J. Math. Anal. Appl. 556 (2026) 130236

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

journal homepage: www.elsevier.com/locate/jmaa

Regular Articles

Higher-order singular-value derivatives of real rectangular matrices ☆

Róisín Luo [a,b,*], Colm O'Riordan [a,b], James McDermott [a,b]

[a] *University of Galway, Ireland*
[b] *Research Ireland – Centre for Research Training in Artificial Intelligence (CRT-AI), Ireland*

A R T I C L E I N F O

Article history:
Received 5 June 2025
Available online 13 November 2025
Submitted by J.A. Ball

Keywords:
Singular-value derivatives
Singular-value Hessian
Spectral perturbation
Perturbation analysis
Matrix analysis
Functional analysis

A B S T R A C T

Higher-order derivatives of singular values in real rectangular matrices arise naturally in both numerical simulation and theoretical analysis, with applications in areas such as statistical physics and optimization in deep learning. Deriving closed-form expressions beyond first order has remained a difficult problem within classical matrix analysis, and no general framework has been available. To address this gap, we present an operator-theoretic framework that extends Kato's analytic perturbation theory from self-adjoint operators to real rectangular matrices, thereby yielding general n-th order Fréchet derivatives of singular values. As a special case, we obtain a closed-form Kronecker-product representation of the singular-value Hessian, not previously found in the literature. This framework bridges abstract perturbation theory with matrix analysis and provides a systematic tool for higher-order spectral analysis.

1. Introduction

Singular values lie at the core of modern matrix analysis, encapsulating key spectral information such as *operator norm*, *conditioning*, *effective rank*, and underpinning applications across numerical linear algebra, data science, control theory, and mathematical physics [6,37,9,32]. In random matrix theory, singular values govern limiting laws such as Marchenko–Pastur distributions [18], edge fluctuations described by Tracy–Widom laws [33], and fine-scale local statistics such as local eigenvalue spacings [19]. In physics and deep learning, higher-order derivatives of singular values are indispensable for rigorous analysis in stochastic dynamical settings, where systems are subject to noise and random perturbations [21].

☆ Reproducibility for numerical experiments: https://github.com/roisincrtai/highorder_spectral_variation_analysis.

* Corresponding author at: University of Galway, Ireland.
E-mail address: roisincrtai@gmail.com (R. Luo).

https://doi.org/10.1016/j.jmaa.2025.130236

For example, let $\boldsymbol{\theta}_t \in \mathbb{R}^{m\times n}$ be a parameter matrix with r non-zero singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$. Its dynamics are characterized by an adapted Itô process [10]

$$\mathrm{dvec}(\boldsymbol{\theta}_t) = G_t \mathrm{d}t + D_t \mathrm{d}W_t, \tag{1}$$

where vec(·) denotes the vectorization operator, $G_t \in \mathbb{R}^{mn}$ is the *drift term*, $D_t \in \mathbb{R}^{mn\times mn}$ is the *diffusion coefficient*, and $\mathrm{d}W_t$ is a high-dimensional Wiener process in $\mathbb{R}^{mn}$. Let

$$\phi_t = \phi(\sigma_1, \sigma_2, \ldots, \sigma_r)(t) \tag{2}$$

denote a spectral functional of the singular values of $\boldsymbol{\theta}_t$, then applying Itô's lemma shows that the rigorous analysis of the spectral dynamics of $\mathrm{d}\phi_t$ requires the second-order derivatives of singular values [21]. Such induced dynamics arise naturally in both physics and deep learning. In physics, the *von Neumann entropy* — a measure of the statistical uncertainty within a quantum system — is a spectral functional of singular values [7], and widely used in the study of quantum entanglement [20]. In deep learning, the Lipschitz continuity of neural networks is a spectral functional of the largest singular value [16]. For more general non-Gaussian drivers in stochastic dynamics, such as Lévy processes, higher-order derivatives of singular values are indispensable for rigorous analysis and for deriving sharp bounds [1].

Although first-order derivatives of singular values are well known in the literature [34,30,31], explicit closed-form expressions for second- and higher-order derivatives are largely absent from the literature. A unified, highly procedural, and systematic framework for their derivation has been lacking, since direct approaches via matrix analysis are challenging due to the intricate interplay among local spectral structures (*e.g.*, spectral gaps), left and right singular subspaces, and the associated null spaces.

To bridge this gap, we present an operator-theoretic framework for deriving arbitrary higher-order derivatives of singular values in a highly procedural approach. Our approach treats matrices as bounded linear operators on Hilbert spaces and extends Kato's analytic perturbation theory [11] beyond the self-adjoint setting. The key step is to embed a non-self-adjoint real rectangular matrix into a self-adjoint operator via the Jordan–Wielandt embedding (*i.e.*, Hermitian dilation trick) [36,30,28]. We then analyze the asymptotic expansions of the resulting eigenvalues by extending Kato's results in eigenvalue expansions, relate these eigenvalue expansions to Fréchet derivatives of singular values, and express the Fréchet derivative tensors with Kronecker-product representation.

1.1. *Perturbation theory*

Classical perturbation theory has developed along several independent traditions. For example, **analytic operator-theoretic perturbation theory** [24,11] treats holomorphic families of operators on Banach or Hilbert spaces, using resolvents, Riesz projectors, and contour integrals to prove the existence of analytic eigenvalue and eigenspace branches and to derive expansion formulas, including trace identities for eigenvalue clusters. **Matrix perturbation theory** [30,9,3] focuses on the finite-dimensional case and derives explicit perturbation formulas via algebraic tools such as characteristic polynomials, Schur forms, and Sylvester equations, typically without explicitly invoking the operator-theoretic machinery. **Rayleigh–Schrödinger perturbation theory** [23,27,26] in quantum mechanics provides basis-dependent expansions in terms of matrix elements and energy gaps; these coincide with the analytic expansions under discreteness and gap assumptions, but are often presented in physics as formal series rather than within Kato's framework. Despite their differences in tool and emphasis, these frameworks are mathematically consistent and recover the same perturbative corrections in overlapping regimes.

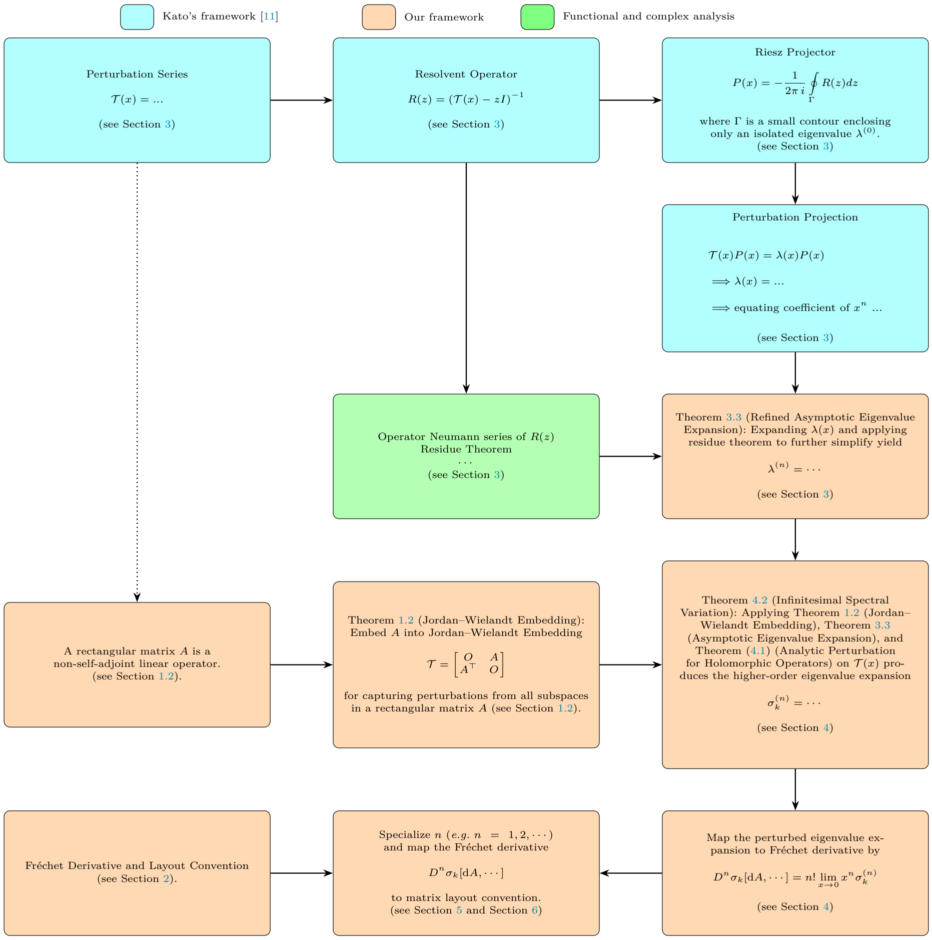


Fig. 1. Theoretical Framework for Infinitesimal Spectral Variations. We extend Kato's analytic perturbation theory for self-adjoint operators to derive arbitrary-order singular-value derivatives [11]. For a rectangular matrix A, we introduce its Jordan–Wielandt embedding $\mathcal{T}$ (Theorem 1.2), a block self-adjoint operator that encodes perturbations across all subspaces (*i.e.*, left-singular, right-singular, left-null, and right-null). By extending Kato's asymptotic eigenvalue expansions to this embedding and expressing them in explicit closed form — computing and simplifying with residue theorem — yields the nth-order expansions of singular values of A. These expansions are then related to Fréchet derivatives, given by analytic perturbation theorem (Theorem 4.1). Finally, by specializing to explicit matrix-layout conventions, we obtain a systematic and constructive procedure for computing arbitrary-order singular-value derivatives of rectangular matrices. Our method is highly procedure for deriving arbitrary-order singular-value derivatives.

1.2. Schematic overview

A schematic overview of the framework is illustrated in Fig. 1. To apply Kato's framework for self-adjoint operators, we first embed a non-self-adjoint $A \in \mathbb{R}^{m\times n}$ (since $A \neq A^\top$) into a self-adjoint operator $\mathcal{T}$ using the Jordan–Wielandt embedding (*i.e.*, Hermitian dilation) [36,30,6,14,2,9,28], taking:

$$\mathcal{T} := \begin{bmatrix} O & A \\ A^\top & O \end{bmatrix} \in \mathbb{R}^{(m+n)\times(m+n)}. \tag{3}$$

It is immediate that $\mathcal{T}$ is self-adjoint, since:

$$\mathcal{T}^\top = \begin{bmatrix} O & A \\ A^\top & O \end{bmatrix}^\top = \begin{bmatrix} O & A \\ A^\top & O \end{bmatrix} = \mathcal{T}. \tag{4}$$

This embedding preserves the complete information regarding the spectrum of A. The spectrum of A is stated in Theorem 1.1 (Matrix Singular Value Decomposition (Full Form)), and the spectrum of $\mathcal{T}$ relates to the spectral structure of A as stated in Theorem 1.2 (Spectrum of Jordan–Wielandt Embedding).

Theorem 1.1 *(Matrix Singular Value Decomposition (Full Form)). Let $A \in \mathbb{R}^{m\times n}$ be a real rectangular matrix. Then A admits a full singular value decomposition (SVD) [9,30] by:*

$$A = U\Sigma V^\top, \tag{5}$$

$$U = \begin{bmatrix} U_r & U_0 \end{bmatrix} \in \mathbb{R}^{m\times m}, \qquad V = \begin{bmatrix} V_r & V_0 \end{bmatrix} \in \mathbb{R}^{n\times n}, \tag{6}$$

$$\Sigma = \left[\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & O_{r\times(n-r)} \\ & & & \sigma_r & \\ \hline & O_{(m-r)\times r} & & & O_{(m-r)\times(n-r)} \end{array}\right], \tag{7}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the non-zero singular values, $U_r \in \mathbb{R}^{m\times r}$ and $V_r \in \mathbb{R}^{n\times r}$ contain the corresponding left and right singular vectors, and U_0 and V_0 span the left and right null spaces of A, respectively.

Theorem 1.2 *(Spectrum of Jordan–Wielandt Embedding). The spectrum of $A \in \mathbb{R}^{m\times n}$ and the spectrum of its Jordan–Wielandt embedding [36,30,14,2,9,28]:*

$$\mathcal{T} := \begin{bmatrix} O & A \\ A^\top & O \end{bmatrix} \in \mathbb{R}^{(m+n)\times(m+n)}, \tag{8}$$

are directly related. Given the SVD of A as stated in Theorem 1.1, the spectrum of A relates to the spectrum of T by:

$$\mathcal{T}\left(\frac{1}{\sqrt{2}}\begin{bmatrix} u_i \\ v_i \end{bmatrix}\right) = \sigma_i\left(\frac{1}{\sqrt{2}}\begin{bmatrix} u_i \\ v_i \end{bmatrix}\right), \tag{9}$$

and:

$$\mathcal{T}\left(\frac{1}{\sqrt{2}}\begin{bmatrix} u_i \\ -v_i \end{bmatrix}\right) = -\sigma_i\left(\frac{1}{\sqrt{2}}\begin{bmatrix} u_i \\ -v_i \end{bmatrix}\right), \tag{10}$$

respectively, where the factor $\frac{1}{\sqrt{2}}$ ensures normalization and hence orthonormality of the eigenvectors. Thus, each singular value σ_i of A corresponds to a pair of eigenvalues:

$$\lambda_i^{(+)} = \sigma_i, \qquad \lambda_i^{(-)} = -\sigma_i, \tag{11}$$

with eigenvectors constructed directly from the singular vector pair (u_i, v_i). The null spaces are also preserved in this embedding:

$$\ker(\mathcal{T}) = \left\{\begin{bmatrix} u_j \\ 0 \end{bmatrix} : u_j \in \ker(A^\top)\right\} \oplus \left\{\begin{bmatrix} 0 \\ v_k \end{bmatrix} : v_k \in \ker(A)\right\}. \tag{12}$$

Remark 1.3. The use of the Jordan–Wielandt embedding to transfer results on Hermitian eigenvalues to singular values of rectangular matrices is well-known in the literature [30,6,14,2,9]. For instance, Stewart and Sun employ the construction in their analysis of singular-value perturbations, using it to extend Weyl–type inequalities [35,8] and sensitivity bounds from Hermitian eigenvalues to singular values [30]. Li and Li also use the embedding to transfer perturbation bounds for Hermitian eigenvalues to singular values of rectangular matrices [14]. Similarly, Horn and Johnson present the Hermitian dilation as a standard device in matrix analysis for proving variational characterizations and interlacing properties of singular values [9]. Unlike these works, which use the Hermitian dilation mainly as a device to transfer known eigenvalue results, our framework exploits it to develop explicit operator-theoretic expansions that yield closed-form higher-order Fréchet derivatives of singular values.

Next, starting from the eigenvalue expansion of reduced resolvent of operator $\mathcal{T}$ and applying the residue theorem to simplify, we derive the asymptotic eigenvalue expansion of $\mathcal{T}$ up to n-th order under holomorphic perturbations (Theorem 3.3). By relating the n-th order term of this expansion with the corresponding n-th order Fréchet derivative, we obtain explicit expressions for higher-order derivatives of singular values. Finally, we deploy the n-th order Fréchet derivative with matrix layout conventions. In particular, the first-order case ($n = 1$) recovers the well-known Jacobian of singular values; while the second-order case ($n = 2$) yields the singular-value Hessian with Kronecker-product representation, which has not appeared previously in the literature. By bridging the abstract operator-theoretic expansions with matrices, our framework provides a toolkit for arbitrary-order singular-value analysis.

1.3. Contributions

This paper makes the following contributions:

1. **Spectral Variations in Rectangular Matrices**. We present an operator-theoretic framework for analyzing n-th order spectral variations in real rectangular matrices (see Fig. 1). This framework provides a **systematic procedure** for deriving higher-order derivatives of singular values in real rectangular matrices.
2. **Singular-Value Hessian**. Specializing to $n = 2$ yields the second-order derivative (Hessian) of singular values, expressed in a Kronecker-product representation that, to the best of our knowledge, has not appeared previously in the literature. This result is particularly essential for analysis of **induced spectral stochastic dynamics**, where second-order derivatives arise naturally in Itô calculus for stochastic differential equations (SDEs) driven by Wiener processes.

2. Fréchet derivative and layout convention

Deploying results from abstract operator theory in matrix settings requires explicit layout conventions, particularly for the representation of derivatives. Before commencing the theoretical analysis, this section introduces the conventions fundamental to our framework. Section 2.1 introduces matrix layout and the differentiability condition; Section 2.3 presents general Fréchet derivatives for matrix-to-matrix maps together with their tensor representations; and Section 2.4 specializes to Fréchet derivatives of matrix-to-scalar functionals and their vectorized Kronecker-product representation [12].

2.1. Matrix and spectral decomposition

Let

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{bmatrix} \in \mathbb{R}^{m\times n} \tag{13}$$

be a real rectangular matrix of rank $r = \text{rank}(A)$, where $A_{i,j}$ denotes its (i, j)-th entry. The A admits a *full* SVD as stated in Theorem 1.1. Specially, the *reduced* or *truncated* SVD of A is given as:

$$A = \sum_{k=1}^{r} \sigma_k u_k v_k^\top, \tag{14}$$

where $r = \text{rank}(A)$, and u_k and v_k are the left and right singular vectors associated with singular value $\sigma_k > 0$.

Lemma 2.1 *(Essential Matrix Identities). Let $x \in \mathbb{R}$ be a scalar, and real matrices A, B, C and V be of such sizes that one can form their products. Then the following identities hold [8,9,15]:*

1. $\text{vec}(x) = x$,
2. $\text{tr}(x) = x$,
3. $\text{vec}(BVA^\top) = (A \otimes B)\,\text{vec}(V)$,
4. $(A \otimes B)^\top = A^\top \otimes B^\top$,
5. $\text{tr}(ABC) = \text{tr}(CAB) = \text{tr}(BCA)$.

2.2. Differentiability condition

To ensure the existence of higher-order differentiability of non-zero singular values and associated singular vectors, we further assume that the non-zero singular values of $A \in \mathbb{R}^{m\times n}$ are simple (*i.e.*, each non-zero singular value has multiplicity one), as stated in Assumption 2.2 (Simplicity Assumption of Non-Zero Singular Values). This *simplicity* assumption is essential for ensuring that non-zero singular value $\sigma_i > 0$ of A and associated singular vectors u_i and v_i depend smoothly on the entries of A, in fact yielding $u_i, v_i, \sigma_i \in C^\infty$ (*i.e.* maps are infinitely continuously differentiable). Under this assumption, non-zero singular values and their associated singular vectors vary smoothly with perturbations of A.

Assumption 2.2 *(Simplicity Assumption of Non-Zero Singular Values).* We assume that the non-zero singular values of A are *simple*, *i.e.*,

$$\sigma_i \neq \sigma_j \quad \text{for all} \quad i \neq j, \tag{15}$$

[11,9].

If this assumption fails, a non-zero singular value may have multiplicity greater than one; singular values then remain continuous but may fail to be differentiable at points of multiplicity, and the associated singular subspaces are well defined whereas individual singular vectors are not unique. In such settings, higher-order derivatives generally do not exist in the classical context, and analysis must instead be carried out in terms of spectral projectors or within the framework of subdifferential calculus [4,13].

2.3. Matrix Fréchet derivative as multilinear operator

We regard matrix Fréchet derivatives as multilinear operators [25,37]. A definition for general Fréchet differentiable real matrix-to-matrix maps and their tensor representation are in Definition 2.3 (α-Times Continuously Fréchet Differentiable Matrix Map). The existence and uniqueness of the Fréchet derivative are stated in Theorem 2.4 (Uniqueness of α-Times Fréchet Derivative [25,29]).

Definition 2.3 *(α-Times Continuously Fréchet Differentiable Matrix Map).* Let

$$F:\, \mathbb{R}^{m\times n} \to \mathbb{R}^{s\times t} \tag{16}$$

be α-times continuously Fréchet differentiable (*i.e.*, $F \in C^{\alpha}$) [25,29,9]. The α-th Fréchet derivative of F is a multilinear map:

$$D^{\alpha}F:\, (\mathbb{R}^{m\times n})^{\alpha} \to \mathbb{R}^{s\times t}. \tag{17}$$

Writing $F_{i,j}$ for the (i,j)-th component of F and $A_{p,q}$ for the (p,q)-th entry of A, with:

$$1 \le i \le s, \qquad 1 \le j \le t, \qquad 1 \le p \le m, \qquad 1 \le q \le n, \tag{18}$$

then the α-th derivative $D^{\alpha}F$ at matrix A is a tensor, defined by:

$$\big[D^{\alpha}F(A)\big]_{i,j\,;\,p_1q_1\,\dots\,p_\alpha q_\alpha} \;=\; \frac{\partial^{\alpha}F_{i,j}(A)}{\partial A_{p_1q_1}\cdots\partial A_{p_\alpha q_\alpha}} \;\in\; \mathbb{R}. \tag{19}$$

The action of tensor $D^{\alpha}F(A)$ on directions $H_1,\dots,H_\alpha \in \mathbb{R}^{m\times n}$ is obtained *component-wise* by contracting tensor $D^{\alpha}F(A)$ with the indices on $H_1,\dots,H_\alpha$:

$$\big[D^{\alpha}F(A)[H_1,\dots,H_\alpha]\big]_{i,j} \;=\; \frac{\partial^{\alpha}F_{i,j}(A)}{\partial A_{p_1q_1}\cdots\partial A_{p_\alpha q_\alpha}}\,(H_1)_{p_1q_1}\cdots(H_\alpha)_{p_\alpha q_\alpha}. \tag{20}$$

Moreover, for $H \in \mathbb{R}^{m\times n}$, the F at A admits a multivariate Taylor expansion:

$$F(A+H) = \sum_{\beta=0}^{\alpha}\frac{1}{\beta!}\,D^{\beta}F(A)[\underbrace{H,\dots,H}_{\beta\ \text{times}}] + o\big(\|H\|^{\alpha}\big), \qquad (\|H\|\to 0), \tag{21}$$

where $\|\cdot\|$ is any norm on $\mathbb{R}^{m\times n}$ (*e.g.*, the Frobenius norm).

Theorem 2.4 *(Uniqueness of α-Times Fréchet Derivative [25,29]). Suppose $F \in C^{\alpha}$ is differentiable up to order α. Then $D^{\alpha}F$ exists, is a symmetric α-linear map, and is* unique*. That is, there is no other α-linear operator satisfying the defining Taylor-remainder condition. This theorem ensures the uniqueness of the derivatives of singular values under the differentiability condition, as stated in Assumption 2.2 (Simplicity Assumption of Non-Zero Singular Values).*

2.4. Representation convention for matrix-valued functionals

We focus on the derivatives of singular values, which are matrix-valued functionals. To obtain matrix representations to facilitate concrete applications, we specialize the general matrix-to-matrix maps of Definition 2.3 to matrix-valued functionals. In general, the α-th Fréchet derivative is a higher-order tensor. To express such tensors in matrix form, we employ vectorization (with a **column-major convention**) together with the Kronecker-product representation [12,17], as established in Corollary 2.5 (Vectorized Kronecker-Product Representation of Fréchet Derivative). As complementary conventions, we also introduce explicit matrix layouts for the Jacobian in Section 2.4.1 (Representation convention for Jacobian of matrix-valued functional) and for the Hessian in Section 2.4.2 (Representation convention for Hessian of matrix-valued functional).

Corollary 2.5 *(Vectorized Kronecker-Product Representation of Fréchet Derivative). Let*

$$f : \mathbb{R}^{m\times n} \to \mathbb{R} \tag{22}$$

*be α-times continuously Fréchet differentiable (*i.e., $f \in C^\alpha$*). For directions* $H_1, \ldots, H_\alpha \in \mathbb{R}^{m\times n}$*, the multilinear action* $D^\alpha f$ *at* $A \in \mathbb{R}^{m\times n}$ *is given by the Frobenius tensor inner product [12,17]:*

$$\begin{aligned} D^\alpha f(A)[H_1, \ldots, H_\alpha] &= \left\langle D^\alpha f(A), H_1 \otimes \cdots \otimes H_\alpha \right\rangle & (23)\\ &= \left\langle \mathrm{vec}\left(D^\alpha f(A)\right), \mathrm{vec}\left(H_1 \otimes \cdots \otimes H_\alpha\right)\right\rangle & (24)\\ &= \mathrm{vec}\left(D^\alpha f(A)\right)^\top \mathrm{vec}\left(H_1 \otimes \cdots \otimes H_\alpha\right), & (25)\end{aligned}$$

*where ⊗ represents Kronecker product (*i.e.*, tensor product) and:*

$$\mathrm{vec} : \mathbb{R}^{m\times n} \mapsto \mathbb{R}^{mn} \tag{26}$$

represents the vectorization operator with the ***column-major convention*** *[12].*

This vectorization is particularly useful for representing arbitrary-order derivatives of matrix-valued functionals in matrix form.

2.4.1. Representation convention for Jacobian of matrix-valued functional

Representing the Jacobian of matrix-valued functionals in matrix form is standard in the literature [9]. For clarity, we introduce a matrix layout as a complementary representation for the Jacobian of matrix-valued functionals. Let

$$f : \mathbb{R}^{m\times n} \mapsto \mathbb{R} \tag{27}$$

be a first-order Fréchet differentiable functional. Then the differential of f admits:

$$\mathrm{d}f = Df(A)[\mathrm{d}A] = \langle Df(A), \mathrm{d}A\rangle = \left(\frac{\partial f}{\partial A}\right)^\top \mathrm{d}A = \mathrm{tr}\left[\left(\frac{\partial f}{\partial A}\right)^\top \mathrm{d}A\right], \tag{28}$$

where $\frac{\partial f}{\partial A}$ and infinitesimal variation $\mathrm{d}A \in \mathbb{R}^{m\times n}$ are piece-wisely defined as:

$$\frac{\partial f}{\partial A} = \begin{bmatrix} \frac{\partial f}{\partial A_{1,1}} & \frac{\partial f}{\partial A_{1,2}} & \cdots & \frac{\partial f}{\partial A_{1,n}} \\ \frac{\partial f}{\partial A_{2,1}} & \frac{\partial f}{\partial A_{2,2}} & \cdots & \frac{\partial f}{\partial A_{2,n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial A_{m,1}} & \frac{\partial f}{\partial A_{m,2}} & \cdots & \frac{\partial f}{\partial A_{m,n}} \end{bmatrix} \tag{29}$$

with **denominator layout convention**, and:

$$\mathrm{d}A = \begin{bmatrix} \mathrm{d}A_{1,1} & \mathrm{d}A_{1,2} & \cdots & \mathrm{d}A_{1,n} \\ \mathrm{d}A_{2,1} & \mathrm{d}A_{2,2} & \cdots & \mathrm{d}A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{d}A_{m,1} & \mathrm{d}A_{m,2} & \cdots & \mathrm{d}A_{m,n} \end{bmatrix}. \tag{30}$$

2.4.2. Representation convention for Hessian of matrix-valued functional

The Hessian of a matrix-valued functional is naturally a higher-order tensor; for instance, it is a fourth-order tensor for matrix-valued functionals [12]. Let

$$f : \mathbb{R}^{m\times n} \to \mathbb{R} \tag{31}$$

be a twice Fréchet differentiable functional. Since

$$D^2 f = D(Df), \tag{32}$$

to obtain a matrix representation of $D^2 f$, we first consider the representation layout of the first-order derivative for a matrix-to-matrix map $F : \mathbb{R}^{m\times n} \to \mathbb{R}^{s\times t}$. We then apply vectorization together with this layout to express the second-order derivatives of matrix-valued functionals in matrix form.

Let

$$F : \mathbb{R}^{m\times n} \mapsto \mathbb{R}^{s\times t} \tag{33}$$

be a first-order Fréchet differentiable matrix-to-matrix map. Then there exists:

$$\frac{\partial \operatorname{vec}(F)}{\partial \operatorname{vec}(A)} \in \mathbb{R}^{p\times q} \quad p = mn \quad \text{and} \quad q = st, \tag{34}$$

piece-wisely defined as:

$$\frac{\partial \operatorname{vec}(F)}{\partial \operatorname{vec}(A)} = \begin{pmatrix} \frac{\partial \operatorname{vec}(F)_1}{\partial \operatorname{vec}(A)_1} & \frac{\partial \operatorname{vec}(F)_2}{\partial \operatorname{vec}(A)_1} & \cdots & \frac{\partial \operatorname{vec}(F)_q}{\partial \operatorname{vec}(A)_1} \\ \frac{\partial \operatorname{vec}(F)_1}{\partial \operatorname{vec}(A)_2} & \frac{\partial \operatorname{vec}(F)_2}{\partial \operatorname{vec}(A)_2} & \cdots & \frac{\partial \operatorname{vec}(F)_q}{\partial \operatorname{vec}(A)_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \operatorname{vec}(F)_1}{\partial \operatorname{vec}(A)_p} & \frac{\partial \operatorname{vec}(F)_2}{\partial \operatorname{vec}(A)_p} & \cdots & \frac{\partial \operatorname{vec}(F)_q}{\partial \operatorname{vec}(A)_p} \end{pmatrix} \tag{35}$$

by using **denominator layout convention** on $\operatorname{vec}(F)$ and $\operatorname{vec}(A)$ [9]. Then the Hessian of the matrix-valued functional f can be defined as:

$$\frac{\partial}{\partial \operatorname{vec}(A)} \operatorname{vec}\left(\frac{\partial f}{\partial A}\right) \tag{36}$$

with vectorized representation.

Relating Vectorized Representation to $D^2 f$. We now relate this vectorized representation to $D^2 f$. By Corollary 2.5 (Vectorized Kronecker-Product Representation of Fréchet Derivative), consider:

$$D^2 f[\mathrm{d}A, \mathrm{d}A] = \langle D^2 f, \mathrm{d}A \otimes \mathrm{d}A\rangle, \tag{37}$$

and use the following identities from Lemma 2.1 (Essential Matrix Identities):

1. $\mathrm{vec}(x) = x$,
2. $\mathrm{vec}(BVA^\top) = (A \otimes B)\,\mathrm{vec}(V)$,

then it yields:

$$D^2 f[\mathrm{d}A, \mathrm{d}A] = \langle D^2 f, \mathrm{d}A \otimes \mathrm{d}A \rangle \tag{38}$$

$$= \mathrm{vec}(D^2 f)^\top \mathrm{vec}(\mathrm{d}A \otimes \mathrm{d}A) \tag{39}$$

$$= \mathrm{vec}(\mathrm{d}A)^\top (D^2 f)^\top \mathrm{vec}(\mathrm{d}A). \tag{40}$$

Relating

$$\frac{\partial}{\partial \mathrm{vec}(A)} \mathrm{vec}\left(\frac{\partial f}{\partial A}\right) \tag{41}$$

with $D^2 f$ yields:

$$D^2 f = \left[\frac{\partial}{\partial \mathrm{vec}(A)} \mathrm{vec}\left(\frac{\partial f}{\partial A}\right)\right]^\top, \tag{42}$$

such that:

$$D^2 f[\mathrm{d}A, \mathrm{d}A] = \mathrm{vec}(\mathrm{d}A)^\top \left[\frac{\partial}{\partial \mathrm{vec}(A)} \mathrm{vec}\left(\frac{\partial f}{\partial A}\right)\right] \mathrm{vec}(\mathrm{d}A). \tag{43}$$

3. Refined asymptotic eigenvalue expansion

Kato's monograph [11] establishes the existence of asymptotic eigenvalue expansions and, in particular, provides a closed-form expression for the *weighted mean* of eigenvalue coefficients, as stated in Theorem 3.2 (Kato's Weighted Mean of Eigenvalue Expansions [11, Ch. II, §2.2]). Nevertheless, Kato's formulation is expressed with an infinite summation of contour integrals involving the perturbed resolvent and does not yield explicit, constructive formulas for the individual coefficients, which limits its direct applicability in our setting. Building on the analytic foundations laid by Kato, and by employing explicit Neumann expansions of resolvents together with the residue theorem, we refine this framework to derive an *explicit, closed-form formula for arbitrary-order eigenvalue coefficients* of holomorphic families of bounded self-adjoint operators. Our main result, Theorem 3.3 (Refined Closed-Form Asymptotic Expansion of Simple Isolated Eigenvalue in Self-Adjoint Operator), goes beyond Kato's weighted mean by furnishing a fully constructive representation of each eigenvalue coefficient. The overall scheme is illustrated in Fig. 1.

Definition 3.1 *(Space of Bounded Linear Operators).* Let

$$\mathcal{L}(X) = \{\, T : X \mapsto X \mid X \text{is a Banach space and } T \text{ is a bounded linear operator}\} \tag{44}$$

be the Banach space of bounded linear operators.

Theorem 3.2 *(Kato's Weighted Mean of Eigenvalue Expansions [11, Ch. II, §2.2]). Let*

$$\mathcal{T}(x) = \mathcal{T}^{(0)} + \sum_{j=1}^{\infty} x^j \mathcal{T}^{(j)} \;\in\; \mathcal{L}(X) \tag{45}$$

be a holomorphic family of bounded operators on a Banach space X [11, Ch. II, §2.1, Eq (2.1)]. Suppose $\lambda^{(0)}$ is an isolated eigenvalue of $\mathcal{T}^{(0)}$ of algebraic multiplicity m.

Let

$$R(z) = (\mathcal{T}(x) - zI)^{-1} \tag{46}$$

be the perturbed resolvent, let

$$P(x) = -\frac{1}{2\pi i} \oint_{\Gamma} R(z) dz, \tag{47}$$

be Riesz projector, and define

$$\widehat{\mathcal{T}}^{(n)} = \sum_{p=1}^{\infty} (-1)^{p-1} \sum_{\substack{i_1+\cdots+i_p=n \\ i_j \geq 1}} \frac{1}{2\pi i} \oint_{\Gamma} R(z)\, \mathcal{T}^{(i_1)}\, R(z) \cdots R(z)\, \mathcal{T}^{(i_p)}\, R(z)\, (z - \lambda^{(0)})\, dz, \tag{48}$$

where Γ is a small contour enclosing only $\lambda^{(0)}$ and no other spectrum [11, Ch. II, §2.1, Eq (2.18)]. Then the weighted mean of the perturbed eigenvalues is:

$$\hat{\lambda}(x) := \frac{1}{m} \operatorname{tr}\big(\mathcal{T}(x) P(x)\big), \tag{49}$$

admits the expansion

$$\hat{\lambda}(x) = \lambda^{(0)} + \sum_{n=1}^{\infty} x^n \hat{\lambda}^{(n)} \tag{50}$$

[11, Ch. II, §2.1, Eq (2.21)], and:

$$\hat{\lambda}^{(n)} = \tfrac{1}{m} \operatorname{tr}\big(\widehat{\mathcal{T}}^{(n)}\big) \tag{51}$$

[11, Ch. II, §2.1, Eq (2.22)].

Sketch to refine Kato's result Following Kato's analytic framework, we also begin with the perturbation series $\mathcal{T}(x)$ of a self-adjoint operator $\mathcal{T}$. Kato's monograph establishes eigenvalue expansions via contour integrals of the resolvent and provides a closed-form expression for the *weighted mean* of eigenvalues, but it does not supply explicit constructive formulas for the individual coefficients. Our approach departs at this point: we expand the resolvent explicitly through its Neumann series, apply the residue theorem on the contour integrals, and simplify the resulting expressions. This yields a closed-form asymptotic eigenvalue expansion for $\mathcal{T}$ with computable coefficients. Unlike Kato's result, our expansion is formulated in terms of the finite summation of the series of unperturbed resolvent, which enables systematic computation and, in particular, facilitates the subsequent derivation of singular-value Fréchet derivatives. As a special case, it also recovers the classical Rayleigh–Schrödinger corrections of quantum mechanics [23,27,26].

Theorem 3.3 *(Refined Closed-Form Asymptotic Expansion of Simple Isolated Eigenvalue in Self-Adjoint Operator). Let*

$$\mathcal{T}(x) = \mathcal{T}^{(0)} + \sum_{j=1}^{\infty} x^j\, \mathcal{T}^{(j)} \;\in\; \mathcal{L}(X) \tag{52}$$

be a holomorphic family of bounded operators on a Banach space X, where $\mathcal{T}^{(0)}$ denotes the unperturbed operator and $x \in \mathbb{C}$ is the perturbation parameter.

Unperturbed Reduced Resolvent. *Define the unperturbed reduced resolvent [25,11,37] of $\mathcal{T}^{(0)}$:*

$$S^{(0)} = \left(\mathcal{T}^{(0)} - \lambda^{(0)} I\right)^{-1} (I - w^{(0)} (w^{(0)})^{*}), \tag{53}$$

*where $\lambda^{(0)}$ is a simple eigenvalue of $\mathcal{T}^{(0)}$ and $w^{(0)}$ is the associated normalized eigenvector (*i.e.*, $\|w^{(0)}\|_2 = 1$).*

Theorem Claim. *Then there exists a unique holomorphic branch $\lambda(x)$ of eigenvalues of $\mathcal{T}(x)$. It admits the power series:*

$$\lambda(x) = \sum_{n=0}^{\infty} x^n \, \lambda^{(n)}, \tag{54}$$

and for each integer $n \geq 1$,

$$\lambda^{(n)} = \sum_{p=1}^{n} (-1)^{p-1} \sum_{\substack{i_1 + \cdots + i_p = n \\ i_j \geq 1}} \left\langle w^{(0)}, \, \mathcal{T}^{(i_1)} \, S^{(0)} \, \mathcal{T}^{(i_2)} \, S^{(0)} \cdots S^{(0)} \, \mathcal{T}^{(i_p)} \, w^{(0)} \right\rangle. \tag{55}$$

Proof 3.4. We begin by presenting a compact and explicit proof of Theorem 3.2 (Kato's Weighted Mean of Eigenvalue Expansions [11, Ch. II, §2.2]) in the case of a simple eigenvalue of a self-adjoint operator, which does not exist in Kato's monograph. Since $\lambda^{(0)}$ is a simple eigenvalue of $\mathcal{T}^{(0)}$, analytic-perturbation theory ensures there exists a unique eigenvalue branch $\lambda(x)$. Let

$$R(z) = (\mathcal{T}(x) - zI)^{-1} \tag{56}$$

be the *resolvent* of operator $\mathcal{T}(x)$, which encapsulates the full spectral information of $\mathcal{T}(x)$, and let

$$S(z) = R(z)(I - P(x)) \tag{57}$$

be the associated reduced resolvent $S(z)$ (*i.e.*, the regular part of the resolvent), where $P(x)$ is the Riesz–Dunford contour integral [5,11], that is:

$$P(x) = -\frac{1}{2\pi i} \oint_{\Gamma} R(z) \, dz = -\frac{1}{2\pi i} \oint_{\Gamma} \left(\mathcal{T}(x) - zI\right)^{-1} dz, \tag{58}$$

where Γ is a small contour enclosing only $\lambda^{(0)}$ and no other eigenvalues. Hence the projection:

$$\mathcal{T}(x) \, P(x) = -\frac{1}{2\pi i} \oint_{\Gamma} \left[I + zR(z)\right] dz = \frac{1}{2\pi i} \oint_{\Gamma} -zR(z) dz = \lambda(x) \, P(x), \tag{59}$$

holds true since the resolvent $R(z)$ for a simple eigenvalue admits the Laurent expansion [11, Ch. I, §5.3, Eq (5.18)]

$$R(z) = \frac{P(x)}{\lambda(x) - z} + S(z) + \hat{R}(z), \quad \hat{R}(z) = \sum_{n=1}^{\infty} \left(z - \lambda(x)\right)^n \left(S(z)\right)^{n+1}, \tag{60}$$

where $\hat{R}(z)$ is the analytic regular remainder of z, and the residue of $-zR(z)$ at $z = \lambda(x)$ is

$$\mathrm{Res}_{z=\lambda(x)}(-zR(z)) = [z-\lambda(x)][-zR(z)]\mid_{z=\lambda(x)} = \lambda(x)P(x). \tag{61}$$

Contour Integral of Perturbed Eigenvalue Series $\lambda(x)$. Starting from the identity in Equation (59),

$$\begin{aligned}
\mathcal{T}(x)P(x) &= \lambda(x)P(x) &(62)\\
\Longrightarrow \mathrm{tr}\big(\mathcal{T}(x)P(x)\big) &= \mathrm{tr}\big(\lambda(x)P(x)\big) &(63)\\
&= \mathrm{tr}\big(\lambda(x)\big)\ \mathrm{tr}\big(P(x)\big) &(64)\\
&= \lambda(x) &(65)\\
\Longrightarrow \lambda(x)-\lambda^{(0)} &= \mathrm{tr}\big(\mathcal{T}(x)P(x)\big)-\lambda^{(0)} &(66)\\
&= \mathrm{tr}\big(\mathcal{T}(x)P(x)\big)-\mathrm{tr}\big(\lambda^{(0)}P(x)\big) &(67)\\
&= \mathrm{tr}\Big(\big(\mathcal{T}(x)-\lambda^{(0)}I\big)P(x)\Big), &(68)
\end{aligned}$$

then substituting $P(x)$ from Equation (59) yields:

$$\begin{aligned}
\lambda(x)-\lambda^{(0)} &= \mathrm{tr}\Big(\big(\mathcal{T}(x)-\lambda^{(0)}I\big)\big(-\frac{1}{2\pi i}\oint_\Gamma R(z)\,dz\big)\Big) &(69)\\
&= -\frac{1}{2\pi i}\oint_\Gamma \mathrm{tr}\left((T(x)-\lambda^{(0)}I)R(z)\right)dz. &(70)
\end{aligned}$$

Considering the resolvent identity:

$$(\mathcal{T}(x)-zI)R(z)=I \tag{71}$$

and:

$$\begin{aligned}
\big(\mathcal{T}(x)-\lambda^{(0)}I\big)R(z) &= \Big(\mathcal{T}(x)-zI+zI-\lambda^{(0)}I\Big)R(z) &(72)\\
&= \Big(\mathcal{T}(x)-zI+(z-\lambda^{(0)})I\Big)R(z) &(73)\\
&= (\mathcal{T}(x)-zI)R(z)+(z-\lambda^{(0)})R(z) &(74)\\
&= I+(z-\lambda^{(0)})R(z), &(75)
\end{aligned}$$

it yields:

$$\begin{aligned}
\lambda(x)-\lambda^{(0)} &= -\frac{1}{2\pi i}\oint_\Gamma \mathrm{tr}\left((T(x)-\lambda^{(0)}I)R(z)\right)dz &(76)\\
&= -\frac{1}{2\pi i}\oint_\Gamma \mathrm{tr}\left(I+(z-\lambda^{(0)})R(z)\right)dz &(77)\\
&= -\frac{1}{2\pi i}\oint_\Gamma \mathrm{tr}\left((z-\lambda^{(0)})R(z)\right)dz &(78)\\
&= -\frac{1}{2\pi i}\oint_\Gamma \mathrm{tr}\left((z-\lambda^{(0)})R(z)\right)\mathrm{tr}\left(1\right)dz &(79)\\
&= -\frac{1}{2\pi i}\oint_\Gamma \mathrm{tr}\left((z-\lambda^{(0)})R(z)\right)\mathrm{tr}\left(P^{(0)}\right)dz &(80)
\end{aligned}$$

$$= -\frac{1}{2\pi i} \oint_\Gamma \operatorname{tr}\left((z-\lambda^{(0)})R(z)P^{(0)}\right) dz. \tag{81}$$

Resolvent Expansion. To expand the resolvent $R(z)$, define the unperturbed resolvent $R^{(0)}(z)$:

$$R^{(0)}(z) = \left(\mathcal{T}^{(0)} - zI\right)^{-1}, \tag{82}$$

then this identity holds:

$$(\mathcal{T}^{(0)} - zI)R^{(0)}(z) = I. \tag{83}$$

Note that:

$$\mathcal{T}(x) - zI = \mathcal{T}^{(0)} - zI + \sum_{j=1}^{\infty} x^j \mathcal{T}^{(j)} \tag{84}$$

$$= \mathcal{T}^{(0)} - zI + I\sum_{j=1}^{\infty} x^j \mathcal{T}^{(j)} \tag{85}$$

$$= \mathcal{T}^{(0)} - zI + (\mathcal{T}^{(0)} - zI)R^{(0)}(z)\sum_{j=1}^{\infty} x^j \mathcal{T}^{(j)} \tag{86}$$

$$= (\mathcal{T}^{(0)} - zI)\Big(I + R^{(0)}(z)\sum_{j=1}^{\infty} x^j \mathcal{T}^{(j)}\Big), \tag{87}$$

so that the operator Neumann series of $R(z)$ expands as:

$$R(z) = \left(\mathcal{T}(x) - zI\right)^{-1} \tag{88}$$

$$= \left[(\mathcal{T}^{(0)} - zI)\Big(I + R^{(0)}(z)\sum_{j=1}^{\infty} x^j \mathcal{T}^{(j)}\Big)\right]^{-1} \tag{89}$$

$$= \Big(I + R^{(0)}(z)\sum_{j=1}^{\infty} x^j \mathcal{T}^{(j)}\Big)^{-1}\Big(\mathcal{T}^{(0)} - zI\Big)^{-1} \tag{90}$$

$$= \Big(I + R^{(0)}(z)\sum_{j=1}^{\infty} x^j \mathcal{T}^{(j)}\Big)^{-1} R^{(0)}(z) \tag{91}$$

$$= \underbrace{\Big(I - \big(-R^{(0)}(z)\sum_{j=1}^{\infty} x^j \mathcal{T}^{(j)}\big)\Big)^{-1}}_{\text{Neumann series}} R^{(0)}(z) \tag{92}$$

$$= \sum_{k=0}^{\infty}(-1)^k \Big(R^{(0)}(z)\sum_{j=1}^{\infty} x^j \mathcal{T}^{(j)}\Big)^k R^{(0)}(z). \tag{93}$$

Asymptotic Eigenvalue Expansion. Expanding the term in Equation (93):

$$\Big(R^{(0)}(z)\sum_{j=1}^{\infty} x^j \mathcal{T}^{(j)}\Big)^k \tag{94}$$

yields:

$$\Big(R^{(0)}(z)\sum_{j=1}^{\infty}x^j\mathcal{T}^{(j)}\Big)^k = \sum_{i_1,\dots,i_k\geq 1} x^{\,i_1+\cdots+i_k}\, R^{(0)}(z)\,\mathcal{T}^{(i_1)}\,R^{(0)}(z)\cdots R^{(0)}(z)\,\mathcal{T}^{(i_k)} \tag{95}$$

$$= \sum_{i_1,\dots,i_k\geq 1} x^{\,i_1+\cdots+i_k}\, R^{(0)}(z)\,R_{i_k}(z), \tag{96}$$

where $R_{i_k}(z)$ represents an operator composition series:

$$R_{i_k}(z) = \mathcal{T}^{(i_1)}\,R^{(0)}(z)\cdots R^{(0)}(z)\,\mathcal{T}^{(i_k)}. \tag{97}$$

Substituting Equation (96) into Equation (93) yields:

$$R(z) = \sum_{k=0}^{\infty}(-1)^k \sum_{i_1,\dots,i_k\geq 1} x^{\,i_1+\cdots+i_k}\, R^{(0)}(z)R_{i_k}(z)R^{(0)}(z) \tag{98}$$

$$= \sum_{k=0}^{\infty}(-1)^k \sum_{i_1,\dots,i_k\geq 1} x^{\,i_1+\cdots+i_k}\, R^{(0)}(z)R_{i_k}(z)R^{(0)}(z). \tag{99}$$

Substituting Equation (99) into the contour integral for $\lambda(x)-\lambda^{(0)}$ in Equation (81):

$$\lambda(x)-\lambda^{(0)} = -\frac{1}{2\pi i}\oint_{\Gamma} \operatorname{tr}\left((z-\lambda^{(0)})R(z)P^{(0)}\right)dz, \tag{100}$$

yields:

$$\lambda^{(n)} = \sum_{k=0}^{\infty}(-1)^k \sum_{\substack{i_1+\cdots+i_k=n\\ i_j\geq 1}} -\frac{1}{2\pi i}\oint_{\Gamma} \operatorname{tr}\Big((z-\lambda^{(0)})R^{(0)}(z)R_{i_k}(z)R^{(0)}(z)P^{(0)}\Big)dz \tag{101}$$

$$= \sum_{k=0}^{\infty}(-1)^{k-1} \sum_{\substack{i_1+\cdots+i_k=n\\ i_j\geq 1}} \frac{1}{2\pi i}\oint_{\Gamma} \operatorname{tr}\Big((z-\lambda^{(0)})R^{(0)}(z)R_{i_k}(z)R^{(0)}(z)P^{(0)}\Big)dz, \tag{102}$$

which recovers Theorem 3.2 (Kato's Weighted Mean of Eigenvalue Expansions [11, Ch. II, §2.2]).

Contracting and Relabeling Indices. We refine Theorem 3.2 (Kato's Weighted Mean of Eigenvalue Expansions [11, Ch. II, §2.2]) further, with the aim of obtaining a constructive, computable and basis-dependent formulation. Note that only the multi-indices satisfying:

$$i_1+\cdots+i_k = n \tag{103}$$

contribute to the coefficient of x^n, and since the terms with:

$$k=0 \quad \text{or} \quad k>n \tag{104}$$

for $n\geq 1$ vanish, we contract the summation to the admissible subset of indices. For clarity, we denote this contracted index set by $p\subseteq k$:

$$\lambda^{(n)} = \sum_{p=1}^{n}(-1)^{p-1} \sum_{\substack{i_1+\cdots+i_p=n\\ i_j\geq 1}} \frac{1}{2\pi i}\oint_{\Gamma} \operatorname{tr}\Big((z-\lambda^{(0)})R^{(0)}(z)R_{i_p}(z)R^{(0)}(z)\,P^{(0)}\Big)dz. \tag{105}$$

Applying Cauchy's Residue Theorem Applying Cauchy's residue theorem via Riesz–Dunford functional calculus [5] for Equation (105) yields:

$$\lambda^{(n)} = \sum_{p=1}^{n} (-1)^{p-1} \sum_{\substack{i_1+\cdots+i_p=n \\ i_j \geq 1}} \mathrm{Res}_{z=\lambda^{(0)}} \left[\mathrm{tr}\Big((z-\lambda^{(0)}) R^{(0)}(z) R_{i_p}(z) R^{(0)}(z) P^{(0)} \Big) \right]. \tag{106}$$

Simplifying Residue. We now aim to compute the residue:

$$\mathrm{Res}_{z=\lambda^{(0)}} \left[\mathrm{tr}\Big((z-\lambda^{(0)}) R^{(0)}(z) R_{i_p}(z) R^{(0)}(z) P^{(0)} \Big) \right]. \tag{107}$$

Note that near $z = \lambda^{(0)}$, the Laurent expansion of the unperturbed resolvent $R^{(0)}(z)$ for a simple eigenvalue $\lambda^{(0)}$ admits [11, Ch. I, §5.3, Eq (5.18)]:

$$R^{(0)}(z) = \frac{P^{(0)}}{\lambda^{(0)} - z} + S^{(0)} + \hat{R}^{(0)}(z), \qquad \hat{R}^{(0)}(z) = \sum_{n=1}^{\infty} \left(z - \lambda^{(0)}\right)^n \left(S^{(0)}\right)^{n+1}, \tag{108}$$

where $P^{(0)} = w^{(0)}(w^{(0)})^*$, $S^{(0)} = R^{(0)}(z)(I - P^{(0)})$ is the unperturbed reduced resolvent of $R^{(0)}(z)$, and $\hat{R}^{(0)}(z)$ is the analytic regular remainder of z. Substitute $R^{(0)}(z)$ into the trace product:

$$\mathrm{tr}\Big((z-\lambda^{(0)}) R^{(0)}(z) R_{i_p}(z) R^{(0)}(z) P^{(0)} \Big), \tag{109}$$

and consider that in the expanded trace product:

1. the terms $P^{(0)}S^{(0)} = S^{(0)}P^{(0)} = 0$ vanish,
2. the terms with higher-order poles vanish, since the denominators are constant operators.

Then only the term with simple pole survives:

$$\mathrm{Res}_{z=\lambda^{(0)}} \left[\mathrm{tr}\Big((z-\lambda^{(0)}) R^{(0)}(z) R_{i_p}(z) R^{(0)}(z) P^{(0)} \Big) \right]$$

$$= \mathrm{Res}_{z=\lambda^{(0)}} \left[\mathrm{tr}\Big((z-\lambda^{(0)}) \frac{P^{(0)}\mathcal{T}^{(i_1)}S^{(0)} \cdots S^{(0)}\mathcal{T}^{(i_p)}P^{(0)}}{(\lambda^{(0)} - z)^2} P^{(0)} \Big) \right] \tag{110}$$

$$= \mathrm{Res}_{z=\lambda^{(0)}} \left[\mathrm{tr}\Big(\frac{P^{(0)}\mathcal{T}^{(i_1)}S^{(0)} \cdots S^{(0)}\mathcal{T}^{(i_p)}P^{(0)}}{z - \lambda^{(0)}} P^{(0)} \Big) \right] \tag{111}$$

$$= (z-\lambda^{(0)})\, \mathrm{tr}\Big(\frac{P^{(0)}\mathcal{T}^{(i_1)}S^{(0)} \cdots S^{(0)}\mathcal{T}^{(i_p)}P^{(0)}}{z - \lambda^{(0)}} P^{(0)} \Big) \,\Big|_{z\to\lambda^{(0)}} \tag{112}$$

$$= \mathrm{tr}\Big(P^{(0)}\mathcal{T}^{(i_1)}S^{(0)} \cdots S^{(0)}\mathcal{T}^{(i_p)}P^{(0)}P^{(0)} \Big) \tag{113}$$

$$= \mathrm{tr}\Big(\mathcal{T}^{(i_1)}S^{(0)} \cdots S^{(0)}\mathcal{T}^{(i_p)}P^{(0)}P^{(0)}P^{(0)} \Big) \tag{114}$$

$$= \mathrm{tr}\Big(\mathcal{T}^{(i_1)}S^{(0)} \cdots S^{(0)}\mathcal{T}^{(i_p)}P^{(0)} \Big), \tag{115}$$

since $P^{(0)}P^{(0)} = P^{(0)}$.

Producing Theorem Claim. Substituting the residue from Equation (115) into Equation (106) yields:

$$\lambda^{(n)} = \sum_{p=1}^{n} (-1)^{p-1} \sum_{\substack{i_1+\cdots+i_p=n \\ i_j \geq 1}} \mathrm{tr}\Big(\mathcal{T}^{(i_1)}S^{(0)} \cdots S^{(0)}\mathcal{T}^{(i_p)}P^{(0)} \Big) \tag{116}$$

$$= \sum_{p=1}^{n} (-1)^{p-1} \sum_{\substack{i_1+\cdots+i_p=n \\ i_j \geq 1}} \operatorname{tr}\Big(\mathcal{T}^{(i_1)} S^{(0)} \cdots S^{(0)} \mathcal{T}^{(i_p)} w^{(0)} (w^{(0)})^*\Big) \tag{117}$$

$$= \sum_{p=1}^{n} (-1)^{p-1} \sum_{\substack{i_1+\cdots+i_p=n \\ i_j \geq 1}} \operatorname{tr}\Big(\big(w^{(0)}\big)^* \mathcal{T}^{(i_1)} S^{(0)} \cdots S^{(0)} \mathcal{T}^{(i_p)} w^{(0)}\Big) \tag{118}$$

$$= \sum_{p=1}^{n} (-1)^{p-1} \sum_{\substack{i_1+\cdots+i_p=n \\ i_j \geq 1}} \Big\langle w^{(0)}, \mathcal{T}^{(i_1)} S^{(0)} \mathcal{T}^{(i_2)} S^{(0)} \cdots S^{(0)} \mathcal{T}^{(i_p)} w^{(0)} \Big\rangle, \tag{119}$$

which is basis-dependent and expressed in terms of unperturbed reduced resolvent. □

Remark 3.5. By refining Theorem 3.2 (Kato's Weighted Mean of Eigenvalue Expansions [11, Ch. II, §2.2]), Theorem 3.3 (Refined Closed-Form Asymptotic Expansion of Simple Isolated Eigenvalue in Self-Adjoint Operator) provides an *explicit, closed-form representation* of the coefficients $\lambda^{(n)}$ in the eigenvalue perturbation series. Classical analytic perturbation theory [11] guarantees the existence of such expansions and gives recursive characterizations of the coefficients, but does not furnish constructive closed forms. In contrast, our formulation expresses each $\lambda^{(n)}$ in terms of finitely many operator products involving the perturbation operators $\mathcal{T}^{(j)}$ and the unperturbed reduced resolvent $S^{(0)}$, making the coefficients directly computable. As a validation, for $n = 1, 2, \ldots$, the expansion specializes to the familiar Rayleigh–Schrödinger corrections of quantum mechanics [23,27,26].

4. Infinitesimal higher-order spectral variations

Guided by the scheme in Fig. 1, and under Assumption 2.2 (Simplicity Assumption of Non-Zero Singular Values), we exploit the spectral correspondence between a rectangular matrix $A \in \mathbb{R}^{m\times n}$ and its Jordan–Wielandt embedding $\mathcal{T}$ as established in Theorem 1.2 (Spectrum of Jordan–Wielandt Embedding). This allows us to derive arbitrary–order Fréchet derivatives of the singular values of A from the asymptotic eigenvalue expansions of $\mathcal{T}$. The argument proceeds by first establishing the correspondence between the perturbation series and Fréchet derivatives as stated in Theorem 4.1 (Analytic Perturbation for Holomorphic Operators), and then applying this relation to obtain higher–order derivatives of singular values as stated in Theorem 4.2 (Higher-Order Infinitesimal Spectral Variation).

Theorem 4.1 *(Analytic Perturbation for Holomorphic Operators). Let X be a Banach space and let $\mathcal{T}(x) : U \subset \mathbb{C} \to \mathcal{L}(X)$ be a holomorphic family —* i.e., *type (A) in the sense of Kato's framework [11], defined in a neighborhood of* 0 *in the operator norm. Then $\mathcal{T}$ is C^∞ in the Fréchet sense at* 0 *and admits the convergent operator–norm expansion:*

$$\mathcal{T}(x) \; = \; \sum_{n=0}^{\infty} \frac{x^n}{n!} \, D^n \mathcal{T}(0), \qquad |x| < \rho, \tag{120}$$

where ρ is the distance from 0 *to the boundary of U. In particular, if one writes the perturbation series as:*

$$\mathcal{T}(x) = \mathcal{T}^{(0)} + \sum_{n=1}^{\infty} x^n \, \mathcal{T}^{(n)}, \tag{121}$$

then the coefficients agree with the Fréchet derivatives, namely:

$$\mathcal{T}^{(n)} = \tfrac{1}{n!} \, D^n \mathcal{T}(0), \qquad n \geq 1. \tag{122}$$

Theorem 4.2 *(Higher-Order Infinitesimal Spectral Variation). Let*

$$A \in \mathbb{R}^{m \times n} \tag{123}$$

be a real rectangular matrix under Assumption 2.2 (Simplicity Assumption of Non-Zero Singular Values).

Matrix Perturbation Series. *Let*

$$A(x) = \sum_{k=0}^{\infty} x^k A^{(k)} \in \mathbb{R}^{m \times n} \tag{124}$$

be holomorphic perturbed operator near $x = 0$ with unperturbed matrix $A^{(0)} = A$.

The unperturbed matrix $A^{(0)}$ admits a full SVD:

$$A^{(0)} = U^{(0)} \Sigma^{(0)} (V^{(0)})^T, \tag{125}$$

as defined in Theorem 1.1 (Matrix Singular Value Decomposition (Full Form)), where ordered $r = \operatorname{rank}(A)$ non-zero singular values are given, under Assumption 2.2 (Simplicity Assumption of Non-Zero Singular Values), as:

$$\sigma_1^{(0)} > \sigma_2^{(0)} > \ldots > \sigma_r^{(0)} > 0, \tag{126}$$

$u_k^{(0)}$ and $v_k^{(0)}$ are singular vectors associated with singular value $\sigma_k^{(0)}$, and:

$$U^{(0)} \in \mathbb{R}^{m \times m}, \quad V^{(0)} \in \mathbb{R}^{n \times n} \tag{127}$$

are orthogonal matrices. For brevity, we also use $\sigma_i = \sigma_i^{(0)}$, $u_i = u_i^{(0)}$, and $v_i = v_i^{(0)}$.

Jordan–Wielandt Perturbation Series Embedding. *Using Theorem 1.2 (Spectrum of Jordan–Wielandt Embedding), we embed the perturbation series $A(x)$ into $\mathcal{T}(x)$ to construct a Jordan–Wielandt embedding:*

$$\mathcal{T}(x) = \begin{bmatrix} 0 & A(x) \\ A(x)^\top & 0 \end{bmatrix}. \tag{128}$$

This embedding admits a perturbation series:

$$\mathcal{T}(x) = \sum_{j=0}^{\infty} x^j \mathcal{T}^{(j)} \tag{129}$$

at x near zero, with the unperturbed operator:

$$\mathcal{T}^{(0)} = \begin{bmatrix} 0 & A^{(0)} \\ (A^{(0)})^\top & 0 \end{bmatrix}, \tag{130}$$

and the perturbations:

$$\mathcal{T}^{(j)} = \begin{bmatrix} 0 & A^{(j)} \\ (A^{(j)})^\top & 0 \end{bmatrix}, \quad j \geq 1. \tag{131}$$

The non-zero eigenvalues $\lambda_i^{(\pm 0)}$ *of* $\mathcal{T}^{(0)}$ *are therefore:*

$$\lambda_i^{(+0)} = +\sigma_i^{(0)}, \quad \lambda_i^{(-0)} = -\sigma_i^{(0)}, \quad \text{for } i = 1, \dots, r, \tag{132}$$

associated with eigenvectors:

$$w_i^{(+0)} = \frac{1}{\sqrt{2}} \begin{bmatrix} u_i^{(0)} \\ v_i^{(0)} \end{bmatrix}, \quad w_i^{(-0)} = \frac{1}{\sqrt{2}} \begin{bmatrix} u_i^{(0)} \\ -v_i^{(0)} \end{bmatrix}, \tag{133}$$

and null eigenvectors,

$$a_j^{(0)} = \begin{bmatrix} u_j^{(0)} \\ 0 \end{bmatrix}, \quad (\text{for } j = r+1, \cdots, m), \tag{134}$$

and:

$$b_k^{(0)} = \begin{bmatrix} 0 \\ v_k^{(0)} \end{bmatrix}, \quad (\text{for } j = r+1, \cdots, n). \tag{135}$$

Since $w_k^{(+0)}$ *and* $w_k^{(-0)}$ *are eigenvalues of* $T^{(0)}$*, hence the identities hold:*

$$\mathcal{T}^{(0)} w_k^{(+0)} = \lambda_k^{(+0)} w_k^{(+0)} \Longrightarrow \mathcal{T}^{(0)} w_k^{(+0)} = \sigma_k^{(0)} w_k^{(+0)}, \tag{136}$$

and:

$$\mathcal{T}^{(0)} w_k^{(-0)} = \lambda_k^{(-0)} w_k^{(-0)} \Longrightarrow \mathcal{T}^{(0)} w_k^{(-0)} = -\sigma_k^{(0)} w_k^{(-0)}. \tag{137}$$

Unperturbed Reduced Resolvent in Embedding. *By definition, the spectral expansion of the reduced resolvent operator associated with the eigenvalue* $\lambda_k^{(+0)} = \sigma_k^{(0)}$ *and associated eigenvector* $w_k^{(+0)}$ *of* $\mathcal{T}(x)$ *is given as:*

$$S_k^{(0)} = \left(\mathcal{T}^{(0)} - \sigma_k^{(0)} I\right)^{-1} \left(I - P_k^{(0)}\right), \quad P_k^{(0)} = w_k^{(+0)} (w_k^{(+0)})^\top, \tag{138}$$

which admits the spectral expansion:

$$S_k^{(0)} = \sum_{i=1,\ i \neq k}^{r} \frac{w_i^{(+0)} (w_i^{(+0)})^\top}{\sigma_i^{(0)} - \sigma_k^{(0)}} + \sum_{i=1, i\neq k}^{r} \frac{w_i^{(-0)} (w_i^{(-0)})^\top}{-\sigma_i^{(0)} - \sigma_k^{(0)}} - \sum_{j=r+1}^{m} \frac{a_j^{(0)} (a_j^{(0)})^\top}{\sigma_k^{(0)}} - \sum_{j=r+1}^{n} \frac{b_j^{(0)} (b_j^{(0)})^\top}{\sigma_k^{(0)}}. \tag{139}$$

Theorem Claim. *For each integer* $n \geq 1$,

$$\sigma_k^{(n)} = \sum_{p=1}^{n} (-1)^{p-1} \sum_{\substack{i_1 + \cdots + i_p = n \\ i_j \geq 1}} \langle w_k^{(+0)}, \mathcal{T}^{(i_1)} S_k^{(0)} \mathcal{T}^{(i_2)} \cdots S_k^{(0)} \mathcal{T}^{(i_p)} w_k^{(+0)} \rangle. \tag{140}$$

By Theorem 4.1 (Analytic Perturbation for Holomorphic Operators), the Fréchet derivative of the singular value is unique and given by:

$$D^n \sigma_k[\mathrm{d}A, \cdots, \mathrm{d}A] = n! \lim_{x \to 0} \left(x^n\, \sigma_k^{(n)}\right), \tag{141}$$

where:

$$\mathrm{d}A = \lim_{x \to 0} x A^{(1)}. \tag{142}$$

Remark 4.3 *(Schematic Procedure of Computing Higher-Order Singular-Value Derivatives).* The suggested schematic procedure of computing arbitrary higher-order singular-value derivatives is as follows:

- **Procedure I – Construct Infinitesimal Perturbation** — constructs an infinitesimal perturbation by $\mathrm{d}A = \lim_{x\to 0} xA^{(1)}$,
- **Procedure II – Specialize n to Obtain Derivative Operator** — specialize n in $\sigma^{(n)}$, and obtain derivative operator $D^n\sigma_k = n!\sigma^{(n)}$,
- **Procedure III – Map Derivative Operator Layout** — map $D^n\sigma_k[\mathrm{d}A, \cdots, \mathrm{d}A]$ to Kronecker-product representation or specific layout.

Proof 4.4. By Theorem 3.3 (Refined Closed-Form Asymptotic Expansion of Simple Isolated Eigenvalue in Self-Adjoint Operator), one eigenvalue $\lambda_k(x)$ of $\mathcal{T}(x)$ admits an asymptotic expansion:

$$\lambda_k(x) = \sum_{n=0}^{\infty} x^n \lambda_k^{(n)}. \tag{143}$$

By Theorem 1.2 (Spectrum of Jordan–Wielandt Embedding), for $n \geq 1$, choosing a positive eigenvalue branch $\sigma_k^{(0)}$ yields the asymptotic singular-value expansion of $A(x)$:

$$\sigma_k^{(n)} = \sum_{p=1}^{n} (-1)^{p-1} \sum_{\substack{i_1+\cdots+i_p=n \\ i_j\geq 1}} \langle w_k^{(+0)}, \mathcal{T}^{(i_1)} S_k^{(0)} \mathcal{T}^{(i_2)} \cdots S_k^{(0)} \mathcal{T}^{(i_p)} w_k^{(+0)} \rangle. \tag{144}$$

By Theorem 4.1 (Analytic Perturbation for Holomorphic Operators), $D^n\sigma_k$ admits:

$$D^n\sigma_k = n!\sigma_k^{(n)}, \tag{145}$$

and its action is given by:

$$D^n\sigma_k[\mathrm{d}A, \cdots, \mathrm{d}A] = n! \lim_{x\to 0} x^n \sigma_k^{(n)}, \tag{146}$$

where

$$\mathrm{d}A = \lim_{x\to 0} xA^{(1)}. \quad \square \tag{147}$$

Remark 4.5. Thanks to Kato's perturbation theory for linear operators, our framework for deriving singular-value derivatives rests on a rigorous analytic foundation and provides a procedural and systematic methodology, resting on a rigorous foundation, and going beyond the *ad hoc* approaches commonly found in classical matrix analysis. In the latter, derivatives are typically obtained through differential identities or perturbation arguments without a fully rigorous treatment of differentiability. For instance, Horn and Johnson [9] present differential identities for spectral functions, but these do not constitute a unified framework for higher-order derivatives.

5. Special case ($n = 1$): closed-form singular-value Jacobian

We now show that Theorem 4.2 (Higher-Order Infinitesimal Spectral Variation) can recover the well-known singular-value Jacobian, stated in Lemma 5.1 [30,17].

Lemma 5.1 *(Closed-Form Singular-Value Jacobian). The Jacobian of a singular value is well-known in the literature [30,17] in the form:*

$$\frac{\partial \sigma_k}{\partial A} = u_k\, v_k^\top, \tag{148}$$

which immediately admits an equivalent result with Kronecker-product presentation:

$$D\sigma_k[\mathrm{d}A] = (v_k \otimes u_k)^\top \mathrm{vec}(\mathrm{d}A). \tag{149}$$

Traditional Method in Matrix Analysis. In classical matrix analysis [30,17], the derivation of singular-value derivatives often begins with the identity

$$\sigma_k = u_k^\top A v_k, \tag{150}$$

and then applies the trace identity

$$\sigma_k = \mathrm{tr}(u_k^\top A v_k), \tag{151}$$

to compute $\mathrm{d}\sigma_k$ and its derivatives. However, this approach is largely ad hoc and does not scale systematically to higher-order derivatives or more general operator settings.

Proof 5.2. We follow the schematic procedure suggested by Remark 4.3 (Schematic Procedure of Computing Higher-Order Singular-Value Derivatives) to recover this first-order singular-value Jacobian by specializing $n = 1$ in Theorem 4.2 (Higher-Order Infinitesimal Spectral Variation).

Procedure I – Construct Infinitesimal Perturbation. We construct a perturbation series on A:

$$A(x) = A + xA^{(1)}, \quad A^{(0)} = A, \tag{152}$$

so it yields:

$$\mathrm{d}A = \lim_{x\to 0} xA^{(1)}. \tag{153}$$

Procedure II – Specialize n to Obtain Derivative Operator. Specializing $n = 1$ in Theorem 4.2 (Higher-Order Infinitesimal Spectral Variation) yields:

$$\sigma_k^{(1)} = \langle w_k^{(+0)}, \mathcal{T}^{(1)} w_k^{(0)} \rangle. \tag{154}$$

Simplifying First-Order Term. Consider:

$$\mathcal{T}^{(1)} w_k^{(+0)} = \begin{bmatrix} 0 & A^{(1)} \\ (A^{(1)})^\top & 0 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} u_k \\ v_k \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix} \tag{155}$$

$$\Longrightarrow \langle w_k^{(+0)}, \mathcal{T}^{(1)} w_k^{(+0)} \rangle = \left(\frac{1}{\sqrt{2}} \begin{bmatrix} u_k \\ v_k \end{bmatrix} \right)^\top \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix} \tag{156}$$

$$\Longrightarrow \langle w_k^{(+0)}, \mathcal{T}^{(1)} w_k^{(+0)} \rangle = \frac{1}{2} \left[u_k^\top A^{(1)} v_k + v_k^\top (A^{(1)})^\top u_k \right] \tag{157}$$

$$\Longrightarrow \langle w_k^{(+0)}, \mathcal{T}^{(1)} w_k^{(+0)} \rangle = u_k^\top A^{(1)} v_k \tag{158}$$

$$\Longrightarrow \sigma_k^{(1)} = u_k^\top A^{(1)} v_k. \tag{159}$$

Procedure III – Map Derivative Operator Layout. By Theorem 4.1 (Analytic Perturbation for Holomorphic Operators), we have:

$$D\sigma_k[\mathrm{d}A] = \sigma_k^{(1)}[\mathrm{d}A] \tag{160}$$
$$= \lim_{x\to 0} \sigma_k^{(1)} x A^{(1)} \tag{161}$$
$$= \lim_{x\to 0} u_k^T x A^{(1)} v_k \tag{162}$$
$$= u_k^\top \mathrm{d}A v_k \in \mathbb{R}. \tag{163}$$

Using following identities from Lemma 2.1 (Essential Matrix Identities):

1. $\mathrm{tr}(x) = x$,
2. $\mathrm{vec}(BVA^\top) = (A \otimes B)\,\mathrm{vec}(V)$,
3. $(A \otimes B)^\top = A^\top \otimes B^\top$,
4. $\mathrm{tr}(ABC) = \mathrm{tr}(CAB) = \mathrm{tr}(BCA)$,

yields:

$$D\sigma_k[\mathrm{d}A] = \mathrm{tr}(u_k^\top \mathrm{d}A v_k) \tag{164}$$
$$= \mathrm{tr}(v_k u_k^\top \mathrm{d}A) \tag{165}$$
$$= (v_k \otimes u_k)^\top \mathrm{vec}(\mathrm{d}A), \tag{166}$$

and:

$$\frac{\partial \sigma_k}{\partial A} = (D\sigma_k)^\top = u_k v_k^\top. \quad \square \tag{167}$$

6. Special case ($n = 2$): closed-form singular-value Hessian

Explicit closed-form expressions for the singular-value Hessian of rectangular matrices are, to the best of our knowledge, not available in the literature. Such a result is essential for applications in stochastic analysis, for example when applying Itô's lemma to stochastic differential equations (SDEs) or stochastic partial differential equations (SPDEs) driven by Wiener processes [21]. We now derive the singular-value Hessian for general real rectangular matrices, under Assumption 2.2 (Simplicity Assumption of Non-Zero Singular Values), as stated in Lemma 6.1 (Closed-Form Singular-Value Hessian), represented in the layout:

$$\mathrm{vec}(\mathrm{d}A)^\top \left(\frac{\partial}{\partial\,\mathrm{vec}(A)} \mathrm{vec}\left(\frac{\partial \sigma_k}{\partial A} \right) \right) \mathrm{vec}(\mathrm{d}A), \tag{168}$$

by specializing Theorem 4.2 (Higher-Order Infinitesimal Spectral Variation) to the case $n = 2$.

Lemma 6.1 *(Closed-Form Singular-Value Hessian). The Hessian of a singular value is given as:*

$$\frac{\partial}{\partial\,\mathrm{vec}(A)} \mathrm{vec}\left(\frac{\partial \sigma_k}{\partial A} \right) = \underbrace{\sum_{i\neq k, i\leq m} \frac{\sigma_k}{\sigma_k^2 - \sigma_i^2} (v_k \otimes u_i)(v_k \otimes u_i)^\top}_{left} + \tag{169}$$

$$\underbrace{\sum_{j\neq k, j\leq n} \frac{\sigma_k}{\sigma_k^2 - \sigma_j^2} (v_j \otimes u_k)(v_j \otimes u_k)^\top}_{right} + \tag{170}$$

$$\underbrace{\sum_{l\neq k, l\leq r} \frac{\sigma_l}{\sigma_k^2-\sigma_l^2}\left[(v_k\otimes u_l)(v_l\otimes u_k)^\top + (v_l\otimes u_k)(v_k\otimes u_l)^\top\right]}_{\textit{left-right interaction}} \tag{171}$$

with Kronecker-product representation.

Proof 6.2. We follow the schematic procedure suggested by Remark 4.3 (Schematic Procedure of Computing Higher-Order Singular-Value Derivatives) to derive this second-order singular-value Hessian by specializing $n = 2$ in Theorem 4.2 (Higher-Order Infinitesimal Spectral Variation).

Procedure I – Construct Infinitesimal Perturbation. We construct a perturbation series on A:

$$A(x) = A + xA^{(1)}, \quad A^{(0)} = A, \tag{172}$$

so it yields:

$$\mathrm{d}A = \lim_{x\to 0} xA^{(1)}. \tag{173}$$

Procedure II – Specialize n to Obtain Derivative Operator. Specializing $n = 2$ in Theorem 4.2 (Higher-Order Infinitesimal Spectral Variation) yields:

$$\sigma_k^{(2)} = \sum_{p=1}^{2}(-1)^{p-1}\sum_{\substack{i_1+\cdots+i_p=2\\ i_j\geq 1}} \langle w_k^{(+0)}, \mathcal{T}^{(i_1)}S_k^{(0)}\mathcal{T}^{(i_2)}\cdots S_k^{(0)}\mathcal{T}^{(i_p)}w_k^{(+0)}\rangle \tag{174}$$

$$= \sigma_k^{(2,p=1)} + \sigma_k^{(2,p=2)} \tag{175}$$

where

$$\sigma_k^{(2,p=1)} := \langle w_k^{(+0)}, \mathcal{T}^{(2)}w_k^{(+0)}\rangle, \tag{176}$$

and

$$\sigma_k^{(2,p=2)} := -\langle w_k^{(+0)}, \mathcal{T}^{(1)}S_k^{(0)}\mathcal{T}^{(1)}w_k^{(+0)}\rangle. \tag{177}$$

Computing Term $\sigma_k^{(2,p=1)}$. We first compute the term with $p = 1$ ($\sigma_k^{(2,p=1)}$). By Theorem 4.2 (Higher-Order Infinitesimal Spectral Variation), we substitute:

$$\mathcal{T}^{(2)} = \begin{bmatrix} 0 & A^{(2)} \\ (A^{(2)})^\top & 0 \end{bmatrix} \quad \text{and} \quad w_k^{(+0)} = \frac{1}{\sqrt{2}}\begin{bmatrix} u_k \\ v_k \end{bmatrix} \tag{178}$$

into:

$$\mathcal{T}^{(2)}w_k^{(+0)}, \tag{179}$$

it yields:

$$\mathcal{T}^{(2)}w_k^{(+0)} = \begin{bmatrix} 0 & A^{(2)} \\ (A^{(2)})^\top & 0 \end{bmatrix}\cdot\frac{1}{\sqrt{2}}\begin{bmatrix} u_k \\ v_k \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} A^{(2)}v_k \\ (A^{(2)})^\top u_k \end{bmatrix}. \tag{180}$$

Substituting

$$\mathcal{T}^{(2)} w_k^{(+0)} = \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(2)} v_k \\ (A^{(2)})^\top u_k \end{bmatrix} \tag{181}$$

into:

$$\sigma_k^{(2,p=1)} = \langle w_k^{(+0)}, \mathcal{T}^{(2)} w_k^{(+0)} \rangle \tag{182}$$

yields:

$$\sigma_k^{(2,p=1)} = \langle w_k^{(+0)}, \mathcal{T}^{(2)} w_k^{(+0)} \rangle \tag{183}$$

$$= \frac{1}{2} \left[u_k^\top A^{(2)} v_k + (v_k)^\top (A^{(2)})^\top u_k \right] \tag{184}$$

$$= u_k^\top A^{(2)} v_k \tag{185}$$

In the construction of $\mathrm{d}A$, there is:

$$A^{(2)} = O, \tag{186}$$

so that:

$$\sigma_k^{(2,p=1)} = u_k^\top A^{(2)} v_k = 0. \tag{187}$$

Sketch of Computing Term $\sigma_k^{(2,p=2)}$. We compute the term with $p = 2$:

$$\sigma_k^{(2,p=2)} = -\langle w_k^{(+0)}, \mathcal{T}^{(1)} S_k^{(0)} \mathcal{T}^{(1)} w_k^{(+0)} \rangle. \tag{188}$$

To simplify, we first compute:

$$S_k^{(0)} \mathcal{T}^{(1)} w_k^{(+0)}, \tag{189}$$

then substitute this result into Equation (188) to produce complete $\sigma_k^{(2,p=2)}$.

Computing Contributions in $S_k^{(0)} \mathcal{T}^{(1)} w_k^{(+0)}$ in $\sigma_k^{(2,p=2)}$. By Theorem 4.2 (Higher-Order Infinitesimal Spectral Variation), the unperturbed reduced resolvent is defined as:

$$S_k^{(0)} = \sum_{i=1, i\neq k}^{r} \frac{w_i^{(+0)} (w_i^{(+0)})^\top}{\sigma_i^{(0)} - \sigma_k^{(0)}} + \sum_{i=1, i\neq k}^{r} \frac{w_i^{(-0)} (w_i^{(-0)})^\top}{-\sigma_i^{(0)} - \sigma_k^{(0)}} - \sum_{j=r+1}^{m} \frac{a_j^{(0)} (a_j^{(0)})^\top}{\sigma_k^{(0)}} - \sum_{j=r+1}^{n} \frac{b_j^{(0)} (b_j^{(0)})^\top}{\sigma_k^{(0)}}. \tag{190}$$

By Theorem 4.2 (Higher-Order Infinitesimal Spectral Variation), substituting non-null eigenvectors of the unperturbed embedding $\mathcal{T}^{(0)}$:

$$w_i^{(+0)} = \frac{1}{\sqrt{2}} \begin{bmatrix} u_i \\ v_i \end{bmatrix}, \quad w_i^{(-0)} = \frac{1}{\sqrt{2}} \begin{bmatrix} u_i \\ -v_i \end{bmatrix}, \tag{191}$$

and null eigenvectors of the unperturbed embedding $\mathcal{T}^{(0)}$:

$$a_j^{(0)} = \begin{bmatrix} u_j \\ 0 \end{bmatrix}, \quad b_j^{(0)} = \begin{bmatrix} 0 \\ v_j \end{bmatrix}, \tag{192}$$

into $S_k^{(0)}$ yields:

$$\begin{aligned} S_k^{(0)} = & \underbrace{\sum_{i=1, i\neq k}^{r} \frac{1}{2} \frac{\begin{bmatrix} u_i \\ v_i \end{bmatrix} \begin{bmatrix} u_i^\top & v_i^\top \end{bmatrix}}{\sigma_i - \sigma_k}}_{\text{positive eigenspaces}} + \underbrace{\sum_{i=1, i\neq k}^{r} \frac{1}{2} \frac{\begin{bmatrix} u_i \\ -v_i \end{bmatrix} \begin{bmatrix} u_i^\top & -v_i^\top \end{bmatrix}}{-\sigma_i - \sigma_k}}_{\text{negative eigenspaces}} \\ & - \underbrace{\sum_{j=r+1}^{m} \frac{\begin{bmatrix} u_j \\ 0 \end{bmatrix} \begin{bmatrix} u_j^\top & 0 \end{bmatrix}}{\sigma_k^{(0)}}}_{\text{left-null eigenspaces}} - \underbrace{\sum_{j=r+1}^{n} \frac{\begin{bmatrix} 0 \\ v_j \end{bmatrix} \begin{bmatrix} 0 & v_j^\top \end{bmatrix}}{\sigma_k^{(0)}}}_{\text{right-null eigenspaces}}, \end{aligned} \tag{193}$$

where:

1. **contributions in positive eigenspaces** ($S_k^{(+0)}$) represents the contribution in the subspaces associated with $w_i^{(+0)}$;
2. **contributions in negative eigenspaces** ($S_k^{(-0)}$) represents the contribution in the subspaces associated with $w_i^{(-0)}$;
3. **contributions in left-null eigenspaces** ($S_k^{(0,a)}$) represents the contribution in the subspaces associated with $a_j^{(0)}$;
4. **contributions in right-null eigenspaces** ($S_k^{(0,b)}$) represents the contribution in the subspaces associated with $b_j^{(0)}$.

Substituting $w_k^{(+0)}$ into:

$$S_k^{(0)} \mathcal{T}^{(1)} w_k^{(+0)} \tag{194}$$

yields:

$$S_k^{(0)} \mathcal{T}^{(1)} w_k^{(+0)} = S_k^{(0)} \mathcal{T}^{(1)} \frac{1}{\sqrt{2}} \begin{bmatrix} u_i \\ v_i \end{bmatrix} = S_k^{(0)} \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix}, \tag{195}$$

then apply the explicit $S_k^{(0)}$ on this result:

$$S_k^{(0)} \mathcal{T}^{(1)} w_k^{(+0)} = S_k^{(0)} \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix} = (S_k^{(+0)} + S_k^{(-0)} + S_k^{(0,a)} + S_k^{(0,b)}) \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix}, \tag{196}$$

and discuss the contributions in terms of subspaces:

1. **Contributions in Positive Eigenspaces**. To compute

$$S_k^{(+0)} \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix} = \sum_{i=1, i\neq k}^{r} \frac{w_i^{(+0)} (w_i^{(+0)})^\top}{\sigma_i - \sigma_k} \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix}, \tag{197}$$

consider:

$$\begin{aligned} (w_i^{(+0)})^\top \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix} &= \frac{1}{\sqrt{2}} \begin{bmatrix} u_i^\top & v_i^\top \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix} \\ &= \frac{1}{2} \left[u_i^\top A^{(1)} v_k + v_i^\top (A^{(1)})^\top u_k \right]. \end{aligned} \tag{198}$$

Note the identity:

$$v_i^\top (A^{(1)})^\top u_k = u_k^\top A^{(1)} v_i \in \mathbb{R}, \tag{199}$$

so that:

$$(w_i^{(+)})^\top \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix} = \frac{1}{2} \left[u_i^\top A^{(1)} v_k + u_k^\top A^{(1)} v_i \right]. \tag{200}$$

Therefore the contributions in positive eigenspaces are given as:

$$S_k^{(+0)} \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix} = \sum_{i=1, i \neq k}^{r} \frac{w_i^{(+0)} (w_i^{(+0)})^\top}{\sigma_i - \sigma_k} \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix} \tag{201}$$

$$= \sum_{i=1, i \neq k}^{r} \frac{1}{\sqrt{2}} \cdot \frac{u_i^\top A^{(1)} v_k + u_k^\top A^{(1)} v_i}{2(\sigma_i - \sigma_k)} \begin{bmatrix} u_i \\ v_i \end{bmatrix}. \tag{202}$$

2. **Contributions in Negative Eigenspaces**.
To compute:

$$S_k^{(-0)} \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix} = \sum_{i=1, i \neq k}^{r} \frac{w_i^{(-0)} (w_i^{(-0)})^\top}{-\sigma_i - \sigma_k} \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix}, \tag{203}$$

consider:

$$(w_i^{(-0)})^\top \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} u_i^\top & -v_i^\top \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix}$$

$$= \frac{1}{2} \left[u_i^\top A^{(1)} v_k - v_i^\top (A^{(1)})^\top u_k \right]. \tag{204}$$

Note the identity:

$$v_i^\top (A^{(1)})^\top u_k = u_k^\top A^{(1)} v_i \in \mathbb{R} \tag{205}$$

so that:

$$(w_i^{(+0)})^\top \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix} = \frac{1}{2} \left[u_i^\top A^{(1)} v_k - u_k^\top A^{(1)} v_i \right] \tag{206}$$

$$= \frac{1}{\sqrt{2}} \cdot \frac{u_i^\top A^{(1)} v_k - u_k^\top A^{(1)} v_i}{2(-\sigma_i - \sigma_k)} \begin{bmatrix} u_i \\ -v_i \end{bmatrix}. \tag{207}$$

Therefore the contributions in negative eigenspaces are given as:

$$S_k^{(-0)} \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix} = \sum_{i=1, i \neq k}^{r} \frac{w_i^{(-0)} (w_i^{(-0)})^\top}{-\sigma_i - \sigma_k} \frac{1}{\sqrt{2}} \begin{bmatrix} A^{(1)} v_k \\ (A^{(1)})^\top u_k \end{bmatrix} \tag{208}$$

$$= \sum_{i=1, i \neq k}^{r} \frac{1}{\sqrt{2}} \cdot \frac{u_i^\top A^{(1)} v_k - u_k^\top A^{(1)} v_i}{2(-\sigma_i - \sigma_k)} \begin{bmatrix} u_i \\ -v_i \end{bmatrix}. \tag{209}$$

3. **Contributions in Left-Null Eigenspaces**.

To compute:

$$S_k^{(0,a)}\frac{1}{\sqrt{2}}\begin{bmatrix}A^{(1)}v_k\\(A^{(1)})^\top u_k\end{bmatrix}=-\sum_{j=r+1}^{m}\frac{a_j^{(0)}(a_j^{(0)})^\top}{\sigma_k}\frac{1}{\sqrt{2}}\begin{bmatrix}A^{(1)}v_k\\(A^{(1)})^\top u_k\end{bmatrix},\tag{210}$$

consider:

$$\begin{aligned}(a_j^{(0)})^\top\cdot\frac{1}{\sqrt{2}}\begin{bmatrix}A^{(1)}v_k\\(A^{(1)})^\top u_k\end{bmatrix}&=\frac{1}{\sqrt{2}}\begin{bmatrix}u_j^\top & 0\end{bmatrix}\cdot\frac{1}{\sqrt{2}}\begin{bmatrix}A^{(1)}v_k\\(A^{(1)})^\top u_k\end{bmatrix}\\&=\frac{1}{2}u_j^\top A^{(1)}v_k.\end{aligned}\tag{211}$$

Therefore the contributions in left-null eigenspaces are given as:

$$S_k^{(0,a)}\frac{1}{\sqrt{2}}\begin{bmatrix}A^{(1)}v_k\\(A^{(1)})^\top u_k\end{bmatrix}=-\sum_{j=r+1}^{m}\frac{a_j^{(0)}(a_j^{(0)})^\top}{\sigma_k}\frac{1}{\sqrt{2}}\begin{bmatrix}A^{(1)}v_k\\(A^{(1)})^\top u_k\end{bmatrix}\tag{212}$$

$$=-\sum_{j=r+1}^{m}\frac{1}{\sqrt{2}}\cdot\frac{u_j^\top A^{(1)}v_k}{\sigma_k}\begin{bmatrix}u_j\\0\end{bmatrix}.\tag{213}$$

4. **Contributions in Right-Null Eigenspaces**.

To compute:

$$S_k^{(0,b)}\frac{1}{\sqrt{2}}\begin{bmatrix}A^{(1)}v_k\\(A^{(1)})^\top u_k\end{bmatrix}=-\sum_{j=r+1}^{n}\frac{b_j^{(0)}(b_j^{(0)})^\top}{\sigma_k}\frac{1}{\sqrt{2}}\begin{bmatrix}A^{(1)}v_k\\(A^{(1)})^\top u_k\end{bmatrix},\tag{214}$$

consider:

$$\begin{aligned}(b_j^{(0)})^\top\cdot\frac{1}{\sqrt{2}}\begin{bmatrix}A^{(1)}v_k\\(A^{(1)})^\top u_k\end{bmatrix}&=\frac{1}{\sqrt{2}}\begin{bmatrix}0 & v_j^\top\end{bmatrix}\cdot\frac{1}{\sqrt{2}}\begin{bmatrix}A^{(1)}v_k\\(A^{(1)})^\top u_k\end{bmatrix}\\&=\frac{1}{2}v_j^\top(A^{(1)})^\top u_k\end{aligned}\tag{215}$$

$$=\frac{1}{2}u_k^\top A^{(1)}v_j.\tag{216}$$

Therefore the contributions in right-null eigenspaces are given as:

$$S_k^{(0,b)}\frac{1}{\sqrt{2}}\begin{bmatrix}A^{(1)}v_k\\(A^{(1)})^\top u_k\end{bmatrix}=-\sum_{j=r+1}^{n}\frac{b_j^{(0)}(b_j^{(0)})^\top}{\sigma_k}\frac{1}{\sqrt{2}}\begin{bmatrix}A^{(1)}v_k\\(A^{(1)})^\top u_k\end{bmatrix}\tag{217}$$

$$=-\sum_{j=r+1}^{n}\frac{1}{\sqrt{2}}\cdot\frac{u_k^\top A^{(1)}v_j}{\sigma_k}\begin{bmatrix}0\\v_j\end{bmatrix}.\tag{218}$$

Computing Inner-Product Contributions in $\langle w_k,\mathcal{T}^{(1)}S_k\mathcal{T}^{(1)}w_k\rangle$. Since the term:

$$S_k^{(0)}\mathcal{T}^{(1)}w_k^{(+0)}\tag{219}$$

is computed above, to further derive:

$$\sigma_k^{(2,p=2)} = -\langle w_k^{(+0)}, \mathcal{T}^{(1)} S_k^{(0)} \mathcal{T}^{(1)} w_k^{(+0)} \rangle, \tag{220}$$

we compute the inner-product contributions in $\langle w_k, \mathcal{T}^{(1)} S_k \mathcal{T}^{(1)} w_k \rangle$ with respect to subspaces:

1. **Inner-Product Contributions in Positive Eigenspaces**. Consider:

$$Z^{(+0)} := \mathcal{T}^{(1)} S_k^{(+0)} \mathcal{T}^{(1)} w_k^{(+0)} \tag{221}$$

$$= \mathcal{T}^{(1)} \sum_{i=1, i\neq k}^{r} \frac{1}{\sqrt{2}} \cdot \frac{u_i^\top A^{(1)} v_k + u_k^\top A^{(1)} v_i}{2(\sigma_i - \sigma_k)} \begin{bmatrix} u_i \\ v_i \end{bmatrix} \tag{222}$$

$$= \sum_{i=1, i\neq k}^{r} \frac{1}{\sqrt{2}} \cdot \frac{u_i^\top A^{(1)} v_k + u_k^\top A^{(1)} v_i}{2(\sigma_i - \sigma_k)} \begin{bmatrix} A^{(1)} v_i \\ (A^{(1)})^\top u_i \end{bmatrix}, \tag{223}$$

so that:

$$\langle w_k, Z^{(+0)} \rangle = \sum_{i\neq k} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} u_k \\ v_k \end{bmatrix} \right)^\top \frac{1}{\sqrt{2}} \cdot \frac{u_i^\top A^{(1)} v_k + u_k^\top A^{(1)} v_i}{2(\sigma_i - \sigma_k)} \begin{bmatrix} A^{(1)} v_i \\ (A^{(1)})^\top u_i \end{bmatrix} \tag{224}$$

$$= \sum_{i=1, i\neq k}^{r} \left(\frac{1}{\sqrt{2}} \right)^2 \cdot \frac{u_i^\top A^{(1)} v_k + u_k^\top A^{(1)} v_i}{2(\sigma_i - \sigma_k)} \cdot \left[u_k^\top A^{(1)} v_i + v_k^\top (A^{(1)})^\top u_i \right] \tag{225}$$

$$= \sum_{i=1, i\neq k}^{r} \frac{1}{4(\sigma_i - \sigma_k)} \left[u_i^\top A^{(1)} v_k + u_k^\top A^{(1)} v_i \right] \left[u_k^\top A^{(1)} v_i + u_i^\top A^{(1)} v_k \right] \tag{226}$$

$$= \sum_{i=1, i\neq k}^{r} \frac{1}{4(\sigma_i - \sigma_k)} \left[u_i^\top A^{(1)} v_k + u_k^\top A^{(1)} v_i \right]^2 \tag{227}$$

$$= \sum_{i=1, i\neq k}^{r} \frac{1}{4(\sigma_i - \sigma_k)} \left[\left[u_i^\top A^{(1)} v_k \right]^2 + 2 u_i^\top A^{(1)} v_k u_k^\top A^{(1)} v_i + \left[u_k^\top A^{(1)} v_i \right]^2 \right]. \tag{228}$$

2. **Inner-Product Contributions in Negative Eigenspaces**. Consider:

$$Z^{(-0)} := \mathcal{T}^{(1)} S_k^{(-0)} \mathcal{T}^{(1)} w_k^{(+0)} \tag{229}$$

$$= \mathcal{T}^{(1)} \sum_{i=1, i\neq k}^{r} \frac{1}{\sqrt{2}} \cdot \frac{u_i^\top A^{(1)} v_k - u_k^\top A^{(1)} v_i}{2(-\sigma_i - \sigma_k)} \begin{bmatrix} u_i \\ -v_i \end{bmatrix} \tag{230}$$

$$= \sum_{i=1, i\neq k}^{r} \frac{1}{\sqrt{2}} \cdot \frac{u_i^\top A^{(1)} v_k - u_k^\top A^{(1)} v_i}{2(-\sigma_i - \sigma_k)} \begin{bmatrix} A^{(1)} (-v_i) \\ (A^{(1)})^\top u_i \end{bmatrix}, \tag{231}$$

so that:

$$\langle w_k, Z^{(-0)} \rangle = \sum_{i=1, i\neq k}^{r} \frac{1}{2} \cdot \frac{u_i^\top A^{(1)} v_k - u_k^\top A^{(1)} v_i}{2(-\sigma_i - \sigma_k)} \left[u_k^\top A^{(1)} (-v_i) + v_k^\top (A^{(1)})^\top u_i \right] \tag{232}$$

$$= \sum_{i=1, i\neq k}^{r} \frac{1}{4(-\sigma_i - \sigma_k)} \left[u_i^\top A^{(1)} v_k - u_k^\top A^{(1)} v_i \right] \left[-(u_k)^\top A^{(1)} v_i + u_i^\top A^{(1)} v_k \right] \tag{233}$$

$$= \sum_{i=1, i\neq k}^{r} \frac{1}{4(-\sigma_i - \sigma_k)} \left[u_i^\top A^{(1)} v_k - u_k^\top A^{(1)} v_i \right]^2 \tag{234}$$

$$= \sum_{i=1, i\neq k}^{r} \frac{1}{4(-\sigma_i - \sigma_k)} \left[\left[u_i^\top A^{(1)} v_k \right]^2 - 2u_i^\top A^{(1)} v_k u_k^\top A^{(1)} v_i + \left[u_k^\top A^{(1)} v_i \right]^2 \right]. \tag{235}$$

3. **Inner-Product Contributions in Left-Null Eigenspaces**. Consider:

$$Z^{(0,a)} := \mathcal{T}^{(1)} S_k^{(0,a)} \mathcal{T}^{(1)} w_k^{(+0)} \tag{236}$$

$$= \mathcal{T}^{(1)} \sum_{j=r+1}^{m} \left(-\frac{1}{\sqrt{2}} \cdot \frac{u_j^\top A^{(1)} v_k}{\sigma_k} \begin{bmatrix} u_j \\ 0 \end{bmatrix} \right) \tag{237}$$

$$= \sum_{j=r+1}^{m} -\frac{1}{\sqrt{2}} \cdot \frac{u_j^\top A^{(1)} v_k}{\sigma_k} \begin{bmatrix} 0 \\ (A^{(1)})^\top u_j \end{bmatrix}, \tag{238}$$

so that:

$$\langle w_k, Z^{(0,a)} \rangle = \sum_{j=r+1}^{m} -\frac{1}{2} \cdot \frac{u_j^\top A^{(1)} v_k}{\sigma_k} \left[u_k^\top \cdot 0 + v_k^\top (A^{(1)})^\top u_j \right] \tag{239}$$

$$= \sum_{j=r+1}^{m} -\frac{1}{2} \cdot \frac{u_j^\top A^{(1)} v_k u_j^\top A^{(1)} v_k}{\sigma_k}. \tag{240}$$

4. **Inner-Product Contributions in Right-Null Eigenspaces**. Consider:

$$Z^{(0,b)} := \mathcal{T}^{(1)} S_k^{(0,b)} \mathcal{T}^{(1)} w_k^{(+0)} \tag{241}$$

$$= \mathcal{T}^{(1)} \sum_{j=r+1}^{n} \left(-\frac{1}{\sqrt{2}} \cdot \frac{u_k^\top A^{(1)} v_j}{\sigma_k} \begin{bmatrix} 0 \\ v_j \end{bmatrix} \right) \tag{242}$$

$$= \sum_{j=r+1}^{n} -\frac{1}{\sqrt{2}} \cdot \frac{u_k^\top A^{(1)} v_j}{\sigma_k} \begin{bmatrix} A^{(1)} v_j \\ 0 \end{bmatrix}, \tag{243}$$

so that:

$$\langle w_k, Z^{(0,b)} \rangle = \sum_{j=r+1}^{n} -\frac{1}{2} \cdot \frac{u_k^\top A^{(1)} v_j}{\sigma_k} \left[u_k^\top A^{(1)} v_j + v_k^\top \cdot 0 \right] \tag{244}$$

$$= \sum_{j=r+1}^{n} -\frac{1}{2} \cdot \frac{\left[u_k^\top A^{(1)} v_j \right]^2}{\sigma_k}. \tag{245}$$

Combining Terms. Since

$$\sigma_k^{(2,p=1)} = 0, \tag{246}$$

thus,

$$\sigma_k^{(2)} = \sigma_k^{(2,p=2)} \tag{247}$$

$$= -\left[\sum_{i=1, i\neq k}^{r} \frac{1}{4(\sigma_i - \sigma_k)} \left[\left[u_i^\top A^{(1)} v_k \right]^2 + 2u_i^\top A^{(1)} v_k u_k^\top A^{(1)} v_i + \left[u_k^\top A^{(1)} v_i \right]^2 \right] \right. \tag{248}$$

$$+ \sum_{i=1, i\neq k}^{r} \frac{1}{4(-\sigma_i - \sigma_k)} \left[\left[u_i^\top A^{(1)} v_k \right]^2 - 2u_i^\top A^{(1)} v_k u_k^\top A^{(1)} v_i + \left[u_k^\top A^{(1)} v_i \right]^2 \right] \tag{249}$$

$$- \sum_{j=r+1}^{m} \frac{1}{2\sigma_k} \left[u_j^\top A^{(1)} v_k\right]^2 - \sum_{j=r+1}^{n} \frac{1}{2\sigma_k} \left[u_k^\top A^{(1)} v_j\right]^2 \Bigg] \tag{250}$$

$$= - \left[\sum_{i=1, i\neq k}^{r} \frac{(\sigma_i + \sigma_k) - (\sigma_i - \sigma_k)}{4(\sigma_i^2 - \sigma_k^2)} \left[\left[u_i^\top A^{(1)} v_k\right]^2 + \left[u_k^\top A^{(1)} v_i\right]^2 \right] \right. \tag{251}$$

$$+ \frac{1}{2} \sum_{i=1, i\neq k}^{r} \frac{(\sigma_i + \sigma_k) + (\sigma_i - \sigma_k)}{\sigma_i^2 - \sigma_k^2} u_i^\top A^{(1)} v_k u_k^\top A^{(1)} v_i \tag{252}$$

$$\left. - \sum_{j=r+1}^{m} \frac{1}{2\sigma_k} \left[u_j^\top A^{(1)} v_k\right]^2 - \sum_{j=r+1}^{n} \frac{1}{2\sigma_k} \left[u_k^\top A^{(1)} v_j\right]^2 \right] \tag{253}$$

$$= \frac{1}{2} \sum_{i=1, i\neq k}^{r} \frac{\sigma_k}{\sigma_k^2 - \sigma_i^2} \left[u_i^\top A^{(1)} v_k\right]^2 + \frac{1}{2} \sum_{i\neq k}^{r} \frac{\sigma_k}{\sigma_k^2 - \sigma_i^2} \left[u_k^\top A^{(1)} v_i\right]^2 \tag{254}$$

$$+ \frac{1}{2} \sum_{i=1, i\neq k}^{r} \frac{\sigma_i}{\sigma_k^2 - \sigma_i^2} u_i^\top A^{(1)} v_k u_k^\top A^{(1)} v_i + \frac{1}{2} \sum_{i=1}^{r} \frac{\sigma_i}{\sigma_k^2 - \sigma_i^2} u_i^\top A^{(1)} v_k u_k^\top A^{(1)} v_i \tag{255}$$

$$+ \frac{1}{2} \sum_{j=r+1}^{m} \frac{\sigma_k}{\sigma_k^2 - \sigma_j^2} \left[u_j^\top A^{(1)} v_k\right]^2 + \frac{1}{2} \sum_{j=r+1}^{n} \frac{\sigma_k}{\sigma_k^2 - \sigma_j^2} \left[u_k^\top A^{(1)} v_j\right]^2 \tag{256}$$

$$\overset{\text{combine indices}}{=\!=\!=\!=} \frac{1}{2} \sum_{i\neq k} \frac{\sigma_k}{\sigma_k^2 - \sigma_i^2} \left[u_i^\top A^{(1)} v_k\right]^2 \tag{257}$$

$$+ \frac{1}{2} \sum_{i\neq k} \frac{\sigma_k}{\sigma_k^2 - \sigma_i^2} \left[u_k^\top A^{(1)} v_i\right]^2 \tag{258}$$

$$+ \frac{1}{2} \sum_{i=1}^{r} \frac{\sigma_i}{\sigma_i^2 - \sigma_k^2} u_i^\top A^{(1)} v_k u_k^\top A^{(1)} v_i \tag{259}$$

$$+ \frac{1}{2} \sum_{i=1}^{r} \frac{\sigma_i}{\sigma_i^2 - \sigma_k^2} u_k^\top A^{(1)} v_i u_i^\top A^{(1)} v_k. \tag{260}$$

Procedure III – Map Derivative Operator Layout. Use following identity from Lemma 2.1 (Essential Matrix Identities):

1. $\operatorname{vec}(BVA^\top) = (A \otimes B) \operatorname{vec}(V)$,

consider:

$$\lim_{x\to 0} x^2 \left[u_i^\top A^{(1)} v_k\right]^2 = \lim_{x\to 0} u_i^\top x A^{(1)} v_k u_i^\top x A^{(1)} v_k \tag{261}$$

$$= \left[u_i^\top \mathrm{d}A v_k\right] \left[u_i^\top \mathrm{d}A v_k\right] \tag{262}$$

$$= \left[v_k^\top \mathrm{d}A u_i\right] \left[u_i^\top \mathrm{d}A v_k\right] \tag{263}$$

$$= \operatorname{vec}\left[v_k^\top \mathrm{d}A u_i\right] \operatorname{vec}\left[u_i^\top \mathrm{d}A v_k\right] \tag{264}$$

$$= \left(u_i^\top \otimes v_k^\top\right) \operatorname{vec}(\mathrm{d}A) \left(v_k^\top \otimes u_i^\top\right) \operatorname{vec}(\mathrm{d}A) \tag{265}$$

$$= \left[\left(u_i^\top \otimes v_k^\top\right) \operatorname{vec}(\mathrm{d}A)\right]^\top \left[\left(v_k^\top \otimes u_i^\top\right) \operatorname{vec}(\mathrm{d}A)\right] \tag{266}$$

$$= \operatorname{vec}(\mathrm{d}A^\top) \left(v_k \otimes u_i\right) \left(v_k \otimes u_i\right)^\top \operatorname{vec}(\mathrm{d}A). \tag{267}$$

Similarly,

$$\lim_{x\to 0} x^2 \left[u_k^\top A^{(1)} v_i\right]^2 = \operatorname{vec}(\mathrm{d}A^\top)\,(v_i \otimes u_k)\,(v_i \otimes u_k)^\top \operatorname{vec}(\mathrm{d}A), \tag{268}$$

$$\lim_{x\to 0} x^2 u_i^\top A^{(1)} v_k u_k^\top A^{(1)} v_i = \lim_{x\to 0} u_i^\top x A^{(1)} v_k u_k^\top x A^{(1)} v_i \tag{269}$$

$$= \lim_{x\to 0} u_i^\top \mathrm{d}A v_k u_k^\top \mathrm{d}A v_i \tag{270}$$

$$= \operatorname{vec}(\mathrm{d}A^\top)\,(v_k \otimes u_i)\,(v_i \otimes u_k)^\top \operatorname{vec}(\mathrm{d}A), \tag{271}$$

and:

$$\lim_{x\to 0} x^2 u_k^\top A^{(1)} v_i u_i^\top A^{(1)} v_k = \lim_{x\to 0} u_k^\top x A^{(1)} v_i u_i^\top x A^{(1)} v_k \tag{272}$$

$$= \lim_{x\to 0} u_k^\top \mathrm{d}A v_i u_i^\top \mathrm{d}A v_k \tag{273}$$

$$= \operatorname{vec}(\mathrm{d}A^\top)\,(v_i \otimes u_k)\,(v_k \otimes u_i)^\top \operatorname{vec}(\mathrm{d}A). \tag{274}$$

Producing Lemma Claim. Hence,

$$D^2\sigma_k[\mathrm{d}A, \mathrm{d}A] = \operatorname{vec}(\mathrm{d}A)^\top \frac{\partial}{\partial \operatorname{vec}(A)} \operatorname{vec}\left(\frac{\partial \sigma_k}{\partial A}\right) \operatorname{vec}(\mathrm{d}A) \tag{275}$$

$$= 2 \lim_{x\to 0} x^2 \sigma_k^{(2)} \tag{276}$$

$$= \sum_{i\neq k} \frac{\sigma_k}{\sigma_k^2 - \sigma_i^2} \lim_{x\to 0} x^2 \left[u_i^\top A^{(1)} v_k\right]^2 \tag{277}$$

$$+ \sum_{i\neq k} \frac{\sigma_k}{\sigma_k^2 - \sigma_i^2} \lim_{x\to 0} x^2 \left[u_k^\top A^{(1)} v_i\right]^2 \tag{278}$$

$$+ \sum_{i=1}^{r} \frac{\sigma_i}{\sigma_i^2 - \sigma_k^2} \lim_{x\to 0} x^2 u_i^\top A^{(1)} v_k u_k^\top A^{(1)} v_i \tag{279}$$

$$+ \sum_{i=1}^{r} \frac{\sigma_i}{\sigma_i^2 - \sigma_k^2} \lim_{x\to 0} x^2 u_k^\top A^{(1)} v_i u_i^\top A^{(1)} v_k. \tag{280}$$

Re-labeling indices yields the claim:

$$D^2\sigma_k[\mathrm{d}A, \mathrm{d}A]$$

$$= \operatorname{vec}(\mathrm{d}A)^\top \Bigg[\underbrace{\sum_{i\neq k, i\leq r} \frac{\sigma_k}{\sigma_k^2 - \sigma_i^2} (v_k \otimes u_i)(v_k \otimes u_i)^\top}_{\text{left}} \tag{281}$$

$$+ \underbrace{\sum_{j\neq k, j\leq n} \frac{\sigma_k}{\sigma_k^2 - \sigma_j^2} (v_j \otimes u_k)(v_j \otimes u_k)^\top}_{\text{right}} \tag{282}$$

$$+ \underbrace{\sum_{l\neq k, l\leq r} \frac{\sigma_l}{\sigma_k^2 - \sigma_l^2} \left[(v_k \otimes u_l)(v_l \otimes u_k)^\top + (v_l \otimes u_k)(v_k \otimes u_l)^\top\right]}_{\text{left-right interaction}} \Bigg] \operatorname{vec}(\mathrm{d}A). \quad \square \tag{283}$$

7. Numerical experiments

We conduct numerical experiments to validate the correctness of the derived special cases $n = 1$ and $n = 2$. Matrix entries are sampled from $\mathcal{N}(0, 1)$ and $U(0, 1)$, respectively. Ground truth is obtained numerically via PyTorch's auto–differentiation framework [22]. The error ϵ is computed by the ℓ_2-norm

$$\epsilon = \|R_{\text{ours}} - R_{\text{gt}}\|_2, \tag{284}$$

where R_{ours} denotes the result from our theoretical computation and R_{gt} the ground truth from auto–differentiation. Singular values are indexed by $k = 1, 2, \ldots, r$ in the reported results. To facilitate the visualization and computation, we choose the dimensions 6×10 in all experiments.

Reproducibility. The random seed is fixed to 1 for reproducibility. All experimental code is available at https://github.com/roisincrtai/highorder_spectral_variation_analysis.

Results of Singular-Value Jacobian. Fig. 2 reports the results for the singular-value Jacobian. Matrix entries are sampled i.i.d. from $\mathcal{N}(0, 1)$ and $U(0, 1)$, respectively. The derivative matrices are visualized using the *viridis* color map. For each singular-value index k, results are shown in pairs: the left panel gives the theoretical computation from Lemma 5.1, while the right panel shows the numerical ground truth obtained from PyTorch's auto–differentiation framework. The reported errors are zero across all experiments.

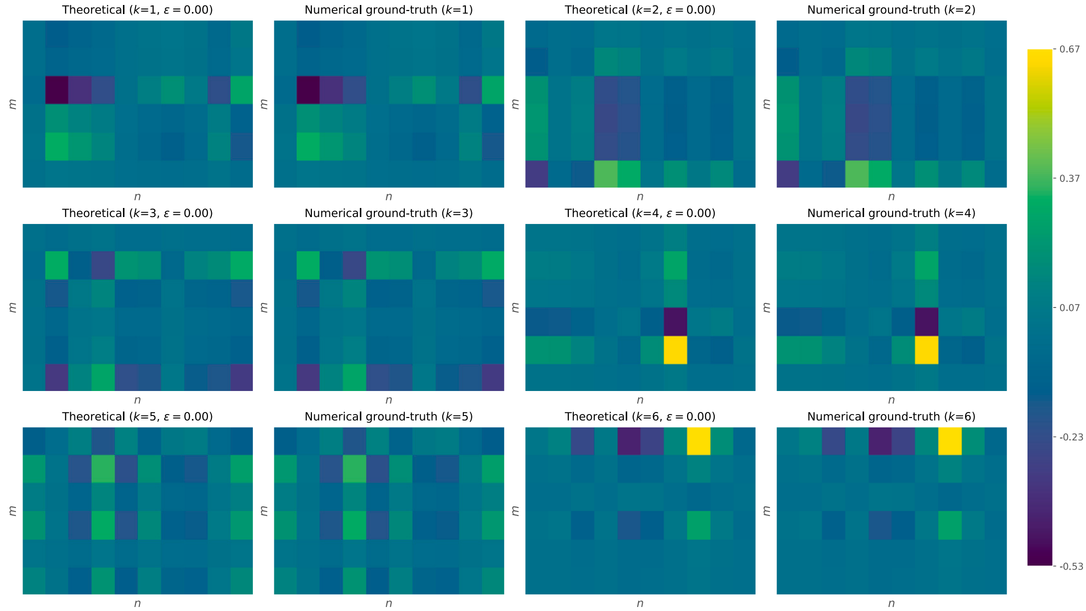
Results of Singular-Value Hessian. Fig. 3 reports the results for the singular-value Hessian. Matrix entries are sampled i.i.d. from $\mathcal{N}(0, 1)$ and $U(0, 1)$, respectively. The derivative matrices are visualized using the *viridis* color map. For each singular-value index k, results are shown in pairs: the left panel gives the theoretical computation from Lemma 6.1, while the right panel shows the numerical ground truth obtained from PyTorch's auto–differentiation framework. The observed errors between theoretical results and numerical ground-truth are on the order of 10^{-14}, confirming that they are numerically negligible (Fig. 4).

8. Conclusion

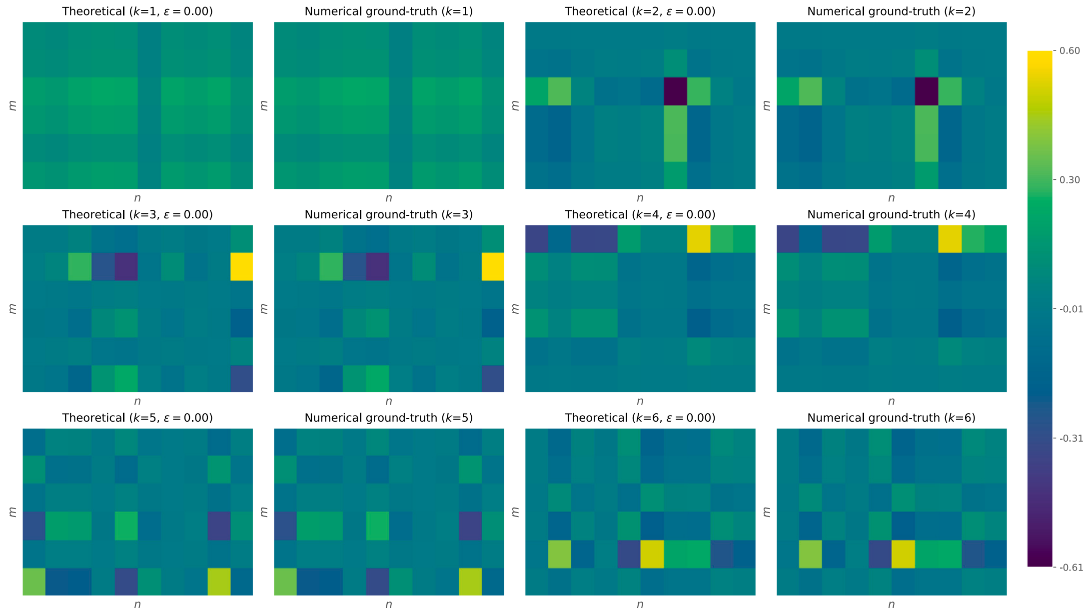
By viewing matrices as compact linear operators and extending Kato's perturbation theory for self-adjoint operators, we present a unified operator-theoretic framework for obtaining closed-form, arbitrary-order derivatives of singular values in real rectangular matrices. In contrast to the ad hoc methods of classical matrix analysis, our approach is systematic and procedural, allowing the derivation of singular-value derivatives of any order. The key step is the Jordan–Wielandt embedding, which maps a real rectangular matrix, usually non-self-adjoint, to a self-adjoint operator, thereby encapsulating its complete spectral information. Based on Kato's framework, we establish a general framework for deriving higher-order singular-value derivatives. Specializing to first order ($n = 1$) recovers the classical singular-value Jacobian, while specializing to second order ($n = 2$) yields a Kronecker-product representation of the singular-value Hessian that, to the best of our knowledge, has not previously appeared in the literature. Beyond these cases, the framework extends to arbitrary order. Higher-order singular-value derivatives are indispensable for analyzing induced spectral dynamics in statistical physics and deep learning.

Acknowledgments

This publication has emanated from research conducted with the financial support of Taighde Éireann - Research Ireland under Grant number 18/CRT/6223.

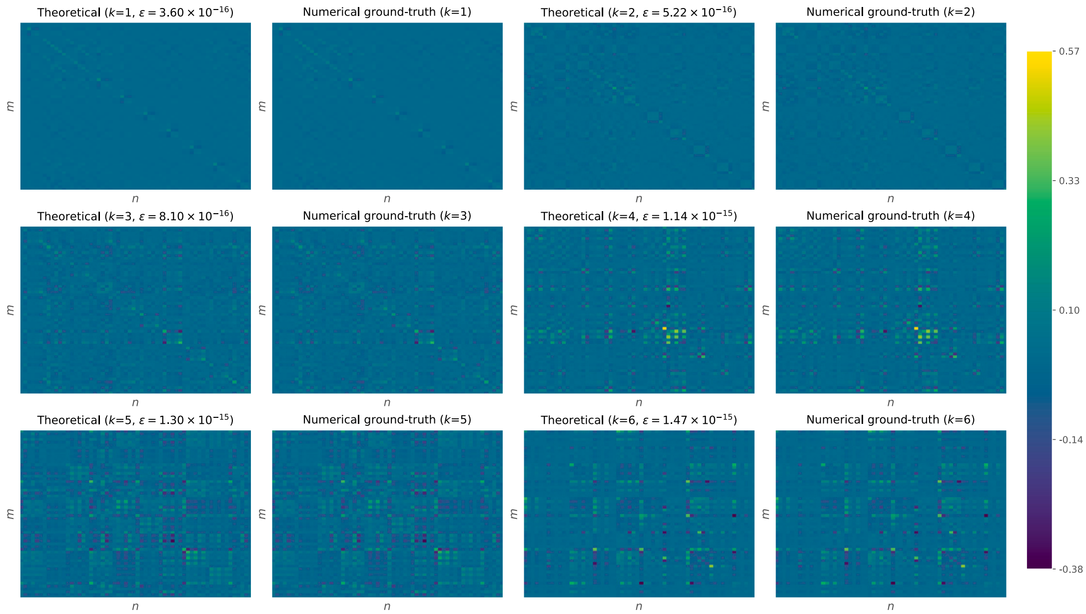


(a) Matrix entries i.i.d. $\sim \mathcal{N}(0, 1)$.

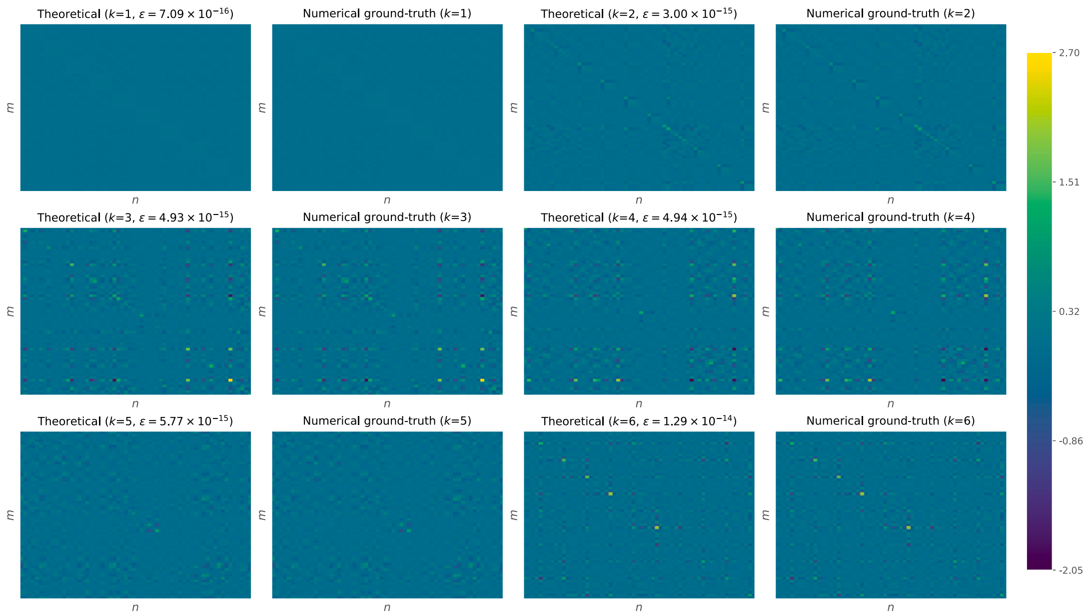


(b) Matrix entries i.i.d. $\sim U[0, 1]$.

Fig. 2. Numerical Experiments for Singular-Value Jacobian. This experiment compares the singular-value Jacobian derived from our framework with that obtained via PyTorch's auto–differentiation. The error ϵ is measured as the ℓ_2-norm between the theoretical and ground-truth results. The error is measured to be zero in these experiments, indicating no difference between the theoretical and ground-truth results. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

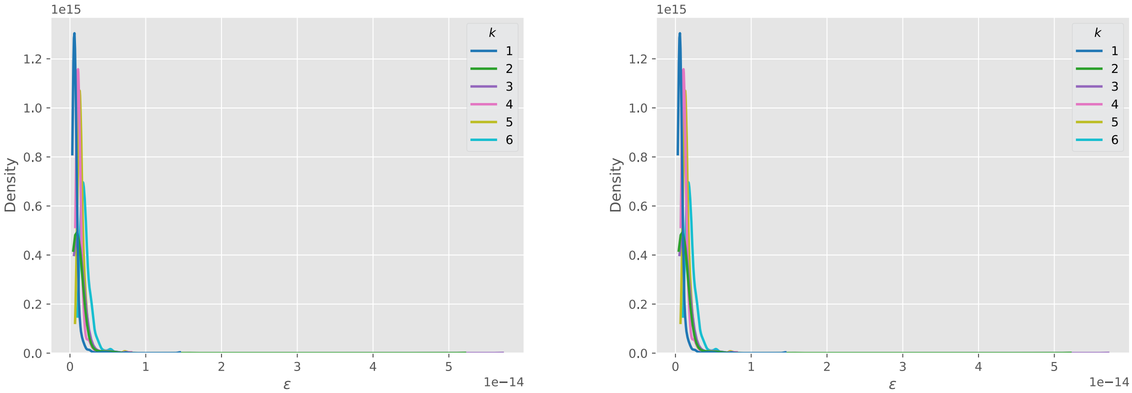


(a) Matrix entries i.i.d. $\sim \mathcal{N}(0, 1)$.



(b) Matrix entries i.i.d. $\sim U[0, 1]$.

Fig. 3. Numerical Experiments for Singular-Value Hessian. This experiment compares the singular-value Hessian derived from our framework with that obtained via PyTorch's auto–differentiation. The error ϵ is measured as the ℓ_2-norm between the theoretical and ground-truth results. The maximum error is measured to be less than 1.3×10^{-14} in these experiments, indicating the difference between the theoretical and ground-truth results is negligible.



(a) Matrix entries i.i.d. $\sim \mathcal{N}(0,1)$. (b) Matrix entries i.i.d. $\sim U[0,1]$.

Fig. 4. Errors for Singular-Value Hessian. Random matrix entries are sampled i.i.d. from $\mathcal{N}(0,1)$ and $U[0,1]$, respectively. For each singular-value index $k = 1, 2, \ldots, r$, the error ϵ is computed over 500 trials and visualized using an unnormalized histogram density. All reported errors are below 6×10^{-14} in these experiments.

References

[1] D. Applebaum, Lévy Processes and Stochastic Calculus, Cambridge University Press, 2009.
[2] Z. Bai, J.W. Silverstein, et al., Spectral Analysis of Large Dimensional Random Matrices, Springer, 2010.
[3] R. Bhatia, Matrix Analysis, vol. 169, Springer Science & Business Media, 2013.
[4] F.H. Clarke, Optimization and Nonsmooth Analysis, SIAM, 1990.
[5] N. Dunford, J.T. Schwartz, Linear Operators, Part 1: General Theory, John Wiley & Sons, 1988.
[6] A. Edelman, N.R. Rao, Random matrix theory, Acta Numer. 14 (2005) 233–297.
[7] U. Fano, Description of states in quantum mechanics by density matrix and operator techniques, Rev. Mod. Phys. 29 (1957) 74.
[8] J.N. Franklin, Matrix Theory, Courier Corporation, 2000.
[9] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, 2012.
[10] K. Itô, On stochastic differential equations, Mem. Am. Math. Soc. (1951) 1–51.
[11] T. Kato, Perturbation Theory for Linear Operators, Springer, 1995.
[12] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, SIAM Rev. 51 (2009) 455–500.
[13] A.S. Lewis, H.S. Sendov, Nonsmooth analysis of singular values. Part i: Theory, Set-Valued Anal. 13 (2005) 213–241.
[14] C.K. Li, R.C. Li, A note on eigenvalues of perturbed Hermitian matrices, Linear Algebra Appl. 395 (2005) 183–190.
[15] S. Liu, G. Trenkler, T. Kollo, D. von Rosen, O.M. Baksalary, Professor Heinz neudecker and matrix differential calculus, Stat. Pap. 65 (2024) 2605–2639.
[16] R. Luo, J. McDermott, C. Gagné, Q. Sun, C. O'Riordan, Optimization-induced dynamics of Lipschitz continuity in neural networks, arXiv preprint, arXiv:2506.18588, 2025.
[17] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, John Wiley & Sons, 2019.
[18] V.A. Marčenko, L.A. Pastur, Distribution of eigenvalues for some sets of random matrices, Math. USSR Sb. 1 (1967) 457.
[19] M.L. Mehta, Random Matrices, vol. 142, Elsevier, 2004.
[20] M.A. Nielsen, I.L. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, 2010.
[21] B. Oksendal, Stochastic Differential Equations: an Introduction with Applications, Springer Science & Business Media, 2013.
[22] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Pytorch: an imperative style, high-performance deep learning library, Adv. Neural Inf. Process. Syst. 32 (2019).
[23] J.W.S.B. Rayleigh, The Theory of Sound, vol. 2, Macmillan, 1896.
[24] F. Rellich, Perturbation Theory of Eigenvalue Problems, CRC Press, 1969.
[25] W. Rudin, Functional Analysis, 2nd ed., 1991.
[26] J.J. Sakurai, J. Napolitano, Modern Quantum Mechanics, Cambridge University Press, 2020.
[27] E. Schrödinger, Quantisierung als eigenwertproblem, Ann. Phys. 385 (1926) 437–490.
[28] O.M. Shalit, Dilation theory: a guided tour, in: Operator Theory, Functional Analysis and Applications, Springer, 2021, pp. 551–623.
[29] M. Spivak, Calculus on Manifolds: a Modern Approach to Classical Theorems of Advanced Calculus, CRC Press, 2018.
[30] G.W. Stewart, J.g. Sun, Matrix Perturbation Theory, Academic Press, 1990.
[31] G. Strang, Linear Algebra and Its Applications, 2012.
[32] T. Tao, Topics in Random Matrix Theory, vol. 132, American Mathematical Soc., 2012.
[33] C.A. Tracy, H. Widom, Level-spacing distributions and the Airy kernel, Commun. Math. Phys. 159 (1994) 151–174.
[34] P.Å. Wedin, Perturbation bounds in connection with singular value decomposition, BIT Numer. Math. 12 (1972) 99–111.

[35] H. Weyl, Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung), Math. Ann. 71 (1912) 441–479.
[36] H. Wielandt, et al., On eigenvalues of sums of normal matrices, Pac. J. Math. 5 (1955) 633–638.
[37] K. Yosida, Functional Analysis, vol. 123, Springer Science & Business Media, 2012.