

20.5 The Hoffman–Wielandt Theorem

Section 04 showed that for a normal $\mathbf{A}$, every eigenvalue of $\mathbf{A} + \mathbf{E}$ lies within $\|\mathbf{E}\|_2$ of **some** eigenvalue of $\mathbf{A}$. That answers “how far”, one perturbed eigenvalue at a time. It does not answer “which goes where”: it never says that the eigenvalues of the two matrices can be **paired off**, each with a partner of its own. This section proves that for two normal matrices they can, with the total squared distance of the pairs bounded by $\|\mathbf{A} - \mathbf{B}\|_F^2$. The proof is short once the right tool is in hand, and the tool is Birkhoff’s theorem from Chapter 19 §07. We then sharpen the result for Hermitian matrices, where the pairing is simply the decreasing order, and show that without normality the inequality fails for every pairing.

Throughout, the field is $\mathbb{C}$, and eigenvalues are listed **with algebraic multiplicity**. For a Hermitian matrix, $\lambda_1(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$ is the decreasing list, as in Chapter 17.

20.5.1 One-sided bounds do not pair eigenvalues

Recall Corollary 20.4.3 (a): if $\mathbf{A} \in M_n(\mathbb{C})$ is normal and $\mathbf{E} \in M_n(\mathbb{C})$ is arbitrary, then every eigenvalue μ of $\mathbf{A} + \mathbf{E}$ satisfies $\min_i |\lambda_i - \mu| \leq \|\mathbf{E}\|_2$, where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $\mathbf{A}$. The statement starts from a perturbed eigenvalue and finds a partner for it. Nothing in it stops two perturbed eigenvalues from choosing the **same** partner.

Here is the smallest case. Let

$$\mathbf{A} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{A} + \mathbf{E} = \begin{pmatrix} \frac{1}{5} & 0 \\ 0 & \frac{2}{5} \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} \frac{1}{5} & 0 \\ 0 & -\frac{3}{5} \end{pmatrix}.$$

Then $\|\mathbf{E}\|_2 = \frac{3}{5}$, and both perturbed eigenvalues are within $\frac{3}{5}$ of an eigenvalue of $\mathbf{A}$, as the corollary promises. But if each one goes to its **nearest** eigenvalue of $\mathbf{A}$, both go to 0: the distances are $\frac{1}{5}$ and $\frac{2}{5}$ to 0, against $\frac{4}{5}$ and $\frac{3}{5}$ to 1. The nearest-eigenvalue map is not a bijection, and the eigenvalue 1 is left without a partner. A pairing does exist here, $0 \leftrightarrow \frac{1}{5}$ and $1 \leftrightarrow \frac{2}{5}$, with distances $\frac{1}{5}$ and $\frac{3}{5}$, but finding it was not the corollary’s doing.

Why ask for a pairing at all? Because it is what “the eigenvalues move a little” should mean. Chapter 16’s Corollary 16.7.4 is a pairing statement: the eigenvalues of a nearby matrix **can be numbered** so that each is close to its counterpart. It gives no rate. For Hermitian matrices, Weyl’s Corollary 17.3.6 gives a pairing with a rate, and the pairing is the decreasing order. For normal matrices with complex eigenvalues there is no decreasing order, and the pairing has to be found.

To say this precisely, let $\lambda_1, \dots, \lambda_n$ and $\mu_1, \dots, \mu_n$ be two lists of complex numbers. A **pairing** of them is a permutation $\sigma \in S_n$, which matches λ_i with $\mu_{\sigma(i)}$. Its cost, in the sense of this section, is the sum of squared distances

$$\sum_{i=1}^n |\lambda_i - \mu_{\sigma(i)}|^2.$$

A one-sided statement, “every μ_j is within r of some λ_i ”, is a statement about sets. A pairing is a statement about lists with multiplicity, and it is strictly stronger.

Quick check. Let $\mathbf{A} = \text{diag}(0, 0, 1)$ and $\mathbf{B} = \text{diag}(0, 1, 1)$. Show that every eigenvalue of $\mathbf{B}$ is an eigenvalue of $\mathbf{A}$ and conversely, so each one-sided distance is 0. What is the smallest possible largest distance $\max_i |\lambda_i - \mu_{\sigma(i)}|$ over all pairings σ ?

Solution. Both matrices have the eigenvalue set $\{0, 1\}$, so every eigenvalue of either one is at distance 0 from an eigenvalue of the other. As lists, $\mathbf{A}$ has $(0, 0, 1)$ and $\mathbf{B}$ has $(0, 1, 1)$. Any pairing must send one of the two 0's of $\mathbf{A}$ to a 1 of $\mathbf{B}$, since $\mathbf{B}$ has only one 0. So every pairing has largest distance at least 1, and the identity pairing attains 1. Sets that agree can still be far apart as lists.

20.5.2 Birkhoff turns weights into a permutation

The proof below will produce a sum of the form $\sum_{i,j} s_{ij} c_{ij}$, in which $c_{ij} = |\lambda_i - \mu_j|^2$ is the cost of pairing λ_i with μ_j , and the weights s_{ij} form a doubly stochastic matrix. Such a sum is a **fractional pairing**: every λ_i spreads a total weight 1 over the μ_j , and every μ_j receives a total weight 1. A genuine pairing is the special case where each weight is 0 or 1, that is, where (s_{ij}) is a permutation matrix. The lemma says that fractional pairings never beat the best genuine one.

Recall Definition 18.8.4: $\mathbf{S} \in M_n(\mathbb{R})$ is doubly stochastic if its entries are nonnegative and every row and every column adds up to 1; the set of such matrices is Ω_n . Recall also Definition 3.4.2: $\mathbf{P}_\sigma$ is the matrix whose j -th column is $\mathbf{e}_{\sigma(j)}$, so its entry in position $(\sigma(j), j)$ is 1 for each j , and all its other entries are 0.

Lemma 20.5.1. A Fractional Pairing Is Never Cheaper

Let $\mathbf{C} = (c_{ij}) \in M_n(\mathbb{R})$ and $\mathbf{S} = (s_{ij}) \in \Omega_n$. Then there is a permutation $\sigma \in S_n$ with

$$\sum_{j=1}^n c_{\sigma(j)j} \leq \sum_{i,j=1}^n s_{ij} c_{ij}.$$

Idea. Write $f(\mathbf{X}) = \sum_{i,j} x_{ij} c_{ij}$, a linear function of $\mathbf{X}$. Birkhoff's theorem writes $\mathbf{S}$ as an average of permutation matrices, so $f(\mathbf{S})$ is the same average of the numbers $f(\mathbf{P}_\sigma)$, and an average is at least its smallest term.

Proof. Let $f(\mathbf{X}) = \sum_{i,j} x_{ij} c_{ij}$ for $\mathbf{X} \in M_n(\mathbb{R})$; it is linear in $\mathbf{X}$. By Theorem 19.7.2, $\mathbf{S} = \sum_{k=1}^m t_k \mathbf{P}_{\sigma_k}$ for some permutations $\sigma_1, \dots, \sigma_m$ and weights $t_k \geq 0$ with $\sum_k t_k = 1$. By linearity,

$$f(\mathbf{S}) = \sum_{k=1}^m t_k f(\mathbf{P}_{\sigma_k}) \geq \left(\sum_{k=1}^m t_k \right) \min_k f(\mathbf{P}_{\sigma_k}) = \min_k f(\mathbf{P}_{\sigma_k}),$$

where the inequality uses $t_k \geq 0$. Let σ be a σ_k attaining the minimum. The only nonzero entries of $\mathbf{P}_\sigma$ are the 1's in positions $(\sigma(j), j)$, so $f(\mathbf{P}_\sigma) = \sum_j c_{\sigma(j)j}$. Hence $\sum_j c_{\sigma(j)j} \leq f(\mathbf{S})$, as claimed. ■

In the language of Chapter 18, this is Corollary 18.8.6 applied to Ω_n . With the Frobenius inner product $\langle \mathbf{X}, \mathbf{C} \rangle = \text{tr}(\mathbf{C}^\top \mathbf{X}) = \sum_{i,j} x_{ij} c_{ij}$, the function f is $\mathbf{X} \mapsto \langle \mathbf{X}, \mathbf{C} \rangle$. The set Ω_n is nonempty, compact, and convex (Chapter 18 §08), so the minimum of f over it is attained at an extreme point, and by Theorem 19.7.2 the extreme points are the permutation matrices. The

proof above uses the second form of Birkhoff's theorem instead, so no compactness argument appears in it.

20.5.3 The theorem

We can now prove that two normal matrices have a pairing of their eigenvalues whose cost is at most the squared Frobenius distance between the matrices.

Theorem 20.5.2. Hoffman–Wielandt Theorem

Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$ be **normal**, with eigenvalues $\lambda_1, \dots, \lambda_n$ and $\mu_1, \dots, \mu_n$ respectively, each listed with algebraic multiplicity in any order. Then there is a permutation $\sigma \in S_n$ with

$$\sum_{i=1}^n |\lambda_i - \mu_{\sigma(i)}|^2 \leq \|\mathbf{A} - \mathbf{B}\|_F^2.$$

Idea. Step roadmap.

① **Diagonalize both.** The spectral theorem gives $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$ and $\mathbf{B} = \mathbf{V}\mathbf{M}\mathbf{V}^*$ with $\mathbf{U}, \mathbf{V}$ unitary and $\mathbf{\Lambda}, \mathbf{M}$ diagonal. The two orthonormal bases of eigenvectors are different, and the unitary $\mathbf{W} = \mathbf{U}^*\mathbf{V}$ records how they sit relative to each other.

② **Compute the distance.** The Frobenius norm does not see unitary factors, so $\|\mathbf{A} - \mathbf{B}\|_F = \|\mathbf{\Lambda}\mathbf{W} - \mathbf{W}\mathbf{M}\|_F$. The (i, j) entry of $\mathbf{\Lambda}\mathbf{W} - \mathbf{W}\mathbf{M}$ is $(\lambda_i - \mu_j)w_{ij}$, so $\|\mathbf{A} - \mathbf{B}\|_F^2 = \sum_{i,j} |w_{ij}|^2 |\lambda_i - \mu_j|^2$. This is a fractional pairing with weights $|w_{ij}|^2$.

③ **Round it.** The weights form a doubly stochastic matrix, because the rows and the columns of a unitary matrix are unit vectors. The lemma replaces the fractional pairing by a genuine one that costs no more.

Proof. By Corollary 12.4.3, since $\mathbf{A}$ and $\mathbf{B}$ are normal, there are unitary $\mathbf{U}, \mathbf{V} \in U(n)$ with

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*, \quad \mathbf{B} = \mathbf{V}\mathbf{M}\mathbf{V}^*,$$

where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$ and $\mathbf{M} = \text{diag}(\mu_1, \dots, \mu_n)$ in the given orders. Put $\mathbf{W} = \mathbf{U}^*\mathbf{V}$, which is unitary as a product of unitary matrices. Since $\mathbf{U}^*\mathbf{U} = \mathbf{I}$ and $\mathbf{V}^*\mathbf{V} = \mathbf{I}$,

$$\mathbf{U}^*(\mathbf{A} - \mathbf{B})\mathbf{V} = \mathbf{\Lambda}\mathbf{U}^*\mathbf{V} - \mathbf{U}^*\mathbf{V}\mathbf{M} = \mathbf{\Lambda}\mathbf{W} - \mathbf{W}\mathbf{M}.$$

By Lemma 11.7.5, applied with the unitary factors $\mathbf{U}^*$ on the left and $\mathbf{V}$ on the right, $\|\mathbf{A} - \mathbf{B}\|_F = \|\mathbf{\Lambda}\mathbf{W} - \mathbf{W}\mathbf{M}\|_F$. The (i, j) entry of $\mathbf{\Lambda}\mathbf{W}$ is $\lambda_i w_{ij}$ and that of $\mathbf{W}\mathbf{M}$ is $w_{ij} \mu_j$, since multiplying by a diagonal matrix on the left scales rows and on the right scales columns. Hence

$$\|\mathbf{A} - \mathbf{B}\|_F^2 = \sum_{i,j=1}^n |w_{ij}|^2 |\lambda_i - \mu_j|^2. \quad (20.5.1)$$

Let $\mathbf{S} = (s_{ij})$ with $s_{ij} = |w_{ij}|^2$. Its entries are nonnegative. Row i adds up to $\sum_j |w_{ij}|^2 = (\mathbf{W}\mathbf{W}^*)_{ii} = 1$, and column j adds up to $\sum_i |w_{ij}|^2 = (\mathbf{W}^*\mathbf{W})_{jj} = 1$, since $\mathbf{W}$ is unitary. So

$\mathbf{S} \in \Omega_n$. Apply [Lemma 20.5.1](#) to $\mathbf{S}$ and the real matrix $\mathbf{C}$ with $c_{ij} = |\lambda_i - \mu_j|^2$. It gives $\tau \in S_n$ with

$$\sum_{j=1}^n |\lambda_{\tau(j)} - \mu_j|^2 \leq \sum_{i,j} s_{ij} c_{ij} = \|\mathbf{A} - \mathbf{B}\|_F^2,$$

the equality being [Eq. 20.5.1](#). Put $\sigma = \tau^{-1}$. As j runs over $1, \dots, n$, so does $i = \tau(j)$, and $j = \sigma(i)$; so the left-hand side is $\sum_i |\lambda_i - \mu_{\sigma(i)}|^2$. This proves the theorem. ■

Both hypotheses were used in the first line and nowhere else: normality is what makes the diagonalizing matrices unitary. The rest of the proof is bookkeeping and Birkhoff's theorem, which is where Chapter 19 §07's theorem earns its keep.

In words: *the eigenvalues of a normal matrix, as a list with multiplicity, are a 1-Lipschitz function of the matrix, when matrices are measured in the Frobenius norm and lists by their best pairing.* In particular every pair in the best pairing is within $\|\mathbf{A} - \mathbf{B}\|_F$, so each eigenvalue of $\mathbf{B}$ has its own partner among those of $\mathbf{A}$ within that distance. This is a matching bound, with a constant, valid for all normal matrices at once. Compare Corollary 16.7.4, which gives a matching with no constant for all matrices, and Corollary 20.4.3, which gives a constant with no matching.

The theorem asserts that **some** pairing is cheap. The proof does not say which, beyond what the lemma gives, but the cheapest of the $n!$ pairings certainly works, since it costs no more than any other. The next example shows that a pairing chosen by any other rule can fail.

Example. A Hermitian and a diagonal matrix

Let

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

Both are normal. Compute $\|\mathbf{A} - \mathbf{B}\|_F^2$, the cost of both pairings of the eigenvalues, and the matrix $\mathbf{S}$ of the proof. Which pairings satisfy the conclusion of [Theorem 20.5.2](#)?

Solution. $\mathbf{A}$ is real symmetric, hence Hermitian and normal, and $\mathbf{B}$ is diagonal, hence normal. The eigenvalues of $\mathbf{A}$ are $\lambda_1 = 3$ and $\lambda_2 = 1$, with unit eigenvectors $\frac{1}{\sqrt{2}}(1, 1)$ and $\frac{1}{\sqrt{2}}(1, -1)$: indeed $\mathbf{A}(1, 1) = (3, 3)$ and $\mathbf{A}(1, -1) = (1, -1)$. Those of $\mathbf{B}$ are $\mu_1 = i$ and $\mu_2 = 2 + i$, with eigenvectors $\mathbf{e}_1, \mathbf{e}_2$.

The distance. $\mathbf{A} - \mathbf{B} = \begin{pmatrix} 2-i & 1 \\ 1 & -i \end{pmatrix}$, so

$$\|\mathbf{A} - \mathbf{B}\|_F^2 = |2-i|^2 + 1 + 1 + |-i|^2 = 5 + 1 + 1 + 1 = 8.$$

The costs. The four squared distances $c_{ij} = |\lambda_i - \mu_j|^2$ are

$$\begin{aligned} c_{11} &= |3 - i|^2 = 10, & c_{12} &= |1 - i|^2 = 2, \\ c_{21} &= |1 - i|^2 = 2, & c_{22} &= |-1 - i|^2 = 2. \end{aligned}$$

The identity pairing $3 \leftrightarrow i, 1 \leftrightarrow 2 + i$ costs $c_{11} + c_{22} = 12$. The swap $3 \leftrightarrow 2 + i, 1 \leftrightarrow i$ costs $c_{12} + c_{21} = 4$.

The weights. With $\mathbf{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ and $\mathbf{V} = \mathbf{I}$, $\mathbf{W} = \mathbf{U}^*$, and every entry of $\mathbf{W}$ has modulus $\frac{1}{\sqrt{2}}$. So $\mathbf{S} = \frac{1}{2}\mathbf{J}$, and Eq. 20.5.1 reads

$$\|\mathbf{A} - \mathbf{B}\|_F^2 = \frac{1}{2}(10 + 2 + 2 + 2) = 8,$$

the average of the two costs, $\frac{1}{2}(12 + 4)$. An average lies between its terms, so the cheaper pairing, the swap with cost 4, satisfies $4 \leq 8$. The identity pairing, with cost $12 > 8$, does not.

Warning. **The theorem does not say that the given numbering works.** In the example, the lists $(3, 1)$ and $(i, 2 + i)$, paired in the order written, cost $12 > 8 = \|\mathbf{A} - \mathbf{B}\|_F^2$. Complex numbers have no natural order to sort by, so, unlike the Hermitian case below, there is no rule fixed in advance that always picks a good pairing. The safe reading is: the **cheapest** pairing satisfies the inequality.

20.5.4 Normality cannot be dropped

The proof used normality once, in its first line, to make the diagonalizing matrices unitary. For diagonalizable matrices $\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}$ and $\mathbf{B} = \mathbf{Y}\mathbf{M}\mathbf{Y}^{-1}$ with $\mathbf{X}$, $\mathbf{Y}$ not unitary, Lemma 11.7.5 no longer applies, and $\|\mathbf{A} - \mathbf{B}\|_F$ is no longer the norm of $\mathbf{\Lambda}\mathbf{W} - \mathbf{W}\mathbf{M}$ with $\mathbf{W} = \mathbf{X}^{-1}\mathbf{Y}$. The failure is real, not an artifact of the proof.

Example. Every pairing fails

Let

$$\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 3 \\ 1 & -1 \end{pmatrix}.$$

Show that $\mathbf{A}$ and $\mathbf{B}$ are diagonalizable but not normal, and that the inequality of Theorem 20.5.2 fails for **every** pairing of their eigenvalues.

Solution. $\mathbf{A}$ is triangular with eigenvalues 1 and -1 . The matrix $\mathbf{B}$ has trace 0 and determinant $-1 - 3 = -4$, so its characteristic polynomial is $x^2 - 4$ and its eigenvalues are 2 and -2 . Each has two distinct eigenvalues, so each is diagonalizable (Theorem 9.4.1). Neither is normal:

$$\begin{aligned} \mathbf{A}^*\mathbf{A} &= \begin{pmatrix} 1 & 3 \\ 3 & 10 \end{pmatrix} \neq \begin{pmatrix} 10 & -3 \\ -3 & 1 \end{pmatrix} = \mathbf{A}\mathbf{A}^*, \\ \mathbf{B}^*\mathbf{B} &= \begin{pmatrix} 2 & 2 \\ 2 & 10 \end{pmatrix} \neq \begin{pmatrix} 10 & -2 \\ -2 & 2 \end{pmatrix} = \mathbf{B}\mathbf{B}^*. \end{aligned}$$

The difference $\mathbf{A} - \mathbf{B}$ has a single nonzero entry, -1 in position $(2, 1)$, so $\|\mathbf{A} - \mathbf{B}\|_F^2 = 1$.

There are two pairings. The one that matches signs, $1 \leftrightarrow 2$ and $-1 \leftrightarrow -2$, costs $1^2 + 1^2 = 2$. The other, $1 \leftrightarrow -2$ and $-1 \leftrightarrow 2$, costs $3^2 + 3^2 = 18$. Both exceed 1, so no pairing satisfies the inequality.

The entry 3 is what makes this work, and the failure is not confined to a perturbation as large

as 1. Put $\delta > 0$ in the corner instead: $\mathbf{B}_\delta = \begin{pmatrix} 1 & 3 \\ \delta & -1 \end{pmatrix}$ has determinant $-1 - 3\delta$ and eigenvalues $\pm\sqrt{1+3\delta}$, each of which moves by $\sqrt{1+3\delta} - 1$, about $\frac{3}{2}\delta$ for small δ . So the cheaper pairing costs about $2(\frac{3}{2}\delta)^2 = \frac{9}{2}\delta^2$, well above $\|\mathbf{A} - \mathbf{B}_\delta\|_F^2 = \delta^2$. With an entry m in place of 3 the cost is about $\frac{1}{2}m^2\delta^2$, so the inequality fails for small δ as soon as $m > \sqrt{2}$. A large entry above the diagonal is exactly what Chapter 12 §11 measured as departure from normality, and it is what makes the eigenvalues sensitive.

A Jordan block shows that no constant can repair the inequality. Let $\mathbf{A} = \mathbf{J}_2(0)$ and, for $\varepsilon > 0$, $\mathbf{B}_\varepsilon = \begin{pmatrix} 0 & 1 \\ \varepsilon & 0 \end{pmatrix}$. Then $\mathbf{A}$ has the eigenvalue 0 twice, $\mathbf{B}_\varepsilon$ has characteristic polynomial $x^2 - \varepsilon$ and eigenvalues $\pm\sqrt{\varepsilon}$, and $\|\mathbf{A} - \mathbf{B}_\varepsilon\|_F^2 = \varepsilon^2$. Both pairings cost the same, $(\sqrt{\varepsilon})^2 + (\sqrt{\varepsilon})^2 = 2\varepsilon$, and

$$\frac{2\varepsilon}{\varepsilon^2} = \frac{2}{\varepsilon} \longrightarrow \infty \quad (\varepsilon \rightarrow 0^+).$$

So there is no constant K with $\min_\sigma \sum_i |\lambda_i - \mu_{\sigma(i)}|^2 \leq K\|\mathbf{A} - \mathbf{B}\|_F^2$ for all pairs of matrices. This is Example (A Jordan block's eigenvalues move like $\varepsilon^{1/k}$) of Chapter 16 read in the language of pairings: the eigenvalues move like $\sqrt{\varepsilon}$ while the matrix moves by ε .

Warning. **Diagonalizable is not enough, and one normal matrix is not enough.** In [Example \(Every pairing fails\)](#) both matrices are diagonalizable, and every pairing fails. Nor does it help to make just one of the two matrices normal: for $\mathbf{A} = \mathbf{J}_2(0)$ and the Hermitian $\mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, with eigenvalues ± 1 , both pairings cost $1 + 1 = 2$, while $\|\mathbf{A} - \mathbf{B}\|_F^2 = 1$. The theorem needs **both** matrices normal.

20.5.5 Hermitian matrices: the pairing is the order

When both matrices are Hermitian, the eigenvalues are real, and there is a natural guess for the best pairing: largest with largest, second with second, and so on. The guess is right, and the reason is a fact about real numbers alone.

Lemma 20.5.3. The Sorted Pairing Is Cheapest

Let $a_1 \geq a_2 \geq \cdots \geq a_n$ and $b_1 \geq b_2 \geq \cdots \geq b_n$ be real numbers. Then for every $\sigma \in S_n$,

$$\sum_{i=1}^n (a_i - b_i)^2 \leq \sum_{i=1}^n (a_i - b_{\sigma(i)})^2.$$

Idea. Expanding the squares, $\sum_i a_i^2$ and $\sum_i b_{\sigma(i)}^2 = \sum_j b_j^2$ do not depend on σ , so the claim is that the cross term $\sum_i a_i b_{\sigma(i)}$ is largest at the identity. If σ puts a smaller b before a larger one in two neighboring positions, swapping those two values can only increase the cross term, and it removes one inversion. Repeating the swap sorts σ into the identity.

Proof. For every σ , since σ is a bijection, $\sum_i b_{\sigma(i)}^2 = \sum_j b_j^2$, and so

$$\begin{aligned}\sum_{i=1}^n (a_i - b_{\sigma(i)})^2 &= \sum_i a_i^2 + \sum_j b_j^2 - 2g(\sigma), \\ g(\sigma) &= \sum_{i=1}^n a_i b_{\sigma(i)}.\end{aligned}$$

It therefore suffices to show $g(\sigma) \leq g(\text{id})$ for every σ . The set S_n is finite, so g attains a maximum on it; among the permutations where it does, choose one, σ , with the fewest inversions (Definition 7.3.1). Suppose $\sigma \neq \text{id}$. Then $\sigma(i) > \sigma(i+1)$ for some i , since a permutation with $\sigma(1) < \sigma(2) < \cdots < \sigma(n)$ is the identity. Let σ' agree with σ except that $\sigma'(i) = \sigma(i+1)$ and $\sigma'(i+1) = \sigma(i)$. Then

$$g(\sigma') - g(\sigma) = (a_i - a_{i+1})(b_{\sigma(i+1)} - b_{\sigma(i)}) \geq 0,$$

because $a_i \geq a_{i+1}$, and $\sigma(i+1) < \sigma(i)$ gives $b_{\sigma(i+1)} \geq b_{\sigma(i)}$. So σ' also maximizes g . Now count inversions. The pair of positions $(i, i+1)$ is an inversion of σ and not of σ' . A pair involving neither i nor $i+1$ sees the same values under both. For a position $p \notin \{i, i+1\}$, the two pairs joining p to i and to $i+1$ compare $\sigma(p)$ with the same two values $\sigma(i), \sigma(i+1)$ under σ and under σ' , from the same side, so together they contain the same number of inversions. Hence σ' has one inversion fewer than σ , contradicting the choice of σ . So $\sigma = \text{id}$ maximizes g , and $g(\sigma) \leq g(\text{id})$ for every σ . This proves the lemma. ■

The move is an **exchange argument**: to show that an arrangement is optimal, show that repairing one local defect never makes things worse. The inequality $\sum_i a_i b_{\sigma(i)} \leq \sum_i a_i b_i$ is known as the rearrangement inequality.

Example. Three numbers, six pairings

Let $\mathbf{a} = (4, 2, 1)$ and $\mathbf{b} = (3, 0, -1)$, both decreasing. Compute $\sum_i (a_i - b_{\sigma(i)})^2$ for all six $\sigma \in S_3$, and find the cheapest and the dearest.

Solution. The nine squared distances $(a_i - b_j)^2$, with i the row and j the column, are

$$\begin{pmatrix} 1 & 16 & 25 \\ 1 & 4 & 9 \\ 4 & 1 & 4 \end{pmatrix}.$$

A pairing picks one entry from each row and each column. Writing σ by its values $(\sigma(1), \sigma(2), \sigma(3))$,

$$\begin{aligned}(1, 2, 3) : 1 + 4 + 4 &= 9, & (1, 3, 2) : 1 + 9 + 1 &= 11, \\ (2, 1, 3) : 16 + 1 + 4 &= 21, & (2, 3, 1) : 16 + 9 + 4 &= 29, \\ (3, 1, 2) : 25 + 1 + 1 &= 27, & (3, 2, 1) : 25 + 4 + 4 &= 33.\end{aligned}$$

The identity, which pairs the two sorted lists in order, costs 9, the least of the six, as [Lemma 20.5.3](#) says. The dearest is the reversed pairing $(3, 2, 1)$, at 33.

One exchange step of the proof is visible here. The permutation $(1, 3, 2)$ has the single inversion $\sigma(2) > \sigma(3)$; undoing it gives the identity and lowers the cost by 2, which is $2(a_2 - a_3)(b_2 - b_3) = 2 \cdot 1 \cdot 1$.

Corollary 20.5.4. Hoffman–Wielandt for Hermitian Matrices

Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$ be Hermitian. Then

$$\sum_{i=1}^n (\lambda_i(\mathbf{A}) - \lambda_i(\mathbf{B}))^2 \leq \|\mathbf{A} - \mathbf{B}\|_F^2.$$

Proof. Hermitian matrices are normal (Proposition 11.6.7 (c)) and have real eigenvalues. By Theorem 20.5.2 applied to the lists $\lambda_1(\mathbf{A}), \dots, \lambda_n(\mathbf{A})$ and $\lambda_1(\mathbf{B}), \dots, \lambda_n(\mathbf{B})$, there is $\sigma \in S_n$ with $\sum_i (\lambda_i(\mathbf{A}) - \lambda_{\sigma(i)}(\mathbf{B}))^2 \leq \|\mathbf{A} - \mathbf{B}\|_F^2$, where $|t|^2 = t^2$ for real t . Both lists are decreasing, so by Lemma 20.5.3 the sorted pairing costs no more than σ . This proves the corollary. ■

This is the Frobenius-norm companion of Weyl’s bound. Put $\mathbf{E} = \mathbf{A} - \mathbf{B}$, Hermitian. By the spectral theorem $\mathbf{E} = \mathbf{U}\mathbf{D}\mathbf{U}^*$ with $\mathbf{U}$ unitary and $\mathbf{D} = \text{diag}(\lambda_1(\mathbf{E}), \dots, \lambda_n(\mathbf{E}))$, so Lemma 11.7.5 gives $\|\mathbf{E}\|_F^2 = \|\mathbf{D}\|_F^2 = \sum_i \lambda_i(\mathbf{E})^2$, while $\|\mathbf{E}\|_2 = \max_i |\lambda_i(\mathbf{E})|$ by Lemma 17.3.5. So Weyl’s Corollary 17.3.6 says

$$\max_i |\lambda_i(\mathbf{A}) - \lambda_i(\mathbf{B})| \leq \max_i |\lambda_i(\mathbf{E})|,$$

and Corollary 20.5.4 says

$$\sum_i (\lambda_i(\mathbf{A}) - \lambda_i(\mathbf{B}))^2 \leq \sum_i \lambda_i(\mathbf{E})^2.$$

In both, a norm of the vector of shifts $\lambda(\mathbf{A}) - \lambda(\mathbf{B})$ is bounded by the same norm of $\lambda(\mathbf{E})$: the ∞ -norm for Weyl, the 2-norm for Hoffman–Wielandt.

Neither implies the other. For $\mathbf{E} = t\mathbf{I}$ with $t > 0$, every eigenvalue shifts by t ; Weyl’s bound t is exact for each shift, while Hoffman–Wielandt alone allows a single shift as large as $\|\mathbf{E}\|_F = \sqrt{n}t$. For a rank-one $\mathbf{E} = t\mathbf{x}\mathbf{x}^*$ with $\|\mathbf{x}\|_2 = 1$, $\|\mathbf{E}\|_F = \|\mathbf{E}\|_2 = t$; Weyl lets **every** eigenvalue move by up to t , while Hoffman–Wielandt says that the squared shifts together use up at most t^2 , so if one eigenvalue moves by t , none of the others moves at all.

Remark. Lidskii’s inequality contains both. Chapter 17 §07 proved $\lambda(\mathbf{A}) - \lambda(\mathbf{B}) \prec \lambda(\mathbf{A} - \mathbf{B})$ (the remark after Theorem 17.7.2), and Weyl’s bound follows from it there, as Corollary 17.7.3 (a). Chapter 21 shows that a majorization $\mathbf{x} \prec \mathbf{y}$ gives $\sum_i \phi(x_i) \leq \sum_i \phi(y_i)$ for every convex ϕ ; with $\phi(t) = t^2$ this is Corollary 20.5.4 again, and Chapter 21 turns the same majorization into one bound for every unitarily invariant norm. Nothing here depends on that route; the proof above goes through Birkhoff’s theorem, Theorem 19.7.2.

Quick check. Let $\mathbf{A} = \text{diag}(3, 1)$ and $\mathbf{B} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. Compute $\|\mathbf{A} - \mathbf{B}\|_F^2$ and both sides of Corollary 20.5.4. What does the reversed pairing $\lambda_1(\mathbf{A}) \leftrightarrow \lambda_2(\mathbf{B})$, $\lambda_2(\mathbf{A}) \leftrightarrow \lambda_1(\mathbf{B})$ cost?

Solution. $\mathbf{A} - \mathbf{B} = \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}$, so $\|\mathbf{A} - \mathbf{B}\|_F^2 = 4$. The eigenvalues of $\mathbf{B}$ are 3 and 1 (as in [Example \(A Hermitian and a diagonal matrix\)](#)), the same as those of $\mathbf{A}$. So the left side of the corollary is $0 \leq 4$: two matrices at Frobenius distance 2 can have identical spectra, and the inequality is only an upper bound. The reversed pairing costs $(3-1)^2 + (1-3)^2 = 8 > 4$. Even for Hermitian matrices, the pairing matters.

20.5.6 Exercises

20.5.6.1 A. Check your understanding

Exercise. A1

- State the Hoffman–Wielandt theorem, with its hypotheses.
- True or false: if $\mathbf{A}$ and $\mathbf{B}$ are normal, the eigenvalues of $\mathbf{B}$ listed in any order satisfy $\sum_i |\lambda_i - \mu_i|^2 \leq \|\mathbf{A} - \mathbf{B}\|_F^2$. Justify your answer.
- True or false: the Hoffman–Wielandt inequality holds for every pair of diagonalizable matrices. Justify your answer.
- For normal $\mathbf{A}$ and $\mathbf{B}$, what bound does the theorem give on the distance between an eigenvalue of $\mathbf{B}$ and its partner?
- Where exactly does Birkhoff’s theorem (Theorem 19.7.2) enter the proof, and what would be missing without it?
- For Hermitian $\mathbf{A}, \mathbf{B}$ with $\mathbf{E} = \mathbf{A} - \mathbf{B}$ of rank one, explain why at most one eigenvalue can move by the full amount $\|\mathbf{E}\|_2$.

Solution.

- If $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$ are normal, with eigenvalues $\lambda_1, \dots, \lambda_n$ and $\mu_1, \dots, \mu_n$ listed with multiplicity, then there is a permutation σ with $\sum_i |\lambda_i - \mu_{\sigma(i)}|^2 \leq \|\mathbf{A} - \mathbf{B}\|_F^2$.
- False. In [Example \(A Hermitian and a diagonal matrix\)](#), the order (3, 1) and $(i, 2+i)$ costs 12 > 8. Only a suitable pairing, for instance the cheapest one, is guaranteed.
- False. In [Example \(Every pairing fails\)](#) both matrices have two distinct eigenvalues, hence are diagonalizable, and both pairings cost more than $\|\mathbf{A} - \mathbf{B}\|_F^2 = 1$.
- Each summand is at most the whole sum, so $|\lambda_i - \mu_{\sigma(i)}| \leq \|\mathbf{A} - \mathbf{B}\|_F$ for every i , for the pairing σ of the theorem.
- The proof writes $\|\mathbf{A} - \mathbf{B}\|_F^2$ as $\sum_{i,j} s_{ij} |\lambda_i - \mu_j|^2$ with (s_{ij}) doubly stochastic, a fractional pairing. Birkhoff’s theorem, through [Lemma 20.5.1](#), replaces it by a genuine pairing that costs no more. Without it, we would know only that a weighted average of the costs $|\lambda_i - \mu_j|^2$ is small, which does not produce a permutation.
- For Hermitian $\mathbf{E}$ of rank one, $\mathbf{E}$ has at most one nonzero eigenvalue, so $\|\mathbf{E}\|_F^2 = \sum_i \lambda_i(\mathbf{E})^2 = \|\mathbf{E}\|_2^2$. By [Corollary 20.5.4](#) the squared shifts add up to at most $\|\mathbf{E}\|_2^2$. If one shift equals $\|\mathbf{E}\|_2$ in absolute value, the others are 0.

20.5.6.2 B. Practice

Exercise. B1: A pair of normal matrices

Let

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

Check that both are normal, compute $\|\mathbf{A} - \mathbf{B}\|_F^2$ and the costs of both pairings of the eigenvalues, and compute the matrix $\mathbf{S}$ of the proof of [Theorem 20.5.2](#). Hence decide which pairings satisfy the Hoffman–Wielandt inequality.

Solution. $\mathbf{A}$ is real symmetric, hence normal, and $\mathbf{B}$ is diagonal, hence normal. $\mathbf{A}(1, 1) = (2, 2)$ and $\mathbf{A}(1, -1) = (-2, 2)$, so $\mathbf{A}$ has eigenvalues $\lambda_1 = 2$, $\lambda_2 = -2$ with unit eigenvectors $\frac{1}{\sqrt{2}}(1, \pm 1)$. $\mathbf{B}$ has eigenvalues $\mu_1 = 1$, $\mu_2 = -1 + 2i$.

$\mathbf{A} - \mathbf{B} = \begin{pmatrix} -1 & 2 \\ 2 & 1 - 2i \end{pmatrix}$, so $\|\mathbf{A} - \mathbf{B}\|_F^2 = 1 + 4 + 4 + 5 = 14$.

The squared distances are $c_{11} = |2 - 1|^2 = 1$, $c_{12} = |3 - 2i|^2 = 13$, $c_{21} = |-3|^2 = 9$, $c_{22} = |-1 - 2i|^2 = 5$. The identity pairing costs $1 + 5 = 6$, and the swap costs $13 + 9 = 22$.

With $\mathbf{U} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ and $\mathbf{V} = \mathbf{I}$, $\mathbf{W} = \mathbf{U}^*$ has all entries of modulus $\frac{1}{\sqrt{2}}$, so $\mathbf{S} = \frac{1}{2}\mathbf{J}$, and indeed $\frac{1}{2}(1 + 13 + 9 + 5) = 14$. Hence the identity pairing satisfies the inequality, $6 \leq 14$, and the swap does not, $22 > 14$.

Exercise. B2: Hermitian, compared with Weyl

Let

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

Compute $\lambda(\mathbf{A})$, $\lambda(\mathbf{B})$, $\|\mathbf{A} - \mathbf{B}\|_F$, and $\|\mathbf{A} - \mathbf{B}\|_2$. Verify [Corollary 20.5.4](#) and [Corollary 17.3.6](#) for this pair. Hence say which of the two bounds is closer to the truth here.

Solution. $\lambda(\mathbf{A}) = (2, 1, 0)$. The top-left block of $\mathbf{B}$ has eigenvalues 3 and 1 (with eigenvectors $(1, 1)$ and $(1, -1)$), and the last entry is -1 , so $\lambda(\mathbf{B}) = (3, 1, -1)$. The shifts are $\lambda_i(\mathbf{B}) - \lambda_i(\mathbf{A}) = 1, 0, -1$.

$\mathbf{E} = \mathbf{B} - \mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$, so $\|\mathbf{E}\|_F^2 = 1 + 1 + 1 + 1 = 4$ and $\|\mathbf{E}\|_F = 2$. Its top-left block has characteristic polynomial $x^2 - x - 1$, with roots $\frac{1 \pm \sqrt{5}}{2}$, and its last eigenvalue is -1 . By [Lemma 17.3.5](#), $\|\mathbf{E}\|_2 = \frac{1 + \sqrt{5}}{2} \approx 1.618$.

Hoffman–Wielandt: $1^2 + 0^2 + (-1)^2 = 2 \leq 4$. Weyl: each $|\text{shift}| \leq 1 \leq 1.618$. Hence both hold. Weyl overestimates the largest shift by a factor of about 1.6; Hoffman–Wielandt overestimates the sum of squares by a factor of 2. Neither is sharp here, and neither dominates.

Exercise. B3: Sorting

Let $\mathbf{a} = (5, 2, -1)$ and $\mathbf{b} = (4, 3, 0)$. Compute $\sum_i (a_i - b_{\sigma(i)})^2$ for all six $\sigma \in S_3$. Hence confirm [Lemma 20.5.3](#), and say which pairing is the most expensive.

Solution. Write σ by its values $(\sigma(1), \sigma(2), \sigma(3))$. Then

$$\begin{aligned} (1, 2, 3) : 1 + 1 + 1 &= 3, & (1, 3, 2) : 1 + 4 + 16 &= 21, \\ (2, 1, 3) : 4 + 4 + 1 &= 9, & (2, 3, 1) : 4 + 4 + 25 &= 33, \\ (3, 1, 2) : 25 + 4 + 16 &= 45, & (3, 2, 1) : 25 + 1 + 25 &= 51. \end{aligned}$$

The identity, which pairs the sorted lists in order, costs 3, the least of the six, as the lemma says. The most expensive is $(3, 2, 1)$, which pairs the largest a with the smallest b and so on: the reversed order. Exercise C1 (b) shows this is no accident.

20.5.6.3 C. Going deeper

Exercise. C1: The most expensive pairing

- (a) Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$ be normal, with eigenvalues λ_i and μ_i . Prove that there is also a permutation σ with $\sum_i |\lambda_i - \mu_{\sigma(i)}|^2 \geq \|\mathbf{A} - \mathbf{B}\|_F^2$.
 (b) Deduce that for Hermitian $\mathbf{A}, \mathbf{B}$,

$$\sum_{i=1}^n (\lambda_i(\mathbf{A}) - \lambda_{n+1-i}(\mathbf{B}))^2 \geq \|\mathbf{A} - \mathbf{B}\|_F^2.$$

Hint for (b): apply [Lemma 20.5.3](#) to the list $-\lambda_{n+1-i}(\mathbf{B})$.

Solution.

- (a) Apply [Lemma 20.5.1](#) to the same $\mathbf{S} \in \Omega_n$ as in the proof of [Theorem 20.5.2](#) and to the matrix $-\mathbf{C}$, where $c_{ij} = |\lambda_i - \mu_j|^2$. It gives τ with $-\sum_j c_{\tau(j)j} \leq -\sum_{i,j} s_{ij} c_{ij} = -\|\mathbf{A} - \mathbf{B}\|_F^2$, by [Eq. 20.5.1](#). Multiplying by -1 and putting $\sigma = \tau^{-1}$, as in that proof, gives $\sum_i |\lambda_i - \mu_{\sigma(i)}|^2 \geq \|\mathbf{A} - \mathbf{B}\|_F^2$.
 (b) Write $a_i = \lambda_i(\mathbf{A})$ and $c_i = \lambda_i(\mathbf{B})$, and let $b_i = -c_{n+1-i}$. Both (a_i) and (b_i) are decreasing, the second because c_{n+1-i} increases with i . Let $\pi \in S_n$, and let $\sigma(i) = n + 1 - \pi(i)$, which is again a permutation, with $b_{\sigma(i)} = -c_{\pi(i)}$. By [Lemma 20.5.3](#), $\sum_i (a_i - b_i)^2 \leq \sum_i (a_i - b_{\sigma(i)})^2$, that is,

$$\sum_i (a_i + c_{n+1-i})^2 \leq \sum_i (a_i + c_{\pi(i)})^2.$$

Expanding both sides, the squares $\sum_i a_i^2$ and $\sum_j c_j^2$ cancel, and what remains is $\sum_i a_i c_{n+1-i} \leq \sum_i a_i c_{\pi(i)}$. Hence

$$\sum_i (a_i - c_{\pi(i)})^2 = \sum_i a_i^2 + \sum_j c_j^2 - 2 \sum_i a_i c_{\pi(i)} \leq \sum_i (a_i - c_{n+1-i})^2.$$

So the reversed pairing $i \mapsto n + 1 - i$ is the most expensive of all. By (a) some pairing costs at least $\|\mathbf{A} - \mathbf{B}\|_F^2$, so the most expensive one does too, which is the claim.

Exercise. C2: Each eigenvalue keeps its own disc

Let $\mathbf{A} \in M_n(\mathbb{C})$ be normal with **distinct** eigenvalues $\lambda_1, \dots, \lambda_n$, and let $\delta = \min_{i \neq j} |\lambda_i - \lambda_j|$. Let $\mathbf{B}$ be normal with $\|\mathbf{A} - \mathbf{B}\|_F < \delta/2$. Prove that each open disc $\{z : |z - \lambda_i| < \delta/2\}$ contains **exactly one** eigenvalue of $\mathbf{B}$, counted with multiplicity. Explain why Corollary 20.4.3 alone does not give this.

Solution. The discs are pairwise disjoint: if $|z - \lambda_i| < \delta/2$ and $|z - \lambda_j| < \delta/2$ with $i \neq j$, the triangle inequality (Theorem 0.6.4) gives $|\lambda_i - \lambda_j| < \delta$, contradicting the definition of δ . By Theorem 20.5.2 there is σ with $\sum_i |\lambda_i - \mu_{\sigma(i)}|^2 \leq \|\mathbf{A} - \mathbf{B}\|_F^2 < \delta^2/4$, so $|\lambda_i - \mu_{\sigma(i)}| < \delta/2$ for every i . Thus the entry $\mu_{\sigma(i)}$ of the list lies in the i -th disc, and, the discs being disjoint, in no other. Since σ is a bijection, every entry of the list $\mu_1, \dots, \mu_n$ is $\mu_{\sigma(i)}$ for exactly one i . So the i -th disc contains exactly one entry of the list, namely $\mu_{\sigma(i)}$, which is the claim.

Corollary 20.4.3 (a), with $\|\mathbf{A} - \mathbf{B}\|_2 \leq \|\mathbf{A} - \mathbf{B}\|_F < \delta/2$ (Proposition 16.3.4, or both norms being 0 if $\mathbf{A} = \mathbf{B}$, the case the proposition excludes), puts every eigenvalue of $\mathbf{B}$ in the union of the discs, but, as in the first example of this section, it allows two of them to fall in the same disc and leave another disc empty.

Exercise. C3: A Jordan block of any size

Let $n \geq 2$, $\varepsilon > 0$, $\mathbf{A} = \mathbf{J}_n(0)$, and $\mathbf{B} = \mathbf{J}_n(0) + \varepsilon \mathbf{E}_{n1}$, where $\mathbf{E}_{n1}$ is the matrix unit.

- Show that the characteristic polynomial of $\mathbf{B}$ is $x^n - \varepsilon$, and hence that every pairing of the eigenvalues of $\mathbf{A}$ and $\mathbf{B}$ costs $n\varepsilon^{2/n}$.
- Deduce that the Hoffman–Wielandt inequality fails for every pairing whenever $0 < \varepsilon < n^{n/(2n-2)}$, and that the ratio of the cost to $\|\mathbf{A} - \mathbf{B}\|_F^2$ tends to ∞ as $\varepsilon \rightarrow 0^+$.

Hint for (a): expand $\det(x\mathbf{I} - \mathbf{B})$ along the first column.

Solution.

- $x\mathbf{I} - \mathbf{B}$ has x on the diagonal, -1 on the superdiagonal, $-\varepsilon$ in position $(n, 1)$, and zeros elsewhere. Expanding along the first column, the $(1, 1)$ entry x multiplies the determinant of an upper triangular matrix with diagonal x , which is x^{n-1} . The $(n, 1)$ entry $-\varepsilon$ carries the sign $(-1)^{n+1}$ and multiplies the determinant of the matrix left after deleting row n and column 1, which is lower triangular with diagonal entries -1 , hence has determinant $(-1)^{n-1}$. So

$$\det(x\mathbf{I} - \mathbf{B}) = x^n + (-1)^{n+1}(-\varepsilon)(-1)^{n-1} = x^n - \varepsilon.$$

Its roots are the n complex numbers z with $z^n = \varepsilon$, each of modulus $\varepsilon^{1/n}$ (these are Chapter 16's computation in Example (A Jordan block's eigenvalues move like $\varepsilon^{1/k}$)). Every eigenvalue of $\mathbf{A}$ is 0, so every pairing costs $\sum_i |0 - \mu_{\sigma(i)}|^2 = n\varepsilon^{2/n}$.

- $\mathbf{A} - \mathbf{B} = -\varepsilon \mathbf{E}_{n1}$, so $\|\mathbf{A} - \mathbf{B}\|_F^2 = \varepsilon^2$. The inequality fails for every pairing exactly when $n\varepsilon^{2/n} > \varepsilon^2$, that is, $\varepsilon^{2-2/n} < n$, that is, $\varepsilon < n^{1/(2-2/n)} = n^{n/(2n-2)}$; for $n = 2$ this is $\varepsilon < 2$, as in the text. The ratio is $n\varepsilon^{2/n}/\varepsilon^2 = n\varepsilon^{-(2n-2)/n}$, and the exponent $-(2n-2)/n$ is negative, so the ratio tends to ∞ as $\varepsilon \rightarrow 0^+$.