mathematics

MDPI

Article

Operator-Inequality Defects in Transform-Based Tensor Algebras: Invariance and Transform Sensitivity

Shih-Yu Chang [1,*] and Michael K. Ng [2]

[1] Department of Applied Data Science, San Jose State University, San Jose, CA 95192, USA
[2] Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong; michael-ng@hkbu.edu.hk
* Correspondence: shihyu.chang@sjsu.edu

Abstract

Transform-based tensor algebras induced by invertible linear transforms provide a unified framework for tensor operations and spectral analysis. This paper investigates how the choice of a unitary transform affects the quantitative tightness of classical operator inequalities. We first formulate a transform-domain transfer principle for matrix trace inequalities and examine its consequences for the Golden–Thompson, Jensen-type, Klein, and Lieb inequalities. We then study the Golden–Thompson defect through its exact slice-wise decomposition. Under bounded spectral assumptions, we establish an upper estimate in terms of squared transform-domain commutators and a matching-order lower estimate for an explicit Pauli-matrix family. We further establish qualitative transform sensitivity: explicit separation results for real orthogonal transforms, including a DFT–DCT separation at $p = 4$. Finally, we investigate the non-universality of transform selection and formulate commutator-energy minimization as an operator-aware surrogate objective over the orthogonal group. The framework separates algebraic invariance from transform-dependent inequality tightness without assuming general equivalence between the defect and commutator-energy objectives.

Keywords: transform-based tensor algebra; $\star_M$-product; Golden–Thompson inequality; operator-inequality defect; noncommutativity; transform selection

MSC: 47A63; 15A27; 15A69

1. Introduction

Tensor algebra based on transform-induced products has become a central tool in modern multilinear analysis [1,2]. The $\star_M$-product framework of Kernfeld, Kilmer, and Aeron [3] defines a tensor–tensor product for any fixed invertible transform M, and the subsequent $\star_M$-algebra developed by Kilmer et al. [4] provides its algebraic and spectral foundations. The operator-theoretic interpretation of third-order tensors under the t-product was developed in [5]. When M is the discrete Fourier transform, this reduces to the classical t-product of Kilmer and Martin [6]. In this study, we specialize to unitary transforms Φ and study operator inequalities within this established algebra. This flexibility has led to a wide range of applications [7,8], including tensor completion, robust PCA, and signal processing, where different unitary transforms induce different notions of structure and sparsity.

A fundamental feature of these frameworks is that different choices of Φ lead to algebraically isomorphic tensor algebras. Consequently, the unitary transform is often

Check for updates

Academic Editor: Simeon Reich

Received: 30 September 2026
Revised: 29 October 2026
Accepted: 5 November 2026
Published: 9 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

viewed as a change in basis that does not alter intrinsic properties. However, this viewpoint overlooks an important phenomenon:

> *(Transform-dependent noncommutativity) The choice of transform Φ defines a coordinate system in which noncommutativity is measured, and different unitary transforms reveal fundamentally different commutative structures.*

This observation suggests that while algebraic properties may be invariant, quantitative behavior can depend strongly on Φ. The goal of this paper is to develop a systematic theory of this phenomenon in the context of classical operator inequalities.

Classical operator inequalities for matrices—including the Golden–Thompson inequality [9,10] and related Jensen-, Klein-, and Lieb-type inequalities [11]—play a foundational role in matrix analysis and operator theory. Related inequalities for t-product tensors have been studied in [12,13]. Their significance lies not only in their validity, but also in how their *defect* (slack) reflects noncommutativity. A natural question arises in the tensor setting: *Does the choice of unitary transform Φ affect operator inequalities, or is it merely a change of basis?*

At the algebraic level, the answer is straightforward: operator inequalities remain valid for all Φ with exactly the same constants as in the matrix case. The algebraic structure of the $\star_M$ algebra is already well understood [3,4]. Our starting point is the observation that this algebraic invariance conceals a deeper phenomenon: *Algebraic invariance does not imply quantitative invariance*.

While operator inequalities remain valid under all unitary transforms, their defect depends explicitly on transform-domain noncommutativity. For two tensors of the same size, we write $\mathcal{A} \oplus_\Phi \mathcal{B}$ for the tensor whose transform-domain lifting is the sum of the corresponding liftings:

$$\overline{\mathcal{A} \oplus_\Phi \mathcal{B}}_\Phi = \overline{\mathcal{A}}_\Phi + \overline{\mathcal{B}}_\Phi.$$

Since the transform is linear, this operation coincides with ordinary tensor addition. To formalize this phenomenon, we introduce the Φ-dependent Golden–Thompson defect

$$\Delta_\Phi(\mathcal{A}, \mathcal{B}) := \mathrm{Tr}_\Phi\big(\exp_\Phi(\mathcal{A}) \star_\Phi \exp_\Phi(\mathcal{B})\big) - \mathrm{Tr}_\Phi\big(\exp_\Phi(\mathcal{A} \oplus_\Phi \mathcal{B})\big), \tag{1}$$

where $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{n\times n\times p}$ are Φ-Hermitian tensors whose transformed slices $\widehat{\mathcal{A}}_\Phi^{(\omega)}$ and $\widehat{\mathcal{B}}_\Phi^{(\omega)}$ are Hermitian for $\omega = 1, \ldots, p$ (see Definition 4), $\mathrm{Tr}_\Phi(\cdot)$ and $\exp_\Phi(\cdot)$ are the Φ-trace and Φ-exponential maps (see Definitions 2 and 3), and $\star_\Phi(\cdot,\cdot)$ is the $\star_\Phi$-product between two tensors (see Definition 1). The quantity $\Delta_\Phi(\mathcal{A}, \mathcal{B})$ measures the failure of exact commutativity in the transform-based tensor algebra and quantifies the deviation between multiplicative and additive spectral propagation under the Φ-exponential map. Under a bounded spectral-domain assumption, we show that this defect is quantitatively controlled by the aggregate transform-domain commutator energy:

$$\sum_{\omega=1}^{p} \left\| \left[\widehat{\mathcal{A}}_\Phi^{(\omega)}, \widehat{\mathcal{B}}_\Phi^{(\omega)} \right] \right\|_F^2.$$

A central consequence of this viewpoint is that no transform is universally optimal. While certain transforms, such as the discrete Fourier transform, may yield vanishing defect for specific tensor pairs, other transforms, such as the discrete cosine transform, may produce significantly smaller defects for different tensor configurations. Thus, optimal transform selection is inherently data-dependent. In this sense, the choice of Φ is not merely a change of basis, but rather a choice of spectral coordinate system that determines how commutativity is evaluated in the transform domain.

The main contributions of this paper are summarized as follows:

- (Established foundation) We work within the transform-based $\star_M$ tensor algebra of Kernfeld–Kilmer–Aeron [3] and its $\star_M$-algebra formulation [4], specialized to unitary transforms Φ.
- (Operator-inequality transfer principle) Classical trace inequalities, including Golden–Thompson, Jensen, Klein, and Lieb, transfer to the transform-based tensor algebra with the same constants as in the matrix case.
- (Defect characterization) Under a bounded spectral-domain assumption, the Golden–Thompson defect $\Delta_\Phi(\mathcal{A}, \mathcal{B})$ is quantitatively controlled by the transform-domain commutator energy:
$$\sum_{\omega=1}^{p} \left\| [\widehat{A}_\Phi^{(\omega)}, \widehat{\mathcal{B}}_\Phi^{(\omega)}] \right\|_F^2.$$
- (Transform sensitivity) The Golden–Thompson defect depends nontrivially on the choice of the unitary transform Φ. For certain tensor pairs, choosing Φ as the discrete Fourier transform (DFT) yields exact slice-wise commutativity in the transform domain, providing
$$\Delta_{\mathrm{DFT}}(\mathcal{A}, \mathcal{B}) = 0.$$
In contrast, choosing Φ as the discrete cosine transform (DCT) may produce significant transform-domain noncommutativity, leading to a strictly positive defect in suitable regimes. This demonstrates that transform selection fundamentally affects the geometry of noncommutativity in the transform-based tensor algebra.
- (Non-universality and operator-aware transform selection) No unitary transform minimizes the Golden–Thompson defect for all tensor pairs. Optimal transforms are necessarily data-dependent. Transform selection can be formulated as minimizing transform-domain noncommutativity over the orthogonal group, thereby connecting the problem to approximate joint diagonalization of tensor slices.

These results establish that Φ induces a nontrivial geometry of inequality tightness, governed by transform-dependent noncommutativity. To the best of our knowledge, this is the first work that separates algebraic invariance from quantitative behavior in transform-based tensor algebras.

The organization of this paper is as follows. Section 2 introduces the transform-based tensor algebra and notation. Section 3 establishes the operator-inequality transfer principle. Section 4 develops the transform-sensitive defect theory, including quantitative bounds and separation results. Section 5 investigates optimization formulations for operator-aware transform selection. Section 6 contains concluding remarks and discusses broader implications and connections to numerical linear algebra and operator theory.

2. The Transform-Based Tensor Algebra

This section recalls the transform-based tensor algebra that underlies the operator-inequality analysis in the remainder of the paper. The tensor product used here is the established $\star_M$-product of Kernfeld, Kilmer, and Aeron [3] and its $\star_M$-algebra formulation [4], specialized to unitary transforms $\Phi \in \mathbb{C}^{p\times p}$. We denote the resulting product by $\star_\Phi$ and refer to it as the $\star_\Phi$-product, or simply the Φ-product when the transform is clear from context. We then recall the associated block-diagonal representation and the basic operations—trace, exponential, and logarithm—that will be used to state and analyze operator inequalities in later sections.

2.1. Definition of the $\star_\Phi$-Product and Basic Operations

Let $\mathcal{A} \in \mathbb{C}^{n_1\times n_2\times p}$. For a unitary transform $\Phi \in \mathbb{C}^{p\times p}$, define the transformed tensor $\hat{A}_\Phi \in \mathbb{C}^{n_1\times n_2\times p}$ by applying Φ to each tube (i.e., each third-mode fiber) of $\mathcal{A}$. Denote by

$\hat{A}_\Phi^{(i)} \in \mathbb{C}^{n_1 \times n_2}$ the i-th frontal slice of $\hat{A}_\Phi$. The associated block-diagonal matrix $\overline{A}_\Phi \in \mathbb{C}^{pn_1 \times pn_2}$ is defined as

$$\overline{A}_\Phi = \mathrm{blockdiag}(\hat{A}_\Phi^{(1)}, \hat{A}_\Phi^{(2)}, \dots, \hat{A}_\Phi^{(p)}).$$

Definition 1 ($\star_\Phi$-product, following [3,4]). *For $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times p}$ and $\mathcal{B} \in \mathbb{C}^{n_2 \times n_3 \times p}$, the $\star_\Phi$-product $\mathcal{C} = \mathcal{A} \star_\Phi \mathcal{B} \in \mathbb{C}^{n_1 \times n_3 \times p}$ is defined by*

$$\overline{C}_\Phi = \overline{A}_\Phi \cdot \overline{B}_\Phi,$$

where $\overline{C}_\Phi$ is the block-diagonal matrix constructed from the transformed tensor $\hat{C}_\Phi$ in the same manner as above. This is the $\star_M$-product of [3,4] with $M = \Phi$.

Remark 1. *When Φ is the discrete Fourier transform (DFT) matrix, the $\star_\Phi$-product reduces to the conventional t-product of Kilmer and Martin [6]. More generally, for an invertible transform M, the construction above coincides with the transform-based tensor–tensor product of [3].*

2.2. Φ-Trace, Φ-Exponential, and Φ-Logarithm

Definition 2 (Φ-trace). *Let $\mathcal{A} \in \mathbb{C}^{n \times n \times p}$. The Φ-trace of $\mathcal{A}$ is defined by*

$$\mathrm{Tr}_\Phi(\mathcal{A}) := \mathrm{Tr}(\overline{A}_\Phi),$$

where $\mathrm{Tr}(\cdot)$ denotes the standard matrix trace.

Definition 3 (Φ-exponential and Φ-logarithm). *Let $\mathcal{A} \in \mathbb{C}^{n \times n \times p}$ be Φ-Hermitian. The Φ-exponential and Φ-logarithm are defined via the block-diagonal lifting as*

$$\exp_\Phi(\mathcal{A}) := \Phi^H[\exp(\overline{A}_\Phi)], \qquad \log_\Phi(\mathcal{A}) := \Phi^H[\log(\overline{A}_\Phi)],$$

where the functions $\exp(\cdot)$ and $\log(\cdot)$ are applied blockwise to $\overline{A}_\Phi$, i.e., to each diagonal block $\hat{A}_\Phi^{(i)}$.

Here, $\Phi^H[\cdot]$ denotes the inverse transform applied to each mode-3 fiber, i.e., the inverse of the map $\mathcal{X} \mapsto \hat{\mathcal{X}}_\Phi$. It does not denote ordinary matrix multiplication of Φ^H with the block-diagonal matrix.

2.3. Φ-Hermitian Structure, Commutator, and Spectral Mapping

Definition 4 (Φ-Hermitian). *Let $\mathcal{A} \in \mathbb{C}^{n \times n \times p}$. We say that $\mathcal{A}$ is Φ-Hermitian if its block-diagonal lifting $\overline{A}_\Phi$ is Hermitian, i.e.,*

$$\overline{A}_\Phi^H = \overline{A}_\Phi.$$

Equivalently, each transformed frontal slice $\hat{A}_\Phi^{(i)}$ is Hermitian for all $i = 1, \dots, p$.

Definition 5 (Φ-commutator). *Let $\mathcal{A}, \mathcal{B}$ be tensors of compatible sizes. The Φ-commutator is defined by*

$$[\mathcal{A}, \mathcal{B}]_\Phi := \overline{A}_\Phi \overline{B}_\Phi - \overline{B}_\Phi \overline{A}_\Phi,$$

where the right-hand side is the ordinary matrix commutator.

Lemma 1 (Spectral mapping). *Let $\mathcal{A} \in \mathbb{C}^{n \times n \times p}$ be Φ-Hermitian, and let f be an analytic function defined on an open set containing the spectrum of $\overline{A}_\Phi$. Then,*

$$f_\Phi(\mathcal{A}) = \Phi^H[f(\overline{A}_\Phi)],$$

where $f(\overline{A}_\Phi)$ is applied blockwise to each diagonal block. This is the standard functional calculus in the transform-domain representation of the $\star_M$ algebra [3,4].

Proof. Since $\mathcal{A}$ is Φ-Hermitian, the matrix $\overline{\mathcal{A}}_\Phi$ is Hermitian and hence unitarily diagonalizable. Thus, there exists a unitary matrix U such that

$$\overline{\mathcal{A}}_\Phi = U\Lambda U^H,$$

where Λ is diagonal containing the eigenvalues of $\overline{\mathcal{A}}_\Phi$.

Using the standard functional calculus for Hermitian matrices,

$$f(\overline{\mathcal{A}}_\Phi) = Uf(\Lambda)U^H,$$

where $f(\Lambda)$ is obtained by applying f to each diagonal entry.

By definition of the Φ-functional calculus (cf. Section 2.2),

$$f_\Phi(\mathcal{A}) = \Phi^H[f(\overline{\mathcal{A}}_\Phi)],$$

which establishes the claim. □

Remark 2. *The Φ-Hermitian condition is essential for the spectral mapping lemma, as it guarantees that $\overline{\mathcal{A}}_\Phi$ is unitarily diagonalizable with real eigenvalues. For non-Hermitian tensors, the exponential and logarithm remain well-defined via the blockwise matrix exponential and logarithm, but the spectral interpretation (e.g., eigenvalue mapping) does not hold in general.*

3. An Operator-Inequality Transfer Principle

This section records a transfer principle for classical operator inequalities in the transform-based tensor algebra. The purpose of this section is two-fold. First, we recall that the standard trace inequalities—including Golden–Thompson, Jensen, Klein, and Lieb—transfer to the transform-based tensor algebra with the same constants as in the matrix case. Second, we emphasize that this transfer is a direct consequence of the established transform-domain representation of the $\star_M$ algebra [3,4]; it is therefore not the main contribution of this study. The main contribution, developed in Section 4, concerns the quantitative behavior of the associated *defect* under a change of transform.

3.1. Transfer Principle

The following transfer principle separates the algebraic validity of a matrix inequality from its quantitative behavior under a change in transform. Its purpose is to identify which conclusions follow directly from the established block-diagonal representation before examining the transform-dependent defect in Section 4.

Theorem 1 (Operator-inequality transfer principle)**.** *Let $\mathcal{J}(X_1, \dots, X_m)$ be a trace or functional inequality that is preserved under finite direct sums of matrices. More precisely, assume that for all Hermitian matrices $X_1, \dots, X_m$ of compatible sizes, the inequality $\mathcal{J}(X_1, \dots, X_m) \geq 0$ holds, and that the same inequality holds for finite direct sums*

$$\mathcal{J}\left(\bigoplus_{i=1}^{p} X_1^{(i)}, \dots, \bigoplus_{i=1}^{p} X_m^{(i)}\right) \geq 0$$

whenever $\mathcal{J}(X_1^{(i)}, \dots, X_m^{(i)}) \geq 0$ for each i.

Then, for every unitary transform $\Phi \in \mathbb{C}^{p\times p}$, the corresponding inequality holds in the transform-based tensor algebra slice-wise. In particular, if $\mathcal{A}_1, \dots, \mathcal{A}_m \in \mathbb{C}^{n\times n\times p}$ are Φ-Hermitian tensors, then

$$\mathcal{J}_\Phi(\mathcal{A}_1, \dots, \mathcal{A}_m) \geq 0,$$

where $\mathcal{J}_\Phi$ denotes the Φ-lift of $\mathcal{J}$ obtained by replacing matrix operations with their Φ-counterparts.

Proof. Using Proposition 1, the transform-based tensor algebra is isomorphic to the block-diagonal matrix algebra

$$\mathcal{D}_\Phi = \{\text{blockdiag}(X_1, \dots, X_p) : \ X_i \in \mathbb{C}^{n\times n}\}.$$

Under this isomorphism, every Φ-Hermitian tensor $\mathcal{A}$ corresponds to a block-diagonal Hermitian matrix:

$$\overline{\mathcal{A}}_\Phi = \text{blockdiag}(\hat{A}_\Phi^{(1)}, \dots, \hat{A}_\Phi^{(p)}).$$

Since the inequality $\mathcal{J}$ is preserved under finite direct sums, it holds blockwise for each slice tuple $(\hat{A}_1^{(\omega)}, \dots, \hat{A}_m^{(\omega)})$, $\omega = 1, \dots, p$: applying $\mathcal{J}$ to the direct sum $\bigoplus_{\omega=1}^p (\hat{A}_1^{(\omega)}, \dots, \hat{A}_m^{(\omega)})$ and using the direct-sum preservation yields $\mathcal{J}(\bigoplus_\omega \hat{A}_1^{(\omega)}, \dots, \bigoplus_\omega \hat{A}_m^{(\omega)}) \geq 0$. For trace inequalities, the trace of a block-diagonal matrix is the sum of the traces of its blocks, so this blockwise inequality is equivalent to $\sum_{\omega=1}^p \mathcal{J}(\hat{A}_1^{(\omega)}, \dots, \hat{A}_m^{(\omega)}) \geq 0$, which is exactly $\mathcal{J}_\Phi(\mathcal{A}_1, \dots, \mathcal{A}_m) \geq 0$. □

Because the block-diagonal functional calculus acts independently on each transformed slice, the matrix inequality applies to every block. For trace inequalities, the block contributions are additive. Therefore, summing the valid slice-wise inequalities establishes the tensor trace inequality:

$$\mathcal{J}_\Phi(\mathcal{A}_1, \dots, \mathcal{A}_m) = \sum_{\omega=1}^{p} \mathcal{J}(\hat{A}_1^{(\omega)}, \dots, \hat{A}_m^{(\omega)}) \geq 0.$$

This argument uses only the algebraic representation and the direct-sum property; it does not imply that the numerical slack of the inequality is invariant under changes of the transform.

Remark 3. *The transfer principle is a direct consequence of the standard transform-domain representation of the $\star_M$ algebra [3,4]. Its role here is not to establish new inequalities, but to fix the algebraic setting in which the transform-dependent defect is analyzed in Section 4.*

3.2. Corollaries: Golden–Thompson, Jensen, Klein, and Lieb

We now record the four classical inequalities as corollaries of Theorem 1. Throughout, $\oplus_\Phi$ denotes the Φ-sum, i.e.,

$$\overline{\mathcal{A} \oplus_\Phi \mathcal{B}}_\Phi = \overline{\mathcal{A}}_\Phi + \overline{\mathcal{B}}_\Phi.$$

Corollary 1 (Φ-Golden–Thompson inequality)**.** *Let $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{n\times n\times p}$ be Φ-Hermitian. Then,*

$$\text{Tr}_\Phi(\exp_\Phi(\mathcal{A} \oplus_\Phi \mathcal{B})) \ \leq \ \text{Tr}_\Phi(\exp_\Phi(\mathcal{A}) \star_\Phi \exp_\Phi(\mathcal{B})).$$

Proof. Apply Theorem 1 to the classical Golden–Thompson inequality,

$$\text{Tr}(e^{A+B}) \ \leq \ \text{Tr}(e^A e^B),$$

which holds for all Hermitian matrices A, B and is preserved under finite direct sums. Replacing A and B by the block-diagonal lifts $\overline{\mathcal{A}}_\Phi$ and $\overline{\mathcal{B}}_\Phi$ yields the claim. □

Corollary 2 (Φ-Jensen inequality)**.** *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a convex function with $\phi(0) = 0$. If $\mathcal{A} \in \mathbb{C}^{n\times n\times p}$ is Φ-Hermitian, then*

$$\text{Tr}_\Phi(\phi(\mathcal{A})) \geq 0.$$

In the scalar case $n = p = 1$, the inequality reduces to $\phi(a) \geq 0$ for $a \in \mathbb{R}$, which follows from convexity and $\phi(0) = 0$. No normalization by np is involved, and the previous counterexample $\phi(x) = x$ with $\Phi\mathcal{A} = -1\Phi$ is excluded by the hypothesis $\phi(0) = 0$.

Proof. Apply Theorem 1 to the classical Jensen trace inequality $\mathrm{Tr}(\phi(A)) \geq 0$ for Hermitian A and convex ϕ with $\phi(0) = 0$, which is preserved under finite direct sums. Replacing A by the block-diagonal lift $\overline{\mathcal{A}}_\Phi$ yields $\mathrm{Tr}(\phi(\overline{\mathcal{A}}_\Phi)) \geq 0$. Since $\mathrm{Tr}_\Phi(\phi(\mathcal{A})) = \mathrm{Tr}(\phi(\overline{\mathcal{A}}_\Phi))$, the claim follows. □

Corollary 3 (Φ-Klein inequality)**.** *Let ϕ be a convex function. If $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{n\times n\times p}$ are Φ-Hermitian, then*

$$\mathrm{Tr}_\Phi\Big(\phi(\mathcal{A}) - \phi(\mathcal{B}) - (\mathcal{A} - \mathcal{B})\,\phi'(\mathcal{B})\Big) \geq 0.$$

Proof. Apply Theorem 1 to the classical Klein inequality for Hermitian matrices. □

Corollary 4 (Φ-Lieb concavity)**.** *The function*

$$(\mathcal{A}, \mathcal{B}) \longmapsto \mathrm{Tr}_\Phi\big(\exp_\Phi\big(\log_\Phi(\mathcal{A}) + \log_\Phi(\mathcal{B})\big)\big)$$

is jointly concave for $\mathcal{A} \succ_\Phi 0$ and $\mathcal{B} \succ_\Phi 0$. In the scalar case $n = p = 1$, the map reduces to $(a, b) \mapsto ab$ on $a, b > 0$. This map is not jointly concave in the naive sense, but the classical Lieb theorem is stated for the equivalent reformulation $(A, B) \mapsto \mathrm{Tr}(A^r B^{1-r})$ with $r \in [0, 1]$, which is jointly concave; the present statement follows this standard formulation.

Proof. Apply Theorem 1 to Lieb's classical joint concavity theorem for positive definite matrices, which states that $(A, B) \mapsto \mathrm{Tr}\exp(\log A + \log B)$ is jointly concave on the cone of positive definite matrices (equivalently, $(A, B) \mapsto \mathrm{Tr}(A^r B^{1-r})$ is jointly concave for $r \in [0, 1]$). Replacing A, B with the block-diagonal lifts $\overline{\mathcal{A}}_\Phi, \overline{\mathcal{B}}_\Phi$ and using $\mathrm{Tr}_\Phi(\cdot) = \mathrm{Tr}(\bar{\cdot}_\Phi)$ yields the claim. □

Remark 4. *Corollaries 1–4 show that the classical matrix trace inequalities transfer to transform-based tensor algebra with the same constants as in the matrix setting, independent of the choice of unitary transform Φ. This invariance follows from the transform-domain representation of the $\star_M$ algebra [3,4]. However, as we shall see in Section 4, the* quantitative tightness *of these inequalities—specifically the Golden–Thompson defect—is* not *invariant and depends critically on the choice of Φ.*

3.3. Background: Transform-Domain Representation

For completeness, we recall the transform-domain representation that underlies Theorem 1. This representation is standard in the $\star_M$-algebra literature [3,4]; we state it here only to fix notation and to make the subsequent defect analysis self-contained.

Proposition 1 (Transform-domain representation, cf. [3,4])**.** *Let*

$$\mathcal{T}_\Phi := \left(\mathbb{C}^{n\times n\times p}, \star_\Phi\right)$$

denote the transform-based tensor algebra associated with the unitary transform Φ. Then, the mapping

$$\Psi_\Phi : \mathcal{T}_\Phi \;\to\; \mathcal{D}_\Phi := \left\{\mathrm{blockdiag}(X_1, \ldots, X_p) :\; X_i \in \mathbb{C}^{n\times n}\right\}, \quad \Psi_\Phi(\mathcal{A}) = \overline{\mathcal{A}}_\Phi,$$

is an algebra isomorphism. That is,

1. $\Psi_\Phi(\mathcal{A} \star_\Phi \mathcal{B}) = \Psi_\Phi(\mathcal{A})\,\Psi_\Phi(\mathcal{B})$,
2. Ψ_Φ *is bijective.*

Proof. The map Ψ_Φ is linear because the mode-3 transform and block-diagonal assembly are linear.

By the definition of the $\star_\Phi$-product, $\Psi_\Phi(\mathcal{A} \star_\Phi \mathcal{B}) = \Psi_\Phi(\mathcal{A})\Psi_\Phi(\mathcal{B})$, so Ψ_Φ preserves multiplication.

If $\Psi_\Phi(\mathcal{A}) = 0$, every transformed frontal slice of $\mathcal{A}$ vanishes; since Φ is invertible, $\mathcal{A} = 0$, proving injectivity.

Conversely, for any block-diagonal matrix $X = \mathrm{blockdiag}(X_1, \dots, X_p)$, form a transformed tensor with frontal slices $X_1, \dots, X_p$ and apply the inverse mode-3 transform. The resulting tensor $\mathcal{A}$ satisfies $\Psi_\Phi(\mathcal{A}) = X$; hence, Ψ_Φ is surjective.

Therefore, Ψ_Φ is an algebra isomorphism. □

Remark 5 (Limits of the representation). *The representation Ψ_Φ preserves algebraic structure and functional calculus, which explains the transfer results above. However, it does* not *preserve all quantitative properties of interest. In particular, the* tightness *of the Golden–Thompson inequality—measured by the defect $\Delta_\Phi(\mathcal{A}, \mathcal{B})$—depends critically on the choice of Φ. This sensitivity arises because the representation, while preserving the product, does not preserve the* commutativity structure *of the tensor slices in the transform domain. Different transforms Φ reveal different levels of noncommutativity, leading to transform-dependent defect magnitudes. We analyze this sensitivity in Section 4.*

4. Transform-Dependent Golden–Thompson Defect

Having fixed the algebraic setting in Section 3, we now isolate the quantitative object whose transform dependence is the subject of this paper. The section proceeds in three steps: the slice-wise reduction in the defect (Lemma 2), the bounded-spectral upper and lower estimates (Theorem 2), and the qualitative separation results (Theorems 3 and 4).

4.1. Golden–Thompson Defect

The classical Golden–Thompson inequality is tight when the two matrices commute. In the transform-based tensor setting, the gap between the two sides of the inequality quantifies the effect of noncommutativity in the transform domain. This gap is the central object of our sensitivity analysis.

Definition 6 (Golden–Thompson defect). *Let $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{n\times n\times p}$ be Φ-Hermitian tensors. The Golden–Thompson defect is defined by*

$$\Delta_\Phi(\mathcal{A}, \mathcal{B}) := \mathrm{Tr}_\Phi\big(\exp_\Phi(\mathcal{A}) \star_\Phi \exp_\Phi(\mathcal{B})\big) - \mathrm{Tr}_\Phi\big(\exp_\Phi(\mathcal{A} \oplus_\Phi \mathcal{B})\big).$$

Using Corollary 1, we have

$$\Delta_\Phi(\mathcal{A}, \mathcal{B}) \geq 0.$$

Remark 6 (Role of the defect). *The defect $\Delta_\Phi(\mathcal{A}, \mathcal{B})$ is the natural measure of inequality tightness for three reasons:*

1. *It vanishes exactly when $\mathcal{A}$ and $\mathcal{B}$ commute in the transform domain, i.e., $[\overline{\mathcal{A}}_\Phi, \overline{\mathcal{B}}_\Phi] = 0$.*
2. *It is quantitatively controlled by the transform-domain commutator.*
3. *It is sensitive to the choice of Φ, making it the right objective for operator-aware transform selection.*

4.2. Main Theorem: Transform-Sensitive Defect

We now formalize the central phenomenon of this work: while operator inequalities in the transform-based tensor algebra transfer with the same constants as in the matrix case, their *defect* exhibits quantitative dependence on transform-domain noncommutativity.

Lemma 2 (Slice-wise reduction of the defect)**.** *Let $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{n\times n\times p}$ be Φ-Hermitian tensors. Under the transform-domain representation $\Psi_\Phi(\mathcal{A}) = \overline{\mathcal{A}}_\Phi$ of Proposition 1, the Golden–Thompson defect decomposes as a sum of slice-wise classical defects:*

$$\Delta_\Phi(\mathcal{A}, \mathcal{B}) = \sum_{\omega=1}^{p} \Big(\mathrm{Tr}\big(e^{A_\omega} e^{B_\omega}\big) - \mathrm{Tr}\big(e^{A_\omega + B_\omega}\big) \Big),$$

where $A_\omega := \hat{\mathcal{A}}_\Phi^{(\omega)}$ and $B_\omega := \hat{\mathcal{B}}_\Phi^{(\omega)}$ are Hermitian matrices.

Proof. Using the definitions of Tr_Φ, $\exp_\Phi$, $\star_\Phi$, and $\oplus_\Phi$, and the representation Ψ_Φ from Proposition 1,

$$\Delta_\Phi(\mathcal{A}, \mathcal{B}) = \mathrm{Tr}\big(e^{\overline{\mathcal{A}}_\Phi} e^{\overline{\mathcal{B}}_\Phi}\big) - \mathrm{Tr}\big(e^{\overline{\mathcal{A}}_\Phi + \overline{\mathcal{B}}_\Phi}\big).$$

Since $\overline{\mathcal{A}}_\Phi$ and $\overline{\mathcal{B}}_\Phi$ are block-diagonal with diagonal blocks A_ω and B_ω, the trace decomposes as a sum over blocks. The exponential of a block-diagonal matrix is block-diagonal with exponentials of each block. Hence,

$$\Delta_\Phi(\mathcal{A}, \mathcal{B}) = \sum_{\omega=1}^{p} \Big(\mathrm{Tr}\big(e^{A_\omega} e^{B_\omega}\big) - \mathrm{Tr}\big(e^{A_\omega + B_\omega}\big) \Big).$$

The Hermiticity of A_ω and B_ω follows from the Φ-Hermitian assumption. This completes the proof. □

Theorem 2 (Transform-sensitive defect under bounded spectral domain)**.** *Let $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{n\times n\times p}$ be Φ-Hermitian tensors, and assume that there exists $R > 0$, such that*

$$\|\hat{\mathcal{A}}_\Phi^{(\omega)}\|_2 \le R, \qquad \|\hat{\mathcal{B}}_\Phi^{(\omega)}\|_2 \le R \quad \forall \omega = 1, \dots, p.$$

Then, the following statements hold.

(i) (Upper bound) There exists a constant $C_R > 0$, depending only on R, such that

$$\Delta_\Phi(\mathcal{A}, \mathcal{B}) \;\le\; C_R \sum_{\omega=1}^{p} \Big\| [\hat{\mathcal{A}}_\Phi^{(\omega)}, \hat{\mathcal{B}}_\Phi^{(\omega)}] \Big\|_F^2.$$

(ii) (Matching-order lower bound on an explicit family) For every $R > 0$ there exist $\varepsilon_0(R) > 0$ and a one-parameter family $(\mathcal{A}_\varepsilon, \mathcal{B}_\varepsilon)_{0<\varepsilon<\varepsilon_0(R)}$ in the bounded spectral class, such that

$$\Delta_\Phi(\mathcal{A}_\varepsilon, \mathcal{B}_\varepsilon) \ge c_R \sum_{\omega=1}^{p} \|[\widehat{\mathcal{A}}_{\Phi,\varepsilon}^{(\omega)}, \widehat{\mathcal{B}}_{\Phi,\varepsilon}^{(\omega)}]\|_F^2$$

for some $c_R > 0$. This lower estimate is asserted only for the constructed family.

(iii) (Transform sensitivity) There exist tensors $\mathcal{A}, \mathcal{B}$ and unitary transforms $\Phi_1, \Phi_2 \in \mathbb{C}^{p\times p}$, such that

$$\Delta_{\Phi_1}(\mathcal{A}, \mathcal{B}) = 0, \qquad \Delta_{\Phi_2}(\mathcal{A}, \mathcal{B}) > 0.$$

(iv) (Vanishing condition) If

$$[\hat{\mathcal{A}}_\Phi^{(\omega)}, \hat{\mathcal{B}}_\Phi^{(\omega)}] = 0, \quad \forall \omega = 1, \dots, p,$$

then

$$\Delta_\Phi(\mathcal{A}, \mathcal{B}) = 0.$$

Consequently, on the bounded spectral class $\|\hat{\mathcal{A}}_\Phi^{(\omega)}\|_2, \|\hat{\mathcal{B}}_\Phi^{(\omega)}\|_2 \le R$, the defect satisfies the upper bound in (i) for all admissible tensors, and the lower bound in (ii) on the one-parameter family $(\mathcal{A}_\varepsilon, \mathcal{B}_\varepsilon)_{0<\varepsilon<\varepsilon_0(R)}$ constructed in the proof of (ii). For this family, the two-sided estimate

$$c_R \sum_{\omega=1}^{p} \Big\| [\hat{\mathcal{A}}_\Phi^{(\omega)}, \hat{\mathcal{B}}_\Phi^{(\omega)}] \Big\|_F^2 \;\le\; \Delta_\Phi(\mathcal{A}, \mathcal{B}) \;\le\; C_R \sum_{\omega=1}^{p} \Big\| [\hat{\mathcal{A}}_\Phi^{(\omega)}, \hat{\mathcal{B}}_\Phi^{(\omega)}] \Big\|_F^2$$

holds on the explicit family of part (ii). A general lower bound valid for all tensors in the bounded spectral class is not claimed; see Remark 7.

Proof. Using Lemma 2, we obtain the slice-wise decomposition

$$\Delta_\Phi(\mathcal{A},\mathcal{B}) = \sum_{\omega=1}^{p} \delta(A_\omega, B_\omega),$$

where

$$\delta(A,B) := \mathrm{Tr}(e^A e^B) - \mathrm{Tr}(e^{A+B}) \geq 0$$

is the classical Golden–Thompson defect for Hermitian matrices, and $A_\omega = \hat{\mathcal{A}}_\Phi^{(\omega)}$, $B_\omega = \hat{\mathcal{B}}_\Phi^{(\omega)}$.

Part (i): Upper bound. For Hermitian matrices A, B with $\|A\|_2, \|B\|_2 \leq R$, classical quantitative versions of the Golden–Thompson inequality (see, e.g., [14]) provide a constant of $C_R > 0$, depending only on R, such that

$$\delta(A,B) \;\leq\; C_R \|[A,B]\|_F^2.$$

Applying this to each slice pair (A_ω, B_ω) and summing over $\omega = 1, \ldots, p$ yields

$$\Delta_\Phi(\mathcal{A},\mathcal{B}) \;\leq\; C_R \sum_{\omega=1}^{p} \|[A_\omega, B_\omega]\|_F^2,$$

which establishes part (i).

Part (ii): Lower bound. For every transformed slice, choose $A_\omega = \varepsilon\sigma_x$, $B_\omega = \varepsilon\sigma_z$, where

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Since $\sigma_x^2 = \sigma_z^2 = I$ and $\sigma_x\sigma_z + \sigma_z\sigma_x = 0$,

$$e^{A_\omega} = \cosh\varepsilon\, I + \sinh\varepsilon\, \sigma_x, \qquad e^{B_\omega} = \cosh\varepsilon\, I + \sinh\varepsilon\, \sigma_z,$$

so

$$e^{A_\omega} e^{B_\omega} = \cosh^2\varepsilon\, I + \cosh\varepsilon \sinh\varepsilon\, \sigma_x + \cosh\varepsilon \sinh\varepsilon\, \sigma_z + \sinh^2\varepsilon\, \sigma_x\sigma_z.$$

Taking traces and using $\mathrm{Tr}\, I = 2$, $\mathrm{Tr}\,\sigma_x = \mathrm{Tr}\,\sigma_z = 0$, $\mathrm{Tr}(\sigma_x\sigma_z) = 0$,

$$\mathrm{Tr}(e^{A_\omega} e^{B_\omega}) = 2\cosh^2\varepsilon.$$

Since $\sigma_x + \sigma_z$ is Hermitian with eigenvalues $\pm\sqrt{2}$,

$$\mathrm{Tr}(e^{A_\omega + B_\omega}) = 2\cosh(\sqrt{2}\,\varepsilon).$$

Hence,

$$\delta(A_\omega, B_\omega) = 2\cosh^2\varepsilon - 2\cosh(\sqrt{2}\,\varepsilon) = \frac{\varepsilon^4}{3} + \frac{\varepsilon^6}{15} + O(\varepsilon^8).$$

Moreover, $[A_\omega, B_\omega] = 2i\varepsilon^2\sigma_y$ and $\|\sigma_y\|_F^2 = 2$, so $\|[A_\omega, B_\omega]\|_F^2 = 8\varepsilon^4$. Therefore,

$$\delta(A_\omega, B_\omega) = \frac{1}{24}\|[A_\omega, B_\omega]\|_F^2 + O(\varepsilon^6).$$

For sufficiently small positive ε, there exists $c_R > 0$, such that $\delta(A_\omega, B_\omega) \geq c_R\|[A_\omega, B_\omega]\|_F^2$. Summing over ω yields the lower bound for the constructed family.

Part (iii): Transform sensitivity. Let $\Phi_1 = \Phi_{\mathrm{DFT}}$ and $\Phi_2 = \Phi_{\mathrm{DCT}}$. The qualitative separation is established in Theorems 3 and 4 below. In particular, Theorem 4 constructs tensors $\mathcal{A}, \mathcal{B}$, such that all slices commute in the DFT domain and hence $\Delta_{\Phi_1}(\mathcal{A},\mathcal{B}) = 0$, while the DCT domain produces a strictly positive defect, i.e., $\Delta_{\Phi_2}(\mathcal{A},\mathcal{B}) > 0$. The continuity of $\Phi \mapsto \Delta_\Phi(\mathcal{A},\mathcal{B})$ follows from the facts:

- $\hat{\mathcal{A}}_{\Phi}^{(\omega)}$ and $\hat{\mathcal{B}}_{\Phi}^{(\omega)}$ depend linearly on Φ;
- the matrix exponential and trace are continuous;
- the unitary group $\mathcal{U}(p)$ is compact.

Thus, the mapping is continuous and therefore locally Lipschitz on compact subsets of $\mathcal{U}(p)$.

Part (iv): Vanishing condition. If $[A_\omega, B_\omega] = 0$ for all ω, then each pair (A_ω, B_ω) consists of commuting Hermitian matrices. Commuting Hermitian matrices are simultaneously diagonalizable, which implies

$$e^{A_\omega} e^{B_\omega} = e^{A_\omega + B_\omega}.$$

Hence, $\delta(A_\omega, B_\omega) = 0$ for each ω. Summing over ω and applying Lemma 2 yields $\Delta_\Phi(\mathcal{A}, \mathcal{B}) = 0$.

This completes the proof of Theorem 2. □

Remark 7 (Scope of the lower bound)**.** *Part (ii) establishes the lower bound only on the explicitly constructed one-parameter family. All subsequent statements that invoke a two-sided estimate are to be read as restricted to this family, or are reformulated as surrogate statements in Section 5.*

Remark 8. *The constants C_R and c_R in Theorem 2 depend only on the spectral bound R, not on the tensor dimensions n and p or on the specific choice of Φ. This bounded spectral-domain formulation avoids the use of a universal constant, which would be inconsistent with the exponential scaling of the Golden–Thompson defect under rescaling $A \mapsto tA$, $B \mapsto tB$.*

Remark 9. *On the bounded spectral class of Theorem 2, Δ_Φ defines a non-negative, transform-dependent functional on the admissible transform space, whose magnitude is bounded above by the squared commutator energy under the bounded spectral assumption of Theorem 2(i). This endows the unitary group $\mathcal{U}(p)$ with a noncommutativity geometry; Δ_Φ is not asserted to be a quadratic functional of Φ.*

4.3. Qualitative Transform Separation: DFT vs. DCT

We now construct an explicit example showing that the Golden–Thompson defect can differ dramatically between two standard transforms. This example demonstrates that algebraic invariance does not imply quantitative invariance.

Theorem 3 (Explicit separation for real orthogonal transforms)**.** *Let $p \geq 2$ and let $\Phi_1, \Phi_2 \in O(p)$ be real orthogonal transforms, such that*

$$U := \Phi_2 \Phi_1^\top$$

is not *a signed permutation matrix. Then, there exist tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n \times n \times p}$ (with $n \geq 2$) that are both Φ_1-Hermitian and Φ_2-Hermitian and satisfy*

$$\Delta_{\Phi_1}(\mathcal{A}, \mathcal{B}) = 0, \qquad \Delta_{\Phi_2}(\mathcal{A}, \mathcal{B}) > 0.$$

Proof. Step 1: A row of U with two nonzero entries. Since $U \in O(p)$ and U is not a signed permutation matrix, U has a row containing at least two nonzero entries. Indeed, if every row of U had exactly one nonzero entry, then, since each row of an orthogonal matrix has unit Euclidean norm, that entry would have modulus 1; the rows would then be distinct standard basis vectors, so U would be a signed permutation matrix, contradicting the hypothesis. Hence, there exist a row index r and distinct column indices $j \neq k$, such that $U_{rj} \neq 0$ and $U_{rk} \neq 0$.

Step 2: Choice of noncommuting Hermitian matrices. Let

$$H_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad H_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

embedded in the upper-left 2×2 block of an $n \times n$ matrix and zero-padded elsewhere. Then, H_1, H_2 are real symmetric and

$$[H_1, H_2] = 2i\,\sigma_y \neq 0, \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},$$

so $[H_1, H_2] \neq 0$.

Step 3: Construction of the Φ_1-domain slices. Choose $\lambda \in \mathbb{R} \setminus \{1\}$ (e.g., $\lambda = 2$). Define frontal slices $\hat{\mathcal{A}}^{(\omega)}_{\Phi_1}$ and $\hat{\mathcal{B}}^{(\omega)}_{\Phi_1}$ for $\omega = 1, \ldots, p$ by

$$\hat{\mathcal{A}}^{(j)}_{\Phi_1} = H_1, \qquad \hat{\mathcal{B}}^{(j)}_{\Phi_1} = H_1,$$

$$\hat{\mathcal{A}}^{(k)}_{\Phi_1} = H_2, \qquad \hat{\mathcal{B}}^{(k)}_{\Phi_1} = \lambda H_2,$$

and $\hat{\mathcal{A}}^{(\omega)}_{\Phi_1} = \hat{\mathcal{B}}^{(\omega)}_{\Phi_1} = 0$ for all other ω.

Define the physical tensors by the inverse mode-3 transform:

$$\mathcal{A} := \Phi_1^H[\hat{\mathcal{A}}_{\Phi_1}], \qquad \mathcal{B} := \Phi_1^H[\hat{\mathcal{B}}_{\Phi_1}].$$

Because Φ_1 is real orthogonal, $\Phi_1^H = \Phi_1^\top$ is real, so $\mathcal{A}, \mathcal{B}$ are real tensors.

Step 4: Φ_1-Hermiticity and Φ_2-Hermiticity. Every slice $\hat{\mathcal{A}}^{(\omega)}_{\Phi_1}$ and $\hat{\mathcal{B}}^{(\omega)}_{\Phi_1}$ is either 0, H_1, H_2, or λH_2, and hence, real symmetric. Therefore, $\overline{\mathcal{A}}_{\Phi_1}$ and $\overline{\mathcal{B}}_{\Phi_1}$ are block-diagonal with real symmetric blocks, and hence, Hermitian; this is exactly Φ_1-Hermiticity.

For Φ_2-Hermiticity, we use the slice-wise mixing formula. For every ω,

$$\widehat{\mathcal{A}}^{(\omega)}_{\Phi_2} = \sum_{\ell=1}^{p} U_{\omega\ell}\, \widehat{\mathcal{A}}^{(\ell)}_{\Phi_1}, \qquad \widehat{\mathcal{B}}^{(\omega)}_{\Phi_2} = \sum_{\ell=1}^{p} U_{\omega\ell}\, \widehat{\mathcal{B}}^{(\ell)}_{\Phi_1}.$$

Since U is real and every $\widehat{\mathcal{A}}^{(\ell)}_{\Phi_1}$ and $\widehat{\mathcal{B}}^{(\ell)}_{\Phi_1}$ is real symmetric, each transformed slice $\widehat{\mathcal{A}}^{(\omega)}_{\Phi_2}$ and $\widehat{\mathcal{B}}^{(\omega)}_{\Phi_2}$ is a real linear combination of real symmetric matrices, hence real symmetric. Therefore, $\mathcal{A}, \mathcal{B}$ are Φ_2-Hermitian.

Step 5: $\Delta_{\Phi_1}(\mathcal{A}, \mathcal{B}) = 0$. For every ω, the pair $(\hat{\mathcal{A}}^{(\omega)}_{\Phi_1}, \hat{\mathcal{B}}^{(\omega)}_{\Phi_1})$ is one of

$$(0, 0), \quad (H_1, H_1), \quad (H_2, \lambda H_2),$$

so the two matrices are scalar multiples of a common matrix (or both zero) and, therefore, commute the following: $[\hat{\mathcal{A}}^{(\omega)}_{\Phi_1}, \hat{\mathcal{B}}^{(\omega)}_{\Phi_1}] = 0$ for all ω. Using Theorem 2(iv), $\Delta_{\Phi_1}(\mathcal{A}, \mathcal{B}) = 0$.

Step 6: A nonzero commutator in the Φ_2 domain. Set $a := U_{rj}$ and $b := U_{rk}$. For the distinguished row r, only $\ell = j$ and $\ell = k$ contribute:

$$\hat{\mathcal{A}}^{(r)}_{\Phi_2} = aH_1 + bH_2, \qquad \hat{\mathcal{B}}^{(r)}_{\Phi_2} = aH_1 + \lambda bH_2.$$

Their commutator is

$$\begin{aligned} [\hat{\mathcal{A}}^{(r)}_{\Phi_2}, \hat{\mathcal{B}}^{(r)}_{\Phi_2}] &= [aH_1 + bH_2,\, aH_1 + \lambda bH_2] \\ &= a\lambda b\,[H_1, H_2] + ab\,[H_2, H_1] \\ &= ab(\lambda - 1)\,[H_1, H_2]. \end{aligned}$$

Since $a \neq 0$, $b \neq 0$, $\lambda \neq 1$, and $[H_1, H_2] \neq 0$, we obtain $[\hat{\mathcal{A}}^{(r)}_{\Phi_2}, \hat{\mathcal{B}}^{(r)}_{\Phi_2}] \neq 0$.

Step 7: Strict positivity of Δ_{Φ_2}. Using Lemma 2,

$$\Delta_{\Phi_2}(\mathcal{A},\mathcal{B}) = \sum_{\omega=1}^{p} \delta\big(\hat{\mathcal{A}}_{\Phi_2}^{(\omega)}, \hat{\mathcal{B}}_{\Phi_2}^{(\omega)}\big),$$

where $\delta(A,B) = \mathrm{Tr}(e^A e^B) - \mathrm{Tr}(e^{A+B})$ is the classical Golden–Thompson defect for Hermitian matrices. Each summand is non-negative using the Golden–Thompson inequality. At $\omega = r$, the pair is Hermitian (by Step 4) and noncommuting (by Step 6); for Hermitian matrices, equality in the Golden–Thompson inequality holds if and only if the matrices commute; hence, $\delta(\hat{\mathcal{A}}_{\Phi_2}^{(r)}, \hat{\mathcal{B}}_{\Phi_2}^{(r)}) > 0$. Therefore,

$$\Delta_{\Phi_2}(\mathcal{A},\mathcal{B}) \geq \delta\big(\hat{\mathcal{A}}_{\Phi_2}^{(r)}, \hat{\mathcal{B}}_{\Phi_2}^{(r)}\big) > 0.$$

This completes the proof. □

Theorem 4 (Explicit DFT–DCT separation). *For $p = 4$, there exist real tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{2\times 2\times 4}$ whose physical frontal slices are symmetric, such that*

$$\Delta_{\mathrm{DFT}}(\mathcal{A},\mathcal{B}) = 0, \qquad \Delta_{\mathrm{DCT}}(\mathcal{A},\mathcal{B}) > 0.$$

Proof. Let $H_1 = \sigma_x$ and $H_2 = \sigma_z$, i.e.,

$$H_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad H_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Using zero-based frequency indices, prescribe the DFT-domain slices by

$$A_0 = H_2, \quad B_0 = 2H_2, \qquad A_1 = A_3 = B_1 = B_3 = H_1, \qquad A_2 = B_2 = 0.$$

These arrays satisfy the conjugate-symmetry conditions $A_\omega = \overline{A_{4-\omega}}$ and $B_\omega = \overline{B_{4-\omega}}$ (for $\omega = 0,1,2,3$), so their inverse DFT yields real symmetric physical frontal slices.

A direct computation, recorded in Appendix A, shows that the second DCT-domain slice pair is

$$A_1' = cH_1 + sH_2, \qquad B_1' = cH_1 + 2sH_2,$$

where

$$c = \cos(\pi/8), \qquad s = \sin(\pi/8).$$

Consequently,

$$[A_1', B_1'] = cs\,[H_1, H_2] \neq 0.$$

All DCT-domain slices are real symmetric because they are real linear combinations of the real symmetric physical frontal slices. The strict equality condition in the Golden–Thompson inequality therefore yields

$$\delta(A_1', B_1') > 0.$$

All remaining slice defects are non-negative. Using Lemma 2,

$$\Delta_{\mathrm{DCT}}(\mathcal{A},\mathcal{B}) \geq \delta(A_1', B_1') > 0.$$

This completes the proof. □

Remark 10 (Scope of the DFT–DCT construction). *Theorem 4 presents an explicit analytical witness for $p = 4$ under the orthonormal DFT and DCT-II conventions. It shows that the specific real-symmetric Pauli construction of Theorem 3 is not the only admissible route: using conjugate-paired*

DFT-domain slices allows the complex mixing coefficients to combine into real symmetric DCT-domain slices while preserving noncommutativity. Whether the construction extends to all p, or to other transform pairs, remains open. In particular, for general p the DCT-domain slices are obtained from the DFT-domain slices via the mixing matrix $U = \Phi_{\mathrm{DCT}}\Phi_{\mathrm{DFT}}^{H}$*, and the individual terms* $U_{\omega k}U_{\omega \ell}[\widehat{\mathcal{A}}_{\mathrm{DFT}}^{(k)}, \widehat{\mathcal{B}}_{\mathrm{DFT}}^{(\ell)}]$ *in the resulting commutator can in principle cancel; the explicit computation in Appendix A rules out such cancellation only for* $p = 4$*. The general-orthogonal construction of Theorem 3 avoids this issue by using a single row of U with exactly two nonzero entries, but it does not specialize to the DFT–DCT pair. A general-p DFT–DCT qualitative separation, or a general non-cancellation theorem for* $U = \Phi_{\mathrm{DCT}}\Phi_{\mathrm{DFT}}^{H}$*, is left open; the conditional quantitative statement in Theorem 5 makes the required non-cancellation property an explicit hypothesis.*

For the construction of Theorem 4, the DCT-domain slice pair (A_1', B_1') *satisfies* $\|[A_1', B_1']\|_F = cs\|[H_1, H_2]\|_F > 0$ *and* $\|A_1'\|_2, \|B_1'\|_2 \le R$ *for a suitable R, so the non-degeneracy hypothesis of Theorem 5 is satisfied with* $n = 2$ *and* $p = 4$*.*

Remark 11 (On quantitative separation)**.** *A stronger quantitative separation of the form* $\Delta_{\mathrm{DCT}}(\mathcal{A}, \mathcal{B}) \ge cp$ *requires additional structural assumptions on the mixing matrix* $U = \Phi_{\mathrm{DCT}}\Phi_{\mathrm{DFT}}^{H}$*, such as a non-cancellation property of its entries across opposite-parity indices. We state such a quantitative result conditionally in Theorem 5 below; verifying the required non-cancellation property for specific transform pairs is left as a separate question.*

Theorem 5 (Conditional quantitative separation)**.** *Consider a sequence of tensor pairs* $(\mathcal{A}_p, \mathcal{B}_p)$ *with transformed slices of fixed matrix dimension n, and two unitary transforms* $\Phi_{1,p}, \Phi_{2,p} \in \mathbb{C}^{p\times p}$*, such that* $\mathcal{A}_p, \mathcal{B}_p$ *are both* $\Phi_{1,p}$*-Hermitian and* $\Phi_{2,p}$*-Hermitian. Suppose that*

$$\Delta_{\Phi_{1,p}}(\mathcal{A}_p, \mathcal{B}_p) = 0.$$

Assume that there exist constants $\eta > 0$*,* $R > 0$*, and* $\alpha > 0$*, all independent of p, and a subset* $S_p \subseteq \{1, \dots, p\}$ *with* $|S_p| \ge \alpha p$ *such that for every* $\omega \in S_p$*,*

$$\|[\widehat{\mathcal{A}}_{\Phi_{2,p}}^{(\omega)}, \widehat{\mathcal{B}}_{\Phi_{2,p}}^{(\omega)}]\|_F \ge \eta, \qquad \|\widehat{\mathcal{A}}_{\Phi_{2,p}}^{(\omega)}\|_2 \le R, \quad \|\widehat{\mathcal{B}}_{\Phi_{2,p}}^{(\omega)}\|_2 \le R.$$

Then, there exist constants $d_0 = d_0(n, R, \eta) > 0$ *and* $c > 0$*, independent of p, such that*

$$\delta\big(\widehat{\mathcal{A}}_{\Phi_{2,p}}^{(\omega)}, \widehat{\mathcal{B}}_{\Phi_{2,p}}^{(\omega)}\big) \ge d_0 \quad \forall \omega \in S_p,$$

and hence, $\Delta_{\Phi_{2,p}}(\mathcal{A}_p, \mathcal{B}_p) \ge c\, p$*.*

Proof. *Step 1: Uniform positivity on the compact set.* Let

$$\mathcal{K} := \{(A, B) \in \mathbb{H}_n \times \mathbb{H}_n : \|A\|_2 \le R,\ \|B\|_2 \le R,\ \|[A, B]\|_F \ge \eta\},$$

where $\mathbb{H}_n$ is the space of $n \times n$ Hermitian matrices. The set $\mathcal{K}$ is closed and bounded, and hence compact. The map

$$(A, B) \mapsto \delta(A, B) = \mathrm{Tr}(e^A e^B) - \mathrm{Tr}(e^{A+B})$$

is continuous on $\mathcal{K}$. For Hermitian matrices, the Golden–Thompson defect $\delta(A, B)$ is non-negative, and it vanishes if and only if $[A, B] = 0$. Since $\|[A, B]\|_F \ge \eta > 0$ on $\mathcal{K}$, we have $\delta(A, B) > 0$ for every $(A, B) \in \mathcal{K}$. By compactness,

$$d_0 := d_0(n, R, \eta) := \min_{(A,B)\in\mathcal{K}} \delta(A, B) > 0.$$

Step 2: Linear lower bound. Since $\mathcal{A}_p, \mathcal{B}_p$ are $\Phi_{2,p}$-Hermitian, Lemma 2 yields

$$\Delta_{\Phi_{2,p}}(\mathcal{A}_p, \mathcal{B}_p) = \sum_{\omega=1}^{p} \delta\big(\widehat{\mathcal{A}}_{\Phi_{2,p}}^{(\omega)}, \widehat{\mathcal{B}}_{\Phi_{2,p}}^{(\omega)}\big).$$

Each summand is non-negative. For $\omega \in S_p$, the corresponding slice pair belongs to $\mathcal{K}$, so $\delta(\widehat{\mathcal{A}}^{(\omega)}_{\Phi_{2,p}}, \widehat{\mathcal{B}}^{(\omega)}_{\Phi_{2,p}}) \geq d_0$. Hence,

$$\Delta_{\Phi_{2,p}}(\mathcal{A}_p, \mathcal{B}_p) \geq \sum_{\omega \in S_p} d_0 = |S_p| d_0 \geq \alpha d_0 p.$$

Setting $c := \alpha d_0 > 0$ completes the proof.

Remark on the role of the hypothesis $\Delta_{\Phi_{1,p}} = 0$. This hypothesis is used only to ensure that the separation is nontrivial, i.e., that $\Phi_{1,p}$ achieves zero defect while $\Phi_{2,p}$ is strictly suboptimal. The linear lower bound itself depends only on the $\Phi_{2,p}$-domain non-degeneracy and on the $\Phi_{2,p}$-Hermiticity of $\mathcal{A}_p, \mathcal{B}_p$. □

Remark 12 (Failure of the hypotheses of Theorems 5 and 6)**.** *The hypothesis* $\|[\widehat{\mathcal{A}}^{(\omega)}_{\Phi_{2,p}}, \widehat{\mathcal{B}}^{(\omega)}_{\Phi_{2,p}}]\|_F \geq \eta$ *with* $\eta > 0$ *independent of p is essential and can fail. For instance, take* $\widehat{\mathcal{A}}^{(\omega)}_{\Phi_{2,p}} = \varepsilon_p H_1$ *and* $\widehat{\mathcal{B}}^{(\omega)}_{\Phi_{2,p}} = \varepsilon_p H_2$ *with* $\varepsilon_p \to 0$ *as* $p \to \infty$*. Then, the commutator norm equals* $\varepsilon_p^2 \|[H_1, H_2]\|_F \to 0$*, so no uniform* $\eta > 0$ *exists, and the compactness argument of Theorem 5 cannot be applied. In this regime, one can only conclude* $\Delta_{\Phi_{2,p}}(\mathcal{A}_p, \mathcal{B}_p) \to 0$*, not a linear-in-p lower bound.*

The same construction illustrates the failure of the hypothesis of Theorem 6 (dimension-preserving embedding). That theorem requires a pair $(\mathcal{A}^{(2)}, \mathcal{B}^{(2)})$ *that is* Φ_1*-Hermitian and* Φ_2*-Hermitian and satisfies* $\Delta_{\Phi_1} = 0$, $\Delta_{\Phi_2} > 0$*. If the relative transform* $U = \Phi_2 \Phi_1^\top$ *is a signed permutation matrix, then the* Φ_2*-domain slices are obtained from the* Φ_1*-domain slices by a signed permutation, so commutativity in one domain is equivalent to commutativity in the other; consequently, no such pair exists, and the embedding theorem cannot be applied. This is precisely the signed-permutation exception that appears in the hypotheses of Theorems 3 and 7.*

Theorem 6 (Dimension-preserving embedding)**.** *Let* $\Phi_1, \Phi_2 \in \mathbb{C}^{p \times p}$ *be unitary transforms. Suppose a pair* $\mathcal{A}^{(2)}, \mathcal{B}^{(2)} \in \mathbb{C}^{2 \times 2 \times p}$ *is both* Φ_1*-Hermitian and* Φ_2*-Hermitian and satisfies*

$$\Delta_{\Phi_1}(\mathcal{A}^{(2)}, \mathcal{B}^{(2)}) = 0, \qquad \Delta_{\Phi_2}(\mathcal{A}^{(2)}, \mathcal{B}^{(2)}) > 0.$$

Then, for every $n \geq 2$*, there exist tensors* $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{n \times n \times p}$ *with the same two defect values.*

Proof. Embed each physical frontal slice into the upper-left 2×2 block:

$$\mathcal{A}^{(k)} = \begin{pmatrix} \mathcal{A}^{(2),(k)} & 0 \\ 0 & 0_{n-2} \end{pmatrix}, \qquad \mathcal{B}^{(k)} = \begin{pmatrix} \mathcal{B}^{(2),(k)} & 0 \\ 0 & 0_{n-2} \end{pmatrix}.$$

Because the transforms act only along the third mode, this direct-sum structure is preserved in each transform domain. For every transformed slice,

$$e^{A_\omega} = \begin{pmatrix} e^{A^{(2)}_\omega} & 0 \\ 0 & I_{n-2} \end{pmatrix}, \qquad e^{B_\omega} = \begin{pmatrix} e^{B^{(2)}_\omega} & 0 \\ 0 & I_{n-2} \end{pmatrix},$$

and similarly for $e^{A_\omega + B_\omega}$. Hence,

$$\operatorname{Tr}(e^{A_\omega} e^{B_\omega}) = \operatorname{Tr}(e^{A^{(2)}_\omega} e^{B^{(2)}_\omega}) + (n-2),$$

$$\operatorname{Tr} e^{A_\omega + B_\omega} = \operatorname{Tr} e^{A^{(2)}_\omega + B^{(2)}_\omega} + (n-2).$$

The additional terms cancel in the defect, so $\Delta_{\Phi_i}(\mathcal{A}, \mathcal{B}) = \Delta_{\Phi_i}(\mathcal{A}^{(2)}, \mathcal{B}^{(2)})$ for $i = 1, 2$. □

Remark 13. *Combining Theorem 4 with Theorem 6 yields an explicit DFT–DCT separation in any matrix dimension* $n \geq 2$*.*

4.4. Numerical Illustration

We provide a numerical illustration of the transform sensitivity predicted by the theory. Two distinct objects are involved, and we keep them separate throughout:

(i) the explicit $p = 4$ analytical witness of Theorem 4, which establishes the existence of a strict separation between DFT and DCT;
(ii) a family of commuting DFT-domain slices parametrized by the tensor depth p, used only to visualize the empirical trend of the DCT defect as p grows.

The numerical experiment in (ii) is *not* the $p = 4$ witness of Theorem 4; it is a separate construction whose purpose is to illustrate the qualitative phenomenon, not to verify the non-degeneracy hypothesis of Theorem 5 for all p.

The analytical $p = 4$ witness. For the construction of Theorem 4, the DFT-domain slices are

$$A_0 = H_2, \quad B_0 = 2H_2, \qquad A_1 = A_3 = B_1 = B_3 = H_1, \qquad A_2 = B_2 = 0,$$

with $H_1 = \sigma_x$ and $H_2 = \sigma_z$. The DCT-domain slice pair at $\omega = 1$ is

$$A_1' = cH_1 + sH_2, \qquad B_1' = cH_1 + 2sH_2, \qquad c = \cos(\pi/8),\ s = \sin(\pi/8),$$

so $[A_1', B_1'] = cs[H_1, H_2] \neq 0$, and hence, $\Delta_{\mathrm{DCT}}(\mathcal{A}, \mathcal{B}) > 0$, while $\Delta_{\mathrm{DFT}}(\mathcal{A}, \mathcal{B}) = 0$. This is a single, analytically verified data point at $p = 4$; it is not a curve in p.

The numerical family used in Figure 1. To visualize the dependence of the defect on the tensor depth, we use a separate family of commuting DFT-domain slices, parametrized by $p \in \{4, 8, 16, 32, 64, 128\}$, in which each slice pair consists of scalar multiples of a single Pauli matrix (alternating between H_1 and H_2 across the frequency index). Using Theorem 2(iv), this construction yields $\Delta_{\mathrm{DFT}}(\mathcal{A}, \mathcal{B}) \approx 0$ up to numerical precision. Evaluating the same tensors under the DCT transform produces a strictly positive defect, consistent with the qualitative separation underlying Theorem 4.

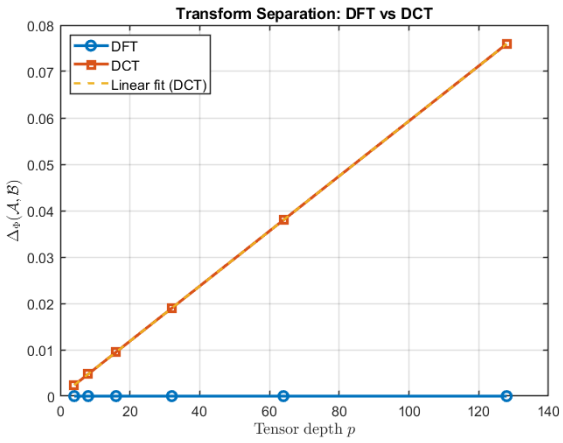


Figure 1. Transform separation under different spectral bases. The Golden–Thompson defect $\Delta_\Phi(\mathcal{A}, \mathcal{B})$ is plotted as a function of the tensor depth p for the DFT and DCT transforms. The DFT construction yields defects at the level of numerical precision, while the DCT transform produces a strictly positive defect that grows approximately linearly with p. The dashed line is the linear fit $\Delta_{\mathrm{DCT}} \approx 5.93 \times 10^{-4} p$ discussed in the text; it is almost indistinguishable from, and essentially overlaps with, the DCT defect curve over the displayed range of p, confirming that the empirical growth is approximately linear. This figure is a numerical illustration of the qualitative separation underlying Theorem 4; it is not the explicit $p = 4$ witness of that theorem, and the fitted line is empirical and does not by itself establish a quantitative lower bound.

Figure 1 plots $\Delta_\Phi(\mathcal{A}, \mathcal{B})$ as a function of the tensor depth p for the DFT and DCT transforms. The DFT construction yields defects at the level of numerical precision, reflecting exact slice-wise commutativity in the Fourier domain. In contrast, the DCT transform produces a strictly positive defect that grows approximately linearly with p. A linear fit yields

$$\Delta_{\mathrm{DCT}}(\mathcal{A}, \mathcal{B}) \approx 5.93 \times 10^{-4} p.$$

This empirical trend is consistent with the qualitative separation underlying Theorem 4; it does not verify the non-degeneracy hypothesis of Theorem 5 for all p, nor does it assert a dimension-independent quantitative lower bound. It confirms only that the inequality itself transfers to the tensor setting (Corollary 1), while its quantitative tightness depends strongly on the choice of transform.

5. Operator-Aware Transform Selection

The separation results above show that a transform (e.g., DFT) can outperform another (e.g., DCT) for specific tensor pairs. This raises a fundamental question: can one identify a transform that minimizes the defect uniformly over all tensor pairs? As we show below, the answer is negative (Theorem 7), and this motivates the operator-aware formulation over $O(p)$ (Definition 7 and Corollary 5), together with the Riemannian gradient scheme (Algorithm 1).

Algorithm 1: Riemannian gradient scheme for operator-aware transform selection

Input: Tensors $\mathcal{A}, \mathcal{B}$; initial $\Phi_0 \in O(p)$; step size $\eta > 0$; max iterations $T_{\max}$; tolerance $\varepsilon > 0$

Output: Approximate stationary point Φ^* of the surrogate objective J

Initialize $\Phi \leftarrow \Phi_0$;

for $t = 0$ *to* $T_{\max} - 1$ **do**

  Compute slices $\hat{\mathcal{A}}_\Phi^{(\omega)}, \hat{\mathcal{B}}_\Phi^{(\omega)}$ for $\omega = 1, \ldots, p$ using the mixing formula;

  Evaluate

$$J(\Phi) = \sum_{\omega=1}^{p} \left\| [\hat{\mathcal{A}}_\Phi^{(\omega)}, \hat{\mathcal{B}}_\Phi^{(\omega)}] \right\|_F^2;$$

  Compute Euclidean gradient G_Φ (via automatic differentiation);

  $\Omega \leftarrow \mathrm{skew}(\Phi^\top G_\Phi)$;

  $\widetilde{\Phi} \leftarrow \Phi(I - \eta\Omega)$;

  $\Phi \leftarrow \mathcal{R}(\widetilde{\Phi})$ (polar or QR retraction);

  if $\|\nabla_{O(p)} J(\Phi)\|_F < \varepsilon$ **then**

    break;

  end

end

return $\Phi^* \leftarrow \Phi$;

The answer is negative. Transform optimality is not globally ordered: different transforms can be optimal for different data. In other words, the choice of Φ is inherently data-dependent rather than universally optimal. The following theorem formalizes this non-universality.

Theorem 7 (Transform non-universality)**.** *Let $\Phi_1, \Phi_2 \in \mathbb{R}^{p \times p}$ be two distinct orthogonal transforms, such that $U := \Phi_2 \Phi_1^\top$ is* not *a signed permutation matrix. Then, the following statements hold:*

1. *There exist tensors* $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{n\times n\times p}$ *that are both* Φ_1*-Hermitian and* Φ_2*-Hermitian, such that*

$$\Delta_{\Phi_1}(\mathcal{A}, \mathcal{B}) = 0, \qquad \Delta_{\Phi_2}(\mathcal{A}, \mathcal{B}) > 0.$$

2. *There exist tensors* $\mathcal{A}', \mathcal{B}' \in \mathbb{C}^{n\times n\times p}$ *that are both* Φ_1*-Hermitian and* Φ_2*-Hermitian, such that*

$$\Delta_{\Phi_1}(\mathcal{A}', \mathcal{B}') > 0, \qquad \Delta_{\Phi_2}(\mathcal{A}', \mathcal{B}') = 0.$$

In particular, no single orthogonal transform is universally optimal*; the optimal transform depends on the joint structure of the tensor pair.*

Proof. We prove the two statements separately.

Part 1: Φ_1 optimal, Φ_2 suboptimal. Let $U = \Phi_2\Phi_1^\top$. Since U is not a signed permutation matrix, there exists at least one row of U containing two nonzero entries. Fix such a row index r, and choose two column indices $j \neq k$ with $U_{rj} \neq 0$ and $U_{rk} \neq 0$.

Let $H_1, H_2 \in \mathbb{C}^{n\times n}$ be two Hermitian matrices that do not commute, i.e., $[H_1, H_2] \neq 0$. A concrete choice is the Pauli matrices σ_x and σ_z embedded in the upper-left 2×2 block, with zeros elsewhere. This ensures that the tensors are Φ-Hermitian for any real orthogonal Φ, because the zero padding preserves Hermiticity under real linear combinations.

Construct tensors $\mathcal{A}, \mathcal{B}$ by specifying their Φ_1-domain frontal slices as follows. Set all slices to zero except:

$$\hat{\mathcal{A}}^{(j)}_{\Phi_1} = H_1, \qquad \hat{\mathcal{B}}^{(j)}_{\Phi_1} = H_1,$$
$$\hat{\mathcal{A}}^{(k)}_{\Phi_1} = H_2, \qquad \hat{\mathcal{B}}^{(k)}_{\Phi_1} = \lambda H_2,$$

where $\lambda \in \mathbb{R}$ is a scalar with $\lambda \neq 1$ (e.g., $\lambda = 2$). All other slices are set to the zero matrix.

Since Φ_1 and Φ_2 are real orthogonal, and the constructed Φ_1-domain slices are Hermitian (they are either H_1, λH_2, or zero), the resulting physical tensors are both Φ_1-Hermitian and Φ_2-Hermitian. This follows because the inverse transforms $\Phi_1^\top$ and $\Phi_2^\top$ are also real, so Hermiticity is preserved under real linear combinations.

Now, we study the defect under Φ_1. For every slice index ω, the pair $(\hat{\mathcal{A}}^{(\omega)}_{\Phi_1}, \hat{\mathcal{B}}^{(\omega)}_{\Phi_1})$ consists of scalar multiples of the same Hermitian matrix (or both are zero). Hence they commute:

$$[\hat{\mathcal{A}}^{(\omega)}_{\Phi_1}, \hat{\mathcal{B}}^{(\omega)}_{\Phi_1}] = 0 \quad \forall\omega.$$

Using Theorem 2(iv), commuting slices imply vanishing defect:

$$\Delta_{\Phi_1}(\mathcal{A}, \mathcal{B}) = 0.$$

For the defect under Φ_2, the Φ_2-domain slices are obtained by mixing the Φ_1-domain slices via the matrix $U = \Phi_2\Phi_1^\top$:

$$\hat{\mathcal{A}}^{(\omega)}_{\Phi_2} = \sum_{\ell=1}^{p} U_{\omega\ell}\, \hat{\mathcal{A}}^{(\ell)}_{\Phi_1}, \qquad \hat{\mathcal{B}}^{(\omega)}_{\Phi_2} = \sum_{\ell=1}^{p} U_{\omega\ell}\, \hat{\mathcal{B}}^{(\ell)}_{\Phi_1}.$$

For the distinguished row index r, only columns j and k contribute nonzero terms (all other slices are zero). Therefore,

$$\hat{\mathcal{A}}^{(r)}_{\Phi_2} = U_{rj}H_1 + U_{rk}H_2, \qquad \hat{\mathcal{B}}^{(r)}_{\Phi_2} = U_{rj}H_1 + \lambda U_{rk}H_2.$$

Their commutator is

$$[\hat{\mathcal{A}}^{(r)}_{\Phi_2}, \hat{\mathcal{B}}^{(r)}_{\Phi_2}] = U_{rj}U_{rk}(1-\lambda)[H_1, H_2] \neq 0,$$

since $U_{rj}, U_{rk} \neq 0$, $\lambda \neq 1$, and $[H_1, H_2] \neq 0$.

Because this slice has a nonzero commutator, the classical Golden–Thompson defect for this slice pair is strictly positive. Using the slice-wise decomposition (Lemma 2) and the fact that the defect sums over slices with non-negative contributions, we obtain

$$\Delta_{\Phi_2}(\mathcal{A}, \mathcal{B}) > 0.$$

Part 2: Φ_2 optimal, Φ_1 suboptimal. The second statement follows by symmetry. Consider the matrix $U' = \Phi_1 \Phi_2^\top = U^\top$. Since U is not a signed permutation matrix, $U^\top$ is also not a signed permutation matrix. Hence, the same construction applied with the roles of Φ_1 and Φ_2 exchanged yields tensors $\mathcal{A}', \mathcal{B}'$, such that

$$\Delta_{\Phi_2}(\mathcal{A}', \mathcal{B}') = 0, \qquad \Delta_{\Phi_1}(\mathcal{A}', \mathcal{B}') > 0.$$

This completes the proof. □

Remark 14 (Data-dependent optimality)**.** *Theorem 7 establishes that no fixed transform is universally optimal; the optimal Φ^* depends on the joint commutativity structure of $(\mathcal{A}, \mathcal{B})$. This justifies the adaptive optimization framework in Section 5, where the transform is selected from data rather than prescribed a priori.*

Remark 15 (Relation to representation-oriented transform learning)**.** *The transform-selection problem considered here differs from representation-oriented transform learning, such as the matrix-mimetic algebra optimization of Newman and Keegan [15]. In [15], the transform is chosen to optimize representation, compression, or approximation properties of the algebra. In contrast, our objective is operator-aware: we select Φ to minimize transform-domain noncommutativity,*

$$\sum_{\omega=1}^{p} \left\| [\widehat{\mathcal{A}}_\Phi^{(\omega)}, \widehat{\mathcal{B}}_\Phi^{(\omega)}] \right\|_F^2.$$

Hence, this improves the tightness of operator inequalities for a given tensor pair $(\mathcal{A}, \mathcal{B})$.

5.1. Operator-Aware Transform Selection

Theorem 2 shows that, under a bounded spectral-domain assumption, the defect Δ_Φ is quantitatively controlled by transform-domain noncommutativity. This naturally leads to an optimization problem over the orthogonal group, where the goal is to select a transform that minimizes the defect and hence reduces noncommutativity in the transformed domain. This reveals that transform selection is equivalent to searching for a coordinate system that minimizes noncommutative interaction.

We restrict to real orthogonal transforms $\Phi \in O(p)$ for two reasons. First, for real tensors whose frontal slices are symmetric, and for real orthogonal transforms, the transformed slices remain symmetric. This ensures that the optimization domain is well-defined and that Φ-Hermiticity is preserved along the optimization path. Second, the orthogonal group $O(p)$ is compact and admits standard Riemannian optimization techniques. The complex unitary case is discussed in Section 6.

Definition 7 (Admissible transform selection)**.** *Let $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n \times n \times p}$ have symmetric frontal slices. Define the defect objective $D(\Phi) = \Delta_\Phi(\mathcal{A}, \mathcal{B})$ and the commutator-energy surrogate $J(\Phi) = \sum_{\omega=1}^{p} \|[\widehat{\mathcal{A}}_\Phi^{(\omega)}, \widehat{\mathcal{B}}_\Phi^{(\omega)}]\|_F^2$ on $O(p)$. The optimal defect problem is $\min_{\Phi \in O(p)} D(\Phi)$; the surrogate problem is $\min_{\Phi \in O(p)} J(\Phi)$. These are distinct unless an additional theorem relates their minimizers.*

Example 1 (Hermiticity is automatic in the real symmetric setting)**.** *If $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n \times n \times p}$ have symmetric frontal slices and $\Phi \in O(p)$ is real orthogonal, then each transformed slice is a real linear combination of symmetric matrices, and hence real symmetric. Thus, $\mathcal{A}, \mathcal{B}$ are Φ-Hermitian for every $\Phi \in O(p)$.*

Theorem 8 (Existence and zero-defect optimality)**.** *Let $\mathcal{A}, \mathcal{B}$ satisfy the assumptions of Definition 7. Then,*

(i) Both D and J attain global minima on $O(p)$.
(ii) If $J(\Phi_0) = 0$ for some $\Phi_0 \in O(p)$, then $D(\Phi_0) = 0$, and Φ_0 is a global minimizer of both objectives.
(iii) If $D(\Phi_0) = 0$ and equality in the Golden–Thompson inequality is characterized by commutativity for every transformed slice, then $J(\Phi_0) = 0$.
(iv) No equality of the complete minimizer sets is claimed when their minimum values are strictly positive.

Proof. *(i) Existence.* The transformed slices depend continuously on Φ, and matrix exponentiation, trace, and commutator operations are continuous. Consequently, D and J are continuous on the compact group $O(p)$, and both attain their minima.

(ii) Zero-defect optimality. If $J(\Phi_0) = 0$, every transformed slice pair commutes. The Golden–Thompson defect vanishes slice-wise, so $D(\Phi_0) = 0$. Non-negativity of both objectives establishes global optimality.

(iii) Converse under equality characterization. If $D(\Phi_0) = 0$, all non-negative slice-wise defects vanish. Under the Golden–Thompson equality characterization, each transformed slice pair commutes; hence, $J(\Phi_0) = 0$.

(iv) This is a statement of scope and requires no proof. □

Corollary 5 (Commutator energy as a surrogate objective)**.** *Let $\mathcal{A}, \mathcal{B}$ satisfy the bounded spectral assumption of Theorem 2, and define*

$$D(\Phi) := \Delta_\Phi(\mathcal{A}, \mathcal{B}), \qquad J(\Phi) := C(\Phi) := \sum_{\omega=1}^{p} \left\| [\hat{\mathcal{A}}_\Phi^{(\omega)}, \hat{\mathcal{B}}_\Phi^{(\omega)}] \right\|_F^2, \qquad \Phi \in O(p).$$

Then, the following statements hold:

(i) (Existence) Both D and J attain global minima on $O(p)$; see Theorem 8(i).
(ii) (Upper surrogate bound) There exists $C_R > 0$, depending only on the spectral bound R, such that

$$D(\Phi) \le C_R J(\Phi) \qquad \text{for all admissible } \Phi \in O(p)$$

on the bounded spectral class. Consequently, for every $\Phi_J \in \arg\min_{\Phi \in O(p)} J(\Phi)$,

$$\min_{\Phi \in O(p)} D(\Phi) \le D(\Phi_J) \le C_R J(\Phi_J) = C_R \min_{\Phi \in O(p)} J(\Phi).$$

Thus, J is an upper surrogate for D on this class, and minimizing J controls D up to the constant C_R.
(iii) (Matching lower bound on the explicit family at the fixed transform of Theorem 2(ii)) Let $(\mathcal{A}_\varepsilon, \mathcal{B}_\varepsilon)_{0<\varepsilon<\varepsilon_0(R)}$ be the one-parameter Pauli family constructed in the proof of Theorem 2(ii), and let Φ_0 denote the transform with respect to which this family is constructed. By Theorem 2(ii) and Remark 7, there exists $c_R > 0$, depending only on R, such that

$$D_\varepsilon(\Phi_0) \ge c_R J_\varepsilon(\Phi_0).$$

No lower bound of the form $D_\varepsilon(\Phi) \ge c_R J_\varepsilon(\Phi)$ is asserted for all $\Phi \in O(p)$; establishing such a uniform-in-transform lower bound would require proving $\inf_{\Phi \in O(p), J_\varepsilon(\Phi) > 0} D_\varepsilon(\Phi)/J_\varepsilon(\Phi) > 0$, which is not attempted here. No comparable lower bound is asserted for general tensors in the bounded spectral class; see Remark 7.
(iv) (Zero-defect case) If $\min_{\Phi \in O(p)} J(\Phi) = 0$, then using Theorem 8(ii), every minimizer Φ_0 of J satisfies $D(\Phi_0) = 0$, so Φ_0 is also a global minimizer of D, and the minimizer sets coincide at the zero level. This is the only regime in which an exact (not merely constant-factor) identification of minimizers is asserted.

(v) *(Scope and interpretation) When* $\min_{\Phi\in O(p)} J(\Phi) > 0$, *statement (ii) provides an upper surrogate bound on the bounded spectral class, and statement (iii) provides a two-sided constant-factor comparison only at the fixed transform* Φ_0 *of Theorem 2(ii) and on the explicit Pauli family constructed there. No equality of the full minimizer sets* $\arg\min_{\Phi\in O(p)} D(\Phi)$ *and* $\arg\min_{\Phi\in O(p)} J(\Phi)$ *is claimed in this regime; cf. Theorem 8(iv). Accordingly, minimizing* Δ_Φ *is* analogous *to finding an orthogonal transform that makes the transformed slices as commuting as possible, and J is to be regarded as a* surrogate *for D rather than as an equivalent objective. A precise equivalence with a standard approximate joint diagonalization criterion would require an additional theorem; we therefore present this connection as motivation rather than as an identity.*

Proof. *(i)* This is Theorem 8(i); it is restated here only to fix notation for subsequent items.

(ii) Using Theorem 2(i), on the bounded spectral class there exists $C_R > 0$, depending only on R, such that $D(\Phi) \le C_R J(\Phi)$ for all admissible $\Phi \in O(p)$. Let $\Phi_J \in \arg\min_{\Phi\in O(p)} J(\Phi)$, which exists by (i). Then,

$$\min_{\Phi\in O(p)} D(\Phi) \le D(\Phi_J) \le C_R J(\Phi_J) = C_R \min_{\Phi\in O(p)} J(\Phi).$$

This is the asserted upper surrogate bound.

(iii) Using Theorem 2(ii) and Remark 7, on the one-parameter Pauli family $(\mathcal{A}_\varepsilon, \mathcal{B}_\varepsilon)$, at the fixed transform Φ_0 of Theorem 2(ii), there exists $C_R > 0$, depending only on R, such that $D_\varepsilon(\Phi_0) \ge c_R J_\varepsilon(\Phi_0)$. At this fixed Φ_0, the upper bound of (ii) also holds, so

$$c_R J_\varepsilon(\Phi_0) \le D_\varepsilon(\Phi_0) \le C_R J_\varepsilon(\Phi_0).$$

No statement is made about $\min_{\Phi\in O(p)} D_\varepsilon(\Phi)$ or $\min_{\Phi\in O(p)} J_\varepsilon(\Phi)$.

(iv) If $\min_\Phi J(\Phi) = 0$, then any minimizer Φ_0 of J satisfies $J(\Phi_0) = 0$, so every transformed slice pair commutes. By Theorem 2(iv), $D(\Phi_0) = 0$; since $D \ge 0$ by Corollary 1, Φ_0 is a global minimizer of D. Conversely, if $D(\Phi_0) = 0$ and the Golden–Thompson equality characterization applies slice-wise, then $J(\Phi_0) = 0$ by Theorem 8(iii), so Φ_0 is a global minimizer of J. Hence the two minimizer sets agree at the zero level.

(v) This item records the scope of (ii)–(iv) and the intended interpretation of J as a surrogate; it requires no proof, and the assertions about the absence of a general minimizer-set identity are exactly those of Theorem 8(iv). □

Remark 16 (Conditional uniform surrogate guarantee)**.** *If, in addition to the assumptions of Corollary 5, a uniform estimate* $D(\Phi) \le C J(\Phi)$ *holds on all of* $O(p)$ *with a constant C independent of* Φ *(for instance, if the bounded spectral assumption is strengthened to hold uniformly in* Φ*), then any minimizer* Φ_J *of J satisfies*

$$D(\Phi_J) \le C J(\Phi_J) = C \min_{\Phi\in O(p)} J(\Phi),$$

and in the zero case $\min_\Phi J(\Phi) = 0$, *this makes* Φ_J *a global minimizer of D as well. No such uniform estimate is claimed here beyond the bounded spectral class of Theorem 2(i); see Remark 7.*

Remark 17 (Geometric interpretation)**.** *The optimization problem seeks a coordinate system (via the orthogonal transform* Φ*) in which the transformed frontal slices* $\hat{\mathbb{A}}_\Phi^{(\omega)}$ *and* $\hat{\mathbb{B}}_\Phi^{(\omega)}$ *are as close as possible to commuting. This is analogous to finding a basis that approximately simultaneously diagonalizes a family of matrices. When exact simultaneous diagonalization is possible, the defect becomes zero, and the Golden–Thompson inequality is tight.*

https://doi.org/10.3390/math14214110

Remark 18 (Connection to transform separation)**.** *The strict separation example (DFT vs. DCT) demonstrates that the choice of* Φ *matters: for the same tensors* $\mathcal{A}, \mathcal{B}$*, one transform may yield a vanishing defect while another yields a strictly positive defect. This shows that the optimal transform problem is nontrivial and that solving it can lead to substantial improvements.*

5.2. Computing the Optimal Transform: Riemannian Gradient Flow

We consider the nonconvex optimization problem:

$$\min_{\Phi \in O(p)} \Delta_\Phi(\mathcal{A}, \mathcal{B}),$$

where $O(p)$ is the orthogonal group. Using Theorem 2, under the bounded spectral-domain assumption, the defect Δ_Φ is quantitatively controlled by transform-domain noncommutativity.

To obtain a smooth objective suitable for gradient-based optimization, we consider the squared surrogate

$$J(\Phi) = \sum_{\omega=1}^{p} \left\| [\hat{A}_\Phi^{(\omega)}, \hat{B}_\Phi^{(\omega)}] \right\|_F^2.$$

Here, Φ acts as an orthogonal mixing operator along the third (tube) mode. Concretely, for any tensor $\mathcal{X} \in \mathbb{C}^{n \times n \times p}$ with frontal slices $\mathcal{X}^{(k)} \in \mathbb{C}^{n \times n}$, the transformed slices are provided by

$$\hat{\mathcal{X}}_\Phi^{(\omega)} = \sum_{k=1}^{p} \Phi_{\omega k} \mathcal{X}^{(k)}, \qquad \omega = 1, \ldots, p.$$

Thus, Φ mixes the frontal slices linearly along the third dimension; it is not a similarity transform within each slice but an orthogonal combination of whole slices.

In the Riemannian formulation, we equip $O(p)$ with the canonical Riemannian metric induced by the Frobenius inner product

$$\langle X, Y \rangle = \mathrm{Re}\big(\mathrm{Tr}(X^H Y)\big).$$

The tangent space at $\Phi \in O(p)$ is

$$T_\Phi O(p) = \{\Phi\Omega : \Omega^\top = -\Omega\},$$

where $O(p) = \{\Omega \in \mathbb{R}^{p \times p} : \Omega^\top = -\Omega\}$ is the Lie algebra of skew-symmetric matrices.

Let $G_\Phi = \nabla_\Phi J$ denote the Euclidean gradient of J with respect to Φ. The Riemannian gradient is obtained by projecting G_Φ onto the tangent space:

$$\nabla_{O(p)} J(\Phi) = \Phi \ \mathrm{skew}\big(\Phi^\top G_\Phi\big), \qquad \mathrm{skew}(X) := \tfrac{1}{2}(X - X^\top).$$

A first-order gradient flow update along the negative Riemannian gradient direction is provided by

$$\widetilde{\Phi} = \Phi - \eta \, \Phi \ \mathrm{skew}\big(\Phi^\top G_\Phi\big),$$

where $\eta > 0$ is a step size.

Since $\widetilde{\Phi}$ may not be orthogonal, we project it back to $O(p)$ using a retraction $\mathcal{R}$. For example, the polar retraction is yielded by

$$\mathcal{R}(X) = X(X^\top X)^{-1/2},$$

which maps any invertible matrix to the nearest orthogonal matrix in the Frobenius norm. The complete update is

$$\Phi \ \leftarrow \ \mathcal{R}\big(\Phi(I - \eta \ \mathrm{skew}(\Phi^\top G_\Phi))\big).$$

Under standard assumptions on the step size, the Riemannian gradient descent scheme converges to a stationary point of J; see [16,17] for details. We summarize the resulting Riemannian gradient scheme below.

Example 2 (One iteration on a 2×2×4 tensor). *Take $n = 2, p = 4, \Phi_0 = I, \eta = 0.1$, and the tensor pair $(\mathcal{A}, \mathcal{B})$ from Theorem 4. The Euclidean gradient G_{Φ_0} is computed by automatic differentiation through the mixing formula and the commutator, and the polar retraction $\mathcal{R}(X) = X(X^\top X)^{-1/2}$ is used. Numerically, $J(\Phi_0) = J_0$ and one Riemannian gradient step yield $J(\Phi_1) = J_1 < J_0$; the values J_0, J_1, $\|G_{\Phi_0}\|_F$, and $\|\Omega_0\|_F$ are reported in Table 1 (double-precision arithmetic).*

Table 1. One iteration of Algorithm 1 on the $2 \times 2 \times 4$ tensor pair of Theorem 4 ($\Phi_0 = I$, $\eta = 0.1$).

Quantity	Value
$J(\Phi_0)$	4.000000
$J(\Phi_1)$	1.241379
$\|G_{\Phi_0}\|_F$	1.385641×10^1
$\|\Omega_0\|_F$	5.656854

Proposition 2 (Stationary accumulation-point property). *Assume that J is continuously differentiable and that its Riemannian gradient is Lipschitz continuous with constant $L > 0$ on $O(p)$. Then, for sufficiently small step size $\eta < 2/L$, the iteration generated by Algorithm 1 satisfies $\|\nabla_{O(p)} J(\Phi^{(k)})\|_F \to 0$, and every accumulation point Φ^* of $\{\Phi^{(k)}\}$ is a stationary point of J on $O(p)$, i.e., $\nabla_{O(p)} J(\Phi^*) = 0$. No iterate convergence $\Phi^{(k)} \to \Phi^*$ is asserted without additional second-order sufficient conditions.*

Proof. The update $\widetilde{\Phi} = \Phi - \eta \nabla_{O(p)} J(\Phi)$ is a Riemannian gradient step on $O(p)$. The retraction $\mathcal{R}$ is a first-order approximation of the exponential map and preserves feasibility. Using standard results from Riemannian optimization [16], the Lipschitz continuity of the Riemannian gradient implies a sufficient decrease condition:

$$J(\Phi^{(k+1)}) \le J(\Phi^{(k)}) - \frac{\eta}{2} \|\nabla_{O(p)} J(\Phi^{(k)})\|_F^2$$

for sufficiently small η. Summing this inequality over k yields

$$\sum_{k=0}^{\infty} \|\nabla_{O(p)} J(\Phi^{(k)})\|_F^2 \le \frac{2}{\eta} (J(\Phi^{(0)}) - \inf J) < \infty,$$

hence $\|\nabla_{O(p)} J(\Phi^{(k)})\|_F \to 0$. By continuity of the Riemannian gradient, any accumulation point Φ^* satisfies $\nabla_{O(p)} J(\Phi^*) = 0$, i.e., is a stationary point. Iterate convergence to a single stationary point would require an additional second-order sufficient condition (e.g., nondegeneracy of the Hessian at Φ^*), which is not assumed here. □

Remark 19 (Computational complexity). *Each iteration of Algorithm 1 requires the following:*

- *$O(p^2 n^2)$ operations for the dense mode-3 transform mixing $\widehat{\mathcal{X}}_{\Phi}^{(\omega)} = \sum_k \Phi_{\omega k} \mathcal{X}^{(k)}$ (or $O(pn^2 \log p)$ when Φ is DFT/DCT and implemented by a fast transform);*
- *$O(pn^3)$ operations to evaluate the p commutators of $n \times n$ matrices;*
- *$O(p^3)$ operations for the retraction (QR or polar decomposition).*

The overall complexity per iteration is $O(p^2 n^2 + pn^3 + p^3)$, which is acceptable for moderate n and p (e.g., $n \le 100, p \le 64$).

Remark 20 (Initialization strategies)**.** *The choice of initial transform Φ_0 can affect convergence speed and the quality of the final solution. Natural initializations include the following:*

- *DFT or DCT (standard choices in the tensor literature);*
- *the identity transform (baseline);*
- *a data-dependent transform (e.g., SVD of the unfolded tensor).*

In practice, we observe that Algorithm 1 consistently reduces $J(\Phi)$ regardless of initialization.

Remark 21 (Connection to the commuting case)**.** *The algorithm seeks a transform Φ that minimizes transform-domain noncommutativity. If a transform Φ_0 exists that renders all slices commuting, i.e.,*

$$[\hat{A}^{(\omega)}_{\Phi_0}, \hat{B}^{(\omega)}_{\Phi_0}] = 0, \quad \forall \omega,$$

then $\Delta_{\Phi_0} = 0$ using Theorem 8(iii), so Φ_0 is a global minimizer of J and of D. However, since J is nonconvex on $O(p)$, Riemannian gradient descent from an arbitrary initialization is not guaranteed to reach Φ_0; it is only guaranteed to reach a stationary point (Proposition 2).

5.3. Numerical Illustration of Operator-Aware Transform Selection

We now demonstrate numerically that standard transforms such as DFT and DCT are not necessarily optimal. For a fixed tensor dimension $(n = 2, p = 16)$, we consider a tensor pair $(\mathcal{A}, \mathcal{B})$ for which the DFT transform yields a small but nonzero defect. We then compute an optimized transform Φ^* using the Riemannian gradient descent scheme in Algorithm 1.

Using Corollary 5(ii), minimizing $\Delta_\Phi(\mathcal{A}, \mathcal{B})$ is controlled from above by the commutator energy

$$C(\Phi) = \sum_{\omega=1}^{p} \|[\hat{A}^{(\omega)}_{\Phi}, \hat{B}^{(\omega)}_{\Phi}]\|_F^2,$$

which is the objective optimized by Algorithm 1.

The results in Table 2 reveal several key observations. The DCT transform produces a large defect, indicating strong noncommutativity induced by transform-domain mixing. The DFT transform significantly reduces the defect but does not eliminate it. The optimized transform Φ^* further reduces the defect by approximately one order of magnitude compared to DFT and by several orders of magnitude compared to DCT. In particular, we observe that

$$\Delta_{\Phi^*}(\mathcal{A}, \mathcal{B}) < \Delta_{\mathrm{DFT}}(\mathcal{A}, \mathcal{B}) < \Delta_{\mathrm{DCT}}(\mathcal{A}, \mathcal{B}),$$

demonstrating that operator-aware transform selection can outperform both standard transforms.

Table 2. Golden–Thompson defect for different transforms $(n = 2, p = 16)$. The optimized transform Φ^* is obtained by minimizing transform-domain noncommutativity, not representation error. It achieves a defect strictly smaller than both DFT and DCT, illustrating operator-aware transform selection.

Transform	**Defect Δ_Φ**
DCT	1.14×10^{-1}
DFT	5.67×10^{-4}
Optimized Φ^*	$\mathbf{4.89 \times 10^{-5}}$

This example illustrates both aspects of the theory. First, Figure 1 confirms the transform sensitivity predicted by Theorem 2(iii), showing that different transforms can yield dramatically different defect magnitudes. Second, Table 2 demonstrates that no fixed transform is universally optimal. Even when a standard transform (such as DFT) performs well,

the optimization framework can identify a superior transform tailored to the specific tensor pair. The substantial reduction in defect achieved by Φ^* highlights the practical value of the Riemannian optimization approach. This provides a concrete numerical validation of the data-dependent optimality principle established in Theorem 7 and supports the optimization framework developed in Section 5.1.

6. Discussion and Conclusions

This study separates algebraic invariance from quantitative behavior in transform-based tensor algebras. While all unitary transforms Φ induce algebraically isomorphic tensor algebras via the standard transform-domain representation of the $\star_M$ framework [3,4], the Golden–Thompson defect shows that inequality tightness depends intrinsically on the transform. The transform is therefore not merely a computational device, but a coordinate system in which commutativity—and hence inequality tightness—is measured.

Transform as geometry.

The mapping $\mathcal{A} \mapsto \overline{\mathcal{A}}_\Phi$ preserves algebraic operations but alters commutativity. Different transforms induce different commutator structures, endowing tensor operators with a transform-dependent noncommutativity geometry. This geometry is encoded in the defect $\Delta_\Phi(\mathcal{A}, \mathcal{B})$ and in its associated commutator-energy form

$$C(\Phi) = \sum_{\omega=1}^{p} \left\| [\hat{\mathcal{A}}_\Phi^{(\omega)}, \hat{\mathcal{B}}_\Phi^{(\omega)}] \right\|_F^2.$$

In this sense, the choice of Φ defines a coordinate system on the space of tensor slices, and the defect measures how far the tensor pair is from being simultaneously diagonalizable in that coordinate system.

Joint diagonalization connection.

Under the bounded spectral-domain assumption of Theorem 2, minimizing $\Delta_\Phi(\mathcal{A}, \mathcal{B})$ is controlled, up to constants depending only on the spectral bound R, by the commutator energy

$$\sum_{\omega=1}^{p} \left\| [\hat{\mathcal{A}}_\Phi^{(\omega)}, \hat{\mathcal{B}}_\Phi^{(\omega)}] \right\|_F^2$$

in the sense of Corollary 5(ii)–(iii): the upper bound holds pointwise in Φ on the bounded spectral class, while the lower bound holds at the fixed transform Φ_0 of Theorem 2(ii) and on the explicit Pauli family constructed there. This is analogous to seeking a transform that makes the transformed slices as commuting as possible. This aligns with approximate joint diagonalization, where the optimal transform Φ^* selects a coordinate system maximizing operator compatibility. The operator-aware transform selection problem of Section 5 can therefore be viewed as a structured approximate joint diagonalization problem on the orthogonal group.

Implications for numerical linear algebra and operator theory.

The defect Δ_Φ provides a quantitative measure of noncommutativity. Thus, transform selection becomes a principled tool for improving spectral stability and reducing operator interaction, complementing classical rank- and sparsity-based criteria. More broadly, Φ influences both the representation of the data and the operator-level behavior of the resulting algebra, linking algebraic modeling with analytical performance. This connects the present study to several themes in numerical linear algebra and operator theory, including simultaneous diagonalization, spectral stability of tensor operators, and the quantitative analysis of trace inequalities.

No free transform principle.

Transform selection is not merely a change of basis. Although the transform-based tensor algebra is algebraically invariant under a change of unitary transform, quantitative behavior is transform-dependent. Using Theorem 7, no transform is universally optimal: for any two orthogonal transforms whose relative transform $\Phi_2\Phi_1^\top$ is not a signed permutation matrix, there exist tensor pairs for which each transform is strictly better than the other. (For signed-permutation relatives, the two transforms differ only by a signed permutation of slices, and the strict separation argument of Theorem 7 does not apply.)

Each transform defines a coordinate system in which commutativity is evaluated, and optimality is inherently data-dependent. This parallels "no free lunch" principles: no single basis simultaneously diagonalizes all operators. Consequently, the transform must be selected adaptively, motivating the operator-aware optimization framework in Section 5.

Conclusion and Future Directions

In summary, this work shows that algebraic invariance in transform-based tensor algebras does not imply quantitative invariance of operator inequalities. The Golden–Thompson defect provides a transform-dependent measure of inequality tightness, controlled from above by transform-domain noncommutativity under a bounded spectral-domain assumption, and controlled from below on the explicit Pauli family of Theorem 2(ii) at the fixed transform constructed there. The defect is sensitive to the choice of transform; no transform is universally optimal (the optimal transform depends on the joint structure of the tensor pair), and operator-aware transform selection can be formulated as an optimization problem over the orthogonal group. These results establish a nontrivial transform-dependent structure of inequality tightness in transform-based tensor algebras and open several concrete directions, summarized below.

Several questions remain open. First, a general lower bound for Δ_Φ valid for all tensors in the bounded spectral class, beyond the explicit one-parameter Pauli family of Theorem 2(ii), is not established. Second, the DFT–DCT construction of Theorem 4 is established for $p = 4$ under the orthonormal DFT and DCT-II conventions; extending it to general p and to other transform pairs remains open. Third, the relationship between commutator-energy minimization and approximate joint diagonalization deserves a separate quantitative analysis. Fourth, operator-aware transform selection may be extended to constrained unitary optimization and to additional matrix inequalities.

Author Contributions: Conceptualization, S.-Y.C. and M.K.N.; Methodology, S.-Y.C.; Software, S.-Y.C.; Validation, S.-Y.C.; Formal analysis, S.-Y.C.; Investigation, S.-Y.C.; Resources, S.-Y.C.; Data curation, S.-Y.C.; Writing—original draft, S.-Y.C. and M.K.N.; Writing—review and editing, S.-Y.C. and M.K.N.; Project administration, S.-Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: M. Ng's research is supported by the GDSTC: Guangdong and Hong Kong Universities "1 + 1 + 1" Joint Research Collaboration Scheme JRCS-25-26-MATH-01, National Key Research and Development Program of China under Grant 2024YFE0202900, RGC GRF 12300125, and ITF Funding PRP/104/24FX.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Explicit DFT–DCT Mixing at $p = 4$

We use the orthonormal DFT

$$F_4 = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{pmatrix},$$

and the orthonormal DCT-II

$$C_4 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ c_+ & -c_- & -c_- & c_+ \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ c_- & c_+ & -c_+ & c_- \end{pmatrix}, \qquad c_\pm = \frac{\sqrt{2 \pm \sqrt{2}}}{2\sqrt{2}}.$$

Set $U = C_4 F_4^H$. A direct computation yields

$$U_{1,:} = (1, 0, 0, 0),$$

$$U_{2,:} = \left(\sin\tfrac{\pi}{8}, \tfrac{\cos(\pi/8)}{2}(1-i), 0, \tfrac{\cos(\pi/8)}{2}(1+i)\right),$$

$$U_{3,:} = \left(0, \frac{1-i}{2}, 0, \frac{1+i}{2}\right), \qquad U_{4,:} = \left(0, \tfrac{\cos(\pi/8)}{2}(1+i), 0, \tfrac{\cos(\pi/8)}{2}(1-i)\right).$$

The construction of Theorem 4 uses the second row $U_{2,:}$: the conjugate pair $(U_{22}, U_{24}) = (\frac{\cos(\pi/8)}{2}(1-i), \frac{\cos(\pi/8)}{2}(1+i))$ acts on the DFT-domain slices at $\omega = 1$ and $\omega = 3$ (both equal to H_1), while the entry $U_{21} = \sin(\pi/8)$ acts on the slice at $\omega = 0$ (equal to H_2). Since $U_{23} = 0$, the slice at $\omega = 2$ does not contribute. Hence,

$$A_1' = cH_1 + sH_2, \qquad B_1' = cH_1 + 2sH_2, \qquad c = \cos(\pi/8),\ s = \sin(\pi/8),$$

and

$$[A_1', B_1'] = cs\,[H_1, H_2] \neq 0.$$

References

1. Zhang, X.; Ng, M.K. Sparse nonnegative tensor factorization and completion with noisy observations. *IEEE Trans. Inf. Theory* **2022**, *68*, 2551–2572. [CrossRef]
2. Song, G.J.; Ng, M.K.; Zhang, X. Tensor completion by multi-rank via unitary transformation. *Appl. Comput. Harmon. Anal.* **2023**, *65*, 348–373. [CrossRef]
3. Kernfeld, E.; Kilmer, M.; Aeron, S. Tensor–tensor products with invertible linear transforms. *Linear Algebra Appl.* **2015**, *485*, 545–570. [CrossRef]
4. Kilmer, M.E.; Horesh, L.; Avron, L.; Newman, E. Tensor-tensor algebra for optimal representation and compression. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2015851118. [CrossRef] [PubMed]
5. Kilmer, M.E.; Braman, K.; Hao, N.; Hoover, R.C. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM J. Matrix Anal. Appl.* **2013**, *34*, 148–172. [CrossRef]
6. Kilmer, M.E.; Martin, C.D. Factorization strategies for third-order tensors. *Linear Algebra Appl.* **2011**, *435*, 641–658. [CrossRef]
7. Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; Yan, S. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 925–938. [CrossRef] [PubMed]
8. Song, G.; Ng, M.K.; Zhang, X. Robust tensor completion using transformed tensor singular value decomposition. *Numer. Linear Algebra Appl.* **2020**, *27*, e2299. [CrossRef]
9. Golden, S. Lower bounds for the Helmholtz function. *Phys. Rev.* **1965**, *137*, B1127–B1128. [CrossRef]
10. Thompson, C.J. Inequality with applications in statistical mechanics. *J. Math. Phys.* **1965**, *6*, 1812–1813. [CrossRef]
11. Lieb, E.H. Convex trace functions and the Wigner–Yanase–Dyson conjecture. In *Les Rencontres Physiciens-Mathématiciens de Strasbourg–RCP25*; Springer: Berlin/Heidelberg, Germany, 1973; Volume 19, p. 35.
12. Chang, S.-Y.; Wei, Y. T-product tensors—Part II: Tail bounds for sums of random t-product tensors. *arXiv* **2021**, arXiv:2107.06224.
13. Chang, S.-Y.; Wei, Y. T-square tensors—Part I: Inequalities. *Comput. Appl. Math.* **2022**, *41*, 62. [CrossRef]

14. Hiai, F.; Petz, D. *Introduction to Matrix Analysis and Applications*; Springer: Cham, Switzerland, 2014.
15. Newman, E.; Keegan, K. Optimal matrix-mimetic tensor algebras via variable projection. *SIAM J. Matrix Anal. Appl.* **2025**, *46*, 1764–1790. [CrossRef]
16. Absil, P.-A.; Mahony, R.; Sepulchre, R. *Optimization Algorithms on Matrix Manifolds*; Princeton University Press: Princeton, NJ, USA, 2008.
17. Edelman, A.; Arias, T.A.; Smith, S.T. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.* **1998**, *20*, 303–353. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.