March 26, 2018

Lecture 12 – Perturbed Eigenvalue Problems

Below $\mathbf{A}$ is an $n \times n$ matrix.
$(\lambda, \mathbf{x})$ is an eigenpair of $\mathbf{A}$: $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ and $\mathbf{x} \neq \mathbf{0}$.
$\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\mathbf{A}$ counted according to multiplicity.
If more matrices play a role we will put $\lambda(\mathbf{A})$ or $\lambda_i(\mathbf{A})$ instead of λ and λ_i, respectively.

The **Rayleigh quotient** ρ is defined by

$$\rho(\mathbf{u}) \equiv \frac{\mathbf{u}^*\mathbf{A}\mathbf{u}}{\mathbf{u}^*\mathbf{u}} \qquad (\mathbf{u} \in \mathbb{C}^n, \mathbf{u} \neq \mathbf{0})$$

The following theorem shows that an approximate eigenpair is an exact eigenpair of a perturbed matrix if Rayleigh quotients are used as approximate eigenvalues. The perturbation can be bounded by the residual. The residual norm is smallest also if Rayleigh quotients are used as approximate eigenvalues.

Theorem 12.1 *Let* $\mathbf{u}$ *be a normalised* n*-vector. Put* $\vartheta \equiv \rho(\mathbf{u})$ *and* $\mathbf{r} \equiv \mathbf{A}\mathbf{u} - \vartheta\mathbf{u}$.
Then $\mathbf{r} \perp \mathbf{u}$ *and* $\|\mathbf{A}\mathbf{u} - \vartheta\mathbf{u}\|_2 \leq \|\mathbf{A}\mathbf{u} - \vartheta'\mathbf{u}\|_2$ *for all* $\vartheta' \in \mathbb{C}$. *Moreover,*

$$(\mathbf{A} + \Delta)\mathbf{u} = \vartheta\mathbf{u} \quad \textit{for} \quad \Delta \equiv -\mathbf{r}\mathbf{u}^* \quad \textit{and} \quad \|\Delta\|_2 = \|\mathbf{r}\|_2.$$

Recall that Δ is the backward error if $(\vartheta, \mathbf{u})$ is an approximate eigenpair and $(\mathbf{A}+\Delta)\mathbf{u} = \vartheta\mathbf{u}$ (cf., Lecture 1.C).

Exercise 12.1.

(a) Prove Theorem 12.1.

(b) Prove that $(\mathbf{A} + \Delta)\mathbf{u} = \vartheta\mathbf{u}$ for the Hermitian perturbation $\Delta \equiv -\mathbf{r}\mathbf{u}^* - \mathbf{u}\mathbf{r}^*$.

Consider the situation where we have a right approximate eigenvector $\mathbf{u}$ as well as a left one, $\mathbf{w}$. Both vectors are normalised.

(c) As approximate eigenvalue ϑ, we now take the two-sided Rayleigh quotient

$$\vartheta \equiv \frac{\mathbf{w}^*\mathbf{A}\mathbf{u}}{\mathbf{w}^*\mathbf{u}}.$$

$\mathbf{r} \equiv \mathbf{A}\mathbf{u} - \vartheta\mathbf{u}$ is the right residual, $\mathbf{s}^* \equiv \mathbf{w}^*\mathbf{A} - \vartheta\mathbf{w}^*$ is the left residual. Show that

$$(\mathbf{A} + \Delta)\mathbf{u} = \vartheta\mathbf{u} \quad \text{and} \quad \mathbf{w}^*(\mathbf{A} + \Delta) = \mathbf{w}^*\vartheta \quad \text{for} \quad \Delta \equiv -\mathbf{r}\mathbf{u}^* - \mathbf{w}\mathbf{s}^*.$$

(d) Compute $\|\mathbf{u}\,\mathbf{w}^*\|_2$.

(e) Compute $\|\mathbf{u}\,\mathbf{w}^* + \mathbf{w}\,\mathbf{u}^*\|_2$.

A The Hermitian eigenvalue problem

In this subsection, $\mathbf{A}$ is Hermitian ($\mathbf{A}^* = \mathbf{A}$). Hence, the eigenvalues are real. We assume that the associated eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are selected to form an orthonormal basis of $\mathbb{C}^n$ and we assume the eigenvalues to be ordered such that

$$\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n. \tag{12.1}$$

Theorem 12.2 (Courant–Fischer) *For any* $j \in \{1, \ldots, n\}$*, we have*

$$\lambda_j = \min_{\mathcal{V}} \max \left\{ \mathbf{u}^*\mathbf{A}\mathbf{u} \,\middle|\, \mathbf{u} \in \mathcal{V},\ \|\mathbf{u}\|_2 = 1 \right\}, \tag{12.2}$$

where the minimum is taken over all linear subspace $\mathcal{V}$ *of* $\mathbb{C}^n$ *of dimension* j *(or* $\geq j$*).*

The following obvious variants can sometimes be more convenient.

$$\lambda_j = \min_{\mathcal{W}} \max \left\{ \mathbf{u}^*\mathbf{A}\mathbf{u} \,\middle|\, \mathbf{u} \perp \mathcal{W},\ \|\mathbf{u}\|_2 = 1 \right\} = \max_{\mathcal{V}} \min \left\{ \mathbf{u}^*\mathbf{A}\mathbf{u} \,\middle|\, \mathbf{u} \in \mathcal{V},\ \|\mathbf{u}\|_2 = 1 \right\}, \tag{12.3}$$

where the minimum is taken over all linear subspaces $\mathcal{W}$ and $\mathcal{V}$ of $\mathbb{C}^n$ of dimension $n-j$ (or $\leq n-j$) and dimension $n-j+1$ (or $\geq n-j+1$), respectively.

Exercise 12.2. ***Proof of Theorem*** **12.2*.***

(a) If $\mathbb{J} \subset \{1,\ldots,n\}$ and $\mathbf{u} \in \text{span}\{\mathbf{v}_j \,\big|\, j \in \mathbb{J}\}, \mathbf{u} \neq \mathbf{0}$, then $\rho(\mathbf{u})$ is a convex mean (weighted average) of the λ_j with $j \in \mathbb{J}$. In particular,

$$\lambda_1 \leq \rho(\mathbf{u}) \leq \lambda_n \qquad (\mathbf{u} \in \mathbb{C}^n, \mathbf{u} \neq \mathbf{0}). \tag{12.4}$$

Let ρ_j be the value of the expression at the right-hand side of (12.2).

(b) Take $\mathcal{V} = \text{span}(\mathbf{x}_1, \ldots, \mathbf{x}_j)$ to see that $\lambda_j \geq \rho_j$.

(c) To show that $\lambda_j \leq \rho_j$ note that $\mathcal{V} \cap \text{span}(\mathbf{x}_j, \ldots, \mathbf{x}_n) \neq \{\mathbf{0}\}$ if $\mathcal{V}$ has dimension $\geq j$.

Theorem 12.3 (Sylvester's law of inertia) *With* $\text{In}(\mathbf{A})$ *the* **inertia** *of* $\mathbf{A}$*, i.e., the triple* (n_-, n_0, n_+) *of the number of negative* (n_-)*, zero* (n_0)*, and positive* (n_+) *eigenvalues, we have*

$$\text{In}(\mathbf{V}^*\mathbf{A}\mathbf{V}) = \text{In}(\mathbf{A}) \quad \textit{for all non-singular matrices } \mathbf{V}. \tag{12.5}$$

Note that $\mathbf{V}^*\mathbf{A}\mathbf{V}$ is Hermitian and that $\mathbf{V}$ is not required to be unitary.

Exercise 12.3. ***Proof of Theorem*** **12.3*.*** Let $\mathbf{V}$ be $n \times n$ non-singular.

(a) Prove $\mathcal{N}(\mathbf{A}) = \mathbf{V}(\mathcal{N}(\mathbf{V}^*\mathbf{A}\mathbf{V}))$. Here, $\mathcal{N}(\mathbf{A}) \equiv \{\mathbf{x} \,\big|\, \mathbf{A}\mathbf{x} = \mathbf{0}\}$.

Let $\mathcal{V}_-(\mathbf{A}) \equiv \text{span}\{\mathbf{x} \,\big|\, \mathbf{A}\mathbf{x} = \lambda\mathbf{x},\ \lambda < 0\}$ and $\mathcal{V}_+(\mathbf{A}) \equiv \text{span}\{\mathbf{x} \,\big|\, \mathbf{A}\mathbf{x} = \lambda\mathbf{x},\ \lambda > 0\}$.

(b) Is $\mathcal{V}_+(\mathbf{A}) = \mathbf{V}(\mathcal{V}_+(\mathbf{V}^*\mathbf{A}\mathbf{V}))$? Are all vector in $\mathcal{V}_+(\mathbf{A})$ eigenvectors of $\mathbf{A}$?
Show that $\mathcal{V}_-(\mathbf{A}) \cap \mathbf{V}(\mathcal{V}_+(\mathbf{V}^*\mathbf{A}\mathbf{V})) = \{\mathbf{0}\}$. Similarly, $\mathcal{V}_+(\mathbf{A}) \cap \mathbf{V}(\mathcal{V}_-(\mathbf{V}^*\mathbf{A}\mathbf{V})) = \{\mathbf{0}\}$.

(c) Prove $\mathbb{C}^n = \mathcal{V}_-(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}) \oplus \mathcal{V}_+(\mathbf{A})$ (with $\oplus$ as in Lecture 0.A). Prove (12.5).

(d) Let $\mathbf{A} = \mathbf{L}\mathbf{U}$ be the LU-decomposition of $\mathbf{A}$ ($\text{diag}(\mathbf{L}) = \mathbf{I}$). Show that $\text{In}(\mathbf{A}) = \text{In}(\text{diag}(\mathbf{U}))$. To simplify notation, put $\text{In}(\mathbf{U}) \equiv \text{In}(\text{diag}(\mathbf{U}))$. Note that this notation is consistent (why?).

(e) Use the above insight to make a method for computing the number of eigenvalues of $\mathbf{A}$ in real intervals $[\alpha, \beta]$. Here, $\alpha, \beta \in \mathbb{R}, \alpha < \beta$. Why might it be helpful to reduce $\mathbf{A}$ to tridiagonal form first (using Householder reflections, cf., Exercise 2.8)?

For an alternative proof of Theorem 12.3, see Exercise 12.11.

The following theorem, Cauchy's interlace Theorem, is particularly useful in case the approximate eigenpairs come from Krylov subspaces. Then $\mathbf{H}$ and $\mathbf{A}$ are the matrices of the projections onto two consecutive Krylov subspaces of the linear map of interest and the eigenvalues of $\mathbf{A}$ and $\mathbf{H}$ are Ritz values. In such a situation, the extension of a projected matrix (as $\mathbf{H}$) is obtained by extending by one new row (at the bottom) and one new column (at the right). According the Cauchy's Theorem, the Ritz values of consecutive order interlace.

Theorem 12.4 (Cauchy's interlace) *Assume* $\mathbf{A}$ *is partitioned as*

$$\mathbf{A} = \begin{bmatrix} \alpha & \mathbf{a}^* \\ \mathbf{a} & \mathbf{H} \end{bmatrix}.$$

Then the eigenvalues of $\mathbf{A}$ *and* $\mathbf{H}$ **interlace***:*

$$\lambda_1(\mathbf{A}) \leq \lambda_1(\mathbf{H}) \leq \lambda_2(\mathbf{A}) \leq \lambda_2(\mathbf{H}) \leq \ldots \leq \lambda_{n-1}(\mathbf{H}) \leq \lambda_n(\mathbf{A}).$$

Proposition 12.5 *Let* $\mathbf{b}$ *be an* n*-vector. The eigenvalues of* $\mathbf{A}$ *and the rank 1 perturbation* $\mathbf{A}' \equiv \mathbf{A} - \mathbf{b}\mathbf{b}^*$ *of* $\mathbf{A}$ *interlace:* $\lambda_1(\mathbf{A}') \leq \lambda_1(\mathbf{A}) \leq \lambda_2(\mathbf{A}') \leq \lambda_2(\mathbf{A}) \leq \ldots \leq \lambda_n(\mathbf{A}') \leq \lambda_n(\mathbf{A})$.

Exercise 12.4. ***Proof of Theorem 12.4 and of Proposition 12.5***.

(a) Use Courant–Fischer to prove Theorem 12.4.

To prove Proposition 12.5, first assume that $\mathbf{A}$ is diagonal, $\mathbf{A} = \mathrm{diag}(\mu_1, \ldots, \mu_n)$, with $\mu_j < \mu_{j+1}$ for all $j = 1, \ldots, n-1$. Put $\mathbf{b} = (b_1, \ldots, b_n)^{\mathrm{T}}$. Consider an eigenvector $\mathbf{x}$ of $\mathbf{A} - \mathbf{b}\mathbf{b}^*$ with eigenvalue λ: $(\mathbf{A} - \mathbf{b}\mathbf{b}^*)\mathbf{x} = \lambda\mathbf{x}$. With $\beta \equiv \mathbf{b}^*\mathbf{x}$, assume $\beta \neq 0$.

(b) Show that $\mathbf{x} = \beta\,(\mathbf{A} - \lambda\,\mathbf{I})^{-1}\mathbf{b}$ and $1 = \mathbf{b}^*(\mathbf{A} - \lambda\,\mathbf{I})^{-1}\mathbf{b}$, whence

$$\mathbf{x} = \beta\left(\frac{b_1}{\mu_1 - \lambda}, \ldots, \frac{b_n}{\mu_n - \lambda}\right)^{\mathrm{T}} \qquad \text{and} \qquad 1 = \sum_{j=1}^{n} \frac{|b_j|^2}{\mu_j - \lambda}. \tag{12.6}$$

Sketch a graph of the function $\zeta \rightsquigarrow \psi(\zeta) \equiv \sum_{j=1}^{n} \frac{|b_j|^2}{\mu_j - \zeta}$ ($\zeta \in \mathbb{R}$) (for, say four μ_j, that is, $n = 4$, and in case all $b_j \neq 0$). Note the vertical asymptotes at the μ_i and the horizontal one at $|\zeta| \to \infty$. Intersect this graph with the line $\zeta \rightsquigarrow 1$ to relate the eigenvalues μ_j of $\mathbf{A}$ and the eigenvalues λ_j of $\mathbf{A} + \mathbf{b}\mathbf{b}^*$. Now, show that these eigenvalues interlace.

(c) What is the effect of having some $b_j = 0$? Discuss the case where $\beta = 0$.

(d) Prove Proposition 12.5 for general Hermitian matrix $\mathbf{A}$.

(e) Relate the eigenvalues of $\mathbf{A}' \equiv \mathbf{A} + \mathbf{b}\mathbf{b}^*$ and $\mathbf{A}$.

$*$ (f) As a variant of the proof suggested in (a), Cauchy's result can also be proved with arguments as in (b): with $\mathbf{A}$ and $\mathbf{H}$ as in Theorem 12.4, assume that $\mathbf{H} = \mathrm{diag}(\mu_1, \ldots, \mu_{n-1})$. With $\mathbf{a} = (a_1, \ldots, a_{n-1})^{\mathrm{T}}$ and $\phi(\zeta) \equiv \prod_{j=1}^{n-1}(\mu_j - \zeta)$ ($\zeta \in \mathbb{R}$), prove that

$$\det(\mathbf{A} - \lambda\,\mathbf{I}) = \phi(\lambda)\,((\alpha - \lambda) - \psi(\lambda)) \quad \text{where} \quad \psi(\zeta) \equiv \sum_{j=1}^{n-1} \frac{|a_j|^2}{\mu_j - \zeta}.$$

Sketch the graph of $\zeta \rightsquigarrow \alpha - \zeta$ and of ψ in one figure. Intersect these graphs to relate the eigenvalues of $\mathbf{A}$ and the eigenvalues μ_j of $\mathbf{H}$.

The second equation in (12.6) is called the **secular equation**. This equation can be efficiently solved for λ using methods as Newton-Raphson. Note also that all eigenvectors can efficiently be computed and stored (only $3n$ scalars are required: the $b_1, \ldots, b_n$, $\mu_1, \ldots, \mu_n$, for all eigenvectors plus, per eigenvector, one additional scalar, the eigenvalue). These insights are exploited in the so-called **Divide and Conquer Algorithm** for computing all eigenpairs of Hermitian tridiagonal matrices: with a well choosen rank 1 perturbation an Hermitian tridiagonal matrix can be divided in two parts (of half dimension). With the above strategy, eigenpairs for the halfs lead to eigenpairs for the unperturbed matrix. The Divide and Conquer Algorithm is particularly efficient on parallel computers. For details, see **??**.

Exercise 12.5. **Cauchy's interlace theorem II.** Suppose $\mathbf{A}$ is partitioned as

$$\mathbf{A} = \begin{bmatrix} \mathbf{U} & \mathbf{F}^* \\ \mathbf{F} & \mathbf{H} \end{bmatrix} \qquad \text{with } \mathbf{U} \in \mathbb{M}_k,\ \mathbf{H} \in \mathbb{M}_{n-k}.$$

(a) Prove that

$$\begin{cases} \lambda_j(\mathbf{A}) \leq \lambda_j(\mathbf{U}) & \text{for all } j = 1, \ldots, k, \\ \lambda_{n-j+1}(\mathbf{A}) \geq \lambda_{k-j+1}(\mathbf{U}) & \text{for all } j = 1, \ldots, k. \end{cases}$$

(b) Let $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_k]$ be an $n \times k$ orthonormal matrix. Then $\sum_{j=1}^{k} \lambda_j \leq \sum_{j=1}^{k} \rho(\mathbf{v}_j)$. (Hint: $\sum_{j=1}^{k} \rho(\mathbf{v}_j) = \mathrm{trace}(\mathbf{V}^*\mathbf{A}\mathbf{V})$.)

(c) Let $\mathbf{u}$ be a normalised n-vector such that $\rho(\mathbf{u}) \leq \frac{1}{2}(\lambda_1 + \lambda_2)$.
Prove that $\widetilde{\mathbf{A}} \equiv (\mathbf{I} - \mathbf{u}\mathbf{u}^*)(\mathbf{A} - \rho(\mathbf{u})\mathbf{I})(\mathbf{I} - \mathbf{u}\mathbf{u}^*)$ is a positive definite map from $\mathbf{u}^\perp$ to $\mathbf{u}^\perp$.

The following tells us that $(\vartheta, \mathbf{u})$ is an accurate approximate eigenpair if the **residual** $\mathbf{A}\mathbf{u} - \vartheta\mathbf{u}$ is sufficiently small.

Theorem 12.6 *Let* $\mathbf{u}$ *be a normalised* n*-vector and let* $\vartheta \in \mathbb{C}$.
Put $\mathbf{r} \equiv \mathbf{A}\mathbf{u} - \vartheta\mathbf{u}$ *and let* λ *be the eigenvalue of* $\mathbf{A}$ *that is nearest to* ϑ. *Then [Eisberg's theorem]*

$$|\vartheta - \lambda| \leq \|\mathbf{r}\|_2. \tag{12.7}$$

Let γ *be the* **spectral gap** *for the eigenvalue* λ*:*

$$\gamma \equiv \min\{|\lambda_j - \lambda| \,\big|\, \lambda_j \neq \lambda\} \quad \textit{and} \quad \widetilde{\gamma} \equiv \min\{|\vartheta - \lambda_j| \,\big|\, \lambda_j \neq \lambda\}.$$

With $\mathcal{V}(\mu) \equiv \{\mathbf{y} \,\big|\, \mathbf{A}\mathbf{y} = \mu\mathbf{y}\}$ $(\mu \in \mathbb{C})$, *we have that*

$$\sin\angle(\mathbf{u}, \mathcal{V}(\lambda)) \leq \frac{\|\mathbf{r}\|_2}{\widetilde{\gamma}} \quad \textit{and} \quad \sin(2\angle(\mathbf{u}, \mathcal{V}(\lambda))) \leq 2\frac{\|\mathbf{r}\|_2}{\gamma}. \tag{12.8}$$

Exercise 12.6. Let $\mathbf{u}$, ϑ, $\mathbf{r}$ and λ be as Theorem 12.6.
(a) Prove **Eisberg's theorem**, i.e., (12.7).
First assume that λ is simple and take $\mathbf{x} = \mathbf{x}_{j_0}$, in particular, $\|\mathbf{x}\|_2 = 1$.
(b) Prove that $s \equiv \sin\angle(\mathbf{x}, \mathbf{u}) = \|(\mathbf{I} - \mathbf{x}\mathbf{x}^*)\mathbf{u}\|_2$ and $c \equiv \cos\angle(\mathbf{x}, \mathbf{u}) = |\mathbf{x}^*\mathbf{u}|$.
(c) Show that $\|(\mathbf{I} - \mathbf{x}\mathbf{x}^*)(\mathbf{A} - \vartheta\mathbf{I})(\mathbf{I} - \mathbf{x}\mathbf{x}^*)\mathbf{t}\|_2 \geq \widetilde{\gamma}\|\mathbf{t}\|_2$ for all $\mathbf{t} \perp \mathbf{x}$.
(d) Show that $\|(\mathbf{I} - \mathbf{u}\mathbf{u}^*)(\mathbf{I} - \mathbf{x}\mathbf{x}^*)\mathbf{t}\|_2 \geq c\|\mathbf{t}\|_2$ for all n-vectors $\mathbf{t}$.
(e) Prove (12.8). (Hint: Show first that $(\mathbf{I} - \mathbf{x}\mathbf{x}^*)(\mathbf{A} - \vartheta\mathbf{I})(\mathbf{I} - \mathbf{x}\mathbf{x}^*)\mathbf{u} = (\mathbf{I} - \mathbf{x}\mathbf{x}^*)\mathbf{r}$ and $(\mathbf{I} - \mathbf{x}\mathbf{x}^*)(\mathbf{A} - \lambda\mathbf{I})(\mathbf{I} - \mathbf{x}\mathbf{x}^*)\mathbf{u} = (\vartheta - \lambda)\mathbf{u} + \mathbf{r}$ and note that $\sin 2\angle(\mathbf{u}, \mathcal{V}(\lambda)) = 2sc$.)
(f) Prove that the simplicity restriction can be removed. (Hint: Restrict $\mathbf{A}$ to $\text{span}\{\mathbf{x}(\lambda_j) \,\big|\, j = 1, \ldots, n\}$, where $\mathbf{x}(\lambda_j)$ is the orthonormal projection of $\mathbf{u}$ onto $\mathcal{V}(\lambda_j)$.)
(g) Conclude that $(\vartheta, \mathbf{u})$ is an approximate eigenpair if the residual $\mathbf{r}$ is sufficiently small.

The accuracy of the eigenvalue is essentially the square of the size of residual if the eigenvalue is approximated by the Rayleigh quotient:

Theorem 12.7 *Let* $\mathbf{u}$ *be a normalised* n*-vector and let* $\vartheta \equiv \rho(\mathbf{u})$.
Put $\mathbf{r} \equiv \mathbf{A}\mathbf{u} - \vartheta\mathbf{u}$ *and let* λ *be the eigenvalue of* $\mathbf{A}$ *that is nearest to* ϑ. *Then*

$$2\|\mathbf{r}\|_2 < \gamma \quad \Rightarrow \quad |\vartheta - \lambda| \leq \frac{\|\mathbf{r}\|_2^2}{\gamma - \|\mathbf{r}\|_2}. \tag{12.9}$$

Exercise 12.7. Let $\mathbf{u}$, ϑ, $\mathbf{r}$ and λ be as Theorem 12.7.
First assume that λ is simple and take $\mathbf{x} = \mathbf{x}_{j_0}$.
(a) Prove $|\rho(\mathbf{u}) - \lambda| \leq \tan\angle(\mathbf{x}, \mathbf{u})\|\mathbf{r}\|_2$ (Hint: $\rho(\mathbf{u}) - \lambda = \mathbf{x}^*\mathbf{r}/\mathbf{x}^*\mathbf{u}$ and $\mathbf{r} = (\mathbf{I} - \mathbf{u}\mathbf{u}^*)\mathbf{r}$).
(b) Prove (12.9).
(c) Prove that the simplicity restriction can be removed (as in (f) of Exercise 12.6.)

The accuracy of the eigenvalue is essentially the square of the accuracy of the eigenvector. Moreover, if the Rayleigh quotient is close to an extreme (smallest or largest) eigenvalue, then then we have a good approximate eigenvector as well (be it, less accurate than the eigenvalue approximation).

Theorem 12.8 *Let* $\mathbf{u}$ *be a normalised* n*-vector and let* $\vartheta \in \mathbb{C}$.
Put $\mathbf{r} \equiv \mathbf{A}\mathbf{u} - \vartheta\mathbf{u}$ *and let* λ *be the eigenvalue of* $\mathbf{A}$ *that is nearest to* ϑ. *Then*

$$|\rho(\mathbf{u}) - \lambda| \leq \sin^2\angle(\mathbf{u}, \mathcal{V}(\lambda)) \cdot \max\{|\lambda_j - \lambda| \,\big|\, j \neq j_0\}. \tag{12.10}$$

Here, $\mathcal{V}(\mu) \equiv \{\mathbf{y} \in \mathbb{C}^n \,\big|\, \mathbf{A}\mathbf{y} = \mu\mathbf{y}\} \quad (\mu \in \mathbb{C})$.
In case $\lambda = \lambda_1$, *we have that* $0 \leq \rho(\mathbf{u}) - \lambda_1$ *and*

$$\sin^2\angle(\mathbf{u}, \mathcal{V}(\lambda_1)) \leq \frac{\rho(\mathbf{u}) - \lambda_1}{\lambda_2 - \lambda_1}. \tag{12.11}$$

If, in addition, $\rho(\mathbf{u}) < \lambda_2$, *then* (12.11) *is equivalent to*

$$\tan^2\angle(\mathbf{u}, \mathcal{V}(\lambda_1)) \leq \frac{\rho(\mathbf{u}) - \lambda_1}{\lambda_2 - \rho(\mathbf{u})}. \tag{12.12}$$

Exercise 12.8. Let $\mathbf{u}$, ϑ, $\mathbf{r}$ and λ be as Theorem 12.8. First assume λ is simple.
(a) Prove (12.10). (Hint: $\rho(\mathbf{u}) - \lambda = \mathbf{u}^*(\mathbf{A} - \lambda\mathbf{I})\mathbf{u}$.)
(b) Prove (12.11).
(c) Generalise the result to the case where λ is not simple. (Hint: see (f) of Exercise 12.6).
(d) Conclude that the accuracy of the eigenvalue approximation (using Ritz values) is basically the square of the accuracy of the eigenvector approximation.
(e) Equation (12.11) is a kind of reverse of (12.10) in case $\lambda = \lambda_1$. Prove or disprove that (12.10) has a 'reverse' in case $\lambda \neq \lambda_1$.

If $\mathcal{V}$ is a linear subspace of $\mathbb{C}^n$, then $(\vartheta, \mathbf{u})$ is a **Ritz pair** of $\mathbf{A}$ with respect to $\mathcal{V}$, if $\mathbf{u} \in \mathcal{V}, \mathbf{u} \neq \mathbf{0}$ and $\vartheta \in \mathbb{C}$ is such that $\mathbf{A}\mathbf{u} - \vartheta\mathbf{u} \perp \mathcal{V}$.

The following theorem tells us that the Ritz vector forms a good approximate eigenvector if the **search subspace** $\mathcal{V}$ contains a good approximation of the eigenvector associated to the smallest eigenvalue and the Ritz vector is associated to the smallest Ritz value. A similar statement holds with respect to the largest eigenvalue.

Theorem 12.9 *Let* $\mathcal{V}$ *be a linear subspace of* $\mathbb{C}^n$.
If λ_1 *is simple and* $\mathbf{u}_1$ *is the Ritz vector with smallest Ritz-values* ϑ_1 *w.r.t.* $\mathcal{V}$, *then*

$$0 \leq \vartheta_1 - \lambda_1 \leq \sin^2\angle(\mathbf{x}_1, \mathcal{V}) \cdot (\lambda_n - \lambda_1), \quad \sin^2\angle(\mathbf{x}_1, \mathbf{u}_1) \leq \sin^2\angle(\mathbf{x}_1, \mathcal{V}) \cdot \frac{\lambda_n - \lambda_1}{\lambda_2 - \lambda_1}. \tag{12.13}$$

Exercise 12.9. Let $\mathcal{V}$ be as in Theorem 12.9.
(a) Show that Ritz values are the Rayleigh quotients, i.e., $\vartheta = \rho(\mathbf{u})$ if $(\vartheta, \mathbf{u})$ is a Ritz pair.
(b) Prove (12.13). (Hint: let $\mathbf{x}_V$ be the orthonormal projection of $\mathbf{x}_1$ onto $\mathcal{V}$. Use Courant-Fischer to show that $\vartheta_1 \leq \rho(\mathbf{x}_V)$. Then use Theorem 12.8.)

B Hermitian matrices with Hermitian perturbations

In this subsection, $\mathbf{A}$ is Hermitian, Δ is $n \times n$ Hermitian perturbation of $\mathbf{A}$. The eigenvalues of $\mathbf{A}$ as well $\mathbf{A} + \Delta$ are ordered in increasing magnitude (cf., (12.1)).

The following result follows from an application of the Implicit Function Theorem. You can use this result; you do not have to prove it here.

Theorem 12.10 *If* F *is an analytic map on a neighbourhood* $\mathcal{U}$ *of a real interval* $[\alpha, \beta]$ *such that* $F(\tau)^* = F(\bar{\tau}) \in \mathbb{M}_n \quad (\tau \in \mathcal{U})$, *then there* n *complex-valued functions* μ_j *on some neighbourhood* $\mathcal{U}'$ *of* $[\alpha, \beta]$ *such that, for each* $\tau \in \mathcal{U}'$, $\mu_1(\tau), \mu_2(\tau), \ldots, \mu_n(\tau)$ *are the eigenvalues of* $F(\tau)$ *counted according to multiplicity.*

Exercise 12.10. Let $\mathbf{x} \in \mathbb{C}^n, \mathbf{x} \neq \mathbf{0}$.

(a) Prove that

$$t \rightsquigarrow \mathbf{x}^*(\mathbf{A} - t\mathbf{I})^{-1}\mathbf{x}$$

is strictly increasing on any interval of $\mathbb{R}$ that does not contain an eigenvalue of $\mathbf{A}$.

Exercise 12.11. Here, we will see an alternative prove of Theorem 12.3 (cf., Exercise 12.3). Let $\mathbf{V}$ be an $n \times n$ non-singular matrix.

(a) Argue that there is a $\zeta \in \mathbb{C}, |\zeta| = 1$ such that $\widetilde{\mathbf{V}} \equiv \zeta\mathbf{V}$ does not have real eigenvalues.

(b) Consider the map $F(\tau) \equiv (\tau\widetilde{\mathbf{V}}^* + 1 - \tau)\mathbf{A}(\tau\widetilde{\mathbf{V}} + 1 - \tau)$ $(\tau \in \mathbb{C})$. Show that the (matrix-valued) map is analytic. Moreover, $F(\tau)^* = F(\bar{\tau})$. This implies that the eigenvalues (counted to multiplicity) depend continuously on τ for $\tau \in [0, 1]$; you can use this without proving it.

(c) Prove (12.5) (you may use (a) of Exercise 12.3).

Exercise 12.12.

(a) Show that the functions $\tau \rightsquigarrow \lambda_j(\mathbf{A} + \tau\Delta)$ are continuous.

(b) Show that the function $\tau \rightsquigarrow \lambda_j(\mathbf{A} + \tau\Delta)$ is analytic round τ_0 if $\lambda_j(\mathbf{A} + \tau_0\Delta)$ is a simple eigenvalue of $\mathbf{A} + \tau_0\Delta$. Is the function also analytic round τ_0 if the eigenvalue is not simple?

(c) Prove that the function $\tau \rightsquigarrow \lambda(\tau) \equiv \lambda_1(\mathbf{A} + \tau\Delta)$ is **convex**, i.e.,

$$\alpha\lambda(\tau_0) + (1 - \alpha)\lambda(\tau_1) \leq \lambda(\alpha\tau_0 + (1 - \alpha)\tau_1) \quad \text{for all } \alpha \in [0, 1] \text{ and all } \tau_1 \leq \tau_2.$$

(Hint: $\lambda(\tau) = \min \mathbf{u}^*(\mathbf{A} + \tau\Delta)\mathbf{u}$, where the minimum is taken over all normalised $\mathbf{u}$.)

Theorem 12.11 (Weyl) *For all $j \in \{1, 2, \ldots, n\}$, we have*

$$-\|\Delta\|_2 \leq \lambda_1(\Delta) \leq \lambda_j(\mathbf{A} + \Delta) - \lambda_j(\mathbf{A}) \leq \lambda_n(\Delta) \leq \|\Delta\|_2. \tag{12.14}$$

Exercise 12.13. ***Proof of Theorem*** **12.11*.*** Assume Δ is $n \times n$ Hermitian.

(a) Prove Theorem 12.11.

(b) Prove the more general result

$$\lambda_{k+m-n}(\mathbf{A} + \Delta) \leq \lambda_m(\mathbf{A}) + \lambda_k(\Delta) \leq \lambda_{k+m-1}(\mathbf{A} + \Delta),$$

with k and m such that the λ_i's in the claim are defined.

Theorem 12.12 (Wielandt–Hofman) *We have that*

$$\sqrt{\sum_{j=1}^{n} |\lambda_j(\mathbf{A} + \Delta) - \lambda_j(\mathbf{A})|^2} \leq \|\Delta\|_{\mathrm{F}}. \tag{12.15}$$

Exercise 12.14.

(a) Compare Theorem 12.12 and Theorem 12.11.

We prove Theorem 12.12.

Let $\mathbf{D} \equiv \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $\mathbf{D}' \equiv \mathrm{diag}(\mu_1, \ldots, \mu_n)$ such that $\lambda_1 < \lambda_2 < \ldots < \lambda_n$ and $\mu_1 < \mu_2 < \ldots < \mu_n$. Consider the following statement

$$\|\mathbf{D} - \mathbf{D}'\|_{\mathrm{F}} < \|\mathbf{D} - \mathbf{Q}^*\mathbf{D}'\mathbf{Q}\|_{\mathrm{F}} \quad \text{for all unitary } n \times n \text{ matrix } \mathbf{Q}. \tag{12.16}$$

(b) Show that (12.16) implies Theorem 12.12.

(c) Show that $\mathbf{Q}_0 = \mathrm{argmin}\{\|\mathbf{D} - \mathbf{Q}^*\mathbf{D}'\mathbf{Q}\|_{\mathrm{F}} \,\big|\, \mathbf{Q} \text{ is unitary } n \times n\}$ exists.

Put $\mathbf{H} \equiv \mathbf{Q}_0^*\mathbf{D}'\mathbf{Q}_0$.

(d) Prove $\mathbf{H}$ is diagonal with increasing diagonal entries in case $n = 2$.

(e) Assume the $(1,2)$-entry of $\mathbf{H}$ is non-zero.
Show there is a rotation $\mathbf{G}$ in the $(1,2)$-plane (i.e., span$(\mathbf{e}_1, \mathbf{e}_2)$: $\mathbf{G}$ is a Givens rotation, see Exercise 3.5) such that the 2×2 left upper block of $\mathbf{G}^*\mathbf{H}\mathbf{G}$ is diagonal with diagonal entries in increasing order.
Show that $\|\mathbf{H}(J,[1,2])\mathbf{G}([1,2],[1,2])\|_{\mathrm{F}} = \|\mathbf{H}(J,[1,2])\|_{\mathrm{F}}$ (see Exercise 4.14(a)). Here, we used MATLAB notation to denote submatrices and $J \equiv [3,4,\ldots,n]$. Conclude that the Frobenius norm of the part of the matrices $\mathbf{H}$ and $\mathbf{G}^*\mathbf{H}\mathbf{G}$ outside the 2×2 upper block are the same.
Conclude that the $(1,2)$-entry of $\mathbf{H}$ is zero.

(f) Prove that $\mathbf{H} = \mathbf{D}'$. Hence, (12.16) is correct.

In case of a cluster of eigenvalues, the eigenvector with eigenvalues in this cluster are very sensitive the perturbations. In such a case, it is more useful to analyse the effect on perturbation on the space spanned by all eigenvalues in the cluster.

Theorem 12.13 (Davis–Kahan) *Assume $\alpha, \beta, \rho \in \mathbb{R}, \alpha < \beta, \rho > 0$ are such that*

$$\lambda_j(\mathbf{A}) \in (-\infty, \alpha - \rho] \cup [\alpha, \beta] \cup [\beta + \rho, \infty) \quad \textit{and} \quad \|\Delta\|_2 < \rho.$$

Let $\mathcal{V}$ be the space spanned by all eigenvectors of $\mathbf{A}$ with eigenvalue in $[\alpha, \beta]$,
let $\widetilde{\mathcal{V}}$ be the space spanned by all eigenvectors of $\mathbf{A} + \Delta$ with eigenvalue in $[\alpha - \|\Delta\|_2, \beta + \|\Delta\|_2]$. Then

$$\sin(\mathcal{V}, \widetilde{\mathcal{V}}) \le \frac{\|\Delta\|_2}{\rho - \|\Delta\|_2}.$$

Exercise 12.15.

(a) Compare the result of Davis–Kahan with (12.8) in Theorem 12.6.

(b) Give an example to illustrate that it is not useful to try to analyse the effect of perturbations on eigenvectors with eigenvalues in a cluster.

Consider the situation of Theorem 12.13.

(c) Prove that $\lambda_j(\mathbf{A}) \in [\alpha, \beta] \quad \Leftrightarrow \quad \lambda_j(\mathbf{A} + \Delta) \in [\alpha - \|\Delta\|_2, \beta + \|\Delta\|_2]$.
Conclude that $\mathcal{V}$ and $\widetilde{\mathcal{V}}$ have the same dimension.

C Hermitian matrices with non-Hermitian perturbations

In this subsection, $\mathbf{A}$ is Hermitian with eigenvalues ordered in increasing magnitude (cf., (12.1)). Δ is $n \times n$ perturbation of $\mathbf{A}$.

Exercise 12.16. We will show that

$$\gamma_n \le \log_2(2n), \quad \text{where } \|\mathbf{U} - \mathbf{L}\|_2 \le \gamma_n \|\mathbf{A}\|_2 \quad \forall \mathbf{A} \in \mathbb{M}_n.$$

Here $\mathbf{U}$ is the strict upper triangular part of the matrix $\mathbf{A}$ and $\mathbf{L}$ is the strict lower triangular part.

With $k = 2^\ell < n$ and ℓ such that $n \le 2^{l+1}$, let the $n \times n$ matrix $\mathbf{B}$ be partitioned as

$$\mathbf{B} = \begin{bmatrix} \mathbf{E} & \mathbf{F}^* \\ \mathbf{G} & \mathbf{K} \end{bmatrix} \quad \text{with } \mathbf{E} \in \mathbb{M}_k, \ \mathbf{K} \in \mathbb{M}_{n-k}.$$

(a) Prove that

$$\max(\|\mathbf{E}\|_2, \|\mathbf{F}\|_2, \|\mathbf{G}\|_2, \|\mathbf{k}\|_2) \le \|\mathbf{B}\|_2 \le \max(\|\mathbf{E}\|_2, \|\mathbf{K}\|_2) + \max(\|\mathbf{F}\|_2, \|\mathbf{G}\|_2).$$

(b) For an $n \times n$ matrix $\mathbf{A}$, let $\Pi(\mathbf{A}) \equiv \mathbf{U} - \mathbf{L}$, where $\mathbf{U}$ is the strict upper triangular part of the matrix $\mathbf{A}$ and $\mathbf{L}$ is the strict lower triangular part. Note that

$$\Pi(\mathbf{B}) = \begin{bmatrix} \Pi(\mathbf{E}) & \mathbf{F}^* \\ -\mathbf{G} & \Pi(\mathbf{K}) \end{bmatrix}$$

and conclude that

$$\|\Pi(\mathbf{B})\|_2 \leq (\gamma_\ell + 1) \max(\|\mathbf{E}\|_2, \|\mathbf{K}\|_2) + \max(\|\mathbf{F}\|_2, \|\mathbf{G}\|_2) \leq (\gamma_\ell + 1)\|\mathbf{B}\|_2.$$

Conclude that $\gamma_{\ell+1} \leq \gamma_\ell + 1$ and $\gamma_n \leq \log_2(2n)$.

Exercise 12.17. **Weyl's theorem for non-Hermitian perturbations.** $\mathbf{A}$ and Δ are $n \times n$, $\mathbf{A}$ is Hermitian, Δ is possibly non-Hermitian.
Let $\lambda_j(\mathbf{A} + \Delta)$ be ordered such that

$$\mathrm{Re}(\lambda_1(\mathbf{A} + \Delta)) \leq \mathrm{Re}(\lambda_2(\mathbf{A} + \Delta)) \leq \ldots \leq \mathrm{Re}(\lambda_n(\mathbf{A} + \Delta)).$$

We will show that, for each $j = 1, \ldots, n$

$$|\lambda_j(\mathbf{A}) - \mathrm{Re}(\lambda_j(\mathbf{A} + \Delta))| \leq (\log_2(4n))\, \|\Delta\|_2 \quad \text{and} \quad |\mathrm{Im}(\lambda_j(\mathbf{A} + \Delta))| \leq \|\Delta\|_2.$$

According to Schur's theorem,

$$\mathbf{A} + \Delta = \mathbf{Q}(\mathbf{D} + \mathbf{U})\mathbf{Q},$$

for some unitary matrix $\mathbf{Q}$, an strictly upper triangular matrix $\mathbf{U}$ and a diagonal matrix $\mathbf{D}$; $\mathbf{D} = \mathrm{diag}(\lambda_1(\mathbf{A} + \Delta), \ldots, \lambda_n(\mathbf{A} + \Delta))$.
Put $\mathbf{H} \equiv \frac{1}{2}(\Delta + \Delta^*)$, $\mathbf{W} \equiv \frac{1}{2}(\mathbf{U} + \mathbf{U}^*)$, and $\mathbf{D}_r \equiv \frac{1}{2}(\mathbf{D} + \overline{\mathbf{D}})$.
(a) Prove that

$$|\lambda_j(\mathbf{A}) - \mathrm{Re}(\lambda_j(\mathbf{A} + \Delta))| \leq \|\mathbf{W}\|_2 + \|\mathbf{H}\|_2 \quad \text{and} \quad \|\mathbf{W}\|_2 \leq \log_2(2n)\, \|\Delta\|_2$$

(b) Prove that

$$|\mathrm{Im}(\lambda_j(\mathbf{A} + \Delta))| \leq \|\mathbf{H}\|_2.$$

D Perturbed eigenvalue problems for general matrices

In this subsection, $\mathbf{A}$ and Δ are $n \times n$ matrices.

Theorem 12.14 (Bauer–Fike) *Suppose $\mathbf{A}$ is diagonizable.*
Let $\mathbf{V}$ be a non-singular matrix that diagonalises $\mathbf{A}$. Put $\mathcal{C}_E \equiv \|\mathbf{V}^{-1}\|_2\, \|\mathbf{V}\|_2$.
For each eigenvalue ϑ of $\mathbf{A} + \Delta$, there is an eigenvalue λ of $\mathbf{A}$ such that $|\vartheta - \lambda| \leq \mathcal{C}_E\, \|\Delta\|_2$.
If $\mathbf{A}$ is Hermitian, then there is a unitary $\mathbf{V}$ and $\mathcal{C}_E$ can be taken 1.

Note that $\mathcal{C}_E = 1$ in case $\mathbf{A}$ is normal.
The theorem is well-known for this particular situation of a normal matrix and with perturbation equal to $\mathbf{r}^*\mathbf{u}$, with $\mathbf{r} \equiv \mathbf{A}\mathbf{u} - \vartheta\mathbf{u}$. Then the theorem tells us (cf., Theorem 12.6, inequality (12.7))

Corollary 12.15 (Eisberg's theorem) *If $\mathbf{A}$ is normal, $(\vartheta, \mathbf{u})$ is an approximate eigenpair with $\|\mathbf{u}\|_2 = 1$, and $\mathbf{r} \equiv \mathbf{A}\mathbf{u} - \vartheta\mathbf{u}$, then there is an eigenvalue λ of $\mathbf{A}$ such that*

$$|\lambda - \vartheta| \leq \|\mathbf{r}\|_2.$$

Exercise 12.18. ***Proof of Theorem*** 12.14.

(a) Prove Theorem 12.14.
Hint: If $(\mathbf{A}+\Delta)\mathbf{u}=\vartheta\mathbf{u}$ with $\|\mathbf{u}\|_2=1$, then $(\mathbf{A}-\vartheta\mathbf{I})\mathbf{u}=-\Delta\mathbf{u}$ and

$$1=\|\mathbf{u}\|_2=\|(\mathbf{A}-\vartheta\mathbf{I})^{-1}\Delta\mathbf{u}\|_2\leq\|(\mathbf{A}-\vartheta\mathbf{I})^{-1}\|_2\,\|\Delta\|_2.$$

(b) Prove Eisberg's theorem (Cor. 12.15).

The matrix $\mathbf{V}$, that is, a basis of eigenvectors, is usually not available. Moreover, if $\mathcal{C}_E$ is large, then the estimate in Bauer–Fike's theorem may suggest that all eigenvalues of $\mathbf{A}$ are sensitive to perturbations, which, in general, is not correct, as we will learn in Theorem 12.19 below (see (12.21)). If $\mathbf{A}$ is not diagonalizable, $\mathbf{V}$ does even not exist. The following theorem covers this situation.

Theorem 12.16 (Henrici) *Let* $\mathbf{A}$ *be an* $n\times n$ *matrix with Schur decomposition* $\mathbf{A}=\mathbf{Q}=\mathbf{QS}$. *Write* $\mathbf{S}=\Lambda+\mathbf{U}$ *with* Λ *diagonal and* $\mathbf{U}$ *strict upper triangular.*
For some $p\leq n$*, and for each eigenvalue* μ *of* $\mathbf{A}+\Delta$*, there is an eigenvalue* λ *of* $\mathbf{A}$ *such that*

$$|\lambda-\mu|\leq\max(\delta,\sqrt[p]{\delta}),\quad \textit{where}\quad \delta\equiv\|\Delta\|_2\sum_{j<p}\|\mathbf{U}\|_2^j.$$

Exercise 12.19. ***Proof of Theorem*** 12.16. Consider the situation of Theorem 12.16. Prove the following claims.

(a) There is a $p\in\mathbb{N}$, $p\leq n$ such that $\mathbf{U}^p=\mathbf{0}$. Let p be the smallest integer.

(b) $(\mathbf{D}-\mathbf{N})^{-1}=\sum_{j<p}(\mathbf{D}^{-1}\mathbf{U})^j\mathbf{D}^{-1}$ for any $n\times n$ diagonal matrix $\mathbf{D}$.

(c) If μ is such that $\|(\Lambda-\mu\mathbf{I}+\mathbf{U})^{-1}\|_2\,\|\Delta\|_2<1$, then $\Lambda+\mathbf{U}-\mu\mathbf{I}$ is non-singular and μ is not an eigenvalue of $\mathbf{A}+\Delta$.

(d) Theorem 12.16 holds.

Henrici's theorem is mainly of theoretical interest: it proves that the eigenvalues depend continuously on perturbations (as stated in Theorem 1.13).

Bauer–Fike nor Henrici's theorem give us readily insight in the sensitivity of eigenvalues to perturbations. Pseudo-eigenvalues do. For non high dimensional matrices, pseudo-eigenvalues can be plotted in the complex plane. These plots show for each individual eigenvalue how sensitive they are to perturbations.

Consider an $n\times n$ matrix $\mathbf{A}$ and an $\varepsilon>0$.
A complex number μ is an ε-**pseudo-eigenvalue** if it is an eigenvalue of $\mathbf{A}+\Delta$ for some $n\times n$ matrix Δ with $\|\Delta\|_2\leq\varepsilon$. $\Lambda_\varepsilon(\mathbf{A})$ is the ε-**pseudo-spectrum** of $\mathbf{A}$, that is, the collection of ε-pseudo-eigenvalue of $\mathbf{A}$.

Theorem 12.17 *Let* $\mathbf{A}$ *be an* $n\times n$ *matrix and let* $\varepsilon>0$.
a) *The following properties are equivalent for a* $\mu\in\mathbb{C}$.

- $\mu\in\Lambda_\varepsilon(\mathbf{A})$
- $\|(\mathbf{A}-\mu\,\mathbf{I})^{-1}\|_2\geq\frac{1}{\varepsilon}$.
- *The smallest singular value of* $\mathbf{A}-\mu\,\mathbf{I}$ *is at most* ε.

b) *If* $|\mu-\lambda|\leq\varepsilon$ *for some* $\lambda\in\Lambda(\mathbf{A})$*, then* $\mu\in\Lambda_\varepsilon(\mathbf{A})$.

c) *If* $(\vartheta,\mathbf{u})$ *is an approximate eigenpair of* $\mathbf{A}$ *and* $\frac{\|\mathbf{Au}-\vartheta\mathbf{u}\|_2}{\|\mathbf{u}\|_2}\leq\varepsilon$*, then* $\vartheta\in\Lambda_\varepsilon(\mathbf{A})$.

The matrix-valued function $\mathbf{R}$ on $\mathbb{C}$, defined by $\mathbf{R}(\mu)\equiv(\mathbf{A}-\mu\mathbf{I})^{-1}$ $(\mu\in\mathbb{C})$ (cf., Theorem 12.17.a), is called the **resolvent** of $\mathbf{A}$.

Exercise 12.20. ***Proof of Theorem*** 12.17.

(a) Use Theorem 12.1 to prove b) and c) of Theorem 12.17.

(b) If $\|(\mathbf{A}-\mu\,\mathbf{I})^{-1}\Delta\|_2<1$, then μ is not an eigenvalue of $\mathbf{A}+\Delta$.
Conclude that $\|(\mathbf{A}-\mu\,\mathbf{I})^{-1}\|_2\geq\frac{1}{\varepsilon}$ if $\|(\mathbf{A}-\mu\,\mathbf{I})^{-1}\|_2\geq\frac{1}{\varepsilon}$.
(c) If $\|(\mathbf{A}-\mu\,\mathbf{I})^{-1}\|_2\geq\frac{1}{\varepsilon}$, then $\|(\mathbf{A}-\mu\,\mathbf{I})^{-1}\mathbf{w}\|_2\geq\frac{1}{\varepsilon}\|\mathbf{w}\|_2$ for some non-trivial vector $\mathbf{w}$. Consider $\mathbf{u}\equiv(\mathbf{A}-\mu\,\mathbf{I})\mathbf{w}$ and conclude that $\|(\mathbf{A}-\mu\,\mathbf{I})^{-1}\|_2\geq\frac{1}{\varepsilon}$.
(d) Complete the proof of Theorem 12.17.

Pseudo-eigenvalues suggest a graphical way to obtain insight in the sensitivity of eigenvalues to perturbations. For $\mu\in\mathbb{C}$, let $\sigma(\mu)$ be the smallest singular value of $\mathbf{A}-\mu\,\mathbf{I}$. For various values of ε, say, $\varepsilon=10^j$ for $j=-14,-13,\ldots,-8$, plot the ε-level curves $\{\mu\in\mathbb{C}\,\big|\,\sigma(\mu)=\varepsilon\}$ of $\sigma(\mu)$ in an part of the complex plane that contains the eigenvalues of $\mathbf{A}$ that are of interest (or plot the spectrum of $\mathbf{A}+\Delta$ for various perturbations Δ of size ε).[1] Around eigenvalues that are not so sensitive, 'circles' will arise with radius of order ε, whereas the sensitive eigenvalues are in 'huge areas'.

Exercise 12.21. Let $\mathbf{S}$ be the $n\times n$ shift matrix: $\mathbf{S}e_i=\mathbf{e}_{i-1}$ for $i=1,\ldots,n-1$.
(a) Compute the spectrum and the ε-pseudo-spectrum of $\mathbf{S}$.
(b) Conclude that for $\mathbf{S}$ and $n\geq 30$, the ε-pseudo-spectrum does not really depend on ε.

Unfortunately, plots of pseudo-eigenvalues are very computationally intensive and are only feasible for matrices of modest dimension. More quantitative statements that are also applicable to high dimensional matrices rely on the following extension of Theorem 1.13. You can use this result. It follows from an application of the Implicit Function Theorem. We will not give details here.

Theorem 12.18 *Assume $\mathbf{F}(\tau)$ is an $n\times n$ matrix for all τ in some subset $\mathcal{I}$ of $\mathbb{C}$ that depends continuously on τ. Let $\mu_1,\ldots,\mu_n$ be continuous complex-valued functions on $\mathcal{I}$ such that $\mu_1(\tau),\ldots,\mu_n(\tau)$ are the eigenvalues of $\mathbf{F}(\tau)$ counted according to multiplicity ($\tau\in\mathcal{I}$).*
If $\mathbf{F}$ is analytic on some neighbourhood of τ_0 in $\mathbb{C}$ and $\mu_j(\tau_0)$ is a simple eigenvalue of $\mathbf{F}(\tau_0)$, then, on some neighbourhood $\mathcal{U}$ of τ_0, μ_j is analytic and for some n-vector valued function $\mathbf{x}_j$ that is analytic on $\mathcal{U}$ we have that $\mathbf{F}(\tau)\mathbf{x}_j(\tau)=\mu_j(\tau)\mathbf{x}_j(\tau)\quad(\tau\in\mathcal{U})$.

Exercise 12.22. Is Theorem 12.18 also correct in case $\mu_j(\tau_0)$ is not simple?

Exercise 12.23. Consider the eigenvalue problem

$$(\mathbf{A}+\tau\Delta)\mathbf{x}_\tau=\lambda_\tau\mathbf{x}_\tau,\quad\text{where}\quad \mathbf{A}=\begin{bmatrix}\mu & \mathbf{a}^*\\ \mathbf{0} & \mathbf{B}\end{bmatrix},\ \Delta=\begin{bmatrix}\nu & \mathbf{f}^*\\ \mathbf{r} & \mathbf{F}\end{bmatrix},\ \mathbf{x}_\tau=\begin{bmatrix}1\\ \mathbf{z}_\tau\end{bmatrix}.$$

Here, $\mathbf{B}$ and $\mathbf{F}$ are $(n-1)\times(n-1)$ matrices and $\mathbf{0},\mathbf{a},\mathbf{f},\mathbf{r}$ and $\mathbf{z}_\tau$ are $(n-1)$-vectors, $\mu,\nu,\tau,\lambda_\tau\in\mathbb{C}$, μ is not an eigenvalue of $\mathbf{B}$.
(a) Note that $\mathbf{u}\equiv\mathbf{e}_1$ is a eigenvector of $\mathbf{A}$ with eigenvalue $\lambda_0=\mu$, μ is a simple eigenvalue of $\mathbf{A}$. Prove that $\|\mathbf{z}_\tau\|_2=\tan\angle(\mathbf{u},\mathbf{x}_\tau)$.
(b) For $|\tau|$ small enough, we may assume λ_τ and $\mathbf{z}_\tau$ to be analytic. Why?
Put $\mathbf{t}\equiv(\mathbf{B}-\mu\mathbf{I})^{-1}\mathbf{r}$ and $\gamma\equiv\|(\mathbf{B}-\mu\mathbf{I})^{-1}\|_2^{-1}$.
(c) Prove that $\|\mathbf{t}\|_2\leq\frac{\|\mathbf{r}\|_2}{\gamma}$.
(d) Show that $\gamma=\min\{|\mu-\lambda_j(\mathbf{A})|\,\big|\,\lambda_j(\mathbf{A})\neq\mu\}$ in case $\mathbf{B}$ is Hermitian.
Put $\phi(\tau)\equiv\lambda_\tau-\mu-\tau\nu$.
(e) Prove that $\phi(\tau)=\mathbf{a}^*\mathbf{z}_\tau+\tau\mathbf{f}^*\mathbf{z}_\tau$ and $(\mathbf{B}-\mu\mathbf{I})\mathbf{z}_\tau=-\tau\mathbf{r}-\tau(\mathbf{F}-\nu\mathbf{I})\mathbf{z}_\tau+\phi(\tau)\mathbf{z}_\tau$.
Put $\alpha\equiv\|\mathbf{a}\|_2$ and $\delta\equiv\|\Delta-\nu\mathbf{I}\|_2$.

[1] In practise, the smallest singular value $\sigma(\mu)$ is computed for all values μ in a grid in a part of $\mathbb{C}$. Then numerical interpolation procedures are applied to find the ε-level curves.

(f) Show that $|\phi(\tau)| \leq (\alpha + \tau\delta)\|\mathbf{z}_\tau\|_2$ and $\|\mathbf{z}_\tau\|_2 \leq \frac{\tau\,\|\mathbf{r}\|_2}{\gamma - \tau\delta - |\phi(\tau)|}$ for τ sufficiently small.

(g) If, for some $\omega, \beta > 0$ and for $t \in [0, \frac{1}{2}\omega]$ we have that $t(\omega - t) \leq \beta$, then $t \leq \frac{\beta}{\omega - \beta}$.
Prove this claim and conclude that $\phi(\tau) = \mathcal{O}(\tau)$ and, if $\mathbf{a} = \mathbf{0}$, $\phi(\tau) = \mathcal{O}(\tau^2)$ for $\tau \to 0$. Note that the order terms can be quantified.

(h) Prove that, for $\tau \to 0$,

$$\begin{cases} \mathbf{z}_\tau = -\tau(\mathbf{B} - \mu\mathbf{I})^{-1}\mathbf{r} + \mathcal{O}(\tau^2), \\ \lambda_\tau = \mu + \tau\nu + \tau\mathbf{a}^*(\mathbf{B} - \mu\mathbf{I})^{-1}\mathbf{r} + \mathcal{O}(\tau^2), \\ \lambda_\tau = \mu + \tau\nu + \tau^2\mathbf{f}^*(\mathbf{B} - \mu\mathbf{I})^{-1}\mathbf{r} + \mathcal{O}(\tau^3) \quad \text{if } \mathbf{a} = \mathbf{0}. \end{cases} \tag{12.17}$$

Theorem 12.19 *Let μ be a simple eigenvalue of $\mathbf{A}$ with normalised eigenvector $\mathbf{u}$.*
Let $(\lambda, \mathbf{x})$ be the eigenpair of $\mathbf{A} + \Delta$ with λ nearest to μ.
Put

$$\mathbf{A}' \equiv (\mathbf{I} - \mathbf{u}\mathbf{u}^*)(\mathbf{A} - \mu\mathbf{I})(\mathbf{I} - \mathbf{u}\mathbf{u}^*) \quad \textit{and} \quad \gamma \equiv \min\{\|\mathbf{A}'\mathbf{t}\|_2 \,\big|\, \mathbf{t} \perp \mathbf{u}, \|\mathbf{t}\|_2 = 1\}. \tag{12.18}$$

0) *Then $\gamma > 0$.*

We assume $\|\Delta\|_2$ to be sufficiently small.

1) *We have*

$$\mathbf{x} = \mathbf{u} + \mathbf{t} + \mathcal{O}(\|\Delta\|_2^2) \quad \textit{and} \quad \tan\angle(\mathbf{x}, \mathbf{u}) \leq \frac{\|\Delta\|_2}{\gamma} + \mathcal{O}(\|\Delta\|_2^2). \tag{12.19}$$

Here,

$$\mathbf{t} \perp \mathbf{u} \quad \textit{solves} \quad \mathbf{A}'\mathbf{t} = -\mathbf{r} \equiv -(\mathbf{I} - \mathbf{u}\mathbf{u}^*)\Delta\mathbf{u}. \tag{12.20}$$

2) *With $\mathbf{w}$ a left eigenvector of $\mathbf{A}$ associated to μ, we have that*

$$\lambda = \mu + \frac{\mathbf{w}^*\Delta\mathbf{u}}{\mathbf{w}^*\mathbf{u}} + \mathcal{O}(\|\Delta\|_2^2) \quad \textit{and} \quad |\lambda - \mu| \leq \frac{\|\Delta\|_2}{\cos\angle(\mathbf{u}, \mathbf{w})} + \mathcal{O}(\|\Delta\|_2^2). \tag{12.21}$$

In particular,
- *for the* **condition number $\mathcal{C}_A(\mathbf{u})$ of the eigenvector $\mathbf{u}$** *of $\mathbf{A}$ and*
- *for the* **condition number $\mathcal{C}_A(\mu)$ of the eigenvalue** μ *of $\mathbf{A}$,*

we have

$$\mathcal{C}_A(\mathbf{u}) = \frac{1}{\gamma} \quad \textit{and} \quad \mathcal{C}_A(\mu) = \frac{\|\mathbf{w}\|_2\,\|\mathbf{u}\|_2}{|\mathbf{w}^*\mathbf{u}|} = \frac{1}{\cos(\mathbf{u}, \mathbf{w})}.$$

Exercise 12.24. Here, we prove Theorem 12.19. We use the results of Exercise 12.23.

(a) Prove that $\gamma > 0$.

(b) Prove (12.19). (Hint: represent the perturbed eigenvalue problem with respect to a basis $\mathbf{u}, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$. Here $\mathbf{v}_2, \ldots, \mathbf{v}_n$ is an orthonormal basis of $\mathbf{u}^\perp$. Then, use (12.17).)

(c) Let $\mathbf{V} = [\mathbf{v}_2, \ldots, \mathbf{v}_n]$ be an orthonormal matrix spanning $\mathbf{w}^\perp$. Show that

$$[\mathbf{u}, \mathbf{V}]^{-1} = \begin{bmatrix} \frac{1}{\mathbf{w}^*\mathbf{u}}\mathbf{w}^* \\ \mathbf{V}^*(\mathbf{I} - \frac{\mathbf{u}\mathbf{w}^*}{\mathbf{w}^*\mathbf{u}}) \end{bmatrix}.$$

(d) Prove (12.21). (Hint: represent the perturbed eigenvalue problem with respect to a basis $\mathbf{u}, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$. Here $\mathbf{v}_2, \ldots, \mathbf{v}_n$ is an orthonormal basis of $\mathbf{w}^\perp$. Note that, with this choice of basis, $\mathbf{a} = \mathbf{0}$. Now, use (12.17).)

Exercise 12.25. Let $\mathbf{u}$ be a normalised approximate eigenvector.
With $\mu \equiv \mathbf{u}^*\mathbf{A}\mathbf{u}$, $\mathbf{r} \equiv \mathbf{A}\mathbf{u} - \mu\mathbf{u}$, and $\Delta \equiv \mathbf{r}\mathbf{u}^*$ we have that $(\mathbf{A} - \Delta)\mathbf{u} = \mu\mathbf{u}$ (cf., Th. 12.1).

(a) Represent $\mathbf{A} - \Delta$ and Δ with respect to an orthonormal basis $\mathbf{u}, \mathbf{v}_2, \ldots, \mathbf{v}_n$. Show that the eigenvalue problem $([\mathbf{A} - \Delta] + \Delta)\mathbf{x} = \lambda\mathbf{x}$ is in the form of Exercise 12.23 (with $\tau = 1$) with $\nu = 0$, $\mathbf{F} = \mathbf{0}$, and $\mathbf{f} = \mathbf{0}$ (if $\Delta \equiv \mathbf{r}\mathbf{u}^* + \mathbf{u}\mathbf{r}^*$ then $\mathbf{f} = \mathbf{r}$).
Can we improve the results in (12.17) for this special situation?

E Krylov search subspaces

In this subsection, we assume that $\mathbf{A}$ is an Hermitian $n \times n$ matrix and we will consider approximate eigenpairs from a sequence of Krylov subspaces $\mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$. In particular, we will continue our discussion on CG as a deflated process (see Prop. 10.2 with preceding text and subsequent exercise). Note that the Krylov subspaces are nested and that the extension when increasing k to $k+1$ is with a vector of the form $\mathbf{A}\mathbf{v}_k$.

Let $(\lambda_1, \mathbf{x}_1), \ldots, (\lambda_n, \mathbf{x}_n)$ be the eigenpairs of $\mathbf{A}$ with $\|\mathbf{x}_i\|_2 = 1$ and increasingly ordered eigenvalues, i.e., $\lambda_1 < \ldots < \lambda_n$. To avoid technical details, we assume that all eigenvalues are simple and the dimension of $\mathcal{K}_n(\mathbf{A}, \mathbf{r}_0)$ is n. For $k \in \{1, 2, \ldots, n\}$, let $(\vartheta_1^{(k)}, \mathbf{u}_1^{(k)}), \ldots, (\vartheta_k^{(k)}, \mathbf{u}_k^{(k)})$ be the kth order Ritz pairs of $\mathbf{A}$, also increasingly ordered Ritz values: $\vartheta_1^{(k)} < \ldots < \vartheta_k^{(k)}$, and, for all $i = 1, \ldots, k$,

$$\mathbf{u}_i^{(k)} \in \mathcal{K}_k(\mathbf{A}, \mathbf{r}_0) \qquad \text{and} \qquad \mathbf{A}\mathbf{u}_i^{(k)} - \vartheta_i^{(k)}\mathbf{u}_i^{(k)} \perp \mathcal{K}_k(\mathbf{A}, \mathbf{r}_0).$$

Proposition 12.20 *Consider a $k < n$. For all $i = 1, \ldots, k$, we have that*

$$\lambda_i < \vartheta_i^{(k+1)} < \vartheta_i^{(k)} \qquad \textit{and} \qquad \vartheta_i^{(k)} \in (\lambda_i, \lambda_{i+1}) \quad \Rightarrow \quad \vartheta_{i'}^{(k)} \in (\lambda_{i'}, \lambda_{i'+1}) \quad (i' < i). \tag{12.22}$$

Moreover, for all $j = 1, \ldots, k$, $\quad \vartheta_{k-j}^{(k)} < \vartheta_{k-j}^{(k+1)} < \lambda_{k-j}$ and

$$\vartheta_{k-j}^{(k)} \in (\lambda_{k-j-1}, \lambda_{k-j}) \quad \Rightarrow \quad \vartheta_{k-j'}^{(k)} \in (\lambda_{k-j'-1}, \lambda_{k-j'}) \quad (j' < j). \tag{12.23}$$

Eventually, for any fixed i, $\vartheta_i^{(k)}$ will 'converge' monotonically decreasing towards λ_i for $k \to \infty$ (usually, if i is small as $i \leq 3$, we already have that $0 < \vartheta_i^{(k)} - \lambda_i \ll \lambda_{i+1} - \lambda_i$ for a modest value of k as $k = 30$). In particular, eventually $\vartheta_i^{(k)}$ will be in its final interval, that is, $\vartheta_i^{(k)}$ will be in the interval $(\lambda_i, \lambda_{i+1})$. From (12.22), we learn that, for $i' < i$, all $\vartheta_{i'}^{(k)}$ will be in there final interval as soon as $\vartheta_i^{(k)}$ is in its final interval. Moreover there is at most one Ritz vale between two consecutive eigenvalues of $\mathbf{A}$. According to (12.23), a similar statement holds for the other end of the spectrum.

Exercise 12.26. *Proof of Prop. 12.20.*

(a) Show that the first statement of (12.22) is a consequence of Courant–Fischer and Cauchy's interlace Theorem.

To prove the second statement in (12.22) consider the projection $\mathbf{A}_v$ of $\mathbf{A}$ onto $\mathbf{v}^\perp$, where $\mathbf{v} \equiv \mathbf{v}_{k+1}$ is the $k+1$th Lanczos vector ($\mathbf{v}_{k+1} = \mathbf{r}_k / \|\mathbf{r}_k\|_2$), i.e., $\mathbf{A}_v = (\mathbf{I} - \mathbf{v}\mathbf{v}^*)\mathbf{A}(\mathbf{I} - \mathbf{v}\mathbf{v}^*)$. Consider $\mathbf{A}_v$ as a map from $\mathbf{v}^\perp$ to $\mathbf{v}^\perp$.

(b) Use Cauchy's interlace to prove that the eigenvalues of $\mathbf{A}$ and of $\mathbf{A}_v$ interlace.

(c) Put $\mathcal{K} \equiv \mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)$ and $\mathcal{W} \equiv \mathcal{K}_{k+1}(\mathbf{A}, \mathbf{r}_0)^\perp$. Note that $\mathbf{v} \perp \mathcal{K}$ as well as $\mathbf{v} \perp \mathcal{W}$.
Show that $\mathbf{A}_v : \mathcal{K} \to \mathcal{K}$ and $\mathbf{A}_v : \mathcal{W} \to \mathcal{W}$.

Conclude that eigenvalues of $\mathbf{A}_v$ as a map from $\mathcal{K}$ to $\mathcal{K}$ together with the eigenvalues of $\mathbf{A}_v$ as a map from $\mathcal{W}$ to $\mathcal{W}$ are the eigenvalues of $\mathbf{A}_v$. Note that the eigenvalues of $\mathbf{A}_v$ as a map from $\mathcal{K}$ to $\mathcal{K}$ are the kth order Ritz values.

(d) Prove the second statement of (12.22).

(e) Replace $\mathbf{A}$ by $-\mathbf{A}$ to prove (12.23).

Prop. 10.2 states that CG from step k on can be viewed as a CG process started with a deflated matrix. We will use the results of the present Lecture to analyse the distribution of the eigenvalues of this deflated matrix and thus explain super-linear convergence of CG.

Here, we use the notation of (10.22) and Prop. 10.2 and we will assume that $\mathbf{A}$ is positive definite. We measure the convergence of CG by the reduction of the residual in the $\mathbf{A}^{-1}$-norm. Therefore, since the approximate solutions of the CG are not needed for this, we can safely use to denote eigenvectors by $\mathbf{x}_i$ in the disicussion below.

Let $\mathbf{A}'$ be the map $\widetilde{\mathbf{A}}$ restricted to $\mathcal{K}^\perp$. As stated in Prop. 10.2, the eigenvalue distribution of $\mathbf{A}'$ determines the convergence of the CG process from the kth step on. The following proposition locates the eigenvalues of $\mathbf{A}'$ depending on the Ritz values of $\mathbf{A}$ and the eigenvalues of $\mathbf{A}$. Except for some mild technical restriction, the proposition basically states that as soon as the ith Ritz value arrives in its final interval, then, from the next step on, the CG process converges as if the eigenvalues $\lambda_1, \ldots, \lambda_i$ have been deflated from the matrix.

Proposition 12.21 *All eigenvalues of $\mathbf{A}'$ are in (λ_1, λ_n).*
Let $i, j < n$. All eigenvalues of $\mathbf{A}'$ are in $(\lambda_{i+1}, \lambda_{n-j-1})$ if the following two properties hold:
a) $\vartheta_i^{(k-1)} \in (\lambda_i, \lambda_{i+1})$ *and* $|(\mathbf{r}_k, \mathbf{x}_{i+1})| < |(\mathbf{r}_{k-1}, \mathbf{x}_{i+1})|$,
b) $\vartheta_{k-j}^{(k-1)} \in (\lambda_{n-j-1}, \lambda_{n-j})$.

Inspection of the proof of Prop. 12.5 (see Exercise 12.4), reveals the folowing generalisation of Prop. 12.5. This result is useful in the proof of the above Proposition (Prop. 12.21).

Proposition 12.22 *With $\mathbf{b}$ be a non-trivial n-vector, let $\vartheta_1, \ldots, \vartheta_{n-1}$ be the increasingly ordered Ritz values of an $n \times n$ Hermitian matrix $\mathbf{A}$ with respect to the space $\mathbf{b}^\perp$. Put $\vartheta_0 \equiv -\infty$ and $\vartheta_n \equiv +\infty$. For each $t \in \mathbb{R}$, let $\lambda_1(t), \ldots, \lambda_n(t)$ be the increasingly ordered eigenvalues of $\mathbf{A} + t\,\mathbf{b}\mathbf{b}^*$. Then, for each $i = 1, \ldots, n$, the function $t \rightsquigarrow \lambda_i(t)$ increases monotonically from ϑ_{i-1} (for $t \to -\infty$) to ϑ_i (for $t \to \infty$), $\lambda_i(0)$ is the ith eigenvalue of $\mathbf{A}$, and $\lambda_i(0) \in [\vartheta_{i-1}, \vartheta_i]$.*

In the proof of Prop. 12.21, we will also consider a matrix family of the form $\mathbf{A} + \frac{1}{t}\mathbf{b}\mathbf{b}^*$ ($t \in \mathbb{R}$). Note that Prop. 12.22 implies that the eigenvalue functions of $\mathbf{A} + \frac{1}{t}\mathbf{b}\mathbf{b}^*$ are decreasing with horizontal asymptots at the eigenvalues of $\mathbf{A}$ and Ritz values at $t = 0$.

Exercise 12.27. *Proof of Prop. 12.21.*

Let $\mathbf{P}$ be the orthogonal projection onto $\mathcal{K}$, and $\mathbf{Q}$ the orthogonal projection onto $\mathcal{K}^\perp$. With $\mathbf{v}_{i+1} \equiv \mathbf{r}_i/\|\mathbf{r}_i\|_2$ for all i, the $\mathbf{v}_i$ form an orthonormal system (of Lanczos vectors). Moreover, $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is a basis of $\mathcal{K}$, $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ is a basis of $\mathcal{K}^\perp$. With $\mathbf{V} \equiv [\mathbf{v}_1, \ldots, \mathbf{v}_k]$ and $\mathbf{W} \equiv [\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n]$, we have that $\mathbf{P} = \mathbf{V}\mathbf{V}^*$ and $\mathbf{Q} = \mathbf{I} - \mathbf{P} = \mathbf{W}\mathbf{W}^*$.

For analysis purposes, we introduce a parameter $t \in \mathbb{R}$ in the definition of $\widetilde{\mathbf{A}}$ of (10.23):

$$\widetilde{\mathbf{A}}(t) \equiv \mathbf{A} - \frac{1}{\sigma_k}\mathbf{c}(t)\mathbf{c}(t)^* \quad \text{with} \quad \mathbf{c}(t) \equiv \sqrt{t}\,\mathbf{r}_{k-1} - \frac{1}{\sqrt{t}}\mathbf{r}_k \qquad (t > 0) \tag{12.24}$$

and

$$\widetilde{\mathbf{A}}(t) \equiv \mathbf{A} + \frac{1}{\sigma_k}\mathbf{c}(t)\mathbf{c}(t)^* \quad \text{with} \quad \mathbf{c}(t) \equiv \sqrt{-t}\,\mathbf{r}_{k-1} + \frac{1}{\sqrt{-t}}\mathbf{r}_k \qquad (t < 0). \tag{12.25}$$

Note that $\widetilde{\mathbf{A}}(1)$ equals $\widetilde{\mathbf{A}}$ of (10.23).

(a) Prove that, for any $t \neq 0$, the eigenvalues of $\mathbf{A}$ and $\widetilde{\mathbf{A}}(t)$ interlace. In particular, with $\lambda_0 \equiv -\infty$, for each $t > 0$, each interval $(\lambda_i, \lambda_{i+1})$ (for $i = 0, \ldots, n-1$) contains at most one eigenvalue of $\widetilde{\mathbf{A}}(t)$, while each interval $[\lambda_i, \lambda_{i+1}]$ contains at least one eigenvalue of $\widetilde{\mathbf{A}}(t)$

(b) Prove that $\mathbf{r}_k^*\widetilde{\mathbf{A}}(t)\mathbf{r}_{k-1} = \mathbf{r}_k^*\widetilde{\mathbf{A}}(1)\mathbf{r}_{k-1}$, whence (by (a) of Exercise 10.17), $\mathbf{r}_k^*\widetilde{\mathbf{A}}(t)\mathbf{r}_{k-1} = 0$. Conclude that, for any t, $\widetilde{\mathbf{A}}(t) : \mathcal{K} \to \mathcal{K}$ and $\widetilde{\mathbf{A}}(t) : \mathcal{K}^\perp \to \mathcal{K}^\perp$.

Note that, with $\rho_i \equiv \|\mathbf{r}_i\|_2^2$, the matrices of these map with respect to the basis $\mathbf{v}_1, \ldots, \mathbf{v}_k$, and $\mathbf{v}_{k+1}, \ldots$, respectively, are given by

$$\mathbf{A}^\uparrow(t) \equiv \mathbf{V}^*\mathbf{A}\mathbf{V} - t\,\frac{\rho_{k-1}}{\sigma_k}\mathbf{e}_k\mathbf{e}_k^* \qquad \text{and} \qquad \mathbf{A}^\downarrow(t) \equiv \mathbf{W}^*\mathbf{A}\mathbf{W} - \frac{1}{t}\,\frac{\rho_k}{\sigma_k}\mathbf{e}_{k+1}\mathbf{e}_{k+1}^* \qquad (t \neq 0).$$

Let $\vartheta_1^\uparrow(t), \ldots, \vartheta_k^\uparrow(t)$ and $\vartheta_1^\downarrow(t), \ldots, \vartheta_{n-k}^\downarrow(t)$ be the increasingly ordered eigenvalues of $\mathbf{A}^\uparrow(t)$ and $\mathbf{A}^\downarrow(t)$, respectively. Then $\vartheta_j^\uparrow(1) = \vartheta_j^{(k)}$ and $\vartheta_i^\downarrow(1)$ are the eigenvalues of $\mathbf{A}'$. Note that the

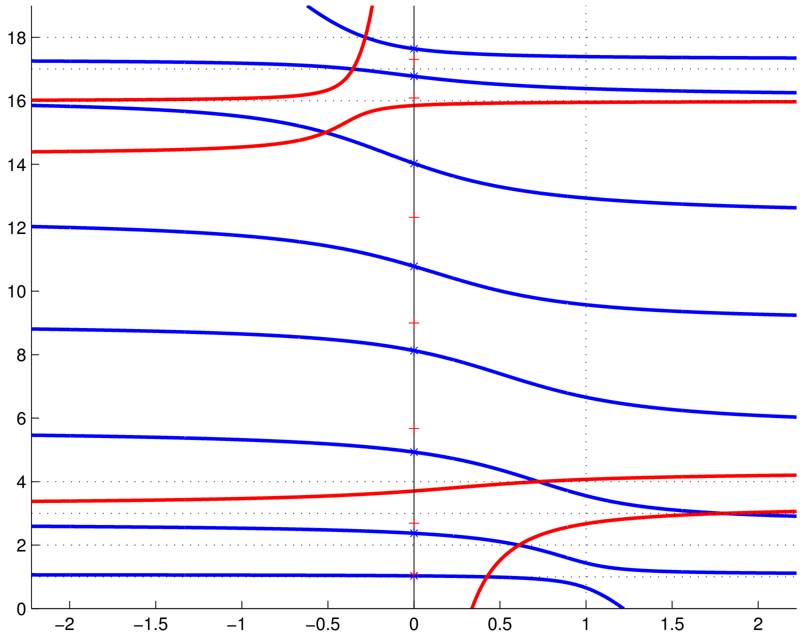

FIGURE 1. The figure shows, for some positive definite matrix $\mathbf{A}$, all $t \rightsquigarrow \vartheta_j^{\uparrow}(t)$ (the blue curves, with t along the horizontal axis) and the extremal $t \rightsquigarrow \vartheta_i^{\downarrow}$ (in red) as defined in (b) of Exercise 12.27. In this example, $n = 18$ and $k = 8$. The blue and the red curves intersect at eigenvalues of $\mathbf{A}$ (indicated by the doted horizontal lines; see Exercise 12.27.(e)).

collection of all $\vartheta_i^{\uparrow}(t)$ together with all $\vartheta_j^{\downarrow}(t)$ form the set of all eigenvalues of $\widetilde{\mathbf{A}}(t)$. See Fig. 1 for a graphical illustration.

(c) Show that for each $i = 1, \ldots, k$, the function $t \rightsquigarrow \vartheta_i^{\uparrow}(t)$ decreases monotonically from $\vartheta_i^{(k-1)}$ (at $t = -\infty$) to $\vartheta_{i-1}^{(k-1)}$ (at $t = \infty$), while $\vartheta_i^{\uparrow}(0) = \vartheta_i^{(k)}$. Here, we assumed $\vartheta_0^{(k-1)}$ and $\vartheta_k^{(k-1)}$ to be defined as $\vartheta_0^{(k-1)} \equiv -\infty$ and $\vartheta_k^{(k-1)} \equiv \infty$. Show that each of the function $\vartheta_j^{\downarrow}$ increases monotonically for $t \in (0, \infty)$. (cf. Fig. 1)

(d) Assume $\vartheta_{k-j}^{(k-1)} \in (\lambda_{n-j-1}, \lambda_{n-j})$. Show that, for each $t > 0$, we have that $\vartheta_{k-j}^{\uparrow}(t) \in (\lambda_{n-j-1}, \lambda_{n-j})$ and $\vartheta_{k-j'}^{\uparrow}(t) \in (\lambda_{n-j'-1}, \lambda_{n-j'})$ for all $j' < j$. Conclude that $\vartheta_{n-k}^{\downarrow}(t) < \lambda_{n-j-1}$. In particular we have that all eigenvalues of $\mathbf{A}'$ are less than λ_{n-j-1}. (The two red top curves in Fig. 1 illustrate the situation as discussed here).

(e) Assume that $\vartheta_{i-1}^{(k-1)} \in (\lambda_{i-1}, \lambda_i)$. Then, there is a $t_i = t > 0$ such that $\vartheta_i^{\uparrow}(t_i) = \lambda_i$. Why? Show that, for $t \in \mathbb{R}$,

$$\vartheta_i^{\uparrow}(t) = \lambda_i \quad \Leftrightarrow \quad \mathbf{x}_i \perp \mathbf{c}(t) \quad \Leftrightarrow \quad t = \frac{\mathbf{r}_k^* \mathbf{x}_i}{\mathbf{r}_{k-1}^* \mathbf{x}_i} \quad \Leftrightarrow \quad \vartheta_1^{\downarrow}(t) = \lambda_i.$$

Conclude that all eigenvalues of $\mathbf{A}'$ are larger than λ_i if, in addition to $\vartheta_{i-1}^{(k-1)} \in (\lambda_{i-1}, \lambda_i)$ we also have that $1 > t_i$. (Compare with two red bottom curves in Fig. 1).

(f) In particular (with the convention that $\vartheta_0^{(k-1)} = -\infty$ and $\lambda_0 = -\infty$), we have that $\vartheta_1^{\uparrow}(t_1) = \lambda_1$, for some $t = t_1 = \frac{\mathbf{r}_k^* \mathbf{x}_1}{\mathbf{r}_{k-1}^* \mathbf{x}_1} > 0$. If we can show that $0 < \mathbf{r}_k^* \mathbf{x}_1 < \mathbf{r}_{k-1}^* \mathbf{x}_1$, then we proved that all eigenvalues of $\mathbf{A}'$ are larger than λ_1.

Note that $\mathbf{r}_k^* \mathbf{x}_1 = p_k(\lambda_1)$, where p_k is the kth CG (FOM) residual polynomial. Conclude from the fact that $p_k(0) = 1$ and $\lambda_1 < \vartheta_1^{(k-1)} < \vartheta_1^{(k)}$ that $0 < \mathbf{r}_k^* \mathbf{x}_1 < \mathbf{r}_{k-1}^* \mathbf{x}_1$.