Total derivatives of eigenvalues and
eigenprojections of symmetric matrices

Karl K. Brustad
Aalto University
May 16, 2019

Abstract

Conditions for existence and formulas for the first- and second
order total derivatives of the eigenvalues, and the first order total
derivatives of the eigenprojections of smooth matrix-valued functions
$H : \Omega \to S(m)$ are given. The eigenvalues and eigenprojections are
considered as functions in the same domain $\Omega \subseteq \mathbb{R}^n$.

1 Introduction

Let $\Omega \subseteq \mathbb{R}^n$ be a domain and assume that $H : \Omega \to S(m)$ is a continuously
differentiable function taking values in the space $S(m)$ of symmetric $m \times m$
matrices. Under what conditions are the eigenvalues and the eigenprojections
of $H$ differentiable, and what are their total derivatives? The eigenprojection
$P_j(x)$, corresponding to the eigenvalue $\lambda_j(x)$ of $H$ at $x \in \Omega$, is the unique
symmetric $m \times m$ projection matrix, i.e. $P_j^T(x) = P_j(x) = P_j^2(x)$, satisfying

$$H(x)P_j(x) = \lambda_j(x)P_j(x)$$

with rank, or dimension, equal to the multiplicity of the eigenvalue.

The Hessian matrix $\mathcal{H}u$ of a function $u \in C^3(\Omega)$ is a motivating special
case. Then $m = n$ and, as a standard example – showing that smooth
matrices need not have differentiable eigenvalues – one may consider the
real part of the analytic function $z^3$ in the plane $\mathbb{C}$: If we set $u(x, y) := \frac{1}{6}(x^3 - 3xy^2)$, then

$$\mathcal{H}u(x, y) = \begin{pmatrix} x & -y \\ -y & -x \end{pmatrix}$$
with eigenvalues
\[ \lambda_1(x, y) = -\sqrt{x^2 + y^2} \quad \text{and} \quad \lambda_2(x, y) = \sqrt{x^2 + y^2}. \]

We see that the problem occurs at the origin where the eigenvalues "cross". This is a well-known phenomenon. The corresponding eigenprojections are not even continuous since \( \text{tr} \, P_1 = \text{tr} \, P_2 = 1 \) away from the origin while \( P_1(0,0) = P_2(0,0) = I \) as \( \mathcal{H} u(0,0) = 0 = 0 \cdot I \).

Perturbations of eigenvalues and eigenvectors of symmetric matrix-valued functions have been studied in various settings. It is shown in [Tor01] that the \( j \)’th eigenvalue of \( H(t) = H_0 + tH_1 + \frac{1}{2} t^2 H_2 \) always has first- and second order one-sided derivatives. This work is partly based on [HUX95] and is developed further in [ZZX13]. Our formulas for the derivatives of \( \lambda_j \) have counterparts in these papers, although the setting is not exactly the same. The total projection for the \( \lambda \)-group – i.e. the sum of projections corresponding to neighbouring eigenvalues – is analyzed in [Kat95]. The expression (3.8) for the derivative of \( P_j \) may be compared with the one found in (Theorem 5.4 [Kat95]). The book by Kato is a standard reference for perturbing matrices depending on a single real or complex parameter. Unfortunately, many of the results therein do not generalize if the matrix depends on several variables. We also mention the papers [LS01] and [ACL93] where, respectively, spectral functions and solutions to nonlinear eigenvalue-eigenvector problems are differentiated.

We shall consider the eigenvalues and eigenprojections as functions in \( \Omega \subseteq \mathbb{R}^n \). The total derivative of an eigenvalue \( \lambda_j \) is, if it exists, the gradient \( \nabla \lambda_j \). For the matrix-valued eigenprojections \( P_j \), the total derivative is a mapping \( \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n \times \Omega \to \mathbb{R} \) linear in the three first arguments. That is, a third order tensor. In order to simplify the notation and minimize the use of indexes, we introduce two different first order matrix valued tensors representing the derivative of matrix functions. The (double-sided) directional derivatives are also studied. In contrast to the one-sided limits, they do not always exist as our example clearly shows.

Our main results Theorem 3.1, Theorem 3.2, and Theorem 4.7 give explicit expressions for the derivatives of \( \lambda_j \) and \( P_j \) in terms of \( H \) and its derivatives. As a little surprise, it turns out that an eigenprojection is continuous only if it is differentiable, and it has constant dimension only if it is directionally differentiable. Moreover, if \( H \) is \( C^2 \) and the eigenprojection is continuous, then the corresponding eigenvalue has a differentiable gradient. We have not been able to find these observations in the literature. A key ingredient in proving the theorems is Lemma 4.1 (Lemma 5.2 [Bel13]). It does not seem to have been used in the other aforementioned papers.
2 Preliminaries

The matrix norm used throughout the paper is \( \|X\| := \sqrt{\text{tr}(X^T X)} \). Even though we treat \( \mathbb{R}^m \) as \( \mathbb{R}^{m \times 1} \) algebraically, the vector norm is denoted by \( |y| := \sqrt{y^T y} \). If \( f: \Omega \to \mathbb{R}^m \) is a differentiable function, its Jacobian matrix is the mapping \( \nabla f: \Omega \to \mathbb{R}^{m \times n} \) satisfying

\[
f(x + y) = f(x) + \nabla f(x) y + o(|y|)
\]
as \( y \to 0 \). In particular, gradients are row vectors.

2.1 Matrix derivatives

**Definition 2.1.** Let \( F: \Omega \to \mathbb{R}^{m \times k} \) be given. The directional derivative \( DF: \mathbb{R}^n \times \Omega \to \mathbb{R}^{m \times k} \) of \( F \) is defined by

\[
D_e F(x) := \lim_{h \to 0} \frac{F(x + he) - F(x)}{h}
\]
whenever the limit exists. When \( F \) is differentiable, the Jacobian derivative \( \nabla F: \mathbb{R}^k \times \Omega \to \mathbb{R}^{m \times n} \) of \( F \) is defined by

\[
\nabla_q F(x) := \nabla[F q](x).
\]

That is, the Jacobian matrix of the vector valued function \( x \mapsto F(x)q \).

It is possible to define the Jacobian in terms of combinations of partial derivatives, but we shall reserve the notation \( \nabla \) and \( \nabla_q \) for functions that are assumed to be differentiable.

Clearly, \( D_e F^T = (D_e F)^T \) and any symmetry of a square matrix \( F \) is therefore preserved. If \( F \) is assumed to be differentiable, the directional derivative satisfies

\[
F(x + y) = F(x) + D_y F(x) + o(|y|) \quad \text{as} \quad y \to 0,
\]
and one can check that

\[
D_e F(x)q = \nabla_q F(x)e \quad \forall e \in \mathbb{R}^n, \ q \in \mathbb{R}^k.
\]

Note that the dimensions match and that the above is an equality in \( \mathbb{R}^m \).

If \( q: \Omega \to \mathbb{R}^k \) and \( e: [a, b] \to \mathbb{R}^n \) are functions, we write

\[
\nabla_{q(x)} F(x) := \nabla_q F(x) \big|_{q=q(x)} \quad \text{and} \quad D_{e(t)} F(x) := D_e F(x) \big|_{e=e(t)}.
\]
Thus if $q$ is differentiable, the product rule yields

$$\nabla[Fq](x) = F(x)\nabla q(x) + \nabla_q F(x),$$

and if $c: [a, b] \to \Omega$ is a differentiable curve, we get, by the chain rule and by using (2.3), that

$$\frac{d}{dt}F(c(t)) = Dc'(t)F(c(t)).$$

Moreover, for vectors $p \in \mathbb{R}^m$ and $q \in \mathbb{R}^k$ we have

$$p^T \nabla_q F(x) = q^T \nabla_p F^T(x).$$

(2.5)

Note again that the dimensions match and that (2.5) is an equality in $\mathbb{R}^{1 \times n}$. Indeed, since $F^T(x)$ is a $k \times m$ matrix, the Jacobian $\nabla p F^T = \nabla[F^T p]$ is of dimension $k \times n$.

If the matrix function $F$ is a Hessian $H_u: \Omega \to S(n)$ of $u \in C^2(\Omega)$, then the directional and the Jacobian derivatives coincide and are again symmetric, that is,

$$D_{\xi} H_u(x) = \nabla_{\xi} H_u(x) \in S(n) \quad \forall \xi \in \mathbb{R}^n.$$  

(2.6)

It is, in fact, the Hessian matrix of the $C^2$ function $x \mapsto \nabla u(x)\xi$ in $\Omega$. Combined with (2.5), this means that

$$\xi_i^T D_{\xi_j} H u \xi_k = \xi_i^T D_{\xi_{\pi(j)}} H u \xi_{\pi(k)}$$

(2.7)

in $\Omega$ for all $\xi_i, \xi_j, \xi_k \in \mathbb{R}^n$ and all permutations $\pi$ on $\{i, j, k\}$. In particular, $e_i^T D_{\xi_j} H u e_k = \partial^3 u / \partial x_i \partial x_j \partial x_k$.

### 2.2 Symmetric matrices

The spectral theorem states that every symmetric $m \times m$ matrix can be diagonalized. For any $X \in S(m)$ there exists an orthogonal $m \times m$ matrix $U$ such that $U^T X U = \text{diag}(\lambda_1, \ldots, \lambda_m)$ where $\lambda_1 \leq \cdots \leq \lambda_m$ are the eigenvalues of $X$. Moreover, the eigenspaces $E_j := \{\xi \in \mathbb{R}^m \mid X \xi = \lambda_j \xi\}$ are $d_j$-dimensional subspaces of $\mathbb{R}^m$ where $d_j$ is the multiplicity of $\lambda_j$. The spaces $E_j$ and $E_k$ are orthogonal whenever $\lambda_j \neq \lambda_k$. Obviously, $E_j = E_k$ if $\lambda_j = \lambda_k$. By writing $U = (\xi_1, \ldots, \xi_m)$, we get that

$$X = U \text{diag}(\lambda_1, \ldots, \lambda_m) U^T = \sum_{i=1}^m \lambda_i \xi_i \xi_i^T$$

(2.8)

and that $E_j = \text{span}\{\xi_i \mid \lambda_i = \lambda_j\}$. 

4
The class of symmetric $m \times m$ projection matrices is denoted by

$$Pr(m) := \{P \in S(m) \mid PP = P\}.$$  

Since their eigenvalues are either 0 or 1, these matrices are on the form

$$P = \sum_{i=1}^{d} \xi_i \xi_i^T = QQ^T,$$

for some $d = 0, 1, \ldots, m$ (with the convention that empty sums are zero) and where $Q^TQ = I_d$. The set $\{\xi_1, \ldots, \xi_d\}$ is an orthonormal basis for the $d$-dimensional subspace

$$P(\mathbb{R}^m) := \{P\xi \mid \xi \in \mathbb{R}^m\} \subseteq \mathbb{R}^m.$$  

Conversely, given a subspace $E$ of $\mathbb{R}^m$, there is a unique symmetric projection $P$ such that $E = P(\mathbb{R}^m)$. Indeed, if $P(\mathbb{R}^m) = E = R(\mathbb{R}^m)$, then $P\xi, R\xi \in E$ for every $\xi \in \mathbb{R}^m$. Thus $RP\xi = P\xi$ and $PR\xi = R\xi$ and $P = P^T = (RP)^T = PR = R$. Note therefore that the factorization (2.9) is not unique as $P = \sum_{i=1}^{d} \eta_i \eta_i^T$ for every orthonormal basis $\{\eta_1, \ldots, \eta_d\}$ of $P(\mathbb{R}^m)$.

In the case of the symmetric matrix $X$ it follows that

$$P_j = \sum_{\lambda_i = \lambda_j}^{m} \xi_i \xi_i^T$$

is the unique eigenprojection corresponding to the $j$’th eigenvalue of $X$, regardless of the choice $U = (\xi_1, \ldots, \xi_m)$ of eigenvectors.

If we let $\alpha : \{1, \ldots, s\} \to \{1, \ldots, m\}$ be a re-indexing that picks out all of the $s := |\{\lambda_1, \ldots, \lambda_m\}|$ distinct eigenvalues of $X$, we may collect the terms in (2.8) with equal coefficients and write

$$X = \sum_{l=1}^{s} \lambda_{\alpha(l)} P_{\alpha(l)}.$$ (2.10)

Now,

$$P_{\alpha(l)}P_{\alpha(k)} = \delta_{lk} P_{\alpha(l)} \quad \text{and} \quad \sum_{l=1}^{s} P_{\alpha(l)} = \sum_{i=1}^{m} \xi_i \xi_i^T = I$$

and (2.10) is the unique representation of $X$ in terms of a complete set of eigenprojections and the unrepeated eigenvalues. However, since $\alpha$ depends on $X$ it is often more convenient to represent $X$ in terms of the repeated
versions $\lambda_i$ and $P_i$ where the indexing goes from 1 to $m$. This is obtained by noticing that

$$\sum_{i=1}^{m} \frac{\lambda_i}{d_i} P_i = \frac{\lambda_{\alpha(l)}}{d_{\alpha(l)}} P_{\alpha(l)} \sum_{i=1}^{m} \lambda_i = \lambda_{\alpha(l)} P_{\alpha(l)}$$

and thus

$$X = \sum_{l=1}^{s} \lambda_{\alpha(l)} P_{\alpha(l)} = \sum_{l=1}^{s} \sum_{\lambda_i = \lambda_{\alpha(l)}} \frac{\lambda_i}{d_i} P_i = \sum_{i=1}^{m} \lambda_i d_i P_i.$$

In [HJ91], the unrepeated eigenprojections are called the Frobenius co-variants and an explicit formula in terms of $X$ and the eigenvalues is given. In our notation

$$P_{\alpha(k)} = \prod_{l \neq k}^{s} \frac{X - \lambda_{\alpha(l)} I}{\lambda_{\alpha(k)} - \lambda_{\alpha(l)}}$$  \hspace{1cm} (2.11)$$

with the convention that an empty product is the identity. The formula can also be verified directly from (2.10).

3 Differentiation of the eigenprojections

By the above discussion, we can write $H(x)$ as

$$H(x) = \sum_{l=1}^{s} \lambda_{\alpha(l)}(x) P_{\alpha(l)}(x) = \sum_{i=1}^{m} \frac{\lambda_i(x)}{d_i(x)} P_i(x)$$

where either representation is unique in its specific sense. Here, $d_i(x) := \text{tr} P_i(x)$ and $s = s(x) \in \{1, \ldots, m\}$ is the number of different eigenvalues of $H$ at $x$.

For $1 \leq j \leq m$, let $A_j : \Omega \to S(m)$ be given by

$$A_j(x) := \sum_{\lambda_{\alpha(l)}(x) \neq \lambda_j(x)} \frac{P_{\alpha(l)}(x)}{\lambda_j(x) - \lambda_{\alpha(l)}(x)} = \sum_{\lambda_i(x) \neq \lambda_j(x)} \frac{P_i(x)/d_i(x)}{\lambda_j(x) - \lambda_i(x)}.$$


It commutes with $H$, $A_j P_j = P_j A_j = 0$ and satisfies
\[
A_j (\lambda_j I - H) = \sum_{l=1}^{s} \frac{P_{\alpha(l)}}{\lambda_j - \lambda_{\alpha(l)}} \sum_{k=1}^{s} (\lambda_j - \lambda_{\alpha(k)}) P_{\alpha(k)}
\]
\[
= \sum_{l=1}^{s} P_{\alpha(l)}
\]
\[
= I - P_j.
\]
It is therefore the pseudoinverse of the singular matrix $\lambda_j I - H$.

**Theorem 3.1** (Total derivative of eigenprojections). Let $H \in C^1(\Omega, S(m))$ with repeated eigenvalues and eigenprojections $\lambda_i(x)$ and $P_i(x)$, $i = 1, \ldots, m$. Let $j \in \{1, 2, \ldots, m\}$ and assume that either

- $P_j$ is continuous
- or $\lambda_j$ is differentiable and $\text{tr} P_j$ is constant

in $\Omega$. Then $P_j$ is differentiable in $\Omega$ with total derivatives

\[
\nabla_q P_j(x) = P_j(x) \nabla_{A_j(x)q} H(x) + A_j(x) \nabla_{P_j(x)q} H(x)
\]

and

\[
D_e P_j(x) = P_j(x) D_e H(x) A_j(x) + A_j(x) D_e H(x) P_j(x)
\]

for all $q \in \mathbb{R}^m$ and all $e \in \mathbb{R}^n$.

An immediate observation is that

\[
\text{tr} (D_e P_j H) = 0 = \text{tr} (D_e P_j).
\]

**Proof.** We shall prove the claim under the latter assumption in (3.1). The proof of the theorem is then completed by Proposition 4.6 which says that these two conditions are equivalent for $C^1$ matrices. It is worth mentioning that the Frobenius formula (2.11) is not directly applicable since the indexing will depend on $x \in \Omega$.

We dropp the subscripts and write $P := P_j$, $d := \text{tr} P$, $A := A_j$, and $\lambda := \lambda_j$. Fix $x \in \Omega$, which we may assume to be the origin, and let $y \in \mathbb{R}^n$ be small. By the differentiability assumptions

\[
0 = (H(y) - \lambda(y) I) P(y)
\]
\[
= (H - \lambda I + D_y H - \nabla \lambda y \cdot I) P(y) + o(|y|)
\]
as \( y \to 0 \). Functions written without an argument are to be understood as evaluated at \( x = 0 \). Multiplying from the left with \( A \) gives

\[
(I - P)P(y) = A(D_yH - \nabla \lambda y \cdot I)P(y) + o(|y|) = O(|y|).
\]

(3.3)

Since \( A = A(I - P) \), it follows from (3.3) that also

\[
(I - P)P(y) = AD_yHP(y) - \nabla \lambda y \cdot A(I - P)P(y) + o(|y|)
\]

\[
= AD_yHP(y) + o(|y|).
\]

(3.4)

It remains to find an estimate for \( PP(y) \). According to (2.9), we can for each \( y \) split \( P(y) \) into a product \( Q(y)Q^T(y) \). Although \( Q(y) \in \mathbb{R}^{m \times d(y)} \) is not unique, it is obviously bounded. Define

\[
R(y) := Q^TQ(y) \in \mathbb{R}^{d \times d(y)}
\]

and write

\[
I_{d(y)} = Q^T(y)Q(y)
\]

\[
= Q^T(y)PQ(y) + Q^T(y)(I - P)Q(y)
\]

\[
= R^T(y)R(y) + O(|y|^2)
\]

(3.5)

where the estimate on the last line is due to (3.3) after multiplying on the right by \( Q(y) \), and by the fact that \( I - P = (I - P)^2 \). This means that the eigenvalues of \( R^T(y)R(y) \in S(d(y)) \) are all in the range \( 1 + O(|y|^2) \). It is therefore invertible, and the inverse is bounded as the eigenvalues of \((R^T(y)R(y))^{-1}\) are again in the range \((1 + O(|y|^2))^{-1} = 1 + O(|y|^2) \). Hence

\[
(R^T(y)R(y))^{-1} = I_{d(y)} + O(|y|^2).
\]

(3.6)

We now use the assumption that \( P(y) \) has constant dimension \( d(y) \equiv d \). It implies that \( R(y) \) is square and, by taking the determinant of (3.5), we see that \( R(y) \) is invertible as well. The left-hand side of (3.6) may therefore be written as \( R^{-1}(y)(R^T(y))^{-1} \), and multiplication from the left with \( QR(y) \) and from the right with \( R^T(y)Q^T \) then yields

\[
P = QR(y)R^T(y)Q^T + O(|y|^2) = PP(y)P + O(|y|^2).
\]

(3.7)

Combining this with the transposed of (3.4) gives

\[
PP(y) = PP(y)(P + I - P) = P + PP(y)D_yHA + o(|y|)
\]

and thus

\[
P(y) = PP(y)(I - P)P(y)
\]

\[
= P + PP(y)D_yHA + AD_yHP(y) + o(|y|).
\]
Since the whole expression is $P + O(|y|)$, the factor $P(y)D_y H$ can be replaced with $(P + O(|y|))D_y H = PD_y H + o(|y|)$ and we finally conclude that

$$P(y) = P + PD_y H A + AD_y HP + o(|y|).$$

In order to get the formula for the directional derivative $D_e P$, substitute $y$ with $he$ where $e \in \mathbb{R}^n$ and let the number $h$ go to zero. As for the Jacobian derivative, the symmetry (2.7) implies that

$$P(y)q - P(0)q = (PD_y H A + AD_y HP)q + o(|y|)$$

$$= (P\nabla_{Aq} H + A\nabla_{Pq} H)y + o(|y|)$$

and $\nabla_q P = P\nabla_{Aq} H + A\nabla_{Pq} H$ being the Jacobian matrix of $x \mapsto P(x)q$ at $x = 0$. \[\square\]

When inspecting the above proof, it becomes clear that the same formula for the directional derivative $D_e P_i$ would have been produced if $y$ had been replaced with $he$ all the way from the beginning. But then the arguments work also for the weaker assumption of constant dimension and directional differentiability of the eigenvalue. Since Proposition 4.5 shows that the former of these two conditions implies the latter, the theorem below follows.

**Theorem 3.2** (Directional derivative of eigenprojections). Let $H \in C^1(\Omega, S(m))$ with repeated eigenvalues and eigenprojections $\lambda_i(x)$ and $P_i(x)$, $i = 1, \ldots, m$. Let $j \in \{1, 2, \ldots, m\}$ and assume that $\text{tr } P_j$ is constant in $\Omega$. Then $P_j$ is directionally differentiable in $\Omega$ with

$$D_e P_j(x) = P_j(x)D_e H(x)A_j(x) + A_j(x)D_e H(x)P_j(x)$$

for all $e \in \mathbb{R}^n$.

The following counterexample settles the question whether differentiability of eigenvalues implies constant dimension of the eigenprojections.

**Example 3.1** (A Hessian matrix with crossing differentiable eigenvalues). Let $u: \mathbb{R}^2 \rightarrow \mathbb{R}$ be the real part of $\frac{1}{12}(x + iy)^4$. That is,

$$u(x, y) = \frac{x^4 - 6x^2y^2 + y^4}{12}.$$  

Then

$$\mathcal{H}u(x, y) = \begin{pmatrix}
x^2 - y^2 & -2xy \\
-2xy & y^2 - x^2
\end{pmatrix}$$
with eigenvalues
\[
\lambda_{\pm}(x, y) = \frac{-\text{tr } H_{u}(x, y)}{2} \pm \frac{1}{2} \sqrt{\text{tr}^2 H_{u}(x, y) - 4 \det H_{u}(x, y)} = \pm \sqrt{-\det H_{u}(x, y)} = \pm (x^2 + y^2)
\]
that meet at the origin while still being differentiable.

One may check that \((y, x)^T\) is an eigenvector corresponding to the smallest eigenvalue. We therefore have that
\[
P_1(x, y) = \frac{1}{|y, x|^2} \begin{pmatrix} y \\ x \end{pmatrix} (y, x) = \frac{1}{x^2 + y^2} \begin{pmatrix} y^2 & xy \\ xy & x^2 \end{pmatrix}
\]
when \(x^2 + y^2 \neq 0\). Likewise,
\[
P_2(x, y) = \frac{1}{|x, -y|^2} \begin{pmatrix} x \\ -y \end{pmatrix} (x, -y) = \frac{1}{x^2 + y^2} \begin{pmatrix} x^2 & -xy \\ -xy & y^2 \end{pmatrix}.
\]

Observe that \(H_{u} = \lambda_1 P_1 + \lambda_2 P_2\) and that \(P_1 P_2 = 0\), and \(P_1 + P_2 = I\) as it should. At the origin, 0 = \(H_{u} = \lambda_{\alpha(1)} P_{\alpha(1)} = \frac{\lambda_1}{2} P_1 + \frac{\lambda_2}{2} P_2\) where, of course, \(\lambda_1 = \lambda_2 = \lambda_{\alpha(1)} = 0\) and \(P_1 = P_2 = P_{\alpha(1)} = I\).

4 Differentiation of the eigenvalues

We now set out to find the conditions that makes the eigenvalues of \(H \in C^1\) differentiable. Recalling (3.2), and since \(d_j(x) \lambda_j(x) = \text{tr } (P_j(x) H(x))\), the directional derivative \(D_e \lambda_j(x) = \frac{d}{dt} \lambda_j(x + te)\bigg|_{t=0}\) is formally given by
\[
D_e \lambda_j = \frac{1}{d_j} \text{tr } (P_j D_e H + D_e P_j H) = \frac{1}{d_j} \text{tr } (P_j D_e H).
\]
Although the above calculation required the assumption of a differentiable eigenprojection, we shall show that the identity still holds true whenever \(P_j(x)\) has merely constant dimension \(d_j\) (Proposition 4.5). This yields partial derivatives, but in order to get the total derivative \(\nabla \lambda_j\) it seems necessary to assume that the eigenprojection also is continuous. Theorem 4.7 summarizes the exact conditions for existence, and present formulas, for the gradient and the Hessian matrix of the eigenvalues.

Our main starting tools are Ky Fan’s minimum principle (Theorem 1 [Fan49]) and a powerful Lemma (Lemma 5.2 [Bel13]) restated below.
Lemma 4.1 (Derivative of a minimum). Let \( \mathcal{M} \) be a smooth compact manifold without boundary. Let \([a, b]\) be an interval of the real line and assume we have a function \( U \in C^1([a, b] \times \mathcal{M}) \). Define the function \( f \) on \([a, b]\) as

\[
f(t) := \min_{p \in \mathcal{M}} U(t, p)
\]

and let \( \Xi = \Xi(t) \) be the set

\[
\Xi(t) := \{ m \in \mathcal{M} \mid U(t, m) = f(t) \}.
\]

Then \( f \) is Lipschitz on \([a, b]\) and the one-sided derivatives exist and are given by

\[
\lim_{h \to 0^+} \frac{f(t + h) - f(t)}{h} = \min_{m \in \Xi(t)} \frac{\partial}{\partial t} U(t, m), \quad t \in [a, b),
\]

\[
\lim_{h \to 0^-} \frac{f(t + h) - f(t)}{h} = \max_{m \in \Xi(t)} \frac{\partial}{\partial t} U(t, m), \quad t \in (a, b].
\]

Lemma 4.2 (Ky Fan’s minimum principle). Let \( X \in S(m) \) with repeated eigenvalues \( \lambda_1 \leq \cdots \leq \lambda_m \). Then for every \( k = 1, \ldots, m \),

\[
\sum_{i=1}^{k} \lambda_i = \min \sum_{i=1}^{k} \xi_i^T X \xi_i
\]

where the minimum is taken over all \( k \)-tuples \( \{\xi_1, \ldots, \xi_k\} \) of vectors in \( \mathbb{R}^m \) such that \( \xi_i^T \xi_j = \delta_{ij} \).

As noted by Fan in the original proof, the right-hand side can be written as the minimum of \( \text{tr} \left( Q^T X Q \right) \) over all matrices \( Q \in \mathbb{R}^{m \times k} \) with \( Q^T Q = I_k \). In light of \([2, 3]\), it is clear that the principle may be restated in terms of projection matrices as

\[
\sum_{i=1}^{k} \lambda_i = \min_{R \in \text{Pr}_k(m)} \text{tr}(RX)
\]

where

\[
\text{Pr}_k(m) := \{ P \in S(m) \mid PP = P, \text{tr} P = k \}, \quad k = 0, 1, \ldots, m,
\]

are the \( k \)-dimensional subclasses of \( P(m) \). Note that \( \text{Pr}_k(m) \) can be identified with the Grassmannian manifold \( \text{Gr}_k(\mathbb{R}^m) \) of \( k \)-dimensional subspaces in \( \mathbb{R}^m \).
Let \( x: [a, b] \rightarrow \Omega \) be a \( C^1 \) curve. By setting \( \mathcal{M} = Pr_k(m) \) and \( U \in C^1([a, b] \times \mathcal{M}) \) as

\[
U(t, R) := \text{tr} \left( R H(x(t)) \right),
\]

the two lemmas imply that the sum of the \( k \) smallest eigenvalues of \( H \in C^1(\Omega, S(m)) \) at \( x(t) \),

\[
\ell_k(t) := \sum_{i=1}^{k} \lambda_i(x(t)) = \min_{R \in Pr_k(m)} U(t, R),
\]
is Lipschitz. Furthermore, by (2.4),

\[
\frac{\partial}{\partial t} U(t, R) = \frac{\partial}{\partial t} \text{tr} \left( R H(x(t)) \right) = \text{tr} \left( R D_{x'(t)} H(x(t)) \right)
\]

and the one-sided derivatives of \( \ell_k(t) \) are then

\[
\lim_{h \to 0^+} \frac{\ell_k(t + h) - \ell_k(t)}{h} = \min_{R \in \Xi_k(t)} \text{tr} \left( R D_{x'(t)} H(x(t)) \right), \quad t \in [a, b].
\]

\[
\lim_{h \to 0^-} \frac{\ell_k(t + h) - \ell_k(t)}{h} = \max_{R \in \Xi_k(t)} \text{tr} \left( R D_{x'(t)} H(x(t)) \right), \quad t \in (a, b],
\]

where

\[
\Xi_k(t) := \{ R \in Pr_k(m) \mid \text{tr}(R H(x(t))) = \ell_k(t) \}.
\]

We therefore want to show that \( \Xi_k(t) \) is a singleton for special values of \( k \). This will imply that the one-sided derivatives are equal and thus making \( \ell_k \) differentiable on \((a, b)\).

First we need a general result about projection matrices.

**Lemma 4.3.** Let \( P, R \in Pr(m) \). Then

\[
0 \leq \text{tr}(RP) \leq \text{tr} P
\]

with equality on the left if and only if \( RP = 0 \) and with equality on the right if and only if \( RP = P \).

**Proof.** Firstly,

\[
0 \leq \|RP\|^2 = \text{tr}(RP(RP)^T) = \text{tr}(RPR) = \text{tr}(RP),
\]

and if \( \text{tr}(RP) = 0 \), then \( RP = 0 \). Secondly,

\[
0 \leq \|RP - P\|^2 = \text{tr} \left( (RP - P)(RP - P)^T \right) = \text{tr}(RPR - RP - PR + P) = \text{tr} P - \text{tr}(RP).
\]

Thus \( \text{tr}(RP) \leq \text{tr} P \) and if they are equal, then \( RP = P \). \(\square\)
Lemma 4.4 (The special index $j^*$). Let $X = \sum_{i=1}^{m} \lambda_i P_i / d_i \in S(m)$ and set $\ell_k := \sum_{i=1}^{k} \lambda_i$ to be the sum of the $k$ smallest repeated eigenvalues. For $j = 1, \ldots, m$, define the indexes

$$j_* := \min \{ i \mid \lambda_i = \lambda_j \} \quad \text{and} \quad j^* := \max \{ i \mid \lambda_i = \lambda_j \}. \quad (4.2)$$

Then the set $\Xi_k := \{ R \in Pr_k(m) \mid \text{tr}(RX) = \ell_k \}$ is a singleton for $k = j^*$. Namely,

$$\Xi_{j^*} = \left\{ \left( \sum_{i=1}^{j^*} \frac{P_i}{d_i} \right) \right\}.$$

Proof. Write $R_j := \sum_{i=1}^{j_*} P_i / d_i$. First of all, since $\text{tr} R_j = \sum_{i=1}^{j_*} 1 = j^*$, and $j^* - d_j = j_* - 1 = (j_* - 1)^*$ (with $0^* := 0$), and

$$R_j = \sum_{i=1}^{j_*-1} \frac{P_i}{d_i} + \sum_{i=j_*}^{j^*} \frac{P_i}{d_i} = R_{j_*-1} + P_j,$$

it follows that $R_j \in Pr_{j^*}(m)$ by induction.

The leftward inclusion is clear since

$$\text{tr} (R_j X) = \text{tr} X R_{j_*-1} + \text{tr} X P_j = \ell_{j_*-1} + d_j \lambda_j = \ell_{j^*}.$$

Now assume that $\text{tr}(RX) = \ell_{j^*}$ for some $R \in Pr_{j^*}(m)$. We want to show that $R = R_j$. Split the matrix $Y := X - \lambda_j I$ into a negative semidefinite and positive semidefinite part as

$$Y = \sum_{i=1}^{j^*} (\lambda_i - \lambda_j) P_i / d_i + \sum_{i=j^*+1}^{m} (\lambda_i - \lambda_j) P_i / d_i.$$

We have $\text{tr}(RY) = \text{tr}(RX) - \lambda_j \text{tr} R = \ell_{j^*} - \lambda_j j^*$, so

$$\ell_{j^*} - \lambda_j j^* = \sum_{i=1}^{j^*} (\lambda_i - \lambda_j) \text{tr}(RP_i) / d_i + \sum_{i=j^*+1}^{m} (\lambda_i - \lambda_j) \text{tr}(RP_i) / d_i$$

$$\geq \sum_{i=1}^{j^*} (\lambda_i - \lambda_j) \text{tr}(RP_i) / d_i + 0$$

$$\geq \sum_{i=1}^{j^*} (\lambda_i - \lambda_j) P_i / d_i = \sum_{i=1}^{j^*} (\lambda_i - \lambda_j) = \ell_{j^*} - \lambda_j j^*$$

since $0 \leq \text{tr}(RP_i) \leq P_i$ by Lemma 4.3. The inequalities are therefore equalities, which in particular means that $\sum_{i=j^*+1}^{m} (\lambda_i - \lambda_j) \text{tr}(RP_i) / d_i = 0$. 

13
Since the coefficients $\lambda_i - \lambda_j$ are positive, we must have that $\text{tr}(RP_i) = 0$ and thus $RP_i = 0$ for $i = j^* + 1, \ldots, m$ by the Lemma. It follows that

$$RR_j = R \left( I - \sum_{i=j^*+1}^{m} P_i/d_i \right) = R,$$

and since $\text{tr}(RR_j) = \text{tr} R = j^* = \text{tr}(R_j)$, we can conclude that also $RR_j = R_j$.

**Proposition 4.5.** Let $H \in C^1(\Omega, S(m))$ and write

$$H(x) = \sum_{i=1}^{m} \lambda_i(x) P_i(x).$$

Assume that the eigenprojection $P_j(x)$ to the $j$'th eigenvalue $\lambda_j(x)$ has constant dimension along a $C^1$ curve $x: [a, b] \to \Omega$. That is,

$$\text{tr} P_j(x(t)) = d_j(x(t)) = d_j = \text{const.}$$

Then $\lambda_j \circ x$ is differentiable on $(a, b)$ and

$$\frac{d}{dt} \lambda_j(x(t)) = \frac{1}{d_j} \text{tr} \left( P_j(x(t)) D_{x(t)} H(x(t)) \right).$$

**Proof.** Let $x \in C^1([a, b], \Omega)$ and let $j \in \{1, \ldots, m\}$. By the definition (4.2), the indexes $j^*$ and $j^* - 1$ are in general functions of $t$. By a continuity argument (Lemma 5.3) one can prove that they are constant, but it turns out that this is insignificant for the proof of the proposition. What matters is that the difference $j^* - (j^* - 1) = d_j$ is constant along the curve.

Since $j^* - 1 = (j^* - 1)^*$ it follows by Lemma 4.4 that the sets

$$\Xi_{j^*}(t) := \left\{ R \in Pr_{j^*}(m) \mid \text{tr} \left( RH(x(t)) \right) = \ell_{j^*}(t) \right\} = \left\{ \sum_{i=1}^{j^*} \frac{P_i(x(t))}{d_i(x(t))} \right\}$$

and

$$\Xi_{j^* - 1}(t) := \left\{ R \in Pr_{j^* - 1}(m) \mid \text{tr} \left( RH(x(t)) \right) = \ell_{j^* - 1}(t) \right\} = \left\{ \sum_{i=1}^{j^* - 1} \frac{P_i(x(t))}{d_i(x(t))} \right\}$$

are singletons for each $t \in [a, b]$. Here,

$$\ell_{j^*}(t) := \sum_{i=1}^{j^*} \lambda_i(x(t)) \quad \text{and} \quad \ell_{j^* - 1}(t) := \sum_{i=1}^{j^* - 1} \lambda_i(x(t)).$$

14
Therefore, by the special case (4.1) of Lemma 4.1, we get that the derivatives of \( \ell_j^* \) and \( \ell_{j-1}^* \) exist and that they are given by

\[
\ell_j^*(t) = \text{tr} \left( \sum_{i=1}^{j^*} \frac{P_i(x(t))}{d_i(x(t))} D_{x(t)} \left( \sum_{i=1}^{j^*} P_i(x(t)) D_{x(t)} H(x(t)) \right) \right)
\]

and

\[
\ell_{j-1}^*(t) = \text{tr} \left( \sum_{i=1}^{j^*-1} \frac{P_i(x(t))}{d_i(x(t))} D_{x(t)} \left( \sum_{i=1}^{j^*-1} P_i(x(t)) D_{x(t)} H(x(t)) \right) \right).
\]

Since

\[
\ell_j^*(t) - \ell_{j-1}^*(t) = \sum_{i=j^*}^{j^*} \lambda_i(x(t)) = d_j \lambda_j(x(t)),
\]

and

\[
\sum_{i=1}^{j^*} \frac{P_i(x(t))}{d_i(x(t))} - \sum_{i=1}^{j^*-1} \frac{P_i(x(t))}{d_i(x(t))} = \sum_{i=j^*}^{j^*} \frac{P_i(x(t))}{d_i(x(t))} = P_j(x(t)),
\]

it follows that

\[
d_j \frac{d}{dt} \lambda_j(x(t)) = \ell_j^*(t) - \ell_{j-1}^*(t) = \text{tr} \left( P_j(x(t)) D_{x(t)} H(x(t)) \right)
\]

which is what we wanted to prove.

This completes the proof of Theorem 3.2 and the next proposition completes the proof of Theorem 3.1 by showing that the two assumptions 3.1 are equivalent.

**Proposition 4.6.** Let \( H \in C^1(\Omega, S(m)) \) and write

\[
H(x) = \sum_{i=1}^{m} \frac{\lambda_i(x)}{d_i(x)} P_i(x).
\]

The following are equivalent for \( j \in \{1, \ldots, m\} \) in \( \Omega \).

(a) \( \lambda_j \) is \( C^1 \) and \( P_j \) has constant dimension.

(b) \( \lambda_j \) is differentiable and \( P_j \) has constant dimension.

(c) \( P_j \) is differentiable.

(d) \( P_j \) is continuous.
Proof. (a) ⇒ (b) is immediate and the proof of Theorem 3.1 gives (b) ⇒ (c). The step (c) to (d) is again trivial, and if assuming (d), the dimension of $P_j$ is of course constant and by Proposition 4.5, the partial derivatives of $\lambda_j$ exists and are given by

$$
\frac{\partial}{\partial x_i} \lambda_j(x) = D_{e_i} \lambda_j(x) = \frac{1}{d_j} \text{tr} \left( P_j(x) D_{e} H(x) \right).
$$

Thus (d) ⇒ (a) since the partial derivatives then are seen to be continuous and $\lambda_j$ is therefore differentiable with a continuous gradient. □

We now gather the various regularity properties for eigenvalues of symmetric matrices. For completeness, we also record some statements valid when $H$ is only continuous.

**Theorem 4.7.** Let $H : \Omega \rightarrow S(m)$ be given and write

$$
H(x) = \sum_{i=1}^{m} \frac{\lambda_i(x)}{d_i(x)} P_i(x)
$$

where $\lambda_1(x) \leq \cdots \leq \lambda_m(x)$ and $d_i(x) = \text{tr} P_i(x)$. The following hold for $j \in \{1, \ldots, m\}$.

(A) Zero’th order properties. Assume that $H \in C(\Omega, S(m))$.

(i) $\lambda_j$ is continuous in $\Omega$.

(ii) If $H \in C^\alpha(\Omega, S(m))$ for some $0 < \alpha \leq 1$, then $\lambda_j \in C^\alpha(\Omega)$ with the same Hölder/Lipchitz-constant as $H$.

(B) First order properties. Assume that $H \in C^1(\Omega, S(m))$.

(i) If $P_j$ has constant dimension $\text{tr} P_j(x) = d_j$ in $\Omega$, then $\lambda_j$ has directional derivatives $D_{e} \lambda_j$ satisfying

$$
D_{e} \lambda_j(x) \cdot P_j(x) = P_j(x) D_{e} H(x) P_j(x)
$$

in every direction $e \in \mathbb{R}^n$. In particular,

$$
D_{e} \lambda_j(x) = \frac{1}{d_j} \text{tr} \left( P_j(x) D_{e} H(x) \right)
$$

and

$$
D_{e} \lambda_j(x) = \xi^T D_{e} H(x) \xi = \xi^T \nabla \xi H(x) e
$$

for any $\xi \in P_j(x)(\mathbb{R}^m) \cap S^{m-1}$. 

16
(ii) Suppose that one (and therefore all) of the conditions (a)-(d) from Proposition 4.6 holds. Then
\[ |\nabla \lambda_j(x)| \leq \frac{1}{\sqrt{d_j}} \max_{e \in S^{m-1}} \|D_e H(x)\|, \]

and
\[ \nabla \lambda_j(x) = \xi^T \nabla \xi H(x) \quad (4.3) \]

for any \( \xi \in P_j(x)(\mathbb{R}^m) \cap S^{m-1} \). Alternatively,
\[ \nabla \lambda_j(x)e = D_e \lambda_j(x) = \frac{1}{d_j} \operatorname{tr} \left( P_j(x)D_e H(x) \right) \quad (4.4) \]

for every \( e \in \mathbb{R}^n \).

(C) Second order properties. Assume that \( H \in C^2(\Omega, S(m)) \).

(i) Let \( a \in \mathbb{R}^n \). If \( P_j \) has constant dimension \( \operatorname{tr} P_j(x) = d_j \) in \( \Omega \), then \( D_a \lambda_j \) exists and has directional derivatives \( D_b D_a \lambda_j \) satisfying
\[
D_b D_a \lambda_j(x) \cdot P_j(x) = P_j(x) \left( D_b D_a H(x) + D_a H(x) A_j(x) D_b H(x) + D_b H(x) A_j(x) D_a H(x) \right) P_j(x)
\]
in every direction \( b \in \mathbb{R}^n \). In particular,
\[
D_b D_a \lambda_j(x) = \frac{1}{d_j} \operatorname{tr} \left( P_j(x) \left[ D_b D_a H(x) + 2 D_a H(x) A_j(x) D_b H(x) \right] \right)
\]

and
\[
D_b D_a \lambda_j(x) = \xi^T \left( D_b D_a H(x) + 2 D_a H(x) A_j(x) D_b H(x) \right) \xi
\]
\[
= a^T \left( \nabla \xi (\nabla \xi H)^T(x) + 2(\nabla \xi H(x))^T A_j(x) \nabla \xi H(x) \right) b
\]

for any \( \xi \in P_j(x)(\mathbb{R}^m) \cap S^{m-1} \).

(ii) Suppose that one (and therefore all) of the conditions (a)-(d) from Proposition 4.6 holds. Then \( \nabla \lambda_j \) is differentiable in \( \Omega \) with Hessian \( H \lambda_j := \nabla (\nabla \lambda_j^T) \) given by
\[
H \lambda_j(x) = \nabla \xi (\nabla \xi H)^T(x) + 2(\nabla \xi H(x))^T A_j(x) \nabla \xi H(x) \quad (4.5)
\]
for any \( \xi \in P_j(x)(\mathbb{R}^m) \cap S^{m-1} \). Alternatively,
\[
a^T \mathcal{H} \lambda_j(x)b = D_bD_a \lambda_j(x) \\
= \frac{1}{d_j} \text{tr} \left( P_j(x) [D_bD_aH(x) + 2D_aH(x)A_j(x)D_bH(x)] \right)
\]
(4.6)
for every \( a, b \in \mathbb{R}^n \).

**Remark 4.1.**

- Note that the eigenvectors \( \xi \) depend on \( x \in \Omega \).
- In (C), the matrices \( \nabla_p(\nabla_qH)^T \) and \( D_bD_aH \) represent the second order derivatives of \( H \). The first one is the Hessian matrix of the \( C^2 \) function \( x \mapsto p^T H(x)q \), and is therefore both in \( S(n) \) and symmetric in \( p \) and \( q \). The latter is the appropriate linear combination of the second order partial derivatives \( D_eD_eH(x) = \frac{\partial^2}{\partial x_i \partial x_k} H(x) \in S(m) \).
- Unlike in the case of the gradient, the expressions for \( \mathcal{H} \lambda_j \) contain the matrix \( A_j \) that cannot be assumed to be continuous or bounded.

**Proof of (A).** It is a well known fact that eigenvalues depend continuously on the matrix. Corollary 6.3.8 in [HJ13] gives the estimate
\[
\sum_{i=1}^m |\lambda_i - \tilde{\lambda}_i|^2 \leq \| E \|^2
\]
whenever \( \lambda_1 \leq \cdots \leq \lambda_m \) are the eigenvalues of \( H_0 \in S(m) \), and \( \tilde{\lambda}_1 \leq \cdots \leq \tilde{\lambda}_m \) are the eigenvalues of \( H_0 + E \in S(m) \). Our results follow by setting \( H_0 = H(x), E = H(x) - H(x+y) \), and by using that \( \| E \| = o(1) \) and \( \| E \| \leq C |y|^\alpha \), respectively.

**Proof of (B).** Part (i): Proposition 4.5 yields the formula \( D_e \lambda_j = \frac{1}{d_j} \text{tr}(P_jD_eH) \). But since also \( P_j \) is directionally differentiable by Theorem 3.2, the derivative of \( \lambda_j P_j = H P_j \) is \( D_e \lambda_j \cdot P_j + \lambda_j D_e P_j = D_e H P_j + HD_e P_j \) and the more general formula
\[
D_e \lambda_j \cdot P_j = P_j D_e H P_j
\]
is obtained by multiplying on the left with \( P_j \).

For Part (ii), assume that the conditions (a)-(d) from Proposition 4.6 hold.
Write the length of the gradient as $\max_{e \in \mathbb{S}^{n-1}} \nabla \lambda_j(x)e$ and use (4.4) to get

$$|\nabla \lambda_j(x)| = \max_{e \in \mathbb{S}^{n-1}} \frac{1}{d_j} \text{tr} \left( P_j(x) D_e H(x) \right)$$

$$\leq \max_{e \in \mathbb{S}^{n-1}} \frac{1}{d_j} \|P_j(x)\| \|D_e H(x)\|$$

$$= \frac{1}{\sqrt{d_j}} \max_{e \in \mathbb{S}^{n-1}} \|D_e H(x)\|.$$ 

Proof of (C). Part (i): Theorem 3.2 and (B) (i) implies that $D_a \lambda_j = \frac{1}{d_j} \text{tr}(P_j D_a H)$ and that it is directionally differentiable when $H$ is $C^2$. The identity $D_a \lambda_j \cdot P_j = P_j D_a HP_j$ can therefore be differentiated yielding

$$D_b D_a \lambda_j \cdot P_j + D_a \lambda_j \cdot D_b P_j = D_b P_j D_a HP_j + P_j D_b D_a H P_j + P_j D_a H D_b P_j.$$ 

Multiply from both sides with $P_j$ and use that $P_j D_b P_j = P_j D_b H A_j$ and $P_j D_b P_j P_j = 0$ to get

$$D_b D_a \lambda \cdot P_j = P_j D_b H A_j D_a H P_j + P_j D_b D_a H P_j + P_j D_a H A_j D_b H P_j$$

$$= P_j \left( D_b D_a H + D_b H A_j D_a H + D_a H A_j D_b H \right) P_j.$$ 

The other identities follow from the cyclic property of the trace and the symmetry of the factors, and by using (2.3) and (2.5).

Part (ii): When the conditions (a)-(d) from Proposition 4.6 hold, formula (4.4) shows that the gradient $\nabla \lambda$ is differentiable when $H$ is $C^2$. The rest follows from (i).

Analogous formulas for one-sided directional derivatives are given in [Tor01]. There the derivatives of $\lambda_j$ are expressed in terms of a specific eigenvalue of certain matrices. For example, and in our notation, the first order derivative is given as a particular eigenvalue of the $d_j \times d_j$ symmetric matrix $Q^T D_e H Q$ where $Q = (\xi_1, \ldots, \xi_{d_j}) \in \mathbb{R}^{m \times d_j}$ is an eigenvector matrix corresponding to $\lambda_j$. This interpretation is valid also for the formulas presented in Theorem 4.7 since $P_j = QQ^T$, and by (B) part (i),

$$Q^T D_e H Q = Q^T P_j D_e H P_j Q = D_e \lambda_j \cdot Q^T P_j Q = D_e \lambda_j \cdot Q^T Q = D_e \lambda_j \cdot I_{d_j}$$

and $Q^T D_e H Q$ is just a scaling of the identity matrix.
5 Asymptotic expansion of the eigenvalues

We conclude the paper by inserting the various expressions for the first- and second order derivatives of the eigenvalues into the expansions

\[
\lambda_j(x + he) = \lambda_j(x) + h D_e \lambda_j(x) + \frac{1}{2} h^2 D_e D_e \lambda_j(x) + o(h^2),
\]

\[
\lambda_j(x + y) = \lambda_j(x) + \nabla \lambda_j(x) y + \frac{1}{2} y^T \mathcal{H} \lambda_j(x) y + o(|y|^2).
\]

Recall that if \( x \mapsto H(x) \in S(m) \) has repeated eigenvalues and eigenprojections \( \lambda_i(x) \in \mathbb{R} \) and \( P_i(x) \in Pr(m), i = 1, \ldots, m \), then the pseudoinverse of \( \lambda_j(x) I - H(x) \) is given by

\[
A_j(x) := \sum_{i=1}^{m, \lambda_i(x) \neq \lambda_j(x)} \frac{P_i(x)/d_i(x)}{\lambda_j(x) - \lambda_i(x)}, \quad d_i(x) = \text{tr} P_i(x).
\]

**Corollary 5.1** (Second order directional expansion). Let \( H : \Omega \to S(m) \) be \( C^2 \) in a domain \( \Omega \subseteq \mathbb{R}^n \) and let \( j \in \{1, \ldots, m\} \). If the eigenprojection \( P_j \) has constant dimension \( d_j = \text{tr} P_j \), then for every \( x \in \Omega \) and for every direction \( e \in \mathbb{R}^n \),

\[
\lambda_j(x + he) = \xi^T \left( \lambda_j I + h D_e H + \frac{1}{2} h^2 D_e D_e H + h^2 D_e H A_j D_e H \right) \xi + o(h^2)
\]

\[
= \lambda_j + h \xi^T \nabla \xi H e + \frac{1}{2} h^2 e^T \left( \nabla \xi (\nabla \xi H)^T + 2(\nabla \xi H)^T A_j \nabla \xi H \right) e + o(h^2)
\]

as \( h \to 0 \) for any \( \xi \in P_j(x)(\mathbb{R}^m) \cap S^{m-1} \). Alternatively,

\[
\lambda_j(x + he) = \frac{1}{d_j} \text{tr} \left( P_j \left[ \lambda_j I + h D_e H + \frac{1}{2} h^2 D_e D_e H + h^2 D_e H A_j D_e H \right] \right) + o(h^2)
\]

The functions on the right-hand sides are all evaluated at \( x \). Of course, the corresponding first order expressions are also valid if \( H \) is \( C^1 \).

The assumption of constant dimension of \( P_j \) is sufficient for directional expansion. But in order to get the total asymptotic behavior we also need to assume that \( \lambda_j \) is differentiable or, equivalently, that \( P_j \) is continuous. We want to add one more condition to this list.

**Proposition 5.2.** Assume that \( H : \Omega \to S(m) \) is \( C^k \) in \( \Omega \) for some \( k = 0, 1, 2, \ldots \). If the number of distinct eigenvalues of \( H \) is constant, then every eigenvalue and eigenprojection of \( H \) is \( C^k \) in \( \Omega \).
Lemma 5.3 (Semicontinuity of some integer-valued functions associated to symmetric matrices). Let $H \in C(\Omega, S(m))$.

$$H(x) = \sum_{i=1}^{m} \frac{\lambda_i(x)}{d_i(x)} P_i(x), \quad d_i(x) = \text{tr} P_i(x).$$

For $j = 1, \ldots, m$, define the indexes $j_*(x) := \min \{i \mid \lambda_i(x) = \lambda_j(x)\}$ and $j^*(x) := \max \{i \mid \lambda_i(x) = \lambda_j(x)\}$, and let $s_j(x) := |\{\lambda_1(x), \ldots, \lambda_j(x)\}|$ be the number of distinct eigenvalues less or equal to $\lambda_j(x)$. Then $j_*$ and $s_j$ are lower semicontinuous, and $j^*$ and $d_j$ are upper semicontinuous in $\Omega$. If $d_j$ is constant, then $j_*$ and $j^*$ are also constant. Furthermore, the number $s_m(x)$ of distinct eigenvalues of $H$ at $x$ satisfies $s_m(x) = \sum_{i=1}^{m} 1/d_i(x)$, and if $s_m$ is constant, then so is every $d_j$.

Proof. We see that $j_*$ and $s_j$ decrease only if two different eigenvalues become equal. Since the eigenvalues of $H$ are continuous (Theorem 4.7 (A)), the superlevels sets $\{x \mid j_*(x) > c\}$ and $\{x \mid s_j(x) > c\}$ are open and the functions are therefore l.s.c. The same reasoning applies when arguing that $j^*$ and $d_j$ are u.s.c. There is however a more instructive proof of the upper semicontinuity of $d_j$. Since

$$0 = (H(x) - \lambda_j(x)I)P_j(x) = (H(x_0) - \lambda_j(x_0)I)P_j(x) + o(1)$$

as $x \to x_0$, multiplying on the left by $A_j(x_0)$ and rearranging gives $P_j(x) = P_j(x_0)P_j(x) + o(1)$. Lemma 4.3 then implies that $d_j(x) \leq d_j(x_0) + o(1)$ and it follows that $\limsup_{x \to x_0} d_j(x) \leq d_j(x_0)$.

If $d_j$ is constant, then $j^* = j_*$ and $d_j - 1$ is also l.s.c. It is therefore continuous and thus constant, which in turn makes $j_*$ constant.

Evaluate $s := s_m = \{|\{\lambda_1, \ldots, \lambda_m\}\|$ at some fixed $x \in \Omega$. As in Section 2 we choose a re-indexing $\alpha : \{1, \ldots, s\} \to \{1, \ldots, n\}$ so that $l \mapsto \lambda_{\alpha(l)}$ is a bijection. Since

$$\sum_{i=1}^{m} \frac{1}{d_i} = \sum_{\lambda_i=\lambda_{\alpha(l)}}^{m} \frac{1}{d_{\alpha(l)}} d_{\alpha(l)} = 1,$$

we get that

$$s = \sum_{l=1}^{s} 1 = \sum_{i=1}^{s} \sum_{\lambda_i=\lambda_{\alpha(l)}}^{m} \frac{1}{d_i} = \sum_{i=1}^{m} \frac{1}{d_i}.$$

Finally, since each $d_i$ is u.s.c., $\frac{1}{d_i}$ is l.s.c., and $-\frac{1}{d_i}$ is u.s.c. So if $s$ is constant, then

$$\frac{1}{d_j(x)} = s - \sum_{i \neq j}^{m} \frac{1}{d_i(x)}$$
is u.s.c. Thus $d_j$ is also l.s.c. and therefore continuous and constant. □

Proof of Proposition 5.2. By the Lemma, every eigenprojection has constant dimension and we can therefore re-index the eigenvalues and eigenprojections independently of $x \in \Omega$. See (2.10). After renaming we can write

$$H(x) = \sum_{i=1}^{s} \lambda_i(x) P_i(x)$$

where

$$\lambda_1(x) < \cdots < \lambda_s(x), \quad P_i(x) P_j(x) = \delta_{ij} P_i(x), \quad \sum_{i=1}^{s} P_i(x) = I,$$

for all $x \in \Omega$. Moreover, the Frobenius formula (2.11) becomes

$$P_j(x) = \prod_{\substack{i=1 \atop i \neq j}}^{s} \frac{H(x) - \lambda_i(x) I}{\lambda_j(x) - \lambda_i(x)}$$  \hspace{1cm} (5.1)

and shows that $P_j$ has (at least) the same regularity as $\bar{\lambda} := (\lambda_1, \ldots, \lambda_s)^T$. In particular, each $P_j$ and $\lambda_j$ is continuous whenever $H$ is continuous by Theorem 4.7 (A).

Assume next that $H$ is $C^k$ for some $k \geq 1$. By the above, $P_j$ is continuous and $\lambda_j$ is $C^1$ (and therefore also $P_j$) by Theorem 4.7 (B).

Since $\nabla \lambda_j(x)e_1 = \frac{1}{d_j} \text{tr}(P_j(x)D_e H(x))$, the derivative of every $\lambda_j$ is a smooth function of $P_j$ – which again is a smooth function of the eigenvalues and $H$ – and the tensor $D H$. In symbols,

$$\nabla \bar{\lambda} = F(\bar{\lambda}, H, DH),$$

and the result follows by induction. □

Corollary 5.4 (Second order total expansion). Let $H: \Omega \to S(m)$ be $C^2$ in a domain $\Omega \subseteq \mathbb{R}^n$ and let $j \in \{1, \ldots, m\}$. Assume that one of the following conditions hold in $\Omega$.

1. The eigenprojection $P_j$ is continuous.
2. $\lambda_j$ is differentiable and $d_j = \text{tr} P_j$ is constant.
3. The number of distinct eigenvalues of $H$ is constant.
Then, for every \( x \in \Omega \),
\[
\lambda_j(x + y) = \xi^T \left( \lambda_j I + D_y H + \frac{1}{2}D_y D_y H + D_y H A_j D_y H \right) \xi + o(|y|^2)
\]
\[
= \lambda_j + \xi^T \nabla \xi H y + \frac{1}{2} y^T \left( \nabla \xi (\nabla \xi H)^T + 2(\nabla \xi H)^T A_j \nabla \xi H \right) y + o(|y|^2)
\]
as \( y \to 0 \) for any \( \xi \in P_j(x)(\mathbb{R}^m) \cap S^{m-1} \). Alternatively,
\[
\lambda_j(x + y) = \frac{1}{d_j} \text{tr} \left( P_j [\lambda_j I + D_y H + \frac{1}{2}D_y D_y H + D_y H A_j D_y H] \right) + o(|y|^2)
\]
Observe that if \( m = 2 \), then constant dimension of an eigenprojection implies (3). Also when \( n = 1 \), the expansions can be written as
\[
\lambda_j(t + h) = \frac{1}{d_j} \text{tr} \left( P_j [\lambda_j I + h H' + \frac{1}{2} h^2 H'' + h^2 H' A_j H] \right) + o(h^2)
\]
\[
= \lambda_j + h \xi^T H' \xi + \frac{1}{2} h^2 \xi^T \left( H'' + 2 A_j H \right) \xi + o(t^2)
\]
and it is again enough to assume a constant \( d_j \) since the directional and total derivatives are equivalent on the real line.

Acknowledgements: Supported by the Academy of Finland (grant SA13316965), and Aalto University. I thank Juha Kinnunen, Peter Lindqvist, and Fredrik Arbo Høeg.

References

[ACL93] Alan L. Andrew, K.-w. Eric Chu, and Peter Lancaster. Derivatives of eigenvalues and eigenvectors of matrix functions. SIAM J. Matrix Anal. Appl., 14(4):903–926, 1993.

[Bel13] Giovanni Bellettini. Lecture notes on mean curvature flow, barriers and singular perturbations, volume 12 of Appunti. Scuola Normale Superiore di Pisa (Nuova Serie) [Lecture Notes. Scuola Normale Superiore di Pisa (New Series)]. Edizioni della Normale, Pisa, 2013.

[Fan49] Ky Fan. On a theorem of Weyl concerning eigenvalues of linear transformations. I. Proc. Nat. Acad. Sci. U. S. A., 35:652–655, 1949.
[HJ91] Roger A. Horn and Charles R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1991.

[HJ13] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013.

[HUY95] J.-B. Hiriart-Urruty and D. Ye. Sensitivity analysis of all eigenvalues of a symmetric matrix. *Numer. Math.*, 70(1):45–72, 1995.

[Kat95] Tosio Kato. *Perturbation theory for linear operators*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[LS01] Adrian S. Lewis and Hristo S. Sendov. Twice differentiable spectral functions. *SIAM J. Matrix Anal. Appl.*, 23(2):368–386, 2001.

[Tor01] Mounir Torki. Second-order directional derivatives of all eigenvalues of a symmetric matrix. *Nonlinear Anal.*, 46(8, Ser. A: Theory Methods):1133–1150, 2001.

[ZZX13] Liwei Zhang, Ning Zhang, and Xiantao Xiao. On the second-order directional derivatives of singular values of matrices and symmetric matrix-valued functions. *Set-Valued Var. Anal.*, 21(3):557–586, 2013.

Karl K. Brustad
Department of Mathematics and System Analysis
Aalto University
FI-00076, Aalto, Finland
karl.brustad@aalto.fi

24