# TAYLOR EXPANSIONS OF THE SPECTRUM OF A SYMMETRIC MATRIX

A Thesis

Presented to

The Faculty of Graduate Studies

of

The University of Guelph

by

BRENDAN P. W. AMES

In partial fulfilment of requirements

for the degree of

Master of Science

April, 2007

ABSTRACT

# TAYLOR EXPANSIONS OF THE SPECTRUM OF A SYMMETRIC MATRIX

Brendan P. W. Ames
University of Guelph, 2007

Advisor(s):
Dr. H.S. Sendov

Optimization problems involving the eigenvalues of a real symmetric matrix are common in many areas of research. In particular, we consider the sensitivity of all the eigenvalues of a symmetric matrix under small perturbations. In this thesis we present several technical results on perturbations of invariant subspaces. Applying these results we show that, for any symmetric matrix $A$, the $m$-th largest eigenvalue of $A + tE(t)$ has a $k$-th order Taylor expansion at $t = 0^+$ when the symmetric matrix $E(t)$ depending on the real parameter $t$ has a $(k-1)$-th order Taylor expansion at $t = 0^+$ and derive a formula for the coefficients of this expansion. Using this formula we calculate the first four directional derivatives of the $m$-th largest eigenvalue of a symmetric matrix $A$ in a fixed, direction $E$.

# Table of Contents

# Chapter 1

## Introduction

The variational analysis of the spectrum of symmetric matrices is a classical subject. It is well-known that the eigenvalues of any matrix are continuous functions on the entries of the matrix since they are the roots of its characteristic polynomial. Suppose that the entries of the matrix depend *smoothly* on one or more parameters. Much more complicated question is to determine how much of that smoothness is transferred onto the eigenvalues. In this work we address issues about the smoothness of the spectrum of a symmetric matrix whose entries depend on one real parameter. If $A$ is an $n \times n$ symmetric matrix, by

$$\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)$$

we denote its $n$ eigenvalues (all of them real) ordered non-increasingly. Often $\lambda_m(A)$ is called the *$m$-th largest* eigenvalue of $A$.

We begin with a simple example. Consider the following $2 \times 2$ symmetric matrix whose entries depend on $t \in \mathbb{R}$:

$$A(t) := \begin{pmatrix} 0 & t \\ t & 0 \end{pmatrix}.$$

Its eigenvalues are $\lambda_1(A(t)) = |t|$ and $\lambda_2(A(t)) = -|t|$. Clearly they are not differentiable at $t = 0$. The problem with non-differentiability at $t = 0$ in this example stems from the fact that we are considering the ordered spectrum of $A(t)$. If, for example, we define $\mu_1(t) := t$ and $\mu_2(t) := -t$, then for any $t$ the spectrum of $A(t)$ is $\{\mu_1(t), \mu_2(t)\}$ and moreover the functions $\mu_1(t)$ and $\mu_2(t)$ are always differentiable. This example has been generalized by F. Rellich [6] in the following theorem.

**Theorem 1.0.1** (Rellich, 1953)**.** *Let $A(t) = (a_{ij}(t))$ be an $n \times n$ symmetric matrix where $a_{ij}(t) \in C^1$ for all $i, j = 1, ..., n$. Then, there are functions $\mu_i(t) \in C^1$ for $i = 1, ..., n$ such that the set $\{\mu_1(t), ..., \mu_n(t)\}$ contains all eigenvalues of $A(t)$ counting multiplicities for all $t$.*

Rellich's theorem shows that there are inherent problems with considering the ordered spectrum when asking questions about the variational properties of the eigenvalues. Unfortunately, Rellich's theorem cannot be strengthened. That is, even if we assume that $a_{ij}(t) \in \mathcal{C}^\infty$ for all $i, j = 1, ..., n$ one cannot guarantee that there are $C^2$ functions $\mu_i(t)$ for $i = 1, ..., n$ such that $\{\mu_1(t), ..., \mu_n(t)\}$ is the spectrum of $A(t)$, see [2, Example 5.3]. In another twist we have the following result, see [2].

**Theorem 1.0.2.** *Let $A(t) = (a_{ij}(t))$ be an $n \times n$ symmetric matrix where $a_{ij}(t)$ is analytic for all $i, j = 1, ..., n$. Then, there are analytic functions $\mu_i(t)$ for $i = 1, ..., n$ such that the set $\{\mu_1(t), ..., \mu_n(t)\}$ contains all eigenvalues of $A(t)$ counting multiplicities for all $t$.*

The behavioral patterns exhibited above reflect the complicated nature of the eigenvalues. In the works of Kato, Rellich and Lancaster [3], the variational theory of the $\lambda$-groups (that is, groups of equal eigenvalues) of symmetric matrices was developed under the assumption of analyticity of $A(t)$. The variational behaviour of the ordered spectrum was considered a difficult problem.

In this work we partially fill the gap between the two extreme cases in Theorem 1.0.1 and Theorem 1.0.2 and examine the variational behaviour of the ordered spectrum.

**Definition 1.0.3.** We say that the symmetric matrix-valued map $t \mapsto A(t)$ has a (one sided) *Taylor expansion of order* $K$ *at* $t = 0^+$ if there exist symmetric matrices $A_0, A_1, ..., A_K$ and $\epsilon > 0$ such that

$$A(t) = A_0 + tA_1 + \cdots + t^K A_K + o(t^K) \text{ for all } t \in [0, \epsilon).$$

In [9], Torki showed that if $A(t)$ is twice differentiable at 0 then $\lambda_m(A(t))$ has a Taylor expansion of order 2 at $t = 0^+$ for all $m = 1, ..., n$. We generalize this result; we show that if $A(t)$ has a $K$-th order Taylor expansion at $t = 0^+$ then so does $\lambda_m(A(t))$ for all $m = 1, ..., n$. We begin by providing the necessary background in Matrix Analysis and Operator Theory in Chapters 2 and 3 respectively. Chapter 4 introduces several results regarding the perturbation of invariant subspaces. In particular, we present a proof of Stewart's Theorem; our main technical tool for deriving the Taylor expansions of the spectrum of $A(t)$.

We derive the main result in Chapter 5. We show that if $A$ is an $n \times n$ symmetric matrix and $t \mapsto E(t)$ is symmetric matrix-valued map with a $K$-th order

Taylor expansion at $t = 0^+$ then the map

$$t \mapsto \lambda_m(A + tE(t))$$

has a $(K + 1)$-th order Taylor expansion at $t = 0^+$ for all $m = 1, ..., n$. We give an inductive formula for computing the Taylor coefficients $A_0, A_1, ..., A_{K+1}$. In particular, if $E(t)$ is analytic at $t = 0^+$ then $\lambda_m(A + tE(t))$ has a Taylor expansion at $t = 0^+$ of any order. As an exercise we compute the first four terms in the Taylor expansion of $\lambda_m(A + tE)$ for any fixed symmetric matrices $A$ and $E$.

# Chapter 2

# Matrix theory

## 2.1 Eigenvalues of matrices

We assume familiarity with basic notions from matrix theory such as determinant, rank, inverse of a matrix, etc.

Denote by $M_{m,n}$ the linear space of all $m \times n$ real matrices.

**Definition 2.1.1.** For any square matrix $A \in M_{n,n}$, $\lambda \in \mathbb{C}$ is an *eigenvalue* of $A$ if there is a nonzero vector, $v \in \mathbb{C}^n$, such that $Av = \lambda v$ ($v$ is called an *eigenvector*). The set of all vectors, $v$, such that $Av = \lambda v$ is called the *eigenspace* associated with $\lambda$. That is, the eigenspace associated with $\lambda$ is $N(A - \lambda I)$, the null space of $A - \lambda I$.

For any $A \in M_{n,n}$, let $\sigma(A)$ denote the set of eigenvalues of $A$, called the *spectrum* of $A$.

**Lemma 2.1.2.** *For any $A \in M_{n,n}$, $\lambda \in \mathbb{C}$ is an eigenvalue of $A$ if and only if $\lambda$ is a root of the characteristic polynomial* $\det(A - \lambda I) = 0$.

**Proof.** By the definition of an eigenvalue of $A$, $\lambda$ is an eigenvalue of $A$ if and only if

$$Ax = \lambda x \text{ (for some } x \neq 0) \Leftrightarrow (A - \lambda I)x = 0 \text{ (for some } x \neq 0)$$

$$\Leftrightarrow A - \lambda I \text{ is singular}$$

$$\Leftrightarrow \det(A - \lambda I) = 0.$$

This completes the proof. $\square$

**Definition 2.1.3.** For any eigenvalue $\lambda$ of the matrix $A \in M_{n,n}$, the *algebraic multiplicity* of $\lambda$ is the number of times $\lambda$ is repeated as a root of the characteristic polynomial $\det(A - \lambda I) = 0$. Let $algmult_A(\lambda)$ denote the algebraic multiplicity of $\lambda$.

We next state the Fundamental theorem of algebra and prove, as a corollary, that every $A \in M_{n,n}$ has exactly $n$ eigenvalues counting algebraic multiplicity. A proof of Theorem 2.1.4 can be found in [7, Section 4.6, Theorem 22].

**Theorem 2.1.4** (Fundamental theorem of algebra)**.** *Every polynomial of degree $n \geq 1$ with complex coefficients has exactly $n$ roots in $\mathbb{C}$ counting multiplicities.*

**Corollary 2.1.5.** *Every $A \in M_{n,n}$ has exactly $n$ eigenvalues counting algebraic multiplicity.*

**Proof.** Consider $\det(A - \lambda I) = 0$. This is a polynomial in $\lambda$ of degree $n$. By the Fundamental theorem of algebra, it has $n$ complex roots (counting multiplicities). $\square$

**Lemma 2.1.6.** *Let $\lambda \in \mathbb{R}$ be an eigenvalue of $A \in M_{n,n}$. Then there exists a real eigenvector of $A$ corresponding to $\lambda$.*

**Proof.** Let $v := u + iw$ be an eigenvector of $A$ corresponding to $\lambda$. Then $Au + iAw = \lambda u + i\lambda w$ and, since $v$ is nonzero, at least one of $u, v$ are nonzero. Consequently, because $\lambda$ is real, $Au = \lambda u$ and $Aw = \lambda w$, and thus there is a real eigenvector corresponding to $\lambda$. $\square$

**Definition 2.1.7.** A matrix $A \in M_{n,n}$ is *symmetric* if $A^T = A$.

Let $\mathcal{S}^n$ denote the set of all $n \times n$ symmetric matrices. If $x$ is a complex vector or a scalar, by $\overline{x}$ we denote its complex conjugate.

**Lemma 2.1.8.** *Every $n \times n$ symmetric matrix $A$ has real eigenvalues.*

**Proof.** Suppose $\lambda$ is an eigenvalue of $A$ with corresponding eigenvector $x = u + iv \in \mathbb{C}^n$. Consider the scalar $q := \overline{x}^T Ax$. Taking the complex conjugate of $q$ gives

$$\overline{q} = \overline{\overline{x}^T Ax} = x^T A\overline{x} = (\overline{x}^T Ax)^T = q,$$

since $A$ is real and $q$ is a scalar. Thus, $q$ is real. Furthermore, since $\lambda$ is an eigenvalue of $A$, $Ax = \lambda x$ and thus

$$q = \overline{x}^T Ax = \lambda(u^T u + v^T v).$$

Note that $(u^T u + v^T v) \neq 0$ since at least one of $u$ and $v$ is nonzero. Thus, $\lambda$ is real showing that symmetric matrices have real eigenvalues. $\square$

Denote by $e$ the vector in $\mathbb{R}^n$ with all entries equal to 1. For any $A \in \mathcal{S}^n$, denote by $\lambda(A) \in \mathbb{R}^n$ the vector of eigenvalues of $A$, counting algebraic multiplicity, in nonincreasing order. That is,

$$\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A).$$

Define the map Diag $: \mathbb{R}^n \to \mathcal{S}^n$ for any $x = (x_1, x_2, ..., x_n)^T \in \mathbb{R}^n$ by

$$\operatorname{Diag}(x) := \begin{bmatrix} x_1 & 0 & \dots & 0 \\ 0 & x_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & x_n \end{bmatrix}.$$

**Definition 2.1.9.** The $k \times k$ *identity matrix* is denoted by $I_k$, or simply $I$ when the size is understood from the context.

**Definition 2.1.10.** An $n \times n$ matrix $U$ is *orthogonal* if $U^T U = I$. Denote by $O^n$ the set of all $n \times n$ orthogonal matrices.

It follows that $U^T = U^{-1}$ and thus $UU^T = I$.

**Definition 2.1.11.** An $n \times n$ *permutation matrix* is a matrix with entries 0 or 1 such that in every row and column there is exactly one 1.

For example, there are two $2 \times 2$ permutation matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

**Theorem 2.1.12.** *Every permutation matrix is orthogonal.*

**Proof.** Let P be an $n \times n$ permutation matrix. Consider the $(i, j)$-th entry of $P^T P$:

$$(P^T P)_{ij} = \sum_{k=1}^{n} P_{ki} P_{kj}.$$

Therefore,

$$(P^T P)_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $P^T P = I$ and consequently $P$ is orthogonal. □

Consider the vector $Px \in \mathbb{R}^n$, where $P$ is an $n \times n$ permutation matrix and $x \in \mathbb{R}^n$. The $i$-th entry of $Px$ is $\sum_{k=1}^{n} P_{ik}x_k = x_r$ where $r$ is the position of the nonzero entry in row $i$. Since each column of $P$ has exactly one 1, then the entries of $Px$ are a permutation of the entries of vector $x$. The following lemma is easy to verify.

**Lemma 2.1.13.** *For any vector $x \in \mathbb{R}^n$, and permutation matrix $P$,*

$$P(\operatorname{Diag} x)P^T = \operatorname{Diag}(Px).$$

## 2.2 Inner products and vector norms

**Definition 2.2.1.** A *norm* on a real vector space $X$ is a function $\|\cdot\| : X \to \mathbb{R}$ that satisfies

(1) $\|x\| \geq 0$,

(2) $\|x\| = 0$ if and only if $x = 0$,

(3) $\|\alpha x\| = |\alpha|\|x\|$ for all $\alpha \in \mathbb{R}$,

(4) $\|x + y\| \leq \|x\| + \|y\|$

for all $x, y \in X$.

**Definition 2.2.2.** An *inner product* on a real vector space $X$ is a function $(x, y) \in X \times X \mapsto \langle x, y\rangle \in \mathbb{R}$ that satisfies

(1) $\langle x, x\rangle \geq 0$ for all $x \in X$ and $\langle x, x\rangle = 0$ if and only if $x = 0$,

(2) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$ for all $\alpha \in \mathbb{R}$ and $x, y \in X$,

(3) $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$ for all $x, y, z \in X$,

(4) $\langle x, y\rangle = \langle y, x\rangle$ for all $x, y \in X$.

**Definition 2.2.3.** A finite dimensional real vector space $X$ endowed with an inner product is called a *Euclidean space.*

In what follows let $E$ be a Euclidean space.

**Lemma 2.2.4** (Cauchy-Schwarz)**.** *Let $E$ be an Euclidean space and consider the function $\|\cdot\| : E \to \mathbb{R}$ defined by $\|x\| := \sqrt{\langle x, x\rangle}$ for all $x \in E$. Then*

$$|\langle x, y\rangle| \leq \|x\|\|y\|$$

*for all $x, y \in E$.*

**Proof.** If $x = 0$ then trivially $|\langle x, y\rangle| = \|x\|\|y\| = 0$. Now suppose $x \neq 0$ and set

$$\alpha := \frac{\langle x, y\rangle}{\langle x, x\rangle}.$$

Then, since $\langle x, \alpha x - y\rangle = 0$, we have

$$0 \leq \|\alpha x - y\|^2 = \langle \alpha x - y, \alpha x - y\rangle = \alpha\langle x, \alpha x - y\rangle - \langle y, \alpha x - y\rangle$$

$$= -\langle y, \alpha x - y\rangle = \langle y, y\rangle - \alpha\langle y, x\rangle = \frac{\|y\|^2\|x\|^2 - \langle x, y\rangle\langle y, x\rangle}{\|x\|^2}.$$

Thus, since $\langle x, y\rangle = \langle y, x\rangle$ it follows that $\langle y, x\rangle\langle x, y\rangle = |\langle y, x\rangle|^2$ and

$$0 \leq \frac{\|y\|^2\|x\|^2 - |\langle y, x\rangle|^2}{\|x\|^2}$$

as required. □

**Proposition 2.2.5.** *The inner product associated with $E$ defines a norm $\|\cdot\|$ on $E$ by $\|x\| := \sqrt{\langle x, x\rangle}$ for all $x \in E$.*

**Proof.** Properties (1), (2) and (3) of a norm follow immediately from Properties (1) and (2) of an inner product. It remains to show that the triangle inequality holds. Indeed, by the Cauchy-Schwarz inequality (Lemma 2.2.4),

$$\begin{aligned}\|x + y\|^2 &= \langle x + y, x + y\rangle = \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle \\ &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.\end{aligned}$$

Thus, since $\|x\| \geq 0$ for all $x \in E$, taking square roots show that that $\|x + y\| \leq \|x\| + \|y\|$. □

**Lemma 2.2.6.** *The vector space $\mathbb{R}^n$ is a Euclidean space with associated inner product defined by*

$$\langle x, y\rangle := x^T y$$

*for all $x, y \in \mathbb{R}^n$. The norm defined on $\mathbb{R}^n$ by this inner product is*

$$\|x\| = \sqrt{x^T x} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2},$$

*called the* Euclidean norm *on $\mathbb{R}^n$.*

Unless otherwise noted, $\|\cdot\|$ will denote the Euclidean norm on $\mathbb{R}^n$.

**Definition 2.2.7.** Denote the *Kronecker delta* symbol by

$$\delta_{ij} := \begin{cases} 1 & \textit{if } i \textit{ is equal to } j \\ 0 & \textit{otherwise.} \end{cases}$$

**Definition 2.2.8.** For any Euclidean space $E$, the subset $\{u^i\}_{i=1}^m \subset E$ is said to be *orthonormal* if $\langle u^i, u^j \rangle = \delta_{ij}$ for all $i, j = 1, 2, ..., m$.

**Lemma 2.2.9.** *If $U := \{u^i\}_{i=1}^m \subset E$ is an orthonormal set, then $U$ is linearly independent.*

**Proof.** Choose $\alpha \in \mathbb{R}^m$ such that

$$0 = \sum_{i=1}^m \alpha_i u^i.$$

Choose $j \in \{1, 2, ..., m\}$. Then

$$0 = \langle u^j, \sum_{i=1}^m \alpha_i u^i \rangle = \sum_{i=1}^m \alpha_i \langle u^j, u^i \rangle = \alpha_j.$$

Since the choice of $j$ was arbitrary, $\alpha = 0$. Thus, $U$ is linearly independent. $\square$

## 2.3 Spectral decomposition theorem

The proof of the following theorem is modified from [5, Section 4.5]. For a matrix $B \in M_{n,p}$, the *range* of $B$ is the linear subspace of $\mathbb{R}^n$ defined by

$$R(B) := \{By \,|\, y \in \mathbb{R}^p\}.$$

**Theorem 2.3.1.** *For any matrices $A \in M_{m,n}$ and $B \in M_{n,p}$*

$$\operatorname{rank}(AB) = \operatorname{rank}(B) - \dim(N(A) \cap R(B)).$$

**Proof.** Let $S = \{x^1, x^2, ..., x^s\}$ be a basis for $N(A) \cap R(B)$. Clearly, since $N(A) \cap R(B) \subset R(B)$, there exists a nonnegative $t$ such that

$$\operatorname{rank}(B) = s + t.$$

Thus, there exists a set of independent vectors $S_t = \{z^1, z^2, ..., z^t\}$ such that $\{S, S_t\}$ forms a basis for $R(B)$.

It remains to show that $\operatorname{rank}(AB) = t$. We will do so by proving that the set $T = \{Az^1, Az^2, ..., Az^t\}$ is a basis for $R(AB)$. We first show that $T$ spans $R(AB)$. Indeed, choose $b \in R(AB)$. Then $b = ABy$ for some $y$ and since $By \in R(B)$ there exists $\alpha \in \mathbb{R}^t$ and $\beta \in \mathbb{R}^s$ such that

$$b = ABy = A\Big(\sum_{i=1}^{t} \alpha_i z^i + \sum_{j=1}^{s} \beta_j x^j\Big) = \sum_{i=1}^{t} \alpha_i (Az^i) + \sum_{j=1}^{s} \beta_j (Ax^j) = \sum_{i=1}^{t} \alpha_i (Az^i)$$

since $x^j \in N(A)$ for each $j = 1, 2, ..., s$. Therefore $R(AB) \subset \operatorname{span} T$.

For the opposite inclusion, since any linear combination of elements of $S_t$ is in $R(B)$, $\operatorname{span} T \subset R(AB)$ and thus $\operatorname{span} T = R(AB)$.

Next, we prove that $T$ is a linearly independent set. Choose $\alpha \in \mathbb{R}^t$ such that

$$0 = \sum_{i=1}^{t} \alpha_i (Az^i) = A \sum_{i=1}^{t} \alpha_i z^i.$$

Clearly

$$\sum_{i=1}^{t} \alpha_i z^i \in N(A) \cap R(B).$$

Thus, there exists $\beta \in \mathbb{R}^s$ such that

$$\sum_{i=1}^{t} \alpha_i z^i = \sum_{j=1}^{s} \beta_j x^j. \tag{2.1}$$

However, since $\{S, S_t\}$ is an independent set, (2.1) implies that $\alpha = 0, \beta = 0$. Therefore, $T = \{Az^1, Az^2, ..., Az^t\}$ is an independent set and a basis for $R(AB)$. $\square$

**Corollary 2.3.2.** *For any matrix $A \in M_{m,n}$*

$$\operatorname{rank}(A^T A) = \operatorname{rank}(A) = \operatorname{rank}(AA^T).$$

**Proof.** It is enough to show the left equality. By Theorem 2.3.1

$$\operatorname{rank}(A^T A) = \operatorname{rank}(A) - \dim(N(A^T) \cap R(A)).$$

We prove the result by showing that $\dim(N(A^T) \cap R(A)) = 0$. Indeed, choose $y \in$

$N(A^T) \cap R(A)$. Then, there exists some $x \in \mathbb{R}^n$ such that $y = Ax$ and thus

$$\|y\|_2^2 = y^T y = x^T A^T y = 0$$

since $y \in N(A^T)$. Therefore, $y = 0$ and thus

$$N(A^T) \cap R(A) = \{0\}.$$

Hence, $\dim(N(A^T) \cap R(A)) = 0$ as required. □

The proof of the following proposition is modified from [5, Section 5.5].

**Proposition 2.3.3** (Gram-Schmidt orthogonalization process)**.** *If $\{x^i\}_{i=1}^n$ is a basis for $\mathbb{R}^n$, then the sequence defined inductively by*

$$u^1 = \frac{x^1}{\|x^1\|_2}$$

*and*

$$u^k = \frac{1}{\alpha_k}\left(x^k - \sum_{i=1}^{k-1}\langle u^i, x^k\rangle u^i\right) \quad k = 2, 3, ..., n, \tag{2.2}$$

*where $\alpha_k = \|x^k - \sum_{i=1}^{k-1}\langle u^i, x^k\rangle u^i\|_2$ for $k = 2, 3, ..., n$ is an orthonormal basis for $\mathbb{R}^n$.*

**Proof.** It suffices to show that $U_k = \{u^1, u^2, ..., u^k\}$ is an orthonormal basis for $S_k := \operatorname{span}\{x^1, x^2, ..., x^k\}$ for all $k = 1, 2, ..., n$. If this holds then $k = n$ shows that $\{u^i\}_{i=1}^n$ is an orthonormal basis for $S_n = \mathbb{R}^n$.

We prove the proposition by induction. Trivially $u^1$ is an orthonormal basis for $\operatorname{span}\{x^1\}$ and thus the base case $k = 1$ holds. Now choose $k > 1$ and suppose that $U_{k-1}$ is an orthonormal basis for $S_{k-1}$. We first show that $U_k$ is an orthonormal set. Observe that

$$\langle u^k, u^k \rangle = \frac{\alpha_k^2}{\alpha_k^2} = 1$$

and, by the induction hypothesis,

$$\langle u^i, u^j \rangle = \delta_{ij}$$

for all $1 \leq i, j < k$. Choose $j < k$. Then

$$\begin{aligned}
\langle u^j, u^k \rangle &= \langle u^j, {\alpha_k}^{-1} \Big( x^k - \sum_{i=1}^{k-1} \langle u^i, x^k \rangle u^i \Big) \rangle \\
&= \alpha_k^{-1} \Big( \langle u^j, x^k \rangle - \sum_{i=1}^{k-1} \langle u^i, x^k \rangle \langle u^i, u^j \rangle \Big) \\
&= \alpha_k^{-1} (\langle u^j, x^k \rangle - \langle u^j, x^k \rangle) = 0
\end{aligned}$$

by the orthogonality of $U_{k-1}$.

We next show that $U_k$ is a basis for $S_k$. Since $U_k$ is an orthonormal set, it is linearly independent and thus it remains to show that $\operatorname{span} U_k = S_k$. Trivially, $u^i \in S_{k-1} \subset S_k$ by the induction hypothesis for all $i = 1, 2, ...k-1$. Since $u^k$ is a linear combination of $U_{k-1} \cup \{x^k\}$, then $u^k \in S_k$. Therefore, $\operatorname{span} U_k \subset S_k$. Now choose $x \in S_k$. There exists $\beta \in \mathbb{R}^k$ such that

$$x = \sum_{i=1}^{k} \beta_i x^i = \sum_{i=1}^{k-1} \beta_i x^i + \beta_k x^k.$$

Observe that $\sum_{i=1}^{k-1} \beta_i x^i \in S_{k-1} = \operatorname{span} U_{k-1} \subset \operatorname{span} U_k$. Furthermore, by (2.2) it follows that

$$x^k = \alpha_k u^k + \sum_{i=1}^{k-1} \langle u^i, x^k \rangle u^i \in \operatorname{span} U_k$$

and thus $x \in \operatorname{span} U_k$. Therefore, $S_k \subset \operatorname{span} U_k$ and thus $\operatorname{span} U_k = S_k$. This completes the proof. $\square$

For each $i = 1, 2, ..., n$, denote by $e^i$ the vector in $\mathbb{R}^n$ such that $e^i_j = \delta_{ij}$ for all $j = 1, 2, ..., n$.

**Theorem 2.3.4** (Spectral decomposition). *If $A \in \mathcal{S}^n$ there is a $U \in O^n$ and $x \in \mathbb{R}^n$ such that*

$$A = U^T(\operatorname{Diag} x)U,$$

*and the entries of $x$ are the eigenvalues of $A$ counting algebraic multiplicity.*

**Proof.** We first show, by induction on $n$, that every symmetric matrix is orthogonally similar to a diagonal matrix. The base case $n = 1$ follows immediately since any $1 \times 1$ matrix is a diagonal matrix.

Now suppose that every $A \in \mathcal{S}^{n-1}$ is orthogonally similar to a diagonal matrix. That is, for all $A \in \mathcal{S}^{n-1}$ there exists $U \in O^{n-1}$ and $x \in \mathbb{R}^{n-1}$ such that

$$A = U^T(\operatorname{Diag} x)U.$$

Now choose $A \in \mathcal{S}^n$. Then, $A$ has at least one eigenvalue $\lambda$ with corresponding unit eigenvector $v^1$ and, by Lemmas 2.1.6 and 2.1.8, we can assume that $\lambda$ and $v^1$ are real. Complete the set $\{v^1\}$ to a basis of $\mathbb{R}^n$ and apply the Gram-Schmidt process to it starting with vector $v^1$ to obtain the orthonormal basis

$$B = \{v^1, v^2, v^3, ..., v^n\}$$

of $\mathbb{R}^n$. Let $X$ be the $n \times n$ real matrix

$$X^T := [v^1, v^2, ..., v^n]$$

with columns $v^1, v^2, ..., v^n$. Clearly $X \in O^n$.

Consider the matrix $\tilde{A} := XAX^T \in \mathcal{S}^n$. The first column of $\tilde{A}$ is

$$(XAX^T)e^1 = XA(X^Te^1) = XAv^1 = X(\lambda v^1) = \lambda(Xv^1) = \lambda(XX^Te^1) = \lambda e^1.$$

Hence, since $\tilde{A} \in \mathcal{S}^n$, it follows that

$$\tilde{A} = \begin{bmatrix} \lambda & 0 \\ 0 & B \end{bmatrix}$$

where $B \in \mathcal{S}^{n-1}$. By the induction hypothesis, there exists $V \in O^{n-1}$ and $x' \in \mathbb{R}^{n-1}$ such that

$$VBV^T = \operatorname{Diag} x'.$$

Let

$$W := \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix}.$$

Then, since $V \in O^{n-1}$, it follows that $W \in O^n$ and

$$WXAX^TW^T = \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} \begin{bmatrix} \lambda & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V^T \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & \operatorname{Diag} x' \end{bmatrix}.$$

Since $X$ and $W$ are orthogonal matrices, the product $U := WX$ is also orthogonal and thus $A$ is orthogonally similar to a diagonal matrix. Therefore, for all $A \in \mathcal{S}^n$ there is a $U \in O^n$ and $x \in \mathbb{R}^n$ such that

$$A = U^T(\operatorname{Diag} x)U. \tag{2.3}$$

It remains to prove that the entries of the vector $x$ are all eigenvalues of $A$. Indeed, using the facts that $U^T = U^{-1}$ and $\det(U^{-1}) = 1/\det(U)$ we compute

$$\begin{aligned}\det(A - \lambda I) &= \det(U^T(\operatorname{Diag} x)U - \lambda I) = \det(U^T(\operatorname{Diag} x - \lambda I)U) \\ &= \det(U^T)\det(\operatorname{Diag} x - \lambda I)\det(U) = \det(\operatorname{Diag} x - \lambda I) \\ &= (x_1 - \lambda)\cdots(x_n - \lambda).\end{aligned}$$

This shows that the roots of $\det(A - \lambda I) = 0$ are $x_1, ..., x_n$, which together with Lemma 2.1.2 concludes the proof. $\square$

Denote the set of all vectors in $\mathbb{R}^n$ with entries in nonincreasing order by

$$\mathbb{R}^n_{\geq} := \{(x_1, x_2, ..., x_n) \in \mathbb{R}^n \mid x_1 \geq x_2 \geq ... \geq x_n\}.$$

**Corollary 2.3.5.** *An ordered spectral decomposition exists for every $A \in \mathcal{S}^n$. That* *is, there exists $U \in O^n$ such that*

$$A = U^T(\operatorname{Diag}\lambda(A))U. \tag{2.4}$$

**Proof.** Let $A$ have spectral decomposition

$$A = V^T(\operatorname{Diag} x)V,$$

for some $V \in O^n$, where $x$ is a vector of the eigenvalues of $A$, counting algebraic multiplicities. Let $P$ be the permutation matrix such that $Px \in \mathbb{R}^n_{\geq}$. That is $Px = \lambda(A)$. It follows, by Lemma 2.1.13, that

$$P(\operatorname{Diag} x)P^T = \operatorname{Diag}\lambda(A).$$

Then, since $P$ is orthogonal, we have

$$\operatorname{Diag} x = P^T(\operatorname{Diag}\lambda(A))P.$$

Substituting into the spectral decomposition of $A$ we have

$$A = V^T P^T(\operatorname{Diag}\lambda(A))PV.$$

Then, since $U := PV$ is an orthogonal matrix, we see that (2.4) holds. □

**Definition 2.3.6.** For any eigenvalue $\lambda \in \mathbb{C}$ of a matrix $A \in M_{n,n}$ The *geometric multiplicity* of $\lambda$ is the dimension of its associated eigenspace, $\dim N(A - \lambda I)$. Let

$geomult_A(\lambda)$ denote the geometric multiplicity of $\lambda$ with respect to $A$.

**Corollary 2.3.7.** *If $\lambda$ is an eigenvalue of a symmetric matrix, $A \in \mathcal{S}^n$, then the algebraic and geometric multiplicities of $\lambda$ are equal.*

**Proof.** Let $algmult_A(\lambda) = a$. Then, by Theorem 2.3.4, there exists an orthogonal $U$ such that

$$UAU^T = \begin{pmatrix} \lambda I_a & 0 \\ 0 & D \end{pmatrix}$$

where $D$ is a diagonal matrix such that $\lambda \notin \sigma(D)$. Then,

$$\operatorname{rank}(A - \lambda I) = \operatorname{rank}\begin{pmatrix} 0 & 0 \\ 0 & D - \lambda I \end{pmatrix} = \operatorname{rank}(D - \lambda I) = n - a.$$

It follows that

$$geomult_A(\lambda) = \dim N(A - \lambda I) = n - \operatorname{rank}(A - \lambda I) = a$$

as required. $\square$

**Definition 2.3.8.** Let $A \in \mathcal{S}^n$ have exactly $r$ nonzero eigenvalues and spectral decomposition

$$A = U^T \begin{pmatrix} D_{r\times r} & 0 \\ 0 & 0 \end{pmatrix} U$$

where $D$ is a diagonal matrix with the nonzero eigenvalues of $A$ on the diagonal.

Then the *Moore-Penrose generalized inverse* of $A$, denoted $A^\dagger$, is

$$A^\dagger := U^T \begin{pmatrix} D_{r\times r}^{-1} & 0 \\ 0 & 0 \end{pmatrix} U.$$

Clearly if $A$ is nonsingular then $A^\dagger = A^{-1}$.

**Lemma 2.3.9.** *For any $A \in S^n$*

$$AA^\dagger A = A$$

*and*

$$A^\dagger A A^\dagger = A^\dagger.$$

**Proof.** Directly from the definition of the Moore-Penrose inverse,

$$A^\dagger A = U^T \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix} UU^T \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} U = U^T \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} U.$$

Therefore,

$$AA^\dagger A = U^T \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} U = A$$

and

$$A^\dagger A A^\dagger = U^T \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} U = A^\dagger$$

as required. $\square$

## 2.4 Matrix norms

**Theorem 2.4.1** (Weierstrass). *Suppose $C \subset \mathbb{R}^n$ is compact and $f : C \to \mathbb{R}$ is a continuous function. Then $f$ has a global minimizer on $C$.*

**Definition 2.4.2.** A *vector norm* on $M_{m,n}$ is a function $\|\cdot\| : M_{m,n} \to \mathbb{R}$ that satisfies

(1) $\|A\| \geq 0$,

(2) $\|A\| = 0$ if and only if $A = 0$,

(3) $\|\alpha A\| = |\alpha|\|A\|$ for all scalars $\alpha$,

(4) $\|A + B\| \leq \|A\| + \|B\|$ for any $A$ and $B$.

If $m = n$ and in addition we have

(5) $\|AB\| \leq \|A\|\|B\|$ for any $A$ and $B$

then the norm is called a *matrix norm.*

**Proposition 2.4.3.** *Any two norms, $\|\cdot\|_m$ and $\|\cdot\|_n$, defined on $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively,* induce *a vector norm on $M_{m,n}$ by setting*

$$\|A\| = \max_{\|x\|_n=1} \|Ax\|_m$$

*for all $A \in M_{m,n}$. If $n = m$ then this is a matrix norm.*

**Proof.** Trivially, from the definition of a vector norm, $\|Ax\|_m \geq 0$ for all $x$ and is equal to 0 if and only if $Ax = 0$. Thus, $\|A\| \geq 0$ and is equal to 0 if and only if

$A = 0$. Therefore, (1) and (2) hold. To show Properties (3) and (4), notice from the sublinearity of the maximum we have

$$\|\alpha A\| = \max_{\|x\|_n=1} \|\alpha Ax\|_m = \max_{\|x\|_n=1} |\alpha|\|Ax\|_m = |\alpha|\|A\|$$

and

$$\|A+B\| = \max_{\|x\|_n=1} \|(A+B)x\|_m \leq \max_{\|x\|_n=1} (\|Ax\|_m + \|Bx\|_m) \leq \|A\| + \|B\|$$

for any $A$ and $B$ in $M_{m,n}$.

Notice that for any $x_0 \in \mathbb{R}^n$ we have

$$\|A\| = \max_{\|x\|_n=1} \|Ax\|_m = \max_{x\neq 0} \left\| A\frac{x}{\|x\|_n} \right\|_m \geq \frac{\|Ax_0\|_m}{\|x_0\|_n}$$

and thus

$$\|Ax_0\|_m \leq \|A\|\|x_0\|_n.$$

To show (5), suppose that $m = n$ and choose $A, B \in M_{n,n}$. Since $x \mapsto \|ABx\|_n$ is a continuous function and the unit ball $\{x \,|\, \|x\|_n = 1\}$ is compact, by Theorem 2.4.1 there exists a unit vector $\bar{x} \in \mathbb{R}^m$ such that $\|AB\bar{x}\|_n = \|AB\|$. Then,

$$\|AB\| = \|AB\bar{x}\|_n \leq \|A\|\|B\bar{x}\|_n \leq \|A\|\|B\|\|\bar{x}\|_n = \|A\|\|B\|.$$

Therefore, (5) holds and thus the norm induced on $M_{n,n}$ by a vector norm is a matrix norm. $\square$

**Lemma 2.4.4.** *For any $A \in M_{m,n}$, the $n \times n$ matrix $A^T A$ has nonnegative real eigenvalues.*

**Proof.** Let $\|\cdot\|_m$ and $\|\cdot\|_n$ be the Euclidean norms on $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively. Since $A^T A$ is a symmetric matrix, $A^T A$ has spectral decomposition

$$A^T A = U^T(\mathrm{Diag}\,\lambda(A^T A))U$$

for some $U \in O^n$. Since $U$ is orthogonal,

$$\|U^T y\|_n = \sqrt{y^T U U^T y} = \sqrt{y^T y} = \|y\|_n$$

for every $y \in \mathbb{R}^n$. Therefore,

$$\begin{aligned}\min_{\|x\|_n=1} \|Ax\|_m^2 = \min_{\|x\|_n=1} x^T U^T(\mathrm{Diag}\,\lambda(A^T A))Ux &= \min_{\|U^T y\|_n=1} y^T(\mathrm{Diag}\,\lambda(A^T A))y \\ = \min_{\|y\|_n=1} y^T(\mathrm{Diag}\,\lambda(A^T A))y &= \min_{\|y\|_n=1} \sum_{i=1}^n \lambda_i(A^T A)y_i^2.\end{aligned}$$

Now observe that for every $y \in \mathbb{R}^n$ such that $\|y\|_n = 1$, we have

$$\sum_{i=1}^n \lambda_i(A^T A)y_i^2 \geq \sum_{i=1}^n \lambda_n(A^T A)y_i^2 = \lambda_n(A^T A)$$

and thus $\|Ay\|_m^2 \geq \lambda_n(A^T A)$. This last inequality holds with equality when $y$ is a unit eigenvector of $A^T A$ corresponding to $\lambda_n(A^T A)$. It follows that

$$\lambda_n(A^T A) = \min_{\|x\|_n=1} \|Ax\|_m \geq 0$$

and, since $\lambda_n(A^TA)$ is the smallest eigenvalue of $A^TA$, every $\lambda \in \sigma(A^TA)$ is nonnegative. $\square$

**Proposition 2.4.5.** *Denote by $\|\cdot\|_2$ the norm on $M_{m,n}$ induced by the Euclidean norms on $\mathbb{R}^n$ and $\mathbb{R}^m$. Then*

$$\|A\|_2 = \max_{\lambda \in \sigma(A^TA)} \sqrt{\lambda}. \tag{2.5}$$

*We call $\|A\|_2$ the spectral norm on $M_{m,n}$. When $m = n$ this is a matrix norm.*

**Proof.** First notice that, by Lemma 2.4.4, the right-hand side of (2.5) is well-defined. By the definition we have

$$\|A\|_2 = \max_{\|x\|_n=1} \|Ax\|_m$$

where $\|\cdot\|_m$ and $\|\cdot\|_n$ are the Euclidean norms on $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively. Since $\|Ax\|_m \geq 0$ for all $x \in \mathbb{R}^n$, in order to evaluate the maximum it suffices to solve the optimization problem

$$\begin{aligned} \max \quad & \|Ax\|_m^2 \\ \text{subject to} \quad & x^Tx = 1. \end{aligned} \tag{2.6}$$

Note that, since $A^TA$ is a symmetric matrix, every eigenvalue of $A^TA$ is real. We show that the optimal value of (2.6) is the largest eigenvalue of $A^TA$.

Observe that since $A^TA$ is symmetric there exists orthogonal $U$ such that

$$A^TA = U^T(\operatorname{Diag}\lambda(A^TA))U.$$

Then, since $x^Tx = x^TU^TUx$ for every $x \in \mathbb{R}^n$, it suffices to solve the equivalent optimization problem

$$\begin{aligned} \max \quad & x^T(\operatorname{Diag}\lambda(A^TA))x \\ \text{subject to} \quad & x^Tx = 1. \end{aligned} \tag{2.7}$$

Now observe that for every $x$ such that $\|x\|_n = 1$ we have

$$x^T(\operatorname{Diag}\lambda(A^TA))x = \sum_{i=1}^{n} \lambda_i(A^TA)x_i^2 \leq \sum_{i=1}^{n} \lambda_1(A^TA)x_i^2 = \lambda_1(A^TA).$$

Therefore, $\|Ax\|_m^2 \leq \lambda_1(A^TA)$ for all $x \in \mathbb{R}^n$ with $\|x\|_n = 1$. When $x$ is a unit eigenvector of $A^TA$ corresponding to $\lambda_1(A^TA)$ the last inequality holds with equality. It follows that $\|A\|_2^2 = \lambda_1(A^TA)$ as required. □

**Lemma 2.4.6.** *For any two matrices $A \in M_{m,n}$, $B \in M_{n,m}$,*

$$\operatorname{tr}(AB) = \operatorname{tr}((AB)^T) = \operatorname{tr}(B^TA^T).$$

**Proof.** Taking the transpose of $AB$ does not alter the diagonal entries of the matrix and thus, since the trace is the sum of the diagonal entries of a matrix, the value of the trace is also unaltered. □

The following lemma is easy to verify directly.

**Lemma 2.4.7.** *For any two matrices $A, B \in \mathrm{M}_{m,n}$,*

$$\operatorname{tr}(A^TB) = \operatorname{tr}(BA^T).$$

**Theorem 2.4.8.** *The vector space $M_{m,n}$ is an inner-product space with*

$$\langle X, Y\rangle = \mathrm{tr}(X^T Y)$$

*for all $X, Y \in M_{m,n}$.*

**Proof.** We first show that $\mathrm{tr}(X^T X) \geq 0$. By definition,

$$\mathrm{tr}(X^T X) = \sum_{i=1}^{n} (X^T X)_{ii} = \sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij}^2.$$

Clearly $\mathrm{tr}(X^T X) \geq 0$ with equality if and only if $X = 0$.

The linearity of the trace function gives $\mathrm{tr}(X^T(\alpha Y)) = \alpha \mathrm{tr}(X^T Y)$ and $\mathrm{tr}(X^T(Y + Z)) = \mathrm{tr}(X^T Y) + \mathrm{tr}(X^T Z)$.

Finally, by Lemma 2.4.6, we have $\mathrm{tr}(X^T Y) = \mathrm{tr}((X^T Y)^T) = \mathrm{tr}(Y^T X)$. Therefore, $\mathrm{tr}(X^T Y)$ is an inner product on $M_{m,n}$. $\square$

**Definition 2.4.9.** The function $\|\cdot\|_F : M_{m,n} \to \mathbb{R}$ defined by

$$\|A\|_F^2 = \mathrm{tr}(A^T A) = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2$$

for all $A \in M_{m,n}$ is called the *Frobenius norm* of $A$.

**Proposition 2.4.10.** *The Frobenius norm is a vector norm on $M_{m,n}$ and a matrix norm on $M_{n,n}$.*

**Proof.** Since $(X, Y) \in M_{n,n} \times M_{n,n} \mapsto \mathrm{tr}(X^T Y) \in \mathbb{R}$ is an inner product, $\|\cdot\|_F$ is a vector norm on $M_{n,n}$.

It remains to show that $\|AB\|_F \leq \|A\|_F \|B\|_F$ for all $A, B \in M_{n,n}$ Indeed, for any two matrices $A, B \in M_{n,n}$,

$$\begin{aligned}\|AB\|_F^2 = \operatorname{tr}((AB)^T AB) = \sum_{i=1}^n \sum_{j=1}^n (AB)_{ij}^2 = \sum_{i=1}^n \sum_{j=1}^n \Big( \sum_{k=1}^n A_{ik} B_{kj} \Big)^2 \\ \leq \sum_{i=1}^n \sum_{j=1}^n \|A_{i*}\|^2 \|B_{*j}\|^2 = \Big( \sum_{i=1}^n \sum_{k=1}^n A_{ik}^2 \Big) \Big( \sum_{j=1}^n \sum_{l=1}^n B_{lj}^2 \Big) = \|A\|_F^2 \|B\|_F^2\end{aligned}$$

by the Cauchy-Schwarz inequality (Lemma 2.2.4). This completes the proof. $\square$

Unless otherwise noted $\|A\|$ will denote the Frobenius norm of $A \in M_{m,n}$.

**Theorem 2.4.11.** *The Frobenius norm on $M_{m,n}$ is invariant under orthogonal similarity transformations:*

$$\|A\|_F = \|UAV\|_F \text{ for all } A \in M_{m,n}, U \in O^m \text{ and } V \in O^n.$$

**Proof.** By Lemma 2.4.7,

$$\|UAV\|_F^2 = \operatorname{tr}(V^T A^T U^T UAV) = \operatorname{tr}(V^T A^T AV) = \operatorname{tr}(A^T AVV^T) = \operatorname{tr}(A^T A) = \|A\|_F^2$$

as required. $\square$

**Lemma 2.4.12.** *For any symmetric matrix $A \in \mathcal{S}^n$*

$$\|A\|_F = \|\lambda(A)\|$$

*where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^n$.*

**Proof.** Let $A$ have spectral decomposition $A = U^T(\operatorname{Diag}\lambda(A))U$. Then,

$$\begin{aligned}\|A\|_F^2 &= \operatorname{tr}(A^TA) = \operatorname{tr}(U^T(\operatorname{Diag}\lambda(A))^T(\operatorname{Diag}\lambda(A))U) \\ &= \operatorname{tr}((\operatorname{Diag}\lambda(A))^T(\operatorname{Diag}\lambda(A))) = \sum_{i=1}^{n}\lambda_i^2(A) = \|\lambda(A)\|^2.\end{aligned}$$

This completes the proof. □

For any $z = x + iy \in \mathbb{C}$ define $|z| := \sqrt{x^2 + y^2} \in \mathbb{R}$.

**Definition 2.4.13.** The function $\rho : M_{n,n} \to \mathbb{R}$ defined by

$$\rho(A) = \max_{\lambda\in\sigma(A)} |\lambda|$$

is called the spectral radius of $A$.

**Proposition 2.4.14.** *For any matrix norm $\|\cdot\| : M_{n,n} \to \mathbb{R}$ and any $A \in M_{n,n}$ we have*

$$\rho(A) \leq \|A\|.$$

**Proof.** Let $\lambda$ be an eigenvalue of $A$ with eigenvector $x$. Defining the $n \times n$ matrix $X$ by

$$X := [\, x \,|\, 0 \,|\, ... \,|\, 0 \,] \neq 0$$

gives $AX = \lambda X$ and thus

$$|\lambda|\|X\| = \|\lambda X\| = \|AX\| \leq \|A\|\|X\|.$$

Therefore, $|\lambda| \leq \|A\|$. Since the eigenvalue $\lambda$ was arbitrary, $\rho(A) \leq \|A\|$. □

# Chapter 3

# Linear operators

## 3.1 Matrix representations of linear operators

Let $X, Y$ be finite dimensional, real vector spaces.

**Definition 3.1.1.** A map $\mathcal{A} : X \to Y$ is *linear* if for all $x, y \in X$ and $\alpha \in \mathbb{R}$, we have

$$\mathcal{A}(\alpha x + y) = \alpha \mathcal{A}x + \mathcal{A}y.$$

When $X = Y$, a linear map $\mathcal{A} : X \to X$ is called a *linear operator*.

**Definition 3.1.2.** Let $\mathcal{A} : X \to Y$ be a linear map. Then the nullspace of $\mathcal{A}$ is the linear subspace of $X$

$$N(\mathcal{A}) := \{x \in X \mid \mathcal{A}x = 0\}$$

and the *range* of $\mathcal{A}$ is the linear subspace of $Y$

$$R(\mathcal{A}) := \{\mathcal{A}x \mid x \in X\}.$$

Denote by $(X, \{x^i\}_{i=1}^n)$ an $n$-dimensional linear space $X$ with a fixed basis

$\{x^i\}_{i=1}^n$.

**Definition 3.1.3.** If $\{x^i\}_{i=1}^n$ and $\{y^i\}_{i=1}^n$ are bases for $X$ then the *matrix representation* of the linear operator $\mathcal{A} : (X, \{x^i\}_{i=1}^n) \to (X, \{y^i\}_{i=1}^n)$ is the $n \times n$ matrix such that

$$\mathcal{A}x^i = \sum_{j=1}^n A_{ji} y^j. \tag{3.1}$$

It follows directly from the definition of the matrix representation of a linear operator that if the coordinates of $x \in X$ with respect to the basis $\{x^i\}_{i=1}^n$ are $(\alpha_1, ..., \alpha_n)^T$ then the coordinates of $\mathcal{A}x$ with respect to $\{y^i\}_{i=1}^n$ are $A(\alpha_1, ..., \alpha_n)^T = (\sum_{i=1}^n A_{1i}\alpha_i, ..., \sum_{i=1}^n A_{ni}\alpha_i)^T$.

**Lemma 3.1.4.** *Let $A \in M_{n,n}$ and let $\{x^i\}_{i=1}^n$ be a basis for $X$. Then $A$ and $\{x^i\}_{i=1}^n$ define a linear operator $\mathcal{A} : X \to X$ by*

$$\mathcal{A}x = \sum_{j=1}^n (A\alpha)_j \, x^j = \sum_{j=1}^n \sum_{i=1}^n A_{ji}\alpha_i x^j$$

*for all $x \in X$ where $\alpha$ is the coordinate vector of $x$ with respect to $\{x^i\}_{i=1}^n$.*

**Proof.** Choose $x, y \in X$ and $c \in \mathbb{R}$. Let $\alpha, \beta \in \mathbb{R}^n$ be the coordinate vectors of $x$ and $y$ respectively. Then,

$$\begin{aligned}
\mathcal{A}(x + cy) &= \sum_{j=1}^n (A(\alpha + c\beta))_j \, x^j \\
&= \sum_{j=1}^n (A\alpha)_j \, x^j + c\sum_{j=1}^n (A\beta)_j \, x^j \\
&= \mathcal{A}x + c\mathcal{A}y.
\end{aligned}$$

This completes the proof. $\square$

**Proposition 3.1.5.** *Let*

$$\mathcal{A} : (X, \{x^i\}_{i=1}^n) \to (X, \{y^i\}_{i=1}^n) \quad \textit{and} \quad \mathcal{B} : (X, \{y^i\}_{i=1}^n) \to (X, \{z^i\}_{i=1}^n)$$

*be two linear operators on $X$. If $A$ and $B$ are the matrix representations of $\mathcal{A}$ and $\mathcal{B}$, respectively, then the matrix of representation of the composition $\mathcal{BA} : (X, \{x^i\}_{i=1}^n) \to (X, \{z^i\}_{i=1}^n)$ is $BA$.*

**Proof.** Fix $i \in \{1, 2, ..., n\}$. Then, since $\mathcal{A}$ and $\mathcal{B}$ are linear operators,

$$\begin{aligned}(\mathcal{BA})x^i = \mathcal{B}(\mathcal{A}x^i) = \mathcal{B}\sum_{j=1}^n A_{ji}y^j = \sum_{j=1}^n A_{ji}(\mathcal{B}y^j) \\ = \sum_{j=1}^n A_{ji}\sum_{k=1}^n B_{kj}z^k = \sum_{k=1}^n\sum_{j=1}^n B_{kj}A_{ji}z^k = \sum_{k=1}^n (BA)_{ki}z^k\end{aligned}$$

as required. □

**Definition 3.1.6.** The *identity* operator on $X$ is defined by $\mathcal{I}d(x) = x$ for all $x \in X$.

The following lemma is trivial.

**Lemma 3.1.7.** *The matrix representation of $\mathcal{I}d : (X, \{x^i\}_{i=1}^n) \to (X, \{x^i\}_{i=1}^n)$ is $I$ for any basis $\{x^i\}_{i=1}^n$ of $X$.*

**Definition 3.1.8.** A linear operator $\mathcal{A} : X \to X$ is said to be *nonsingular* if $\mathcal{A}x = 0$ implies $x = 0$.

It is clear that if $\mathcal{A}$ is nonsingular then it is injective. Since $X$ is finite dimensional it is easy to see that, in that case, $\mathcal{A}$ is onto. Then, $\mathcal{A}$ is invertible and the inverse is denoted by $\mathcal{A}^{-1}$.

**Proposition 3.1.9.** *Let the nonsingular operator* $\mathcal{A} : (X, \{x^i\}_{i=1}^n) \to (X, \{y^i\}_{i=1}^n)$ *have matrix representation* $A$. *Then* $A^{-1}$ *is the matrix representation of* $\mathcal{A}^{-1} : (X, \{y^i\}_{i=1}^n) \to (X, \{x^i\}_{i=1}^n)$.

**Proof.** The composition $\mathcal{A}(\mathcal{A}^{-1}) = \mathcal{I}d$ has matrix representation $I$ with respect to $\{x^i\}_{i=1}^n$ by Lemma 3.1.7. Denote by $B$ the matrix representation of $\mathcal{A}^{-1}$. It follows from Proposition 3.1.5 that $BA = I$ and thus $B = A^{-1}$ as required. $\square$

The following corollary is now easy to see.

**Corollary 3.1.10.** *A linear operator* $\mathcal{A} : X \to X$ *is nonsingular if and only if it has a matrix representation that is nonsingular.*

**Corollary 3.1.11.** *Let* $\{x^i\}_{i=1}^n$ *and* $\{y^i\}_{i=1}^n$ *be two different bases of* $X$. *Let* $\mathcal{I}d : (X, \{x^i\}_{i=1}^n) \to (X, \{y^i\}_{i=1}^n)$ *have matrix representation* $B$. *Then, the matrix representation of* $\mathcal{I}d : (X, \{y^i\}_{i=1}^n) \to (X, \{x^i\}_{i=1}^n)$ *is* $B^{-1}$.

**Proposition 3.1.12.** *Let* $A_x, A_y$ *be the matrix representations of the operators* $\mathcal{A} : (X, \{x^i\}_{i=1}^n) \to (X, \{x^i\}_{i=1}^n)$ *and* $\mathcal{A} : (X, \{y^i\}_{i=1}^n) \to (X, \{y^i\}_{i=1}^n)$ *respectively. Then,*

$$A_x = BA_yB^{-1}$$

*where* $B$ *is defined as in Corollary 3.1.11.*

**Proof.** The linear operator $\mathcal{A} : (X, \{x^i\}_{i=1}^n) \to (X, \{x^i\}_{i=1}^n)$ can be represented as the composition

$$(X, \{x^i\}_{i=1}^n) \xrightarrow{\mathcal{I}d} (X, \{y^i\}_{i=1}^n) \xrightarrow{\mathcal{A}} (X, \{y^i\}_{i=1}^n) \xrightarrow{\mathcal{I}d} (X, \{x^i\}_{i=1}^n).$$

Thus, it follows from Proposition 3.1.5 and Corollary 3.1.11 that

$$A_x = BA_yB^{-1}$$

as required. $\square$

**Definition 3.1.13.** Let $\mathcal{A} : X \to X$ be a linear operator. A subspace $\mathcal{X} \subset X$ is called *invariant* under $\mathcal{A}$ if

$$\mathcal{A}\mathcal{X} \subset \mathcal{X}.$$

For any linear operator $\mathcal{A} : X \to X$ and an invariant subspace $\mathcal{X} \subset X$, we denote by $\mathcal{A}_{|\mathcal{X}}$ the linear operator $\mathcal{A}_{|\mathcal{X}} : \mathcal{X} \to \mathcal{X}$ defined by $\mathcal{A}_{|\mathcal{X}}(x) := \mathcal{A}(x)$ for all $x \in \mathcal{X}$. We call $\mathcal{A}_{|\mathcal{X}}$ the *restriction of* $\mathcal{A}$ *to* $\mathcal{X}$.

**Proposition 3.1.14.** *Fix the standard basis* $\{e^i\}_{i=1}^n$ *in* $\mathbb{R}^n$ *and let* $A$ *be the matrix representation of a linear operator* $\mathcal{A} : (\mathbb{R}^n, \{e^i\}_{i=1}^n) \to (\mathbb{R}^n, \{e^i\}_{i=1}^n)$. *Let* $\mathcal{X}$ *be a subspace of* $\mathbb{R}^n$ *of dimension* $m$. *Let* $\{y^j\}_{j=1}^m$ *be a basis for* $\mathcal{X}$. *Suppose that the coordinates of* $y^j$ *with respect to* $\{e^i\}_{i=1}^n$ *are the entries of the* $j$*-th column of the matrix* $X \in M_{n,m}$. *Then* $\mathcal{X}$ *is invariant under* $\mathcal{A}$ *if and only if there exists a matrix* $L \in M_{m,m}$ *such that*

$$AX = XL.$$

*In fact,* $L$ *is unique and is the matrix representation of*

$$\mathcal{A}_{|\mathcal{X}} : (\mathcal{X}, \{y^j\}_{j=1}^m) \to (\mathcal{X}, \{y^j\}_{j=1}^m).$$

*Moreover,*

$$\sigma(L) \subset \sigma(A).$$

**Proof.** Let $\{y^j\}_{j=1}^m$ be a fixed basis for $\mathcal{X}$ and let the coordinates of the basis vectors with respect to $\{e^i\}_{i=1}^n$ be the columns of the $n \times m$ matrix $X$. Denote the $j$-th column of $X$ by $x^j$, for each $j = 1, 2, ..., m$.

Suppose $\mathcal{X}$ is invariant under $\mathcal{A}$. Then, by the comments after Definition 3.1.3

$$\mathcal{A}y^j = \sum_{i=1}^n (Ax^j)_i e^i.$$

On the other hand, since $\mathcal{A}y^j \in \mathcal{X}$, there exists a unique $\alpha^j \in \mathbb{R}^m$ such that

$$\mathcal{A}y^j = \sum_{k=1}^m \alpha_k^j y^k = \sum_{k=1}^m \alpha_k^j \Big( \sum_{i=1}^n x_i^k e^i \Big) = \sum_{i=1}^n \Big( \sum_{k=1}^m x_i^k \alpha_k^j \Big) e^i.$$

Equating the two expressions for $\mathcal{A}y^j$ and using that $\{e^i\}_{i=1}^n$ is a basis, we obtain $Ax^j = \sum_{k=1}^m x^k \alpha_k^j$ for all $j = 1, 2, ..., m$. Then,

$$AX = [X\alpha^1 \; X\alpha^2 \; ... \; X\alpha^m] = XL$$

where $L$ is the $m \times m$ matrix with columns $\alpha^1, \alpha^2, ..., \alpha^m$.

Conversely suppose there exists an $m \times m$ matrix $L = [\alpha^1 \; \alpha^2 \; ... \; \alpha^m]$ such that $AX = XL$. Then, for each $j = 1, 2, ..., m$ we have

$$\begin{aligned}\mathcal{A}y^j &= \sum_{i=1}^n (Ax^j)_i e^i = \sum_{i=1}^n (X\alpha^j)_i e^i = \sum_{i=1}^n \Big( \sum_{k=1}^n x_i^k \alpha_k^j \Big) e^i = \sum_{k=1}^n \alpha_k^j \Big( \sum_{i=1}^n x_i^k e^i \Big) \\ &= \sum_{k=1}^n \alpha_k^j y^k.\end{aligned}$$

This shows $\mathcal{A}\mathcal{X} \subset \mathcal{X}$ and that the matrix representation of $\mathcal{A}_{|\mathcal{X}}$ with respect to the basis $\{y^j\}_{j=1}^m$ is $L$. Therefore, the matrix $L$ with the property $AX = XL$ is unique.

It remains to prove that $\sigma(L) \subset \sigma(A)$. Indeed, suppose $\lambda \in \sigma(L)$. Then, there exists a nonzero $v \in \mathbb{C}^m$ such that $Lv = \lambda v$. Multiplying each side of this equality on the left by $X$ gives $XLv = A(Xv) = \lambda(Xv)$. Since $X$ has a full column rank, $Xv \neq 0$. Therefore $\lambda$ is an eigenvalue of $A$. $\square$

## 3.2 The adjoint of a linear operator

Let $E$ be a Euclidean space.

**Lemma 3.2.1.** *Let $g : E \to \mathbb{R}$ be a linear function. Then there exists a unique vector $y \in E$ such that $g(x) = \langle x, y \rangle$ for all $x \in E$.*

**Proof.** Let $\{x^i\}_{i=1}^n$ be an orthonormal basis for $E$. Let $y := \sum_{i=1}^n g(x^i)x^i$ and define a linear function $f : E \to \mathbb{R}$ by $f(x) = \langle x, y \rangle$. Then,

$$f(x^j) = \langle x^j, \sum_{i=1}^n g(x^i)x^i \rangle = \sum_{i=1}^n g(x^i)\langle x^j, x^i \rangle = g(x^j).$$

Then, $f = g$ on $\{x^i\}_{i=1}^n$ and by the linearity of $f$ and $g$, $f = g$ on $E$.

It remains to prove the uniqueness of $y$. Suppose $g(x) = \langle x, z \rangle$ for all $x \in E$. Then, $\langle x, y \rangle = \langle x, z \rangle$ for all $x \in E$. Taking $x = y - z$ shows that $\|y - z\| = 0$ and thus $y = z$. $\square$

**Proposition 3.2.2.** *Let $\mathcal{A} : E \to E$ be a linear map. There exists a unique linear map $\mathcal{A}^* : E \to E$ such that*

$$\langle \mathcal{A}x, y \rangle = \langle x, \mathcal{A}^* y \rangle$$

*for all $x, y \in E$. We call the linear map $\mathcal{A}^* : E \to E$ the* adjoint *of $A$.*

**Proof.** We first show that $\mathcal{A}^*$ exists. Fix $y \in E$. Then, the function $f : E \to \mathbb{R}$ defined by

$$f(x) = \langle \mathcal{A}x, y \rangle \quad \forall x \in E$$

is linear by the linearity of $\mathcal{A}$ and the inner product. Thus, by Lemma 3.2.1, there exists unique $v \in E$ such that

$$\langle \mathcal{A}x, y \rangle = \langle x, v \rangle.$$

Define $\mathcal{A}^* : E \to E$ by $\mathcal{A}^* y = v$.

We next show that $\mathcal{A}^*$ is linear. Indeed, for any $y_1, y_2 \in E$ and $\alpha \in \mathbb{R}$. Then,

$$\begin{aligned}
\langle x, \mathcal{A}^*(y_1 + \alpha y_2) \rangle &= \langle \mathcal{A}x, y_1 + \alpha y_2 \rangle \\
&= \langle \mathcal{A}x, y_1 \rangle + \alpha \langle \mathcal{A}x, y_2 \rangle \\
&= \langle x, \mathcal{A}^* y_1 \rangle + \langle x, \alpha \mathcal{A}^* y_2 \rangle \\
&= \langle x, \mathcal{A}^* y_1 + \alpha \mathcal{A}^* y_2 \rangle
\end{aligned}$$

and thus $\mathcal{A}^*$ is linear.

It remains to show that $\mathcal{A}^*$ is unique. Suppose that

$$\langle \mathcal{A}x, y \rangle = \langle x, \mathcal{A}_1^* y \rangle = \langle x, \mathcal{A}_2^* y \rangle$$

for all $x \in E, y \in E$. Then,

$$0 = \langle x, (\mathcal{A}_1^* - \mathcal{A}_2^*)y \rangle \quad \forall\, x, y \in E$$

implying that $\mathcal{A}_1^* = \mathcal{A}_2^*$ as required. $\square$

**Lemma 3.2.3.** *If $\mathcal{A} : E \to E$ is a nonsingular linear operator then $\mathcal{A}^* : E \to E$ is nonsingular as well.*

**Proof.** Suppose that $y \in E$ is such that $\mathcal{A}^* y = 0$. From the definition of the adjoint operator, for every $x \in E$ we have $0 = \langle x, \mathcal{A}^* y \rangle = \langle \mathcal{A}x, y \rangle$. Since $\mathcal{A}$ is onto, we conclude that $y = 0$. $\square$

**Lemma 3.2.4.** *For any nonsingular linear operator $\mathcal{A} : E \to E$ we have*

$$(\mathcal{A}^{-1})^* = (\mathcal{A}^*)^{-1}.$$

**Proof.** Since $\mathcal{A}$ is nonsingular, $\mathcal{A}^*$ is nonsingular by Lemma 3.2.3 and thus $(\mathcal{A}^*)^{-1}$ is well-defined. Choose $x \in E$ and $u \in E$ and define $y := \mathcal{A}x$, $v := \mathcal{A}^* u$. By the definition of the adjoint

$$\langle \mathcal{A}x, u \rangle = \langle x, \mathcal{A}^* u \rangle = \langle \mathcal{A}^{-1} y, v \rangle = \langle y, (\mathcal{A}^{-1})^* v \rangle = \langle \mathcal{A}x, (\mathcal{A}^{-1})^* v \rangle$$

and thus

$$u = (\mathcal{A}^*)^{-1} v = (\mathcal{A}^{-1})^* v.$$

Therefore, since the choice of $x$ and $u$ was arbitrary and $\mathcal{A}^*$ is onto, we obtain $(\mathcal{A}^*)^{-1} = (\mathcal{A}^{-1})^*$ as required. $\square$

**Lemma 3.2.5.** *Let $\{x^i\}_{i=1}^k$ be an orthonormal subset of $E$. Then every vector $y$ in* span $\{x^1, x^2, ..., x^k\}$ *can be expressed as*

$$y = \sum_{i=1}^{k} \langle y, x^i \rangle x^i.$$

**Proof.** Fix $y \in \text{span}\,\{x^1, x^2, ..., x^k\}$. There exists $\alpha \in \mathbb{R}^k$ such that $y = \sum_{i=1}^{k} \alpha_i x^i$. Then, for $j = 1, 2, ...k$, it follows that

$$\langle y, x^i \rangle = \langle \sum_{j=1}^{k} \alpha_j x^j, x^i \rangle = \alpha_i.$$

Thus,

$$y = \sum_{i=1}^{k} \alpha_i x^i = \sum_{i=1}^{k} \langle y, x^i \rangle x^i$$

as required. □

**Corollary 3.2.6.** *Let $\{x^i\}_{i=1}^n$ and $\{y^i\}_{i=1}^n$ be orthonormal bases for $E$. If $A$ is the matrix representation of $\mathcal{A} : (E, \{x^i\}_{i=1}^n) \to (E, \{y^i\}_{i=1}^n)$ then*

$$A_{ij} = \langle \mathcal{A} x^j, y^i \rangle$$

*for all $i, j = 1, ..., n$.*

**Proof.** By Lemma 3.2.5,

$$\mathcal{A} x^j = \sum_{i=1}^{n} \langle \mathcal{A} x^j, y^i \rangle y^i.$$

However, since $A$ is the matrix representation of $\mathcal{A}$, we have

$$\mathcal{A}x^j = \sum_{i=1}^{n} A_{ij} y^i$$

and thus $A_{ij} = \langle \mathcal{A}x^j, y^i \rangle$ as required. □

**Proposition 3.2.7.** *Let $\{x^i\}_{i=1}^n$ and $\{y^i\}_{i=1}^n$ be orthonormal bases for $E$. Then if $\mathcal{A} : (E, \{x^i\}_{i=1}^n) \to (E, \{y^i\}_{i=1}^n)$ has matrix representation $A$ then the matrix representation of $\mathcal{A}^* : (E, \{y^i\}_{i=1}^n) \to (E, \{x^i\}_{i=1}^n)$ is $A^T$.*

**Proof.** Let $\mathcal{A}^* : (E, \{y^i\}_{i=1}^n) \to (E, \{x^i\}_{i=1}^n)$ have matrix representation $B$. Then, by Corollary 3.2.6

$$A_{ij} = \langle \mathcal{A}x^j, y^i \rangle = \langle x^j, \mathcal{A}^* y^i \rangle = B_{ji}$$

and thus $A^T = B$. □

**Definition 3.2.8.** Let $E$ be a Euclidean space. Then the linear operator $\mathcal{A} : E \to E$ is called *self-adjoint* if $\mathcal{A} = \mathcal{A}^*$.

**Lemma 3.2.9.** *Let $\{x^i\}_{i=1}^n$ be an orthonormal basis for $E$ and let $\mathcal{A} : (E, \{x^i\}_{i=1}^n) \to (E, \{x^i\}_{i=1}^n)$ be a self-adjoint linear operator. Then the matrix representation of $\mathcal{A}$ is a symmetric matrix.*

**Proof.** Let $\mathcal{A}$ have matrix representation $A$. Then, since $\mathcal{A} = \mathcal{A}^*$, it follows from Proposition 3.2.7 that $A = A^T$. □

**Lemma 3.2.10.** *Let $A \in \mathcal{S}^n$ and let $\{x^i\}_{i=1}^n$ be an orthonormal basis for the Euclidean space $E$. Then $A$ and $\{x^i\}_{i=1}^n$ define a self-adjoint linear operator by*

$$\mathcal{A}x = \sum_{j=1}^{n} (\mathcal{A}\alpha)_j x^j = \sum_{j=1}^{n} \sum_{i=1}^{n} A_{ji} \alpha_i x^j$$

*for all $x \in X$ where $\alpha$ is the coordinate vector of $x$ with respect to $\{x^i\}_{i=1}^n$.*

**Proof.** By Lemma 3.1.4 we have that $\mathcal{A}$ and $\{x^i\}_{i=1}^n$ define a linear operator $\mathcal{A} : (E, \{x^i\}_{i=1}^n) \to (E, \{x^i\}_{i=1}^n)$ such that

$$\mathcal{A}x = \sum_{j=1}^{n} (\mathcal{A}\alpha)_j x^j = \sum_{j=1}^{n} \sum_{i=1}^{n} A_{ji} \alpha_i x^j$$

for all $x \in E$ where $\alpha$ is the coordinate vector of $x$ with respect to $\{x^i\}_{i=1}^n$.

It remains to show that the operator $\mathcal{A}$ is self-adjoint. Indeed, $\mathcal{A}$ has matrix representation $A$ and thus, by Corollary 3.2.6, $A_{ij} = \langle \mathcal{A}x^i, x^j \rangle$ for all $i, j = 1, 2, ..., n$. Then, since $A$ is symmetric,

$$A_{ij} = \langle \mathcal{A}x^i, x^j \rangle = \langle \mathcal{A}x^j, x_i \rangle = \langle \mathcal{A}^* x^i, x^j \rangle = A_{ji}.$$

Therefore $\mathcal{A} = \mathcal{A}^*$ as required. $\square$

**Lemma 3.2.11.** *Let $\mathcal{X} \subset \mathbb{R}^n$ be invariant under the self-adjoint operator $\mathcal{A} : (\mathbb{R}^n, \{e^i\}_{i=1}^n) \to (\mathbb{R}^n, \{e^i\}_{i=1}^n)$. Let $\{y^i\}_{i=1}^m$ be an orthonormal basis for $\mathcal{X}$. Then the operator $\mathcal{A}_{|\mathcal{X}} : (\mathcal{X}, \{y^i\}_{i=1}^m) \to (\mathcal{X}, \{y^i\}_{i=1}^m)$ is self-adjoint.*

**Proof.** Let $A$ and $L$ be the matrix representations of $\mathcal{A}$ and $\mathcal{A}_{|\mathcal{X}}$ respectively. Let the coordinates of $\{y^i\}_{i=1}^m$ with respect to $\{e^i\}_{i=1}^n$ be the columns of the $n \times m$ matrix $X$. Then, by Proposition 3.1.14,

$$AX = XL. \tag{3.2}$$

Since $\{y^i\}_{i=1}^m$ is an orthonormal set in $\mathbb{R}^n$, $X^T X = I_m$ and thus multiplying (3.2) on the left by $X^T$ gives

$$X^T A X = L.$$

Consequently, since $A$ is symmetric by Lemma 3.2.9, $L$ is a symmetric matrix. Then, by Corollary 3.2.6,

$$L_{ij} = \langle (\mathcal{A}_{|\mathcal{X}}) y^i, y^j \rangle = \langle (\mathcal{A}_{|\mathcal{X}}) y^j, y^i \rangle = \langle (\mathcal{A}_{|\mathcal{X}})^* y^i, y^j \rangle = L_{ji}$$

for each $i, j = 1, 2, ..., m$ and thus $\mathcal{A}_{|\mathcal{X}} = (\mathcal{A}_{|\mathcal{X}})^*$ as required. □

## 3.3 Eigenvalues and norms of linear operators

**Definition 3.3.1.** For any linear operator $\mathcal{A} : X \to X$, $\lambda \in \mathbb{R}$ is an *eigenvalue* of $\mathcal{A}$ if $\mathcal{A}x = \lambda x$ for some nonzero $x \in X$; $x$ is called an *eigenvector* of $\mathcal{A}$.

For any linear operator $\mathcal{A} : X \to X$, let $\sigma(\mathcal{A})$ denote the set of eigenvalues of $\mathcal{A}$, called the *spectrum* of $\mathcal{A}$.

**Lemma 3.3.2.** *Let $\{x^i\}_{i=1}^n$ and $\{y^i\}_{i=1}^n$ be bases of $X$. Let $A$ be the matrix representation of the linear operator $\mathcal{A} : (X, \{x^i\}_{i=1}^n) \to (X, \{y^i\}_{i=1}^n)$. Then $\lambda \in \sigma(\mathcal{A})$ if and only if $\lambda \in \sigma(A)$.*

**Proof.** If $\mathcal{A}$ has matrix representation $A$ then $\mathcal{A} - \lambda \mathcal{I}d$ has matrix representation $A - \lambda I$. Then,

$$\begin{aligned} \lambda \in \sigma(\mathcal{A}) &\Leftrightarrow \mathcal{A} - \lambda \mathcal{I}d \text{ is singular} \\ &\Leftrightarrow A - \lambda I \text{ is singular} \\ &\Leftrightarrow \det(A - \lambda I) = 0 \end{aligned}$$

$$\Leftrightarrow \lambda \in \sigma(A)$$

by Corollary 3.1.10 and Lemma 2.1.2. □

The following two properties are easy to verify.

**Lemma 3.3.3.** *Let $\mathcal{A} : X \to X$ be a linear operator. Then $\sigma(\mathcal{A}^2) = \{\lambda^2 \mid \lambda \in \sigma(\mathcal{A})\}$ and if $\mathcal{A}$ is invertible $\sigma(\mathcal{A}^{-1}) = \{\lambda^{-1} \mid \lambda \in \sigma(\mathcal{A})\}$.*

Let $E$ and $Y$ be Euclidean spaces. Let $\mathcal{L}(E, Y)$ denote the real vector space of all linear maps from $E$ to $Y$.

**Definition 3.3.4.** Let $\|\cdot\|_E$ and $\|\cdot\|_Y$ be vector norms on $E$ and $Y$ respectively. Then the function $\|\cdot\| : \mathcal{L}(E, Y) \to \mathbb{R}$ defined by

$$\|\mathcal{A}\| = \max_{\|x\|_E=1} \|\mathcal{A}x\|_Y \tag{3.3}$$

for all $\mathcal{A} \in \mathcal{L}(E, Y)$ is called the *operator norm* induced by $\|\cdot\|_E$ and $\|\cdot\|_Y$.

The proof of the following lemma is analogous to the proof of Proposition 3.3.6.

**Lemma 3.3.5.** *Let $\|\cdot\|_E$ and $\|\cdot\|_Y$ be vector norms on $E$ and $Y$ respectively. Equation (3.3) defines a vector norm on $\mathcal{L}(E, Y)$ such that for any $\mathcal{A} \in \mathcal{L}(E, Y)$ and any $x \in E$*

$$\|\mathcal{A}x\|_Y \leq \|\mathcal{A}\|\|x\|_E.$$

**Proposition 3.3.6.** *The operator norm on $\mathcal{L}(M_{m,n}, M_{m,n})$, induced by the Frobenius norm $\|\cdot\|_F$ on $M_{m,n}$ is*

$$\|\mathcal{A}\|_2 := \max_{\lambda \in \sigma(\mathcal{A}^*\mathcal{A})} \sqrt{\lambda}.$$

*We call this the spectral norm on $\mathcal{L}(M_{m,n}, M_{m,n})$.*

**Proof.** Consider the standard basis $\{E^{ij}\}_{i,j=1}^{m,n}$ in $M_{m,n}$, defined by $E^{ij}_{kl} := \delta_{ik}\delta_{jl}$ for all $i, k = 1, 2, ..., m$ and $j, l = 1, 2, ..., n$. Let $A \in M_{mn,mn}$ be the representation of $\mathcal{A}$ with respect to this basis, then $\mathcal{A}X = AX$ for all $X \in M_{m,n}$. By Proposition 2.4.5, we obtain

$$\|\mathcal{A}\|_2 = \max_{\|X\|_F=1} \|\mathcal{A}X\|_F = \max_{\|X\|_F=1} \|AX\|_F = \max_{\lambda\in\sigma(A^TA)} \sqrt{\lambda} = \max_{\lambda\in\sigma(\mathcal{A}^*\mathcal{A})} \sqrt{\lambda},$$

where in the last equality we used Lemma 3.3.2. $\square$

# Chapter 4

# Stewart's theorem

## 4.1 The operator $X \mapsto XA - BX$

The proof of the following lemma is modified from [8, Chapter V, Theorem 1.3]

**Lemma 4.1.1.** *For any $A \in \mathcal{S}^n, B \in \mathcal{S}^m$, the linear map $\mathcal{T} : M_{m,n} \to M_{m,n}$ defined by*

$$\mathcal{T}(X) = XA - BX \quad \forall\, X \in M_{m,n}$$

*is nonsingular if and only if*

$$\sigma(A) \cap \sigma(B) = \emptyset,$$

*where $\sigma(A), \sigma(B)$ are the spectrums of $A$ and $B$ respectively.*

**Proof.** Suppose $\lambda \in \sigma(A) \cap \sigma(B)$. Then, there exists nonzero $p \in \mathbb{R}^n$ and $q \in \mathbb{R}^m$ such that $Ap = \lambda p$ and $Bq = \lambda q$. Taking the transpose of the first equation gives $p^T A = \lambda p^T$ since $A$ is symmetric. Let $X = qp^T$. Then,

$$\mathcal{T}(X) = (qp^T)A - B(qp^T) = \lambda qp^T - \lambda qp^T = 0$$

and thus $\mathcal{T}$ is singular since $X$ is nonzero.

Now assume $\sigma(A) \cap \sigma(B) = \emptyset$. It remains to show that the system

$$XA - BX = C \tag{4.1}$$

has a unique solution for every $C \in M_{m,n}$. Fix $C \in M_{m,n}$ and let the spectral decomposition of $A$ be $A = V^T(\operatorname{Diag} \lambda(A))V$. Then, setting $Y = XV^T = [y^1 ... y^n]$ and $D = CV^T = [d^1 ... d^n]$ and multiplying (4.1) on the right by $V^T$ gives the equivalent system

$$Y(\operatorname{Diag} \lambda(A)) - BY = D. \tag{4.2}$$

Since $\operatorname{Diag} \lambda(A)$ is a diagonal matrix, the $i$-th column on each side of (4.2) is

$$(\lambda_i(A)I - B)y^i = d^i \tag{4.3}$$

for each $i = 1, ..., n$. Since $\lambda_i(A) \notin \sigma(B)$, the matrix $\lambda_i(A)I - B$ is nonsingular. Therefore, $y^i$ is the unique solution of (4.3). Thus, there is a unique solution $X = YV$ of (4.1) implying that $\mathcal{T}$ is nonsingular. $\square$

**Corollary 4.1.2.** *For any $A \in \mathcal{S}^n, B \in \mathcal{S}^m$,*

$$\sigma(\mathcal{T}) = \sigma(A) - \sigma(B).$$

*Furthermore, if $\mathcal{T}$ is nonsingular then,*

$$\|\mathcal{T}^{-1}\|_2 = \Big( \min_{\substack{i=1,...,n \\ j=1,...,m}} |\lambda_i(A) - \lambda_j(B)| \Big)^{-1} \tag{4.4}$$

*where* $\|\cdot\|_2$ *is the spectral norm on* $\mathcal{L}(M_{m,n}, M_{m,n})$.

**Proof.** If $v \in \sigma(\mathcal{T})$, then there is an $X \neq 0$ such that $XA - BX = vX$, or $XA-(B+vI)X = 0$. That is, the operator $X \mapsto XA-(B+vI)X$ is singular. Then, by Lemma 4.1.1, $\sigma(A)$ and $\sigma(B+vI)$ have a common element. That is, $v = \lambda - \mu$ for some $\lambda \in \sigma(A)$ and $\mu \in \sigma(B)$. Therefore, $\sigma(\mathcal{T}) \subset \sigma(A) - \sigma(B)$.

Now choose $v \in \sigma(A) - \sigma(B)$. That is, $v = \lambda - \mu$, for some $\lambda \in \sigma(A)$ and $\mu \in \sigma(B)$. Then, $\sigma(A)$ and $\sigma(B+vI)$ have the common element $\lambda$, and $XA - (B+vI)X = 0$ for some nonzero $X$. Rearranging gives $XA - BX = vX$ and thus $v \in \sigma(\mathcal{T})$.

It remains to show that (4.4) holds. Observe that $\mathcal{T}$ is self-adjoint with respect to the inner product $\langle X, Y\rangle = \mathrm{tr}(X^T Y)$ since

$$\begin{aligned}\langle \mathcal{T}(X), Y\rangle = \langle XA - BX, Y\rangle = \langle XA, Y\rangle - \langle BX, Y\rangle = \mathrm{tr}(AX^TY) - \mathrm{tr}(X^TBY)\\ = \mathrm{tr}(X^T(YA - BY)) = \langle X, YA - BY\rangle = \langle X, \mathcal{T}(Y)\rangle\end{aligned}$$

for any $X, Y \in M_{m,n}$. Thus, by Lemmas 3.2.4 and 3.3.6, it follows that

$$\begin{aligned}\|\mathcal{T}^{-1}\|_2 &= \max\{\sqrt{\lambda} \mid \lambda \in \sigma((\mathcal{T}^{-1})^*\mathcal{T}^{-1})\} = \max\{\sqrt{\lambda} \mid \lambda \in \sigma((\mathcal{T}^*)^{-1}\mathcal{T}^{-1})\}\\ &= \max\{|\lambda| \mid \lambda \in \sigma(\mathcal{T}^{-1})\} = \max\{|\lambda|^{-1} \mid \lambda \in \sigma(\mathcal{T})\}\\ &= \Big(\min_{\substack{i=1,\ldots,n\\ j=1,\ldots,m}} |\lambda_i(A) - \lambda_j(B)|\Big)^{-1},\end{aligned}$$

where we used Lemma 3.3.3. $\square$

## 4.2 Stewart's theorem

In the next theorem we use the convention $1/0 = \infty$. The proof of the following theorem is modified from [8, Chapter V, Theorem 2.11]

**Theorem 4.2.1.** *Let $\mathcal{T} : M_{m,n} \to M_{m,n}$ be an invertible linear operator and let $\varphi : M_{m,n} \to M_{m,n}$ be a map that satisfies $\varphi(0) = 0$ and for some $\eta \geq 0$*

$$\|\varphi(X) - \varphi(Y)\|_F \leq \eta \max\{\|X\|_F, \|Y\|_F\}\|X - Y\|_F \quad \textit{for all} \quad X, Y \in M_{m,n}. \tag{4.5}$$

*For any $G \in M_{m,n}$, with $\|G\|_F \leq 1/(4\eta\|\mathcal{T}^{-1}\|_2^2)$, the equation*

$$\mathcal{T}(X) = G + \varphi(X) \tag{4.6}$$

*has a unique solution in the open ball $B(0, 1/(\eta\|\mathcal{T}^{-1}\|_2))$. The sequence defined by $X_0 = 0$ and*

$$X_{k+1} = \mathcal{T}^{-1}(G + \varphi(X_k)), \;\; k = 0, 1, ... \tag{4.7}$$

*converges to that unique solution $X$ satisfying, in addition, the inequality*

$$\|X\|_F \leq 2\|G\|_F\|\mathcal{T}^{-1}\|_2.$$

**Proof.** Let $\delta := \|\mathcal{T}^{-1}\|_2^{-1}$ and $\gamma := \|G\|_F$. To show the uniqueness part, note first that the case $\eta = 0$ is trivial. Suppose that $\eta > 0$ and that (4.6) has two distinct solutions $X$ and $Y$ both in the open ball $B(0, \delta/\eta)$. Subtracting

$$X = \mathcal{T}^{-1}(G + \varphi(X)) \quad \text{and} \quad Y = \mathcal{T}^{-1}(G + \varphi(Y))$$

from each other and taking norms on both sides gives

$$\|X - Y\|_F \leq \|\mathcal{T}^{-1}\|_2 \|\varphi(X) - \varphi(Y)\|_F \leq \eta \|\mathcal{T}^{-1}\|_2 \max\{\|X\|_F, \|Y\|_F\} \|X - Y\|_F$$

by Lemma 3.3.5. Equivalently, $\delta/\eta \leq \max\{\|X\|_F, \|Y\|_F\}$ thus reaching a contradiction.

Notice that from $\varphi(0) = 0$ and (4.5) we have $\|\varphi(X)\|_F \leq \eta \|X\|_F^2$ for any $X \in M_{m,n}$. Using the recursive relationship (4.7) gives

$$\|X_{k+1}\|_F \leq \|\mathcal{T}^{-1}\|_2 (\|G\|_F + \|\varphi(X_k)\|_F) \leq \frac{\gamma}{\delta} + \frac{\eta}{\delta} \|X_k\|_F^2,$$

again by Lemma 3.3.5. Consequently, if we set $\xi_0 = 0$ and

$$\xi_{k+1} = \frac{\gamma}{\delta} + \frac{\eta}{\delta} \xi_k^2, \quad k = 0, 1, ... \tag{4.8}$$

then $\|X_k\|_F \leq \xi_k$. Moreover, the sequence $\{\xi_k\}$ is nondecreasing. Indeed, note that $\xi_0 = 0$ and $\xi_1 = \gamma/\delta \geq 0$. Supposing that $\xi_k = \xi_{k-1} + \beta$, for some $\beta \geq 0$, we see that

$$\xi_{k+1} = \frac{\gamma}{\delta} + \frac{\eta}{\delta} \xi_k^2 = \frac{\gamma}{\delta} + \frac{\eta}{\delta} (\xi_{k-1} + \beta)^2 \geq \frac{\gamma}{\delta} + \frac{\eta}{\delta} \xi_{k-1}^2 = \xi_k,$$

completing the inductive argument. Next, if $\delta^2 - 4\eta\gamma \geq 0$, the quadratic equation

$$\frac{\gamma}{\delta} + \frac{\eta}{\delta}\xi^2 = \xi \tag{4.9}$$

has nonnegative roots—the fixed points of the function on the left-hand side. Notice that the condition $\delta^2 - 4\eta\gamma \geq 0$ is exactly the condition $\|G\|_F \leq 1/(4\eta\|\mathcal{T}^{-1}\|_2^2)$. Next, denote by $\xi_* := (\delta - \sqrt{\delta^2 - 4\gamma\eta})/(2\eta)$ the smaller of the two roots. If $\xi_k \leq \xi_*$, then

$$\xi_{k+1} = \frac{\gamma}{\delta} + \frac{\eta}{\delta}\xi_k^2 \leq \frac{\gamma}{\delta} + \frac{\eta}{\delta}\xi_*^2 = \xi_*$$

and since $0 = \xi_0 \leq \xi_*$, the sequence $\{\xi_k\}$ is bounded above by $\xi_*$. Hence, it converges to some $\bar{\xi} \leq \xi_*$. Taking the limit on both sides of (4.8) as $k \to \infty$ shows that $\bar{\xi}$ is a root of (4.9). Since $\xi_*$ is the smallest root of (4.9) we obtain that $\xi_* \leq \bar{\xi} \leq \xi_*$. That is, the sequence $\{\xi_k\}$ converges to $\xi_*$ and

$$\|X_k\|_F \leq \xi_* \leq \frac{\delta}{2\eta}. \tag{4.10}$$

The last inequality is necessary to show that the sequence $\{X_k\}$ converges since

$$\begin{aligned}
\|X_{k+1} - X_k\|_F &\leq \|T^{-1}\|_2 \|\varphi(X_k) - \varphi(X_{k-1})\|_F \\
&\leq \frac{\eta}{\delta} \max\{\|X_k\|_F, \|X_{k-1}\|_F\} \|X_k - X_{k-1}\|_F \\
&\leq \frac{1}{2}\|X_k - X_{k-1}\|_F.
\end{aligned}$$

Hence

$$\|X_{k+1} - X_k\|_F \leq \frac{1}{2^k}\|X_1 - X_0\|_F \leq \frac{\delta}{\eta 2^{k+1}}.$$

Therefore $\{X_k\}$ is a Cauchy sequence and must have a limit $X$. Inequality (4.10) shows that $X$ is in the open ball $B(0, \delta/\eta)$. It also shows that

$$\|X\|_F \leq \xi_* = \frac{\delta - \sqrt{\delta^2 - 4\gamma\eta}}{2\eta} = \frac{2\gamma}{\delta + \sqrt{\delta^2 - 4\gamma\eta}} \leq \frac{2\gamma}{\delta} = 2\|G\|_F\|\mathcal{T}^{-1}\|_2.$$

This concludes the proof of the theorem. □

Let the eigenvalues of $A \in \mathcal{S}^n$, counting multiplicities, be denoted by

$$\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A).$$

Denote by $r_m$ the multiplicity of the $m$-th largest eigenvalue, $\lambda_m(A)$. Let $m', m''$ be the largest and smallest index respectively such that

$$\lambda_{m'}(A) > \lambda_m(A) > \lambda_{m''}(A).$$

If the left and/or the right inequality is never satisfied we let $m' := 0$ and/or $m'' := n+1$ respectively. Thus $r_m = m'' - m' - 1$. Define

$$i_m := m - m'. \tag{4.11}$$

In other words, $i_m$ is the number of eigenvalues, ranking before $\lambda_m$ and including $\lambda_m$, which are equal to $\lambda_m$.

**Definition 4.2.2.** Let $E$ and $Y$ be Euclidean spaces and let $\|\cdot\|_E$ and $\|\cdot\|_Y$ be the norms induced by the inner products on $E$ and $Y$ respectively. Then, for any $f : \mathbb{R} \to E$ and $g : \mathbb{R} \to Y$,

$$f(x) = O(g(x)) \text{ as } x \to a$$

if there exist constants $C$ and $\epsilon > 0$ such that

$$\|f(x)\|_E < C\|g(x)\|_Y \quad \text{for all} \quad x \in (a-\epsilon, a+\epsilon).$$

**Definition 4.2.3.** Let $E$ and $Y$ be Euclidean spaces and let $\|\cdot\|_E$ and $\|\cdot\|_Y$ be the norms induced by the inner products on $E$ and $Y$ respectively. Then, for any $f : \mathbb{R} \to E$ and $g : \mathbb{R} \to Y$,

$$f(x) = o(g(x)) \text{ as } x \to a$$

if for every constant $C > 0$, there is an $\epsilon > 0$ such that

$$\|f(x)\|_E \leq C\|g(x)\|_Y \quad \text{for all} \quad x \in (a-\epsilon, a+\epsilon).$$

**Theorem 4.2.4** (Stewart)**.** *Let $A \in \mathcal{S}^n$ and $X = (X_1\ X_2) \in O^n$ be such that*

$$(X_1\ X_2)^T A (X_1\ X_2) = \begin{pmatrix} L_1 & 0 \\ 0 & L_2 \end{pmatrix}, \tag{4.12}$$

*where* $L_1 = \lambda_m(A) I_{r_m}$ *and* $L_2 = \mathrm{Diag}\,(\lambda_1(A), ..., \lambda_{m'}(A), \lambda_{m''}(A), ..., \lambda_n(A))$. *Given a perturbation matrix* $E \in \mathcal{S}^n$, *let*

$$\begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix} := (X_1\ X_2)^T E (X_1\ X_2). \tag{4.13}$$

*For all* $E$ *close enough to zero, there exists a matrix* $P = O(E)$ *solving the equation*

$$P(L_1 + E_{11}) - (L_2 + E_{22})P = E_{21} - P E_{12} P. \tag{4.14}$$

*Furthermore, the columns of* $\hat{X} := (X_1 + X_2 P)(I_{r_m} + P^T P)^{-1/2}$ *form an orthonormal basis of* $R(\hat{X})$, *an invariant subspace of* $A + E$.[1] *The matrix representation of the restriction of* $A + E$ *to* $R(\hat{X})$, *with respect to the basis determined by the columns of* $\hat{X}$, *is*

$$\hat{L} = (I_{r_m} + P^T P)^{1/2} (L_1 + E_{11} + E_{12} P)(I_{r_m} + P^T P)^{-1/2}. \tag{4.15}$$

*The matrix* $\hat{L}$ *is symmetric.*

**Proof.** Define the linear operator $\mathcal{T} : M_{n-r_m, r_m} \to M_{n-r_m, r_m}$ by

$$\mathcal{T}(P) := P(L_1 + E_{11}) - (L_2 + E_{22})P.$$

By Corollary 4.1.2, $\sigma(\mathcal{T}) = \sigma(L_1 + E_{11}) - \sigma(L_2 + E_{22})$ and using the continuity of the eigenvalues, whenever $E$ is close to 0, all eigenvalues of $\mathcal{T}$ are nonzero, thus, it is invertible.

[1] Here, we treat $A + E$ as the matrix representation of a linear operator on $\mathbb{R}^n$ with respect to the standard basis $\{e^i\}_{i=1}^n$.

Next, define the quadratic map $\varphi : M_{n-r_m,r_m} \to M_{n-r_m,r_m}$ by

$$\varphi(P) := -PE_{12}P.$$

With $\eta := 2\|E_{12}\|$, for any $X, Y \in M_{n-r_m,r_m}$ we get

$$\begin{aligned}
\|\varphi(X) - \varphi(Y)\|_F &= \|XE_{12}X - YE_{12}Y\|_F \\
&\leq \|(X - Y)E_{12}X\|_F + \|YE_{12}(X - Y)\|_F \\
&= \|X - Y\|_F\|E_{12}\|_F\|X\|_F + \|Y\|_F\|E_{12}\|_F\|X - Y\|_F \\
&\leq \eta \max\{\|X\|_F, \|Y\|_F\}\|X - Y\|_F.
\end{aligned}$$

For any matrix $E$ with $\|E\|_F \leq 1/(2\|\mathcal{T}^{-1}\|_2)$ we have $\|E_{21}\|_F \leq 1/(4\eta\|\mathcal{T}^{-1}\|_2^2)$ since $\eta = 2\|E_{21}\|_F$ and $2\|E_{21}\|_F^2 \leq \|E\|_F^2$, using that the matrix $E$ is symmetric. Thus, by Theorem 4.2.1, for all $E$ with $\|E\|_F \leq 1/(2\|\mathcal{T}^{-1}\|_2)$, Equation (4.14) has a solution $P$ with $\|P\|_F \leq 2\|E_{21}\|_F\|\mathcal{T}^{-1}\|_2 \leq 2\|E\|_F\|\mathcal{T}^{-1}\|_2$. In other words, $P = O(E)$.

Suppose now that $\|E\|_F \leq 1/(2\|\mathcal{T}^{-1}\|_2)$ and $P$ is a solution of (4.14). It is easy to see that $\hat{X}^T\hat{X} = I_{r_m}$, since $X = (X_1\, X_2)$ is orthogonal, showing that the columns of $\hat{X}$ are orthonormal.

We now prove that $R(\hat{X})$ is an invariant subspace of $A + E$. By Proposition 3.1.14, it is enough to show that the matrix $\hat{L}$ satisfies $(A+E)\hat{X} = \hat{X}\hat{L}$. Indeed, by (4.12) we have $A = X_1L_1X_1^T + X_2L_2X_2^T$ and by (4.13) we have

$$E = X_1E_{11}X_1^T + X_2E_{21}X_1^T + X_1E_{12}X_2^T + X_2E_{22}X_2^T.$$

Substituting those two formulae below, using that $X = (X_1\, X_2)$ is orthogonal and that $P$ satisfies (4.14), we find

$$
\begin{aligned}
(A+E)\hat{X}(I_{r_m} + P^T P)^{1/2} &= (A+E)(X_1 + X_2 P) \\
&= X_1 L_1 + X_2 L_2 P + X_1 E_{11} + X_2 E_{21} + X_1 E_{12} P + X_2 E_{22} P \\
&= X_1(L_1 + E_{11} + E_{12}P) + X_2(E_{21} + (L_2 + E_{22})P) \\
&= X_1(L_1 + E_{11} + E_{12}P) + X_2(P(L_1 + E_{11}) + PE_{12}P) \\
&= X_1(L_1 + E_{11} + E_{12}P) + X_2 P(L_1 + E_{11} + E_{12}P) \\
&= \hat{X}\hat{L}(I_{r_m} + P^T P)^{1/2}.
\end{aligned}
$$

With the standard basis in $\mathbb{R}^n$ fixed, the symmetric matrix $A+E$ defines a self-adjoint operator, by Lemma 3.2.10. Its restriction to the invariant subspace $R(\hat{X})$ is also self-adjoint, by Lemma 3.2.11. Matrix $\hat{L}$ is the matrix representation of that restriction with respect to the basis defined by the columns of $\hat{X}$, by Proposition 3.1.14. Since the columns of $\hat{X}$ form an orthonormal basis of $R(\hat{X})$, matrix $\hat{L}$ is symmetric, by Lemma 3.2.9. $\square$

# Chapter 5

# Taylor expansions of the spectrum

## 5.1 Auxiliary results

We begin this chapter with an important definition.

**Definition 5.1.1.** Let $K \geq 0$. We say that a map $f : \mathbb{R} \to Y$ has a (one sided) *Taylor expansion of order* $K$ *at* $t = 0^+$ if there exists $a_0, a_1, ..., a_K \in Y$ and $\epsilon > 0$ such that

$$f(t) = a_0 + ta_1 + \cdots + t^K a_K + o(t^K) \text{ for all } t \in [0, \epsilon).$$

We say that a map $f : \mathbb{R} \to Y$ is *analytic at* $t = 0$, if it can be represented by an infinite power series for all $t$ close to 0.

Notice that, a Taylor expansion of order 0 implies that $f(t)$ has a limit as $t \downarrow 0$.

**Lemma 5.1.2.** *Let* $A \in \mathcal{S}^n$ *and let* $1 \leq m \leq n$. *Then, for all* $E \in \mathcal{S}^n$ *close enough to zero we have*

$$\lambda_m(A + E) = \lambda_{i_m}(\hat{L}),$$

*where* $i_m$ *is defined by (4.11) and* $\hat{L}$ *by (4.15).*

**Proof.** Since, by Proposition 3.1.14, we have the inclusion $\sigma(\hat{L}) \subset \sigma(A+E)$ and since the spectrum of $A+E$ is real, it is sufficient to prove that

$$\lambda_{m'}(A+E) > \lambda_1(\hat{L}) \text{ and } \lambda_{r_m}(\hat{L}) > \lambda_{m''}(A+E).$$

Without loss of generality assume that $1 \leq m', m'' \leq n$.[1] Since $\lambda_i$ is a continuous function for all $i = 1, ..., n$,

$$\lambda_i(A) = \lambda_i(A+E) + o(1) \quad \text{as} \quad E \to 0$$

and, in addition, since $P = O(E)$ we have

$$\lambda_i((I_{r_m} + P^T P)^{1/2}(E_{11} + E_{12}P)(I_{r_m} + P^T P)^{-1/2}) = o(1) \quad \text{as} \quad E \to 0.$$

Then, by the fact that $L_1 = \lambda_m(A) I_{r_m}$ we have

$$\begin{aligned}
\lambda_1(\hat{L}) &= \lambda_1(L_1 + (I_{r_m} + P^T P)^{1/2}(E_{11} + E_{12}P)(I_{r_m} + P^T P)^{-1/2}) \\
&= \lambda_m(A) + \lambda_1((I_{r_m} + P^T P)^{1/2}(E_{11} + E_{12}P)(I_{r_m} + P^T P)^{-1/2}) \\
&= \lambda_m(A+E) + o(1) \quad \text{as} \quad E \to 0.
\end{aligned}$$

Thus, choosing $E$ close to zero gives

$$\lambda_{m'}(A+E) > \lambda_m(A+E) + o(1) = \lambda_1(\hat{L}).$$

[1] If $m' = 0$ we treat the left inequality as automatically valid and/or if $m'' = n+1$ we treat the right one similarly.

The second inequality follows in an analogous way. □

**Lemma 5.1.3.** *Suppose that the map $t \in [0, \delta) \mapsto E(t) \in S^n$ has a $K$-th order Taylor expansion at $t = 0^+$ (where $\delta > 0$). Then, there is an $\epsilon \in (0, \delta)$ such that for all $t \in [0, \epsilon)$ the equation*

$$P(L_1 + tE_{11}(t)) - (L_2 + tE_{22}(t))P = tE_{21}(t) - tPE_{12}(t)P \tag{5.1}$$

*has a solution $P = P(t)$ where the map $t \in [0, \epsilon) \mapsto P(t)$ has a $K$-th order Taylor expansion at $t = 0^+$ and $P(0) = 0$.*

*In particular, if $E(t)$ is analytic at $t = 0^+$, then $P(t)$ has a Taylor expansion of any order at $t = 0^+$.*

**Proof.** Recall that $L_1 = \lambda_m(A)I_{r_m}$ and, since $\lambda_m(A)$ is not among the eigenvalues of the diagonal matrix $L_2 \in M_{n-r_m}$, we obtain $P(0) = 0$ by Lemma 4.1.1. Moreover, from $P = O(tE(t))$ we have that $P(t) \to 0$ as $t \downarrow 0$, so $P(t)$ is continuous at $t = 0^+$.

Since the difference $\lambda_m(A)I_{n-r_m} - L_2$ is invertible, the matrix

$$B := (\lambda_m(A)I_{n-r_m} - L_2)^{-1}$$

is well-defined and Equation (5.1) can be expressed as

$$\begin{aligned} P(t) &= B\big(tE_{21}(t) - tP(t)E_{11}(t) + tE_{22}(t)P(t) - tP(t)E_{12}(t)P(t)\big) \qquad (5.2) \\ &= tBE_{21}(t) + o(t). \end{aligned}$$

Substituting the expression on the right-hand side of (5.2) recursively into itself we obtain

$$P(t) = tBE_{21}(t) - t^2 BBE_{21}(t)E_{11}(t) + t^2 BE_{22}(t)BE_{21}(t) + o(t^2).$$

Substituting the expression for $P(t)$ recursively twice into the right-hand side of (5.2) results in the formula

$$\begin{aligned} P(t) = {} & tBE_{21}(t) - t^2 BBE_{21}(t)E_{11}(t) + t^2 BE_{22}(t)BE_{21}(t) \\ & + t^3 BBBE_{21}(t)E_{11}(t)E_{11}(t) - t^3 BBE_{22}(t)BE_{21}(t)E_{11}(t) \\ & - t^3 BE_{22}(t)BBE_{21}(t)E_{11}(t) + t^3 BE_{22}(t)BE_{22}(t)BE_{21}(t) \\ & - t^3 BBE_{21}(t)E_{12}(t)BE_{21}(t) + o(t^3). \end{aligned}$$

Analogously, after $K-1$ recursive substitutions of $P(t)$ into the right-hand side of (5.2) we can approximate $P(t)$ by an expression involving $B$ and $E(t)$ with precision $o(t^K)$. Substituting, the $K$-th order Taylor expansion of $E(t)$ at $t = 0^+$ into that approximating formula, we see that $P(t)$ also has a $K$-th order Taylor expansion at $t = 0^+$. $\square$

## 5.2 Additional notation and assumptions

Before deriving expressions for the $K$-th order Taylor expansions of $\lambda_m$ at $A \in \mathcal{S}^n$ along a path, we need additional notation.

- Let $\{n_k\}_{k=0}^{\infty}$ be a non-increasing sequence of integers such that $1 \le n_k \le n$ for all $k = 0, 1, ....$
- Let $\{m_k\}_{k=0}^{\infty}$ be a non-increasing sequence of integers such that $1 \le m_k \le n_k$

for all $k = 0, 1, \ldots$.

- Let $\{A^k\}_{k=0}^{\infty}$ be a sequences of matrices such that $A^k \in \mathcal{S}^{n_k}$.

- Let $\{E^k(t)\}_{k=0}^{\infty}$ be a sequence of matrix-valued maps $t \in [0, \delta_k) \to E^k(t) \in \mathcal{S}^{n_k}$ for all $k = 0, 1, \ldots$, where $\delta_k > 0$.

Our goal is to inductively define the sequences $\{n_k\}_{k=0}^{\infty}$, $\{m_k\}_{k=0}^{\infty}$, $\{A^k\}_{k=0}^{\infty}$, $\{E^k(t)\}_{k=0}^{\infty}$, while making sure that the following assumption is satisfied.

**Assumption 1.** *The map $t \mapsto E^k(t)$ has a limit as $t \downarrow 0$ for all $k = 0, 1, \ldots$.*

All other pieces of notation introduced in this section are derived from the above four sequences. Thus, for all $k = 0, 1, \ldots$:

- Let $m'_k$ and $m''_k$ be the largest and smallest integers respectively, such that

$$\lambda_{m'_k}(A^k) > \lambda_{m_k}(A^k) > \lambda_{m''_k}(A^k).$$

  If the left and/or the right inequality is never satisfied we let $m'_k := 0$ and/or $m''_k := n_k + 1$ respectively.

- Let $r_{m_k}$ be the multiplicity of $\lambda_{m_k}$ in $A^k$, that is, $r_{m_k} = m''_k - m'_k - 1$.

- Let $i_{m_k}$ be defined by

$$i_{m_k} := m_k - m'_k. \tag{5.3}$$

  In other words, $i_{m_k}$ is the number of eigenvalues of $A^k$, ranking before $\lambda_{m_k}$, including $\lambda_{m_k}$, which are equal to $\lambda_{m_k}$.

- Let $X^k = (X_1^k\, X_2^k) \in O^{n_k}$ (where $X_1^k \in M_{n_k, r_{m_k}}$ and $X_2^k \in M_{n_k, n_k - r_{m_k}}$) be an orthogonal matrix such that

$$(X_1^k\, X_2^k)^T A^k (X_1^k\, X_2^k) = \begin{pmatrix} L_1^k & 0 \\ 0 & L_2^k \end{pmatrix},$$

where

$$L_1^k := \lambda_{i_{m_k}}(A^k) I_{r_{m_k}},$$

$$L_2^k := \mathrm{Diag}\,(\lambda_1(A^k), ..., \lambda_{m'_k}(A^k), \lambda_{m''_k}(A^k), ..., \lambda_{n_k}(A^k)).$$

- Let

$$\begin{pmatrix} E_{11}^k(t) & E_{12}^k(t) \\ E_{21}^k(t) & E_{22}^k(t) \end{pmatrix} := (X_1^k\ X_2^k)^T E^k(t) (X_1^k\ X_2^k).$$

Assumption 1, together with Theorem 4.2.4, ensures that there is an $\epsilon_k \in [0, \delta_k)$ such that for all $t \in [0, \epsilon_k)$ the equation

$$P^k(L_1^k + tE_{11}^k(t)) - (L_2^k + tE_{22}^k(t))P^k = tE_{21}^k(t) - tP^k E_{12}^k(t) P^k. \tag{5.4}$$

has a solution $P^k = P^k(t)$. By Theorem 4.2.4, the map

$$t \in [0, \epsilon_k) \mapsto P^k(t) \in M_{n_k - r_{m_k}, r_{m_k}},$$

satisfies

$$P^k(t) = O(t) \text{ for all } k = 0, 1, ....$$

Define

$$V^k(t) := (I_{r_{m_k}} + (P^k(t))^T P^k(t))^{1/2}, \;\; k = 0, 1, ... \text{ and } t \in [0, \epsilon_k).$$

Then, by Theorem 4.2.4, the columns of

$$\hat{X}^k(t) := (X_1^k + X_2^k P^k(t))(V^k(t))^{-1}$$

form a basis of $R(\hat{X}^k(t))$ and the matrix representation of the restriction of $A^k + tE^k(t)$ to $R(\hat{X}^k(t))$ with respect to the columns of $\hat{X}^k(t)$ is

$$\hat{L}^k(t) := V^k(t)(L_1^k + tE_{11}^k(t) + tE_{12}^k(t)P^k(t))(V^k(t))^{-1}.$$

**Lemma 5.2.1.** *For every $k = 0, 1, ...,$ and all $t$ close to $0$ we have*

$$V^k(t) = I_{r_{m_k}} + \sum_{j=1}^{N} \binom{1/2}{j} \big((P^k(t))^T P^k(t)\big)^j + O(t^{2N+2}) \quad \textit{and}$$
$$(V^k(t))^{-1} = I_{r_{m_k}} + \sum_{j=1}^{N} \binom{-1/2}{j} \big((P^k(t))^T P^k(t)\big)^j + O(t^{2N+2}).$$

**Proof.** Let $\mu_1, ..., \mu_{r_{m_k}}$ be the eigenvalues of $(P^k(t))^T P^k(t)$. They are all nonnegative by Lemma 2.4.4. Since $P^k(t) = O(t)$, Lemma 2.4.12 implies that $\mu_i = O(t^2)$. The spectral decomposition

$$I_{r_{m_k}} + (P^k(t))^T P^k(t) = U^T(\mathrm{Diag}\,(1 + \mu_1, ..., 1 + \mu_{r_{m_k}}))U,$$

where $U \in O^{r_{m_k}}$ and the definition of $V^k(t)$, show that

$$V^k(t) = U^T(\mathrm{Diag}\,((1+\mu_1)^{1/2}, ..., (1+\mu_{r_{m_k}})^{1/2}))U.$$

Since, $\sqrt{1+x} = 1 + \sum_{j=1}^N \binom{1/2}{j} x^j + O(x^{N+1})$ for $x$ close to 0, then for $t$ close to 0:

$$\begin{aligned} V^k(t) &= I_{r_{m_k}} + \sum_{j=1}^N \binom{1/2}{j} U^T(\mathrm{Diag}\,(\mu_1^j, ..., \mu_{r_{m_k}}^j))U + O(t^{2N+2}) \\ &= I_{r_{m_k}} + \sum_{j=1}^N \binom{1/2}{j} ((P^k(t))^T P^k(t))^j + O(t^{2N+2}) \end{aligned}$$

as required. The second formula follows in an analogous way. $\square$

**Lemma 5.2.2.** *Suppose that the map $t \mapsto P^k(t)$ has a $K$-th order Taylor expansion at $t = 0^+$. Then, so do the maps $t \mapsto V^k(t)$ and $t \mapsto (V^k(t))^{-1}$.*

**Proof.** Consider the separable function $(1+x_1)^{3/2} + \cdots + (1+x_{n_k})^{3/2}$ defined on the set $\{x \in \mathbb{R}^{n_k} \,|\, x_i \geq -1, i = 1, 2, ..., n_k\}$. Notice that it is analytic in a neighbourhood of the origin. Therefore, by [10, Theorem 2.1], the spectral function $F(X)$ defined by

$$X \in \mathcal{S}^{n_k} \mapsto (1 + \lambda_1(X))^{3/2} + \cdots + (1 + \lambda_{n_k}(X))^{3/2} \in \mathbb{R}$$

is analytic in a neighbourhood of $X = 0$. By [4, Theorem 1.1], we can find the gradient of $F(X)$:

$$\nabla F(X) = \frac{3}{2}(I + X)^{1/2},$$

which is also analytic in a neighbourhood of $X = 0$. Thus, if $t \mapsto P^k(t)$ has a $K$-th order Taylor expansion at $t = 0^+$, so does $t \mapsto \nabla F((P^k(t))^T P^k(t))$.

The second part of the lemma is analogous. □

## 5.3 The proof of the main result

Fix a matrix $A \in \mathcal{S}^n$ and an index $m$ such that $1 \leq m \leq n$. Let $t \in [0, \epsilon) \in \mathbb{R} \to E(t) \in \mathcal{S}^n$ be a map having $K$-th order Taylor expansion at $t = 0^+$. In this section, we show that

$$t \mapsto \lambda_m(A + tE(t))$$

has a $(K+1)$-th order Taylor expansion at $t = 0^+$. We proceed by induction and start by setting

$$n_0 := n, \;\; m_0 := m, \;\; A^0 := A, \;\; E^0(t) := E(t).$$

Clearly, the map $t \mapsto E^0(t)$ has a $K$-th order Taylor expansion at $t = 0^+$. Let $\epsilon_0 := \epsilon$.

Let $k$ be such that $1 \leq k \leq K$ and suppose that the quantities $n_{k-1}$, $m_{k-1}$, $A^{k-1}$ and the map $t \in [0, \epsilon_{k-1}) \mapsto E^{k-1}(t) \in \mathcal{S}^{n_{k-1}}$ have been defined and that the map $t \mapsto E^{k-1}(t)$ has $(K - k + 1)$-th order Taylor expansion at $t = 0^+$. We define

$$\begin{aligned} n_k &:= r_{m_{k-1}}, \\ m_k &:= i_{m_{k-1}}, \\ A^k &:= \lim_{t \downarrow 0} E_{11}^{k-1}(t). \end{aligned} \tag{5.5}$$

In addition, for $t \in [0, \epsilon_{k-1})$ we define

$$E^k(t) := \frac{1}{t}(V^{k-1}(t)(E_{11}^{k-1}(t) + E_{12}^{k-1}(t)P^{k-1}(t))(V^{k-1}(t))^{-1} - A^k), \text{ if } t \neq 0 \quad (5.6)$$

and

$$E^k(0) := \lim_{t\downarrow 0} \frac{1}{t}(V^{k-1}(t)(E_{11}^{k-1}(t) + E_{12}^{k-1}(t)P^{k-1}(t))(V^{k-1}(t))^{-1} - A^k).$$

The limit in the definition of $A^k$ is well-defined, by the induction hypothesis. In addition, we have $A^k \in \mathcal{S}^{n_k}$ and it is easy to see that

$$A^k = \lim_{t\downarrow 0} V^{k-1}(t)(E_{11}^{k-1}(t) + E_{12}^{k-1}(t)P^{k-1}(t))(V^{k-1}(t))^{-1}.$$

**Proposition 5.3.1.** *If the map $t \mapsto E^{k-1}(t)$ has $(K-k+1)$-th order Taylor expansion at $t = 0^+$, then the map $t \mapsto E^k(t)$ has $(K-k)$-th order Taylor expansion at $t = 0^+$.*

**Proof.** Since the map $t \mapsto E^{k-1}(t)$ has $(K-k+1)$-th order Taylor expansion at $t = 0^+$, by Lemma 5.1.3 and Lemma 5.2.2 the maps $t \mapsto P^{k-1}(t)$, $t \mapsto V^{k-1}(t)$ and $t \mapsto (V^{k-1}(t))^{-1}$ have $(K-k+1)$-th order Taylor expansion at $t = 0^+$ with constant terms 0, $I_{n_{k-1}}$ and $I_{n_{k-1}}$ respectively. On the other hand, the constant term of the $(K-k+1)$-th order Taylor expansion of $t \mapsto E_{11}^{k-1}(t)$ is, by definition, $A^k$. Thus, $t \mapsto E^k(t)$ has $(K-k)$-th order Taylor expansion and $E^k(0)$ is well-defined. $\square$

By Theorem 4.2.4, the matrix

$$V^{k-1}(t)(E_{11}^{k-1}(t) + E_{12}^{k-1}(t)P^{k-1}(t))(V^{k-1}(t))^{-1}$$

is symmetric (as the difference of two symmetric matrices) and immediately from the definitions we have that

$$A^k + tE^k(t) = V^{k-1}(t)(E_{11}^{k-1}(t) + E_{12}^{k-1}(t)P^{k-1}(t))(V^{k-1}(t))^{-1}. \tag{5.7}$$

**Theorem 5.3.2.** *Suppose that $A \in \mathcal{S}^n$ and $t \in [0, \epsilon) \in \mathbb{R} \mapsto E(t) \in \mathcal{S}^n$ is a map having $K$-th order Taylor expansion at $t = 0^+$. Then, for any $m \in \{1, ..., n\}$ the function*

$$t \in [0, \epsilon) \mapsto \lambda_m(A + tE(t)) \in \mathbb{R}$$

*has $(K+1)$-th order Taylor expansion at $t = 0^+$:*

$$\lambda_m(A + tE(t)) = \lambda_m(A) + \Big(\sum_{k=1}^{K+1} \lambda_{m_k}(A^k)t^k\Big) + o(t^{K+1}),$$

*where the symmetric matrices $A^k$ are defined by (5.5) and the indexes $i_{m_k}$ by (5.3).*

**Proof.** Let $n_0 := n$, $m_0 := m$, $A^0 := A$, $E^0(t) := E(t)$ and $\epsilon_0 := \epsilon$. By Lemma 5.1.2, there is an $\epsilon_1 > 0$ such that for all $t \in [0, \epsilon_1)$

$$\begin{aligned}
\lambda_m(A + tE(t)) &= \lambda_{i_m}(\hat{L}) \\
&= \lambda_m(A) + \lambda_{i_m}(V^0(t)(E_{11}^0(t) + E_{12}^0(t)P^0(t))(V^0(t))^{-1})t \\
&= \lambda_m(A) + \lambda_{m_1}(A^1 + tE^1(t))t,
\end{aligned}$$

where we used the definition $m_1 = i_{m_0} = i_m$. By Proposition 5.3.1, $t \mapsto E^1(t)$ has a $(K-1)$-th order Taylor expansion at $t = 0^+$ and the process can be repeated. Thus, there is an $\epsilon_2 > 0$ such that for all $t \in [0, \epsilon_2)$

$$\lambda_m(A + tE(t)) = \lambda_m(A) + \lambda_{m_1}(A^1)t + \lambda_{m_2}(A^2 + tE^2(t))t^2.$$

After, $K$ repetitions we obtain, for all $t \in [0, \epsilon_K)$

$$\lambda_m(A + tE(t)) = \lambda_m(A) + \Big(\sum_{k=1}^{K-1} \lambda_{m_k}(A^k)t^k\Big) + \lambda_{m_K}(A^K + tE^K(t))t^K,$$

where the map $t \mapsto E^K(t)$ has a limit as $t \to 0$. This allows us to repeat the process two more times! The first is

$$\begin{aligned}\lambda_{m_K}(A^K + tE^K(t)) &= \lambda_{i_{m_K}}(\hat{L}^K)\\ &= \lambda_{m_K}(A^K) + \lambda_{m_{K+1}}(V^K(t)(E_{11}^K(t) + E_{12}^K(t)P^K(t))(V^K(t))^{-1})t.\end{aligned}$$

Since the map $t \mapsto E^K(t)$ has a limit as $t \to 0$ so do the maps $t \mapsto E_{11}^K(t)$ and $t \mapsto E_{12}^K(t)$. Recalling that $P^K(t) = O(tE^K(t))$ as $t \to 0$ we can define

$$A^{K+1} := \lim_{t \downarrow 0}(V^K(t)(E_{11}^K(t) + E_{12}^K(t)P^K(t))(V^K(t))^{-1}) = \lim_{t \downarrow 0} E_{11}^K(t),$$

and

$$E^{K+1}(t) := (V^K(t)(E_{11}^K(t) + E_{12}^K(t)P^K(t))(V^K(t))^{-1}) - A^{K+1}.$$

Thus, $E^{K+1}(t)$ approaches 0 as $t \downarrow 0$ and there is an $\epsilon_{K+1} > 0$ such that for all

$t \in [0, \epsilon_{K+1})$ the solution, $P^{K+1} = P^{K+1}(t)$, of

$$P^{K+1}(L_1^{K+1} + E_{11}^{K+1}(t)) - (L_2^{K+1} + E_{22}^{K+1}(t))P^{K+1} = E_{21}^{K+1}(t) - P^{K+1}E_{12}^{K+1}(t)P^{K+1}$$

exists and $P^{K+1}(t) = O(E^{K+1}(t))$ as $E^{K+1}(t) \to 0$. Therefore,

$$\begin{aligned}
\lambda_{m_{K+1}}(V^K(t)(E_{11}^K(t) + E_{12}^K(t)P^K(t))(V^K(t))^{-1}) &= \lambda_{m_{K+1}}(A^{K+1} + E^{K+1}(t)) \\
&= \lambda_{m_{K+1}}(A^{K+1}) + \lambda_{m_{K+2}}(V^{K+1}(t)(E_{11}^{K+1}(t) + E_{12}^{K+1}(t)P^{K+1}(t))(V^{K+1}(t))^{-1}) \\
&= \lambda_{m_{K+1}}(A^{K+1}) + o(1),
\end{aligned}$$

where we used the continuity of the eigenvalues and the fact that the expression in the argument of $\lambda_{m_{K+2}}$ converges to zero. Putting everything together concludes the proof. □

**Corollary 5.3.3.** *Suppose that $A \in \mathcal{S}^n$ and that the map $t \in [0, \epsilon) \in \mathbb{R} \mapsto E(t) \in \mathcal{S}^n$ is analytic at $t = 0^+$. Then, for any $m \in \{1, ..., n\}$, the function*

$$t \in [0, \epsilon) \mapsto \lambda_m(A + tE(t)) \in \mathbb{R}$$

*has a Taylor expansion at $t = 0^+$ of any order.*

## 5.4 The higher-order directional derivatives of $\lambda_m(A)$

Of particular interest is the case when $E(t) \equiv E$. By Corollary 5.3.3, for any fixed $m \in \{1, ..., n\}$ the function $t \mapsto \lambda_m(A + tE)$ has a Taylor expansion at $t = 0^+$ of any order. Denote its $k$-th term by $\lambda_m^{(k)}(A; E)$, $k = 0, 1, ....$ We call it the *$k$-th order directional derivative of $\lambda_m(A)$ in the direction $E$.* For $k = 1, 2, 3$, we interchangeably use the notation $\lambda'_m(A;E)$, $\lambda''_m(A;E)$ and $\lambda'''_m(A;E)$ respectively. Clearly, $\lambda_m^{(0)}(A;E) = \lambda_m(A)$.

In this section, we compute the first four directional derivatives of $\lambda_m(A)$ in the direction $E$. Clearly, $\lambda'_m(A;E)$ is the usual directional derivative

$$\lambda'_m(A;E) = \lim_{t \downarrow 0} \frac{\lambda_m(A + tE) - \lambda_m(A)}{t}$$

which has been computed several times in the literature, see [1, Theorem 3.12] and [9, Theorem 1.5, Remark 3]. In general we have

$$\lambda_m^{(k)}(A;E) = \lim_{t \downarrow 0} \frac{\lambda_m(A + tE) - \sum_{l=0}^{k-1} \lambda_m^{(l)}(A;E) t^l}{t^k}$$

with Corollary 5.3.3 guaranteeing that the limit exists for all $k = 0, 1, ....$

**Proposition 5.4.1.** *With the notation from Theorem 4.2.4, the directional derivative of $\lambda_m$ at $A \in \mathcal{S}^n$ in a direction $E \in \mathcal{S}^n$ is*

$$\lambda'_m(A;E) = \lambda_{m_1}(X_1^T E X_1).$$

**Proof.** By Theorem 5.3.2 we have $\lambda'_m(A;E) = \lambda_{m_1}(A^1)$, where, by (5.5), $A^1 = \lim_{t \downarrow 0} E_{11}(t) = E_{11} = X_1^T E X_1$, using the definition $E^0(t) = E$. □

In order to present a succinct formula for the next several directional derivatives, we need a bit more notation. The matrices

$$B^k := (\lambda_{m_k}(A) I_{n_k - r_{m_k}} - L_2^k)^{-1}, \quad \text{for } k = 0, 1, ...$$

play important part in expanding the map $t \mapsto P^k(t)$ in a Taylor series at $t = 0^+$. Define also

$$\tilde{A}^k := (\lambda_{m_k}(A^k)I_{n_k} - A^k)^\dagger, \quad \text{for } k = 0, 1, ....$$

Furthermore, denote

$$U^k := X_1^0 X_1^1 ... X_1^k \quad \text{for } k = 0, 1, ....$$

We continue to write $B$ for $B^0$, $\tilde{A}$ for $\tilde{A}^0$ and $U$ for $U^0$.

**Lemma 5.4.2.** *We have $X_2^k B^k (X_2^k)^T = \tilde{A}^k$ for all $k = 0, 1, ....$*

**Proof.** We have

$$\begin{aligned}
(\lambda_{m_k}(A^k)I_{n_k} - A^k)^\dagger &= (X_1^k \, X_2^k) \begin{pmatrix} 0 & 0 \\ 0 & (\lambda_{m_k}(A)I_{n_k - r_{m_k}} - L_2^k)^{-1} \end{pmatrix} \begin{pmatrix} (X_1^k)^T \\ (X_2^k)^T \end{pmatrix} \\
&= (X_1^k \, X_2^k) \begin{pmatrix} 0 & 0 \\ 0 & B^k \end{pmatrix} \begin{pmatrix} (X_1^k)^T \\ (X_2^k)^T \end{pmatrix} = X_2^k B^k (X_2^k)^T
\end{aligned}$$

as required. □

**Proposition 5.4.3.** *With the notation from Theorem 4.2.4, the second order directional derivative of $\lambda_m$ at $A \in \mathcal{S}^n$ in the direction $E \in \mathcal{S}^n$ is*

$$\lambda''_m(A; E) = \lambda_{m_2}\big((U^1)^T E \tilde{A} E (U^1)\big).$$

**Proof.** By Theorem 5.3.2, $\lambda''_m(A; E) = \lambda_{m_2}(A^2)$, where, by (5.5) $A^2 = \lim_{t \downarrow 0} E_{11}^1(t)$.

By (5.6) we have

$$E^1(t) = \frac{1}{t}\big(V(t)(E_{11} + E_{12}P(t))(V(t))^{-1} - A^1\big).$$

Substituting $P(t) = tBE_{21} + o(t)$, by Lemma 5.1.3, and $V(t) = I_{r_m} + O(t^2)$, by Lemma 5.2.1, gives

$$E^1(t) = \frac{1}{t}(E_{11} + E_{12}(tBE_{21} + o(t)) - A^1 + O(t^2)) = E_{12}BE_{21} + o(1)$$

since $A^1 = E_{11}$ by Proposition 5.4.1. Therefore

$$A^2 = (U^1)^T EX_2BX_2^T EU^1 = (U^1)^T E\tilde{A}EU^1$$

using Lemma 5.4.2. □

In order to better capture the increasing complexity of the expressions, we introduce the *Jordan product* between two square matrices of the same size:

$$X \bullet Y := (XY^T + YX^T)/2.$$

Notice that $X \bullet Y$ is always a symmetric matrix.

**Proposition 5.4.4.** *With the notation from Theorem 4.2.4, the third order directional derivative of $\lambda_m$ at $A \in \mathcal{S}^n$ in the direction $E \in \mathcal{S}^n$ is*

$$\lambda_m'''(A; E) = \lambda_{m_3}(A^3)$$

*where*

$$A^3 = (U^2)^T\big(E\tilde{A}EX_1\tilde{A}^1X_1^TE\tilde{A}E + E\tilde{A}E\tilde{A}E - (E\tilde{A}\tilde{A}E)\bullet(EX_1X_1^T)\big)U^2.$$

**Proof.** By Theorem 5.3.2, $\lambda_m'''(A;E) = \lambda_{m_3}(A^3)$, where, by (5.5) $A^3 = \lim_{t\downarrow 0} E_{11}^2(t)$. By (5.6) we have

$$E^2(t) = \frac{1}{t}(V^1(t)(E_{11}^1(t) + E_{12}^1(t)P^1(t))(V^1(t))^{-1} - A^2).$$

In order to find $A^3$ we first calculate a first order expansion of $E^1(t)$. By Lemma 5.2.1, we find

$$V(t) = I_{r_m} + (P(t))^TP(t)/2 + O(t^4) \quad \text{and} \quad (V(t))^{-1} = I_{r_m} - (P(t))^TP(t)/2 + O(t^4)$$

and using the fact that (see the proof of Lemma 5.1.3 with $E(t) = E$)

$$P(t) = tBE_{21} - t^2BBE_{21}E_{11} + t^2BE_{22}BE_{21} + o(t^2),$$

it follows that

$$\begin{aligned} E^1(t) &= \frac{1}{t}(V(t)(E_{11} + E_{12}P(t))(V(t))^{-1} - A^1) \\ &= \frac{1}{t}\Big(E_{12}P(t) + \frac{P(t)^TP(t)}{2}E_{11} - E_{11}\frac{P(t)^TP(t)}{2} + o(t^2)\Big) \\ &= E_{12}BE_{21} + t\big(E_{12}BE_{22}BE_{21} - (E_{12}BBE_{21})\bullet E_{11}\big) + o(t). \end{aligned}$$

We now continue with the derivation of $A^3$ by observing that

$$\lim_{t\downarrow 0} E^2(t) = \lim_{t\downarrow 0} \frac{V^1(t)E^1_{11}(t)(V^1(t))^{-1} - A^2}{t} + \lim_{t\downarrow 0} \frac{V^1(t)E^1_{12}(t)P^1(t)(V^1(t))^{-1}}{t}. \quad (5.8)$$

Consider the two limits on the right-hand side separately. To evaluate the first limit, use the fact that $V^1(t) = I_{r_{m_1}} + O(t^2) = (V^1(t))^{-1}$ and $P^1(t) = tB^1E^1_{21}(t) + o(t)$ by Lemma 5.1.3 and Lemma 5.2.1, substituting the first order expansion of $E^1(t)$:

$$\begin{aligned}
\lim_{t\downarrow 0}\frac{V^1(t)E^1_{11}(t)(V^1(t))^{-1} - A^2}{t} &= \lim_{t\downarrow 0} \frac{E^1_{11}(t) - A^2}{t} \\
&= (X^1_1)^T\big(E_{12}BE_{22}BE_{21} - (E_{12}BBE_{21}) \bullet E_{11}\big)X^1_1 \\
&= (U^1)^T\big(E\tilde{A}E\tilde{A}E - (E\tilde{A}\tilde{A}E) \bullet (EX_1X_1^T)\big)(U^1).
\end{aligned}$$

Analogously, the second limit in the right-hand side of (5.8) is

$$\begin{aligned}
&\lim_{t\downarrow 0}\frac{V^1(t)E^1_{12}(t)P^1(t)(V^1(t))^{-1}}{t} = \lim_{t\downarrow 0} \frac{E^1_{12}(t)P^1(t)}{t} = \lim_{t\downarrow 0} E^1_{12}(t)B^1E^1_{21}(t) \\
&= \lim_{t\downarrow 0}(X^1_1)^T E^1(t)\tilde{A}^1E^1(t)X^1_1 = (U^1)^T\big(E\tilde{A}EX_1\tilde{A}^1X_1^TE\tilde{A}E\big)(U^1),
\end{aligned}$$

using Lemma 5.4.2. Putting everything together, the proposition follows. $\square$

**Proposition 5.4.5.** *With the notation from Theorem 4.2.4, the fourth order directional derivative of $\lambda_m$ at $A \in \mathcal{S}^n$ in the direction $E \in \mathcal{S}^n$ is*

$$\lambda^{(4)}_m(A; E) = \lambda_{m_4}(A^4)$$

*where*

$$A^4 := (U^3)^T \begin{pmatrix} E\tilde{A}EX_1\tilde{A}^1X_1^TE\tilde{A}EX_1\tilde{A}^1X_1^TE\tilde{A}E \\ -(E\tilde{A}EX_1\tilde{A}^1\tilde{A}^1X_1^TE\tilde{A}E) \bullet (E\tilde{A}EU^1(U^1)^T) \\ +2(E\tilde{A}EX_1\tilde{A}^1X_1^T) \bullet (E\tilde{A}E\tilde{A}E - (E\tilde{A}\tilde{A}E) \bullet (EX_1X_1^T)) \\ +(E\tilde{A}\tilde{A}\tilde{A}EX_1X_1^T - E\tilde{A}\tilde{A}E\tilde{A}) \bullet (EX_1X_1^TE) \\ -(E\tilde{A}\tilde{A}EX_1X_1^T + EX_1X_1^TE\tilde{A}\tilde{A}) \bullet (E\tilde{A}E) \\ +E\tilde{A}E\tilde{A}E\tilde{A}E \end{pmatrix} U^3$$

$$+(U^3)^T \begin{pmatrix} E\tilde{A}E\tilde{A}E \\ -(E\tilde{A}\tilde{A}E) \bullet (EX_1X_1^T) \\ +E\tilde{A}EX_1\tilde{A}^1X_1^TE\tilde{A}E \end{pmatrix} U^1\tilde{A}^2(U^1)^T \begin{pmatrix} E\tilde{A}E\tilde{A}E \\ -(E\tilde{A}\tilde{A}E) \bullet (EX_1X_1^T) \\ +E\tilde{A}EX_1\tilde{A}^1X_1^TE\tilde{A}E \end{pmatrix} U^3.$$

**Proof.** By Theorem 5.3.2 we have $\lambda_m^{(4)}(A;E) = \lambda_{m_4}(A^4)$ where, by (5.5),

$$A^4 = \lim_{t \downarrow 0} E_{11}^3(t).$$

By (5.6) we have

$$E^3(t) = \frac{1}{t}(V^2(t)(E_{11}^2(t) + E_{12}^2(t)P^2(t))(V^2(t))^{-1} - A^3).$$

In order to calculate the limit

$$\lim_{t \downarrow 0} E^3(t),$$

we must calculate a first order expansion of $E^2(t)$ which itself requires a second order expansion of $E^1(t)$. Since $V^k(t)$ and $(V^k(t))^{-1}$ have the second order expansions

$$V^k(t) = I_{r_{m_k}} + (P^k(t))^T P^k(t)/2 + O(t^4) \tag{5.9}$$

$$(V^k(t))^{-1} = I_{r_{m_k}} - (P^k(t))^T P^k(t)/2 + O(t^4) \tag{5.10}$$

for $k = 0, 1$ it follows that

$$\begin{aligned} E^1(t) &= \frac{1}{t}(V^0(t)(E_{11} + E_{12}P(t))(V^0(t))^{-1} - A^1) \\ &= \Big(I_{r_m} + \frac{P(t)^T P(t)}{2}\Big)\frac{E_{12}P(t)}{t}\Big(I_{r_m} - \frac{P(t)^T P(t)}{2}\Big) + \frac{P(t)^T P(t)}{2t}E_{11} \\ &\quad - E_{11}\frac{P(t)^T P(t)}{2t} + O(t^3). \end{aligned}$$

Observe that

$$\begin{aligned} \frac{P(t)^T P(t)}{2}E_{11} &= t^2 \begin{pmatrix} \frac{1}{2}E_{12}BBE_{21} - tE_{12}BBBE_{21} \bullet E_{11} \\ +t(E_{12}BBE_{22}B) \bullet E_{12} \end{pmatrix} E_{11} + o(t^3) \\ &= X_1^T \begin{pmatrix} \frac{1}{2}E\tilde{A}\tilde{A}E + t(E\tilde{A}\tilde{A}E) \bullet (E\tilde{A}) \\ -t(E\tilde{A}\tilde{A}\tilde{A}E) \bullet (EX_1X_1^T) \end{pmatrix} X_1X_1^T EX_1 + o(t^3) \end{aligned} \tag{5.11}$$

and

$$E_{11}\frac{P(t)^T P(t)}{2} = X_1^T EX_1X_1^T \begin{pmatrix} \frac{1}{2}E\tilde{A}\tilde{A}E + t(E\tilde{A}\tilde{A}E) \bullet (E\tilde{A}) \\ -t(E\tilde{A}\tilde{A}\tilde{A}E) \bullet (EX_1X_1^T) \end{pmatrix} X_1 + o(t^3) \tag{5.12}$$

using the second order expansion of $P(t)$ given in Lemma 5.1.3. Moreover, using the third order expansion of $P(t)$ calculated in Lemma 5.1.3 we get

$$E_{12}P(t) = tX_1^t \begin{pmatrix} E\tilde{A}E - tE\tilde{A}\tilde{A}EX_1X_1^TE + tE\tilde{A}E\tilde{A}E \\ +t^2E\tilde{A}\tilde{A}\tilde{A}EX_1X_1^TEX_1X_1^TE \\ -t^2(E\tilde{A}\tilde{A}E\tilde{A} + E\tilde{A}E\tilde{A}\tilde{A})EX_1X_1^TE \\ -t^2(E\tilde{A}\tilde{A}EX_1X_1^TE\tilde{A}E - E\tilde{A}E\tilde{A}E\tilde{A}E) \end{pmatrix} X_1 + o(t^3), \quad (5.13)$$

and substituting the first order expansion of $P(t)$ from Lemma 5.1.3 gives

$$\begin{aligned} -E_{12}P(t)\Big(\frac{P(t)^TP(t)}{2}\Big) &= -\frac{1}{2}E_{12}(tBE_{21} + o(t))(t^2E_{12}BBE_{21} + o(t^2)) \\ &= -\frac{t^3}{2}X_1^TE\tilde{A}EX_1X_1^TE\tilde{A}\tilde{A}EX_1 + o(t^3) \end{aligned} \quad (5.14)$$

and

$$\Big(\frac{P(t)^TP(t)}{2}\Big)E_{12}P(t) = \frac{t^3}{2}X_1^TE\tilde{A}\tilde{A}EX_1X_1^TE\tilde{A}EX_1 + o(t^3). \quad (5.15)$$

Therefore, combining (5.11), (5.12), (5.13), (5.14) and (5.15) gives

$$E^1(t) = X_1^T \begin{pmatrix} E\tilde{A}E + t(E\tilde{A}E\tilde{A}E + (E\tilde{A}\tilde{A}E) \bullet (EX_1X_1^T)) \\ +t^2(E\tilde{A}\tilde{A}\tilde{A}EX_1X_1^T - E\tilde{A}\tilde{A}E\tilde{A}) \bullet (EX_1X_1^TE) \\ -t^2(E\tilde{A}\tilde{A}EX_1X_1^T + EX_1X_1^TE\tilde{A}\tilde{A}) \bullet (E\tilde{A}E) \\ +t^2E\tilde{A}E\tilde{A}E\tilde{A}E \end{pmatrix} X_1 + o(t^2). \quad (5.16)$$

We next compute the first order expansion of $E^2(t)$ using the second order expansion of $E^1(t)$ given by (5.16). Observe that

$$\begin{aligned} E^2(t) &= \frac{1}{t}(V^1(t)(E^1_{11}(t) + E^1_{12}(t)P^1(t))(V^1(t))^{-1} - A^1) \\ &= \frac{1}{t}\left(\begin{array}{c} E^1_{11}(t) + E^1_{12}(t)P^1(t) - A^1 \\ +\frac{(P^1(t))^T P^1(t)}{2}E^1_{11}(t) - E^1_{11}(t)\frac{(P^1(t))^T P^1(t)}{2} \end{array}\right) + o(t) \end{aligned}$$

using (5.9) and (5.10). Then, substituting (5.16) and using the second order expansion of $P^1(t)$ gives

$$E^2 = (U^1)^T \left(\begin{array}{c} E\tilde{A}E\tilde{A}E - (E\tilde{A}\tilde{A}E) \bullet (EX_1X_1^T) \\ +E^1(t)\tilde{A}^1E^1(t) + tX_1E^1(t)\tilde{A}^1E^1(t)\tilde{A}^1E^1(t)X_1^T \\ +t(E\tilde{A}\tilde{A}\tilde{A}EX_1X_1^T - E\tilde{A}\tilde{A}E\tilde{A}) \bullet (EX_1X_1^TE) \\ -t(E\tilde{A}\tilde{A}EX_1X_1^T + EX_1X_1^TE\tilde{A}\tilde{A}) \bullet (E\tilde{A}E) \\ +tE\tilde{A}E\tilde{A}E\tilde{A}E \\ -tX_1(E^1(t)\tilde{A}^1\tilde{A}^1E^1(t)) \bullet (E^1(t)X_1^1(X_1^1)^T)X_1^T \end{array}\right) U^1 + o(t). \quad (5.17)$$

It remains to substitute this representation of $E^2(t)$ into the limit

$$\begin{aligned} \lim_{t\downarrow 0} E^3(t) &= \lim_{t\downarrow 0} \frac{V^2(t)E^2_{11}(t)(V^2)^{-1} - A^3}{t} + \lim_{t\downarrow 0} V^2(t)E^2_{12}(t)\frac{P^2(t)}{t}(V^2(t))^{-1} \\ &= \lim_{t\downarrow 0} \frac{E^2_{11}(t) - A^3}{t} + \lim_{t\downarrow 0} E^2_{12}(t)\frac{P^2(t)}{t} \end{aligned} \quad (5.18)$$

since $V^2(t) = I_{r_{m_2}} + O(t^2)$ by Lemma 5.2.1. Since $P^2(t) = tB^2E^2_{21}(t) + o(t)$ by Lemma 5.1.3, substituting (5.17) into the first limit on the right-hand side of (5.18) gives

$$
\lim_{t\downarrow 0}\frac{E_{11}^2(t)-A^3}{t}
$$

$$
=(U^2)^T\begin{pmatrix}
E\tilde{A}EX_1\tilde{A}^1X_1^TE\tilde{A}EX_1\tilde{A}^1X_1^TE\tilde{A}E\\
-(E\tilde{A}EX_1\tilde{A}^1\tilde{A}^1X_1^TE\tilde{A}E)\bullet(E\tilde{A}EU^1(U^1)^T)\\
+2(E\tilde{A}EX_1\tilde{A}^1X_1^T)\bullet(E\tilde{A}E\tilde{A}E-(E\tilde{A}\tilde{A}E)\bullet(EX_1X_1^T))\\
+(E\tilde{A}\tilde{A}\tilde{A}EX_1X_1^T-E\tilde{A}\tilde{A}E\tilde{A})\bullet(EX_1X_1^TE)\\
-(E\tilde{A}\tilde{A}EX_1X_1^T+EX_1X_1^TE\tilde{A}\tilde{A})\bullet(E\tilde{A}E)\\
+E\tilde{A}E\tilde{A}E\tilde{A}E
\end{pmatrix}U^2.
$$

and substituting (5.17) in the second limit on the right-hand side of (5.18) gives

$$
\lim_{t\downarrow 0}E_{12}^2(t)\frac{P^2(t)}{t}=\lim_{t\downarrow 0}E_{12}^2(t)B^2E_{21}^2(t)
$$

$$
=(U^2)^T\begin{pmatrix}
E\tilde{A}E\tilde{A}E\\
-(E\tilde{A}\tilde{A}E)\bullet(EX_1X_1^T)\\
+E\tilde{A}EX_1\tilde{A}^1X_1^TE\tilde{A}E
\end{pmatrix}U^1\tilde{A}^2(U^1)^T\begin{pmatrix}
E\tilde{A}E\tilde{A}E\\
-(E\tilde{A}\tilde{A}E)\bullet(EX_1X_1^T)\\
+E\tilde{A}EX_1\tilde{A}^1X_1^TE\tilde{A}E
\end{pmatrix}U^2.
$$

Putting everything together gives the desired form of $A^4$. □

# Bibliography

[1] J.B. HIRIART-URRUTY and D. YE. Sensitivity analysis of all eigenvalues of a symmetric matrix. *Numerische Mathematik*, 70:45–72, 1995.

[2] T. KATO. *A Short Introduction to Perturbation Theory for Linear Operators*. Springer-Verlag, Berlin, 1976.

[3] P. LANCASTER. On eigenvalues of matrices dependent on a parameter. *Numerische Mathematik*, 6:377–387, 1964.

[4] A.S. LEWIS. Derivatives of spectral functions. *Mathematics of Operations Research*, 21:576–588, 1996.

[5] C. MEYER. *Matrix Analysis and Applied Linear Algebra*. The Soceity for Industriral and Applied Mathematics, PA, 2000.

[6] F. RELLICH. *Perturbation Theory of Eigenvalue Problems*. Lecture Notes. New York University, 1953.

[7] E.B. SAFF and A.D. SNIDER. *Fundamentals of Complex Analysis with Applications to Engineering, Science, and Mathematics (3rd Edition)*. Pearson Education, Inc., 2003.

[8] G.W. STEWART and JI-GUANG SUN. *Matrix Perturbation Theory*. Computer Science and Scientific Computing. Academic Press Inc., Boston, MA, 1990.

[9] M. TORKI. Second-order directional derivatives of all eigenvalues of a symmetric matrix. *Nonlinear Analysis, Ser. A: Theory Methods*, 46(8):1133–1150, 2001.

[10] N. TSING, M.K.H. FAN, and E.I. VERRIEST. On analyticity of functions involving eigenvalues. *Linear Algebra and its Applications*, 207:159–180, 1994.