

21.8 Operator Convex Functions

Chapter 17 §11 defined two notions side by side and then spent itself on one of them. Operator monotonicity got four families of examples, a warning, and a classification to look forward to; operator convexity got one proposition — that $t \mapsto t^2$ is operator convex (Proposition 17.11.8) — and then the chapter moved on. This section repays the debt. It collects the examples that are within reach, and it proves the one structural theorem the notion has: an inequality of Jensen’s type that survives being squeezed by a contraction. That theorem is what makes operator convexity useful rather than merely curious.

Throughout, $F = \mathbb{R}$ or $F = \mathbb{C}$, every matrix whose spectrum is named is Hermitian, and I denotes an **interval** of real numbers. The letter I is plain italic and is not the identity matrix $\mathbf{I}$. For $f: I \rightarrow \mathbb{R}$ and Hermitian $\mathbf{A}$ with $\text{spec}(\mathbf{A}) \subseteq I$, the matrix $f(\mathbf{A})$ is the one built by the functional calculus of Chapter 12 §08 (Definition 12.8.2), and it is Hermitian because f is real-valued (Theorem 12.8.3 (c)).

21.8.1 What operator convexity asks

Recall the definition. A function $f: I \rightarrow \mathbb{R}$ is **operator convex on** I if, for **every** size $n \geq 1$, all Hermitian $\mathbf{A}, \mathbf{B} \in M_n(F)$ with spectra in I , and every $\theta \in [0, 1]$,

$$f(\theta \mathbf{A} + (1 - \theta) \mathbf{B}) \preceq \theta f(\mathbf{A}) + (1 - \theta) f(\mathbf{B})$$

in the Loewner order (Definition 17.11.1 (b), Definition 13.5.1). It is **operator concave** when $-f$ is operator convex. Chapter 17 §11 checked that the left-hand side makes sense: the spectrum of a convex combination of two Hermitian matrices with spectra in I again lies in I .

Taking $n = 1$ turns every matrix into a number and the Loewner order into $\geq$, so an operator convex function is convex in the sense of Definition 18.10.1. As with monotonicity, the content lies in the sizes $n \geq 2$, and the clause “for every n ” is what makes the class small.

Two facts from Chapter 17 §11 are the starting stock. Affine functions are operator convex and operator concave at once, with equality in the definition (Proposition 17.11.2), and $t \mapsto t^2$ is operator convex on $\mathbb{R}$, because

$$\theta \mathbf{A}^2 + (1 - \theta) \mathbf{B}^2 - (\theta \mathbf{A} + (1 - \theta) \mathbf{B})^2 = \theta(1 - \theta)(\mathbf{A} - \mathbf{B})^2$$

and the square of a Hermitian matrix is positive semidefinite (Proposition 17.11.8).

Before adding to the stock we record three bookkeeping identities. Each says that a familiar operation on matrices passes through the functional calculus untouched, and each is used below without further comment.

Lemma 21.8.1. Three Ways to Move the Functional Calculus

Let $I \subseteq \mathbb{R}$ be an interval and $f: I \rightarrow \mathbb{R}$.

- (a) **(Shift.)** Let $s \in \mathbb{R}$, let $g(t) = f(t + s)$ on $I - s$, and let $\mathbf{A}$ be Hermitian with $\text{spec}(\mathbf{A}) \subseteq I - s$. Then $g(\mathbf{A}) = f(\mathbf{A} + s\mathbf{I})$.
- (b) **(Unitary conjugation.)** Let $\mathbf{A}$ be Hermitian with $\text{spec}(\mathbf{A}) \subseteq I$ and let $\mathbf{U}$ be unitary of the same size. Then $\text{spec}(\mathbf{U}^* \mathbf{A} \mathbf{U}) = \text{spec}(\mathbf{A})$ and

$$f(\mathbf{U}^* \mathbf{A} \mathbf{U}) = \mathbf{U}^* f(\mathbf{A}) \mathbf{U}.$$

- (c) **(Direct sums.)** Let $\mathbf{A}, \mathbf{B}$ be Hermitian with spectra in I . Then $f(\mathbf{A} \oplus \mathbf{B}) = f(\mathbf{A}) \oplus f(\mathbf{B})$.

$f(\mathbf{B})$.

Proof. Write the spectral resolution $\mathbf{A} = \sum_j \mu_j \mathbf{P}_j$ with distinct μ_j (Definition 12.8.1, Theorem 12.8.1).

- (a) The numbers $\mu_j + s$ are distinct, and $\mathbf{A} + s\mathbf{I} = \sum_j (\mu_j + s) \mathbf{P}_j$ by Theorem 12.8.1 (c) and (d). The $\mathbf{P}_j$ are nonzero self-adjoint idempotents, pairwise annihilating, and summing to $\mathbf{I}$ (Theorem 12.8.1 (b), (c)), so this is the spectral resolution of $\mathbf{A} + s\mathbf{I}$ by Theorem 12.8.2. Hence $f(\mathbf{A} + s\mathbf{I}) = \sum_j f(\mu_j + s) \mathbf{P}_j = \sum_j g(\mu_j) \mathbf{P}_j = g(\mathbf{A})$, using Definition 12.8.2 twice.
- (b) Put $\mathbf{Q}_j = \mathbf{U}^* \mathbf{P}_j \mathbf{U}$. Each $\mathbf{Q}_j$ is self-adjoint, $\mathbf{Q}_j^2 = \mathbf{U}^* \mathbf{P}_j \mathbf{U} \mathbf{U}^* \mathbf{P}_j \mathbf{U} = \mathbf{Q}_j$, and $\mathbf{Q}_i \mathbf{Q}_j = \mathbf{O}$ for $i \neq j$, all because $\mathbf{U} \mathbf{U}^* = \mathbf{I}$; moreover $\sum_j \mathbf{Q}_j = \mathbf{U}^* \mathbf{I} \mathbf{U} = \mathbf{I}$ and $\mathbf{Q}_j \neq \mathbf{O}$ since $\mathbf{U}$ is invertible. Also $\mathbf{U}^* \mathbf{A} \mathbf{U} = \sum_j \mu_j \mathbf{Q}_j$. So this is the spectral resolution of $\mathbf{U}^* \mathbf{A} \mathbf{U}$ by Theorem 12.8.2, its distinct eigenvalues are the μ_j , and

$$f(\mathbf{U}^* \mathbf{A} \mathbf{U}) = \sum_j f(\mu_j) \mathbf{Q}_j = \mathbf{U}^* \left(\sum_j f(\mu_j) \mathbf{P}_j \right) \mathbf{U} = \mathbf{U}^* f(\mathbf{A}) \mathbf{U}.$$

- (c) Let $\nu_1, \dots, \nu_r$ be the distinct elements of $\text{spec}(\mathbf{A}) \cup \text{spec}(\mathbf{B})$. For each ν , let $\mathbf{P}_\nu^{\mathbf{A}}$ be the orthogonal projection onto $E_\nu(\mathbf{A})$, understood as $\mathbf{O}$ when $\nu \notin \text{spec}(\mathbf{A})$, and likewise $\mathbf{P}_\nu^{\mathbf{B}}$. A vector $(\mathbf{x}, \mathbf{y})$ satisfies $(\mathbf{A} \oplus \mathbf{B})(\mathbf{x}, \mathbf{y}) = \nu(\mathbf{x}, \mathbf{y})$ exactly when $\mathbf{A}\mathbf{x} = \nu\mathbf{x}$ and $\mathbf{B}\mathbf{y} = \nu\mathbf{y}$, so $E_\nu(\mathbf{A} \oplus \mathbf{B}) = E_\nu(\mathbf{A}) \oplus E_\nu(\mathbf{B})$, whose orthogonal projection is $\mathbf{P}_\nu^{\mathbf{A}} \oplus \mathbf{P}_\nu^{\mathbf{B}}$. Hence $\text{spec}(\mathbf{A} \oplus \mathbf{B}) = \{\nu_1, \dots, \nu_r\} \subseteq I$ and, by Definition 12.8.2,

$$f(\mathbf{A} \oplus \mathbf{B}) = \sum_\nu f(\nu) (\mathbf{P}_\nu^{\mathbf{A}} \oplus \mathbf{P}_\nu^{\mathbf{B}}) = f(\mathbf{A}) \oplus f(\mathbf{B}).$$

■

21.8.2 The examples we can prove

Operator convex functions form a cone, exactly as the operator monotone ones do, and for the same reason.

Proposition 21.8.2. Combining Operator Convex Functions

Let f and g be operator convex on I .

- (a) For all $c \geq 0$ and $\alpha, \beta \in \mathbb{R}$, the functions $f + g$ and $\alpha + \beta t + cf(t)$ are operator convex on I .
- (b) For every $s \in \mathbb{R}$, the translate $t \mapsto f(t + s)$ is operator convex on $I - s$.

Proof.

- (a) Fix Hermitian $\mathbf{A}, \mathbf{B}$ of the same size with spectra in I , fix $\theta \in [0, 1]$ and put $\mathbf{C} = \theta \mathbf{A} + (1 - \theta) \mathbf{B}$. By Theorem 12.8.3 (b), the defect

$$D_h := \theta h(\mathbf{A}) + (1 - \theta) h(\mathbf{B}) - h(\mathbf{C})$$

satisfies $D_{f+g} = D_f + D_g$ and $D_{cf} = c D_f$. Both D_f and D_g are $\succeq \mathbf{O}$ by hypothesis,

so for every $\mathbf{x}$ the quadratic forms add and scale to something ≥ 0 , giving $D_{f+g} \succeq \mathbf{O}$ and $D_{cf} \succeq \mathbf{O}$. Adding an affine function changes nothing, since $D_h = \mathbf{O}$ for affine h by Proposition 17.11.2.

- (b) Let $g(t) = f(t+s)$ and let $\mathbf{A}, \mathbf{B}$ be Hermitian with spectra in $I-s$. Then $\mathbf{A} + s\mathbf{I}$ and $\mathbf{B} + s\mathbf{I}$ have spectra in I , and $\theta(\mathbf{A} + s\mathbf{I}) + (1-\theta)(\mathbf{B} + s\mathbf{I}) = \mathbf{C} + s\mathbf{I}$. By Lemma 21.8.1 (a) applied three times,

$$g(\mathbf{C}) = f(\mathbf{C} + s\mathbf{I}) \preceq \theta f(\mathbf{A} + s\mathbf{I}) + (1-\theta)f(\mathbf{B} + s\mathbf{I}) = \theta g(\mathbf{A}) + (1-\theta)g(\mathbf{B}).$$

■

The new example is the reciprocal. Chapter 17 §11 left it as an exercise; part (d) of the theorem below rests on it, so it is proved here. The proof is the one the exercise's hint pointed at, and the whole content is that a positive definite matrix and its inverse sit in a single positive semidefinite block matrix.

Theorem 21.8.3. Four Operator Convex Functions

- (a) Every affine function $t \mapsto \alpha + \beta t$ is operator convex and operator concave on $\mathbb{R}$.
- (b) $t \mapsto t^2$ is operator convex on $\mathbb{R}$, hence on every interval.
- (c) $t \mapsto 1/t$ is operator convex on $(0, \infty)$; equivalently, $t \mapsto -1/t$ is operator concave there.
- (d) For every $c > 0$, the function $t \mapsto 1/(t+c)$ is operator convex on $(0, \infty)$, and $t \mapsto t/(t+c)$ is operator **concave** there.

Idea. Only (c) is new. For numbers, $1/t$ is convex because $1/t$ is the smallest u with $tu \geq 1$, and “ $tu \geq 1$ ” is a condition that survives averaging. The matrix version of “ $tu \geq 1$ ” is that the block matrix $\begin{pmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{I} & \mathbf{U} \end{pmatrix}$ is positive semidefinite, and Chapter 13 §06 reads that off a Schur complement. Averaging two such blocks averages the corners, and reading the Schur complement backwards turns the average back into an inequality between matrices.

Proof.

- (a) is Proposition 17.11.2 and (b) is Proposition 17.11.8. Restricting to a smaller interval only shrinks the supply of matrices, so both persist on every interval.
- (b) **Claim.** For every $\mathbf{G} \succ \mathbf{O}$ in $M_n(F)$, the Hermitian block matrix

$$\mathbf{M}_{\mathbf{G}} = \begin{pmatrix} \mathbf{G} & \mathbf{I}_n \\ \mathbf{I}_n & \mathbf{G}^{-1} \end{pmatrix}$$

is positive semidefinite.

Proof. The inverse of a Hermitian invertible matrix is Hermitian, since $(\mathbf{G}^{-1})^* = (\mathbf{G}^*)^{-1} = \mathbf{G}^{-1}$, so $\mathbf{M}_{\mathbf{G}}$ is Hermitian. Its corner $\mathbf{G}$ is $\succ \mathbf{O}$, and its Schur complement is $\mathbf{M}_{\mathbf{G}}/\mathbf{G} = \mathbf{G}^{-1} - \mathbf{I}_n \mathbf{G}^{-1} \mathbf{I}_n = \mathbf{O} \succeq \mathbf{O}$. By Theorem 13.6.1 (b), $\mathbf{M}_{\mathbf{G}} \succeq \mathbf{O}$.

■

Now let $\mathbf{A}, \mathbf{B} \in M_n(F)$ be Hermitian with spectra in $(0, \infty)$, so $\mathbf{A} \succ \mathbf{O}$ and $\mathbf{B} \succ \mathbf{O}$ by

Theorem 13.1.5 (b), and let $\theta \in [0, 1]$. Put $\mathbf{C} = \theta\mathbf{A} + (1 - \theta)\mathbf{B}$. For $\mathbf{x} \neq \mathbf{0}$,

$$\langle \mathbf{C}\mathbf{x}, \mathbf{x} \rangle = \theta \langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle + (1 - \theta) \langle \mathbf{B}\mathbf{x}, \mathbf{x} \rangle > 0,$$

since both terms are ≥ 0 and at least one coefficient is positive; hence $\mathbf{C} \succ \mathbf{O}$. By the Claim, $\mathbf{M}_{\mathbf{A}} \succeq \mathbf{O}$ and $\mathbf{M}_{\mathbf{B}} \succeq \mathbf{O}$, so their convex combination

$$\theta\mathbf{M}_{\mathbf{A}} + (1 - \theta)\mathbf{M}_{\mathbf{B}} = \begin{pmatrix} \mathbf{C} & \mathbf{I}_n \\ \mathbf{I}_n & \theta\mathbf{A}^{-1} + (1 - \theta)\mathbf{B}^{-1} \end{pmatrix}$$

has nonnegative quadratic form at every vector and is therefore $\succeq \mathbf{O}$. Its corner $\mathbf{C}$ is $\succ \mathbf{O}$, so Theorem 13.6.1 (b) applies in the other direction and gives

$$\theta\mathbf{A}^{-1} + (1 - \theta)\mathbf{B}^{-1} - \mathbf{C}^{-1} \succeq \mathbf{O}.$$

Since $t \mapsto 1/t$ sends a positive definite matrix to its inverse (Definition 12.8.2, as recalled in Chapter 17 §11), this is exactly the inequality required of $f(t) = 1/t$. The size n was arbitrary.

- (d) By (c) and Proposition 21.8.2 (b), $t \mapsto 1/(t + c)$ is operator convex on $(0, \infty) - c = (-c, \infty)$, hence on the smaller interval $(0, \infty)$. Finally $t/(t + c) = 1 - c \cdot \frac{1}{t+c}$, which is an affine function minus a nonnegative multiple of an operator convex function, so it is operator concave by Proposition 21.8.2 (a) applied to $-f$.

■

Remark. Compare the list with Chapter 17 §11's list of operator monotone functions. The two overlap only in the affine ones: t^2 is operator convex and not operator monotone, $-1/t$ is operator monotone and not even convex, and $t/(t + c)$ is both operator monotone (Corollary 17.11.5) and operator concave. Two further examples, $-\log t$ and $t \log t$, are operator convex on $(0, \infty)$; both proofs need an integral representation that §10 of this chapter supplies, so they are postponed to there.

21.8.3 Convex is not operator convex

The gap between convexity and operator convexity is as wide as the gap between increasing and operator monotone, and a single cube exposes it.

Example. The Cube Is Convex and Not Operator Convex

Let

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Show that $\mathbf{A} \succeq \mathbf{O}$ and $\mathbf{B} \succ \mathbf{O}$, and that with $\theta = \frac{1}{2}$ the inequality required of $f(t) = t^3$ on $[0, \infty)$ fails.

Solution. $\mathbf{A}$ is diagonal with nonnegative entries, so $\mathbf{A} \succeq \mathbf{O}$ with spectrum $\{0, 1\}$. $\mathbf{B}$ is real symmetric with leading principal minors 2 and $2 - 1 = 1$, both positive, so $\mathbf{B} \succ \mathbf{O}$ by

Sylvester's criterion (Theorem 13.1.5 (d)). Both spectra lie in $[0, \infty)$, and so does that of $\frac{1}{2}(\mathbf{A} + \mathbf{B})$, a convex combination of two positive semidefinite matrices.

Now compute. $\mathbf{A}^3 = \mathbf{A}$. From $\mathbf{B}^2 = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix}$ we get $\mathbf{B}^3 = \begin{pmatrix} 13 & 8 \\ 8 & 5 \end{pmatrix}$. Also $\mathbf{A} + \mathbf{B} = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}$, whose square is $\begin{pmatrix} 10 & 4 \\ 4 & 2 \end{pmatrix}$ and whose cube is $\begin{pmatrix} 34 & 14 \\ 14 & 6 \end{pmatrix}$. Hence, clearing the denominator by a factor 8,

$$8\left(\frac{1}{2}(\mathbf{A}^3 + \mathbf{B}^3) - \left(\frac{1}{2}(\mathbf{A} + \mathbf{B})\right)^3\right) = 4(\mathbf{A}^3 + \mathbf{B}^3) - (\mathbf{A} + \mathbf{B})^3 = \begin{pmatrix} 22 & 18 \\ 18 & 14 \end{pmatrix}.$$

Its determinant is $22 \cdot 14 - 18^2 = 308 - 324 = -16 < 0$, so it is not positive semidefinite (Theorem 13.1.4 (e)). A witness needs no eigenvalues: with $\mathbf{x} = (2, -3)$,

$$\mathbf{x}^\top \begin{pmatrix} 22 & 18 \\ 18 & 14 \end{pmatrix} \mathbf{x} = 4 \cdot 22 - 12 \cdot 18 + 9 \cdot 14 = 88 - 216 + 126 = -2 < 0.$$

By Theorem 12.8.3 (a), $t \mapsto t^3$ sends a Hermitian matrix to its cube, so t^3 is not operator convex on $[0, \infty)$. It is nevertheless convex there in the ordinary sense, since its second derivative $6t$ is ≥ 0 (Theorem 18.10.6).

Warning. Convexity of f is not enough, already at size 2. Every t^p with $p \geq 1$ is convex on $[0, \infty)$, and the exponents 3 and 4 already fail to be operator convex. The trace does not notice the difference: §02's Theorem 21.2.8 makes $\mathbf{A} \mapsto \operatorname{tr} f(\mathbf{A})$ convex for **every** convex f , and in [Example \(The Cube Is Convex and Not Operator Convex\)](#) the trace of the displayed matrix is $22 + 14 = 36 > 0$, exactly as that theorem predicts, while the matrix itself is indefinite. A convex function controls the sum of the eigenvalues; an operator convex function controls the matrix.

Quick check. Is $t \mapsto t^4$ operator convex on $[0, \infty)$, because $t^4 = (t^2)^2$ is a composite of two operator convex functions?

Solution. No. [Proposition 21.8.2](#) covers sums and nonnegative multiples, never composites, and there is no reason it should: operator convexity of g constrains g on convex combinations of matrices, while $(g \circ f)(\mathbf{A}) = g(f(\mathbf{A}))$ asks about g at the matrix $f(\mathbf{A})$, which is a different question. The same pair as in [Example \(The Cube Is Convex and Not Operator Convex\)](#) settles it. From $\mathbf{A}^4 = \mathbf{A}$, $\mathbf{B}^4 = \begin{pmatrix} 34 & 21 \\ 21 & 13 \end{pmatrix}$, and $(\mathbf{A} + \mathbf{B})^4 = \begin{pmatrix} 116 & 48 \\ 48 & 20 \end{pmatrix}$,

$$4\left(\frac{1}{2}(\mathbf{A}^4 + \mathbf{B}^4) - \left(\frac{1}{2}(\mathbf{A} + \mathbf{B})\right)^4\right) = \begin{pmatrix} 41 & 30 \\ 30 & 21 \end{pmatrix},$$

whose determinant is $861 - 900 = -39 < 0$. A witness is $\mathbf{x} = (3, -4)$, giving $9 \cdot 41 - 24 \cdot 30 + 16 \cdot 21 = -15 < 0$.

21.8.4 Contractions and their dilations

The theorem this section is heading for compares $f(\mathbf{V}^* \mathbf{A} \mathbf{V})$ with $\mathbf{V}^* f(\mathbf{A}) \mathbf{V}$. The matrices $\mathbf{V}$ for which this can possibly work are the ones that do not stretch: otherwise $\mathbf{V}^* \mathbf{A} \mathbf{V}$ can leave

the interval where f lives.

A *contraction* is a matrix that never lengthens a vector.

Definition 21.8.1. Contraction

Let $\mathbf{V} \in M_{n \times m}(F)$. Then $\mathbf{V}$ is a **contraction** if $\mathbf{V}^* \mathbf{V} \preceq \mathbf{I}_m$, that is, if $\|\mathbf{V}\mathbf{x}\| \leq \|\mathbf{x}\|$ for every $\mathbf{x} \in F^m$. It is an **isometry** if $\mathbf{V}^* \mathbf{V} = \mathbf{I}_m$.

The word is overloaded: Chapter 15 §04 uses **contraction** for the pairing $V^* \otimes V \rightarrow F$ (Definition 15.4.1), which is a different object entirely, and the two never appear together. The two phrasings above agree because $\langle (\mathbf{I}_m - \mathbf{V}^* \mathbf{V})\mathbf{x}, \mathbf{x} \rangle = \|\mathbf{x}\|^2 - \|\mathbf{V}\mathbf{x}\|^2$. Examples: every unitary matrix; every orthogonal projection $\mathbf{P}$, since $\mathbf{P}^* \mathbf{P} = \mathbf{P}^2 = \mathbf{P} \preceq \mathbf{I}$ (Definition 13.5.1); the matrix $\frac{1}{2}\mathbf{I}$; and any matrix whose columns are an orthonormal list, which is an isometry. A nonexample by minimal change: $2\mathbf{I}$ fails, since $\mathbf{I} - (2\mathbf{I})^*(2\mathbf{I}) = -3\mathbf{I}$ is not positive semidefinite.

The key structural fact is that a contraction is always the corner of a unitary matrix of twice the size. This is the **dilation**, and it is how a statement about contractions gets converted into a statement about unitary conjugation, which the functional calculus understands.

Lemma 21.8.4. Unitary Dilation of a Contraction

Let $\mathbf{V} \in M_{n \times m}(F)$ be a contraction, and set

$$\mathbf{D}_1 = (\mathbf{I}_m - \mathbf{V}^* \mathbf{V})^{1/2}, \quad \mathbf{D}_2 = (\mathbf{I}_n - \mathbf{V} \mathbf{V}^*)^{1/2}.$$

Then $\mathbf{I}_m - \mathbf{V}^* \mathbf{V} \succeq \mathbf{O}$ and $\mathbf{I}_n - \mathbf{V} \mathbf{V}^* \succeq \mathbf{O}$, so both square roots exist, and the $(n+m) \times (n+m)$ matrix

$$\mathbf{U} = \begin{pmatrix} \mathbf{V} & \mathbf{D}_2 \\ \mathbf{D}_1 & -\mathbf{V}^* \end{pmatrix}$$

is unitary.

Idea. Only one of the four block identities needs an idea: $\mathbf{V}^* \mathbf{D}_2 = \mathbf{D}_1 \mathbf{V}^*$, which says that $\mathbf{V}^*$ intertwines the two square roots. It holds for $\mathbf{V}^*(\mathbf{I} - \mathbf{V} \mathbf{V}^*) = (\mathbf{I} - \mathbf{V}^* \mathbf{V})\mathbf{V}^*$ by one line of algebra, hence for every polynomial in the two matrices, and the positive square root **is** a polynomial (Theorem 13.2.1 (a)) — provided one polynomial can be chosen to serve both matrices at once, which Lagrange interpolation allows.

Proof. Write $\mathbf{M} = \mathbf{I}_m - \mathbf{V}^* \mathbf{V}$ and $\mathbf{N} = \mathbf{I}_n - \mathbf{V} \mathbf{V}^*$. By hypothesis $\mathbf{M} \succeq \mathbf{O}$, which says $\|\mathbf{V}\mathbf{x}\| \leq \|\mathbf{x}\|$ for every $\mathbf{x} \in F^m$. For $\mathbf{N}$, fix $\mathbf{y} \in F^n$ and put $\mathbf{z} = \mathbf{V}^* \mathbf{y}$. Then, by Cauchy–Schwarz (Theorem 11.1.5) and the contraction property applied to $\mathbf{z}$,

$$\|\mathbf{z}\|^2 = \langle \mathbf{V}\mathbf{z}, \mathbf{y} \rangle \leq \|\mathbf{V}\mathbf{z}\| \|\mathbf{y}\| \leq \|\mathbf{z}\| \|\mathbf{y}\|,$$

the first equality because $\langle \mathbf{V}\mathbf{z}, \mathbf{y} \rangle = \mathbf{y}^* \mathbf{V} \mathbf{V}^* \mathbf{y} = \|\mathbf{V}^* \mathbf{y}\|^2$. If $\mathbf{z} \neq \mathbf{0}$ we may divide by $\|\mathbf{z}\| > 0$ to get $\|\mathbf{V}^* \mathbf{y}\| \leq \|\mathbf{y}\|$, and if $\mathbf{z} = \mathbf{0}$ that inequality is immediate. Hence $\langle \mathbf{N}\mathbf{y}, \mathbf{y} \rangle = \|\mathbf{y}\|^2 - \|\mathbf{V}^* \mathbf{y}\|^2 \geq 0$ for every $\mathbf{y}$, so $\mathbf{N} \succeq \mathbf{O}$. Both square roots now exist and are Hermitian, by Theorem 13.2.1.

Claim. $\mathbf{V}^* \mathbf{N}^{1/2} = \mathbf{M}^{1/2} \mathbf{V}^*$.

Proof. First, $\mathbf{V}^*\mathbf{N} = \mathbf{V}^* - \mathbf{V}^*\mathbf{V}\mathbf{V}^* = \mathbf{M}\mathbf{V}^*$. Hence $\mathbf{V}^*\mathbf{N}^k = \mathbf{M}^k\mathbf{V}^*$ for every integer $k \geq 0$, by induction: the case $k = 0$ is trivial, and $\mathbf{V}^*\mathbf{N}^{k+1} = (\mathbf{V}^*\mathbf{N})\mathbf{N}^k = \mathbf{M}(\mathbf{V}^*\mathbf{N}^k) = \mathbf{M}^{k+1}\mathbf{V}^*$. Therefore $\mathbf{V}^*q(\mathbf{N}) = q(\mathbf{M})\mathbf{V}^*$ for every polynomial q with real coefficients.

The set $\text{spec}(\mathbf{M}) \cup \text{spec}(\mathbf{N})$ is finite and contained in $[0, \infty)$, so by Lagrange interpolation (Theorem 6.6.1) there is a real polynomial q with $q(\nu) = \sqrt{\nu}$ for every ν in it. By Theorem 12.8.3 (a) and Definition 12.8.2, $q(\mathbf{M}) = \sum_j q(\mu_j)\mathbf{P}_j = \sum_j \sqrt{\mu_j}\mathbf{P}_j = \mathbf{M}^{1/2}$, and likewise $q(\mathbf{N}) = \mathbf{N}^{1/2}$. Substituting q into the displayed identity proves the claim. ■

Taking adjoints in the claim, and using that $\mathbf{M}^{1/2}$ and $\mathbf{N}^{1/2}$ are Hermitian, gives $\mathbf{N}^{1/2}\mathbf{V} = \mathbf{V}\mathbf{M}^{1/2}$. Now compute, in blocks (Theorem 8.1.1),

$$\begin{aligned} \mathbf{U}^*\mathbf{U} &= \begin{pmatrix} \mathbf{V}^* & \mathbf{D}_1 \\ \mathbf{D}_2 & -\mathbf{V} \end{pmatrix} \begin{pmatrix} \mathbf{V} & \mathbf{D}_2 \\ \mathbf{D}_1 & -\mathbf{V}^* \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{V}^*\mathbf{V} + \mathbf{D}_1^2 & \mathbf{V}^*\mathbf{D}_2 - \mathbf{D}_1\mathbf{V}^* \\ \mathbf{D}_2\mathbf{V} - \mathbf{V}\mathbf{D}_1 & \mathbf{D}_2^2 + \mathbf{V}\mathbf{V}^* \end{pmatrix}. \end{aligned}$$

The top-left block is $\mathbf{V}^*\mathbf{V} + \mathbf{M} = \mathbf{I}_m$ and the bottom-right is $\mathbf{N} + \mathbf{V}\mathbf{V}^* = \mathbf{I}_n$, both by the definition of $\mathbf{M}$ and $\mathbf{N}$. The top-right block is $\mathbf{O}$ by the Claim, and the bottom-left is $\mathbf{O}$ by its adjoint form. So $\mathbf{U}^*\mathbf{U} = \mathbf{I}_{m+n}$, and since $\mathbf{U}$ is square this makes $\mathbf{U}$ unitary (Theorem 3.4.4). ■

Quick check. Take $m = n = 1$ and $\mathbf{V} = (c)$ with c real, $|c| \leq 1$. What is the dilation?

Solution. $\mathbf{D}_1 = \mathbf{D}_2 = (\sqrt{1-c^2})$, so $\mathbf{U} = \begin{pmatrix} c & \sqrt{1-c^2} \\ \sqrt{1-c^2} & -c \end{pmatrix}$. Writing $c = \cos \varphi$ with $\varphi \in [0, \pi]$, so that $\sin \varphi \geq 0$ and $\sqrt{1-c^2} = \sin \varphi$, this is the reflection matrix $\begin{pmatrix} \cos \varphi & \sin \varphi \\ \sin \varphi & -\cos \varphi \end{pmatrix}$, which is orthogonal with determinant -1 . So the lemma is the statement that every number in $[-1, 1]$ is the corner of a 2×2 reflection — the sentence “ $\cos^2 \varphi + \sin^2 \varphi = 1$ ”, read as a matrix identity.

21.8.5 Jensen’s inequality for operators

Here is the theorem. In words: **squeezing first and then applying f is at most applying f first and then squeezing**, provided f is operator convex and does not reward the squeezing at the origin.

Theorem 21.8.5. Hansen–Pedersen–Jensen Inequality

Let $I \subseteq \mathbb{R}$ be an interval with $0 \in I$, let $f: I \rightarrow \mathbb{R}$ be operator convex on I , let $\mathbf{A} \in M_n(F)$ be Hermitian with $\text{spec}(\mathbf{A}) \subseteq I$, and let $\mathbf{V} \in M_{n \times m}(F)$ be a contraction. Then $\text{spec}(\mathbf{V}^*\mathbf{A}\mathbf{V}) \subseteq I$, and

$$f(\mathbf{V}^*\mathbf{A}\mathbf{V}) \preceq \mathbf{V}^*f(\mathbf{A})\mathbf{V} + f(0)(\mathbf{I}_m - \mathbf{V}^*\mathbf{V}).$$

In particular:

- (a) if $f(0) \leq 0$, then $f(\mathbf{V}^* \mathbf{A} \mathbf{V}) \preceq \mathbf{V}^* f(\mathbf{A}) \mathbf{V}$;
- (b) if $\mathbf{V}$ is an isometry, then $f(\mathbf{V}^* \mathbf{A} \mathbf{V}) \preceq \mathbf{V}^* f(\mathbf{A}) \mathbf{V}$, with no condition on $f(0)$.

Idea. Dilate. Once $\mathbf{V}$ sits in the corner of a unitary $\mathbf{U}$, the matrix $\mathbf{V}^* \mathbf{A} \mathbf{V}$ sits in the corner of $\mathbf{U}^*(\mathbf{A} \oplus \mathbf{O})\mathbf{U}$, which is a unitary conjugate of $\mathbf{A} \oplus \mathbf{O}$ and therefore has the same spectrum. Conjugating a second time by the sign matrix $\mathbf{W} = \mathbf{I}_m \oplus (-\mathbf{I}_n)$ flips the off-diagonal blocks, so **averaging the two conjugates deletes them**: the average is block diagonal, with $\mathbf{V}^* \mathbf{A} \mathbf{V}$ in the corner. Operator convexity applied to that average, and then reading off the top-left block, is the whole proof. The term $f(0)(\mathbf{I} - \mathbf{V}^* \mathbf{V})$ is the residue of the zero block we padded with.

Proof. The spectrum. Let $\mathbf{x} \in F^m$ be a unit eigenvector of the Hermitian matrix $\mathbf{V}^* \mathbf{A} \mathbf{V}$, with eigenvalue μ . If $\mathbf{V}\mathbf{x} = \mathbf{0}$, then $\mu = \langle \mathbf{V}^* \mathbf{A} \mathbf{V} \mathbf{x}, \mathbf{x} \rangle = 0 \in I$. Otherwise put $\mathbf{y} = \mathbf{V}\mathbf{x} / \|\mathbf{V}\mathbf{x}\|$ and $\tau = \|\mathbf{V}\mathbf{x}\|^2$, so $0 < \tau \leq 1$ because $\mathbf{V}$ is a contraction, and

$$\mu = \langle \mathbf{A} \mathbf{V} \mathbf{x}, \mathbf{V} \mathbf{x} \rangle = \tau \langle \mathbf{A} \mathbf{y}, \mathbf{y} \rangle.$$

By Lemma 13.5.2, $\langle \mathbf{A} \mathbf{y}, \mathbf{y} \rangle$ lies between $\lambda_n(\mathbf{A})$ and $\lambda_1(\mathbf{A})$, hence in I , since I is an interval containing the spectrum. Then $\mu = \tau \langle \mathbf{A} \mathbf{y}, \mathbf{y} \rangle + (1 - \tau) \cdot 0$ is a convex combination of two points of I , so $\mu \in I$. The same argument applies to $\mathbf{D}_2 = (\mathbf{I}_n - \mathbf{V} \mathbf{V}^*)^{1/2}$: it is Hermitian with $\mathbf{D}_2^* \mathbf{D}_2 = \mathbf{I}_n - \mathbf{V} \mathbf{V}^* \preceq \mathbf{I}_n$, because $\mathbf{V} \mathbf{V}^* \succeq \mathbf{O}$, so it is a contraction. Hence $\text{spec}(\mathbf{D}_2 \mathbf{A} \mathbf{D}_2) \subseteq I$.

The two conjugates. Let $\mathbf{U}$ be the dilation of Lemma 21.8.4 and put

$$\mathbf{M} = \mathbf{A} \oplus \mathbf{O}_m, \quad \mathbf{W} = \mathbf{I}_m \oplus (-\mathbf{I}_n),$$

so that $\mathbf{W}$ is unitary and Hermitian, and $\text{spec}(\mathbf{M}) = \text{spec}(\mathbf{A}) \cup \{0\} \subseteq I$. Multiplying out in blocks,

$$\mathbf{U}^* \mathbf{M} \mathbf{U} = \begin{pmatrix} \mathbf{V}^* \mathbf{A} \mathbf{V} & \mathbf{V}^* \mathbf{A} \mathbf{D}_2 \\ \mathbf{D}_2 \mathbf{A} \mathbf{V} & \mathbf{D}_2 \mathbf{A} \mathbf{D}_2 \end{pmatrix},$$

and conjugating by $\mathbf{W}$ negates the two off-diagonal blocks. Hence, with $\mathbf{X} = \mathbf{U}^* \mathbf{M} \mathbf{U}$ and $\mathbf{Y} = (\mathbf{U} \mathbf{W})^* \mathbf{M} (\mathbf{U} \mathbf{W}) = \mathbf{W}^* \mathbf{X} \mathbf{W}$,

$$\frac{1}{2}(\mathbf{X} + \mathbf{Y}) = (\mathbf{V}^* \mathbf{A} \mathbf{V}) \oplus (\mathbf{D}_2 \mathbf{A} \mathbf{D}_2). \quad (*)$$

Both $\mathbf{X}$ and $\mathbf{Y}$ are unitary conjugates of $\mathbf{M}$, so both are Hermitian with spectrum $\text{spec}(\mathbf{M}) \subseteq I$ by Lemma 21.8.1 (b).

Convexity. Apply Definition 17.11.1 (b) with $\theta = \frac{1}{2}$ to $\mathbf{X}$ and $\mathbf{Y}$, then Lemma 21.8.1 (b) twice:

$$f\left(\frac{1}{2}(\mathbf{X} + \mathbf{Y})\right) \preceq \frac{1}{2}(\mathbf{U}^* f(\mathbf{M}) \mathbf{U} + \mathbf{W}^* \mathbf{U}^* f(\mathbf{M}) \mathbf{U} \mathbf{W}).$$

By Lemma 21.8.1 (c), $f(\mathbf{M}) = f(\mathbf{A}) \oplus f(0)\mathbf{I}_m$. The same block computation as before, now with $f(\mathbf{A})$ in place of $\mathbf{A}$ and $f(0)\mathbf{I}_m$ in place of $\mathbf{O}_m$, gives

$$\mathbf{U}^* f(\mathbf{M}) \mathbf{U} = \begin{pmatrix} \mathbf{V}^* f(\mathbf{A}) \mathbf{V} + f(0)\mathbf{D}_1^2 & * \\ * & * \end{pmatrix},$$

and averaging with its $\mathbf{W}$ -conjugate again deletes the off-diagonal blocks. On the left, Lemma 21.8.1 (c) and (*) give $f(\frac{1}{2}(\mathbf{X} + \mathbf{Y})) = f(\mathbf{V}^*\mathbf{A}\mathbf{V}) \oplus f(\mathbf{D}_2\mathbf{A}\mathbf{D}_2)$. So the inequality reads, between two block diagonal matrices,

$$f(\mathbf{V}^*\mathbf{A}\mathbf{V}) \oplus f(\mathbf{D}_2\mathbf{A}\mathbf{D}_2) \preceq (\mathbf{V}^*f(\mathbf{A})\mathbf{V} + f(0)\mathbf{D}_1^2) \oplus \mathbf{Z}$$

for some Hermitian $\mathbf{Z}$. The difference of the two sides is positive semidefinite, so its top-left $m \times m$ block is too, by Theorem 13.6.1 (a). Since $\mathbf{D}_1^2 = \mathbf{I}_m - \mathbf{V}^*\mathbf{V}$, that block inequality is the displayed statement of the theorem.

- (a) If $f(0) \leq 0$, then $f(0)(\mathbf{I}_m - \mathbf{V}^*\mathbf{V}) \preceq \mathbf{O}$, because $\mathbf{I}_m - \mathbf{V}^*\mathbf{V} \succeq \mathbf{O}$ and a nonpositive multiple of a positive semidefinite matrix has nonpositive quadratic form. Adding it can only decrease, so $\mathbf{V}^*f(\mathbf{A})\mathbf{V} + f(0)(\mathbf{I}_m - \mathbf{V}^*\mathbf{V}) \preceq \mathbf{V}^*f(\mathbf{A})\mathbf{V}$, and transitivity of $\preceq$ (Proposition 13.5.1) finishes.
- (b) If $\mathbf{V}^*\mathbf{V} = \mathbf{I}_m$ the extra term is $\mathbf{O}$.

■

The special case worth remembering separately takes $\mathbf{V}$ to be a projection. It says that **an operator convex function commutes with compression, up to the right inequality**.

Corollary 21.8.6. Compression

Let I be an interval with $0 \in I$, let $f: I \rightarrow \mathbb{R}$ be operator convex with $f(0) \leq 0$, let $\mathbf{A} \in M_n(F)$ be Hermitian with $\text{spec}(\mathbf{A}) \subseteq I$, and let $\mathbf{P} \in M_n(F)$ be the matrix of an orthogonal projection. Then $\text{spec}(\mathbf{PAP}) \subseteq I$ and

$$f(\mathbf{PAP}) \preceq \mathbf{P}f(\mathbf{A})\mathbf{P}.$$

Proof. $\mathbf{P}$ is Hermitian with $\mathbf{P}^*\mathbf{P} = \mathbf{P}^2 = \mathbf{P} \preceq \mathbf{I}$, so $\mathbf{P}$ is a contraction (Definition 21.8.1). Apply Theorem 21.8.5 (a) with $\mathbf{V} = \mathbf{P}$, noting $\mathbf{P}^*\mathbf{A}\mathbf{P} = \mathbf{PAP}$.

■

Warning. The condition $f(0) \leq 0$ is not decoration. Take $f(t) = t^2 + 1$, operator convex on $\mathbb{R}$ by Proposition 21.8.2 (a), with $f(0) = 1 > 0$. Take $\mathbf{A} = \mathbf{O}_n$ and $\mathbf{V} = \mathbf{O}_{n \times m}$, a contraction. Then $f(\mathbf{V}^*\mathbf{A}\mathbf{V}) = f(\mathbf{O}_m) = \mathbf{I}_m$ while $\mathbf{V}^*f(\mathbf{A})\mathbf{V} = \mathbf{O}$, so the conclusion of Theorem 21.8.5 (a) fails as badly as it can. The full inequality of the theorem survives: its right-hand side is $\mathbf{O} + 1 \cdot \mathbf{I}_m = \mathbf{I}_m$, with equality.

Quick check. Why can the hypothesis “ $\mathbf{V}$ is a contraction” not be dropped in Corollary 21.8.6 by rescaling — that is, why does the corollary for $\mathbf{P}$ not give a version for $c\mathbf{P}$ with $c > 1$?

Solution. Because both sides scale differently. For $f(t) = t^2$ and $\mathbf{V} = c\mathbf{I}$ the claimed inequality would read $c^4\mathbf{A}^2 \preceq c^2\mathbf{A}^2$, which fails for every $c > 1$ and every $\mathbf{A}$ with $\mathbf{A}^2 \neq \mathbf{O}$. The contraction hypothesis is what keeps the left-hand side from being inflated, and it enters

the proof twice: it puts $\text{spec}(\mathbf{V}^*\mathbf{A}\mathbf{V})$ back inside I , which is what makes $f(\mathbf{V}^*\mathbf{A}\mathbf{V})$ defined at all, and it gives the square roots of [Lemma 21.8.4](#) something to be square roots of.

21.8.6 Where this sits

Two convexity statements about matrices are now on the table, and they are of very different strengths.

- §02's Theorem 21.2.8: for **every** convex f , the function $\mathbf{A} \mapsto \text{tr } f(\mathbf{A})$ is convex on the Hermitian matrices with spectrum in I . This is a statement about one real number attached to the matrix.
- Operator convexity: for a much smaller class of f , the matrices themselves are ordered. [Example \(The Cube Is Convex and Not Operator Convex\)](#) shows t^3 in the first class and not the second.

The Hansen–Pedersen–Jensen inequality is the reason the smaller class is worth isolating. A trace inequality cannot be compressed: knowing $\text{tr } f(\mathbf{A})$ says nothing about $\text{tr } f(\mathbf{PAP})$, since the compression changes the eigenvalues in ways the trace does not track. What [Corollary 21.8.6](#) compares is not the traces but the matrices, $f(\mathbf{PAP})$ with $\mathbf{P}f(\mathbf{A})\mathbf{P}$, and it is the step that later turns an operator convex function into an inequality between matrices of different sizes. Section 10 of this chapter adds $-\log t$ and $t \log t$ to the list of operator convex functions, and §11 uses the operator monotonicity of $t^{1/2}$ to build the geometric mean of two positive definite matrices.

21.8.7 Exercises

21.8.7.1 A. Check your understanding

Exercise. A1

- State what it means for $f: I \rightarrow \mathbb{R}$ to be operator convex, and say what the clause “for every n ” ranges over.
- Explain why an operator convex function is convex, and name a convex function that is not operator convex.
- Decide whether the following is correct, and justify your answer: if f is operator convex on I and $c < 0$, then cf is operator convex on I .
- State the Hansen–Pedersen–Jensen inequality, and say which hypothesis fails for $f(t) = t^2 + 1$.
- Is $t \mapsto -1/t$ operator convex on $(0, \infty)$? Is it operator concave there?

Solution.

- For every $n \geq 1$, all Hermitian $\mathbf{A}, \mathbf{B} \in M_n(F)$ with spectra in I and every $\theta \in [0, 1]$, $f(\theta\mathbf{A} + (1 - \theta)\mathbf{B}) \preceq \theta f(\mathbf{A}) + (1 - \theta)f(\mathbf{B})$ (Definition 17.11.1 (b)). The clause ranges over all matrix sizes at once.
- Taking $n = 1$ turns the matrices into numbers and $\preceq$ into $\leq$, which is Definition 18.10.1. The function $t \mapsto t^3$ is convex on $[0, \infty)$ and not operator convex there ([Example \(The Cube Is Convex and Not Operator Convex\)](#)).
- Incorrect: [Proposition 21.8.2](#) (a) requires $c \geq 0$. For $c = -1$ and $f(t) = t^2$, the claim would make $-t^2$ operator convex, that is t^2 operator concave; but [Proposition 17.11.8](#)

gives $\theta \mathbf{A}^2 + (1 - \theta) \mathbf{B}^2 - (\theta \mathbf{A} + (1 - \theta) \mathbf{B})^2 = \theta(1 - \theta)(\mathbf{A} - \mathbf{B})^2$, which is nonzero for $\mathbf{A} \neq \mathbf{B}$ and $0 < \theta < 1$, so the reverse inequality fails.

- (d) [Theorem 21.8.5](#): for f operator convex on an interval containing 0, $\mathbf{A}$ Hermitian with spectrum in that interval and $\mathbf{V}$ a contraction, $f(\mathbf{V}^* \mathbf{A} \mathbf{V}) \preceq \mathbf{V}^* f(\mathbf{A}) \mathbf{V} + f(0)(\mathbf{I} - \mathbf{V}^* \mathbf{V})$. The last term vanishes when $\mathbf{V}$ is an isometry, and when $f(0) \leq 0$ it is $\preceq \mathbf{O}$ and may be discarded, which leaves $f(\mathbf{V}^* \mathbf{A} \mathbf{V}) \preceq \mathbf{V}^* f(\mathbf{A}) \mathbf{V}$ in both cases. For $f(t) = t^2 + 1$ the failing hypothesis is $f(0) \leq 0$.
- (e) It is not operator convex: at $n = 1$ the scalar function $-1/t$ is not convex, since at 1, 3 and their midpoint 2 the value $-\frac{1}{2}$ exceeds the chord's $-\frac{2}{3}$. It is operator concave, by [Theorem 21.8.3](#) (c).

21.8.7.2 B. Practice

Exercise. B1: Which are operator convex?

Determine which of the following functions are operator convex on the interval given. Justify your answer.

- (i) $f(t) = 7 - 3t$ on $\mathbb{R}$.
- (ii) $f(t) = 2t^2 + 5t - 1$ on $\mathbb{R}$.
- (iii) $f(t) = \frac{4}{t+3}$ on $(0, \infty)$.
- (iv) $f(t) = \frac{t}{t+3}$ on $(0, \infty)$.
- (v) $f(t) = t^3$ on $[0, \infty)$.

Solution.

- (i) Operator convex (and concave), by [Theorem 21.8.3](#) (a).
- (ii) Operator convex: it is 2 times t^2 plus the affine function $5t - 1$, so [Proposition 21.8.2](#) (a) applies with $c = 2 \geq 0$.
- (iii) Operator convex, by [Theorem 21.8.3](#) (d) with $c = 3$ and [Proposition 21.8.2](#) (a) with the factor $4 \geq 0$.
- (iv) Not operator convex; it is operator **concave** by [Theorem 21.8.3](#) (d). Operator convexity would force ordinary convexity at $n = 1$, and $t/(t+3) = 1 - 3/(t+3)$ has second derivative $-6/(t+3)^3 < 0$ on $(0, \infty)$, so it is strictly concave there ([Theorem 18.10.6](#)).
- (v) Not operator convex, by [Example \(The Cube Is Convex and Not Operator Convex\)](#).

Exercise. B2: A dilation by hand

Let $\mathbf{V} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix}$.

- (a) Verify that $\mathbf{V}$ is a contraction and compute $\mathbf{D}_1 = (\mathbf{I}_2 - \mathbf{V}^* \mathbf{V})^{1/2}$ and $\mathbf{D}_2 = (\mathbf{I}_2 - \mathbf{V} \mathbf{V}^*)^{1/2}$.

- (b) Write down the 4×4 matrix $\mathbf{U}$ of [Lemma 21.8.4](#) and verify directly that its first column is a unit vector orthogonal to its third column.

Solution.

- (a) $\mathbf{V}^*\mathbf{V} = \mathbf{V}\mathbf{V}^* = \text{diag}(\frac{1}{4}, 0)$, so $\mathbf{I}_2 - \mathbf{V}^*\mathbf{V} = \text{diag}(\frac{3}{4}, 1) \succeq \mathbf{O}$, and $\mathbf{V}$ is a contraction. Both $\mathbf{I}_2 - \mathbf{V}^*\mathbf{V}$ and $\mathbf{I}_2 - \mathbf{V}\mathbf{V}^*$ equal $\text{diag}(\frac{3}{4}, 1)$, a diagonal matrix with nonnegative entries, so

$$\mathbf{D}_1 = \mathbf{D}_2 = \text{diag}(\frac{\sqrt{3}}{2}, 1),$$

the unique positive semidefinite square root (Theorem 13.2.1).

- (b) With the blocks in the order of the lemma,

$$\mathbf{U} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & 0 & 1 \\ \frac{\sqrt{3}}{2} & 0 & -\frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Its first column is $(\frac{1}{2}, 0, \frac{\sqrt{3}}{2}, 0)$, of squared length $\frac{1}{4} + \frac{3}{4} = 1$. Its third column is $(\frac{\sqrt{3}}{2}, 0, -\frac{1}{2}, 0)$, and the inner product of the two is $\frac{1}{2} \cdot \frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{2} \cdot (-\frac{1}{2}) = 0$.

Exercise. B3: Compression of a square

Let

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Compute $(\mathbf{PAP})^2$ and $\mathbf{PA}^2\mathbf{P}$, and verify the inequality of [Corollary 21.8.6](#) for $f(t) = t^2$. Hence state what the inequality says about the entry in position $(1, 1)$.

Solution. $\mathbf{P}$ is the orthogonal projection onto $\text{span}(\mathbf{e}_1)$, and $\mathbf{PAP} = \text{diag}(2, 0)$, so $(\mathbf{PAP})^2 = \text{diag}(4, 0)$. From $\mathbf{A}^2 = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix}$ we get $\mathbf{PA}^2\mathbf{P} = \text{diag}(5, 0)$. Hence

$$\mathbf{PA}^2\mathbf{P} - (\mathbf{PAP})^2 = \text{diag}(1, 0) \succeq \mathbf{O},$$

as [Corollary 21.8.6](#) predicts for $f(t) = t^2$, which is operator convex with $f(0) = 0$. In position $(1, 1)$ the inequality says $(\mathbf{A}^2)_{11} \geq (a_{11})^2$, that is $5 \geq 4$: the $(1, 1)$ entry of $\mathbf{A}^2$ is $\sum_j |a_{1j}|^2$, which exceeds $|a_{11}|^2$ by the contribution of the off-diagonal entries.

21.8.7.3 C. Going deeper

Exercise. C1: The square, without the dilation

Let $\mathbf{A} \in M_n(F)$ be Hermitian and let $\mathbf{V} \in M_{n \times m}(F)$ be a contraction.

- (a) Prove that $(\mathbf{V}^*\mathbf{A}\mathbf{V})^2 \preceq \mathbf{V}^*\mathbf{A}^2\mathbf{V}$ directly, without using [Theorem 21.8.5](#).
 (b) Which hypothesis of [Theorem 21.8.5](#) does your proof use, and which does it not?

Hint for (a): insert $\mathbf{I} - \mathbf{V}\mathbf{V}^*$.

Solution.

- (a) Compute the difference and factor it:

$$\mathbf{V}^* \mathbf{A}^2 \mathbf{V} - (\mathbf{V}^* \mathbf{A} \mathbf{V})^2 = \mathbf{V}^* \mathbf{A} \mathbf{A} \mathbf{V} - \mathbf{V}^* \mathbf{A} \mathbf{V} \mathbf{V}^* \mathbf{A} \mathbf{V} = \mathbf{V}^* \mathbf{A} (\mathbf{I}_n - \mathbf{V} \mathbf{V}^*) \mathbf{A} \mathbf{V}.$$

As in the proof of [Lemma 21.8.4](#), $\mathbf{I}_n - \mathbf{V} \mathbf{V}^* \succeq \mathbf{O}$ because $\mathbf{V}$ is a contraction. Writing $\mathbf{S} = \mathbf{A} \mathbf{V}$, the right-hand side is $\mathbf{S}^* (\mathbf{I}_n - \mathbf{V} \mathbf{V}^*) \mathbf{S}$, which is $\succeq \mathbf{O}$ by Proposition 13.1.2 (a). Hence $(\mathbf{V}^* \mathbf{A} \mathbf{V})^2 \preceq \mathbf{V}^* \mathbf{A}^2 \mathbf{V}$.

- (b) It uses the contraction hypothesis, in the single step $\mathbf{I}_n - \mathbf{V} \mathbf{V}^* \succeq \mathbf{O}$. It does not use $f(0) \leq 0$ — but it does not need to, since $f(t) = t^2$ has $f(0) = 0$.

Exercise. C2: Shrinking the matrix

Let I be an interval with $0 \in I$, let f be operator convex on I with $f(0) \leq 0$, and let $\mathbf{A} \in M_n(F)$ be Hermitian with $\text{spec}(\mathbf{A}) \subseteq I$.

- (a) Prove that $f(\theta \mathbf{A}) \preceq \theta f(\mathbf{A})$ for every $\theta \in [0, 1]$, assuming $\text{spec}(\theta \mathbf{A}) \subseteq I$.
 (b) Give an operator convex f with $f(0) > 0$ for which (a) fails, and say at which θ .
 (c) Let $\mathbf{V} \in M_{n \times 1}(F)$ be a unit column vector and $g(t) = t^2 + 1$ on $\mathbb{R}$. Write out what [Theorem 21.8.5](#) (b) says here, and verify it for $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ and $\mathbf{V} = \mathbf{e}_1$.

Hint for (a): a scalar multiple of the identity is a contraction when the scalar is small.

Solution.

- (a) Put $\mathbf{V} = \sqrt{\theta} \mathbf{I}_n$, which is Hermitian with $\mathbf{V}^* \mathbf{V} = \theta \mathbf{I}_n \preceq \mathbf{I}_n$, so $\mathbf{V}$ is a contraction ([Definition 21.8.1](#)). Then $\mathbf{V}^* \mathbf{A} \mathbf{V} = \theta \mathbf{A}$ and $\mathbf{V}^* f(\mathbf{A}) \mathbf{V} = \theta f(\mathbf{A})$. By [Theorem 21.8.5](#) (a),

$$f(\theta \mathbf{A}) = f(\mathbf{V}^* \mathbf{A} \mathbf{V}) \preceq \mathbf{V}^* f(\mathbf{A}) \mathbf{V} = \theta f(\mathbf{A}).$$

- (b) Take $f(t) = t^2 + 1$, operator convex on $\mathbb{R}$ by [Proposition 21.8.2](#) (a), with $f(0) = 1 > 0$. At $\theta = 0$ the claim reads $f(\mathbf{O}) \preceq \mathbf{O}$, that is $\mathbf{I}_n \preceq \mathbf{O}$, which is false. In fact it fails at every $\theta < 1$: the claim is $\theta^2 \mathbf{A}^2 + \mathbf{I} \preceq \theta \mathbf{A}^2 + \theta \mathbf{I}$, which rearranges to $(1 - \theta) \mathbf{I} \preceq (\theta - \theta^2) \mathbf{A}^2$, and for $\mathbf{A} = \mathbf{O}$ the right-hand side is $\mathbf{O}$ while the left-hand side is $\succ \mathbf{O}$.
 (c) A unit column $\mathbf{V}$ satisfies $\mathbf{V}^* \mathbf{V} = (1) = \mathbf{I}_1$, so it is an isometry and (b) of the theorem applies with no condition on $g(0) = 1$. It says $g(\mathbf{V}^* \mathbf{A} \mathbf{V}) \preceq \mathbf{V}^* g(\mathbf{A}) \mathbf{V}$, which for 1×1 matrices is the number inequality $(\mathbf{V}^* \mathbf{A} \mathbf{V})^2 + 1 \leq \mathbf{V}^* (\mathbf{A}^2 + \mathbf{I}) \mathbf{V}$. With $\mathbf{V} = \mathbf{e}_1$: $\mathbf{e}_1^* \mathbf{A} \mathbf{e}_1 = 2$, so the left side is 5; and $\mathbf{A}^2 = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix}$, so the right side is $5 + 1 = 6$. Indeed $5 \leq 6$.

Exercise. C3: What the trace does not see

Let $\mathbf{A} = \text{diag}(1, 0)$ and $\mathbf{B} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ be the matrices of [Example \(The Cube Is Convex and Not Operator Convex\)](#), and let $f(t) = t^3$ on $[0, \infty)$.

- (a) Compute $\operatorname{tr}(\frac{1}{2}(\mathbf{A}^3 + \mathbf{B}^3)) - \operatorname{tr}((\frac{1}{2}(\mathbf{A} + \mathbf{B}))^3)$, and say which theorem makes its sign predictable.
- (b) The same defect matrix is indefinite. Explain, in one sentence, why (a) and that fact are consistent.
- (c) Give a convex function g on $[0, \infty)$ for which the defect matrix at this pair $\mathbf{A}, \mathbf{B}$ is positive semidefinite, and name the reason.

Solution.

- (a) From [Example \(The Cube Is Convex and Not Operator Convex\)](#) the defect is $\frac{1}{8} \begin{pmatrix} 22 & 18 \\ 18 & 14 \end{pmatrix}$, whose trace is $\frac{1}{8}(22 + 14) = \frac{9}{2} > 0$. Its sign is predictable from Theorem 21.2.8: $t \mapsto t^3$ is convex on $[0, \infty)$, so $\mathbf{A} \mapsto \operatorname{tr}(\mathbf{A}^3)$ is a convex function of the Hermitian matrix, and the defect of a convex function at a midpoint is ≥ 0 .
- (b) A Hermitian 2×2 matrix can have positive trace and a negative eigenvalue; the trace only sees the **sum** of the eigenvalues, and here they are approximately 4.55 and -0.055 .
- (c) Take $g(t) = t^2$. By Proposition 17.11.8 the defect equals $\frac{1}{4}(\mathbf{A} - \mathbf{B})^2$, a nonnegative multiple of the square of a Hermitian matrix, hence $\succeq \mathbf{O}$. Explicitly $\mathbf{A} - \mathbf{B} = \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix}$, so the defect is $\frac{1}{4} \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \succeq \mathbf{O}$.