SIAM J. MATRIX ANAL. APPL.
Vol. 40, No. 3, pp. 918–942

ON THE SENSITIVITY OF SINGULAR AND ILL-CONDITIONED LINEAR SYSTEMS*

ZHONGGANG ZENG†

Abstract. Solving a singular linear system for an individual vector solution is an ill-posed problem with a condition number infinity. From an alternative perspective, however, the general solution of a singular system is of a bounded sensitivity as a unique element in an affine Grassmannian. If a singular linear system is given through empirical data that are sufficiently accurate with a tight error bound, a properly formulated general numerical solution uniquely exists in the same affine Grassmannian, enjoys Lipschitz continuity, and approximates the underlying exact solution with an accuracy in the same order as the data. Furthermore, any backward accurate numerical solution vector is an accurate approximation to one of the solutions of the underlying singular system.

Key words. condition number, linear system, Grassmannian

AMS subject classifications. 65F22, 65F35, 15A12, 15A06

DOI. 10.1137/18M1197990

1. Introduction. Solving linear systems in the matrix-vector form $A\mathbf{x} = \mathbf{b}$ is one of the most fundamental problems in scientific computing. In the literature of numerical analysis, linear systems are always assumed to be nonsingular with few exceptions. Numerical solutions of singular systems are almost never mentioned directly in textbooks. The following rare remark in Meyer's textbook [25, pp. 217–218] accurately reflects the current state of knowledge:

> If $\mathbf{A}$ is singular, ... even a stable algorithm can result in a significant loss of information. ... the small perturbation $\mathbf{E}$ due to roundoff makes the possibility that $rank(\mathbf{A}+\mathbf{E}) > rank(\mathbf{A})$ very likely. *The moral is to avoid floating-point solutions of singular systems*" [emphasis added].

In applications such as deblurring images and discrete inverse problems, rank-deficient and highly ill-conditioned linear systems are approached using the Tikhonov regularization [11, 12, 13, 27]. As Neumaier states [27, p. 637]:

> Though frequently needed in applications, the adequate handling of such ill-posed linear problems is hardly ever touched upon in numerical analysis text books.

Singular linear systems are unavoidable in scientific computing and often need to be solved without knowledge of the exact matrices and vectors, as shown in our case studies in section 3. The obvious difficulty in solving a singular linear system from empirical data is the condition number infinity, and thus the error is unbounded when solving for an individual vector solution. While this error analysis is in itself

*Received by the editors July 2, 2018; accepted for publication (in revised form) by J. L. Barlow April 24, 2019; published electronically August 1, 2019. This work was performed by an employee of the U.S. Government or under U.S. Government contract. The U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. Copyright is owned by SIAM to the extent not limited by these rights.
https://doi.org/10.1137/18M1197990
Funding: This work was partially supported by National Science Foundation grant DMS-1620337.
†Department of Mathematics, Northeastern Illinois University, Chicago, IL 60625 (zzeng@neiu.edu).

impeccable, the solution of a singular system is more than an individual vector. The very notion of the numerical solution to a given system $\tilde{A}\mathbf{x} = \tilde{\mathbf{b}}$ needs clarification when entries of $(\tilde{A}, \tilde{\mathbf{b}})$ serve as empirical data for an underlying singular linear system $A\mathbf{x} = \mathbf{b}$.

This paper attempts to analyze the accuracy and sensitivity of solving singular linear systems from the following different perspective: The solution of a singular linear system is either an empty set or an affine subspace as a unique element in an affine Grassmannian rather than a vector. Using this point of view, the condition number becomes bounded. A properly formulated general numerical solution in a certain affine Grassmannian is of a sensitivity proportional to $\|A\|_2 \|A^\dagger\|_2$, never infinity, with respect to either constrained or arbitrary perturbations, where $A^\dagger$ is the Moore–Penrose inverse of A. Such a numerical solution of a perturbed system $\tilde{A}\mathbf{x} = \tilde{\mathbf{b}}$ within a viable error tolerance accurately solves the underlying singular system $A\mathbf{x} = \mathbf{b}$, and the ratio of solution accuracy to the data error is bounded by a factor of $\|A\|_2 \|A^\dagger\|_2$, not $\|\tilde{A}\|_2 \|\tilde{A}^{-1}\|_2$, assuming the data error is small with an attainable tight bound.

We shall further demonstrate that the sensitivity of a singular linear system $A\mathbf{x} = \mathbf{b}$ is measured by $\|A\|_2 \|A^\dagger\|_2$ rather than infinity from multiple perspectives, including homogeneous cases, under constrained perturbations preserving the singularity and consistency, solving for the general numerical solutions in an affine Grassmannian, and solving for a single particular solution. Furthermore, every backward accurate numerical (vector) solution of a singular consistent linear system accurately approximates a particular exact solution regardless of the algorithm used. The "error" largely falls harmlessly in the kernel of A. This result extends what Peters and Wilkinson discovered in [29] beyond inverse power iterations. While any numerical (single-vector) solution may be inaccurate to a linear system that is genuinely nonsingular and highly ill-conditioned, we shall prove that a stable numerical (affine subspace) solution may exist and contain an accurate approximation to the exact solution. For practical computation, efficient and robust algorithms already exist for general numerical solutions in affine Grassmannians. Regularization algorithms, such as the Tikhonov method and truncated singular value decomposition (TSVD) [9, sect. 5.5.4], [10], produce the accurate vector component, and numerical rank-revealing algorithms [2, 7], [9, sect. 5.4.6], [18, 19, 20, 31] provide the numerical kernel as the remaining component.

For continuity of presentation, lemmas and long proofs are listed in the appendices. Additional computational results and software demonstration are given in the supplementary material.

2. Preliminaries. Column vectors are denoted by boldface lowercase letters such as $\mathbf{b}$, $\mathbf{x}$, $\mathbf{y}$, etc., with $\mathbf{0}$ being a zero vector whose dimension can be derived from the context. The vector space of n-dimensional complex column vectors is denoted by $\mathbb{C}^n$. The vector space of $m \times n$ matrices with complex entries is denoted by $\mathbb{C}^{m\times n}$. Matrices are denoted by uppercase letters such as A, B, X, etc., while O and I denote a zero matrix and an identity matrix. respectively. The range, kernel, rank, and Hermitian transpose of a matrix A are denoted by $\mathcal{R}ange(A)$, $\mathcal{K}ernel(A)$, $rank(A)$, and A^{H}, respectively. In this paper, we consider general $m \times n$ linear systems in the form of $A\mathbf{x} = \mathbf{b}$, and we say the system is *singular* when $rank(A) < n$ so that $\mathcal{K}ernel(A) \neq \{\mathbf{0}\}$, including nonsquare cases where $m < n$ or $m > n$. The system is *consistent* if $\mathbf{b} \in \mathcal{R}ange(A)$.

For any matrix $A \in \mathbb{C}^{m\times n}$, the jth largest singular value of a matrix A is denoted by $\sigma_j(A)$. The *numerical rank* of a matrix A *within an error tolerance* $\theta > 0$ is defined as

$$\mathit{rank}_\theta\,(\,A\,) \quad := \quad \min_{\|B-A\|_2<\theta} \mathit{rank}\,(\,B\,) \quad \equiv \quad \max_{\sigma_j(A)>\theta} j,$$

assuming θ is not equal to any singular value of A. Let $U\,\Sigma\,V^{\mathsf{H}}$ be the SVD of A, where $U = [\mathbf{u}_1, \ldots, \mathbf{u}_m]$ and $V = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$. If $\mathit{rank}_\theta\,(\,A\,) = r$ within θ, then the θ*-projection* A_θ of A is defined as

$$A_\theta \quad := \quad \sigma_1(A)\,\mathbf{u}_1\,\mathbf{v}_1^{\mathsf{H}} + \cdots + \sigma_r(A)\,\mathbf{u}_r\,\mathbf{v}_r^{\mathsf{H}} \quad = \quad \sum_{\sigma_j(A)>\theta} \sigma_j(A)\,\mathbf{u}_j\,\mathbf{v}_j^{\mathsf{H}}.$$

In this case, the *numerical kernel* of A within θ is $\mathcal{K}\mathit{ernel}\,(A_\theta) = \mathit{span}\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$, where $\mathit{span}\{\ldots\}$ denotes the vector space spanned by vectors in the list. The entities $\mathit{rank}_\theta\,(\,A\,)$, A_θ, and $\mathcal{K}\mathit{ernel}\,(A_\theta)$ are undefined if θ is a singular value of A. The *Moore–Penrose inverse* of A, denoted by $A^\dagger$, is the unique matrix satisfying the Moore–Penrose conditions $A\,A^\dagger\,A = A$, $A^\dagger\,A\,A^\dagger = A^\dagger$, $(A\,A^\dagger)^{\mathsf{H}} = A\,A^\dagger$, and $(A^\dagger\,A)^{\mathsf{H}} = A^\dagger\,A$. Using the singular value decomposition as above and assuming $\mathit{rank}\,(\,A\,) = r$, we see that the identity [9, sect. 5.5.2]

$$A^\dagger \quad \equiv \quad \frac{1}{\sigma_1(A)}\,\mathbf{v}_1\,\mathbf{u}_1^{\mathsf{H}} + \cdots + \frac{1}{\sigma_r(A)}\,\mathbf{v}_r\,\mathbf{u}_r^{\mathsf{H}} \quad = \quad \sum_{\sigma_j(A)>0} \frac{1}{\sigma_j(A)}\,\mathbf{v}_j\,\mathbf{u}_j^{\mathsf{H}}$$

holds, and $X = A^\dagger$ is the minimum Frobenius norm matrix such that $A\,X$ and $X\,A$ are orthogonal projections from $\mathbb{C}^m$ onto $\mathcal{R}\mathit{ange}(A)$ and from $\mathbb{C}^n$ onto $\mathcal{R}\mathit{ange}(A^{\mathsf{H}})$, respectively. We shall frequently use $\left\|A^\dagger\right\|_2^{-1}$ as an alternative notation for the smallest positive singular value $\sigma_r(A)$ of A with rank r.

The set of k-dimensional subspaces of $\mathbb{C}^n$ is called the *Grassmannian* [6], [17, p. 52] of index k of $\mathbb{C}^n$ denoted by $\mathcal{G}_k(\mathbb{C}^n)$. For any $\mathcal{P}, \mathcal{Q} \in \mathcal{G}_k(\mathbb{C}^n)$, let $P, Q, \in \mathbb{C}^{n\times k}$ be matrices whose columns form orthonormal bases for $\mathcal{P}$, $\mathcal{Q}$, respectively, while $\hat{P}, \hat{Q} \in \mathbb{C}^{n\times(n-k)}$ such that $[P,\hat{P}]^{\mathsf{H}}\,[P,\hat{P}] = [Q,\hat{Q}]^{\mathsf{H}}\,[Q,\hat{Q}] = I$. The Grassmannian $\mathcal{G}_k(\mathbb{C}^n)$ is a metric space with the distance [9, sect. 2.5.3]

$$\mathit{dist}\,(\,\mathcal{P},\mathcal{Q}\,) \quad := \quad \left\|P\,P^{\mathsf{H}} - Q\,Q^{\mathsf{H}}\right\|_2 \quad \equiv \quad \left\|P^{\mathsf{H}}\,\hat{Q}\right\|_2 \quad \equiv \quad \left\|Q^{\mathsf{H}}\,\hat{P}\right\|_2.$$

The set of k-dimensional affine subspaces of $\mathbb{C}^n$ is called the *affine Grassmannian* [16, sect. 7.1], [21, 22] of index k of $\mathbb{C}^n$ and is denoted by

$$\mathcal{A}_k(\mathbb{C}^n) \quad := \quad \left\{\mathbf{u} + \mathcal{V} \subset \mathbb{C}^n \;\middle|\; \mathbf{u} \in \mathbb{C}^n,\; \mathcal{V} \in \mathcal{G}_k(\mathbb{C}^n)\right\}.$$

Here, for any vector $\mathbf{u} \in \mathbb{C}^n$ and subspace $\mathcal{V} \in \mathcal{G}_k(\mathbb{C}^n)$, the *affine subspace*

$$\mathbf{u} + \mathcal{V} \quad := \quad \{\mathbf{u} + \mathbf{v} \in \mathbb{C}^n \mid \mathbf{v} \in \mathcal{V}\}$$

can be written as $\hat{\mathbf{u}} + \mathcal{V}$ with a unique $\hat{\mathbf{u}} \in \mathcal{V}^\perp \cap (\mathbf{u} + \mathcal{V})$ of the minimum norm, where $(\cdot)^\perp$ denotes the unitary complement of any subspace $(\cdot)$. The metric

$$\begin{aligned} &\mathit{dist}\,(\,\mathbf{u}_1 + \mathcal{V}_1,\, \mathbf{u}_2 + \mathcal{V}_2\,) \\ &\qquad := \max_{\hat{\mathbf{u}}_j \in \mathcal{V}_j^\perp \cap (\mathbf{u}_j + \mathcal{V}_j),\, j=1,2} \left\{\|\hat{\mathbf{u}}_1 - \hat{\mathbf{u}}_2\|_2,\; \mathit{dist}\,(\,\mathcal{V}_1,\,\mathcal{V}_2\,)\right\} \end{aligned} \tag{1}$$

for every $\mathbf{u}_1 + \mathcal{V}_1, \mathbf{u}_2 + \mathcal{V}_2 \in \mathcal{A}_k(\mathbb{C}^n)$ is a distance in $\mathcal{A}_k(\mathbb{C}^n)$. For every $(A,\mathbf{b}) \in \mathbb{C}^{m\times n} \times \mathbb{C}^m$, denote the set of vector solutions to the system $A\,\mathbf{x} = \mathbf{b}$ by

$$\mathit{sol}\,(A,\mathbf{b}) \quad := \quad \{\mathbf{u} \in \mathbb{C}^n \mid A\,\mathbf{u} = \mathbf{b}\}.$$

For $r = \mathit{rank}(A)$, the set $\mathit{sol}(A, \mathbf{b})$ as *the* solution of $A\mathbf{x} = \mathbf{b}$ uniquely exists as either $\emptyset$ or an element in the affine Grassmannian $\mathcal{A}_{n-r}(\mathbb{C}^n)$. The *dimension* of $\mathit{sol}(A, \mathbf{b})$ is either $n-r$ if it is in $\mathcal{A}_{n-r}(\mathbb{C}^n)$ or -1 if it is empty [3, p. 6]. We define $\mathit{dist}(\emptyset, \emptyset) = 0$ so that the deviation of solutions can be measured if and only if they are of the same dimension.

The *condition number* of a square matrix A in the context of solving a linear system $A\mathbf{x} = \mathbf{b}$ is well known to be $\kappa(A) = \|A\|_2 \|A^{-1}\|_2$, with a convention $\kappa(A) = \infty$ when A is singular [9, p. 87]. This condition number is based on the attainable error of the solution as an individual vector. The infinity convention can be justified by $\limsup_{G\to A} \|G\|_2 \|G^\dagger\|_2 = \infty$ when A is singular, and by the interpretation of the condition number as the reciprocal of the distance to the singularity [14, Theorem 6.5]. For a rectangular matrix A, it is natural to generalize the condition number as $\kappa(A) = \|A\|_2 \|A^\dagger\|_2$ (see, e.g., [14, p. 382]). We present arguments from multiple perspectives that the infinity convention may be unnecessary even if A is square and singular.

It is easy to see that $\kappa(A) = \|A\|_2 \|A^\dagger\|_2$ is discontinuous at any rank deficient matrix A and cannot be approximated from empirical data $\tilde{A}$ since $\kappa(\tilde{A}) = \|\tilde{A}\|_2 \|\tilde{A}^\dagger\|_2$ can be arbitrarily large when $\|\Delta A\|_2 = \|\tilde{A} - A\|_2$ is small. For any error tolerance θ with $0 < \theta < \|A^\dagger\|_2^{-1}$, however, the asymptotic bound

$$\kappa(A) - 2\|\Delta A\|_2 + O(\|\Delta A\|_2^2) \quad \leq \quad \kappa(\tilde{A}_\theta) \quad \leq \quad \kappa(A) + 2\|\Delta A\|_2 + O(\|\Delta A\|_2^2)$$

follows from [9, Corollary 8.6.2] when the data matrix $\tilde{A}$ is sufficiently accurate so that $\|\Delta A\|_2 < \|A^\dagger\|_2^{-1} - \theta$. Assuming an error bound $\beta > \|\Delta A\|_2$ is attainable and is sufficiently tight so that $\beta < \|A^\dagger\|_2^{-1} - \|\Delta A\|_2$, the condition number $\kappa(\tilde{A}_\theta)$ of the θ-projection $\tilde{A}_\theta$ of the data matrix $\tilde{A}$, not $\kappa(\tilde{A})$, is an approximation to the underlying condition number $\kappa(A) = \|A\|_2 \|A^\dagger\|_2 < \infty$.

3. Models of singular linear systems. We shall elaborate some case studies to show that solving singular linear systems is not only unavoidable in scientific computing but also crucial in many applications. It may even be beneficial for the systems to be singular. Moreover, singular linear systems are often not known with exact matrices and right-hand side vectors in practical computation, and need to be solved from empirical data.

Example 1 (multiplicity of a singular solution to a nonlinear system). For a system of nonlinear equations in the form of $\mathbf{f}(\mathbf{x}) = \mathbf{0}$, where $\mathbf{f} = (f_1, \ldots, f_m)$ and $f_j : \mathbb{C}^n \to \mathbb{C}$ is an analytic function for $j = 1, \ldots, m$, a zero $\mathbf{x}_*$ of $\mathbf{f}$ is multiple if the Jacobian of $\mathbf{f}$ at $\mathbf{x}_*$ is rank-deficient. At such a multiple $\mathbf{x}_*$, there is a vector space called the dual space $\mathcal{D}_{\mathbf{f},\mathbf{x}_*}$ that forms the multiplicity structure of the zero and the dimension of $\mathcal{D}_{\mathbf{f},\mathbf{x}_*}$ is the multiplicity. The multiplicity structure can be determined by solving a sequence of homogeneous linear systems

$$S_\alpha(\mathbf{x}_*)\,\mathbf{c} \;=\; \mathbf{0} \quad \text{for} \quad \alpha = 1, 2, \ldots, \tag{2}$$

where $S_\alpha(\mathbf{x}_*)$ is the Macaulay matrix whose entries are derivatives of f_j's of orders up to α evaluated at $\mathbf{x}_*$. The solution $\mathit{sol}\big(S_\alpha(\mathbf{x}_*), \mathbf{0}\big)$ of (2) in a proper Grassmannian is isomorphic to the desired dual space $\mathcal{D}_{\mathbf{f},\mathbf{x}_*}$ when α reaches the so-called depth. See, e.g., [4] for detailed elaborations and the supplementary material for a computing demo. The exact Macaulay matrix $S_\alpha(\mathbf{x}_*)$ is almost never available since $\mathbf{x}_*$ is generally known approximately through a certain $\tilde{\mathbf{x}} \approx \mathbf{x}_*$ within an error bound. The model is to solve the singular system (2) for the solution in a Grassmannian rather than individual vectors from empirical data matrix $S_\alpha(\tilde{\mathbf{x}}) \approx S_\alpha(\mathbf{x}_*)$.

Example 2 (Sylvester equation). This is an application arising in control and system theory [1] in the form of the Sylvester matrix equation $A(t)\,X + X\,B(t) = C(t)$, where $A(t)$, $B(t)$, and $C(t)$ are matrices depending on a parameter t. The system may inevitably become singular when the parameter t varies continuously and passes through a certain t_* whose value may only be obtained approximately. The following illustrative example is slightly modified from [1] (cf. the supplementary material). Let

$$A(t) = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \quad B(t) = \begin{bmatrix} -\frac{5}{3}+t & 1 \\ -1 & -\frac{1}{3}+2\,t \end{bmatrix}, \quad \text{and} \quad C(t) = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}. \tag{3}$$

When t varies continuously, the system becomes singular but still is consistent when t hits the value $t_* = \frac{2}{3}$ with the general solution

$$X_* = \tfrac{1}{4}\begin{bmatrix} 1 & -1 \\ -3 & -1 \end{bmatrix} + \alpha_1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}, \qquad \alpha_1, \alpha_2 \in \mathbb{C}. \tag{4}$$

Suppose we know that $\tilde{t} \approx 0.6666$ with an error bound 0.0001. Can we find a numerical solution $\tilde{X}$ of the perturbed system at the parameter value $t = \tilde{t}$ approximating X_* in (4) of the underlying system at $t = t_*$ with an accuracy $\|\tilde{X} - X_*\|_2$ roughly 0.0001?

Example 3 (Bézout coefficients). For polynomials $f_1, \ldots, f_n$, with a greatest common divisor g, there exist polynomials $u_1, \ldots, u_n$, known as the Bézout coefficients (see, e.g., [26, sect. 1.3], [36]) such that the Bézout identity

$$u_1\,f_1 + \cdots + u_n\,f_n = g \tag{5}$$

holds. Solving the linear equation (5) for the Bézout coefficients appears in many applications, such as computing the Smith normal form in linear control theory [36], and the systems are often singular for $n \geq 3$. Denote by $\mathbb{P}_k$ the vector space of polynomials with degrees up to k. For instance, let f_1, f_2, f_3 be polynomials of degrees, say 4, 7, 6, with degree of g, say 2. Equation (5) for $(u_1, u_2, u_3) \in \mathbb{P}_3 \times \mathbb{P}_1 \times \mathbb{P}_2$ is consistent and rank-deficient by 2. The rank-deficiency is, in fact, a blessing because it transforms the general solution

$$(u_1, u_2, u_3) = (u_{01}, u_{02}, u_{03}) + t_1\,(u_{11}, u_{12}, u_{13}) + t_2\,(u_{21}, u_{22}, u_{23})$$

into an invertible transformation

$$\begin{bmatrix} u_{01} & u_{02} & u_{03} \\ u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} g \\ 0 \\ 0 \end{bmatrix}. \tag{6}$$

The exact coefficients of polynomial parameters $f_1, \ldots, f_n$ and g may be unknown beyond their empirical data, say,

$$\begin{aligned}
\tilde{f}_1 &= 2.5714 + 3.8571\,x - 3\,x^2 - 6.4286\,x^3 - 2.1429\,x^4, \\
\tilde{f}_2 &= -1.7143 - 1.7143\,x + 0.4286\,x^2 + 0.4286\,x^3 - 3.4286\,x^5 - 5.1429\,x^6 - 1.7143\,x^7, \\
\tilde{f}_3 &= 0.8571 + 1.2857\,x + 2.1429\,x^2 + 2.5714\,x^3 + 3.4286\,x^4 + 3.8571\,x^5 + 1.2857\,x^6, \\
\tilde{g} &= 4.6667 + 7\,x + 2.3333\,x^2,
\end{aligned}$$

with coefficientwise error bound $\varepsilon = 0.5 \times 10^{-4}$. Can we accurately calculate the general solution for $(u_1, u_2, u_3) \in \mathbb{P}_3 \times \mathbb{P}_1 \times \mathbb{P}_2$ of (5) using the imperfect data $\tilde{f}_1$, $\tilde{f}_2$, $\tilde{f}_3$, and $\tilde{g}$ within an error in the same order of the data? A computation/software demo for this example is given in the supplementary material. The matrix-vector representation $\tilde{A}\,\mathbf{x} = \tilde{\mathbf{b}}$ of (5) in the given data is

$$
\begin{bmatrix}
2.5714 & 0 & 0 & 0 & -1.7143 & 0 & 0.8571 & 0 & 0 \\
3.8571 & 2.5714 & 0 & 0 & -1.7143 & -1.7143 & 1.2857 & 0.8571 & 0 \\
-3.0000 & 3.8571 & 2.5714 & 0 & 0.4286 & -1.7143 & 2.1429 & 1.2857 & 0.8571 \\
-6.4286 & -3.0000 & 3.8571 & 2.5714 & 0.4286 & 0.4286 & 2.5714 & 2.1429 & 1.2857 \\
-2.1429 & -6.4286 & -3.0000 & 3.8571 & 0 & 0.4286 & 3.4286 & 2.5714 & 2.1429 \\
0 & -2.1429 & -6.4286 & -3.0000 & -3.4286 & 0 & 3.8571 & 3.4286 & 2.5714 \\
0 & 0 & -2.1429 & -6.4286 & -5.1429 & -3.4286 & 1.2857 & 3.8571 & 3.4286 \\
0 & 0 & 0 & -2.1429 & -1.7143 & -5.1429 & 0 & 1.2857 & 3.8571 \\
0 & 0 & 0 & 0 & 0 & -1.7143 & 0 & 0 & 1.2857
\end{bmatrix}
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{bmatrix}
= [\ 4.6667 \quad 7.0000 \quad 2.3333 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0\ ]^\top, \tag{7}
$$

with respect to monomial bases, and the condition number $\kappa(\tilde{A}) \gtrapprox 2.29 \times 10^6$. The system (7) in the conventional sense is highly ill-conditioned since $\varepsilon\,\kappa(\tilde{A}) > 1$.

Applications are abundant involving singular linear systems. The output regulation problem arises in the application of neural networks [23] for finding the matrix pair (X, U) satisfying the so-called regulator equations, whose solutions are not necessarily unique. An illustrative example is as follows (cf. the supplementary material):

$$
\left\{
\begin{array}{rcl}
X \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} & = & \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -1 & 0 \end{bmatrix} X + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} U + \begin{bmatrix} 2 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix} \\
{[\ 0 \quad 0\ ]} & = & [\ 1 \quad 0 \quad -1\ ]\, X + [\ -1 \quad 0\ ],
\end{array}
\right. \tag{8}
$$

where the unknowns X and U are matrices. The system is rank-deficient by one. Furthermore, the matrix parameters are not known exactly but given by estimation. For a matrix A with a defective eigenvalue λ_* and an associated eigenvector $\mathbf{z}_*$, a generalized eigenvector satisfies the singular system $(A - \lambda_*\, I)\,\mathbf{x} = \mathbf{z}_*$ for $\mathbf{x} \in \mathbb{C}^n$. The value of λ_* and $\mathbf{z}_*$ generally can only be known approximately. The problem is to solve the underlying system by solving $(\tilde{A} - \tilde{\lambda}\, I)\,\mathbf{x} = \tilde{\mathbf{z}}$ from the data $\tilde{A} \approx A$, $\tilde{\lambda} \approx \lambda_*$, and $\tilde{\mathbf{z}} \approx \mathbf{z}_*$. More applications include solving the singular homogeneous linear systems of Ruppert matrices in numerical factorization of polynomials [8, 37], numerical elimination of polynomial variables [38], etc. The generalized Lyapunov equation $E^{\mathsf{H}}\, A\, X + A^{\mathsf{H}}\, X\, E = -G$ with given matrices A, E, and G is singular when E is rank-deficient [33]. A singular linear system that models the atmospheric path delay and the water vapor constant estimation is given in [30]. Linear systems derived from discretizing the Fredholm and Volterra integral equations can be considered empirical data of singular systems in the presence of annihilators [12, sect. 2.4 and p. 83] (cf. an example in the supplementary material).

4. Homogeneous systems with empirical data. A problem is *well-posed* if its solution satisfies existence, uniqueness, and Lipschitz continuity with respect to the data or, otherwise, it is an *ill-posed problem.* For an $m \times n$ singular homogeneous linear system $A\,\mathbf{x} = \mathbf{0}$, the problem

$$
\text{Solve } A\,\mathbf{x} = \mathbf{0} \text{ for a single-vector solution } \mathbf{x} \text{ in } \mathbb{C}^n \tag{9}
$$

is obviously ill-posed, as its solutions are not unique. However, the problem (9) is not precisely the problem to be solved in standard linear algebra, where all the solutions are in question. There is a unique solution to the problem

$$
\text{Solve } A\,\mathbf{x} = \mathbf{0} \text{ for the solution } \mathcal{sol}(A, \mathbf{0}) \text{ in the Grassmannian } \mathcal{G}_{n-r}(\mathbb{C}^n), \tag{10}
$$

where $r = \mathcal{rank}(\,A\,)$. The problem may become somewhat confounding when the exact A is unknown but given through empirical data in $\tilde{A}$, as illustrated in Example 1. What really is at stake is a nontrivial solution $\mathcal{sol}(A, \mathbf{0}) \equiv \mathcal{Kernel}(A)$ in the Grassmannian $\mathcal{G}_{n-r}(\mathbb{C}^n)$, but the data system $\tilde{A}\,\mathbf{x} = \mathbf{0}$ is almost always nonsingular with $\mathit{sol}(\tilde{A}, \mathbf{0}) = \{\mathbf{0}\} \in \mathcal{G}_0(\mathbb{C}^n)$ when $m \geq n$. The condition number $\kappa(\tilde{A}) = O(\|A - \tilde{A}\|_2^{-1})$ can be huge as well if $r < \min\{m, n\}$. The very problem of solving a homogeneous linear system from empirical data needs clarification.

Problem 1 (numerical solution of a homogeneous linear system). *Let $\tilde{A}$ be an $m \times n$ matrix serving as empirical data for an underlying homogeneous system $A\,\mathbf{x} = \mathbf{0}$, where entries of A may or may not be known exactly. Identify the rank r of A using $\tilde{A}$, and find a numerical solution of $\tilde{A}\,\mathbf{x} = \mathbf{0}$ in the Grassmannian $\mathcal{G}_{n-r}(\mathbb{C}^n)$ in the form of an orthonormal basis $\{\mathbf{z}_1, \ldots, \mathbf{z}_{n-r}\}$ so that*

$$\mathit{dist}\left(\,\mathit{span}\{\mathbf{z}_1, \ldots, \mathbf{z}_{n-r}\},\ \mathit{sol}(A, \mathbf{0})\,\right) \;=\; O\left(\tfrac{\|A-\tilde{A}\|_2}{\|A\|_2}\right). \tag{11}$$

From Wedin's perturbation analysis [34], the numerical kernel $\mathit{Kernel}(\tilde{A}_\theta)$ within a proper error tolerance $\theta > 0$ is an approximation to $\mathit{Kernel}(A) = \mathit{sol}(A, \mathbf{0})$ in $\mathcal{G}_{n-r}(\mathbb{C}^n)$ (cf. Lemma 11 in Appendix A). For every $G \in \mathbb{C}^{m\times n}$, we define

$$\mathit{sol}_\theta(G, \mathbf{0}) \;:=\; \mathit{Kernel}(G_\theta) \;\equiv\; \mathit{sol}(G_\theta, \mathbf{0})$$

as the *numerical solution* of the homogeneous system $G\,\mathbf{x} = \mathbf{0}$ in the Grassmannian $\mathcal{G}_{n-r}(\mathbb{C}^n)$ *within an error tolerance* θ, where $r = \mathit{rank}_\theta\,(G)$. Numerical methods for computing $\mathit{Kernel}(G_\theta)$ as $\mathit{sol}_\theta(G, \mathbf{0})$ are well established, including the SVD and other numerical rank-revealing methods (see, e.g., [11, 20]). The following theorem summarizes the properties of the numerical solution as a generalization of the exact solution to the homogeneous system and as a well-posed computing problem that solves the underlying system in Problem 1. The essence and underlying substance of Theorem 1 are based on Wedin [34].

Theorem 1. *Let $A \in \mathbb{C}^{m\times n}$. The following properties hold for the numerical solution of a homogeneous system:*

(i) The exact solution is a special case of the numerical solution

$$\mathit{sol}(A, \mathbf{0}) \;\equiv\; \mathit{sol}_\theta(A, \mathbf{0}) \quad \textit{for all} \quad \theta \in \left(0, \|A^\dagger\|_2^{-1}\right).$$

(ii) Computing the numerical solution is a well-posed problem: *If $\mathit{sol}_\theta(A, \mathbf{0})$ is well-defined within $\theta > 0$, then $\mathit{sol}_\theta(A + \Delta A, \mathbf{0})$ uniquely exists in the same Grassmannian as $\mathit{sol}_\theta(A, \mathbf{0})$ and enjoys Lipschitz continuity with*

$$\begin{aligned}&\mathit{dist}\,(\,\mathit{sol}_\theta(A + \Delta A,\, \mathbf{0}),\, \mathit{sol}_\theta(A, \mathbf{0})\,) \\ &\quad\leq\; \frac{\|A_\theta\|_2\,\|A_\theta^\dagger\|_2}{1 - \|A_\theta^\dagger\|_2\,(\|A - A_\theta\|_2 + \|\Delta A\|_2)}\,\frac{\|\Delta A\|_2}{\|A\|_2}\end{aligned} \tag{12}$$

for all ΔA with sufficiently small $\|\Delta A\|_2$ satisfying

$$\|\Delta A\|_2 \;\leq\; \min\left\{\tfrac{1}{2}\left(\|A_\theta^\dagger\|_2^{-1} - \|A - A_\theta\|_2\right),\, \theta - \|A - A_\theta\|_2,\, \|A_\theta^\dagger\|_2^{-1} - \theta\right\}.$$

(iii) A homogeneous system can be solved from empirical data with an accuracy in the same order as the data: *For any $A + \Delta A$ serving as empirical data of A with $\|\Delta A\|_2 < \frac{1}{2}\,\|A^\dagger\|_2^{-1}$, there exist $\mu, \eta > 0$ with*

$$\mu \;\leq\; \|\Delta A\|_2 \;<\; \|A^\dagger\|_2^{-1} - \|\Delta A\|_2 \;\leq\; \eta \tag{13}$$

such that the numerical solution $\mathit{sol}_\theta(A+\Delta A,\,\mathbf{0})$ *within any error tolerance* $\theta \in (\mu,\,\eta)$ *is in the same Grassmannian as the exact solution* $\mathit{sol}(A,\mathbf{0})$ *and*

$$\text{(14)} \qquad \mathit{dist}\,(\,\mathit{sol}_\theta(A+\Delta A,\,\mathbf{0}),\,\mathit{sol}(A,\mathbf{0})\,) \quad \leq \quad \frac{\|A\|_2\,\|A^\dagger\|_2}{1-\|A^\dagger\|_2\,\|\Delta A\|_2}\,\frac{\|\Delta A\|_2}{\|A\|_2}.$$

Proof. The proof is straightforward from Wedin's error bound [34] on singular subspaces (see Lemma 11 in Appendix A) along with the identity $A_\theta \equiv A$ for $0 < \theta < \|A^\dagger\|_2^{-1} = \sigma_r(A)$, where $r = \mathit{rank}\,(\,A\,)$, $\mu = \sigma_{r+1}(\tilde{A}) \leq \|\Delta A\|_2$, and $\eta = \sigma_r(\tilde{A}) \geq \sigma_r(A) - \|\Delta A\|_2$. ∎

By Theorem 1, Problem 1 is solvable if the data are sufficiently accurate and a tight error bound on data is attainable, as asserted in the following corollary.

COROLLARY 2. *Let the matrices* A *and* $\tilde{A}$ *be as in Problem* 1*. Assume the data in* $\tilde{A}$ *are sufficiently accurate such that* $\|A-\tilde{A}\|_2 < \frac{1}{2}\,\|A^\dagger\|_2^{-1}$*. Further assume a data error bound* $\beta > \|A-\tilde{A}\|_2$ *is known and is sufficiently tight so that* $\beta < \|A^\dagger\|_2^{-1} - \|A-\tilde{A}\|_2$*. Then Problem* 1 *is solvable by setting the error tolerance* $\theta = \beta$ *and finding an orthonormal basis for the numerical solution* $\mathit{sol}_\theta(\tilde{A},\mathbf{0}) = \mathcal{K}\mathit{ernel}(\tilde{A}_\theta)$ *within* θ*.*

Proof. The proof is straightforward from Theorem 1. ∎

The error tolerance θ in Theorem 1 is an operational parameter that needs to be set up for solving Problem 1. If we assume that the underlying application allows the data error to a certain extent, say $\|A-\tilde{A}\|_2 < \hat{\theta}$, the data error bound β in Corollary 2 is expected to be below $\hat{\theta}$. The inequality (13) ensures there is a window $(\mu,\,\eta)$ for setting the operational error tolerance θ at β or slightly larger. Using the notation of Problem 1, it is reasonable to assume that the data error bound β on $\|A-\tilde{A}\|_2$ is known or can be estimated. The crucial criterion for operational purpose is to set θ at or slightly above $\|A-\tilde{A}\|_2$ according to Theorem 1(iii). The error tolerance θ should not exceed $\|A^\dagger\|_2^{-1} - \|A-\tilde{A}\|_2$, whose exact value or estimation is not needed if the data error bound β is sufficiently tight. See the supplementary material for examples of setting error tolerances.

For a rank-r matrix A, the sensitivity of solving $A\,\mathbf{x} = \mathbf{0}$ for $\mathit{sol}_\theta(\tilde{A},\mathbf{0})$ in the Grassmannian $\mathcal{G}_{n-r}(\mathbb{C}^n)$ from a perturbed data matrix $\tilde{A}$ is

$$\|A\|_2\,\|A^\dagger\|_2 \quad = \quad \frac{\sigma_1(A)}{\sigma_r(A)} \quad \approx \quad \|\tilde{A}_\theta\|_2\,\|\tilde{A}_\theta^\dagger\|_2$$

from (12) and (14), not infinity or $\kappa(\tilde{A})$. The convention $\kappa(A) = \infty$ for the square singular case and $\kappa(\tilde{A}) = \|\tilde{A}\|_2\,\|\tilde{A}^\dagger\|_2$ may overestimate the sensitivity substantially. Problem 1 may not be solvable if the data error is large beyond, say, $\frac{1}{2}\,\|A^\dagger\|_2^{-1}$, or may not be solved accurately if the data error bound is unknown or the inherent sensitivity $\|A\|_2\,\|A^\dagger\|_2$ is high.

For solving $A\,\mathbf{x} = \mathbf{0}$ with $A \in \mathbb{C}^{m\times n}$, there are differences between the cases $m < n$ and $m \geq n$. The solution is of a positive dimension when $m < n$ regardless of perturbations and, if $\mathit{rank}\,(\,A\,) = m$, the condition $\|A\|_2\,\|A^\dagger\|_2$ is continuous with respect to small perturbations. When $m \geq n$ and $\mathit{sol}(A,\,\mathbf{0})$ is nontrivial, however, the dimension of $\mathit{sol}(A+\Delta A,\,\mathbf{0})$ degrades to zero for almost all perturbations ΔA, and the condition $\|A\|_2\,\|A^\dagger\|_2$ is discontinuous. The assertions of Theorem 1 remain the same no matter whether $m < n$ or $m \geq n$.

5. Sensitivity of a consistent singular system. Solving a singular system for an individual vector solution is known to have an unbounded sensitivity under arbitrary perturbations. From a different perspective, the infinity condition number is not the sensitivity of the *singular system* if the singularity is not maintained. There is an intrinsic stability in solving $A\mathbf{x} = \mathbf{b}$ when the rank and consistency are preserved. This point of view originated in [15], where Kahan suggested that the perceived hypersensitivity of multiple roots may be a "misconception" without maintaining the multiplicity.

A consistent $m \times n$ linear system $A\mathbf{x} = \mathbf{b}$ with $\mathit{rank}(A) = r$ has a unique solution $\mathit{sol}(A, \mathbf{b}) = \mathbf{x}_0 + \mathit{Kernel}(A)$ in the affine Grassmannian $\mathcal{A}_{n-r}(\mathbb{C}^n)$, where $\mathbf{x}_0$ is any particular solution. The sensitivity of the linear system $A\mathbf{x} = \mathbf{b}$ can be based on the deviation of the solution $\mathit{sol}(A, \mathbf{b})$ in $\mathcal{A}_{n-r}(\mathbb{C}^n)$ with respect to perturbations of $(A, \mathbf{b}) \in \mathbb{C}^{m\times n} \times \mathbb{C}^n$. From (1), the difference between solutions of two consistent systems of the same rank can be measured by the metric (1), namely,

$$\begin{aligned} &\mathit{dist}\,(\,\mathit{sol}(A,\mathbf{b}),\ \mathit{sol}(B,\mathbf{d})\,) \\ &\qquad = \max\left\{\|A^\dagger\,\mathbf{b} - B^\dagger\,\mathbf{d}\|_2,\ \mathit{dist}\,(\,\mathit{Kernel}(A),\ \mathit{Kernel}(B)\,)\,\right\}. \end{aligned} \tag{15}$$

Notice that the component $\mathit{dist}\,(\,\mathit{Kernel}(A),\ \mathit{Kernel}(B)\,) \le 1$ in (15), but the other component $\|A^\dagger\,\mathbf{b} - B^\dagger\,\mathbf{d}\|_2$ can be large or small. One way to avoid an imbalance is to put a weight factor ω on the component $\|A^\dagger\,\mathbf{b} - B^\dagger\,\mathbf{d}\|_2$. We choose not to use weights for the sake of simplicity and because the weight ω can be used to scale the linear system instead so that we can solve $A\,(\omega\,\mathbf{x}) = \omega\,\mathbf{b}$ equivalently. For convenience, we adopt a specific norm

$$\|(A, \mathbf{b})\| \;:=\; \sqrt{\|A\|_2^2 + \|\mathbf{b}\|_2^2} \tag{16}$$

in the product space $\mathbb{C}^{m\times n} \times \mathbb{C}^n$. The theories in this paper can be adapted to other norms.

With these notation and metrics, the solution $\mathit{sol}(A, \mathbf{b})$ of a singular consistent $m \times n$ linear system $A\mathbf{x} = \mathbf{b}$ uniquely exists in the affine Grassmannian $\mathcal{A}_{n-r}(\mathbb{C}^n)$, and the sensitivity is proportional to $\|A\|_2\,\|A^\dagger\|_2$ rather than infinity when the rank and consistency are preserved, as established in the following theorem.

Theorem 3. The solution of a consistent linear system is Lipschitz continuous when the rank and consistency are preserved. *Let* $A \in \mathbb{C}^{m\times n}$ *and* $\mathbf{b} \in \mathit{Range}(A)$. *Assume the perturbation* $(\Delta A,\, \Delta\mathbf{b})$ *is constrained such that* $\mathit{rank}\left(\tilde{A}\right) = \mathit{rank}(A)$ *and* $\tilde{\mathbf{b}} \in \mathit{Range}(\tilde{A})$, *where* $\tilde{A} = A + \Delta A$ *and* $\tilde{\mathbf{b}} = \mathbf{b} + \Delta\mathbf{b}$. *Then,*

$$\begin{aligned} &\mathit{dist}\left(\,\mathit{sol}(\tilde{A},\,\tilde{\mathbf{b}}),\ \mathit{sol}(A,\,\mathbf{b})\,\right) \\ &\quad\le\; \|A\|_2\,\|A^\dagger\|_2 \cdot \frac{\sqrt{2\,\|\mathbf{x}_*\|_2^2 + 1}}{\|A\|_2 - \sqrt{2}\,\|A\|_2\,\|A^\dagger\|_2\,\|\Delta A\|_2}\,\|(\Delta A,\,\Delta\mathbf{b})\|, \end{aligned} \tag{17}$$

where $\mathbf{x}_* = A^\dagger\,\mathbf{b}$ *whenever* $\sqrt{2}\,\|A^\dagger\|_2\,\|\Delta A\|_2 < 1$.

Proof sketch. The kernel component of the distance in (17) is bounded by Wedin's error estimate [34] (see Lemma 11 in Appendix A). Let N be a matrix whose columns form an orthonormal basis for $\mathit{Kernel}(A)$. Then the minimum norm solution $\mathbf{x}_*$ is the unique least squares solution of the system

$$\begin{bmatrix} \mu\,N^{\mathsf{H}} \\ A \end{bmatrix} \mathbf{x} \;=\; \begin{bmatrix} \mathbf{0} \\ \mathbf{b} \end{bmatrix} \qquad \text{for any} \qquad \mu > 0,$$

and the standard error bound [24, Theorem 1.4.6] applies. A detailed proof is given in Appendix B. ∎

As a result of (17), the intrinsic sensitivity of solving a singular system $A\mathbf{x} = \mathbf{b}$ for the general solution $\mathit{sol}(A, \mathbf{b})$ is a constant multiple of

$$\|A\|_2 \left\|A^\dagger\right\|_2 \;=\; \frac{\sigma_1(A)}{\sigma_r(A)} \;<\; \infty$$

when the rank and consistency are preserved.

As a by-product of establishing Theorem 3, the following corollary improves the standard normwise error bound [9, Theorem 5.6.1] on the minimum norm solution of a full rank underdetermined linear system by reducing a factor from 2 to $\sqrt{2}$.

Corollary 4. *Let $A \in \mathbb{C}^{m\times n}$ with $\mathit{rank}(A) = m < n$ and $\mathbf{b} \in \mathbb{C}^m$. If $\mathbf{x}_*$ and $\tilde{\mathbf{x}}$ are minimum norm solutions of the underdetermined linear systems $A\mathbf{x} = \mathbf{b}$ and $(A + \Delta A)\mathbf{x} = \mathbf{b} + \Delta\mathbf{b}$, respectively, with $\sqrt{2}\,\|A^\dagger\|_2 \left\|\Delta A\right\|_2 < 1$, then*

$$\frac{\|\tilde{\mathbf{x}} - \mathbf{x}_*\|_2}{\|\mathbf{x}_*\|_2} \;\le\; \|A\|_2 \,\|A^\dagger\|_2 \left(\sqrt{2}\,\frac{\|\Delta A\|_2}{\|A\|_2} + \frac{\|\Delta\mathbf{b}\|_2}{\|\mathbf{b}\|_2}\right) + O(\|(\Delta A, \Delta\mathbf{b})\|^2). \tag{18}$$

Proof. The inequality (18) follows from (45) in Appendix B. ∎

Remark 1. The subset of all rank-r matrices is a complex analytic manifold in the topological space $\mathbb{C}^{m\times n}$ [5], with the topology derived from the Frobenius norm. Similarly, the subset $\mathcal{M}_r^{m\times n} := \left\{(A, \mathbf{b}) \in \mathbb{C}^{m\times n} \times \mathbb{C}^m \,\middle|\, \mathit{rank}(A) = r, \mathbf{b} \in \mathit{Range}(A)\right\}$ is a complex analytic manifold in $\mathbb{C}^{m\times n} \times \mathbb{C}^m$. Although in general the problem of solving a singular linear system is ill-posed, Theorem 3 implies that the problem of solving $A\mathbf{x} = \mathbf{b}$ for $\mathit{sol}(A, \mathbf{b})$ in $\mathcal{A}_{n-r}(\mathbb{C}^n)$ is well-posed on the manifold $\mathcal{M}_r^{m\times n}$.

6. The general numerical solution.

When a rank-deficient $m \times n$ linear system $A\mathbf{x} = \mathbf{b}$ is given through empirical data $(\tilde{A}, \tilde{\mathbf{b}})$, the perturbed matrix $\tilde{A}$ is almost always of full rank and highly ill-conditioned. Furthermore, the conventional single-vector solution of the data system $\tilde{A}\mathbf{x} = \tilde{\mathbf{b}}$ is in $\mathbb{C}^n$, while the general solution of the underlying system is in the completely different space $\mathcal{A}_{n-r}(\mathbb{C}^n)$. What the problem precisely is and what the numerical solution really means need to be clarified.

Problem 2 (numerical solution of a linear system). *For given $\tilde{A}$ and $\tilde{\mathbf{b}}$ serving as empirical data for an underlying linear system $A\mathbf{x} = \mathbf{b}$ to be solved, find a numerical solution of $\tilde{A}\mathbf{x} = \tilde{\mathbf{b}}$ that can be identified as the exact solution $\mathit{sol}(\hat{A}, \hat{\mathbf{b}})$ of $\hat{A}\mathbf{x} = \hat{\mathbf{b}}$ such that both the backward error and the forward error,*

$$\left\|(\tilde{A}, \tilde{\mathbf{b}}) - (\hat{A}, \hat{\mathbf{b}})\right\| \;=\; O\big(\left\|(\tilde{A}, \tilde{\mathbf{b}}) - (A, \mathbf{b})\right\|\big), \tag{19}$$

$$\mathit{dist}\Big(\mathit{sol}(\hat{A}, \hat{\mathbf{b}}),\, \mathit{sol}(A, \mathbf{b})\Big) \;=\; O\big(\left\|(\tilde{A}, \tilde{\mathbf{b}}) - (A, \mathbf{b})\right\|\big), \tag{20}$$

are in the same order of the data accuracy.

The accuracy requirement (20) stipulates that both $\mathit{sol}(A, \mathbf{b})$ and $\mathit{sol}(\hat{A}, \hat{\mathbf{b}})$ are either in the same affine Grassmannian or empty. It is natural to choose $\hat{A} = \tilde{A}_\theta$ within a proper θ and $\hat{\mathbf{b}} = \tilde{\mathbf{b}}_\theta := \tilde{A}_\theta \tilde{A}_\theta^\dagger \tilde{\mathbf{b}}$ as the orthogonal projection of $\tilde{\mathbf{b}}$ onto $\mathit{Range}(\tilde{A}_\theta)$. The solution $\mathit{sol}(\tilde{A}_\theta, \tilde{\mathbf{b}}_\theta)$ is acceptable as the numerical solution of $\tilde{A}\mathbf{x} = \tilde{\mathbf{b}}$ if its backward error is below the error tolerance or the empty set.

Definition 5 (general numerical solution). *Let $G \in \mathbb{C}^{m\times n}$, $\mathbf{d} \in \mathbb{C}^m$, and $\theta > 0$ be an error tolerance within which $\mathit{rank}_\theta(G)$ is well defined. With respect* *to a norm* $\|\cdot\|$ *on* $\mathbb{C}^{m\times n}\times\mathbb{C}^m$, *the* general numerical solution *of the linear system* $G\,\mathbf{x} = \mathbf{d}$ within θ *is defined as*

$$\mathit{sol}_\theta(G,\mathbf{d}) \;:=\; \begin{cases} \mathit{sol}(G_\theta,\mathbf{d}_\theta) & \text{if } \|(G,\mathbf{d})-(G_\theta,\mathbf{d}_\theta)\| < \theta,\\ \emptyset & \text{if } \|(G,\mathbf{d})-(G_\theta,\mathbf{d}_\theta)\| > \theta,\end{cases}$$

where G_θ *is the* θ*-projection of* G, *and* $\mathbf{d}_\theta = G_\theta\,G_\theta^\dagger\,\mathbf{d}$ *is the orthogonal projection of* $\mathbf{d}$ *onto the range* $\mathit{Range}(G_\theta)$ *of* G_θ.

The solution $\mathit{sol}_\theta(G,\mathbf{d})$ is undefined if θ is equal to a singular value of G or $\theta = \|(G,\mathbf{d})-(G_\theta,\mathbf{d}_\theta)\|$. We can now establish the following theorem on the general numerical solution.

THEOREM 6. *At any* $(A,\,\mathbf{b}) \in \mathbb{C}^{m\times n}\times\mathbb{C}^m$, *the following properties of the general numerical solution hold with respect to the norm* (16)*:*

(i) An exact general solution is a special case of general numerical solution: *The identity* $\mathit{sol}(A,\,\mathbf{b}) \equiv \mathit{sol}_\theta(A,\,\mathbf{b})$ *holds for all* $\theta < \|A^\dagger\|_2^{-1}$ *if* $\mathbf{b}\in\mathit{Range}(A)$, *or* $\theta < \min\big\{\|A^\dagger\|_2^{-1},\,\|\mathbf{b}-A\,A^\dagger\,\mathbf{b}\|_2\big\}$ *otherwise.*

(ii) Computing the general numerical solution is a well-posed problem: *Assume* $\mathit{sol}_\theta(A,\mathbf{b})$ *is well-defined within a certain* $\theta > 0$. *There is a* $\xi > 0$ *depending on* A, $\mathbf{b}$, *and* θ *such that for every* $(\Delta A,\,\Delta\mathbf{b})$ *with a sufficiently small norm, there exists a unique* $\mathit{sol}_\theta(A+\Delta A,\,\mathbf{b}+\Delta\mathbf{b})$ *satisfying the Lipschitz continuity*

$$\mathit{dist}\,\big(\,\mathit{sol}_\theta(A+\Delta A,\,\mathbf{b}+\Delta\mathbf{b}),\,\mathit{sol}_\theta(A,\mathbf{b})\,\big) \;\le\; \xi\,\big\|(\Delta A,\,\Delta\mathbf{b})\big\|. \tag{21}$$

(iii) A singular linear system can be solved from empirical data with an accuracy in the same order as the data:

(a) *Assume* $\mathbf{b}\in\mathit{Range}(A)$, *and let* $\mathbf{x}_* = A^\dagger\,\mathbf{b}$. *For any empirical data pair* $(\tilde{A},\,\tilde{\mathbf{b}}) = (A+\Delta A,\,\mathbf{b}+\Delta\mathbf{b})$ *satisfying* $\big\|(\Delta A,\,\Delta\mathbf{b})\big\| < \big((\omega+1)\,\|A^\dagger\|_2\big)^{-1}$, *where* $\omega = \sqrt{4\,\|A^\dagger\|_2^2\,\|\mathbf{b}\|_2^2+2}$, *and for any error tolerance* θ *satisfying*

$$\omega\,\big\|(\Delta A,\,\Delta\mathbf{b})\big\| \;<\; \theta \;<\; \|A^\dagger\|_2^{-1} - \big\|(\Delta A,\,\Delta\mathbf{b})\big\|, \tag{22}$$

there exists a unique general numerical solution $\mathit{sol}_\theta(\tilde{A},\tilde{\mathbf{b}})$ *with a backward error bound* $\omega\,\big\|(\Delta A,\,\Delta\mathbf{b})\big\|$ *and a forward error bound*

$$\begin{aligned} &\mathit{dist}\Big(\,\mathit{sol}_\theta(\tilde{A},\tilde{\mathbf{b}}),\,\mathit{sol}(A,\mathbf{b})\,\Big)\\ &\quad\le\; \|A\|_2\,\|A^\dagger\|_2\,\frac{\sqrt{4\,\|\mathbf{x}_*\|_2^2+1}}{\|A\|_2-\|A\|_2\,\|A^\dagger\|_2\,\|\Delta A\|_2}\,\big\|(\Delta A,\,\Delta\mathbf{b})\big\|. \end{aligned} \tag{23}$$

(b) *Assume* $\mathit{sol}(A,\mathbf{b}) = \emptyset$. *For any* $\theta < \min\big\{\tfrac{1}{2}\,\|A^\dagger\|_2^{-1},\,\|\mathbf{b}-A\,A^\dagger\,\mathbf{b}\|_2\big\}$, *there is a constant* $\rho\in(0,\theta)$ *such that* $\mathit{sol}_\theta(\tilde{A},\,\tilde{\mathbf{b}}) = \mathit{sol}(A,\mathbf{b}) = \emptyset$ *at any empirical data pair* $(\tilde{A},\,\tilde{\mathbf{b}})$ *satisfying* $\big\|(\tilde{A},\,\tilde{\mathbf{b}})-(A,\,\mathbf{b})\big\| < \rho$.

Proof sketch. Assertion (i) and the unique existence in assertion (ii) directly follow from Definition 5. The Lipschitz continuity (21) is a variation of the error estimate for the truncated SVD solution by Hansen [10, inequality (26a)] as an extension of Wedin error analysis [35]. The bound on the minimum norm solution component of the distance in the inequality (23) follows from Hansen [10, inequality (27a)], and the bound on the numerical kernel is established by Wedin [34]. A detailed proof is given in Appendix B. ∎

For Problem 2, assume the underlying linear system $A\mathbf{x} = \mathbf{b}$ in Problem 2 is known to be consistent in applications such as those in Example 3; the solvability of the system from empirical data $(\tilde{A}, \tilde{\mathbf{b}})$ is given in the following corollary of Theorem 6.

Corollary 7. *Let $(A, \mathbf{b})$ and $(\tilde{A}, \tilde{\mathbf{b}})$ be as in Problem 2, where the underlying linear system $A\mathbf{x} = \mathbf{b}$ is consistent. Assume the data matrix $\tilde{A}$ is sufficiently accurate with $\|A - \tilde{A}\|_2 < \frac{1}{2}\|A^\dagger\|_2^{-1}$. Further assume that an error bound $\beta > \|A - \tilde{A}\|_2$ is attainable and is sufficiently tight so that $\beta < \|A^\dagger\|_2^{-1} - \|A - \tilde{A}\|_2$. Then Problem 2 is solvable by calculating $\mathit{sol}(\tilde{A}_\theta, \tilde{\mathbf{b}}_\theta)$ with the error tolerance $\theta = \beta$, where $\tilde{\mathbf{b}}_\theta$ is the orthogonal projection of $\tilde{\mathbf{b}}$ onto $\mathcal{Range}(\tilde{A}_\theta)$. Furthermore,*

$$\begin{aligned} &\mathit{dist}\Big(\mathit{sol}(\tilde{A}_\theta, \tilde{\mathbf{b}}_\theta),\ \mathit{sol}(A, \mathbf{b})\Big) \\ &\quad \leq \|A\|_2 \|A^\dagger\|_2 \cdot \frac{\sqrt{4\|A^\dagger \mathbf{b}\|_2^2 + 1}}{\|A\|_2 - \|A\|_2 \|A^\dagger\|_2 \|A - \tilde{A}\|_2} \|(\tilde{A}, \tilde{\mathbf{b}}) - (A, \mathbf{b})\|. \end{aligned}$$

We reiterate that the sensitivity of solving $A\mathbf{x} = \mathbf{b}$ from empirical data $(\tilde{A}, \tilde{\mathbf{b}})$ is measured by

$$\|A\|_2 \|A^\dagger\|_2 \;=\; \frac{\sigma_1(A)}{\sigma_r(A)} \;\approx\; \|\tilde{A}_\theta\|_2 \|\tilde{A}_\theta^\dagger\|_2,$$

not infinity or $\kappa(\tilde{A})$, when the underlying matrix A is singular, where $r = \mathit{rank}(A)$. Problem 2 may still be difficult if data are inaccurate, if the intrinsic condition $\|A\|_2 \|A^\dagger\|_2$ is large, or if the window for setting the error tolerance is too narrow. The general numerical solution can be computed using existing rank-revealing tools, such as those in [18, 20] and UTV/ULV decomposition [9, sect. 5.4.6], in the following template:

> set the error tolerance θ at or slightly above the error bound $\beta \gtrapprox \|\Delta A\|_2$
> `if` $r = \mathit{rank}_\theta(A) \approx n$ `then`
> – calculate $N \in \mathbb{C}^{n\times(n-r)}$, whose columns form an orthonormal basis for the numerical kernel $\mathcal{Kernel}(A_\theta)$
> – solve $A\mathbf{x} = \mathbf{b}$ for a particular solution $\mathbf{x} = \mathbf{x}_*$ by any backward accurate method, such as $\mathbf{x}_* = (A^{\mathsf{H}}A + \mu^2 N N^{\mathsf{H}})^{-1}A^{\mathsf{H}}\mathbf{b}$ or Tikhonov regularization
> – `output` $\mathit{sol}_\theta(A, \mathbf{b}) = \mathbf{x}_* + \mathcal{Range}(N)$.
> `else`
> – calculate a decomposition $U S V^{\mathsf{H}} = A_\theta$ with $S \in \mathbb{C}^{r\times r}$, $U^{\mathsf{H}}U = I$, and $V^{\mathsf{H}}V = I$
> – solve $S\mathbf{y} = U^{\mathsf{H}}\mathbf{b}$ for $\mathbf{y} = \mathbf{y}_*$ and obtain the truncated SVD solution $\mathbf{x}_* = V\mathbf{y}_*$
> – `output` $\mathit{sol}_\theta(A, \mathbf{b}) = \mathbf{x}_* + \mathcal{Range}(V)^\perp$
> `end if`

As we shall establish in section 7, the particular solution component $\mathbf{x}_*$ of $\mathit{sol}_\theta(A, \mathbf{b})$ in the above template can be computed by any backward accurate numerical algorithm, including Tikhonov regularization and truncated SVD. Computation of the general numerical solution is implemented in the MATLAB package NAClab [40] as the functionality `LinearSolve` (cf. [39] and the supplementary material). The general guideline for the error tolerance is to set it at or slightly larger than a known data error bound $\beta > \|A - \tilde{A}\|_2$ if the application allows such an adjustment. We conclude this section with the following example.

Example 4. Revisiting the linear system in Example 3, we see that the data error bound can be estimated as $\|\Delta A\|_2 \leq \|\Delta A\|_F \leq 4.5\times 10^{-4}$, where A is the underlying matrix since the entrywise error bound is 5×10^{-5}. The error tolerance θ can be set at or slightly larger than the error bound, say $\theta = 0.0005$. The numerical solution of the system (5) within 0.0005 in the affine Grassmannian $\mathcal{A}_7(\mathbb{C}^9)$ is a representation of

$$
\begin{aligned}
&(u_1,\, u_2, u_3)\\
&\qquad = \big(.90710 + .33322\,x + .71029\,x^2 + .59968\,x^3,\ \ -.79946 + .06694\,x,\ \ 1.12433 - .06648\,x + .08926\,x^2\big)\\
&\quad + t_1\,\big(-.27897 - .08391\,x - .17878\,x^2 + .08424\,x^3,\ \ -.35739 - .47261\,x,\ \ .12212 - .33612\,x - .63016\,x^2\big)\\
&\quad + t_2\,\big(-.21387 + .29319\,x - .18465\,x^2 + .46503\,x^3,\ \ -.55471 + .18011\,x,\ \ -.46785 + .03542\,x + .24016\,x^2\big)
\end{aligned}
$$

(cf. the supplementary material). The general numerical solution $\mathit{sol}_\theta(\tilde{A}, \tilde{\mathbf{b}})$ is of a healthy sensitivity $\|\tilde{A}_\theta\|_2\,\|\tilde{A}_\theta^\dagger\|_2 \approx 17.19$—not the infinite $\kappa(A)$ or the large $\kappa(\tilde{A}) \approx 2.29 \times 10^6$. The three components of $\mathit{sol}_\theta(\tilde{A})$ form an invertible polynomial transformation matrix as shown in (6) with the numerical inverse

$$
\begin{bmatrix}
.55101 - .91839\,x^2, & -2.33985 + .70128\,x + .30047\,x^2 + .00001\,x^3, & .71342 + 1.83982\,x + 1.12986\,x^2 + .00002\,x^3\\
-.36735 + .18367\,x - .73471\,x^5, & -.43135 - 1.09668\,x + .33663\,x^2 - 1.72768\,x^3 + .56105\,x^4 + 0.24037\,x^5, & -.9954 + .44186\,x + .8831\,x^2 + 1.11306\,x^3 + 1.47187\,x^4 + .90389\,x^5\\
.18366 + .36734\,x^2 + .55103\,x^4, & 1.58108 - .53294\,x + 1.17553\,x^2 - .42079\,x^3 - .18027\,x^4, & -1.28338 - 1.39825\,x - 1.28672\,x^2 - 1.10389\,x^3 - .6779\,x^4
\end{bmatrix}.
$$

Remark 2. An $m \times n$ system $A\mathbf{x} = \mathbf{b}$ with $m > n$ is inconsistent for almost all $\mathbf{b} \in \mathbb{C}^m$, and its least squares solution is usually studied in the literature. In fact, the least squares solution can be considered whenever $\mathbf{b} \notin \mathit{Range}(A)$ even if $m \leq n$. There are substantial differences between the conventional solution and the least squares solution. In Theorem 6 and throughout this paper, our elaboration is restricted to the conventional solution so that $\mathit{sol}(A, \mathbf{b}) = \emptyset$ for inconsistent systems, and the nonempty set of least squares solutions is beyond the scope of this paper. The sensitivity of the least squares solution is well known to be $\kappa(A)^2$ (see, e.g., [14, sect. 20.1]) in contrast to $\kappa(A)$ for the (conventional) solution in Theorem 6.

7. Particular solution of a singular linear system. There are many applications where only a particular solution is needed from among the infinitely many solutions of a singular linear system $A\mathbf{x} = \mathbf{b}$, and it makes little difference which particular solution is obtained. For such applications, the problem of finding a *numerical particular solution* can be stated as follows.

PROBLEM 3 (numerical particular solution). *Assume a linear system $A\mathbf{x} = \mathbf{b}$ is consistent where the entries of A and $\mathbf{b}$ may be known through empirical data of limited accuracy. Find a numerical particular solution $\tilde{\mathbf{x}}$ that approximates an exact solution $\mathbf{x}_* \in \mathit{sol}(A, \mathbf{b})$ with the error $\|\tilde{\mathbf{x}} - \mathbf{x}_*\|_2$ at an acceptable level.*

There are regularization approaches, such as the Tikhonov method [9, sect. 6.1.5], [11, 27] that can produce approximate particular solutions with high backward accuracy. For any backward accurate numerical solution $\tilde{\mathbf{x}}$ of the system $A\mathbf{x} = \mathbf{0}$, in the sense that there is a pair $(\tilde{A}, \tilde{\mathbf{b}})$ such that $\tilde{A}\tilde{\mathbf{x}} = \tilde{\mathbf{b}}$ and that $\|(A, \mathbf{b}) - (\tilde{A}, \tilde{\mathbf{b}})\|$ is at an acceptable level, we call $\tilde{\mathbf{x}}$ a *numerical particular solution* of $A\mathbf{x} = \mathbf{b}$. The following theorem asserts that every numerical particular solution approximates one of the exact solutions.

THEOREM 8. *Let $A \in \mathbb{C}^{m\times n}$ with $\mathit{rank}(A) < n$ and $\mathbf{b} \in \mathit{Range}(A)$. Assume $\tilde{\mathbf{x}} \in \mathbb{C}^n$ is a backward accurate numerical solution of $A\mathbf{x} = \mathbf{b}$ in the sense that $\tilde{\mathbf{x}}$ is* *an exact solution of* $\tilde{A}\,\mathbf{x} = \tilde{\mathbf{b}}$ *with* $\|\tilde{A} - A\|_2 \leq .46\,\|A^\dagger\|_2^{-1}$*. Then* $\tilde{\mathbf{x}}$ *approximates an exact solution* $\mathbf{x}_* \in \mathit{sol}(A, \mathbf{b})$ *with an error bound*

$$\frac{\|\tilde{\mathbf{x}} - \mathbf{x}_*\|_2}{\|\mathbf{x}_*\|_2} \;\leq\; \frac{\|A\|_2\,\|A^\dagger\|_2}{1 - \|A^\dagger\|_2\,\|\Delta A\|_2}\left(2\sqrt{2}\,\frac{\|\Delta A\|_2}{\|A\|_2} + \frac{\|\Delta \mathbf{b}\|_2}{\|\mathbf{b}\|_2}\right), \tag{24}$$

assuming $\mathbf{b} \neq \mathbf{0}$ *where* $\Delta A = A - \tilde{A}$*,* $\Delta\mathbf{b} = \mathbf{b} - \tilde{\mathbf{b}}$*, or*

$$\|\tilde{\mathbf{x}} - \mathbf{x}_*\|_2 \;\leq\; \frac{\|A\|_2\,\|A^\dagger\|_2}{1 - \|A^\dagger\|_2\,\|\Delta A\|_2}\left(\|\tilde{\mathbf{x}}\|_2\,\frac{\|\Delta A\|_2}{\|A\|_2} + \frac{\|\Delta \mathbf{b}\|_2}{\|A\|_2}\right) \tag{25}$$

if $\mathbf{b} = \mathbf{0}$.

Proof sketch. Let $r = \mathit{rank}(A)$ and $\sigma_{r+1}(\tilde{A}) < \theta < \sigma_r(\tilde{A})$. Write $\tilde{\mathbf{x}} = \tilde{\mathbf{x}}_1 + \tilde{\mathbf{x}}_2$, where $\tilde{\mathbf{x}}_1 = \tilde{A}_\theta^\dagger\,\tilde{A}_\theta\,\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}_2 = (I - \tilde{A}_\theta^\dagger\,\tilde{A}_\theta)\,\tilde{\mathbf{x}}$. Choose a particular solution $\mathbf{x}_* = A^\dagger\,\mathbf{b} + (I - A^\dagger\,A)\,\tilde{\mathbf{x}}_2$ from $\mathit{sol}(A, \mathbf{b})$. Since $\tilde{\mathbf{x}}_1 = \tilde{A}_\theta^\dagger\,\tilde{\mathbf{b}}$ approximates $A^\dagger\,\mathbf{b}$, $\tilde{\mathbf{x}}_2 \in \mathit{Kernel}(\tilde{A}_\theta)$, $(I - A^\dagger\,A)\,\tilde{\mathbf{x}}_2 \in \mathit{Kernel}(A)$, and $\mathit{Kernel}(\tilde{A}_\theta)$ approximates $\mathit{Kernel}(A)$, and hence $\tilde{\mathbf{x}}$ is an approximation to the particular solution $\mathbf{x}_*$ of $A\,\mathbf{x} = \mathbf{b}$, so the theorem holds. Detailed proofs of (24) and (25) are given in Appendix B. ∎

For the case $\mathbf{b} = \mathbf{0}$ in Theorem 8, the objective is to solve the homogeneous system $A\,\mathbf{x} = \mathbf{0}$. The inequality (25) includes the following three cases:

Case (i): $\tilde{\mathbf{b}} = \mathbf{0}$ and $\tilde{\mathbf{x}} = \mathbf{0}$. The inequality (25) is trivial and perhaps meaningless since $\tilde{\mathbf{x}} = \mathbf{x}_* = \mathbf{0}$.

Case (ii): $\tilde{\mathbf{b}} = \mathbf{0}$ and $\tilde{\mathbf{x}} \neq \mathbf{0}$. Then we can normalize $\tilde{\mathbf{x}}$ to be a unit vector so that (25) becomes

$$\min_{\mathbf{z} \in \mathit{Kernel}(A)} \|\tilde{\mathbf{x}} - \mathbf{z}\|_2 \;\leq\; \|\tilde{\mathbf{x}} - \mathbf{x}_*\|_2 \;\leq\; \frac{\|A\|_2\,\|A^\dagger\|_2}{1 - \|A^\dagger\|_2\,\|\Delta A\|_2}\,\frac{\|\Delta A\|_2}{\|A\|_2}. \tag{26}$$

Case (iii): $\tilde{\mathbf{b}} \neq \mathbf{0}$. This case is relevant in practical computation by setting the right-hand side $\tilde{\mathbf{b}}$ as a nonzero random vector of a moderate norm and obtaining a numerical particular solution $\tilde{\mathbf{x}}$ as an exact solution of $\tilde{A}\,\mathbf{x} = \tilde{\mathbf{b}}$ with a small $\|A - \tilde{A}\|_2$, leading to the inverse power iteration. The norm $\|\tilde{\mathbf{x}}\|_2$ is almost always large due to the condition number $\kappa(\tilde{A}) = O(\|A - \tilde{A}\|_2^{-1})$. As it pleasantly turns out, the large $\|\tilde{\mathbf{x}}\|_2$ is exactly what is needed as (25) becomes

$$\left\|\frac{\tilde{\mathbf{x}}}{\|\tilde{\mathbf{x}}\|_2} - \frac{\mathbf{x}_*}{\|\tilde{\mathbf{x}}\|_2}\right\|_2 \;\leq\; \frac{\|A\|_2\,\|A^\dagger\|_2}{1 - \|A^\dagger\|_2\,\|\Delta A\|_2}\,\frac{1}{\|A\|_2}\left(\|\Delta A\|_2 + \frac{\|\tilde{\mathbf{b}}\|_2}{\|\tilde{\mathbf{x}}\|_2}\right). \tag{27}$$

The larger the norm achieved by $\|\tilde{\mathbf{x}}\|_2$, the higher the accuracy of $\frac{\tilde{\mathbf{x}}}{\|\tilde{\mathbf{x}}\|_2}$ to a particular nontrivial solution of the homogeneous system $A\,\mathbf{x} = \mathbf{0}$. Once again, the sensitivity of solving a singular linear system $A\,\mathbf{x} = \mathbf{b}$ is $\|A\|_2\,\|A^\dagger\|_2 = \frac{\sigma_1(A)}{\sigma_r(A)}$—not infinity in the sense of finding a numerical particular solution.

Particular solutions of $A\,\mathbf{x} = \mathbf{b}$ can vary arbitrarily, but their deviations can only stretch in $\mathit{Kernel}(A)$. As the following corollary states, the high sensitivity is near a direction in $\mathit{Kernel}(A)$, and such a sensitivity may be harmless after all.

COROLLARY 9. *Let* $A \in \mathbb{C}^{m\times n}$ *with* $\mathit{rank}(A) < n$ *and* $\mathbf{b} \in \mathit{Range}(A)$*. Assume* $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *are both backward accurate numerical particular solutions of* $A\,\mathbf{x} = \mathbf{b}$ *in* *the sense that* $A_1\,\mathbf{x}_1 = \mathbf{b}_1$ *and* $A_2\,\mathbf{x}_2 = \mathbf{b}_2$ *with sufficiently small* $\|(A_1,\, \mathbf{b}_1) - (A,\, \mathbf{b})\|$ *and* $\|(A_2,\, \mathbf{b}_2) - (A,\, \mathbf{b})\|$. *Then there is an* $\mathbf{x}_* \in \mathscr{Kernel}(A)$ *such that*

$$\begin{aligned} \|(\mathbf{x}_1 - \mathbf{x}_2) - \mathbf{x}_*\|_2 \quad &\leq \quad \|A\|_2 \left\|A^\dagger\right\|_2 \\ &\times \left(\frac{\|\mathbf{b} - \mathbf{b}_1\|_2}{\|A\|_2} + \frac{\|\mathbf{b} - \mathbf{b}_2\|_2}{\|A\|_2} + \frac{\|A - A_1\|_2}{\|A\|_2}\|\mathbf{x}_1\|_2 + \frac{\|A - A_2\|_2}{\|A\|_2}\|\mathbf{x}_2\|_2 \right). \end{aligned} \tag{28}$$

Proof. Apply the inequality (25) on $\tilde{A} = A$ and $\tilde{\mathbf{x}} = \mathbf{x}_1 - \mathbf{x}_2$ that satisfy $A\,(\mathbf{x}_1 - \mathbf{x}_2) = (\mathbf{b}_1 - \mathbf{b}) + (\mathbf{b} - \mathbf{b}_2) + (A - A_1)\,\mathbf{x}_1 + (A_2 - A)\,\mathbf{x}_2$. ∎

Theorem 8 extends the accuracy result for the inverse iteration in spite of the large condition number. In [29, pp. 339–340], Peters and Wilkinson described what they called "exaggerated fears" in the early days of the computer age when the inverse iteration

$$(A - \lambda I)\,\mathbf{x}_{k+1} \quad = \quad \mathbf{x}_k \qquad \text{for} \quad k = 0,\, 1,\, \ldots \tag{29}$$

at an approximation λ to an eigenvalue λ_* of A was proposed for calculating an eigenvector $\mathbf{x}_*$ as a nontrivial solution to the homogeneous system $(A - \lambda_* I)\,\mathbf{x} = \mathbf{0}$:

> Although [inverse iteration is] basically a simple concept its numerical properties have not been widely understood. If λ really is very close to an eigenvalue, the matrix $(A - \lambda I)$ is almost singular and hence a typical step in the iteration involves the solution of a very ill-conditioned set of equations. ... The period when inverse iteration was first considered was notable for exaggerated fears concerning the instability of direct methods for solving linear systems and *ill-conditioned* systems were a source of particular anxiety. ... few numerical analysts discuss inverse iteration with any confidence.

It is counterintuitive, but pleasantly surprising nonetheless, that ill-conditioning is not harmful in computing the eigenvector. As pointed out in [29] and by Parlett [28, sect. 4.3] as follows, errors mainly lie in $\mathscr{Kernel}(A - \lambda_*\, I)$ and are not really errors at all:

> ... roundoff errors can give rise to completely erroneous "solutions" to very ill-conditioned systems of equations. ... Indeed some textbooks have cautioned users not to take [λ] too close to any eigenvalue. ... Fortunately these fears are groundless and furnish a nice example of confusing ends with means. ... the error $\mathbf{e}$ [$= \mathbf{x}_{k+1} - \mathbf{x}_*$], which may be almost as large as the exact solution of [$(A - \lambda I)^{-1}\,\mathbf{x}_k$], is almost entirely in the direction of [the eigenvector]. ... This result is alarming if we had hoped for an accurate solution of [(29)] (the means) but is a delight in the search for [the eigenvector] (the end).

Theorem 8 concludes, in fact, that the fear of solving a highly ill-conditioned linear system may also be exaggerated for nonhomogeneous systems when the underlying system $A\,\mathbf{x} = \mathbf{b}$ is consistent and singular, as long as the numerical solution is backward accurate and the intrinsic sensitivity measure $\|A\|_2\,\|A^\dagger\|_2$ is moderate. The variation between any two numerical particular solutions can be large, but the difference falls harmlessly in the kernel of A. In other words, the "error" is actually a part of the solution.

Example 5. The system $\tilde{A}\,\mathbf{x} = \tilde{\mathbf{b}}$ in (7) is a representation of the underlying system $A\,\mathbf{x} = \mathbf{b}$ with $\|\Delta A\|_2 \leq \|\Delta A\|_F \leq 4.5 \times 10^{-4} = \theta$ from the entrywise error bound 0.5×10^{-5}. Rounded to five digits after the decimal point, two numerical particular solutions $\tilde{\mathbf{x}}_0$ and $\mathbf{x}_1$ by truncated SVD $\tilde{A}_\theta^\dagger \tilde{\mathbf{b}}$ and MATLAB backslash "\" operator, respectively, are

$$\begin{aligned}
\tilde{\mathbf{x}}_0 &= [\, 0.90711,\ 0.33322,\ 0.71029,\ 0.59968,\ -0.79946,\ 0.06694,\ 1.12433,\ -0.06648,\ 0.08926]^{\mathsf{H}},\\
\tilde{\mathbf{x}}_1 &= [-0.78366,\ 0.47296,\ -0.45954,\ 1.83637,\ -3.47453,\ -1.81379,\ 0.84635,\ -1.57209,\ -2.41843]^{\mathsf{H}},
\end{aligned}$$

with both residuals $\|\tilde{A}\,\tilde{\mathbf{x}}_0 - \tilde{\mathbf{b}}\| \approx 8.1 \times 10^{-5}$ and $\|\tilde{A}\,\tilde{\mathbf{x}}_1 - \tilde{\mathbf{b}}\| \approx 5.3 \times 10^{-5}$ roughly within the data error bound. The two numerical particular solutions are far apart with $\|\tilde{\mathbf{x}}_0 - \tilde{\mathbf{x}}_1\| \approx 5.01$ as predicted by the large condition number $\kappa(\tilde{A}) \approx 2.29 \times 10^6$. However, the underlying system is consistent and singular with a healthy sensitivity $\|A\|_2\, \|A^\dagger\|_2 \approx \|\tilde{A}_\theta\|_2\, \|\tilde{A}_\theta^\dagger\|_2 \lessapprox 17.19$. Both $\tilde{\mathbf{x}}_0$ and $\tilde{\mathbf{x}}_1$ are accurate approximations to different exact solutions with estimate error bounds 0.00186 and 0.00177, respectively, and actual relative errors are 4.49×10^{-5} and 0.93×10^{-5} in the same level of the data error.

8. Bona fide ill-conditioned linear systems. A linear system $A\,\mathbf{x} = \mathbf{b}$ is truly ill-conditioned when $\|A\|_2\, \|A^\dagger\|_2$ is large regardless of its rank. When A is of full column rank, $\mathbf{b} \in \mathcal{Range}(A)$, and the condition number $\kappa(A)$ is huge, the solution uniquely exists but in general cannot be computed accurately from perturbed data using any algorithm. The system is de facto rank-deficient in a practical sense. Even in such cases, a stable general numerical solution may still be attainable in an affine Grassmannian from empirical data, and the underlying solution can be accurately approximated by a vector in the affine subspace as the general numerical solution.

THEOREM 10. *Assume $A \in \mathbb{C}^{m\times n}$, $\mathbf{x}_* \in \mathbb{C}^n$, and $\mathbf{b} = A\,\mathbf{x}_*$. Let r be any integer with $\sigma_r(A) > \sigma_{r+1}(A)$. For any $(\tilde{A}, \tilde{\mathbf{b}}) = (A + \Delta A,\, \mathbf{b} + \Delta\mathbf{b})$ serving as empirical data of $(A,\, \mathbf{b})$ with*

$$\|\Delta A\|_2 \;<\; \min\{\sigma_r(A) - \sigma_{r+1}(A),\, (2\sqrt{3} - 3)\,\sigma_r(A)\}, \tag{30}$$

there is an $\tilde{\mathbf{x}} \in \mathcal{sol}_\theta(\tilde{A}, \tilde{\mathbf{b}})$ with $\sigma_{r+1}(\tilde{A}) < \theta < \sigma_r(\tilde{A})$ such that

$$\frac{\|\tilde{\mathbf{x}} - \mathbf{x}_*\|_2}{\|\mathbf{x}_*\|_2} \;\le\; \frac{\sigma_1(A)}{\sigma_r(A)}\, \frac{1}{1 - \frac{\sigma_{r+1}(A) - \|\Delta A\|_2}{\sigma_r(A)}} \left((2 + \sqrt{2})\, \frac{\|\Delta A\|_2}{\|A\|_2} + \frac{\|\Delta \mathbf{b}\|_2}{\|\mathbf{b}\|_2} \right). \tag{31}$$

Proof sketch. Since $\mathbf{x}_*$ is a backward accurate solution of the linear system $\tilde{A}_\theta\,\mathbf{x} = \tilde{\mathbf{b}}_\theta$, Theorem 8 applies from a reversed perspective. The detailed proof is given in Appendix B. ∎

In the following example, the underlying system is ill-conditioned but truly nonsingular. All known numerical algorithms, including regularization methods, produce solutions that are inaccurate as single vectors but highly accurate as the vector component of a general numerical solution that is perfectly conditioned and contains accurate approximations to the underlying exact solution.

Example 6. Consider the polynomial division problem in the form of the equation

$$(x + 10)\, q + \rho \;=\; \tfrac{1}{3}\, x^8 + 4\, x^7 + \tfrac{23}{3}\, x^6 + \tfrac{34}{3}\, x^5 + 15\, x^4 + \tfrac{56}{3}\, x^3 + \tfrac{67}{3}\, x^2 + 26\, x + \tfrac{89}{3}$$

for the quotient q and the constant remainder ρ. There is a unique solution which consists of $q = \frac{1}{3}\,(x^7 + 2\,x^6 + \cdots + 7\,x + 8)$ and $\rho = 3$. The corresponding linear system is of the form $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & & & \\ 10 & 1 & & \\ & \ddots & \ddots & \\ & & 10 & 1 \end{bmatrix}, \quad \mathbf{b} = \tfrac{1}{3}\begin{bmatrix} 1 \\ 12 \\ \vdots \\ 89 \end{bmatrix},$$

with the exact solution $\mathbf{x}_* = \frac{1}{3}[1, 2, \ldots, 9]^{\mathsf{H}}$ that is attainable in symbolic computation using the exact data in rational number format. In MATLAB single precision arithmetic, the system is represented as perturbed data $\tilde{A}\mathbf{x} = \tilde{\mathbf{b}}$, where $\tilde{A} = A$ and $\tilde{\mathbf{b}} = [.3333333,\ 4.0,\ 7.6666665,\ 11.333333,\ 15.0,\ 18.666666,\ 22.333334,\ 26.0,\ 29.666666]^{\mathsf{H}}$. The singular values $10.9461079 > 10.7891169 > \cdots > 9.0683689 > 9.9 \times 10^{-9}$ indicate that it is practically impossible to calculate the single-vector solution with any meaningful accuracy using such data. Table 1 shows three sample numerical solutions: $\mathbf{x}_1$ by a straightforward application of the MATLAB command `A\b`, a Tikhonov regularization solution $\mathbf{x}_2 = (A^{\mathsf{H}}A + \alpha^2 I)^{-1}A^{\mathsf{H}}\tilde{\mathbf{b}}$ at, say, $\alpha = 0.001$, and the truncated SVD solution $\mathbf{x}_3 = A_\theta^\dagger \tilde{\mathbf{b}}$ with an error tolerance that is roughly $\theta = \|\mathbf{b}\|_2\,\varepsilon \approx 3.18 \times 10^{-6}$, where ε is the unit roundoff. As expected from the condition number $\kappa(A) \approx 1.1 \times 10^9$, none of the solutions can be considered accurate *as a single vector*. On the other hand, the general numerical solution $\mathit{sol}_\theta(\tilde{A}, \tilde{\mathbf{b}})$ is almost perfectly conditioned at $\|\tilde{A}_\theta\|_2 \|\tilde{A}_\theta^\dagger\|_2 \approx 1.21$. The three solutions $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ that are inaccurate as individual vectors are all accurate as the component $\tilde{\mathbf{u}}$ of the general numerical solution $\mathit{sol}_\theta(\tilde{A}, \tilde{\mathbf{b}}) = \tilde{\mathbf{u}} + \mathcal{K}\mathit{ernel}(\tilde{A}_\theta)$ with $\mathcal{K}\mathit{ernel}(\tilde{A}_\theta) = \mathit{span}\{\tilde{\mathbf{v}}\}$, where

$$\tilde{\mathbf{v}} = [.0,\ -.0000001,\ .0000010,\ -.0000099,\ .0000995,\ -.0009950,\ .0099499,\ -.0994987,\ .9949875]^{\mathsf{H}}. \tag{32}$$

All $\mathbf{x}_j + \mathcal{K}\mathit{ernel}(A_\theta)$ for $j = 1, 2, 3$ are nearly identical in the affine Grassmannian $\mathcal{A}_1(\mathbb{C}^9)$, and each contains a particular vector $\hat{\mathbf{x}}_j$ that is an accurate approximation to the exact solution $\mathbf{x}_*$, as shown in the bottom part of Table 1. The errors $\frac{\|\hat{\mathbf{x}}_j - \mathbf{x}_*\|_2}{\|\mathbf{x}_*\|_2}$ are all within the bound 8.28×10^{-7} predicted by (31).

TABLE 1
For $A\mathbf{x} = \tilde{\mathbf{b}}$ in Example 6, numerical solutions $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ by MATLAB "\", Tikhonov regularization, and truncated SVD, respectively, in comparison with the exact solution $\mathbf{x}_*$. Also shown are the accuracies of $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ as a component of the general numerical solution.

Solution type	Numerical (single-vector) solution with incorrect digits crossed out	Error $\frac{\Vert\mathbf{x}_j - \mathbf{x}_*\Vert_2}{\Vert\mathbf{x}_*\Vert_2}$
8 digits of $\mathbf{x}_*$	.3333333 .6666667 1.0000000 1.3333333 1.6666667 2.0000000 2.3333333 2.6666667 3.0000000	
MATLAB "\" $\mathbf{x}_1$	.3333333 .666666~~5~~ 1.00000~~14~~ 1.3333~~187~~ 1.666~~8129~~ 1.998~~5371~~ 2.3~~479633~~ 2.~~5203667~~ ~~4.4629993~~	0.2612916
Tikhonov $\mathbf{x}_2$	.3333333 .666667~~0~~ 0.99999~~71~~ 1.3333~~607~~ 1.666~~3934~~ 2.00~~27277~~ 2.3~~060572~~ 2.~~9394238~~ ~~0.2724303~~	0.4871437
trunc. SVD $\mathbf{x}_3$	.333333~~5~~ 0.666666~~9~~ 0.99999~~67~~ 1.3333~~603~~ 1.666~~3938~~ 2.00~~27270~~ 2.3~~060613~~ 2.~~9393935~~ ~~0.2727296~~	0.4870902
	Particular $\hat{\mathbf{x}}_j = \mathbf{x}_j + t_j\tilde{\mathbf{v}} \in \mathbf{x}_j + \mathcal{K}\mathit{ernel}(A_\theta)$ nearest to $\mathbf{x}_*$ with $\tilde{\mathbf{v}}$ in (32)	Error $\frac{\Vert\hat{\mathbf{x}}_j - \mathbf{x}_*\Vert_2}{\Vert\mathbf{x}_*\Vert_2}$
$\mathbf{x}_1 + \mathcal{K}\mathit{ernel}(A_\theta)$	$\hat{\mathbf{x}}_1 = \mathbf{x}_1 + t_1\tilde{\mathbf{v}} \approx \mathbf{x}_*$ with $t_1 = -1.4703701$	8.8×10^{-8}
$\mathbf{x}_2 + \mathcal{K}\mathit{ernel}(A_\theta)$	$\hat{\mathbf{x}}_2 = \mathbf{x}_2 + t_2\tilde{\mathbf{v}} \approx \mathbf{x}_*$ with $t_2 = 2.7413113$	1.5×10^{-7}
$\mathbf{x}_3 + \mathcal{K}\mathit{ernel}(A_\theta)$	$\hat{\mathbf{x}}_3 = \mathbf{x}_3 + t_3\tilde{\mathbf{v}} \approx \mathbf{x}_*$ with $t_3 = 2.7410104$	1.9×10^{-7}

The linear system in Example 6 is nonsingular in theory but practically underdetermined in numerical computation. Suppose an additional piece of information becomes available, say the remainder $\rho = 3$. One can impose such a constraint on the general numerical solution $\{\tilde{\mathbf{u}} + t\tilde{\mathbf{v}} \mid t \in \mathbb{C}\}$ at the trailing component as $0.2727296 + .9949875\,t = 3$, obtaining $t = 2.7410097$ corresponding to a numerical solution with a relative error 1.79×10^{-7} in the same order of the data.

Appendix A. Lemmas.

LEMMA 11. *Let $A, \tilde{A} \in \mathbb{C}^{m\times n}$ with $\Delta A = \tilde{A} - A$. Assume $\sigma_r(A) > \sigma_{r+1}(A)$.*

(i) (Wedin) *If $\|\Delta A\|_2 < \frac{1}{2}\big(\sigma_r(A) - \sigma_{r+1}(A)\big)$, then*

$$\begin{aligned} &\mathit{dist}\Big(\mathit{Kernel}(A_\theta),\, \mathit{Kernel}(\tilde{A}_\theta) \Big) \\ &\le \frac{\sigma_1(A)}{\sigma_r(A)} \frac{1}{1 - \frac{\sigma_{r+1}(A) + \|\Delta A\|_2}{\sigma_r(A)}} \frac{\|\Delta A\|_2}{\|A\|_2} \\ &\le \frac{\sigma_1(A)}{\sigma_r(A)} \frac{2}{1 - \frac{\sigma_{r+1}(A)}{\sigma_r(A)}} \frac{\|\Delta A\|_2}{\|A\|_2} \end{aligned} \tag{33}$$

for any $\theta \in \big(\sigma_{r+1}(A),\, \sigma_r(A)\big) \cap \big(\sigma_{r+1}(\tilde{A}),\, \sigma_r(\tilde{A})\big) \neq \emptyset$.

(ii) *If $\mathit{rank}(A) = \mathit{rank}\big(\tilde{A}\big) = r$ and $\|\Delta A\| < \sigma_r(A)$, then*

$$\mathit{dist}\Big(\mathit{Kernel}(A),\, \mathit{Kernel}(\tilde{A}) \Big) \le \frac{\sigma_1(A)}{\sigma_r(A)} \frac{\|\Delta A\|_2}{\|A\|_2}. \tag{34}$$

Proof. Assertion (i) is established by Wedin [34] (also see [32, Theorem 4.4] and [10, Theorem 3.3]). To prove (ii), let the SVDs of A and $\tilde{A}$ be

$$A = [U_1, U_2] \begin{bmatrix} \Sigma_1 & \\ & O \end{bmatrix} [V_1, V_2]^{\mathsf{H}} \quad \text{and} \quad \tilde{A} = [\tilde{U}_1, \tilde{U}_2] \begin{bmatrix} \tilde{\Sigma}_1 & \\ & O \end{bmatrix} [\tilde{V}_1, \tilde{V}_2]^{\mathsf{H}},$$

respectively, where $\Sigma_1, \tilde{\Sigma}_1 \in \mathbb{C}^{r\times r}$. Then,

$$-\tilde{V}_2^{\mathsf{H}} \Delta A^{\mathsf{H}} = \tilde{V}_2^{\mathsf{H}} \tilde{A}^{\mathsf{H}} - \tilde{V}_2^{\mathsf{H}} \Delta A^{\mathsf{H}} = \tilde{V}_2^{\mathsf{H}} A^{\mathsf{H}} = (\tilde{V}_2^{\mathsf{H}} V_1)(\Sigma_1^{\mathsf{H}} U_1^{\mathsf{H}}),$$

and thus,

$$\mathit{dist}\Big(\mathit{Kernel}(A),\, \mathit{Kernel}(\tilde{A}) \Big) = \big\|\tilde{V}_2^{\mathsf{H}} V_1\big\|_2 \le \frac{\|\Delta A\|_2}{\sigma_r(A)} \le \frac{\sigma_1(A)}{\sigma_r(A)} \frac{\|\Delta A\|_2}{\|A\|_2}. \qquad \square$$

LEMMA 12. *Let $\mathcal{U}$ be a subspace of $\mathbb{C}^n$, and let U be a matrix whose columns form an orthonormal basis for $\mathcal{U}$. For every subspace $\mathcal{V}$ of $\mathbb{C}^n$ of the same dimension as $\mathcal{U}$ with $\mathit{dist}(\mathcal{U}, \mathcal{V}) < 1$, there is a matrix V whose columns form a basis for $\mathcal{V}$ such that $\|U - V\|_2 \le \mathit{dist}(\mathcal{U}, \mathcal{V})$.*

Proof. Let G be any matrix whose columns form an orthonormal basis for $\mathcal{V}$, and let $[G, \hat{G}]$ be a unitary matrix. Then, for any unit vector $\mathbf{x} \in \mathbb{C}^n$,

$$1 = \big\|[G, \hat{G}]^{\mathsf{H}} U \mathbf{x}\big\|_2^2 = \|(G^{\mathsf{H}} U)\mathbf{x}\|_2^2 + \|(\hat{G}^{\mathsf{H}} U)\mathbf{x}\|_2^2 \le \|(G^{\mathsf{H}} U)\mathbf{x}\|_2^2 + \mathit{dist}(\mathcal{U}, \mathcal{V})^2$$

leading to $\|(G^{\mathsf{H}} U)\mathbf{x}\|_2^2 \ge 1 - \mathit{dist}(\mathcal{U}, \mathcal{V})^2 > 0$, implying $G^{\mathsf{H}} U$ is invertible so that columns of $V = G(G^{\mathsf{H}} U)$ form a basis for $\mathcal{V}$, and $\|U - V\|_2 = \big\|(U U^{\mathsf{H}} - G G^{\mathsf{H}}) U\big\|_2$ that is less than or equal to $\mathit{dist}(\mathcal{U}, \mathcal{V})$. $\square$

LEMMA 13. *Let $A \in \mathbb{C}^{m\times n}$ with $\sigma_r(A) > \theta > \sigma_{r+1}(A)$.*

(i) *For every* $\mu \in [\sigma_r(A), \sigma_1(A)]$, *let* $N \in \mathbb{C}^{n\times(n-r)}$ *be a matrix whose columns form an orthonormal basis for* $\mathcal{K}ernel(A_\theta)$. *Then*

$$\left\| \begin{bmatrix} \mu N^{\mathsf{H}} \\ A \end{bmatrix} \right\|_2 = \max\left\{\sigma_1(A),\ \sqrt{\mu^2+\sigma_{r+1}(A)^2}\right\} \in \left[\|A\|_2,\ \sqrt{2}\,\|A\|_2\right) \tag{35}$$

$$\left\| \begin{bmatrix} \mu N^{\mathsf{H}} \\ A \end{bmatrix}^\dagger \right\|_2 = \max\left\{\frac{1}{\sigma_r(A)},\ \frac{1}{\sqrt{\mu^2+\eta^2}}\right\} \ \le\ \|A_\theta^\dagger\|_2, \tag{36}$$

where $\eta = \sigma_n(A)$ *if* $m \ge n$ *or* $\eta = 0$ *otherwise.*

(ii) *Assume columns of* $N \in \mathbb{C}^{n\times(n-r)}$ *span* $\mathcal{K}ernel(A_\theta)$. *For any* $\mu > 0$, *let* $\mathbf{b} \in \mathbb{C}^m$, *and let* $\mathbf{x}_*$ *be the least squares solution of the linear system*

$$\begin{bmatrix} \mu N^{\mathsf{H}} \\ A \end{bmatrix} \mathbf{x} \ = \ \begin{bmatrix} \mathbf{0} \\ \mathbf{b} \end{bmatrix}. \tag{37}$$

Then $\mathbf{x}_* = A^\dagger \mathbf{b}$ *if* $\mathit{rank}(A) = r$ *or* $A_\theta\,\mathbf{x}_* = \mathbf{b}_\theta$ *if* $\mathit{rank}(A) > r$, *where* $\mathbf{b}_\theta = A_\theta\, A_\theta^\dagger\, \mathbf{b}$ *is the orthogonal projection of* $\mathbf{b}$ *onto* $\mathcal{R}ange(A_\theta)$,

$$\left\| \begin{bmatrix} \mu N^{\mathsf{H}} \\ A \end{bmatrix} \mathbf{x}_* - \begin{bmatrix} \mathbf{0} \\ \mathbf{b} \end{bmatrix} \right\|_2 \ \le\ \|\mathbf{b}-\mathbf{b}_\theta\|_2, \tag{38}$$

$$\mathbf{x}_* - T\,T^{\mathsf{H}}\,\mathbf{x}_* \ = \ A_\theta^\dagger\,\mathbf{b}_\theta, \tag{39}$$

for any $T \in \mathbb{C}^{n\times(n-r)}$ *with* $\mathcal{R}ange(T) = \mathcal{K}ernel(A_\theta)$ *and* $T^{\mathsf{H}}T = I$.

Proof. For the case of $m \ge n$, we can write A in its singular value expansion $A = \sigma_1\,\mathbf{u}_1\,\mathbf{v}_1^{\mathsf{H}} + \cdots + \sigma_n\,\mathbf{u}_n\,\mathbf{v}_n^{\mathsf{H}}$ and $N = [\mathbf{v}_{r+1},\ldots,\mathbf{v}_n]\,G$ where $\mathbf{u}_1,\ldots,\mathbf{u}_m$ and $\mathbf{v}_1,\ldots,\mathbf{v}_n$ are left and right singular vectors, respectively, with a unitary matrix $G \in \mathbb{C}^{(n-r)\times(n-r)}$. Write $\mathbf{x} = x_1\,\mathbf{v}_1 + \cdots + x_n\,\mathbf{v}_n$. Then,

$$\begin{aligned} \left\| \begin{bmatrix} \mu N^{\mathsf{H}} \\ A \end{bmatrix} \mathbf{x} \right\|_2^2 \ &= \ \left\| \begin{bmatrix} G^{\mathsf{H}} & \\ & I \end{bmatrix} \begin{bmatrix} \mu\,[\mathbf{v}_{r+1},\cdots,\mathbf{v}_n]^{\mathsf{H}} \\ A \end{bmatrix} \mathbf{x} \right\|_2^2 \\ &= \sigma_1^2\,|x_1|^2 + \cdots + \sigma_r\,|x_r|^2 + (\mu^2+\sigma_{r+1}^2)\,|x_{r+1}|^2 + \cdots + (\mu^2+\sigma_n^2)\,|x_n|^2, \end{aligned}$$

whose extrema subject to $\|\mathbf{x}\|_2 = 1$ are $\max\{\sigma_1^2,\ \mu^2+\sigma_{r+1}^2\}$ and $\min\{\sigma_r^2,\ \mu^2+\sigma_n^2\}$, leading to (35) and (36) in assertion (i). The case $m < n$ is similar.

To prove (ii), write the SVD $A = U_1\,\Sigma_1\,V_1^{\mathsf{H}} + U_2\,\Sigma_2\,V_2^{\mathsf{H}}$, where $\Sigma_1 \in \mathbb{C}^{r\times r}$ and $\Sigma_2 \in \mathbb{C}^{(m-r)\times(n-r)}$. Then $\mathbf{x}_*$ is the solution of the normal equation $\mu^2 N N^{\mathsf{H}}\mathbf{x}_* + A^{\mathsf{H}}A\mathbf{x}_* - A^{\mathsf{H}}\mathbf{b} = \mathbf{0}$. Namely, we have an orthogonal decomposition

$$\begin{aligned} &(V_1\,\Sigma_1^{\mathsf{H}}\,\Sigma_1\,V_1^{\mathsf{H}}\,\mathbf{x}_* - V_1\,\Sigma_1^{\mathsf{H}}\,U_1^{\mathsf{H}}\,\mathbf{b}) + (V_2\,\Sigma_2^{\mathsf{H}}\,\Sigma_2\,V_2^{\mathsf{H}}\,\mathbf{x}_* - V_2\,\Sigma_2^{\mathsf{H}}\,U_2^{\mathsf{H}}\,\mathbf{b} + \mu^2\,N\,N^{\mathsf{H}}\,\mathbf{x}_*) \\ &\quad = \ \mathbf{0}, \end{aligned} \tag{40}$$

implying $V_1\,\Sigma_1^{\mathsf{H}}\,\Sigma_1\,V_1^{\mathsf{H}}\,\mathbf{x}_* - V_1\,\Sigma_1^{\mathsf{H}}\,U_1^{\mathsf{H}}\,\mathbf{b} \ = \ \mathbf{0}$ and thus $V_1^{\mathsf{H}}\,\mathbf{x}_* \ = \ \Sigma_1^{-1}\,U_1^{\mathsf{H}}\,\mathbf{b}$. Since $\mathbf{x}_* = V_1\,V_1^{\mathsf{H}}\,\mathbf{x}_* + V_2\,V_2^{\mathsf{H}}\,\mathbf{x}_*$, we have $A_\theta\,\mathbf{x}_* = (U_1\,\Sigma_1\,V_1^{\mathsf{H}})\,(V_1\,V_1^{\mathsf{H}}\,\mathbf{x}_*) = U_1\,U_1^{\mathsf{H}}\,\mathbf{b} = \mathbf{b}_\theta$. Namely $\mathbf{x}_*$ is a particular solutions of the system $A_\theta\,\mathbf{x} = \mathbf{b}_\theta$. Also,

$$A_\theta^\dagger\,\mathbf{b}_\theta \ = \ (V_1\,\Sigma_1^{-1}\,U_1^{\mathsf{H}})\,(U_1\,U_1^{\mathsf{H}}\,\mathbf{b}) \ = \ (V_1\,\Sigma_1^{-1}\,U_1^{\mathsf{H}})\,\mathbf{b} \ = \ V_1\,V_1^{\mathsf{H}}\mathbf{x}_* \ = \ (I - T\,T^{\mathsf{H}})\,\mathbf{x}_*.$$

Furthermore,

$$\left\| \begin{bmatrix} \mu N^{\mathsf{H}} \\ A \end{bmatrix} \mathbf{x}_* - \begin{bmatrix} \mathbf{0} \\ \mathbf{b} \end{bmatrix} \right\|_2 \ \le \ \left\| \begin{bmatrix} \mu N^{\mathsf{H}} \\ A \end{bmatrix} V_1\,V_1^{\mathsf{H}}\,\mathbf{x}_* - \begin{bmatrix} \mathbf{0} \\ \mathbf{b} \end{bmatrix} \right\|_2 \ = \ \|\mathbf{b}-\mathbf{b}_\theta\|_2.$$

Finally, if $\mathit{rank}(A) = r$, then $A_\theta = A$ so $\Sigma_2 = O$ in (40), implying $N^{\mathsf{H}}\,\mathbf{x}_* = \mathbf{0}$. Consequently $\mathbf{x}_* = A^\dagger\,\mathbf{b}$ from (39). ∎

The following lemma is a variation of Theorem 5.1 in [35] by Wedin and its extension in Theorem 3.4 in [10] by Hansen.

LEMMA 14 (Wedin [35], Hansen [10]). *Let $A \in \mathbb{C}^{m\times n}$ and $\mathbf{b} \in \mathbb{C}^m$. Assume, for a $\theta > 0$, we have $\mathit{rank}_\theta(A) = r$ and $\|\mathbf{b} - A_\theta A_\theta^\dagger \mathbf{b}\|_2 < \theta$. There is a constant*

$$\zeta \;=\; \|A_\theta^\dagger \mathbf{b}\|_2 + \frac{1 + \|A_\theta^\dagger \mathbf{b}\|_2}{1 - \|A_\theta^\dagger\|_2 \, \|A - A_\theta\|_2} \tag{41}$$

such that, for any $\tilde{A} = A + \Delta A \in \mathbb{C}^{m\times n}$ and $\tilde{\mathbf{b}} = \mathbf{b} + \Delta\mathbf{b} \in \mathbb{C}^m$ with

$$\|\Delta A\|_2 \;<\; \min\left\{\tfrac{1}{2}\big(\sigma_r(A) - \sigma_{r+1}(A)\big),\ \sigma_r(A) - \theta,\ \theta - \sigma_{r+1}(A)\right\}, \tag{42}$$

the following inequality holds:

$$\left\|A_\theta^\dagger \mathbf{b} - \tilde{A}_\theta^\dagger \tilde{\mathbf{b}}\right\|_2 \;\leq\; \frac{\sigma_1(A)}{\sigma_r(A)}\left(\zeta\,\frac{\|\Delta A\|_2}{\|A\|_2} + \frac{\|\Delta \mathbf{b}\|_2}{\|A\|_2}\right) + O(\|(\Delta A,\, \Delta\mathbf{b})\|^2). \tag{43}$$

As a special case, further assume $\mathit{rank}(A) = r$ and $\mathbf{b} \in \mathcal{Range}(A)$. Then

$$\left\|A^\dagger \mathbf{b} - \tilde{A}_\theta^\dagger \tilde{\mathbf{b}}\right\|_2 \;\leq\; \frac{\sigma_1(A)}{\sigma_r(A)}\,\frac{1}{1 - \frac{\|\Delta A\|_2}{\sigma_r(A)}}\left(2\,\|A^\dagger \mathbf{b}\|_2\,\frac{\|\Delta A\|_2}{\|A\|_2} + \frac{\|\Delta \mathbf{b}\|_2}{\|A\|_2}\right). \tag{44}$$

Proof. The assumption $\mathit{rank}_\theta(A) = r$ implies $\sigma_{r+1}(A) < \theta < \sigma_r(A)$, and thus $\sigma_{r+1}(\tilde{A}) < \theta < \sigma_r(\tilde{A})$, following (42) so that $\mathit{rank}_\theta\left(\tilde{A}\right) = r$ as well. Then it is straightforward to verify (43) from the inequality (26a) in [10] using

$$\|A\,(A_\theta^\dagger \mathbf{b}) - \mathbf{b}\|_2 \;=\; \|A_\theta A_\theta^\dagger \mathbf{b} - \mathbf{b}\|_2 \;<\; \theta \;<\; \sigma_r(A).$$

The inequality (44) follows from [10, inequality (27a)] and $\mathbf{b} \in \mathcal{Range}(A)$. ∎

LEMMA 15. *At any $(A, \mathbf{b}) \in \mathbb{C}^{m\times n} \times \mathbb{C}^n$ and $\theta > 0$ within which $\mathit{sol}_\theta(A, \mathbf{b})$ is well-defined, there is a $\delta > 0$ such that $\mathit{sol}_\theta(A + \Delta A, \mathbf{b} + \Delta\mathbf{b})$ is well-defined with the same dimension as $\mathit{sol}_\theta(A, \mathbf{b})$ if $\|(\Delta A, \Delta\mathbf{b})\| < \delta$.*

Proof. Write $\tilde{A} = A + \Delta A$ and $\tilde{\mathbf{b}} = \mathbf{b} + \Delta\mathbf{b}$. Since $r = \mathit{rank}_\theta(A)$ is well-defined, we have $\sigma_{r+1} < \theta < \sigma_r(A)$. Thus $\|\Delta A\|_2 < \min\{\sigma_r - \theta, \theta - \sigma_{r+1}\}$ ensures $\mathit{rank}_\theta\left(\tilde{A}\right) = r$. Let $P = I - A_\theta A_\theta^\dagger$ and $\tilde{P} = I - \tilde{A}_\theta \tilde{A}_\theta^\dagger$. Then

$$\tilde{A} - \tilde{A}_\theta \;=\; \tilde{P}\,\tilde{A} \;=\; \tilde{P}\,\Delta A + (\tilde{P} - P)\,A + (A - A_\theta),$$

and by (33),

$$\|P - \tilde{P}\|_2 \;=\; \mathit{dist}\left(\mathcal{Range}(A_\theta),\, \mathcal{Range}(\tilde{A}_\theta)\right) \;\leq\; \eta\,\frac{\|\Delta A\|_2}{\|A\|_2},$$

where, assuming $\|\Delta A\|_2 \leq \frac{1}{2}(\sigma_r(A) - \sigma_{r+1}(A))$,

$$\eta \;=\; \frac{\sigma_1(A)}{\sigma_r(A)}\,\frac{2}{1 - \frac{\sigma_{r+1}(A)}{\sigma_r(A)}} \;=\; \frac{2\,\|A_\theta\|_2\,\|A_\theta^\dagger\|_2}{1 - \|A_\theta^\dagger\|_2\,\|A - A_\theta\|_2},$$

implying

$$\|A - A_\theta\|_2 - (\eta + 1)\,\|\Delta A\|_2 \;\leq\; \|\tilde{A} - \tilde{A}_\theta\|_2 \;\leq\; \|A - A_\theta\|_2 + (\eta + 1)\,\|\Delta A\|_2,$$

and similarly

$$\begin{aligned}\|\mathbf{b}-\mathbf{b}_\theta\|_2 - \eta\,\frac{\|\Delta A\|_2}{\|A\|_2}\,\|\mathbf{b}\|_2 - \|\Delta\mathbf{b}\|_2 \quad &\le \quad \|\tilde{\mathbf{b}}-\tilde{\mathbf{b}}_\theta\|_2 \\ &\le \quad \|\mathbf{b}-\mathbf{b}_\theta\|_2 + \eta\,\frac{\|\Delta A\|_2}{\|A\|_2}\,\|\mathbf{b}\|_2 + \|\Delta\mathbf{b}\|_2,\end{aligned}$$

where $\mathbf{b}_\theta = \mathbf{b} - P\,\mathbf{b}$ and $\tilde{\mathbf{b}}_\theta = \tilde{\mathbf{b}} - \tilde{P}\,\tilde{\mathbf{b}}$. If $\mathit{sol}_\theta(A,\mathbf{b})$ is empty, then $\|(A,\mathbf{b})-(A_\theta,\mathbf{b}_\theta)\| > \theta$, and thus $\|(\tilde{A},\tilde{\mathbf{b}})-(\tilde{A}_\theta,\tilde{\mathbf{b}}_\theta)\| > \theta$ when $\|(\Delta A,\,\Delta\mathbf{b})\|$ is sufficiently small so that $\mathit{sol}_\theta(\tilde{A},\tilde{\mathbf{b}}) = \emptyset$ as well. When $\|(A,\mathbf{b})-(A_\theta,\mathbf{b}_\theta)\| < \theta$ and $\|(\Delta A,\,\Delta\mathbf{b})\|$ is sufficiently small, we also have $\|(\tilde{A},\tilde{\mathbf{b}})-(\tilde{A}_\theta,\tilde{\mathbf{b}}_\theta)\| < \theta$ and $\sigma_{r+1}(\tilde{A}) < \theta < \sigma_r(\tilde{A})$ so that $\mathit{sol}_\theta(\tilde{A},\tilde{\mathbf{b}}) = \tilde{A}_\theta^\dagger\,\tilde{\mathbf{b}}_\theta + \mathcal{K}ernel(\tilde{A}_\theta)$ has the identical dimension $n-r$ as $\mathit{sol}_\theta(A,\mathbf{b})$. $\square$

Appendix B. Proofs of theorems and corollaries.

Proof of Theorem 3. Let $N \in \mathbb{C}^{n\times(n-r)}$, whose columns form an orthonormal basis for $\mathcal{K}ernel(A)$. By Lemma 12, there is an $\tilde{N} \in \mathbb{C}^{n\times(n-r)}$, whose columns form a basis for $\mathcal{K}ernel(\tilde{A})$ such that $\|N-\tilde{N}\|_2 \le \mathit{dist}\Big(\mathcal{K}ernel(A),\ \mathcal{K}ernel(\tilde{A})\Big)$. For $\zeta = \sigma_r(A)$, denote $B = \left[\begin{smallmatrix}\zeta N^{\mathsf{H}}\\ A\end{smallmatrix}\right]$ and $\tilde{B} = \left[\begin{smallmatrix}\zeta \tilde{N}^{\mathsf{H}}\\ \tilde{A}\end{smallmatrix}\right]$. Then $A^\dagger\,\mathbf{b} = B^\dagger\left[\begin{smallmatrix}\mathbf{0}\\ \mathbf{b}\end{smallmatrix}\right]$ and $\tilde{A}^\dagger\,\tilde{\mathbf{b}} = \tilde{B}^\dagger\left[\begin{smallmatrix}\mathbf{0}\\ \tilde{\mathbf{b}}\end{smallmatrix}\right]$ by Lemma 13(ii). By $\|B-\tilde{B}\|_2 \le \sqrt{2}\,\|\Delta A\|_2$ from Lemma 11(ii), [24, Theorem 1.4.6, p. 30], and Lemma 13(i),

$$\begin{aligned}\left\|B^\dagger\left[\begin{smallmatrix}\mathbf{0}\\ \mathbf{b}\end{smallmatrix}\right]-\tilde{B}^\dagger\left[\begin{smallmatrix}\mathbf{0}\\ \tilde{\mathbf{b}}\end{smallmatrix}\right]\right\|_2 \;&\le\; \frac{\sigma_1(B)}{\sigma_n(B)}\,\frac{1}{1-\frac{\|B-\tilde{B}\|_2}{\sigma_n(B)}}\left(\|\mathbf{x}_*\|_2\,\frac{\|B-\tilde{B}\|_2}{\|B\|_2}+\frac{\|\Delta\mathbf{b}\|_2}{\|B\|_2}\right) \\ &\le\; \frac{\sigma_1(A)}{\sigma_r(A)}\,\frac{1}{1-\frac{\sqrt{2}\,\|\Delta A\|_2}{\sigma_r(A)}}\,\frac{\sqrt{2}\,\|\mathbf{x}_*\|_2\,\|\Delta A\|_2+\|\Delta\mathbf{b}\|_2}{\|A\|_2}, \qquad (45)\end{aligned}$$

leading to (17). $\square$

Proof Theorem 6. Assertion (i) is true because $A_\theta = A$ and $\mathbf{b}_\theta = \mathbf{b}$ for $\theta \in (0,\sigma_r(A))$. If $\mathbf{b} \in \mathcal{R}ange(A)$, then

$$\mathit{sol}_\theta(A,\mathbf{b}) \;=\; \mathit{sol}(A,\mathbf{b}) \;=\; A^\dagger\mathbf{b} + \mathcal{K}ernel(A).$$

Otherwise $\mathit{sol}_\theta(A,\mathbf{b}) = \mathit{sol}(A,\mathbf{b}) = \emptyset$ if $\theta < \min\big\{\sigma_r(A),\ \|A\,A^\dagger\,\mathbf{b}-\mathbf{b}\|_2\big\}$. Assertion (ii) directly follows from Lemmas 14 and 15 with

$$\xi \quad = \quad \|A_\theta\|_2\,\|A_\theta^\dagger\|_2\,\frac{\sqrt{\zeta^2+1}}{\|A\|_2-\|A\|_2\,\|A_\theta^\dagger\|_2\,\|A-A_\theta\|_2} + \varepsilon$$

for any $\varepsilon > 0$.

We now prove assertion (iii)(a). Let $\tilde{\mathbf{b}}_\theta = \tilde{A}_\theta\,\tilde{A}_\theta^\dagger\,\tilde{\mathbf{b}}$, $P = I - A\,A^\dagger$, and $\tilde{P} = I - \tilde{A}_\theta\,\tilde{A}_\theta^\dagger$. From $\|\tilde{A}-\tilde{A}_\theta\|_2 = \min_{\mathit{rank}(\,B\,)=r}\|\tilde{A}-B\|_2 \le \|\Delta A\|_2$, we have

$$\begin{aligned}\|\tilde{\mathbf{b}}-\tilde{\mathbf{b}}_\theta\|_2 \;&=\; \|\tilde{P}\,\tilde{\mathbf{b}}\|_2 \;=\; \|\tilde{P}\,\tilde{\mathbf{b}}-P\,\mathbf{b}\|_2 \;\le\; \|\tilde{P}\|_2\,\|\tilde{\mathbf{b}}-\mathbf{b}\|_2+\|\tilde{P}-P\|_2\,\|\mathbf{b}\|_2 \\ &\le\; \|\Delta\mathbf{b}\|_2 + \mathit{dist}\Big(\mathcal{R}ange(\tilde{A}_\theta),\ \mathcal{R}ange(A)\Big)\,\|\mathbf{b}\|_2 \\ \text{(by (33))} \qquad &\le\; \|\Delta\mathbf{b}\|_2 + \frac{\sigma_1(A)}{\sigma_r(A)}\,\frac{2\,\|\mathbf{b}\|_2}{\|A\|_2}\,\|\Delta A\|_2 \\ &\le\; \sqrt{4\,\|A^\dagger\|_2^2\,\|\mathbf{b}\|_2^2+1}\;\|(\Delta A,\,\Delta\mathbf{b})\| \;=\; \sqrt{\omega^2-1}\;\|(\Delta A,\,\Delta\mathbf{b})\|\end{aligned}$$

and

$$\begin{aligned}
\big\|(\tilde{A},\,\tilde{\mathbf{b}})-(\tilde{A}_\theta,\,\tilde{\mathbf{b}}_\theta)\big\| &\leq \sqrt{\|\Delta A\|_2^2+(\omega^2-1)\,(\|\Delta A\|_2^2+\|\Delta\mathbf{b}\|_2^2)}\\
&\leq \omega\,\big\|(\Delta A,\,\Delta\mathbf{b})\big\| \;<\; \sigma_r(A)-\big\|(\Delta A,\,\Delta\mathbf{b})\big\|. \qquad (46)
\end{aligned}$$

Then, for any θ satisfying (22),

$$\sigma_{r+1}(\tilde{A}) \;\leq\; \big\|(\Delta A,\,\Delta\mathbf{b})\big\| \;<\; \theta \;<\; \sigma_r(A)-\big\|(\Delta A,\,\Delta\mathbf{b})\big\| \;\leq\; \sigma_r(\tilde{A})$$

so $\mathit{rank}_\theta\left(\tilde{A}\right) = r$, $\big\|(\tilde{A},\,\tilde{\mathbf{b}})-(\tilde{A}_\theta,\,\tilde{\mathbf{b}}_\theta)\big\| < \theta$, and thus $\mathit{sol}_\theta(\tilde{A},\,\tilde{\mathbf{b}})$ is of the same dimension as $\mathit{sol}(A,\,\mathbf{b})$. Since $\mathit{sol}_\theta(\tilde{A},\,\tilde{\mathbf{b}}) = \mathit{sol}(\tilde{A}_\theta,\tilde{\mathbf{b}}_\theta)$, the backward error of $\mathit{sol}_\theta(\tilde{A},\,\tilde{\mathbf{b}})$ is bounded above by $\omega\,\big\|(\Delta A,\,\Delta\mathbf{b})\big\|$ from (46). Thus (23) follows from (33) in Lemma 11 and (44) in Lemma 14, leading to assertion (iii).

We now prove assertion (iii)(b). If $\mathit{sol}(A,\mathbf{b})$ is empty, then $\mathit{sol}_\theta(A,\mathbf{b}) = \emptyset$ whenever $\theta < \min\{\sigma_r(A),\,\|\mathbf{b}-A\,A^\dagger\,\mathbf{b}\|_2\}$. By Lemma 15, there is a $\delta_1 > 0$ such that $\mathit{sol}_\theta(\tilde{A},\,\tilde{\mathbf{b}}) = \emptyset$ for every $(\tilde{A},\,\tilde{\mathbf{b}})$ with $\big\|(\tilde{A},\,\tilde{\mathbf{b}})-(A,\,\mathbf{b})\big\| < \delta_1$. Thus $\mathit{sol}_\theta(\tilde{A},\,\tilde{\mathbf{b}}) = \mathit{sol}(A,\mathbf{b})$ with both backward and forward errors as zero. □

Proof of Theorem 8. Let $\tilde{A} = \tilde{U}_1\,\tilde{\Sigma}_1\,\tilde{V}_1^{\mathsf{H}}+\tilde{U}_2\,\tilde{\Sigma}_2\,\tilde{V}_2^{\mathsf{H}}$ be the SVD where $\tilde{\Sigma}_1$ is $r\times r$ with $r = \mathit{rank}\,(A)$. Then $\tilde{\mathbf{x}}$ is a solution to $\tilde{A}\,\mathbf{x} = \tilde{\mathbf{b}}$, implying $\tilde{U}_1\,\tilde{\Sigma}_1\,\tilde{V}_1^{\mathsf{H}}\,\tilde{\mathbf{x}}_1 = \tilde{U}_1\,\tilde{U}_1^{\mathsf{H}}\,\tilde{\mathbf{b}}$ and $\tilde{U}_2\,\tilde{\Sigma}_2\,\tilde{V}_2^{\mathsf{H}}\,\tilde{\mathbf{x}}_2 = \tilde{U}_2\,\tilde{U}_2^{\mathsf{H}}\,\tilde{\mathbf{b}}$ where $\tilde{\mathbf{x}}_1 = \tilde{V}_1\,\tilde{V}_1^{\mathsf{H}}\,\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}_2 = \tilde{V}_2\,\tilde{V}_2^{\mathsf{H}}\,\tilde{\mathbf{x}}$. Then $\tilde{\mathbf{x}} = \tilde{\mathbf{x}}_1+\tilde{\mathbf{x}}_2$ with $\tilde{\mathbf{x}}_1 = \tilde{A}_\theta^\dagger\,\tilde{\mathbf{b}}$ for any θ between $\sigma_{r+1}(\tilde{A})$ and $\sigma_r(A)-\|\Delta A\|_2$. By Lemma 14 with $\hat{\mathbf{x}} = A^\dagger\,\mathbf{b}$, we have

$$\big\|\tilde{\mathbf{x}}_1-\hat{\mathbf{x}}\big\|_2 \;\leq\; \frac{\sigma_1(A)}{\sigma_r(A)}\,\frac{\|\hat{\mathbf{x}}\|_2}{1-\frac{\|\Delta A\|_2}{\sigma_r(A)}}\,\left(2\,\frac{\|\Delta A\|}{\|A\|_2}+\frac{\|\Delta\mathbf{b}\|_2}{\|\mathbf{b}\|_2}\right).$$

Let columns of N form an orthonormal basis for $\mathit{Kernel}(A)$. Since $\tilde{\mathbf{x}}_2 \in \mathit{Kernel}(\tilde{A}_\theta)$,

$$\begin{aligned}
\|N\,N^{\mathsf{H}}\,\tilde{\mathbf{x}}_2-\tilde{\mathbf{x}}_2\|_2 &= \min_{\mathbf{u}\in\mathit{Kernel}(A)}\|\mathbf{u}-\tilde{\mathbf{x}}_2\|_2\\
&\leq \mathit{dist}\,(\,\mathit{Kernel}(A_\theta),\;\mathit{Kernel}(A)\,)\;\|\tilde{\mathbf{x}}_2\|_2\\
&\leq \frac{\sigma_1(A)}{\sigma_r(A)}\,\frac{1}{1-\frac{\|\Delta A\|_2}{\sigma_r(A)}}\,\frac{\|\Delta A\|_2}{\|A\|_2}\,\|\tilde{\mathbf{x}}_2\|_2 \qquad (47)
\end{aligned}$$

by Lemma 11, which, combined with $\|\Delta A\|_2 \leq .46\,\sigma_r(A) < (2\sqrt{3}-3)\,\sigma_r(A)$, implies $\mathit{dist}\,(\,\mathit{Kernel}(A_\theta),\;\mathit{Kernel}(A)\,) < \frac{\sqrt{3}}{2}$, and thus

$$\begin{aligned}
\|N\,N^{\mathsf{H}}\,\tilde{\mathbf{x}}_2\|_2 &= \|N^{\mathsf{H}}\,\tilde{\mathbf{x}}_2\|_2 = \|(N^{\mathsf{H}}\,\tilde{V}_2)\,\tilde{V}_2^{\mathsf{H}}\,\tilde{\mathbf{x}}_2\|_2\\
&\geq \sqrt{1-\mathit{dist}\,(\,\mathit{Kernel}(A_\theta),\;\mathit{Kernel}(A)\,)^2}\,\|\tilde{V}_2^{\mathsf{H}}\,\tilde{\mathbf{x}}_2\|_2 \;\geq\; \tfrac{1}{2}\,\|\tilde{\mathbf{x}}_2\|_2.
\end{aligned}$$

Let $\mathbf{x}_* = \hat{\mathbf{x}}+N\,N^{\mathsf{H}}\,\tilde{\mathbf{x}}_2$. Then $\mathbf{x}_*$ is a particular solution to $A\,\mathbf{x} = \mathbf{b}$ and $\|\mathbf{x}_*\|_2^2 = \|\hat{\mathbf{x}}\|_2^2+\|N\,N^{\mathsf{H}}\,\tilde{\mathbf{x}}_2\|_2^2$. We have

$$\begin{aligned}
\|\tilde{\mathbf{x}}-\mathbf{x}_*\| &\leq \|\tilde{\mathbf{x}}_1-\hat{\mathbf{x}}\|_2+\|\tilde{\mathbf{x}}_2-N\,N^{\mathsf{H}}\,\tilde{\mathbf{x}}_2\|_2\\
&\leq \frac{\sigma_1(A)}{\sigma_r(A)}\,\frac{1}{1-\frac{\|\Delta A\|_2}{\sigma_r(A)}}\,\left(\frac{\|\Delta A\|}{\|A\|_2}\big(2\,\|\hat{\mathbf{x}}\|_2+2\|N\,N^{\mathsf{H}}\,\tilde{\mathbf{x}}\|_2\big)+\|\mathbf{x}_*\|\,\frac{\|\Delta\mathbf{b}\|_2}{\|\mathbf{b}\|_2}\right)\\
&\leq \frac{\sigma_1(A)}{\sigma_r(A)}\,\frac{\|\mathbf{x}_*\|_2}{1-\frac{\|\Delta A\|_2}{\sigma_r(A)}}\,\left(2\sqrt{2}\,\frac{\|\Delta A\|_2}{\|A\|_2}+\frac{\|\Delta\mathbf{b}\|_2}{\|\mathbf{b}\|_2}\right),
\end{aligned}$$

leading to (24). For the case $\mathbf{b} = \mathbf{0}$, the bound (25) follows from (47)

$$\|\tilde{\mathbf{x}}_1\|_2 \quad \le \quad \frac{\|U_1^{\mathsf{H}}\,\tilde{\mathbf{b}}\|_2}{\sigma_r(\tilde{A})} \quad \le \quad \frac{\sigma_1(A)}{\sigma_r(A)}\,\frac{1}{1-\frac{\|\Delta A\|_2}{\sigma_r(A)}}\,\frac{\|\Delta\tilde{\mathbf{b}}\|_2}{\|A\|_2}. \qquad \square$$

Proof of Theorem 10. Let $U_1\,\Sigma_1\,V_1^{\mathsf{H}}+U_2\,\Sigma_2\,V_2^{\mathsf{H}}$ and $\tilde{U}_1\,\tilde{\Sigma}_1\,\tilde{V}_1^{\mathsf{H}}+\tilde{U}_2\,\tilde{\Sigma}_2\,\tilde{V}_2^{\mathsf{H}}$ be SVDs of A and $\tilde{A}$, respectively, where $\Sigma_1,\ \tilde{\Sigma}_1 \in \mathbb{C}^{r\times r}$. Denote $A_1 = U_1\,\Sigma_1\,V_1^{\mathsf{H}}$, $\tilde{A}_1 = \tilde{U}_1\,\tilde{\Sigma}_1\,\tilde{V}_1^{\mathsf{H}}$, $\mathbf{x}_1 = A_1^{\dagger}\,\mathbf{b}$, $\mathbf{x}_2 = \mathbf{x}_* - \mathbf{x}_1$, $\tilde{\mathbf{x}}_1 = \tilde{A}_1^{\dagger}\,\tilde{\mathbf{b}}$, and $\mathbf{r} = A\,\mathbf{x}_1 - \mathbf{b}$. Then, with $\mathbf{r} = U_2\,U_2^{\mathsf{H}}\,\mathbf{b} = U_2\,\Sigma_2\,V_2^{\mathsf{H}}\,\mathbf{x}_2$,

$$\begin{aligned}
\tilde{\mathbf{x}}_1 - \mathbf{x}_1 \quad &= \quad \tilde{A}_1^{\dagger}\,(\mathbf{b}+\Delta\mathbf{b}) - \mathbf{x}_1 \quad = \quad \tilde{A}_1^{\dagger}\,(A\,\mathbf{x}_1 - \mathbf{r} + \Delta\mathbf{b}) - \mathbf{x}_1 \\
&= \quad \tilde{A}_1^{\dagger}\,((\tilde{A}-\Delta A)\,\mathbf{x}_1 - \mathbf{r} + \Delta\mathbf{b}) - \mathbf{x}_1 \\
&= \quad \tilde{A}_1^{\dagger}\,(-\Delta A\,\mathbf{x}_1 - \mathbf{r} + \Delta\mathbf{b}) - (I - \tilde{A}_1^{\dagger}\,\tilde{A}_1)\,\mathbf{x}_1 \\
&= \quad \tilde{A}_1^{\dagger}\,(-\Delta A\,\mathbf{x}_1 + \Delta\mathbf{b}) - \tilde{V}_1\,\tilde{\Sigma}_1^{-1}\,\tilde{U}_1^{\mathsf{H}}\,U_2\,\Sigma_2\,V_2^{\mathsf{H}}\,\mathbf{x}_2 - \tilde{V}_2\,\tilde{V}_2^{\mathsf{H}}\,\mathbf{x}_1,
\end{aligned}$$

leading to

$$\begin{aligned}
&\|(\tilde{\mathbf{x}}_1 + \tilde{V}_2\,\tilde{V}_2^{\mathsf{H}}\,\mathbf{x}_1) - \mathbf{x}_1\|_2 \\
&\quad \le \quad \|\tilde{A}_1^{\dagger}\|_2\,(\|\Delta A\|_2\,\|\mathbf{x}_1\|_2 + \|\Delta\mathbf{b}\|_2) + \|\tilde{\Sigma}_1^{-1}\|_2\,\|\Sigma_2\|_2\,\|\tilde{U}_1^{\mathsf{H}}\,U_2\|_2\,\|\mathbf{x}_2\|_2 \\
&\quad \le \quad \|\tilde{A}_1^{\dagger}\|_2\,(\|\Delta A\|_2\,\|\mathbf{x}_1\|_2 + \|\Delta\mathbf{b}\|_2) + \mathit{dist}\Big(\mathit{Range}(U_1),\ \mathit{Range}(\tilde{U}_1)\Big)\,\|\mathbf{x}_2\|_2 \\
&\quad \le \quad \frac{\sigma_1(A)}{\sigma_r(A)}\,\frac{\|\mathbf{x}_*\|_2}{1-\frac{\sigma_{r+1}+\|\Delta A\|_2}{\sigma_r(A)}}\,\left(\sqrt{2}\,\frac{\|\Delta A\|_2}{\|A\|_2} + \frac{\|\Delta\mathbf{b}\|_2}{\|\mathbf{b}\|_2}\right).
\end{aligned}$$

Let $\tilde{\mathbf{x}} = \tilde{\mathbf{x}}_1 + \tilde{\mathbf{x}}_2 \in \mathit{sol}_{\theta}(\tilde{A},\, \tilde{\mathbf{b}})$ with

$$\tilde{\mathbf{x}}_2 \quad = \quad \tilde{V}_2\,\tilde{V}_2^{\mathsf{H}}\,\mathbf{x}_1 + \tilde{V}_2\,\tilde{V}_2^{\mathsf{H}}\,\mathbf{x}_2 \in \mathit{Kernel}\,(\tilde{A}_{\theta}).$$

Then

$$\|\tilde{\mathbf{x}} - \mathbf{x}_*\|_2 \quad \le \quad \|(\tilde{\mathbf{x}}_1 + \tilde{V}_2\,\tilde{V}_2^{\mathsf{H}}\,\mathbf{x}_1) - \mathbf{x}_1\|_2 + \|\tilde{V}_2\,\tilde{V}_2^{\mathsf{H}}\,\mathbf{x}_2 - \mathbf{x}_2\|_2$$

while, similar to the proof of Theorem 8 from (30),

$$\|\tilde{V}_2\,\tilde{V}_2^{\mathsf{H}}\,\mathbf{x}_2 - \mathbf{x}_2\|_2 \quad \le \quad \frac{\sigma_1(A)}{\sigma_r(A)}\,\frac{1}{1-\frac{\sigma_{r+1}+\|\Delta A\|_2}{\sigma_r(A)}}\,\frac{\|\Delta A\|_2}{\|A\|_2}\,2\,\|\tilde{V}_2\,\tilde{V}_2^{\mathsf{H}}\,\mathbf{x}_2\|_2$$

leading to (31). $\square$

REFERENCES

[1] K. E. Avrachenkov and J. B. Lasserre, *Analytic perturbation of Sylvester matrix equations*, IEEE Trans. Automat. Control, 47 (2002), pp. 1116–1119.

[2] J. L. Barlow, H. Erbay, and I. Slapničar, *An alternative algorithm for the refinement of ULV decompositions*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 198–211, https://doi.org/10.1137/S0895479801372621.

[3] M. Coornaert, *Topological Dimension and Dynamical Systems*, Springer, Switzerland, 2015.

[4] B. Dayton, T.-Y. Li, and Z. Zeng, *Multiple zeros of nonlinear systems*, Math. Comp., 80 (2011), pp. 2143–2168.

[5] J. W. Demmel and A. Edelman, *The dimension of matrices (matrix pencils) with given Jordan (Kronecker) canonical forms*, Linear Algebra Appl., 230 (1995), pp. 61–87.

[6] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 303–353, https://doi.org/10.1137/S0895479895290954.
[7] R. D. Fierro, P. C. Hansen, and P. S. K. Hansen, *UTV Tools: Matlab templates for rank-revealing UTV decompositions*, Numer. Algorithms, 20 (1999), pp. 165–194.
[8] S. Gao, *Factoring multivariate polynomials via partial differential equations*, Math. Comp., 72 (2003), pp. 801–822.
[9] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., The Johns Hopkins University Press, Baltimore, 2013.
[10] P. C. Hansen, *The truncated SVD as a method for regularization*, BIT, 27 (1987), pp. 534–553.
[11] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, Math. Model. Comput. 4, SIAM, Philadelphia, 1998, https://doi.org/10.1137/1.9780898719697.
[12] P. C. Hansen, *Discrete Inverse Problems. Insight and Algorithms*, Fund. Alg. 7, SIAM, Philadelphia, 2010, https://doi.org/10.1137/1.9780898718836.
[13] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images: Matrices, Spectra, and Filtering*, Fund. Alg. 3, SIAM, Philadelphia, 2006, https://doi.org/10.1137/1.9780898718874.
[14] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002, https://doi.org/10.1137/1.9780898718027.
[15] W. Kahan, *Conserving Confluence Curbs Ill-Condition*, Technical Report 6, Computer Science, University of California, Berkeley, 1972.
[16] D. A. Klain and G.-C. Rota, *Introduction to Geometric Probability*, Cambridge University Press, Cambridge, 1997.
[17] V. Lakshmibai and J. Brown, *The Grassmannian Variety*, Springer, New York, 2015.
[18] T.-L. Lee, T.-Y. Li, and Z. Zeng, *A rank-revealing method with updating, downdating, and applications, Part* II, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 503–525, https://doi.org/10.1137/07068179X.
[19] T.-L. Lee, T.-Y. Li, and Z. Zeng, *RankRev—A Matlab package for computing the numerical rank and updating/downdating*, Numer. Algorithms, 77 (2018), pp. 559–576.
[20] T.-Y. Li and Z. Zeng, *A rank-revealing method with updating, downdating and applications*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 918–946, https://doi.org/10.1137/S0895479803435282.
[21] L.-H. Lim, K. S.-W. Wong, and K. Ye, *Numerical Algorithms on the Affine Grassmannian*, preprint, https://arxiv.org/abs/1607.01833, 2018.
[22] L.-H. Lim, K. S.-W. Wong, and K. Ye, *The Grassmannian of Affine Subspaces*, preprint, https://arxiv.org/abs/1807.10883, 2018.
[23] T. Liu and J. Huang, *A discrete-time recurrent neural network for solving rank-deficient matrix equations with an application to output regulation of linear systems*, IEEE Trans. Neural Netw. Learn. Syst., 29 (2018), pp. 2271–2277.
[24] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996, https://doi.org/10.1137/1.9781611971484.
[25] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, 2000.
[26] T. Mora, *Solving Polynomial Equation Systems* I: *The Kronecker-Duval Philosophy*, Cambridge University Press, London, 2003.
[27] A. Neumaier, *Solving ill-conditioned and singular linear systems: A tutorial on regularization*, SIAM Rev., 40 (1998), pp. 636–666, https://doi.org/10.1137/S0036144597321909.
[28] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, NJ, 1980; republished as Classics Appl. Math. 20, SIAM, 1998, https://doi.org/10.1137/1.9781611971163.
[29] G. Peters and J. H. Wilkinson, *Inverse iteration, ill-conditioned equations and Newton's method*, SIAM Rev., 21 (1979), pp. 339–360, https://doi.org/10.1137/1021052.
[30] A. Saqellari-Likoka and V. Karathanassi, *An approach for solving rank-deficient systems that enable atmospheric path delay and water vapor content estimation*, IEEE Trans. Geosci. Remote Sensing, 46 (2008), pp. 3187–3195.
[31] G. W. Stewart, *UTV decompositions*, in Numerical Analysis 1993 (Dundee, 1993), D. F. Griffiths and G. A. Watson, eds., Pitman Res. Notes Math. Ser., Longman Sci. Tech., Harlow, 1994, pp. 225–236.
[32] G. W. Stewart and J. Sun, *Matrix Perturbation Theory*, Academic Press, New York, 1990.
[33] T. Stykel, *Numerical solution and perturbation theory for generalized Lyapunov equations*, Linear Algebra Appl., 349 (2002), pp. 155–185.

[34] P.-Å. Wedin, *Perturbation bounds in connection with singular value decomposition*, BIT, 12 (1972), pp. 99–111.

[35] P.-Å. Wedin, *Perturbation theory for pseudo-inverses*, BIT, 13 (1973), pp. 217–232.

[36] J. Wilkening and J. Yu, *A local construction of the Smith normal form of a matrix polynomial*, J. Symbolic Comput., 46 (2011), pp. 1–12.

[37] W. Wu and Z. Zeng, *The numerical factorization of polynomials*, J. Found. Comput. Math., 17 (2017), pp. 259–286.

[38] Z. Zeng, *A polynomial elimination method for numerical computation*, Theoret. Comput. Sci., 409 (2008), pp. 318–331.

[39] Z. Zeng, *Intuitive interface for solving linear and nonlinear system of equations*, in Mathematical Software — ICMS 2018, J. H. Davenport, M. Kauers, G. Labahn, and J. Urban, eds., LNCS 10931, Springer International AG, 2018, pp. 495–506.

[40] Z. Zeng and T.-Y. Li, *NAClab: A Matlab toolbox for numerical algebraic computation*, ACM Commun. Comput. Algebra, 47 (2013), pp. 170–173.