Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Operator-based Analysis of Numerical Projection Perturbations in Grassmann Manifold Optimization

**Pascal KUBWIMANA**
University of Dodoma https://orcid.org/0009-0008-5610-6855
**B. Prabhakar Reddy**
prabhakar.bijjula@gmail.com
University of Dodoma https://orcid.org/0000-0002-1776-5064
**Jason M. Mkenyeleye**
University of Dodoma

**Research Article**

**Keywords:** Orthogonal projectors, Grassmann manifold, Matrix perturbation theory, Backward error analysis, Structured perturbations, Riemannian optimization

**Posted Date:** July 27th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10428281/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** The authors declare no competing interests.

# OPERATOR-BASED ANALYSIS OF NUMERICAL PROJECTION PERTURBATIONS IN GRASSMANN MANIFOLD OPTIMIZATION

**Pascal KUBWIMANA, B. Prabhakar Reddy, Jason M. Mkenyeleye**

*Department of Mathematics and Statistics, The University of Dodoma, P. O. Box 338, Dodoma, Tanzania*

**Abstract**

This paper develops a complete perturbation analysis for orthogonal projectors under tangent-structured perturbations arising in Grassmann manifold optimization. For a tangent perturbation $E$ of an exact rank-$p$ orthogonal projector $P$, the perturbed projector $\tilde{P} = P + E$ satisfies the exact identity $\tilde{P}^2 - \tilde{P} = E^2$, demonstrating that the idempotency defect is intrinsically quadratic. The sharp constant bound $\| \tilde{P}^2 - \tilde{P} \|_F \leq \epsilon^2/\sqrt{2}$ is proved, with equality if and only if the reduced perturbation has rank one. The closed-form spectrum of $\tilde{P}$ is derived: eigenvalues $\frac{1}{2}(1 \pm \sqrt{1 + 4\sigma_i^2})$ for each singular value $\sigma_i$ of the reduced perturbation, plus 0 and 1. The residual satisfies $\| \tilde{P}^2 - \tilde{P} \|_F = \sqrt{2\sum\sigma_i^4}$ with sharp two-sided bounds $\epsilon^2/\sqrt{2r} \leq \cdot \leq \epsilon^2/\sqrt{2}$, where $\epsilon = \| E \|_F$ and $r = \mathrm{rank}(E)/2$. The nearest exact orthogonal projector $P^*$ satisfies $\| P^* - \tilde{P} \|_F = \| \tilde{P}^2 - \tilde{P} \|_F + O(\epsilon^6)$. Non-tangent perturbations are shown to yield linear $O(\epsilon)$ defect, proving quadratic cancellation is unique to tangent structure. Propagation through retractions remains first-order. Rank-dependent ($\epsilon < \sqrt{u/p}$) and rank-independent ($\epsilon \lesssim 2^{1/4}\sqrt{u}$) stability criteria are provided. Numerical experiments on random projectors and on a real-data CIFAR-10 projector confirm all theoretical predictions to machine precision.

***Keywords:*** Orthogonal projectors; Grassmann manifold; Matrix perturbation theory; Backward error analysis; Structured perturbations; Riemannian optimization.

## 1. Introduction

The notion of optimizing on the Grassman manifold $\mathrm{Gr}(p, n)$ is one of the fundamental computational primitives used across modern data science, machine learning and signal processing. The challenge of optimizing over p-dimensional subspaces of $\mathbb{R}^n$.has direct implications for a plethora of applications ranging from low-rank matrix completion, subspace tracking and even deep learning models imposing orthogonality constraints or seeking to perform reduced-order modeling of dynamical systems (Mishra et al., 2014; Candès & Recht, 2009; Recht et al., 2010). Such algorithms exploit the compact structure of the Grassmann

manifold, leading to subspace-optimized techniques outperforming unconstrained methods (Dai & Yuan, 1999; Absil et al., 2008; Sato ,2022).

In numerical implementations of Grassmann optimization points on the manifold are almost universally represented by orthogonal projectors symmetric idempotent matrices $P$ satisfying $P^2 = P$, $P^\top = P$, and $\mathrm{rank}(P) = p$ (Stewart, 1973; Golub & Van Loan, 2013; Davis & Kahan, 1970). The representation is canonial and has better properties numerically if it's accurate, and it's easier to define vector transports and retractions (Absil & Malick, 2012; Zimmermann & Hüper, 2022; Bendokat et al., 2024). Classical perturbation theory for orthogonal projectors, dating back to the foundational work of Parlett (1998) and Wedin (1972) establishes that when an exact projector $P$ is perturbed to $\tilde{P} = P + E$ by an arbitrary symmetric error $E$, the resulting subspace error satisfies linear bounds of the form $\| \tilde{P} - P^* \|_F \leq C \| E \|_F$ for some constant $C$ that may depend on $n$, $p$, or the spectral gap (Stewart & Sun, 1990; Saad , 2011). More recent improvements to these constants by Chen et al. (2016), Vannieuwenhoven (2024), and Li et al. (2024) give tighter results but do not fundamentally alter the fact that perturbations grow linearly with backward error. O'Rourke et al. (2018, 2024) showed that random perturbations give optimal estimators for the classical bounds if one makes certain probabilistic assumptions, but the constant order remains linear in $\| E \|_F$. Finally, Damle and Sun (2020) produced uniform bounds for invariant subspace perturbations which hold over certain intervals of matrix parameters, again without using the tangent space. In contrast, this traditional linear sensitivity analysis assumes the perturbations are arbitrary symmetric matrices, thus, does not differentiate the directions that slide along the Grassmann manifold and those that throw the iterate out of it (Absil et al., 2007; Boumal, 2023). In fact, most of the perturbations in Riemannian optimization algorithms-coming from the gradient descent steps, retraction error or round-off errors-are highly tangent structured: they satisfy $E = P\Delta(I - P) + (I - P)\Delta^\top P$ for some $\Delta \in \mathbb{R}^{p\times(n-p)}$ (Edelman et al., 1998; Bendokat et al., 2024). These tangent perturbations preserve symmetry but, crucially, have a block structure that classical theory ignores (Stewart, 1973; Wedin, 1972). By treating all symmetric perturbations uniformly, the literature has missed a potential structural cancellation: the possibility that the idempotency defect $\| \tilde{P}^2 - \tilde{P} \|_F$ might grow not linearly but quadratically in $\| E \|_F$ when $E$ respects the tangent space. Moreover, existing bounds do not exploit the fact that for tangent perturbations, the first-order term in the expansion of $\tilde{P}^2 - \tilde{P}$ may vanish identically, leaving only second-order contributions (Stewart & Sun, 1990; Golub & Van Loan, 2013).

The core question is if by confining perturbations to the tangent space of the Grassmann manifold the sensitivity of the projector structure is essentially modified, and if that modification can be accounted for by a dimension-independent backward error bound (Stewart & Sun, 1990; Li et al., 2024). This absence of a tight theoretical bound has real consequences. For example, in floating point implementation of Grassmann optimization, the size of rounding errors in the tangent space will determine the frequency at which iterates have to be re-orthogonalised or projected back to the manifold (Nocedal et al., 2006; Boumal, 2023). A quadratic increase in structural faithfulness means that for small enough, the projector is kept exactly idempotent to machine precision with no projection step being explicitly necessary. This can be a major advantage in, e.g., subspace tracking and low-rank recovery on large scale systems (O'Rourke et al., 2018; Damle & Sun, 2020).

This study resolves the gap by developing an operator-based perturbation framework that exactly characterizes the effect of tangent-structured perturbations on orthogonal projectors (Edelman et al., 1998; Bendokat et al., 2024).We establish that for any tangent perturbation $E$, the perturbed projector $\tilde{P} = P + E$ satisfies the exact identity $\tilde{P}^2 - \tilde{P} = E^2$, demonstrating that the first-order term cancels identically and the idempotency defect is purely quadratic in $E$, with sharp constant bound $\| \tilde{P}^2 - \tilde{P} \|_F \leq \epsilon^2/\sqrt{2}$ attained if and only if the reduced perturbation has rank one. We derive the complete closed-form spectrum of $\tilde{P}$: eigenvalues $\frac{1}{2}(1 \pm \sqrt{1 + 4\sigma_i^2})$ for each singular value $\sigma_i$ of the reduced perturbation, together with unperturbed 0 and 1. The exact residual $\| \tilde{P}^2 - \tilde{P} \|_F = \sqrt{2 \sum_i \sigma_i^4}$ is computed and sharp two-sided bounds $\epsilon^2/\sqrt{2r} \leq \| \tilde{P}^2 - \tilde{P} \|_F \leq \epsilon^2/\sqrt{2}$ are proved, where $r = \text{rank}(E)/2$, which are dimension-independent and provide practical guidance for numerical stability. The nearest exact orthogonal projector $P^*$ is identified and it is shown that $\| P^* - \tilde{P} \|_F = \| \tilde{P}^2 - \tilde{P} \|_F + O(\epsilon^6)$, giving an optimal-constant backward error bound.

This study proves that for arbitrary symmetric perturbations, the idempotency defect is linear in $\epsilon$, establishing that quadratic cancellation is unique to tangent structure, while projector perturbations propagate through smooth retractions at linear order, clarifying the scope of the quadratic gain. Finally, we provide rank-dependent ($\epsilon < \sqrt{u/p}$) and rank-independent ($\epsilon \lesssim 2^{1/4}\sqrt{u}$) stability criteria for double-precision computation. All theoretical results are validated on random orthogonal projectors and on a real-data projector derived from the CIFAR-10 image dataset, confirming all theoretical predictions to machine precision.

## 2. Mathematical Framework

### 2.1 Notation and Preliminaries

Throughout this paper, $\|\cdot\|_F$ denotes the Frobenius norm. For a matrix $X$, $\mathrm{rank}(X)$ is its rank. The symbol $I_k$ denotes the $k \times k$ identity matrix. The unit roundoff in double precision is denoted $u \approx 2.22 \times 10^{-16}$(Higham, 2002; Wilkinson, 1965).

The analysis draws upon the spectral theorem for symmetric matrices (Horn & Johnson, 2013; Bhatia, 1997) and standard results from the theory of matrix perturbations (Stewart & Sun, 1990; Demmel, 1997).We also rely on the classical Wedin (1972) and Davis–Kahan (1970) perturbation bounds for subspaces, which provide the baseline against which our tangent-structured improvements are compared.

### 2.2 Grassmann Manifold Representation

Let $\mathrm{Gr}(p, n)$ denote the Grassmann manifold of $p$-dimensional subspaces of $\mathbb{R}^n$. An orthogonal projector $P \in \mathbb{R}^{n\times n}$ satisfies

$$P^2 = P, \quad P^\top = P, \quad \mathrm{rank}(P) = p. \tag{1}$$

By the spectral theorem, there exists an orthogonal $U \in \mathbb{R}^{n\times n}$ such that

$$P = U \begin{bmatrix} I_p & 0 \\ 0 & 0_{n-p} \end{bmatrix} U^\top. \tag{2}$$

Every element of Gr(p,n) is in fact uniquely identified to such a projector of rank p (Edelman et al., 1998). Such a view can be used to represent the manifold as a sub-manifold of the space of symmetric matrices, which allows for matrix-analytic study of geometric intrinsic quantities (Absil et al., 2004; Bendokat et al., 2024). It is actually a quite standard view point used for example for numerical linear algebra and optimization on manifolds (Golub & Van Loan, 2013; Absil et al., 2008).

The tangent space at $P$ admits the structured characterization

$$T_P\mathrm{Gr}(p,n) = \left\{\xi \in \mathbb{R}^{n\times n} : \xi = P\Delta(I-P) + (I-P)\Delta^\top P,\ \Delta \in \mathbb{R}^{p\times(n-p)}\right\}.$$
(3)

This characterization follows from the standard derivation of the Grassmann tangent space (Edelman et al., 1998; Absil et al., 2008; Bendokat et al., 2024). The tangent space dimension is $p(n-p)$, matching the manifold dimension.

### 2.3 Operator-Based Perturbation Model

For a perturbed projector $\tilde{P} = P + E$ with $\| E \|_F = \epsilon$, define the operator residual

$$R(P, \tilde{P}) = \| \tilde{P}^2 - \tilde{P} \|_F. \tag{4}$$

The analysis restricts to tangent perturbations of the form

$$E = P\Delta(I - P) + (I - P)\Delta^\top P, \quad \Delta \in \mathbb{R}^{p \times (n-p)}. \tag{5}$$

This satisfies perturbations staying aligned to the geometry of the Grassmann manifold and matches with the notion of admissible directions of the projection-based retractions (Absil & Malick, 2012; Zimmermann & Hüper, 2022). The tangent perturbation E is by construction symmetric, and, in the spectral basis of P, its blocks have a zero diagonal which we will find very convenient for quadratic cancellations (Bendokat et al., 2024).

### 2.4 Quadratic Residual Identity

The fundamental identity that drives all subsequent results is now established.

**Lemma 1** (Quadratic Residual Structure). Let P be an exact orthogonal projector and E a tangent perturbation with $\| E \|_F = \epsilon$. Define $\tilde{P} = P + E$. Then

1. $\tilde{P}^\top = \tilde{P}$ *(symmetry),*
2. $\tilde{P}^2 - \tilde{P} = E^2$,
3. $\| \tilde{P}^2 - \tilde{P} \|_F = \| E^2 \|_F \leq \epsilon^2$.

*Proof.* (i) Symmetry follows from the symmetry of $P$ and $E$. (ii) Expanding,

$$\tilde{P}^2 = (P + E)^2 = P^2 + PE + EP + E^2 = P + PE + EP + E^2, \tag{6}$$

so

$$\tilde{P}^2 - \tilde{P} = (P + PE + EP + E^2) - (P + E) = PE + EP + E^2 - E. \tag{7}$$

Using the tangent form (5), one computes $PE = P\Delta(I - P)$ and $EP = (I - P)\Delta^\top P$, so $PE + EP = E$; hence $\tilde{P}^2 - \tilde{P} = E^2$. (iii) By submultiplicativity of the Frobenius norm, $\| E^2 \|_F \leq \| E \|_F^2 = \epsilon^2$. □

This identity reveals that the first-order term cancels exactly, leaving only the quadratic contribution $E^2$. This cancellation is unique to tangent-structured perturbations and does not

occur for general symmetric perturbations, as shown in Theorem 3.7. The result generalizes classical perturbation bounds (Ring & Wirth, 2012; Wedin, 1972) by exploiting the special structure of tangent directions (Edelman et al., 1998). The next lemma sharpens the bound using the singular value structure of the tangent perturbation.

**Lemma 2** (Sharp Residual Constant). Under the hypotheses of Lemma 1, the residual satisfies the sharper bound

$$\| \widetilde{\mathrm{P}}^2 - \widetilde{\mathrm{P}} \|_{\mathrm{F}} \leq \frac{\epsilon^2}{\sqrt{2}}, \tag{8}$$

with equality if and only if the tangent block $\Delta$ (in the spectral basis of P) has rank one.

Proof. In the spectral basis where $\mathrm{P} = \mathrm{U}\mathrm{diag}(\mathrm{I_p}, 0)\mathrm{U}^{\top}$, the tangent perturbation has block form

$$U^{\top}EU = \begin{bmatrix} 0 & \Delta \\ \Delta^{\top} & 0 \end{bmatrix}, \tag{9}$$

so $\| E \|_F^2 = 2 \| \Delta \|_F^2$ and hence $\| \Delta \|_F^2 = \epsilon^2/2$. Squaring,

$$U^{\top}(\tilde{P}^2 - \tilde{P})U = U^{\top}E^2U = \begin{bmatrix} \Delta\Delta^{\top} & 0 \\ 0 & \Delta^{\top}\Delta \end{bmatrix}, \tag{10}$$

hence

$$\| \tilde{P}^2 - \tilde{P} \|_F^2 = \| \Delta\Delta^{\top} \|_F^2 + \| \Delta^{\top}\Delta \|_F^2. \tag{11}$$

Let $\sigma_1 \geq \cdots \geq \sigma_r > 0$ be the singular values of $\Delta$. Then

$$\| \Delta\Delta^{\top} \|_F^2 = \| \Delta^{\top}\Delta \|_F^2 = \sum_{i=1}^{r} \sigma_i^4, \quad \| \Delta \|_F^2 = \sum_i \sigma_i^2. \tag{12}$$

By the power-mean (Cauchy-Schwarz) inequality, $\sum_i \sigma_i^4 \leq (\sum_i \sigma_i^2)^2$, with equality iff at most one $\sigma_i$ is nonzero, i.e., $\Delta$ has rank one. Therefore

$$\| \tilde{P}^2 - \tilde{P} \|_F^2 = 2\sum_i \sigma_i^4 \leq 2(\sum_i \sigma_i^2)^2 = 2(\frac{\epsilon^2}{2})^2 = \frac{\epsilon^4}{2}, \tag{13}$$

so $\| \tilde{P}^2 - \tilde{P} \|_F \leq \epsilon^2/\sqrt{2}$.

This sharpens the sub multiplicativity bound $\epsilon^2$(Lemma 1) by a factor of $1/\sqrt{2}$. , and the equality condition identifies rank-one perturbations as the worst case. This result is closely related to classical perturbation bounds (Stewart & Sun, 1990; Davis & Kahan, 1970) but improves them by exploiting the tangent structure (Edelman et al., 1998; Bendokat et al., 2024).

### 2.5 Quadratic Backward-Error Bound

From the residual identity, a dimension-independent backward-error bound is obtained.

**Theorem 3** (Quadratic Backward Error). Let P be an exact rank-p orthogonal projector and let E be a tangent perturbation with $\| \mathrm{E} \|_{\mathrm{F}} = \epsilon$ and $\epsilon < 1/2$. Let $\widetilde{\mathrm{P}} = \mathrm{P} + \mathrm{E}$, and let $\mathrm{P}^*$ be the orthogonal projector onto the invariant subspace spanned by the p largest eigenvalues of the symmetric matrix $\widetilde{\mathrm{P}}$. Then $\mathrm{P}^*$ has rank p and

$$\| \tilde{P} - P^* \|_F \leq 2 \| \tilde{P}^2 - \tilde{P} \|_F \leq 2\epsilon^2. \tag{14}$$

*Proof.* $\tilde{P}$ is symmetric, with real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. Since $\| E \|_2 \leq \| E \|_F = \epsilon < 1/2$, Weyl's inequality places every eigenvalue of $\tilde{P}$ within $\epsilon$ of an eigenvalue of $P$, i.e., within $\epsilon$ of $\{0,1\}$. Thus exactly $p$ eigenvalues lie in $(1/2, 1+\epsilon)$ and $n-p$ lie in $(-\epsilon, 1/2)$; $P^*$ (rounding these eigenvalues to 1 and 0 in the eigenbasis of $\tilde{P}$) is the nearest orthogonal projector in Frobenius norm and has rank $p$ (Golub & Van Loan, 2013; Stewart & Sun, 1990).

Writing the residual through the eigenvalues,

$$\| \tilde{P}^2 - \tilde{P} \|_F^2 = \sum_i \lambda_i^2 (\lambda_i - 1)^2, \tag{15}$$

while $\| \tilde{P} - P^* \|_F^2 = \sum_i d(\lambda_i)^2$ with $d(\lambda) = \min(|\lambda|, |\lambda - 1|)$. For $\lambda$ within $1/2$ of $\{0,1\}$, the larger of $|\lambda|, |\lambda - 1|$ is at least $1/2$, so

$$d(\lambda) \leq \frac{|\lambda(\lambda-1)|}{1/2} = 2|\lambda(\lambda - 1)|. \tag{16}$$

Squaring and summing gives $\| \tilde{P} - P^* \|_F \leq 2 \| \tilde{P}^2 - \tilde{P} \|_F$, and Lemma 1(iii) gives the bound $2\epsilon^2$. □

This theorem establishes that the backward error to the nearest exact projector is quadratic in the perturbation norm, with constant 2. Classical bounds (Stewart, 1973; Wedin, 1972) would give a linear bound with a constant depending on the spectral gap, which may be arbitrarily large. The quadratic bound is therefore significantly stronger for small perturbations and, crucially, is dimension independent (Bendokat et al., 2024; Damle & Sun, 2020).

### 2.6 Exact Spectral and Metric Structure

A complete characterization beyond bounds is now provided. Throughout this section, work in the spectral basis where $P = U\mathrm{diag}(I_p, 0)U^\top$ and the tangent perturbation takes the block form

$$U^\top E U = \begin{bmatrix} 0 & \Delta \\ \Delta^\top & 0 \end{bmatrix}, \quad \Delta \in \mathbb{R}^{p \times (n-p)}. \tag{17}$$

Let $\sigma_1 \geq \cdots \geq \sigma_r > 0$ be the singular values of $\Delta$, where $r = \mathrm{rank}(\Delta) \leq \min(p, n-p)$, and write $\epsilon = \| E \|_F$, so that $\sum_i \sigma_i^2 = \frac{1}{2}\epsilon^2$. (Equivalently, $2r = \mathrm{rank}(E)$.)

**Theorem 4** (Closed-Form Spectrum). *The eigenvalues of* $\tilde{P} = P + E$ *are*

$$\lambda_{\pm}(\sigma_i) = \frac{1}{2}\left(1 \pm \sqrt{1 + 4\sigma_i^2}\right), \quad i = 1, \ldots, r, \tag{18}$$

together with the eigenvalue 1 of multiplicity $\mathrm{p} - \mathrm{r}$ and the eigenvalue 0 of multiplicity $\mathrm{n} - \mathrm{p} - \mathrm{r}$.

*Proof.* For each singular triple $(\sigma_i, w_i, v_i)$ of $\Delta$, with $\Delta v_i = \sigma_i w_i$ and $\Delta^{\top} w_i = \sigma_i v_i$, the two-dimensional subspace spanned by $(w_i, 0)$ and $(0, v_i)$ is invariant under $U^{\top}\tilde{P}U$, on which it acts as the $2 \times 2$ matrix

$$\begin{bmatrix} 1 & \sigma_i \\ \sigma_i & 0 \end{bmatrix}. \tag{19}$$

Its eigenvalues satisfy $\lambda^2 - \lambda - \sigma_i^2 = 0$, giving $\frac{1}{2}(1 \pm \sqrt{1 + 4\sigma_i^2})$. The $p - r$ left-singular directions annihilated by $\Delta^{\top}$ give eigenvalue 1, and the $n - p - r$ right-singular directions annihilated by $\Delta$ give eigenvalue 0. The multiplicities sum to $n$ and the eigenvalues sum to $p = \mathrm{tr}(\tilde{P})$. □

This theorem shows that the spectrum departs from $\{0,1\}$ at exactly second order, as $\lambda_+(\sigma) = 1 + \sigma^2 + O(\sigma^4)$ and $\lambda_-(\sigma) = -\sigma^2 + O(\sigma^4)$. This closed-form characterization is more precise than general perturbation bounds (Stewart, 1973; Davis & Kahan, 1970) and provides exact structural information (Bendokat et al., 2024; Stewart & Sun, 1990).

**Theorem 5** (Exact Residual and Sharp Bounds). With the notation above,

$$\| \tilde{P}^2 - \tilde{P} \|_F = \sqrt{2\sum_{i=1}^{r} \sigma_i^4}. \tag{20}$$

Consequently,

$$\frac{\epsilon^2}{\sqrt{2r}} \leq \| \tilde{P}^2 - \tilde{P} \|_F \leq \frac{\epsilon^2}{\sqrt{2}}, \tag{21}$$

with the upper bound attained iff $\Delta$ has rank one and the lower bound attained iff all $\sigma_i$ are equal.

*Proof.* From the block form,

$$U^\top \tilde{P} U = \begin{bmatrix} I_p & \Delta \\ \Delta^\top & 0 \end{bmatrix}. \tag{22}$$

Squaring and subtracting gives

$$U^\top (\tilde{P}^2 - \tilde{P}) U = \begin{bmatrix} \Delta\Delta^\top & 0 \\ 0 & \Delta^\top\Delta \end{bmatrix}. \tag{23}$$

Since the Frobenius norm is invariant under orthogonal similarity,

$$\| \tilde{P}^2 - \tilde{P} \|_F^2 = \| \Delta\Delta^\top \|_F^2 + \| \Delta^\top\Delta \|_F^2 = 2\sum_{i=1}^{r} \sigma_i^4. \tag{24}$$

With the constraint $\sum_i \sigma_i^2 = \epsilon^2/2$, the inequalities

$$\sqrt{\sum_i \sigma_i^4} \in \left[ \frac{\sum_i \sigma_i^2}{\sqrt{r}},\ \sum_i \sigma_i^2 \right] \tag{25}$$

follow from the Cauchy-Schwarz inequality and the fact that for fixed $\ell^2$ norm, the $\ell^4$ norm is minimized when all components are equal and maximized when only one component is nonzero. Substituting $\sum_i \sigma_i^2 = \epsilon^2/2$ yields the stated bounds.

These bounds are sharp and dimension-independent, unlike classical bounds that depend on the spectral gap or the dimensions *n* and *p* (Stewart, 1973; Wedin, 1972; Chen et al., 2016). The lower bound $1/\sqrt{2}r$ shows that the residual can be significantly smaller when the perturbation is spread across many singular directions (Bendokat et al., 2024).

**Theorem 6** (Nearest Projector and Backward Error). Let $P^*$ be the orthogonal projector onto the eigenvectors of the p largest eigenvalues of $\widetilde{P}$ (the r values $\lambda_+(\sigma_i)$ together with the $p - r$ unit eigenvalues). Then $P^*$ is a nearest rank-p orthogonal projector to $\widetilde{P}$ in the Frobenius norm, and

$$\| P^* - \tilde{P} \|_F = \sqrt{2\sum_{i=1}^{r} \lambda_-(\sigma_i)^2}, \tag{26}$$

where $\lambda_-(\sigma) = \frac{1}{2}(1 - \sqrt{1 + 4\sigma^2})$. Moreover,

$$\| P^* - \tilde{P} \|_F = \| \tilde{P}^2 - \tilde{P} \|_F + O(\epsilon^6) \le \frac{\epsilon^2}{\sqrt{2}} + O(\epsilon^6). \tag{27}$$

*Proof.* $\tilde{P}$ is symmetric (Lemma 1); by the Eckart-Young-Fan characterization of the best low-rank approximation in the Frobenius norm (Eckart & Young, 1936; Golub & Van Loan, 2013), the nearest rank-$p$ orthogonal projector is obtained by capping the $p$ largest eigenvalues of $\tilde{P}$ at 1 and setting the remaining $n - p$ eigenvalues to 0, while keeping the eigenvectors

unchanged. By Theorem 4, the $p$ largest eigenvalues are the $\lambda_+(\sigma_i)$ (which are $> 1/2$) together with the $p - r$ eigenvalues equal to 1. The $n - p$ smallest eigenvalues are 0 (multiplicity $n - p - r$) and $\lambda_-(\sigma_i)$ (which are negative).

In the eigenbasis of $\tilde{P}$, the difference $P^* - \tilde{P}$ is diagonal with entries:

- $1 - \lambda_+(\sigma_i) = \lambda_-(\sigma_i)$ on the $r$ "+" modes,
- $0 - \lambda_-(\sigma_i) = -\lambda_-(\sigma_i)$ on the $r$ "−" modes,
- 0 on the remaining $p - r$ and $n - p - r$ modes.

Hence $\| P^* - \tilde{P} \|_F^2 = 2\sum_i \lambda_-(\sigma_i)^2$. The expansion $\lambda_-(\sigma) = -\sigma^2 + O(\sigma^4)$ follows from the Taylor series $\sqrt{1 + 4\sigma^2} = 1 + 2\sigma^2 - 2\sigma^4 + O(\sigma^6)$. Thus $\lambda_-(\sigma)^2 = \sigma^4 + O(\sigma^6)$, and summing over $i$ gives $\sum_i \lambda_-(\sigma_i)^2 = \sum_i \sigma_i^4 + O(\epsilon^6)$. Therefore

$$\| P^* - \tilde{P} \|_F = \sqrt{2\sum_i \sigma_i^4 + O(\epsilon^6)} = \| \tilde{P}^2 - \tilde{P} \|_F + O(\epsilon^6), \tag{28}$$

and the bound follows from Lemma 2.

This theorem shows that the distance to the nearest exact projector equals the idempotency residual to leading order, confirming that the backward error is quadratic with the same optimal constant $1/\sqrt{2}$. This result is consistent with the classical theory of best approximations (Golub & Van Loan, 2013; Stewart & Sun, 1990) but provides much sharper constants for tangent-structured perturbations.

**2.7 Non-Tangent Perturbations Restore Linear Growth**

To demonstrate that the quadratic cancellation is unique to tangent structure, general symmetric perturbations are analyzed.

**Theorem 7** (Linear Defect for Non-Tangent Perturbations). *Let $E$ be an arbitrary symmetric perturbation, decomposed in the spectral basis of $P$ as*

$$U^\top E U = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}, \tag{29}$$

*where $A \in \mathbb{R}^{p\times p}$, $C \in \mathbb{R}^{(n-p)\times(n-p)}$, and $B \in \mathbb{R}^{p\times(n-p)}$. Write $E_N$ for the normal (block-diagonal) part: $E_N = \begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}$. Then*

$$U^\top(\tilde{P}^2 - \tilde{P})U = \begin{bmatrix} A & 0 \\ 0 & -C \end{bmatrix} + U^\top E^2 U, \tag{30}$$

*and consequently*

$$\| \tilde{P}^2 - \tilde{P} \|_F \geq \| E_N \|_F - \| E \|_F^2. \tag{31}$$

*In particular, if* $\| E_N \|_F = c\epsilon$ *with* $c > 0$*, then* $\| \tilde{P}^2 - \tilde{P} \|_F \geq c\epsilon - \epsilon^2 = \Theta(\epsilon)$*: the idempotency defect is first order (linear) in* $\epsilon$*.*

*Proof.* Expanding $\tilde{P}^2 - \tilde{P} = PE + EP - E + E^2$. In the spectral basis where $P = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$, we have

$$PE = \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix}, \quad EP = \begin{bmatrix} A & 0 \\ B^\top & 0 \end{bmatrix}, \tag{32}$$

so

$$PE + EP - E = \begin{bmatrix} 2A & B \\ B^\top & 0 \end{bmatrix} - \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & -C \end{bmatrix}. \tag{33}$$

Adding $E^2$ gives the identity. The norm of the first term is

$$\| \begin{bmatrix} A & 0 \\ 0 & -C \end{bmatrix} \|_F = \sqrt{\| A \|_F^2 + \| C \|_F^2} = \| E_N \|_F. \tag{34}$$

By the reverse triangle inequality and $\| E^2 \|_F \leq \| E \|_F^2 = \epsilon^2$,

$$\| \tilde{P}^2 - \tilde{P} \|_F \geq \| E_N \|_F - \epsilon^2. \tag{35}$$

If $\| E_N \|_F = c\epsilon$, then the right-hand side is $c\epsilon - \epsilon^2$, which is $\Theta(\epsilon)$ for sufficiently small $\epsilon$.

This theorem shows that any component of the perturbation lying in the normal space (i.e., *A* or *C* nonzero) destroys the quadratic cancellation and reintroduces a linear defect. Classical perturbation bounds (Stewart, 1973; Wedin, 1972; Davis & Kahan, 1970) are recovered in this regime.

**Corollary 8**. A perturbation is tangent precisely when $\mathrm{A} = 0$ and $\mathrm{C} = 0$, i.e., $\mathrm{E_N} = 0$; then the linear term vanishes and Lemma 1 ($\widetilde{\mathrm{P}}^2 - \widetilde{\mathrm{P}} = \mathrm{E}^2$) is recovered. Thus, the quadratic cancellation of Lemma 1 is exactly the tangent case, and any normal component restores linear growth. This result explains the numerical observation that generic symmetric perturbations yield $\mathrm{O}(\epsilon)$ backward error while tangent perturbations yield $\mathrm{O}(\epsilon^2)$.

This corollary clarifies that the tangent structure is both necessary and sufficient for the quadratic improvement. The result generalizes the classical tangent space characterization (Edelman et al., 1998; Bendokat et al., 2024).

### 2.8 Propagation Through Retractions is First-Order

While the projector structure enjoys quadratic fidelity, the quadratic gain does not automatically carry to optimization iterates when a retraction is applied (Absil et al., 2008; Boumal, 2023).

**Theorem 9** (First-Order Retraction Propagation). Let R be a retraction on the Grassmann manifold, twice continuously differentiable on a neighborhood of the zero section, with $\mathrm{R_P}(0) = \mathrm{P}$ and $\mathrm{dR_P}(0) = \mathrm{id}$ (the defining first-order property of a retraction). Let $\xi$ be a tangent vector and $\eta$ a tangent perturbation with $\| \eta \|_{\mathrm{F}} = \epsilon$. Then there exists a local constant $\mathrm{L_R} = \| \mathrm{dR_P}(\xi) \|$ (a bound on the first derivative of R near $\xi$) such that

$$\| R_P(\xi + \eta) - R_P(\xi) \|_F \leq L_R \epsilon + O(\epsilon^2). \tag{36}$$

Hence the propagated error is first order in $\epsilon$.

*Proof.* $R_P$ is $C^2$ on a neighborhood of $\xi$, so by Taylor's theorem with the mean-value form of the remainder,

$$R_P(\xi + \eta) - R_P(\xi) = dR_P(\xi)[\eta] + \frac{1}{2} d^2R_P(\zeta)[\eta, \eta] \tag{37}$$

for some $\zeta$ on the segment $[\xi, \xi + \eta]$. Taking norms,

$$\| R_P(\xi + \eta) - R_P(\xi) \|_F \leq \| dR_P(\xi) \| \cdot \| \eta \|_F + \frac{1}{2} \sup \| d^2R_P \| \cdot \| \eta \|_F^2 = L_R \epsilon + O(\epsilon^2), \tag{38}$$

with $L_R = \| dR_P(\xi) \|$. □

**Consequence.** Because the leading term is $L_R \epsilon$ and not $O(\epsilon^2)$, the quadratic gain established for the idempotency defect (Theorem 3) is not, in general, inherited by the optimization iterate after a retraction step. A genuinely quadratic end-to-end guarantee would require a structure-preserving retraction (one for which the linear term in the relevant error also cancels); this is left as future work. This theorem demonstrates that although the projector representation of a point on the Grassmann manifold is quadratically stable under tangent perturbations, the propagation of such perturbations through a retraction, the standard tool for updating iterates in Riemannian optimization remains first-order (Absil et al., 2008; Wen & Yin , 2013; Absil &

Malick, 2012). This clarifies the scope of the quadratic gain and suggests that structure-preserving retractions (Zimmermann & Hüper, 2022; Bendokat et al., 2024) may be needed to fully exploit the quadratic fidelity.

### 2.9 Practical Stability Criteria

Two complementary stability criteria for practitioners are provided, based on the theoretical bounds derived above and the standard analysis of floating-point arithmetic (Higham, 2002; Wilkinson, 1965).

**Heuristic 10** (Rank-Dependent Stability Criterion). A practical guideline for retaining projector structure to working precision in double precision is

$$\epsilon < \sqrt{\frac{u}{p}}, \tag{39}$$

where $u \approx 2.22 \times 10^{-16}$ is the unit roundoff (Higham, 2002). For $p = 10$ this gives $\epsilon < \sqrt{2.22 \times 10^{-17}} \approx 4.71 \times 10^{-9}$. This criterion marks where the $O(\epsilon^2)$ structural defect sinks beneath the rounding floor of the computation (Wilkinson, 1965; Demmel, 1997). It is heuristic and not a convergence guarantee.

**Corollary 11** (Rank-Independent Stability Criterion). By Theorem 6, the backward error satisfies

$$\| P^* - \tilde{P} \|_F \leq \frac{\epsilon^2}{\sqrt{2}} + O(\epsilon^6). \tag{40}$$

Hence, to guarantee that the backward error stays below the unit roundoff u, it suffices that

$$\epsilon \lesssim (\sqrt{2} \cdot u)^{1/2} = 2^{1/4}\sqrt{u}. \tag{41}$$

Unlike a spectral-gap criterion, this threshold is independent of the projector rank p: the worst-case constant $1/\sqrt{2}$ in Theorem 5 does not grow with dimension. For IEEE double precision ($u \approx 2.22 \times 10^{-16}$) this gives $\epsilon \lesssim 1.3 \times 10^{-8}$; for half precision ($u \approx 4.9 \times 10^{-4}$) it gives $\epsilon \lesssim 2.6 \times 10^{-2}$, quantifying how much less perturbation the low-precision formats can absorb before projector structure is lost (Higham, 2002; Wilkinson, 1965).

These criteria are sufficient but not necessary (Higham, 2002). They provide practical guidance for controlling the accumulation of rounding errors in Grassmann optimization algorithms (Boumal, 2023; Absil et al., 2008).

## 3. Numerical Experiments

### 3.1 Experimental Setup

$P = QQ^{\top}$: random rank-p orthogonal projectors, with Q from a QR factorization of a random Gaussian matrix of rank p such that $Q \in \mathbb{R}^{n\times p}$ has orthonormal columns. Default parameters were n=100 and p=10. T: random tangent perturbation with distribution G and scaled s.t. $\| E \|_F = \epsilon$(Edelman et al., 1998; Bendokat et al., 2024). Average of each setup on 100 different runs: only mean and standard deviation are reported. Computations are carried out in IEEE double precision ($u \approx 2.22 \times 10^{-16}$) (Higham, 2002; Wilkinson, 1965). Setups follow typical practices in numerical linear algebra (Demmel, 1997; Trefethen & Bau, 1997).

### 3.2 Random Projector Results

For $\epsilon \gtrsim 10^{-7}$ the residual follows the quadratic law with a stable, dimension-independent constant $\approx 0.237$, and the backward error tracks it closely confirming Theorem 3 (the bound $2\epsilon^2$ is satisfied with room to spare). For $\epsilon \lesssim 10^{-7}$ the measured residual saturates at the rounding floor ($\approx 10^{-15}$); values below this cannot be observed in double precision (Higham, 2002; Wilkinson, 1965).

For fixed $\| E \|_F$, the backward error decreases as $p$ grows. This occurs because spreading the perturbation over more singular directions reduces $\| \Delta\Delta^{\top} \|_F$ relative to $\| \Delta \|_F^2$. This behavior is consistent with the dimension-independent bound of Theorem 3 and the sharp lower bound of Theorem 5, which scales as $1/\sqrt{r}$ (Bendokat et al., 2024; Stewart & Sun, 1990).

At $\epsilon = 10^{-6}$ the tangent perturbation produces a backward error roughly six orders of magnitude smaller than a generic symmetric perturbation of the same norm, directly exhibiting the quadratic-versus-linear distinction predicted by Theorem 7 (Stewart, 1973; Wedin, 1972; Davis & Kahan, 1970).

Table 1. Operator Residual and Backward Error vs $\epsilon$ for $n = 100$, $p = 10$, 100 Trials

| $\epsilon$ | $\| \tilde{P}^2 - \tilde{P} \|_F$ | ratio/$\epsilon^2$ | $\| \tilde{P} - P^* \|_F$ | ratio/$\epsilon^2$ |
|---|---|---|---|---|
| $10^{-2}$ | $2.36 \times 10^{-5}$ | 0.236 | $2.36 \times 10^{-5}$ | 0.236 |
| $10^{-4}$ | $2.37 \times 10^{-9}$ | 0.237 | $2.37 \times 10^{-9}$ | 0.237 |
| $10^{-6}$ | $2.37 \times 10^{-13}$ | 0.237 | $2.37 \times 10^{-13}$ | 0.237 |
| $10^{-8}$ | $1.3 \times 10^{-15}$ (floor) | - | $2.8 \times 10^{-15}$ (floor) | - |
| $10^{-10}$ | $1.3 \times 10^{-15}$ (floor) | - | $2.9 \times 10^{-15}$ (floor) | - |
| $10^{-12}$ | $1.3 \times 10^{-15}$ (floor) | - | $2.8 \times 10^{-15}$ (floor) | - |

Table 2. Backward Error vs Rank $p$ at $\epsilon = 10^{-6}$, $n = 100$

| $p$ | $\| \tilde{P} - P^* \|_F$ | ratio/$\epsilon^2$ |
|---|---|---|
| 1 | $7.1 \times 10^{-13}$ | 0.707 |
| 5 | $3.2 \times 10^{-13}$ | 0.325 |
| 10 | $2.4 \times 10^{-13}$ | 0.237 |
| 20 | $1.8 \times 10^{-13}$ | 0.178 |
| 40 | $1.5 \times 10^{-13}$ | 0.145 |

For $p = 1$, the perturbation matrix $\Delta$ is a row vector and must have rank one. The observed ratio $0.707$ exactly matches the theoretical upper bound $1/\sqrt{2}$ from Lemma 2 and Theorem 5, directly verifying the equality condition. For larger $p$, the ratio decreases as the perturbation rank $r$ grows, consistent with the lower bound scaling as $1/\sqrt{r}$ (Bendokat et al., 2024; Chen et al., 2016).

*Table 3. Tangent vs Non-Tangent Perturbation at $\epsilon = 10^{-6}$, $n = 100$, $p = 10$*

| **Configuration** | $\| \tilde{P} - P^* \|_F$ | **Order** |
|---|---|---|
| Tangent (this work) | $2.4 \times 10^{-13}$ | $O(\epsilon^2)$ |
| Non-tangent (block-diagonal) | $9.1 \times 10^{-7}$ | $O(\epsilon)$ |

The non-tangent perturbation in Table 3 was constructed by adding a random block-diagonal matrix (normal component) to the tangent perturbation, scaled to preserve $\| E \|_F = 10^{-6}$. This directly demonstrates the loss of quadratic behavior when the perturbation deviates from the tangent space. This directly demonstrates the loss of quadratic behavior when the perturbation is not tangent-structured (Stewart, 1977; Stewart & Sun, 1990). The backward error for the non-tangent perturbation is approximately $3.9 \times 10^6$ times larger than the tangent case, confirming the linear growth predicted by Theorem 7.

### 3.3 Validation on a Real-Data Projector (CIFAR-10)

To confirm that the analysis applies to projectors arising from real data, the experiment was repeated with $P$ taken as the orthogonal projector onto the top-$p$ principal subspace of the CIFAR-10 image set (Krizhevsky, 2009).A sample of 8,000 CIFAR-10 images was converted to grayscale $32 \times 32$ vectors ($n = 1024$), centered, and their covariance matrix eigendecomposed; the leading $p$ eigenvectors define $P$. The top ten principal components

capture 68.6% of the pixel variance. Tangent perturbations were applied to this data-derived projector exactly as in Section 3, averaged over 50 trials in IEEE double precision (Higham, 2002; Demmel, 1997). This validation approach follows standard practice in numerical experiments with real data (Candès & Recht, 2009; Recht et al., 2010; Vandereycken , 2013).

*CIFAR-10 PCA Projector ($n = 1024$, $p = 10$, 50 Trials): Residual and Backward Error vs $\epsilon$*

| $\epsilon$ | $\| \tilde{P}^2 - \tilde{P} \|_F$ | ratio/$\epsilon^2$ | $\| \tilde{P} - P^* \|_F$ | ratio/$\epsilon^2$ |
|---|---|---|---|---|
| $10^{-2}$ | $2.25 \times 10^{-5}$ | 0.225 | $2.25 \times 10^{-5}$ | 0.225 |
| $10^{-4}$ | $2.25 \times 10^{-9}$ | 0.225 | $2.25 \times 10^{-9}$ | 0.225 |
| $10^{-6}$ | $2.26 \times 10^{-13}$ | 0.226 | $2.26 \times 10^{-13}$ | 0.226 |
| $10^{-8}$ | $3.3 \times 10^{-15}$ (floor) | — | $4.4 \times 10^{-15}$ (floor) | — |
| $10^{-10}$ | $3.2 \times 10^{-15}$ (floor) | — | $4.2 \times 10^{-15}$ (floor) | — |

The quadratic law is reproduced almost exactly: the residual constant ($\approx 0.225$) matches the random-projector value ($\approx 0.237$) to within the variation expected from the differing subspace geometry, and the double-precision floor appears near $3 \times 10^{-15}$ for $n = 1024$. The rank dependence is again decreasing, and a generic non-tangent symmetric perturbation of the same norm yields a backward error of $9.9 \times 10^{-7}$about six orders of magnitude larger than the tangent value $2.3 \times 10^{-13}$ at $\epsilon = 10^{-6}$. Since Lemma 1 and Theorem 3 are algebraic identities valid for every orthogonal projector, this agreement is expected; the CIFAR-10 experiment confirms that the predicted behavior is not an artifact of randomly generated projectors but also holds for subspaces estimated from real image statistics (Krizhevsky, 2009; Bendokat et al., 2024).

The double-precision floor at approximately $10^{-15}$ arises from the absolute error in floating-point operations such as squaring and subtraction, where quantities of order $\epsilon^2$ fall below the machine epsilon when $\epsilon \lesssim 10^{-8}$ (Higham, 2002; Wilkinson, 1965; Demmel, 1997).

## 4. Discussion

The experiments confirm the exact formulas of Section 3 rather than merely a scaling law. The measured residual and backward error agree with $\sqrt{2\sum\sigma_i^4}$ and $\sqrt{2\sum\lambda_-(\sigma_i)^2}$ to machine precision, and the two coincide, as Theorem 6 predicts (Stewart & Sun, 1990; Golub & Van

Loan, 2013). The relative residual is constant in $\epsilon$ at $\approx 0.236$ and sits inside the sharp envelope $[1/\sqrt{2r}, 1/\sqrt{2}]$; as the perturbation rank $r$ grows it decreases toward the lower envelope, while the worst case $1/\sqrt{2}$ attained by a rank-one perturbation is independent of dimension (Bendokat et al., 2024; Chen et al., 2016). The ablation (Table 3) isolates tangent structure as the mechanism: a non-tangent perturbation of equal norm produces an $O(\epsilon)$ residual, about $3.9 \times 10^6$ times larger at $\epsilon = 10^{-6}$ (Stewart, 1977; Wedin, 1972).

The quadratic identity $\tilde{P}^2 - \tilde{P} = E^2$ reveals that tangent-structured perturbations enjoy a cancellation not available to generic symmetric perturbations. This explains why the numerical experiments show a six-order-of-magnitude improvement for tangent perturbations at $\epsilon = 10^{-6}$: the linear term vanishes exactly, leaving only the quadratic contribution (Edelman et al., 1998; Bendokat et al., 2024). Classical linear bounds (Stewart, 1973; Wedin, 1972; Davis & Kahan, 1970) cannot capture this effect because they are designed for worst-case symmetric perturbations.

The observed constant $\approx 0.237$ lies between the submultiplicativity bound $\epsilon^2$ (which would give ratio 1.0) and the tighter spectral bound $\epsilon^2/\sqrt{2} \approx 0.707\epsilon^2$ derived in Lemma 2. The actual value is smaller because random Gaussian $\Delta$ produces singular value distributions for which $\sum\sigma_i^4$ is significantly smaller than $(\sum\sigma_i^2)^2$ (Golub & Van Loan, 2013; Demmel, 1997). As $p$ increases, this effect becomes more pronounced (Table 2), explaining why the backward error decreases with rank despite the bound being dimension-independent (Bendokat et al., 2024; Stewart & Sun, 1990).

Compared with classical linear bounds for unstructured perturbations (Stewart, 1973; Wedin, 1972; Davis & Kahan, 1970) and with first-order projection-perturbation bounds (Stewart, 1977; Chen et al., 2016)), the tangent-structured analysis is quadratic with an optimal constant, which moves the usable stability threshold from the order of a spectral gap to a rank-independent $\epsilon \lesssim 2^{1/4}\sqrt{u}$ (Higham, 2002; Wilkinson, 1965). For double precision this admits tangent-structured perturbations up to $\epsilon \approx 10^{-8}$, far larger than a worst-case linear bound would permit (Boumal, 2023; Demmel, 1997).

## 5. Conclusion

This paper developed a complete, sharp perturbation analysis for orthogonal projectors under tangent-structured perturbations. The main results are summarized as follows.

- Lemma 1: $\tilde{P}^2 - \tilde{P} = E^2$ (exact quadratic residual identity).
- Lemma 2: $\| \tilde{P}^2 - \tilde{P} \|_F \leq \epsilon^2/\sqrt{2}$ (sharp constant, attained for rank-one $\Delta$).
- Theorem 4: Closed-form spectrum: $\frac{1}{2}(1 \pm \sqrt{1 + 4\sigma_i^2})$, 1, 0.
- Theorem 5: $\| \tilde{P}^2 - \tilde{P} \|_F = \sqrt{2\sum\sigma_i^4}$ with sharp bounds $\epsilon^2/\sqrt{2r} \leq \cdot \leq \epsilon^2/\sqrt{2}$.
- Theorem 6: $\| P^* - \tilde{P} \|_F = \| \tilde{P}^2 - \tilde{P} \|_F + O(\epsilon^6)$.
- Theorem 7: Non-tangent perturbations give linear $O(\epsilon)$ defect.
- Theorem 9: Retraction propagation remains first-order.
- Heuristic 10: Rank-dependent criterion $\epsilon < \sqrt{u/p}$.
- Corollary 11: Rank-independent criterion $\epsilon \lesssim 2^{1/4}\sqrt{u}$.

Numerical experiments with random projectors, and with a projector on real-data (CIFAR-10) reproduce the quadratic law above and show the double-precision floor below (Krizhevsky, 2009; Higham, 2002; Wilkinson, 1965). These results can be of interest to structured matrix perturbation theory for orthogonal projectors, and they can provide a basis for studying the effect of rounding errors in Riemannian optimization for Grassmann manifolds (Edelman et al., 1998; Absil et al., 2008; Boumal, 2023).

**Limitations of the study:**

Even though our perturbation analysis is exact and tight, there are a number of limitations to take into account.

- The analysis assumes exact orthogonal projector and tangent-structured perturbations. General symmetric perturbations (Theorem 7) or non-orthogonal projector are not addressed here.
- The quadratic gain is not preserved through retractions (Theorem 9), so our result may not directly be applicable for iterates unless using structure-preserving retractions are designed (Absil et al., 2008; Boumal, 2023; Zimmermann & Hüper, 2022).
- The numerical assessment employed Gaussian random perturbations; the behaviour of other type of perturbations (e.g., adversarial perturbations, structured rounding errors) could potentially differ (O'Rourke et al., 2018, 2024; Demmel, 1997).

- Our proposed conditions for stability (Heuristic 10, Corollary 11) are only sufficient and not necessary, and stronger bounds might be possible for certain types of perturbations (Higham, 2002; Stewart & Sun, 1990).
- Extensions of our results to the Stiefel manifold and simplistic Grassmann manifold are left as future research work (Absil et al., 2008; Bendokat et al., 2024; Edelman et al., 1998).

**Future research directions:**

- Extension to non-orthogonal projectors, which exhibit a different tangent structure.
- Development of structure-preserving retractions that would translate the quadratic gain to iterates (Zimmermann & Hüper, 2022; Absil & Malick, 2012).
- Study of tangent-structure error accumulation in an optimization sequence (Boumal, 2023; Bendokat et al., 2024).
- Application to other manifolds with comparable tangent-space structure, like the Stiefel manifold and the simplistic Grassmann manifold (Edelman et al., 1998; Absil et al., 2008).

**Acknowledgements**

The authors thank the University of Dodoma for providing the academic environment and resources that made this work possible.

**Declaration of competing interest.** The authors declare that have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Funding Statements.** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Data and code availability.** The experiments use synthetically generated random orthogonal projectors and the publicly available CIFAR-10 dataset (Krizhevsky, 2009). All code generated during this study is available from the corresponding author upon reasonable request.

## References

Absil, P.-A., Baker, C. G., & Gallivan, K. A. (2007). Trust-region methods on Riemannian manifolds. Foundations of Computational Mathematics,7(3), 303–330. https://doi.org/10.1007/s10208-005-0179-9.

Absil, P.-A., Mahony, R., & Sepulchre, R. (2004). Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. Acta Applicandae Mathematicae, 80(2), 199–220. https://doi.org/10.1023/B:ACAP.0000013855.14971.9a.

Absil, P.-A., Mahony, R., & Sepulchre, R. (2008). Optimization algorithms on matrix manifolds. Princeton University Press.

Absil, P.-A., & Malick, J. (2012). Projection-like retractions on matrix manifolds. SIAM Journal on Optimization, 22(1), 135–158. https://doi.org/10.1137/100802529.

Bendokat, T., Zimmermann, R., & Absil, P.-A. (2024). A Grassmann manifold handbook: Basic geometry and computational aspects. Advances in Computational Mathematics, 50(1), Article 6. https://doi.org/10.1007/s10444-023-10090-8.

Bhatia, R. (1997). Matrix analysis. Springer-Verlag. https://doi.org/10.1007/978-1-4612-0653-8.

Boumal, N. (2023). An introduction to optimization on smooth manifolds. Cambridge University Press.

Candès, E. J., & Recht, B. (2009). Exact matrix completion via convex optimization. Foundations of Computational Mathematics, 9(6), 717–772. https://doi.org/10.1007/s10208-009-9045-5.

Chen, Y.-M., Chen, X.-S., & Li, W. (2016). On perturbation bounds for orthogonal projections. Numerical Algorithms, 73(2), 433–444. https://doi.org/10.1007/s11075-016-0103-5.

Dai, Y.-H., & Yuan, Y.-X. (1999). A nonlinear conjugate gradient method with a strong global convergence property. SIAM Journal on Optimization, 10(1), 177–182. https://doi.org/10.1137/S1052623497318992.

Damle, A., & Sun, Y. (2020). Uniform bounds for invariant subspace perturbations. SIAM Journal on Matrix Analysis and Applications. Advance online publication. https://doi.org/10.1137/19M1281992.

Davis, C., & Kahan, W. M. (1970). The rotation of eigenvectors by a perturbation. III. SIAM Journal on Numerical Analysis, 7(1), 1–46. https://doi.org/10.1137/0707001.

Demmel, J. W. (1997). Applied numerical linear algebra. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611971446.

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. Psychometrika, 1(3), 211–218. https://doi.org/10.1007/BF02288367.

Edelman, A., Arias, T. A., & Smith, S. T. (1998). The geometry of algorithms with orthogonality constraints. SIAM Journal on Matrix Analysis and Applications, 20(2), 303–353. https://doi.org/10.1137/S0895479895290954.

Golub, G. H., & Van Loan, C. F. (2013). Matrix computations (4th ed.). Johns Hopkins University Press.

Higham, N. J. (2002). Accuracy and stability of numerical algorithms (2nd ed.). Society for Industrial and Applied Mathematics.

Horn, R. A., & Johnson, C. R. (2013). Matrix analysis (2nd ed.). Cambridge University Press.

Krizhevsky, A. (2009). Learning multiple layers of features from tiny images. University of Toronto.

Li, R.-C., Truhar, N., & Zhang, L.-H. (2024). On Stewart's perturbation theorem for SVD. arXiv. https://arxiv.org/abs/2406.06901.

Mishra, B., Meyer, G., & Sepulchre, R. (2014). Low-rank optimization for distance matrix completion. IEEE Transactions on Signal Processing, 62(12), 3212–3223. https://doi.org/10.1109/TSP.2014.2324995.

Nocedal, J., & Wright, S. J. (2006). Numerical optimization (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-40065-5.

O'Rourke, S., Vu, V., & Wang, K. (2018). Random perturbation of low-rank matrices: Improving classical bounds. *Linear Algebra and Its Applications*, 540, 26–59. https://doi.org/10.1016/j.laa.2017.12.009.

O'Rourke, S., Vu, V., & Wang, K. (2024). Matrices with Gaussian noise: Optimal estimates for singular subspace perturbation. *IEEE Transactions on Information Theory*, 70(3), 1978–2002. https://doi.org/10.1109/TIT.2023.3328044.

Parlett, B. N. (1998). The symmetric eigenvalue problem. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611971163.

Recht, B., Fazel, M., & Parrilo, P. A. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Review, 52(3), 471–501. https://doi.org/10.1137/070697835.

Ring, W., & Wirth, B. (2012). Optimization methods on Riemannian manifolds and their application to shape space. SIAM Journal on Optimization, 22(2), 596–627. https://doi.org/10.1137/110831822.

Saad, Y. (2011). Numerical methods for large eigenvalue problems (2nd ed.). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611970739.

Sato, H. (2022). Riemannian optimization and its applications. Journal of the Operations Research Society of Japan, 65(3), 143–163. https://doi.org/10.15807/jorsj.65.143.

Stewart, G. W. (1973). Error and perturbation bound for subspaces associated with certain eigenvalue problems. SIAM Review, 15(4), 727–764. https://doi.org/10.1137/1015098.

Stewart, G. W. (1977). On the perturbation of pseudo-inverses, projections and linear least squares problems. SIAM Review, 19(4), 634–662. https://doi.org/10.1137/1019089.

Stewart, G. W., & Sun, J. G. (1990). Matrix perturbation theory. Academic Press..

Trefethen, L. N., & Bau, D. (1997). Numerical linear algebra. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898719574.

Vandereycken, B. (2013). Low-rank matrix completion by Riemannian optimization. SIAM Journal on Optimization, 23(2), 1214–1236. https://doi.org/10.1137/110845768.

Vannieuwenhoven, N. (2024). The condition number of singular subspaces, revisited. Linear Algebra and Its Applications. Advance online publication. https://doi.org/10.1016/j.laa.2024.08.012.

Wedin, P.-A. (1972). Perturbation bounds in connection with singular value decomposition. BIT Numerical Mathematics, 12(1), 99–111. https://doi.org/10.1007/BF01932678.

Wen, Z., & Yin, W. (2013). A feasible method for optimization with orthogonality constraints. Mathematical Programming, 142(1), 397–434. https://doi.org/10.1007/s10107-012-0584-9.

Wilkinson, J. H. (1965). The algebraic eigenvalue problem. Clarendon Press.

Zimmermann, R., & Hüper, K. (2022). Computing the Riemannian logarithm on the Stiefel manifold: Metrics, methods, and performance. SIAM Journal on Matrix Analysis and Applications, 43(2), 953–980. https://doi.org/10.1137/21M1426137.