Differentiating Generalized Eigen and Singular Value Decompositions

Jan de Leeuw

February 20, 2025

The derivatives of eigenvalues and eigenvectors (and singular values and vectors) are used in many places in multivariate data analysis. This paper reviews formulas for these derivatives and discusses several applications. R code implementing the basic formulas and applications is included. The results extend, generalize, correct, and improve the results of De Leeuw (2007).

Table of contents

Note: This is a working manuscript which will be updated frequently. All suggestions for improvement are welcome. All qmd, tex, html, pdf, R, and C files are in the public domain. Attribution will be appreciated, but is not required. Files can be found at https://github.com/deleeuw/generalized-eigen-derivatives

1 Introduction

1.1 GEV Systems

Suppose A and B are real symmetric matrices of order n, with B positive definite. Generalized eigenvalues and eigenvectors are defined as the solutions (x, λ) of the system of equations

$$Ax = \lambda Bx, \tag{1a}$$
$$x'Bx = 1. \tag{1b}$$

We call this a GEV system, short for Generalized EigenValue system.

The properties of the solutions of the system (1a),(1b) can be found in any textbook on matrix algebra, for example in Wilkinson (1965). We briefly summarise them here.

Solving equation (1a) is equivalent to solving the determinantal equation $\det(A - \lambda B) = 0$ for λ. If λ is a solution then $A - \lambda B$ is singular and has a non-trivial null-space. Any vector x in that null space satisfies (1a).

The determinant $\det(A - \lambda B)$ is a polynomial in λ of degree n, and consequently, by the Fundamental Theorem of Algebra, has n roots. Because A is symmetric and B is positive definite all n roots are real. Note that A can be indefinite and/or singular, which means that some roots can be negative or zero.

If $\lambda_s \neq \lambda_t$ are two solutions of the determinantal equation then the corresponding eigenvectors, which are defined up to a scale factor, x_s and x_t are B-orthogonal, i.e. $x_s'Bx_t = 0$. If λ_s is a root of multiplicity p then there are p corresponding eigenvectors, spanning a p-dimensional subspace of $\mathbb{R}^n$, and these p eigenvectors can be chosen to be B-orthogonal as well. If we use the normalization in (1b) it follows that there exists a non-singular matrix X and a diagonal Λ such that $X'BX = I$ and $AX = BX\Lambda$, which implies $X'AX = \Lambda$. If all roots are different the solution (X, Λ) is unique up to a permutation of the columns of X, and we can eliminate this non-uniqueness by requiring that $\lambda_1 > \lambda_2 > \cdots > \lambda_n$. If there are multiple roots, and we require $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, the solution is unique up to a rotation within each of the subspaces associated with multiple roots.

If λ_s is a simple eigenvalue, i.e. it is different from all other roots, then both λ_s and x_s are differentiable at (A, B) (see Wilkinson (1965), chapter 2, and for much more detail Kato (1976)). Suppose the matrices Δ_A and Δ_B are real and symmetric perturbations. Define[1][2]

$$A(\epsilon) := A + \epsilon\Delta_A + o(\epsilon), \tag{2a}$$
$$B(\epsilon) := B + \epsilon\Delta_B + o(\epsilon). \tag{2b}$$

[1]Here $o(\epsilon)$ is any function of ϵ satisfying $\lim_{\epsilon\rightarrow 0} o(\epsilon)/\epsilon = 0$.
[2]The symbol $:=$ is used for definitions.

Differentiability of λ_s and x_s implies that the differentials

$$d\lambda_s := \lim_{\epsilon\rightarrow 0}\frac{\lambda_s(A(\epsilon),B(\epsilon))-\lambda_s(A,B)}{\epsilon}, \tag{3a}$$

and

$$dx_s := \lim_{\epsilon\rightarrow 0}\frac{x_s(A(\epsilon),B(\epsilon))-x_s(A,B)}{\epsilon} \tag{3b}$$

exist. Our notation surpresses the dependence of $d\lambda_s$ and dx_s on (A,B) and on (Δ_A,Δ_B) because for our purposes these are just fixed constants.

1.2 GSV Systems

Suppose F is an $n\times m$ matrix, G is a positive definite matrix of order n, and H is a positive definite matrix of order m. We suppose without loss of generality that $n\geq m$. The generalized singular value problem for the triple (F,G,H) finds solutions to the system

$$Fy = \lambda Gx, \tag{4a}$$
$$F'x = \lambda Hy, \tag{4b}$$
$$x'Gx+y'Hy = 1, \tag{4c}$$

We refer to this as a GSV system, short for Generalized Singular Value system.

Now consider the GEV system

$$\begin{bmatrix}0&F\\F'&0\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\lambda\begin{bmatrix}G&0\\0&H\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}, \tag{5a}$$

with normalization

$$x'Gx+y'Hy=1. \tag{5b}$$

It is easy to see that x,y and λ satisfy (5a) and (5b) if and only if they satisfy (4a)-(4c). Note that in GSV systems the normalization constraint (4c) is often replaced by the constraint $x'Gx = y'Hy = 1$. We now show that this does not change the solutions, except for multiplying the singular vectors with a scale factor $\frac{1}{2}\sqrt{2}$.

The GSV system has m solutions with $\Lambda\geq 0$, $Y'HY = I$, and $X'GX = I$. Of these m solutions there are $r=\text{rank}(F)$ solutions with $\lambda_s>0$. For each of these m solutions (λ_s,x_s,y_s) there is a mirror solution $(-\lambda_s,x_s,-y_s)$, and these $2m$ solutions are also solutions of the GEV system. Because in GEV two solutions with different eigenvalues are B-orthogonal it follows that for a pair of mirror solutions with non-zero eigenvalue $x_s'Gx_s-y_s'Hy_s=0$, and thus, using (5b), $x_s'Gx_s=y_s'Hy_s=\frac{1}{2}$. In addition there are $n-m$ solutions with $\lambda_s=0$ and $Y=0$. The $n-m$ eigenvectors $X_\perp$ satisfy $F'X_\perp=0$ and $X_\perp'GX_\perp=I$. Note that if we replace (4c) by $x'Gx = y'Hy = 1$ then the $n - m$ solutions of (5a)-(5b) with $y_s = 0$ are no longer solutions of (4a)-(4c).

In summary, the solutions of the GEV and GSV systems for the ordered eigenvalues are

$$\begin{bmatrix}\Lambda & 0 & -\Lambda\end{bmatrix}, \tag{6a}$$

and for the corresponding eigenvectors

$$\begin{bmatrix}X & X_\perp & X \\ Y & 0 & -Y\end{bmatrix}, \tag{6b}$$

with $X_\perp$ a G-orthonormal basis for the null-space of F' and $Y_\perp$ an H-orthonormal basis for the null-space of F. The three distinct parts of the solutions in (6a) and (6b) are of orders $m, n - m$, and m. We are only really interested in the first r solutions, for which the singular values can be chosen to be positive.

2 Perturbation

2.1 Perturbed GEV Systems

We study the effect of symmetric perturbations $A + \epsilon\Delta_A + o(\epsilon)$ and $B + \epsilon\Delta_B + o(\epsilon)$ of A and B on the eigenvalues λ_s and their corresponding eigenvectors x_s. Throughout we assume that λ_s is a simple eigenvalue, i.e. that $\lambda_t \neq \lambda_s$ if $t \neq s$. Note that this does not mean all eigenvalues need to be different. Also note that if ϵ is small enough then $B(\epsilon)$ is still positive definite and $\lambda_s(A(\epsilon), B(\epsilon))$ is still a simple eigenvalue.

To find $d\lambda_s$ and dx_s we must solve the equations

$$\begin{multline} (A + \epsilon\Delta_A + o(\epsilon))(x_s + \epsilon dx_s + o(\epsilon)) = \\ (B + \epsilon\Delta_B + o(\epsilon)))(x_s + \epsilon dx_s + o(\epsilon))(\lambda_s + \epsilon d\lambda_s + o(\epsilon)), \tag{7a} \end{multline}$$

and

$$(x_s + \epsilon dx_s + o(\epsilon))'(B + \epsilon\Delta_B + o(\epsilon))(x_s + \epsilon dx_s + o(\epsilon)) = 1. \tag{7b}$$

Expand (7a) and (7b) and only keep the first order terms. This gives

$$A(dx_s) + \Delta_A x_s = d\lambda_s Bx_s + \lambda_s B(dx_s) + \lambda_s \Delta_B x_s, \tag{8a}$$

$$x_s'\Delta_B x_s + 2x_s' B(dx_s) = 0. \tag{8b}$$

Premultiply equation (8a) with x_s'. After some simplification this gives

$$d\lambda_s = x_s'(\Delta_A - \lambda_s\Delta_B)x_s, \tag{9}$$

We next solve for dx_s. Write $dx_s = X\alpha$, where X is any complete set of eigenvectors from $AX = BX\Lambda$, normalized by $X'BX = I$. Then (8a) becomes

$$BX(\Lambda - \lambda_s I)\alpha = d\lambda_s Bx_s - (\Delta_A x_s - \lambda_s\Delta_B)x_s. \tag{10}$$

Premultiplying by X' gives

$$(\Lambda - \lambda_s I)\alpha = (d\lambda_s)e_s - X'(\Delta_A - \lambda_s\Delta_B)x_s, \tag{11}$$

with e_s a unit vector.[3]

Premultiply both sides of (11) by e_s'. The result on the left hand side is zero, and using (9) the right hand side is also zero. Thus (11) cannot be used to find α_s. For $t \neq s$ we obtain

$$\alpha_t = -\frac{x_t'(\Delta_A - \lambda_s\Delta_B)x_s}{\lambda_t - \lambda_s}, \tag{12}$$

[3] A unit vector e_s has element s equal to one and all other elements equal to zero.

and, using (8b),

$$\alpha_s = -\frac{1}{2} x_s' \Delta_B x_s. \tag{13}$$

Thus

$$dx_s = -\sum_{t \neq s}^{n} \frac{x_t'(\Delta_A - \lambda_s \Delta_B) x_s}{\lambda_t - \lambda_s} x_t - \frac{1}{2}(x_s' \Delta_B x_s) x_s. \tag{14}$$

The two equations (9) and (14) are the basic tools we use in this paper.

Suppose that for there is an eigenvalue $\lambda \neq \lambda_s$ with multiplicity $r > 1$. In that case the corresponding eigenvectors are not uniquely defined, and it may seem that (14) may give different results for different choices of these eigenvectors. We briefly investigate this case. Suppose T is the index set with $\lambda_t = \lambda$ for all $t \in T$. There is an $n \times r$ matrix Y with $AY = \lambda BY$, $Y'BY = I$, unique up to a rotation. From (14)

$$\sum_{t \in T} \frac{y_t'(\Delta_A - \lambda_s \Delta_B) x_s}{\lambda_t - \lambda_s} y_t = \frac{YY'(\Delta_A - \lambda_s \Delta_B) x_s}{\lambda - \lambda_s}, \tag{15}$$

and the right hand side of is invariant under rotations of Y. Thus it does not matter how we choose the r eigenvectors corresponding to the multiple eigenvalue, the result of equation (2.1), and thus of (14), will always be the same. As long as λ_s is simple, we will have differentiability of both the eigenvalue λ_s and the corresponding eigenvector x_s.

Equations (9) and (14) simplify in some important special cases. For example, we can perturb A but not B. Thus $\Delta_B = 0$, and

$$d\lambda_s = x_s' \Delta_A x_s, \tag{16a}$$

and

$$dx_s = -\sum_{t \neq s}^{n} \frac{x_t' \Delta_A x_s}{\lambda_t - \lambda_s} x_t. \tag{16b}$$

If, in addition, $B = I$ we have perturbation equations for an SEV or simple eigenvalue problem. The case in which we perturb B and not A is handled in the same way. Other special cases and simplifications will be treated next.

2.2 Perturbed GSV Systems

We now have enough information to apply our previous perturbation results to GSV systems. The perturbations we consider have the form

$$\Delta_A = \begin{bmatrix} 0 & \Delta_F \\ \Delta_F' & 0 \end{bmatrix}, \tag{17a}$$

and

$$\Delta_B = \begin{bmatrix}\Delta_G & 0 \\ 0 & \Delta_H\end{bmatrix}, \tag{17b}$$

so that the perturbed system is still a GSV system.

We are only interested in perturbations of the first m singular values and vectors. For the singular values we find

$$d\lambda_s = 2x_s'\Delta_F y_s - \lambda_s(x_s'\Delta_G x_s + y_s'\Delta_H y_s). \tag{18}$$

Now suppose $s \leq m$ and $\lambda_s > 0$ is a simple singular value, not equal to any of the other λ_t. From (14) and from (17a)-(17b) we have

$$\begin{aligned}\begin{bmatrix}dx_s \\ dy_s\end{bmatrix} = &-\sum_{\substack{t=1\\t\neq s}}^{m}\frac{x_t'\Delta_F y_s + x_s'\Delta_F y_t - \lambda_s(x_s'\Delta_G x_t + y_s'\Delta_H y_t)}{\lambda_t - \lambda_s}\begin{bmatrix}x_t \\ y_t\end{bmatrix} \\ &+ \sum_{t=m+1}^{m+n}\frac{x_t'\Delta_F y_s - \lambda_s x_s'\Delta_G x_t}{\lambda_s}\begin{bmatrix}x_t \\ 0\end{bmatrix} \\ &+ \sum_{\substack{t=1\\t\neq s}}^{m}\frac{x_t'\Delta_F y_s - x_s'\Delta_F y_t + \lambda_s(x_s'\Delta_G x_t - y_s'\Delta_H y_t)}{\lambda_t + \lambda_s}\begin{bmatrix}x_t \\ -y_t\end{bmatrix} \\ &- \frac{1}{2}(x_s'\Delta_G x_s + y_s'\Delta_H y_s)\begin{bmatrix}x_s \\ y_s\end{bmatrix}\end{aligned} \tag{19}$$

2.3 Parametric Perturbations

If A and B are differentiable functions of a vector of q parameters θ then

$$A(\theta + \epsilon d\theta) = A(\theta) + \epsilon\sum_{r=1}^{q}(d\theta_r)\mathcal{D}_r A + o(\epsilon), \tag{20a}$$

$$B(\theta + \epsilon d\theta) = B(\theta) + \epsilon\sum_{r=1}^{q}(d\theta_r)\mathcal{D}_r B + o(\epsilon). \tag{20b}$$

Here the $\mathcal{D}_r A$ is a matrix with partial derivatives of A with respect to θ_r, evaluated at θ, and $d\theta_r$ is element r of the perturbation $d\theta$. And similarly for B.

We now can apply equations (9) and (14) with

$$\Delta_A = \sum_{r=1}^{q}(d\theta_r)\mathcal{D}_r A, \tag{21a}$$

$$\Delta_B = \sum_{r=1}^{q}(d\theta_r)\mathcal{D}_r B. \tag{21b}$$

If A and B depend on two different sets of parameters then we can use the same equations with some of the $\mathcal{D}_r A$ and some of the $\mathcal{D}_r B$ equal to zero.

2.3.1 Linear Perturbations

In an important special case A and B are linear in θ. So

$$A(\theta) = \sum_{r=1}^{q} \theta_r A_r, \tag{22a}$$

$$B(\theta) = \sum_{r=1}^{q} \theta_r B_r. \tag{22b}$$

In that case $\Delta_A = \mathcal{D}_r A = A_r$ and $\Delta_B = \mathcal{D}_r B = B_r$.

2.3.2 Elementwise Perturbations

In an important special case of the linear case the parameters are all the $n(n+1)$ elements of A and B on and above the diagonal. We have

$$A = \sum\sum_{1\leq i<j\leq n} a_{ij} E_{ij} + \sum_{i=1}^{n} a_{ii} E_i, \tag{23a}$$

$$B = \sum\sum_{1\leq i<j\leq n} b_{ij} E_{ij} + \sum_{i=1}^{n} b_{ii} E_i, \tag{23b}$$

with $E_{ij} := e_i e_j' + e_j e_i'$ and $E_i := e_i e_i'$.

2.4 Perturbation Code

The code in Section 6 has the function perturbGeigen(), written in R (R Core Team (2024)), which has arguments A, B, Δ_A, Δ_B and p. The argument p is a subset of $\{1, 2, \cdots, n\}$, with $1 \leq m \leq n$ elements. The program computes $d\lambda_s$ and dx_s for all $s \in p$ and returns them, respectively, in a vector with m elements and an $n \times m$ matrix. The computations use (16a) and (16b).

The code section also has the function perturbCheck(), whichb has the same arguments as perturbGeigen(), plus the additional argument eps, the value of ϵ. The function computes generalized eigenvalues and eigenvectors of the pair $(A + \epsilon\Delta_A, B + \epsilon\Delta_B)$ and compares them with the output of perturbGeigen().

We do not give numerical examples using these functions, because they just generate a large number of values that have no meaning. Suffices it to say that our checks indicate the perturbation functions work well and our approximations are as close as can be expected. Readers can use the code to try their their own examples. We will give numerical examples in Section 4, where they are more meaningful.

3 Partial Derivatives

3.1 Basic Partial Derivatives

The parametric perturbation results can be easily translated into the notation for partial derivatives. Let's introduce the notation first. If f is a function of a vector θ then the partial derivative with respect to θ_r is $\mathcal{D}_r f$, defined by

$$\mathcal{D}_r f(\theta) := \lim_{\epsilon \to 0} \frac{f(\theta + \epsilon e_r) - f(\theta)}{\epsilon}, \tag{24}$$

with e_r a unit vector.

To find partial derivatives we only perturb θ_r by setting $d\theta = e_r$ in our perturbation equations (21a) and (21b). We then have $\Delta_A = \mathcal{D}_r A$ and $\Delta_B = \mathcal{D}_r B$, and thus

$$\mathcal{D}_r \lambda_s = x_s'(\mathcal{D}_r A - \lambda_s \mathcal{D}_r B)x_s, \tag{25a}$$

and

$$\mathcal{D}_r x_s = -\sum_{t \neq s}^{n} \frac{x_t'(\mathcal{D}_r A - \lambda_s \mathcal{D}_r B)x_s}{\lambda_t - \lambda_s} x_t - \frac{1}{2}(x_s' \mathcal{D}_r B x_s)x_s. \tag{25b}$$

For linear parametric perturbations we have the same equations for the partial derivatives with $\mathcal{D}_r A = A_r$ and $\mathcal{D}_r B = B_r$.

3.2 Elementwise Perturbations

For elementwise perturbations there are some useful simplifications. If we apply (23a) and (23b) to (25a) and (25b) we get

$$\mathcal{D}_{ij}^A \lambda_s = \begin{cases} 2x_{is}x_{js} & \text{if } i \neq j, \\ x_{is}^2 & \text{if } i = j. \end{cases} \tag{26a}$$

$$\mathcal{D}_{ij}^B \lambda_s = \begin{cases} -2\lambda_s x_{is}x_{js} & \text{if } i \neq j, \\ -\lambda_s x_{is}^2 & \text{if } i = j. \end{cases} \tag{26b}$$

$$\mathcal{D}_{ij}^A x_s = \begin{cases} -\sum_{t \neq s}^n \frac{x_{is}x_{jt} + x_{js}x_{it}}{\lambda_t - \lambda_s} x_t & \text{if } i \neq j, \\ -\sum_{t \neq s}^n \frac{x_{is}x_{it}}{\lambda_t - \lambda_s} x_t & \text{if } i = j. \end{cases} \tag{26c}$$

$$\mathcal{D}_{ij}^B x_s = \begin{cases} \lambda_s \sum_{t \neq s}^n \frac{x_{is}x_{jt} + x_{js}x_{it}}{\lambda_t - \lambda_s} x_t - x_{is}x_{js}x_s & \text{if } i \neq j, \\ \lambda_s \sum_{t \neq s}^n \frac{x_{is}x_{it}}{\lambda_t - \lambda_s} x_t - \frac{1}{2}x_{is}^2 x_s & \text{if } i = j. \end{cases} \tag{26d}$$

In (26a)-(26d) we use the somewhat ad-hoc notation $\mathcal{D}^A_{ij}$ and $\mathcal{D}^B_{ij}$ for the partial derivatives with respect to a_{ij} and b_{ij}.

As an aside, instead of deriving (26a)-(26d) from (25a) and (25b) we could also have used the chain rule to derive (25a) and (25b) from (26a)-(26d). This looks like

$$\mathcal{D}_r\lambda_s = \sum\sum_{1\leq i\leq j\leq n} \mathcal{D}^A_{ij}\lambda_s\mathcal{D}_r a_{ij} + \sum\sum_{1\leq i\leq j\leq n} \mathcal{D}^B_{ij}\lambda_s\mathcal{D}_r b_{ij}, \tag{27a}$$

$$\mathcal{D}_r x_s = \sum\sum_{1\leq i\leq j\leq n} \mathcal{D}^A_{ij}x_s\mathcal{D}_r a_{ij} + \sum\sum_{1\leq i\leq j\leq n} \mathcal{D}^B_{ij}x_s\mathcal{D}_r b_{ij}. \tag{27b}$$

3.3 Second Order Partials

For various purposes in data analysis, such as Newton's method or asymptotic bias correction, we need the second derivatives of the eigenvalues and eigenvectors.

We start by differentiating equation (25a) with respect to θ_u, using the abbreviations $C_{rs} := \mathcal{D}_r A - \lambda_s\mathcal{D}_r B$ and $C_{rus} := \mathcal{D}_{ru}A - \lambda_s\mathcal{D}_{ru}B$. This gives

$$\mathcal{D}_{ur}\lambda_s = 2(d_u x_s)'C_{rs}x_s + x_s'C_{rus}x_s - (d_u\lambda_s)x_s'\mathcal{D}_r Bx_s, \tag{28}$$

We could expand this further by substituting $d_u\lambda_s$ and $d_u x_s$ from (25a) and (25b). But in our computations we will use (28) as is, even though (28) does not show immediately that for each s the Hessian with elements $\mathcal{D}_{ru}\lambda_s$ is a symmetric matrix of order q. Note that in the linear case $\mathcal{D}_{ru}A = \mathcal{D}_{ru}B = 0$, so the middle term on the right disappears.

The logical next step is to differentiate (25b) with respect to θ_u. We compute the second partials for one single element x_{ks} of X at a time. For each element x_{ks} the Hessian will be a symmetric matrix of order q. From (25b)

$$\mathcal{D}_r x_{ks} = -\sum_{t\neq s}^{n}\frac{x_t'C_{rs}x_s}{\lambda_t-\lambda_s}x_{kt} - \frac{1}{2}(x_s'\mathcal{D}_r Bx_s)x_{ks}. \tag{29}$$

We start by working on the first term on the right of (29). Differentiating with respect to θ_u gives

$$\mathcal{D}_u\left\{\frac{x_t'C_{rs}x_s}{\lambda_t-\lambda_s}x_{kt}\right\} = \frac{x_t'C_{rs}x_s}{\lambda_t-\lambda_s}d_u x_{kt} + \mathcal{D}_u\left\{\frac{x_t'C_{rs})x_s}{\lambda_t-\lambda_s}\right\}x_{kt}. \tag{30}$$

The derivative in the second term on the right of (30) evaluates to

$$\mathcal{D}_u\left\{\frac{x_t'C_{rs}x_s}{\lambda_t-\lambda_s}\right\} = \frac{(\lambda_t-\lambda_s)\mathcal{D}_u\{x_t'C_{rs}x_s\} - x_t'C_{rs}x_s(d_u\lambda_t - d_u\lambda_s)}{(\lambda_t-\lambda_s)^2}. \tag{31}$$

The derivative in the first term of the numerator on the right of (31) is

$$\mathcal{D}_u\{x_t'C_{rs}x_s\} = (d_u x_t)'C_{rs}x_s + x_t'C_{rs}d_u x_s + x_t'C_{rus}x_s - (d_u\lambda_s)x_t'(\mathcal{D}_r B)x_s. \tag{32}$$

And finally, differentiating the last term in (29),

$$\mathcal{D}_u\{(x_s'\mathcal{D}_r Bx_s)x_{ks}\} = \{2(d_u x_s)'\mathcal{D}_r Bx_s + x_s'\mathcal{D}_{ru}Bx_s\}x_{ks} + (x_s'\mathcal{D}_r Bx_s)d_u x_{ks}. \tag{33}$$

We combine the results in (30)-(33) into one big equation. The resulting formula is rather ugly, but think of it as a recipe for calculation, not as a beautiful object in its own right.

$$\begin{aligned}\mathcal{D}_{ru}x_{ks} = &-\sum_{t\neq s}^{n}\left\{\frac{x_t'C_{rs}x_s}{\lambda_t-\lambda_s}d_u x_{kt} + \right.\\ &+\frac{(d_u x_t)'C_{rs}x_s + x_t'C_{rs}(d_u x_s) + x_t'C_{rus}x_s - (d_u\lambda_s)x_t'(\mathcal{D}_r B)x_s}{\lambda_t-\lambda_s}x_{kt} + \\ &\left.-\frac{x_t'C_{rs}x_s(d_u\lambda_t - d_u\lambda_s)}{(\lambda_t-\lambda_s)^2}x_{kt}\right\} + \\ &-\{(d_u x_s)'D_r Bx_s + x_s'\mathcal{D}_{ru}Bx_s\}x_{ks} - \frac{1}{2}(x_s'\mathcal{D}_r Bx_s)d_u x_{ks}.\end{aligned} \tag{34}$$

There are simplifications in the linear case, where the second derivatives of A and B are zero, and we can replace $\mathcal{D}_r A$ and $\mathcal{D}_r B$ in (34) by A_r and B_r. Thus $C_{rs} = A_r - \lambda_s B_r$. Also $C_{rus} = 0$ and $\mathcal{D}_{ru}B = 0$. Thus equations (28) and (34) become

$$\mathcal{D}_{ru}\lambda_s = 2(d_u x_s)'C_{rs}x_s - d_u\lambda_s x_s'B_r x_s, \tag{35}$$

and

$$\begin{aligned}\mathcal{D}_{ru}x_{ks} = &-\sum_{t\neq s}\left\{\frac{x_t'C_{rs}x_s}{\lambda_t-\lambda_s}d_u x_{kt}\right.\\ &+\frac{(d_u x_t)'C_{rs}x_s + x_t'C_{rs}d_u x_s - (d_u\lambda_s)x_t'B_r x_s}{\lambda_t-\lambda_s}x_{kt}\\ &\left.-\frac{x_t'C_{rs}x_s(d_u\lambda_t - d_u\lambda_s)}{(\lambda_t-\lambda_s)^2}x_{kt}\right\}\\ &-((d_u x_s)'B_r x_s)x_{ks} - \frac{1}{2}(x_s'B_r x_s)d_u x_{ks}.\end{aligned} \tag{36}$$

There are more simplifications for SEV problems where B does not depend on θ and thus $\mathcal{D}B = 0$.

3.4 Partial Derivative Code

The functions partialGeigen() and partialCheck() can be used for linear perturbations. They have both have arguments $theta, a, b, s$, where a and b are lists of matrices of lengths q qnd p and $theta$ is a vector of length $p+q$. The last q matrices in the list a are zero, as are the first p matrices in the list b. The index $1 \leq s \leq n$ dictates which eigen-pair we study.

partialGeigen() uses the formulas (25a) and (25b), while partialCheck computes numerical derivatives using grad() and jacobian() from the numDeriv package (Gilbert and Varadhan (2019)). For our example we use the same A and B as before, and we use elementwise perturbation. For the dominant eigenvalue partialGeigen() gives the derivatives with respect to the elements of A and B as

We also used partialCheck() to compute numerical derivatives. The maximum absolute difference between the numerical and analytical partials of the eigenvalue is

The partialGeigen() function also gives the partials of the dominant eigenvector. The partials of the three eigenvector elements with respect to the elements of A are

and those with respect to B are

The maximum absolute differences between the numerical and analytical partials of the eigenvector are, for A and B,

In elementwise perturbation partialGeigen() uses equations (25a) and (25b) and consequently needs lists of binary sparse matrices as arguments. This is very wasteful, both in memory and speed, given the fact that we also have the compact equations (26a) and (26). We have added the more specialized function partialElement() that uses these compact equations. It does not use θ and the lists with the A_r and B_r. It gives the same results as partialGeigen(), but is much faster.

The function hessianGeigenEval() computes the second partials of the eigenvalues for linear perturbations. It has the same arguments as partialGeigen(). The function returns a list with the second partials of the eigenvalues.

hessianCheckEval() that computes numerical second partials. Even in our small example the Hessian is already of order 12, so we do not show the actual matrix. We do show that the maximum absolute difference between the numerical and analytical second partials of the eigenvalue is

3.5 Partial Derivatives

4 Applications

4.1 Generalized Canonical Analysis

Suppose we have m variables, with variable j having k_j categories or levels. This defines $q := \prod_{j=1}^m k_j$ profiles, which are binary vectors of length $\sum_{j=1}^m k_j$. The data are the relative frequencies of the profiles (cf. Gifi (1990), chapter 2). If the number of variables is large the number of profiles will be very large, even if the number of categories for each variable is moderate. The profile frequencies will tend to be small and many of them will be zero.

In

$$A = \sum_{r=1}^{q} p_r g_r g_r', \tag{37a}$$

$$B = \sum_{r=1}^{q} p_r B_r, \tag{37b}$$

where p_r is the relative frequency of profile r, and g_r is the profile vector. Matrix B_r is diagonal, with g_r on the diagonal.

This is a linear parametric model, and consequently we can apply the formulas from Section 2.3.1 to find the derivatives of the eigenvalues and eigenvectors with respect to the p_r.

First

$$\mathcal{D}_r\lambda_s = x_s'(g_r g_r' - \lambda_s B_r)x_s, \tag{38a}$$

and

$$\mathcal{D}_r x_s = -\sum_{t\neq s}\frac{x_t'(g_r g_r' - \lambda_s D_r)x_s}{\lambda_t - \lambda_s}x_t - \frac{1}{2}(x_s' D_r x_s)x_s. \tag{38b}$$

We now apply the delta method (Mann and Wald (1943)), assuming that p is a realization of a multinomial random variable $\underline{p}$[4] with parameters (N, π) from a multinomial with probabilities π_r. It follows that the eigenvalues $N^{-\frac{1}{2}}(\underline{\lambda} - \lambda)$ converges in distribution to a multivariate normal with mean zero and covariances

$$\text{ACOV}(\underline{\lambda}_s, \underline{\lambda}_t) = \sum_{r=1}^{R} \pi_r x_s'(g_r g_r' - \lambda_s B_r)x_s x_t'(g_r g_r' - \lambda_t B_r)x_t,$$

[4] We use the "Dutch convention" (Hemelrijk, 1966) of underlining random variables

assuming that λ_s and λ_t are simple eigenvalues.

$$\mathcal{D}_{ur}\lambda_s = 2(d_u x_s)' C_{rs} x_s + x_s' C_{rus} x_s - (d_u \lambda_s) x_s' \mathcal{D}_r B x_s, \tag{39}$$

4.2 Principal Component Analysis

If we think of PCA as a function of the elements of covariance matrix then using our results is real simple. The elementwise perturbation results and partial derivative results from Section 3.2 apply directly, using $B = I$ and $\Delta_B = 0$. If we interpret PCA as a function of the correlation matrix matters become slightly more complicated. A is still the covariance matrix, but B is now the diagonal of the covariance matrix. The $\frac{1}{2}n(n+1)$ covariances are still the parameters, but the diagonal elements contribute to both A and B. Thus there are $\frac{1}{2}n(n+1)$ design matrices A_r and B_r, where the B_r corresponding with off-diagonal elements are zero.

It is perhaps more interesting to interpret PCA as a function of the data matrix. Suppose we have m numerical variables, with variable j having k_j possible values. This defines $q := \prod_{j=1}^m k_j$ possible profiles, which are vectors of length m with all combinations of the variable values. The data are the relative frequencies of these profiles, collected in a vector p of length q. Write G for the $q \times m$ matrix of profiles.

The GEV problem for the principal component analysis (PCA) of a covariance matrix has

$$A = G'(P - pp')G, B = I$$

while for a PCA of the correlation matrix

$$A = G'(P - pp')G, B = \text{diag}(G'(P - pp')G)$$

From ...

$$\mathcal{D}_r A = g_r g_r' - (\mu e_r' + e_r \mu'),$$

with $\mu := Gp$.

From ... $\mathcal{D}_r B = 0$. From ...

$$\mathcal{D}_r B = \text{diag}(g_r g_r') - 2\mu_r$$

These can be used in (25b) and (25b).

$$\mathcal{D}_r \lambda_s = x_s' \{g_r g_r' - (\mu e_r' + e_r \mu') - \lambda_s \{\text{diag}(g_r g_r') - 2\mu_r\} x_s$$

Alternative data matrix Z and PCA is an SVD of JZ, which

Jackknife, Infinitesimal Jackknife

4.3 Canonical Analysis

In Canonical Analysis
$$A=\begin{bmatrix}0&F'G\\G'F&0\end{bmatrix},$$
and
$$B=\begin{bmatrix}F'F&0\\0&G'G\end{bmatrix}$$
Perturb F and G, which gives
$$\Delta_A=\begin{bmatrix}0&\Delta_F'G+F'\Delta_G\\G'\Delta_F+F\Delta_G&0\end{bmatrix}$$
and
$$\Delta_B=\begin{bmatrix}\Delta_F'F+F'\Delta_F&0\\0&\Delta_G'G+G'\Delta_G\end{bmatrix}$$
We could introduce Multiple Correspondence Analysis (MCA) as a form of canonical analysis and use the perturbation results from Section 4.3. Instead we go directly to a parametric approach.

4.4 Classical Multidimensional Scaling

In classical multidimensional scaling (MDS) we have a symmetric matrix D of squared dissimilarities.

$$\xi_{ij}=-\frac{1}{2}\left\{\theta_{ij}^2-\frac{1}{n}\sum_{l=1}^n\theta_{il}^2-\frac{1}{n}\sum_{l=1}^n\theta_{lj}^2+\frac{1}{n^2}\sum_{k=1}^n\sum_{l=1}^n\theta_{kl}^2\right\}$$
$$A=\sum\sum_{1\leq i<j\leq n}\xi_{ij}E_{ij}+\sum_{i=1}^n\xi_{ii}E_i$$

4.5 Factor Analysis

4.6 Low rank Matrix Approximation

5 Discussion

If B is a singular, if B is indefinite. If A and B are not symmetric.

6 Code

References

De Leeuw, J. 2007. "Derivatives of Generalized Eigen Systems with Applications." Preprint Series 528. Los Angeles, CA: UCLA Department of Statistics. https://jansweb.netlify.app/publication/deleeuw-r-07-c/deleeuw-r-07-c.pdf.

Gifi, A. 1990. *Nonlinear Multivariate Analysis*. New York, N.Y.: Wiley.

Gilbert, P., and R. Varadhan. 2019. *numDeriv: Accurate Numerical Derivatives*. https://CRAN.R-project.org/package=numDeriv.

Kato, T. 1976. *Perturbation Theory for Linear Operators*. Second Edition. Springer.

Mann, H. B., and A. Wald. 1943. "On Stochastic Limit and Order Relationships." *Annals of Mathematical Statistics* 14: 217–26.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wilkinson, J. H. 1965. *The Algebraic Eigenvalue Problem*. Clarendon Press.