

20.6 Hermitian Perturbations

For Hermitian matrices, Chapter 17 already gave the basic perturbation bound: each eigenvalue moves by at most $\|\mathbf{E}\|_2$, and no smaller constant is possible. This section asks what that bound does **not** tell us, and answers two such questions. The first is practical. A computation rarely hands us $\mathbf{E}$; it hands us an approximate eigenvector, and the question is how good the eigenvalue estimate it yields is. The answer is a residual bound, and when the eigenvalue is separated from the others it improves from linear to **quadratic** in the residual; that is the Kato–Temple theorem. The second concerns small eigenvalues. An absolute bound of size $\|\mathbf{E}\|_2$ says nothing about an eigenvalue much smaller than $\|\mathbf{E}\|_2$, and for perturbations of the form $\mathbf{A} \mapsto \mathbf{SAS}^*$ we prove a **relative** bound instead, Ostrowski’s theorem, which is Sylvester’s law of inertia made quantitative.

Throughout, $F = \mathbb{C}$, matrices whose eigenvalues are ordered are Hermitian, and $\lambda_1(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$ are the eigenvalues of $\mathbf{A}$ in decreasing order, with multiplicity. Vectors are measured in $\|\cdot\|_2$, written $\|\cdot\|$ when no other norm is in sight, and $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* \mathbf{x}$.

20.6.1 What Weyl’s bound leaves open

Recall Corollary 17.3.6: if $\mathbf{A}$ and $\mathbf{E}$ are Hermitian, then $|\lambda_i(\mathbf{A} + \mathbf{E}) - \lambda_i(\mathbf{A})| \leq \|\mathbf{E}\|_2$ for every i . By Corollary 17.3.7 the constant 1 cannot be lowered, since $\mathbf{E} = t\mathbf{I}$ moves every eigenvalue by exactly $t = \|\mathbf{E}\|_2$. As a bound on **absolute** error, for an **unknown** Hermitian perturbation of known size, there is nothing more to be said.

The two assumptions in bold are where there is room. First, in practice $\mathbf{E}$ is not known at all. An iterative method stops with a unit vector $\mathbf{x}$ and a number μ that it believes to be an eigenpair, and what can be measured is the **residual** $\mathbf{r} = \mathbf{Ax} - \mu\mathbf{x}$. Second, an absolute error bound is only informative for eigenvalues larger than the error. For the matrix $\text{diag}(1, 10^{-6})$ and a Hermitian perturbation of size 10^{-2} , Weyl’s bound allows the eigenvalue 10^{-6} to end up anywhere in an interval of radius 10^{-2} , including negative numbers. Whether that is a fair description depends on the kind of perturbation, and for an important kind it is not.

20.6.2 Residual bounds

Suppose $\mathbf{x}$ is a unit vector and μ a real number, and $\mathbf{r} = \mathbf{Ax} - \mu\mathbf{x}$ is small. If $\mathbf{r} = \mathbf{0}$, then μ is an eigenvalue. The next theorem says that if $\mathbf{r}$ is merely small, μ is **near** an eigenvalue, with the size of $\mathbf{r}$ as the distance.

Theorem 20.6.1. Residual Bound for Hermitian Matrices

Let $\mathbf{A} \in M_n(\mathbb{C})$ be Hermitian, let $\mathbf{x} \in \mathbb{C}^n$ with $\|\mathbf{x}\| = 1$, let $\mu \in \mathbb{R}$, and put $\mathbf{r} = \mathbf{Ax} - \mu\mathbf{x}$. Then some eigenvalue λ of $\mathbf{A}$ satisfies $|\lambda - \mu| \leq \|\mathbf{r}\|$; that is, the interval $[\mu - \|\mathbf{r}\|, \mu + \|\mathbf{r}\|]$ contains an eigenvalue of $\mathbf{A}$.

This is the normal case of §04’s Theorem 20.4.4, since a Hermitian matrix is normal. We give the direct proof from the spectral theorem instead, because the expansion it uses is the one Kato–Temple needs.

Proof. By Corollary 12.4.3 there is an orthonormal basis $(\mathbf{q}_1, \dots, \mathbf{q}_n)$ of $\mathbb{C}^n$ with $\mathbf{A}\mathbf{q}_i = \lambda_i\mathbf{q}_i$, where $\lambda_i = \lambda_i(\mathbf{A})$. Write $\mathbf{x} = \sum_i c_i \mathbf{q}_i$, with $\sum_i |c_i|^2 = \|\mathbf{x}\|^2 = 1$ by Theorem 11.2.1 (c). Then

$\mathbf{r} = \sum_i c_i(\lambda_i - \mu)\mathbf{q}_i$, and by the same result

$$\|\mathbf{r}\|^2 = \sum_{i=1}^n |c_i|^2 (\lambda_i - \mu)^2 \geq \left(\min_i (\lambda_i - \mu)^2 \right) \sum_{i=1}^n |c_i|^2 = \min_i (\lambda_i - \mu)^2.$$

Taking square roots, $\min_i |\lambda_i - \mu| \leq \|\mathbf{r}\|$, as claimed. ■

The bound holds for every real μ , and one choice of μ makes the residual smallest: the **Rayleigh quotient** $R_{\mathbf{A}}(\mathbf{x}) = \mathbf{x}^* \mathbf{A} \mathbf{x}$ of Definition 17.1.1, which for a unit vector needs no denominator. With that choice the residual is orthogonal to $\mathbf{x}$, since

$$\langle \mathbf{r}, \mathbf{x} \rangle = \mathbf{x}^* \mathbf{A} \mathbf{x} - \mu \mathbf{x}^* \mathbf{x} = \mu - \mu = 0,$$

and Pythagoras (Theorem 11.1.3) applied to $\mathbf{A} \mathbf{x} = \mu \mathbf{x} + \mathbf{r}$ gives

$$\|\mathbf{r}\|^2 = \|\mathbf{A} \mathbf{x}\|^2 - \mu^2 \quad (\mu = R_{\mathbf{A}}(\mathbf{x}), \|\mathbf{x}\| = 1). \quad (20.6.1)$$

Exercise B3 shows that any other μ gives a larger residual. From now on, μ is the Rayleigh quotient and $\varepsilon = \|\mathbf{A} \mathbf{x} - \mu \mathbf{x}\|$ its residual.

Quick check. Let $\mathbf{A} = \text{diag}(1, 3)$ and $\mathbf{x} = \frac{1}{\sqrt{2}}(1, 1)$. Compute $\mu = R_{\mathbf{A}}(\mathbf{x})$, the residual $\mathbf{r}$, and $\varepsilon = \|\mathbf{r}\|$. Which eigenvalues does the interval $[\mu - \varepsilon, \mu + \varepsilon]$ contain?

Solution. $\mathbf{A} \mathbf{x} = \frac{1}{\sqrt{2}}(1, 3)$, so $\mu = \mathbf{x}^* \mathbf{A} \mathbf{x} = \frac{1}{2}(1 + 3) = 2$ and $\mathbf{r} = \frac{1}{\sqrt{2}}(1 - 2, 3 - 2) = \frac{1}{\sqrt{2}}(-1, 1)$, with $\varepsilon = 1$. As a check, Eq. 20.6.1 gives $\|\mathbf{A} \mathbf{x}\|^2 - \mu^2 = 5 - 4 = 1$. The interval is $[1, 3]$, and it contains both eigenvalues, at its two ends. The bound is attained: a vector halfway between two eigenvectors is as bad an approximation as it can be.

20.6.3 The Kato–Temple bound

The residual bound is linear in ε , and in general no better is possible: in the Quick check, $|\lambda - \mu| = \varepsilon$ for both eigenvalues. But that example had $\mathbf{x}$ equally far from both eigenvectors. When $\mathbf{x}$ is close to one eigenvector, the Rayleigh quotient is a much better estimate than ε suggests. Exercise (C1: Why the quotient estimates eigenvalues so well) of Chapter 17 found that an eigenvector error of size δ costs only about δ^2 in the Rayleigh quotient. The trouble is that the eigenvector error cannot be measured, while the residual can. The next theorem restates the δ^2 phenomenon in terms of the residual, and the price is a hypothesis on how far the other eigenvalues are.

Here is where the gap comes in. If no eigenvalue of $\mathbf{A}$ lies strictly between two real numbers λ and β , then $(t - \lambda)(t - \beta) \geq 0$ at every eigenvalue t , so the matrix $(\mathbf{A} - \lambda \mathbf{I})(\mathbf{A} - \beta \mathbf{I})$ is positive semidefinite. Evaluating its quadratic form at $\mathbf{x}$ gives an inequality that involves only μ , ε , λ , and β .

Theorem 20.6.2. Kato–Temple Theorem

Let $\mathbf{A} \in M_n(\mathbb{C})$ be Hermitian, let $\mathbf{x} \in \mathbb{C}^n$ with $\|\mathbf{x}\| = 1$, let $\mu = \mathbf{x}^* \mathbf{A} \mathbf{x}$, and let $\varepsilon = \|\mathbf{A} \mathbf{x} - \mu \mathbf{x}\|$.

- (a) Let $\beta > \mu$, and let $\lambda < \beta$ be an eigenvalue of $\mathbf{A}$ such that **no** eigenvalue of $\mathbf{A}$ lies in the open interval (λ, β) . Then

$$\lambda \geq \mu - \frac{\varepsilon^2}{\beta - \mu}.$$

- (b) Let $\alpha < \mu$, and let $\lambda > \alpha$ be an eigenvalue of $\mathbf{A}$ such that **no** eigenvalue of $\mathbf{A}$ lies in the open interval (α, λ) . Then

$$\lambda \leq \mu + \frac{\varepsilon^2}{\mu - \alpha}.$$

In particular, if $\alpha < \mu < \beta$ and λ is the **only** eigenvalue of $\mathbf{A}$ in (α, β) , then

$$\mu - \frac{\varepsilon^2}{\beta - \mu} \leq \lambda \leq \mu + \frac{\varepsilon^2}{\mu - \alpha}.$$

Idea. Compute the number $\mathbf{x}^*(\mathbf{A} - \lambda \mathbf{I})(\mathbf{A} - \beta \mathbf{I})\mathbf{x}$ two ways. ① In the eigenbasis it is $\sum_i |c_i|^2 (\lambda_i - \lambda)(\lambda_i - \beta)$, and the gap hypothesis makes every term nonnegative: each eigenvalue lies on or outside the interval between λ and β . ② Expanding the product, it is $\|\mathbf{A} \mathbf{x}\|^2 - (\lambda + \beta)\mu + \lambda\beta$, and [Eq. 20.6.1](#) turns this into $\varepsilon^2 + (\mu - \lambda)(\mu - \beta)$. So $(\mu - \lambda)(\beta - \mu) \leq \varepsilon^2$; divide by the positive number $\beta - \mu$. Part (b) is the same with α in place of β .

Proof.

- (a) As in the proof of [Theorem 20.6.1](#), write $\mathbf{x} = \sum_i c_i \mathbf{q}_i$ in an orthonormal eigenbasis, $\mathbf{A} \mathbf{q}_i = \lambda_i \mathbf{q}_i$, $\sum_i |c_i|^2 = 1$. Then $(\mathbf{A} - \lambda \mathbf{I})(\mathbf{A} - \beta \mathbf{I})\mathbf{x} = \sum_i c_i (\lambda_i - \lambda)(\lambda_i - \beta) \mathbf{q}_i$, and by [Theorem 11.2.1 \(b\)](#),

$$\mathbf{x}^*(\mathbf{A} - \lambda \mathbf{I})(\mathbf{A} - \beta \mathbf{I})\mathbf{x} = \sum_{i=1}^n |c_i|^2 (\lambda_i - \lambda)(\lambda_i - \beta).$$

Each eigenvalue λ_i satisfies $\lambda_i \leq \lambda$ or $\lambda_i \geq \beta$, because none lies in (λ, β) . In the first case both factors are ≤ 0 , in the second both are ≥ 0 , since $\lambda < \beta$. So every term is ≥ 0 , and the left-hand side is ≥ 0 .

On the other hand, expanding the product and using $\mathbf{x}^* \mathbf{x} = 1$, $\mathbf{x}^* \mathbf{A} \mathbf{x} = \mu$, and $\mathbf{x}^* \mathbf{A}^2 \mathbf{x} = (\mathbf{A} \mathbf{x})^* (\mathbf{A} \mathbf{x}) = \|\mathbf{A} \mathbf{x}\|^2$, where $\mathbf{A}^* = \mathbf{A}$,

$$\begin{aligned} \mathbf{x}^*(\mathbf{A} - \lambda \mathbf{I})(\mathbf{A} - \beta \mathbf{I})\mathbf{x} &= \|\mathbf{A} \mathbf{x}\|^2 - (\lambda + \beta)\mu + \lambda\beta \\ &= \varepsilon^2 + \mu^2 - (\lambda + \beta)\mu + \lambda\beta \\ &= \varepsilon^2 + (\mu - \lambda)(\mu - \beta), \end{aligned}$$

where the second line is [Eq. 20.6.1](#). Hence $\varepsilon^2 + (\mu - \lambda)(\mu - \beta) \geq 0$, that is, $(\mu - \lambda)(\beta - \mu) \leq \varepsilon^2$. Since $\beta - \mu > 0$, dividing gives $\mu - \lambda \leq \varepsilon^2 / (\beta - \mu)$, which is (a).

- (b) Apply (a) to $-\mathbf{A}$, the unit vector $\mathbf{x}$, the eigenvalue $-\lambda$, and the number $-\alpha$. The Rayleigh quotient of $-\mathbf{A}$ at $\mathbf{x}$ is $-\mu$, its residual is $-(\mathbf{A} \mathbf{x} - \mu \mathbf{x})$, of the same norm ε ,

and $-\alpha > -\mu$. The eigenvalues of $-\mathbf{A}$ are the negatives of those of $\mathbf{A}$, so $-\lambda < -\alpha$ and none lies in $(-\lambda, -\alpha)$. Part (a) gives $-\lambda \geq -\mu - \varepsilon^2/(-\alpha + \mu)$, which is (b).

The final statement follows: if λ is the only eigenvalue in (α, β) , then no eigenvalue lies in (λ, β) or in (α, λ) , so (a) and (b) both apply. This proves the theorem. ■

Compare the two bounds on the error $|\lambda - \mu|$. The residual bound gives ε . Kato–Temple gives ε^2 divided by the distance from μ to the neighboring part of the spectrum. When ε is small compared with that gap, the square wins by a wide margin: a residual of 10^{-4} and a gap of 1 give an error of at most 10^{-8} . *The Rayleigh quotient's error is quadratic in the residual, provided there is a gap.* That is the reason the Rayleigh quotient is the eigenvalue estimate of choice, and Chapter 24 builds an iteration on it.

The theorem needs a number α or β that is known to lie beyond the neighboring eigenvalue, and in practice that information has to come from somewhere. The next example gets it from interlacing.

Example. Bracketing the top eigenvalue

Let

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}, \quad \mathbf{x} = \frac{1}{\sqrt{3}}(1, 1, 1).$$

Estimate $\lambda_1(\mathbf{A})$ from $\mathbf{x}$: compute μ and ε , the interval given by [Theorem 20.6.1](#), and the interval given by [Theorem 20.6.2](#), using only information that can be read off $\mathbf{A}$ without computing its eigenvalues. Compare with the true value $\lambda_1(\mathbf{A}) = 2 + \sqrt{2}$.

Solution. Rayleigh quotient and residual. $\mathbf{A}(1, 1, 1) = (3, 4, 3)$, so $\mathbf{Ax} = \frac{1}{\sqrt{3}}(3, 4, 3)$ and

$$\mu = \mathbf{x}^* \mathbf{Ax} = \frac{1}{3}(3 + 4 + 3) = \frac{10}{3}, \quad \|\mathbf{Ax}\|^2 = \frac{1}{3}(9 + 16 + 9) = \frac{34}{3}.$$

By [Eq. 20.6.1](#), $\varepsilon^2 = \frac{34}{3} - \frac{100}{9} = \frac{2}{9}$, so $\varepsilon = \frac{\sqrt{2}}{3} \approx 0.471$. Directly: $\mathbf{r} = \frac{1}{\sqrt{3}}(3 - \frac{10}{3}, 4 - \frac{10}{3}, 3 - \frac{10}{3}) = \frac{1}{3\sqrt{3}}(-1, 2, -1)$, of squared norm $\frac{6}{9} = \frac{2}{9}$.

Residual bound. Some eigenvalue lies in $[\frac{10}{3} - 0.471, \frac{10}{3} + 0.471] \approx [2.862, 3.805]$.

Kato–Temple. By Proposition 17.1.1, $\lambda_1(\mathbf{A}) \geq \mu = \frac{10}{3}$. For an upper bound we need $\alpha < \mu$ with no eigenvalue in (α, λ_1) , which says $\lambda_2(\mathbf{A}) \leq \alpha$. Deleting the second row and column leaves $\mathbf{B} = \text{diag}(2, 2)$, and Cauchy's interlacing theorem (Theorem 17.4.2, with $I = \{1, 3\}$) gives $\lambda_2(\mathbf{A}) \leq \lambda_1(\mathbf{B}) = 2$. So $\alpha = 2$ is allowed: $\alpha < \mu$, and no eigenvalue lies in $(2, \lambda_1)$, because $\lambda_2 \leq 2$ and $\lambda_1 \geq \frac{10}{3} > 2$. Part (b) gives

$$\lambda_1(\mathbf{A}) \leq \frac{10}{3} + \frac{2/9}{10/3 - 2} = \frac{10}{3} + \frac{2}{9} \cdot \frac{3}{4} = \frac{10}{3} + \frac{1}{6} = \frac{7}{2}.$$

So $\frac{10}{3} \leq \lambda_1(\mathbf{A}) \leq \frac{7}{2}$, an interval of width $\frac{1}{6} \approx 0.167$, against width $2\varepsilon \approx 0.943$ for the residual bound.

Check. The true value is $2 + \sqrt{2} \approx 3.414$, which lies in both intervals; the other eigenvalues

are 2 and $2 - \sqrt{2}$. The actual error of the Rayleigh quotient is $2 + \sqrt{2} - \frac{10}{3} \approx 0.081$, about a sixth of ε .

Kato–Temple cannot be improved in general: for $\mathbf{A} = \text{diag}(\lambda_1, \lambda_2)$ with $\lambda_1 > \lambda_2$, and $\alpha = \lambda_2$ the other eigenvalue, the upper bound in (b) is an **equality** at every real unit $\mathbf{x}$ that is not an eigenvector for λ_2 (Exercise C3). At such an eigenvector the hypothesis $\alpha < \mu$ fails, since there $\mu = \lambda_2$.

Warning. The gap hypothesis is not decoration. Take $\mathbf{A} = \text{diag}(1, 0)$ and $\mathbf{x} = \frac{1}{\sqrt{2}}(1, 1)$, so $\mu = \frac{1}{2}$ and $\mathbf{r} = \frac{1}{\sqrt{2}}(\frac{1}{2}, -\frac{1}{2})$, $\varepsilon^2 = \frac{1}{4}$. With $\lambda = 1$ and $\alpha = 0$, part (b) gives $1 \leq \frac{1}{2} + \frac{1/4}{1/2} = 1$, true with equality. With $\alpha = -\frac{1}{2}$ instead, the eigenvalue 0 lies in (α, λ) , the hypothesis fails, and the formula would claim $1 \leq \frac{1}{2} + \frac{1/4}{1} = \frac{3}{4}$, which is false. A number α chosen too far from μ , past an unseen eigenvalue, gives a bound that is simply wrong.

20.6.4 Relative perturbations and congruence

Now the second question: small eigenvalues. Weyl’s bound is absolute, and an absolute bound is useless for an eigenvalue smaller than itself. Some perturbations deserve better. In a change of variables $\mathbf{x} = \mathbf{S}^*\mathbf{y}$ with $\mathbf{S}$ invertible, the quadratic form $\mathbf{x}^*\mathbf{A}\mathbf{x}$ becomes $\mathbf{y}^*(\mathbf{S}\mathbf{A}\mathbf{S}^*)\mathbf{y}$; the matrix undergoes a **congruence**, as in Chapter 14 and in Chapter 17 §10. When $\mathbf{S}$ is close to $\mathbf{I}$, for instance when it comes from a slightly inaccurate elimination step or from rounding in a factor, $\mathbf{S}\mathbf{A}\mathbf{S}^*$ is a perturbation of $\mathbf{A}$, but a very structured one. Its effect on the eigenvalues is **multiplicative**.

Two facts are known already. By Theorem 17.10.1 (b), applied with $\mathbf{P} = \mathbf{S}^*$, a congruence preserves the numbers of positive, negative, and zero eigenvalues; it cannot change a sign. And by Chapter 14 §05’s warning, it can still move eigenvalues a long way: with $\mathbf{P} = \text{diag}(2, 3)$, $\mathbf{P}^\top \text{diag}(1, -1)\mathbf{P} = \text{diag}(4, -9)$. Ostrowski’s theorem measures how far.

Theorem 20.6.3. Ostrowski’s Theorem

Let $\mathbf{A} \in M_n(\mathbb{C})$ be Hermitian and let $\mathbf{S} \in M_n(\mathbb{C})$ be invertible. Then $\mathbf{S}\mathbf{S}^*$ is positive definite, and for each $k = 1, \dots, n$ there is a real number θ_k with

$$\lambda_k(\mathbf{S}\mathbf{A}\mathbf{S}^*) = \theta_k \lambda_k(\mathbf{A}), \quad \lambda_n(\mathbf{S}\mathbf{S}^*) \leq \theta_k \leq \lambda_1(\mathbf{S}\mathbf{S}^*).$$

In particular $\theta_k > 0$, and

$$|\lambda_k(\mathbf{S}\mathbf{A}\mathbf{S}^*) - \lambda_k(\mathbf{A})| \leq \|\mathbf{S}\mathbf{S}^* - \mathbf{I}\|_2 |\lambda_k(\mathbf{A})|.$$

Idea. Step roadmap. Fix k and put $\mu = \lambda_k(\mathbf{A})$.

① **Shift**, so the k -th eigenvalue is 0. $\mathbf{A} - \mu\mathbf{I}$ has at most $k - 1$ positive eigenvalues and at least k nonnegative ones. A congruence preserves both counts (Sylvester’s law), so $\mathbf{S}(\mathbf{A} - \mu\mathbf{I})\mathbf{S}^*$ has the same property, and that forces its k -th eigenvalue to be exactly 0.

② **Split**. $\mathbf{S}\mathbf{A}\mathbf{S}^* = \mathbf{S}(\mathbf{A} - \mu\mathbf{I})\mathbf{S}^* + \mu\mathbf{S}\mathbf{S}^*$.

③ **Weyl**. Adding $\mu\mathbf{S}\mathbf{S}^*$ moves the k -th eigenvalue from 0 to somewhere between the small-

est and the largest eigenvalue of $\mu\mathbf{S}\mathbf{S}^*$, which are $\mu\lambda_n(\mathbf{S}\mathbf{S}^*)$ and $\mu\lambda_1(\mathbf{S}\mathbf{S}^*)$ in some order, depending on the sign of μ .

Proof. For $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^*\mathbf{S}\mathbf{S}^*\mathbf{x} = \|\mathbf{S}^*\mathbf{x}\|^2 > 0$, because $\mathbf{S}^*$ is invertible. So $\mathbf{S}\mathbf{S}^*$, which is Hermitian, is positive definite, and its eigenvalues are positive by Theorem 13.1.5; write $m = \lambda_n(\mathbf{S}\mathbf{S}^*) > 0$ and $M = \lambda_1(\mathbf{S}\mathbf{S}^*)$.

Fix k and put $\mu = \lambda_k(\mathbf{A})$. The eigenvalues of the Hermitian matrix $\mathbf{A} - \mu\mathbf{I}$ are $\lambda_i(\mathbf{A}) - \mu$, in the same decreasing order. Those with $i \geq k$ are ≤ 0 , so at most $k - 1$ of them are positive; those with $i \leq k$ are ≥ 0 , so at least k are nonnegative. In the notation of Definition 17.10.1, $n_+(\mathbf{A} - \mu\mathbf{I}) \leq k - 1$ and $n_+(\mathbf{A} - \mu\mathbf{I}) + n_0(\mathbf{A} - \mu\mathbf{I}) \geq k$. Put $\mathbf{Y} = \mathbf{S}(\mathbf{A} - \mu\mathbf{I})\mathbf{S}^*$. This is $\mathbf{P}^*(\mathbf{A} - \mu\mathbf{I})\mathbf{P}$ with $\mathbf{P} = \mathbf{S}^*$ invertible, so by Theorem 17.10.1 (b), $\mathbf{Y}$ is Hermitian with the same inertia triple. Hence $\mathbf{Y}$ has at most $k - 1$ positive eigenvalues, so $\lambda_k(\mathbf{Y}) \leq 0$; and at least k nonnegative ones, so $\lambda_k(\mathbf{Y}) \geq 0$. Thus $\lambda_k(\mathbf{Y}) = 0$.

Now $\mathbf{S}\mathbf{A}\mathbf{S}^* = \mathbf{Y} + \mu\mathbf{S}\mathbf{S}^*$, a sum of two Hermitian matrices. By Corollary 17.3.4,

$$\lambda_n(\mu\mathbf{S}\mathbf{S}^*) \leq \lambda_k(\mathbf{S}\mathbf{A}\mathbf{S}^*) - \lambda_k(\mathbf{Y}) \leq \lambda_1(\mu\mathbf{S}\mathbf{S}^*),$$

and $\lambda_k(\mathbf{Y}) = 0$. If $\mu \geq 0$, multiplying by μ keeps the order of the eigenvalues, so $\lambda_n(\mu\mathbf{S}\mathbf{S}^*) = \mu m$ and $\lambda_1(\mu\mathbf{S}\mathbf{S}^*) = \mu M$. If $\mu < 0$, then $\mu\mathbf{S}\mathbf{S}^* = -(|\mu|\mathbf{S}\mathbf{S}^*)$, and by Lemma 17.3.2 its extreme eigenvalues are $\lambda_n = -|\mu|M = \mu M$ and $\lambda_1 = -|\mu|m = \mu m$. In both cases $\lambda_k(\mathbf{S}\mathbf{A}\mathbf{S}^*)$ lies between μm and μM . If $\mu \neq 0$, put $\theta_k = \lambda_k(\mathbf{S}\mathbf{A}\mathbf{S}^*)/\mu$; dividing by μ (and reversing the inequalities when $\mu < 0$) gives $m \leq \theta_k \leq M$. If $\mu = 0$, both ends are 0, so $\lambda_k(\mathbf{S}\mathbf{A}\mathbf{S}^*) = 0$, and $\theta_k = m$ will do. In either case $\theta_k \geq m > 0$.

For the last inequality, $|\lambda_k(\mathbf{S}\mathbf{A}\mathbf{S}^*) - \lambda_k(\mathbf{A})| = |\theta_k - 1||\mu|$, and $\theta_k - 1$ lies in $[m - 1, M - 1]$. The numbers $m - 1$ and $M - 1$ are the smallest and largest eigenvalues of the Hermitian matrix $\mathbf{S}\mathbf{S}^* - \mathbf{I}$, so by Lemma 17.3.5 both have absolute value at most $\|\mathbf{S}\mathbf{S}^* - \mathbf{I}\|_2$, and so does everything between them. This proves the theorem. ■

This is a quantitative Sylvester's law of inertia. Sylvester's law says that a congruence cannot change the sign of an eigenvalue; Ostrowski's theorem says that it multiplies the k -th eigenvalue by a factor θ_k that is always positive and always between the extreme eigenvalues of $\mathbf{S}\mathbf{S}^*$. These are $\lambda_1(\mathbf{S}\mathbf{S}^*) = \max_{\|\mathbf{y}\|=1} \|\mathbf{S}^*\mathbf{y}\|^2$ and $\lambda_n(\mathbf{S}\mathbf{S}^*) = \min_{\|\mathbf{y}\|=1} \|\mathbf{S}^*\mathbf{y}\|^2$, by Proposition 17.1.1, since $\mathbf{y}^*\mathbf{S}\mathbf{S}^*\mathbf{y} = \|\mathbf{S}^*\mathbf{y}\|^2$: the most and the least that the change of variables stretches a vector, squared. When $\mathbf{S}$ is close to a unitary matrix, both are close to 1, and **every** eigenvalue, however small, keeps most of its significant digits.

Quick check. Chapter 14 §05's warning used $\mathbf{A} = \text{diag}(1, -1)$ and $\mathbf{P} = \text{diag}(2, 3)$, with $\mathbf{P}^\top \mathbf{A} \mathbf{P} = \text{diag}(4, -9)$. Take $\mathbf{S} = \mathbf{P}^\top = \mathbf{P}$. Find θ_1 , θ_2 , and the interval $[\lambda_n(\mathbf{S}\mathbf{S}^*), \lambda_1(\mathbf{S}\mathbf{S}^*)]$. Can the interval in Theorem 20.6.3 be shortened?

Solution. $\mathbf{S}\mathbf{S}^* = \text{diag}(4, 9)$, so the interval is $[4, 9]$. The eigenvalues of $\mathbf{S}\mathbf{A}\mathbf{S}^* = \text{diag}(4, -9)$ in decreasing order are 4 and -9 , and those of $\mathbf{A}$ are 1 and -1 . So $\theta_1 = 4/1 = 4$ and $\theta_2 = -9/(-1) = 9$. They sit at the two ends of the interval, so neither end can be moved

inwards: the interval is sharp.

The next example shows the difference between the absolute and the relative view on a matrix whose eigenvalues differ greatly in size.

Example. A tiny eigenvalue under a change of variables

Let $\mathbf{A} = \text{diag}(1, 10^{-6})$ and $\mathbf{S} = \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}$ with $t = 10^{-2}$, the change of variables that adds t times the first coordinate to the second. Compute $\mathbf{SAS}^*$ and $\mathbf{E} = \mathbf{SAS}^* - \mathbf{A}$. What do Corollary 17.3.4 and Theorem 20.6.3 say about $\lambda_2(\mathbf{SAS}^*)$, and what is its true value?

Solution. Write $\eta = 10^{-6}$. Multiplying out,

$$\mathbf{SAS}^* = \begin{pmatrix} 1 & t \\ t & t^2 + \eta \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} 0 & t \\ t & t^2 \end{pmatrix},$$

with $t^2 + \eta = 1.01 \times 10^{-4}$.

Weyl. $\mathbf{E}$ has trace t^2 and determinant $-t^2$, so its eigenvalues are $\frac{1}{2}(t^2 \pm t\sqrt{t^2 + 4})$, approximately 0.01005 and -0.00995 . By Corollary 17.3.4,

$$10^{-6} - 0.00995 \leq \lambda_2(\mathbf{SAS}^*) \leq 10^{-6} + 0.01005.$$

This interval is about 10^4 times wider than the eigenvalue it is meant to locate, and it contains negative numbers.

Ostrowski. $\mathbf{SS}^* = \begin{pmatrix} 1 & t \\ t & 1+t^2 \end{pmatrix}$ has trace $2 + t^2$ and determinant 1, so its eigenvalues are $\frac{1}{2}(2 + t^2 \pm t\sqrt{4 + t^2})$, approximately 1.01005 and 0.99005. By Theorem 20.6.3,

$$0.99005 \times 10^{-6} \leq \lambda_2(\mathbf{SAS}^*) \leq 1.01005 \times 10^{-6} :$$

the eigenvalue is positive and known to within about 1%.

The truth. $\det(\mathbf{SAS}^*) = \det(\mathbf{S})^2 \det \mathbf{A} = \eta$, since $\det \mathbf{S} = 1$, and the trace is $1 + t^2 + \eta$. So the two eigenvalues multiply to η , and the larger is $1.0001\dots$; hence $\lambda_2(\mathbf{SAS}^*) = \eta/\lambda_1(\mathbf{SAS}^*) \approx 0.99990 \times 10^{-6}$. The relative change is about 10^{-4} , well inside Ostrowski's 1%; Weyl's bound, by contrast, cannot even decide the sign.

Warning. Weyl's bound is sharp and still useless for small eigenvalues. It is the best possible bound on **absolute** error for **arbitrary** Hermitian $\mathbf{E}$ of a given norm, and for exactly that reason it cannot distinguish an eigenvalue of size 10^{-6} from 0 once $\|\mathbf{E}\|_2 \geq 10^{-6}$. A general Hermitian $\mathbf{E}$ really can destroy a small eigenvalue: $\mathbf{E} = -10^{-6}\mathbf{e}_2\mathbf{e}_2^\top$ sends the eigenvalue 10^{-6} of $\text{diag}(1, 10^{-6})$ to 0. Relative accuracy needs a structured perturbation, such as a congruence $\mathbf{SAS}^*$ with $\mathbf{S}$ near $\mathbf{I}$.

20.6.5 Exercises

20.6.5.1 A. Check your understanding

Exercise. A1

- (a) State the Kato–Temple theorem, with its hypotheses.
- (b) True or false: for Hermitian $\mathbf{A}$, a unit vector $\mathbf{x}$, and real μ , the eigenvalue of $\mathbf{A}$ closest to μ is within $\|\mathbf{A}\mathbf{x} - \mu\mathbf{x}\|$ of μ . Justify your answer.
- (c) True or false: for Hermitian $\mathbf{A}$ and invertible $\mathbf{S}$, $\lambda_k(\mathbf{SAS}^*)$ and $\lambda_k(\mathbf{A})$ have the same sign for every k . Justify your answer.
- (d) Which of the three bounds of this section need information that is not contained in $\mathbf{A}$, $\mathbf{x}$, and μ ? What information?
- (e) Why can the constant 1 in Corollary 17.3.6 not be lowered, and why does that not contradict Ostrowski’s theorem?

Solution.

- (a) Let $\mathbf{A}$ be Hermitian, $\mathbf{x}$ a unit vector, $\mu = \mathbf{x}^*\mathbf{A}\mathbf{x}$, $\varepsilon = \|\mathbf{A}\mathbf{x} - \mu\mathbf{x}\|$. If $\beta > \mu$ and $\lambda < \beta$ is an eigenvalue with no eigenvalue in (λ, β) , then $\lambda \geq \mu - \varepsilon^2/(\beta - \mu)$. If $\alpha < \mu$ and $\lambda > \alpha$ is an eigenvalue with no eigenvalue in (α, λ) , then $\lambda \leq \mu + \varepsilon^2/(\mu - \alpha)$.
- (b) True. [Theorem 20.6.1](#) gives some eigenvalue within $\|\mathbf{A}\mathbf{x} - \mu\mathbf{x}\|$ of μ , and the closest one is at least as close.
- (c) True. By [Theorem 20.6.3](#), $\lambda_k(\mathbf{SAS}^*) = \theta_k \lambda_k(\mathbf{A})$ with $\theta_k > 0$; a positive multiple has the same sign, and 0 goes to 0.
- (d) The residual bound needs nothing more. Kato–Temple needs a number α or β beyond the neighboring eigenvalue, that is, a lower bound on the gap; in [Example \(Bracketing the top eigenvalue\)](#) it came from interlacing. Ostrowski’s theorem is not about an approximate eigenvector, but it needs $\mathbf{S}$, or at least the extreme eigenvalues of $\mathbf{SS}^*$.
- (e) $\mathbf{E} = t\mathbf{I}$ moves every eigenvalue by exactly $t = \|\mathbf{E}\|_2$. Ostrowski’s theorem is about the perturbations $\mathbf{SAS}^* - \mathbf{A}$, which form a special class; $\mathbf{A} + t\mathbf{I}$ is not of the form $\mathbf{SAS}^*$ in general (for $\mathbf{A} = \mathbf{O}$ it never is, when $t \neq 0$). A better bound for a smaller class contradicts nothing.

20.6.5.2 B. Practice

Exercise. B1: Three bounds for one estimate

Let $\mathbf{A} = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}$ and $\mathbf{x} = \frac{1}{\sqrt{10}}(3, 1)$.

- (a) Compute $\mu = \mathbf{x}^*\mathbf{A}\mathbf{x}$ and $\varepsilon = \|\mathbf{A}\mathbf{x} - \mu\mathbf{x}\|$, and the interval given by [Theorem 20.6.1](#).
- (b) Using Theorem 17.4.2 to bound $\lambda_2(\mathbf{A})$, apply [Theorem 20.6.2](#) to bracket $\lambda_1(\mathbf{A})$.
- (c) Find the eigenvalues of $\mathbf{A}$ exactly. Hence compare the actual error of μ with ε and with the Kato–Temple bound.

Solution.

- (a) $\mathbf{A}(3, 1) = (17, 8)$, so $\mu = \frac{1}{10}(3 \cdot 17 + 1 \cdot 8) = \frac{59}{10} = 5.9$. The residual is $\frac{1}{\sqrt{10}}((17, 8) - 5.9(3, 1)) = \frac{1}{\sqrt{10}}(-0.7, 2.1)$, so $\varepsilon^2 = \frac{1}{10}(0.49 + 4.41) = 0.49$ and $\varepsilon = 0.7$. Some eigenvalue lies in $[5.2, 6.6]$.
- (b) Deleting the first row and column leaves (2) , so Theorem 17.4.2 gives $\lambda_2(\mathbf{A}) \leq 2$. Also $\lambda_1(\mathbf{A}) \geq \mu = 5.9$ by Proposition 17.1.1. So with $\alpha = 2 < \mu$, no eigenvalue lies in $(2, \lambda_1)$, and part (b) of Theorem 20.6.2 gives

$$\lambda_1(\mathbf{A}) \leq 5.9 + \frac{0.49}{5.9 - 2} = 5.9 + \frac{0.49}{3.9} \approx 6.026.$$

Hence $5.9 \leq \lambda_1(\mathbf{A}) \leq 6.026$.

- (c) $\mathbf{A}$ has trace 7 and determinant 6, so its eigenvalues are 6 and 1; indeed $\mathbf{A}(2, 1) = (12, 6)$. The actual error of μ is $6 - 5.9 = 0.1$. The residual bound allows 0.7, seven times too much; Kato–Temple allows 0.126, within about 26% of the truth.

Exercise. B2: Ostrowski's factors

Let $\mathbf{A} = \text{diag}(2, -1)$ and $\mathbf{S} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Compute $\mathbf{SAS}^*$ and its eigenvalues, the eigenvalues of $\mathbf{SS}^*$, and the factors θ_1, θ_2 of Theorem 20.6.3. Hence verify the theorem for this pair.

Solution. $\mathbf{AS}^* = \text{diag}(2, -1)\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ -1 & -1 \end{pmatrix}$, and

$$\mathbf{SAS}^* = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ -1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix},$$

with trace 0 and determinant -2 , so eigenvalues $\pm\sqrt{2}$. Next $\mathbf{SS}^* = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, with trace 3 and determinant 1, so eigenvalues $\frac{3 \pm \sqrt{5}}{2} \approx 2.618$ and 0.382 .

Then $\theta_1 = \lambda_1(\mathbf{SAS}^*)/\lambda_1(\mathbf{A}) = \sqrt{2}/2 \approx 0.707$ and $\theta_2 = (-\sqrt{2})/(-1) = \sqrt{2} \approx 1.414$. Both lie in $[0.382, 2.618]$, as the theorem says, and both are positive, so the signs $+$, $-$ are preserved.

Exercise. B3: The Rayleigh quotient minimizes the residual

Let $\mathbf{A}$ be Hermitian and $\mathbf{x}$ a unit vector, with $\rho = \mathbf{x}^* \mathbf{A} \mathbf{x}$. Prove that for every real ν ,

$$\|\mathbf{Ax} - \nu \mathbf{x}\|^2 = \|\mathbf{Ax} - \rho \mathbf{x}\|^2 + (\nu - \rho)^2.$$

Hence deduce that among all real ν , the residual bound of Theorem 20.6.1 is sharpest at $\nu = \rho$.

Solution. Write $\mathbf{Ax} - \nu \mathbf{x} = (\mathbf{Ax} - \rho \mathbf{x}) + (\rho - \nu)\mathbf{x}$. The first vector is orthogonal to $\mathbf{x}$, since $\langle \mathbf{Ax} - \rho \mathbf{x}, \mathbf{x} \rangle = \mathbf{x}^* \mathbf{A} \mathbf{x} - \rho = 0$. By Theorem 11.1.3,

$$\|\mathbf{Ax} - \nu \mathbf{x}\|^2 = \|\mathbf{Ax} - \rho \mathbf{x}\|^2 + (\rho - \nu)^2 \|\mathbf{x}\|^2 = \|\mathbf{Ax} - \rho \mathbf{x}\|^2 + (\nu - \rho)^2.$$

Hence $\|\mathbf{Ax} - \nu \mathbf{x}\| \geq \|\mathbf{Ax} - \rho \mathbf{x}\|$, with equality only for $\nu = \rho$. The interval $[\nu - \|\mathbf{Ax} - \nu \mathbf{x}\|, \nu + \|\mathbf{Ax} - \nu \mathbf{x}\|]$ of Theorem 20.6.1 therefore has its smallest radius at $\nu = \rho$.

20.6.5.3 C. Going deeper

Exercise. C1: A residual that guarantees an eigenvalue

Let $\mathbf{A}$ be Hermitian, $\mathbf{x}$ a unit vector, $\mu = \mathbf{x}^* \mathbf{A} \mathbf{x}$, and $\varepsilon = \|\mathbf{A} \mathbf{x} - \mu \mathbf{x}\|$.

- (a) Let $\alpha < \mu < \beta$. Prove that if $\varepsilon^2 < (\mu - \alpha)(\beta - \mu)$, then $\mathbf{A}$ has an eigenvalue in (α, β) .
- (b) Deduce that $\mathbf{A}$ has an eigenvalue in $[\mu - \varepsilon, \mu + \varepsilon]$, which is [Theorem 20.6.1](#) again, for this choice of μ .

Hint for (a): if there is none, consider $\mathbf{x}^*(\mathbf{A} - \alpha \mathbf{I})(\mathbf{A} - \beta \mathbf{I})\mathbf{x}$.

Solution.

- (a) Suppose no eigenvalue lies in (α, β) . Then every eigenvalue t satisfies $t \leq \alpha$ or $t \geq \beta$, so $(t - \alpha)(t - \beta) \geq 0$. As in the proof of [Theorem 20.6.2](#), with α in place of λ ,

$$0 \leq \mathbf{x}^*(\mathbf{A} - \alpha \mathbf{I})(\mathbf{A} - \beta \mathbf{I})\mathbf{x} = \varepsilon^2 + (\mu - \alpha)(\mu - \beta),$$

so $\varepsilon^2 \geq (\mu - \alpha)(\beta - \mu)$, contradicting the hypothesis. Hence some eigenvalue lies in (α, β) .

- (b) Let d be the smallest distance from μ to an eigenvalue of $\mathbf{A}$, and suppose $d > \varepsilon$. Choose s with $\varepsilon < s < d$, and apply (a) with $\alpha = \mu - s$ and $\beta = \mu + s$: the hypothesis holds, since $(\mu - \alpha)(\beta - \mu) = s^2 > \varepsilon^2$. So some eigenvalue lies in $(\mu - s, \mu + s)$, at distance less than $s < d$ from μ , contradicting the choice of d . Hence $d \leq \varepsilon$, which is the claim.

Exercise. C2: Relative bounds need structured perturbations

- (a) Let $\mathbf{A}$ be Hermitian and $\mathbf{S} = \mathbf{I} + \mathbf{F}$ with $\mathbf{F} \in M_n(\mathbb{C})$. Prove that

$$|\lambda_k(\mathbf{SAS}^*) - \lambda_k(\mathbf{A})| \leq (2\|\mathbf{F}\|_2 + \|\mathbf{F}\|_2^2) |\lambda_k(\mathbf{A})|$$

for every k , whenever $\mathbf{S}$ is invertible.

- (b) Show that there is **no** constant c such that $|\lambda_k(\mathbf{A} + \mathbf{E}) - \lambda_k(\mathbf{A})| \leq c\|\mathbf{E}\|_2 |\lambda_k(\mathbf{A})|$ for all Hermitian $\mathbf{A}$, $\mathbf{E}$ and all k .

Hint for (a): expand $\mathbf{SS}^* - \mathbf{I}$, and show first that $\|\mathbf{F}^*\|_2 = \|\mathbf{F}\|_2$.

Solution.

- (a) First, $\|\mathbf{F}^*\|_2 \leq \|\mathbf{F}\|_2$: for every $\mathbf{x}$, by [Theorem 11.1.5](#) and [Theorem 16.3.2 \(a\)](#),

$$\|\mathbf{F}^* \mathbf{x}\|^2 = \langle \mathbf{F} \mathbf{F}^* \mathbf{x}, \mathbf{x} \rangle \leq \|\mathbf{F} \mathbf{F}^* \mathbf{x}\| \|\mathbf{x}\| \leq \|\mathbf{F}\|_2 \|\mathbf{F}^* \mathbf{x}\| \|\mathbf{x}\|,$$

so $\|\mathbf{F}^* \mathbf{x}\| \leq \|\mathbf{F}\|_2 \|\mathbf{x}\|$ (trivially if $\mathbf{F}^* \mathbf{x} = \mathbf{0}$), and the reverse inequality follows by applying this to $\mathbf{F}^*$, since $(\mathbf{F}^*)^* = \mathbf{F}$. Now $\mathbf{SS}^* - \mathbf{I} = \mathbf{F} + \mathbf{F}^* + \mathbf{F} \mathbf{F}^*$, so by the triangle inequality for $\|\cdot\|_2$, which is a norm by [Theorem 16.3.2 \(b\)](#), and by part (d) of the same theorem,

$$\|\mathbf{SS}^* - \mathbf{I}\|_2 \leq \|\mathbf{F}\|_2 + \|\mathbf{F}^*\|_2 + \|\mathbf{F}\|_2 \|\mathbf{F}^*\|_2 = 2\|\mathbf{F}\|_2 + \|\mathbf{F}\|_2^2.$$

The last inequality of [Theorem 20.6.3](#) then gives the claim.

- (b) Take $\mathbf{A} = \text{diag}(1, 0)$ and $\mathbf{E} = \text{diag}(0, t)$ with $t > 0$. Then $\lambda_2(\mathbf{A}) = 0$ and $\lambda_2(\mathbf{A} + \mathbf{E}) = \min(1, t) > 0$, so the left side is positive while the right side is $ct \cdot 0 = 0$. No constant c works. By (a), such an $\mathbf{E}$ is not of the form $\mathbf{SAS}^* - \mathbf{A}$: congruence can never move the eigenvalue 0.

Exercise. C3: Kato–Temple is sharp

Let $\lambda_1 > \lambda_2$ be real, $\mathbf{A} = \text{diag}(\lambda_1, \lambda_2)$, and $\mathbf{x} = (\cos \phi, \sin \phi)$ with $0 < \phi < \pi/2$.

- (a) Compute $\mu = \mathbf{x}^* \mathbf{A} \mathbf{x}$ and show that $\varepsilon^2 = (\lambda_1 - \lambda_2)^2 \cos^2 \phi \sin^2 \phi$.
 (b) Show that [Theorem 20.6.2](#) (b), with $\lambda = \lambda_1$ and $\alpha = \lambda_2$, holds with **equality**. Hence the constant in the Kato–Temple bound cannot be improved.

Solution. Write $c = \cos \phi$, $s = \sin \phi$, and $d = \lambda_1 - \lambda_2 > 0$.

- (a) $\mathbf{Ax} = (\lambda_1 c, \lambda_2 s)$, so $\mu = \lambda_1 c^2 + \lambda_2 s^2$. Using $c^2 + s^2 = 1$, $\lambda_1 - \mu = ds^2$ and $\lambda_2 - \mu = -dc^2$. The residual is $\mathbf{Ax} - \mu \mathbf{x} = ((\lambda_1 - \mu)c, (\lambda_2 - \mu)s) = (ds^2 c, -dc^2 s)$, so

$$\varepsilon^2 = d^2 s^4 c^2 + d^2 c^4 s^2 = d^2 c^2 s^2 (s^2 + c^2) = d^2 c^2 s^2.$$

- (b) Here $\alpha = \lambda_2 < \mu$, since $\mu - \lambda_2 = dc^2 > 0$, and no eigenvalue lies in (λ_2, λ_1) . The bound reads

$$\mu + \frac{\varepsilon^2}{\mu - \lambda_2} = \mu + \frac{d^2 c^2 s^2}{dc^2} = \mu + ds^2 = \lambda_1.$$

So equality holds for every $\phi \in (0, \pi/2)$, and no smaller multiple of $\varepsilon^2/(\mu - \alpha)$ would give a valid bound.