta_0^\top \zeta_\Delta}\right)\rho_\Delta\end{aligned}$$

Then by linearity,

$$\begin{aligned}T'(\zeta_0, \rho_0)\begin{bmatrix}\zeta_1 + \zeta_2 \\ \rho_2\end{bmatrix} &= \begin{bmatrix}\zeta_0^\top \zeta_1 \\ y_2\end{bmatrix} + \begin{bmatrix}y_1 - \zeta_0^\top \zeta_1 \\ 0\end{bmatrix} \\ &= y.\end{aligned}$$

Since y was arbitrary, $T'(\zeta_0, \rho_0)$ is surjective. □

Proof of Lemma 5.

Let $\widehat{v}_k^H = [u_n]_k^H(\widehat{\delta M}_k - \delta M_k + \Delta R_k)(I - V_k V_k^H)$ and let $\widehat{u}_k$ satisfy $\widehat{u}_k^H(M - \delta M_k - R_k) = \widehat{v}_k^H$; such a $\widehat{u}_k$ exists since $\widehat{v}_k$ is in the row space of $M - \delta M_k - R_k$. Consider the product

$$\begin{aligned}([u_n]_k &+ \alpha \widehat{u}_k)^H(M - \delta M_k - R_k - \alpha(\widehat{\delta M}_k - \delta M_k + \Delta R_k)) \\ &= [u_n]_k^H(M - \delta M_k - R_k) \\ &\quad - \alpha\left(\widehat{v}_k^H - [u_n]_k^H(\widehat{\delta M}_k - \delta M_k + \Delta R_k)\right) \\ &\quad - \alpha^2 \widehat{u}_k^H(\widehat{\delta M}_k - \delta M_k + \Delta R_k) \\ &= (1 - \alpha)L_{uV}(M - \delta M_k - R_k)V_k^H \\ &\quad + \alpha L_{uV}(M - \widehat{\delta M}_k - R_k - \Delta R_k)V_k^H \\ &\quad - \alpha^2 \widehat{u}_k^H(\widehat{\delta M}_k - \delta M_k + \Delta R_k).\end{aligned} \tag{71}$$

Since $L_{uV\mathcal{S}}$ is surjective, step 6 of Algorithm 1 guarantees $L_{uV}(M - \widehat{\delta M}_k - R_k - \Delta R_k) = 0$. Recall that for all $u \in \mathbb{C}^n$, $\widetilde{M} \in \mathbb{C}^{n\times m}$, $u^H \widetilde{M} \geq \sigma_n(\widetilde{M})\|u\|$ (See [25, Corollary 9.6.7]). Combining this with the fact that $[u_n]_k$ and $\widehat{u}_k$ are orthogonal, implying $\|[u_n]_k + \alpha \widehat{u}_k\| \geq \|[u_n]_k\| = 1$, the norm of the left-hand side of (71) upper bounds $\sigma_n(M - \delta M_k - R_k - \alpha(\widehat{\delta M}_k - \delta M_k + \Delta R_k))$. Taking the norm of both sides of (71) and applying the triangle inequality results in the statement of the lemma. □

Proof of Lemma 6.

By definition of $\overline{\delta M}_k$, $\widehat{\delta M}_k = (Q_2 \circ Q_1)(\delta M_k) + \overline{\delta M}_k$. If $\delta M_k = 0$, $(Q_2 \circ Q_1)(\delta M_k) = 0 = \overline{\delta M}_k$ and (63) holds trivially. Assume $\delta M_k \neq 0$. Then since Q_1 and Q_2 are orthogonal projections $\|(Q_2 \circ Q_1)(\delta M_k)\|_F \leq \|\delta M_k\|_F$. Hence,

$$\begin{aligned}\|(1 - \alpha)&\delta M_k + \alpha(Q_2 \circ Q_1)(\delta M_k)\|_F - \|\delta M_k\|_F \\ &= \frac{\|(1 - \alpha)\delta M_k + \alpha(Q_2 \circ Q_1)(\delta M_k)\|_F^2 - \|\delta M_k\|_F^2}{\|(1 - \alpha)\delta M_k + \alpha(Q_2 \circ Q_1)(\delta M_k)\|_F + \|\delta M_k\|_F} \\ &\leq \frac{\|(1 - \alpha)\delta M_k + \alpha(Q_2 \circ Q_1)(\delta M_k)\|_F^2 - \|\delta M_k\|_F^2}{2\|\delta M_k\|_F}.\end{aligned} \tag{72}$$

Since $\widehat{\delta M}_k = (Q_2 \circ Q_1)(\delta M_k) + \overline{\delta M}_k$, the triangle inequality implies that

$$\begin{aligned}\|(1 - \alpha)&\delta M_k + \alpha \widehat{\delta M}_k\|_F \\ &\leq \|(1 - \alpha)\delta M_k + \alpha(Q_2 \circ Q_1)(\delta M_k)\|_F + \alpha\|\overline{\delta M}_k\|_F.\end{aligned}$$

Applying (72) yields the desired result. □

Proof of Lemma 7.

To prove necessity, assume R and δM satisfy necessary conditions i) and ii) of Theorem 1. Then by i)

$$\begin{aligned}Q_1(\delta M) &= (L_{uV\mathcal{S}}^\dagger \circ L_{uV\mathcal{S}})\delta M \\ &= (L_{uV\mathcal{S}}^\dagger \circ L_{uV})(M - R) \\ &= \delta M,\end{aligned}$$

i.e., $Q_1(\delta M) = \delta M$. By necessary condition ii),

$$0 = \Delta R \triangleq (L_{uV\mathcal{S}}^\dagger \circ L_{uV\mathcal{P}})^\dagger (L_{uV\mathcal{S}}^\dagger \circ L_{uV})(M - R).$$

Using the definitions of Q_1 and Q_2, we have

$$\begin{aligned}(Q_2 \circ Q_1)(\delta M) &= Q_1(\delta M) + (L_{uV\mathcal{S}}^\dagger \circ L_{uV\mathcal{P}})\Delta R \\ &= \delta M.\end{aligned}$$

Thus $(Q_2 \circ Q_1)(\delta M) = \delta M$ as desired.

Now for sufficiency, assume that $(Q_2 \circ Q_1)(\delta M) = \delta M$. Since Q_1 and Q_2 are orthogonal projections

$$\|\delta M\|_F = \|(Q_2 \circ Q_1)(\delta M)\|_F \leq \|Q_1(\delta M)\|_F \leq \|\delta M\|_F,$$

implying that equality holds. Thus $\|\delta M\|_F = \|Q_1(\delta M)\|_F$ and this implies that $\delta M = Q_1(\delta M)$ since Q_1 is an orthogonal projection. Similarly, we can show that $\delta M = Q_2(\delta M)$ since

$$\|\delta M\|_F = \|(Q_2 \circ Q_1)(\delta M)\|_F = \|Q_2(\delta M)\|_F \leq \|\delta M\|_F.$$

Since $Q_1(\delta M) = \delta M$, δM satisfies the first necessary condition in Theorem 1. What remains is to show that $\Delta R = 0$. Since $Q_2(\delta M) = \delta M$ and $L_{uV}(M - R) = L_{uV\mathcal{S}}\delta M$,

$$\begin{aligned}\Delta R &= (L_{uV\mathcal{S}}^\dagger \circ L_{uV\mathcal{P}})^\dagger \delta M \\ &= (L_{uV\mathcal{S}}^\dagger \circ L_{uV\mathcal{P}})^\dagger Q_2(\delta M).\end{aligned}$$

Thus by definition of Q_2, $\Delta R = T^\dagger \delta M - T^\dagger T T^\dagger \delta M = 0$, where $T = L_{uV\mathcal{S}}^\dagger \circ L_{uV\mathcal{P}}$ and $T^\dagger T T^\dagger = T^\dagger$ follows from the definition of the Moore Penrose pseudoinverse. □

B. Justification of Assumption 4

The following Lemma proves that given Assumption 1, if rank$[M-\delta M-R]=n-1$ then there exists a neighborhood of $M-\delta M-R$ for which rank reducing perturbations exist with a Frobenius norm bounded by the growth of the n^{th} singular value in this neighborhood, justifying Assumption 4. In fact, [6] shows how the assumption can be removed.

Lemma 8. *Let $\delta M_0 \in \mathcal{S}$ and $R_0 \in \mathcal{P}$ be such that $\widetilde{M} \triangleq M - \delta M_0 - R_0 \in \mathbb{C}^{n\times m}$ satisfies* rank$[\widetilde{M}] = n-1$ *and let L_{uV} be the operator associated with the SVD of $\widetilde{M}$. If $L_{uV\mathcal{S}}$ is surjective then there exists constants c and K such that for every $N \in \mathbb{C}^{n\times m}$ with $\|N\|_F < c$ there is a $\delta M \in \mathcal{S}$ satisfying $\|\delta M\|_F \le K\sigma_n(\widetilde{M}-N)$ and $L_{\tilde{u}\widetilde{V}}(\widetilde{M}-N-\delta M)=0$, where $L_{\tilde{u}\widetilde{V}}$ is the operator associated with the SVD of $\widetilde{M}-N$.*

Proof. Let $\{S_1, S_2, \ldots, S_k\}$ be an orthonormal basis of $\mathcal{S}$, let $B_{\mathcal{S}} = [\mathrm{vec}(S_1), \mathrm{vec}(S_2), \ldots, \mathrm{vec}(S_k)]$ and define

$$Z(u,V) = \begin{bmatrix} \mathrm{Re}\left[(V^\top \otimes u^H)B_{\mathcal{S}}\right] \\ \mathrm{Im}\left[(V^\top \otimes u^H)B_{\mathcal{S}}\right] \end{bmatrix}$$

Since $L_{uV\mathcal{S}}$ is surjective,

$$\mathrm{rank}\, Z(u,V) = 2(m-n+1). \tag{73}$$

Given a matrix $\widetilde{M}-N$, let $\tilde{u}$ be the n^{th} lsv of $\widetilde{M}-N$ and $\widetilde{V} \in \mathbb{C}^{m\times(m-n+1)}$ have columns equal to the last $m-n+1$ rsv of $\widetilde{M}-N$. Given that the last singular value of $\widetilde{M}$ is simple, there exists a ball with a radius d around $\widetilde{M}$ wherein all matrix valued functions $N(\alpha)$ which depend analytically on the real scalar α and satisfy $\|N(\cdot)\|_F < d$ have $\tilde{u}(\alpha)$ and $\widetilde{V}(\alpha)$ which can be chosen to be analytic functions of α (see [22], [26] for more details). As a result, there exists $c > 0$ small enough for which there exists $\epsilon = \epsilon(c) > 0$ such that for each N with $\|N\|_F < c$, $\tilde{u}$ and $\widetilde{V}$ satisfy

$$\sigma_{2(m-n+1)}(Z(\tilde{u},\widetilde{V})) \ge \epsilon > 0. \tag{74}$$

Consider one specific N satisfying $\|N\|_F < c$ and take the corresponding $\tilde{u}$ and $\widetilde{V}$ for $\widetilde{M}-N$. Set $x_0 \in \mathbb{C}^m$ to be

$$x_0^T = L_{\tilde{u}\widetilde{V}}(\widetilde{M}-N).$$

We will now construct a perturbation $\delta M \in \mathcal{S}$ such that $L_{\tilde{u}\widetilde{V}\mathcal{S}}(\delta M) = x_0^T$. Note that $\mathrm{vec}(\delta M) = B_{\mathcal{S}}\zeta_0$ for some $\zeta_0 \in \mathbb{R}^k$, so

$$\begin{bmatrix} \mathrm{Re}(L_{\tilde{u}\widetilde{V}\mathcal{S}}(\delta M))^T \\ \mathrm{Im}(L_{\tilde{u}\widetilde{V}\mathcal{S}}(\delta M))^T \end{bmatrix} = Z(\tilde{u},\widetilde{V})\zeta_0.$$

If $x_0 = 0$, then $\delta M = 0$ satisfies the conditions of the lemma. If $x_0 \neq 0$, then $L_{\tilde{u}\widetilde{V}\mathcal{S}}(\delta M) = x_0^T$ if and only if

$$\begin{bmatrix} \mathrm{Re}\, x_0 \\ \mathrm{Im}\, x_0 \end{bmatrix} = Z(\tilde{u},\widetilde{V})\zeta_0. \tag{75}$$

Since $Z(\tilde{u},\widetilde{V})$ has full row rank by (74), it has a right inverse $Z(\tilde{u},\widetilde{V})^\dagger$. Therefore, we can compute ζ_0 from (75):

$$\zeta_0 = Z(\tilde{u},\widetilde{V})^\dagger \begin{bmatrix} \mathrm{Re}\, x_0 \\ \mathrm{Im}\, x_0 \end{bmatrix}.$$

Thus $L_{\tilde{u}\widetilde{V}}(\widetilde{M}-N-\delta M) = 0$.

What remains is to find the bound on $\|\delta M\|_F$. Since $B_{\mathcal{S}}$ has orthonormal columns, $\|\delta M\|_F = \|\zeta_0\|_2$. Hence

$$\begin{aligned} \|\delta M\|_F &\le \sigma_1(Z(\tilde{u},\widetilde{V})^\dagger) \left\| \begin{bmatrix} \mathrm{Re}\, x_0 \\ \mathrm{Im}\, x_0 \end{bmatrix} \right\|_2 \\ &\le \frac{1}{\epsilon}\|L_{\tilde{u}\widetilde{V}}(\widetilde{M}-N)\|_2. \end{aligned}$$

But according to Proposition 1, $\|L_{\tilde{u}\widetilde{V}}(\widetilde{M}-N)\|_2 = \sigma_n(\widetilde{M}-N)$. Letting $K = 1/\epsilon$, we obtain the desired bound $\|\delta M\| \le K\sigma_n(\widetilde{M}-N)$. Observe that the last inequality implies that $\frac{\sigma_n(\widetilde{M}-N)}{2\|\delta M\|} \ge \frac{\epsilon}{2} > 0$. This is precisely the expression that appears in g_k as defined in step 8 of Algorithm 1.

□

C. Additional Example

The following example is a simple illustration of some of the issues that occur when rank$(M-\delta M_* - R_*) < n-1$. It also exhibits repeated singular values for all $R \in \mathcal{P}$ and $\delta M \in \mathcal{S}$:

Example 4. *Consider the problem:*

$$M = \begin{bmatrix} 1+\epsilon & 1 \\ -1 & 1 \end{bmatrix},\ \mathcal{R} = \{\gamma I : \gamma \in \mathbb{R}\},$$
$$\mathcal{S} = \left\{ \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} : \alpha, \beta \in \mathbb{R} \right\},$$

where $0 \le \epsilon \ll 1$. We seek to minimize $\alpha^2 + \beta^2$ subject to

$$\det(M - \delta M - R) = \det\left(\begin{bmatrix} 1+\epsilon-\gamma-\alpha & 1-\beta \\ -1+\beta & 1-\gamma-\alpha \end{bmatrix} \right) = 0.$$

The singular values of $M-\delta M-R$ are repeated for all values of α, β, and γ. The global minimum occurs at $\gamma_*(\epsilon) = 1+\epsilon/2$, $\alpha_* = 0$, and $\beta_*(\epsilon) = 1-\epsilon/2$, for which

$$M - \delta M_* - R_* = \begin{bmatrix} \epsilon/2 & \epsilon/2 \\ -\epsilon/2 & -\epsilon/2 \end{bmatrix}$$

and the norm of the minimum $\delta M_*(\epsilon)$ is $\sqrt{2}(1-\epsilon/2)$.

Consider $\epsilon = 0$, implying $M - \delta M_* - R_* = 0$ with the dimension of the singular space equal to two. This example shows that the necessary condition in Theorem 1 (condition 1a) is *not* satisfied for an arbitrary left/right singular vector pair when L_{uV} is not surjective. Since $M - \delta M_* - R_* = 0$, any vector in $\mathbb{R}^2$ is both a right and a left singular vector, so the choice of singular vectors is arbitrary at $M - \delta M_* - R_*$. However, the choice is not arbitrary in a neighborhood of $M - \delta M_* - R_*$, even though the singular vectors are repeated. The left and right singular vectors must be chosen as a pair from one of many possible SVD factorizations of $M-\delta M-R$, i.e., not independently. Consider an arbitrary left singular vector $u^T = [u_1, u_2]$ and a right singular vector $v^T = [v_1, v_2]$. Then $L_{uv}(M-R) = u_2v_1 - u_1v_2$ and $L_{uv}(\delta M) = \beta(u_2v_1 - u_1v_2)$ so Theorem 1 is satisifed except in the case that $u = \pm v$, in which case L_{uv} is not surjective. Other choices exist without this problem, even in this restrictive example where $\mathcal{S}$ has dimension 1.

Now consider the case where $0 < \epsilon \ll 1$, and examine what happens to the singular vectors and the minimizing solution as $\epsilon \to 0$. The singular vectors at the solution $M - \delta M_* - R_*$ span a one-dimensional space (independent of ϵ) and cannot be chosen arbitrarily. One choice is $u^T = [1, 1]$ and $v^T = [1, -1]$. With these vectors, $L_{uv\mathcal{P}}(\Delta R) = 0$ for all $\Delta R \in \mathcal{P}$; $L_{uv\mathcal{S}}(M - R_*) = -2 + \epsilon$, and $L_{uv\mathcal{S}}(\delta M) = -2\beta$. The minimizer $\beta_*(\epsilon) = 1 - \epsilon/2$ satisfies the condition of Theorem 1 (1a), and $\Delta R = 0$ is the minimum norm minimizer of (2a) since $L_{uv\mathcal{P}}(\Delta R) = 0$. Note that $\beta_*(\epsilon)$ is continuous at $\epsilon = 0$ so a sequence of solutions parameterized by $\epsilon_k \to 0$ approaches the solution at $\epsilon = 0$.

One point of this example is there *always* exists a problem with a solution where $\text{rank}(M - \delta M_* - R_*) = n - 1$ arbitrarily close to every problem where the solution satisfies $\text{rank}(M - \delta M_* - R_*) = n - 2$. Any problem can be explicitly perturbed to one which satisfies Assumption 3. In the presence of finite arithmetic, M will be implicitly perturbed from its original value, and not in any particular direction.

References

[1] C.-T. Chen, *Linear System Theory and Design*, 3rd ed. New York, NY, USA: Oxford University Press, Inc., 1998.

[2] R. Eising, "Between controllable and uncontrollable," *Syst. Control Lett.*, vol. 4, no. 5, pp. 263–264, 1984.

[3] M. Wicks and R. DeCarlo, "Computing the distance to an uncontrollable system," *IEEE Trans. Automat. Contr.*, vol. 36, no. 1, pp. 39–49, 1991.

[4] C. Kenney and A. J. Laub, "Controllability and stability radii for companion form systems," *Math. Control. Signals, Syst.*, vol. 1, no. 3, pp. 239–256, 1988.

[5] M. Wicks, "Matrix rank-robustness problems in systems and control: theory and computation," Ph.D. dissertation, Purdue University, 1992.

[6] M. A. Wicks and R. A. DeCarlo, "Rank robustness of complex matrices with respect to real perturbations," *SIAM Journal on Matrix Analysis and Applications*, vol. 15, no. 4, pp. 1182–1207, 1994.

[7] M. Karow and D. Kressner, "On the structured distance to uncontrollability," *Systems & Control Letters*, vol. 58, no. 2, pp. 128–132, 2009.

[8] S. R. Khare, H. K. Pillai, and M. N. Belur, "Computing the radius of controllability for state space systems," *Syst. Control Lett.*, vol. 61, no. 2, pp. 327–333, 2012.

[9] S. C. Johnson and R. A. DeCarlo, "Bounding the distance to the nearest unobservable switched linear time-invariant system," in *2015 Am. Control Conf.* Chicago, IL: IEEE, 2015.

[10] C. He, "Estimating the distance to uncontrollability: A fast method and a slow one," *Syst. Control Lett.*, vol. 26, no. 4, pp. 275–281, 1995.

[11] G. Hu and E. Davison, "Real controllability/stabilizability radius of lti systems," *IEEE Trans. Automat. Contr.*, vol. 49, no. 2, pp. 254–257, 2004.

[12] J. V. Burke, A. S. Lewis, and M. L. Overton, "Pseudospectral components and the distance to uncontrollability," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 2, pp. 350–361, 2004.

[13] M. Gu, E. Mengi, M. L. Overton, J. Xia, and J. Zhu, "Fast methods for estimating the distance to uncontrollability," *SIAM J. Matrix Anal. Appl.*, vol. 28, no. 2, pp. 477–502, 2006.

[14] N. K. Son and D. D. Thuan, "The structured controllability radii of higher order systems," *Linear Algebra Appl.*, vol. 438, no. 6, pp. 2701–2716, 2013.

[15] J. Clotet and M. D. Magret, "Upper bounds for the distance between a controllable switched linear system and the set of uncontrollable ones," *Math. Probl. Eng.*, vol. 2013, no. 4, pp. 1–9, 2013.

[16] S. Lam and E. J. Davison, "Computation of the real controllability radius and minimum-norm perturbations of higher-order, descriptor, and time-delay lti systems," *IEEE Trans. Automat. Contr.*, vol. 59, no. 8, pp. 2189–2195, 2014.

[17] I. Markovsky, *Low Rank Approximation: Algorithms, Implementation, Applications*. Springer, 2011.

[18] N. Guglielmi and I. Markovsky, "An ODE-Based Method for Computing the Distance of Coprime Polynomials to Common Divisibility," *SIAM Journal on Numerical Analysis*, vol. 55, no. 3, pp. 1456–1482, Jan. 2017.

[19] M. Wicks and R. DeCarlo, "Computing robustness of system properties with respect to structured real matrix perturbations," in *[1992] Proc. 31st IEEE Conf. Decis. Control.* IEEE, 1992, pp. 1909–1914.

[20] S. C. Johnson, R. A. DeCarlo, and M. Zefran, "Set-transition observability of switched linear systems," in *2014 Am. Control Conf.* IEEE, 2014, pp. 3267–3272.

[21] S. C. Johnson, "Observability and observer design for switched linear systems," Ph.D. dissertation, Purdue University, 2016.

[22] B. De Moor and S. Boyd, "Analytic Properties of Singular Values and Vectors," Tech. Report 1989-28, SISTA, E. E. Dept.-ESAT, K. U. Leuven,, Tech. Rep., 1989.

[23] D. G. Luenberger, *Optimization by Vector Space Methods*. John Wiley & Sons, Inc., 1969.

[24] J. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, no. 9, pp. 772–781, 1978.

[25] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear System Theory*. Princeton University Press, 2005.

[26] T. Kato, *A short introduction to perturbation theory for linear operators*. Springer New York, 2012.

Scott C. Johnson (S'12), native of Fort Wayne, IN, USA received the B.S. degree in mathematics and physics from The College of Idaho, Caldwell, ID, USA, in 2011, the M.S. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 2013, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 2016.

He is currently a Senior Controls Engineer at John Deere Electronic Solutions in Fargo, ND, USA. His research interests include observer design, fault detection, electric motor control, linear algebra, and switched systems.

Mark Wicks (SM'10) is currently the Lead Data Scientist at Xometry, where he uses machine learning techniques to set prices for parts that are manufactured on demand. Dr. Wicks completing the first half of his career in academia on the faculty at Kettering University, where he had been a department chair and the associate vice president for academic affairs. He is a past associate editor for IEEE Control Systems and a past member of the NCEES FE exam committee. His research interests include machine learning, predictive modeling, numerical linear algebra, and switched-system theory.

Miloš Žefran completed his undergraduate studies in Electrical Engineering and Mathematics at the University of Ljubljana, Slovenia, where he also received a M.S. in Electrical Engineering. He received a M.S. in Mechanical Engineering and a Ph.D. in Computer Science from the University of Pennsylvania in 1995 and 1996, respectively.

From 1997 to 1999 he was a NSF Postdoctoral Scholar at the California Institute of Technology and afterwards he joined the Rensselaer Polytechnic Institute, Troy, NY, USA. Since 1999 he has been with the Department of Electrical and Computer Engineering at the University of Illinois at Chicago where he is currently a Professor. His research interests include human-robot interaction, runtime monitoring of cyber-physical systems, model-predictive control for switched and hybrid systems, and distributed control of robot networks. Dr. Žefran received the NSF Career Award in 2001 and is an Associate Editor of the IEEE Transactions on Control Systems Technology.

Raymond DeCarlo (F'89) received a B.S. (1972) and M.S. (1974) in Electrical Engineering from the University of Notre Dame and his Ph.D. (1976) under the direction of Dr. Richard Saeks from Texas Tech University.He joined Purdue as an Assistant Professor of Electrical Engineering in 1977, becoming Associate and full Professors in 1982 and 2005. He worked at the General Motors Research Labs during the summers of 1985 and 1986. He is a Fellow of the IEEE (1989), past Associate Editor for Technical Notes and Correspondence and past Associate Editor for Survey and Tutorial Papers, both for the *IEEE Transactions on Automatic Control*. He was secretary-administrator of the IEEE Control Systems Society, a member of the Board of Governors from 1986-1992 and 1999-2003, Program Chair for the 1990 IEEE CDC, General Chair of the 1993 IEEE CDC, andCSS VP for Financial Activities 2001-2002. His awards include the CSS distinguished member award (1990), the IEEE Third Millennium Medal (2000), the Motorola Excellence in Teaching Award (2006, 2011, 2016), and best Theoretical Paper in *Automatica* in 2008. He has coauthored three books, has numerous journal and conference articles, and several book chapters/reprints. His research interests are diverse. He is a former Chair of the Purdue University Senate and a CIC Faculty Fellow.