Appendix A: Matrix Algebra

As a two-dimensional array we define a quadratic and rectangular matrix. *First*, we review *matrix algebra* with respect to two *inner and one external relation*, namely multiplication of a matrix by a scalar, addition of matrices of the same order, matrix multiplication of type *Cayley, Kronecker-Zehfuss, Khatri-Rao and Hadamard*. *Second*, we introduce *special matrices* of type symmetric, antisymmetric, diagonal, unity, null, idempotent, normal, orthogonal, orthonormal (special facts of representing a 2×2 orthonormal matrix, a general $n \times n$ orthonormal matrix, the *Helmert representation of an orthonormal matrix* with examples, special facts about the representation of a *Hankel matrix* with examples, the definition of a *Vandermonde matrix*), the *permutation* matrix, the *commutation* matrix. *Third*, scalar measures like *rank*, *determinant*, *trace* and *norm*. In detail, we review the *Inverse Partitional Matrix /IPM/* and the *Cayley inverse of the sum of two matrices*. We summarize the notion of a *division algebra*. A special paragraph is devoted to *vector-valued* matrix forms like vec, vech and veck. *Fifth*, we introduce the notion of *eigenvalue-eigenvector decomposition* (analysis *versus* synthesis) and the *singular value decomposition*. *Sixth*, we give details of generalized inverse, namely *g-inverse, reflexive g-inverse*, *reflexive symmetric g-inverse*, *pseudo inverse*, *Zlobec formula, Bjerhammar formula, rank factorization, left and right inverse, projections, bordering, singular value representation* and the theory solving linear equations.

A1 Matrix-Algebra

A matrix is a *rectangular* or a *quadratic array of numbers,*

$$\mathbf{A} := [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m-1} & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m-1} & a_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n-11} & a_{n-12} & \dots & a_{n-1m-1} & a_{n-1m} \\ a_{n1} & a_{n2} & \dots & a_{nm-1} & a_{nm} \end{bmatrix}, \quad a_{ij} \in \mathbb{R}, [a_{ij}] \in \mathbb{R}^{n \times m}.$$

The format or *"order"* of **A** is given by the number n of *rows* and the number of the *columns,*

$$O(\mathbf{A}) := n \times m.$$

Fact:
Two matrices are identical if they have identical format and if at each place (i, j) are identical numbers, namely

$$\mathbf{A} = \mathbf{B} \Leftrightarrow a_{ij} = b_{ij} \begin{bmatrix} i \in \{1, \dots, n\} \\ j \in \{1, \dots, m\}. \end{bmatrix}$$

Beside the identity of two matrices the transpose of an $m \times n$ matrix $\mathbf{A} = [a_{ij}]$ is the $m \times n$ matrix $\mathbf{A}' = [a_{ji}]$ whose *ij* element is a_{ji}.

Fact:

$$(\mathbf{A}')' = \mathbf{A}.$$

A *matrix algebra* is defined by the following operations:

- multiplication of a matrix by a *scalar* (external relation)
- addition of two matrices of the same order (*internal relation*)
- multiplication of two matrices (*internal relation*)

Definition (matrix additions and multiplications):

(1) Multiplication by a scalar

$$\mathbf{A} = [a_{ij}],\ \alpha \in \mathbb{R} \Rightarrow \alpha\mathbf{A} = \mathbf{A}\alpha = [\alpha a_{ij}].$$

(2) Addition of two matrices of the same order

$$\mathbf{A} = [a_{ij}],\ \mathbf{B} = [b_{ij}] \Rightarrow \mathbf{A} + \mathbf{B} := [a_{ij} + b_{ij}]$$

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \quad (\textit{commutativity})$$

$$(\mathbf{A}+\mathbf{B})+\mathbf{C} = \mathbf{A}+(\mathbf{B}+\mathbf{C})\ (\textit{associativity})$$

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-1)\mathbf{B}\ (\textit{inverse addition}).$$

Compatibility

$$\left.\begin{aligned}(\alpha+\beta)\mathbf{A} &= \alpha\mathbf{A} + \beta\mathbf{A}\\ \alpha(\mathbf{A}+\mathbf{B}) &= \alpha\mathbf{A} + \alpha\mathbf{B}\end{aligned}\right] \ \textit{distributivity}$$

$$(\mathbf{A}+\mathbf{B})' = \mathbf{A}' + \mathbf{B}'.$$

(3) Multiplication of matrices

3(i) "Cayley-product" ("matrix-product")

$$\begin{bmatrix}\mathbf{A} = [a_{ij}],\ O(\mathbf{A}) = n \times l\\ \mathbf{B} = [b_{ij}],\ O(\mathbf{B}) = l \times m\end{bmatrix} \Rightarrow$$

$$\Rightarrow \mathbf{C} := \mathbf{A} \cdot \mathbf{B} = [c_{ij}] := \sum_{k=1}^{l} a_{ik} b_{kl},\ O(\mathbf{C}) = n \times m$$

3(ii) "Kronecker-Zehfuss-product"

$$\left.\begin{aligned}\mathbf{A} &= [a_{ij}],\ O(\mathbf{A}) = n \times m\\ \mathbf{B} &= [b_{ij}],\ O(\mathbf{B}) = k \times l\end{aligned}\right] \Rightarrow$$

With permission from the author: Extract from "Erik W. Grafarend: Linear and nonlinear models. Fixed effects, random effects, and mixed models"

$$\Rightarrow \mathbf{C} := \mathbf{B} \otimes \mathbf{A} = [c_{ij}],\ \mathbf{B} \otimes \mathbf{A} := [b_{ij}\mathbf{A}],\ O(\mathbf{C}) = O(\mathbf{B} \otimes \mathbf{A}) = kn \times l$$

3(iii) *"Khatri-Rao-product"*
(of two rectangular matrices of identical column number)

$$\left.\begin{array}{r} \mathbf{A} = [a_1, ..., a_m],\ O(\mathbf{A}) = n \times m \\ \mathbf{B} = [b_1, ..., b_m],\ O(\mathbf{B}) = k \times m \end{array}\right] \Rightarrow$$

$$\Rightarrow \mathbf{C} := \mathbf{B} \odot \mathbf{A} := [b_1 \otimes a_1, \ldots, b_m \otimes a_m],\ O(\mathbf{C}) = kn \times m$$

3(iv) *"Hadamard-product"*
(of two rectangular matrices of the same order; elementwise product)

$$\left.\begin{array}{r} \mathbf{G} = [g_{ij}],\ O(\mathbf{G}) = n \times m \\ \mathbf{H} = [h_{ij}],\ O(\mathbf{H}) = n \times m \end{array}\right] \Rightarrow$$

$$\Rightarrow \mathbf{K} := \mathbf{G} * \mathbf{H} = [k_{ij}],\ k_{ij} := g_{ij} h_{ij},\ O(\mathbf{K}) = n \times m\,.$$

The existence of the product $\mathbf{A} \cdot \mathbf{B}$ does not imply the existence of the product $\mathbf{B} \cdot \mathbf{A}$. If both products exist, they are in general not equal. Two *quadratic* matrices A and B, for which holds $\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A}$, are called *commutative.*

Laws

(i)

$$(\mathbf{A} \cdot \mathbf{B}) \cdot \mathbf{C} = \mathbf{A} \cdot (\mathbf{B} \cdot \mathbf{C})$$
$$\mathbf{A} \cdot (\mathbf{B} + \mathbf{C}) = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \mathbf{C}$$
$$(\mathbf{A} + \mathbf{B}) \cdot \mathbf{C} = \mathbf{A} \cdot \mathbf{C} + \mathbf{B} \cdot \mathbf{C}$$
$$(\mathbf{A} \cdot \mathbf{B})' = (\mathbf{B}' \cdot \mathbf{A}')\,.$$

(ii)

$$(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) = \mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C}$$
$$(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = (\mathbf{A} \otimes \mathbf{B}) + (\mathbf{B} \otimes \mathbf{C})$$
$$\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) + (\mathbf{A} \otimes \mathbf{C})$$
$$(\mathbf{A} \otimes \mathbf{B}) \cdot (\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A} \cdot \mathbf{C}) \otimes (\mathbf{B} \cdot \mathbf{D})$$
$$(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'\,.$$

(iii)

$$(\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C} = \mathbf{A} \odot (\mathbf{B} \odot \mathbf{C}) = \mathbf{A} \odot \mathbf{B} \odot \mathbf{C}$$
$$(\mathbf{A} + \mathbf{B}) \odot \mathbf{C} = (\mathbf{A} \odot \mathbf{C}) + (\mathbf{B} \odot \mathbf{C})$$
$$\mathbf{A} \odot (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \odot \mathbf{B}) + (\mathbf{A} \odot \mathbf{C})$$
$$(\mathbf{A} \cdot \mathbf{C}) \odot (\mathbf{B} \cdot \mathbf{D}) = (\mathbf{A} \odot \mathbf{B}) \cdot (\mathbf{C} \odot \mathbf{D})$$
$$\mathbf{A} \odot (\mathbf{B} \cdot \mathbf{D}) = (\mathbf{A} \odot \mathbf{B}) \cdot \mathbf{D},\ \text{if}\ d_{ij} = 0\ \text{for}\, i \neq j.$$

The transported *Khatri-Rao-product* generates a *row product* which we do not follow here.

(iv)
$$\mathbf{A}*\mathbf{B}=\mathbf{B}*\mathbf{A}$$
$$(\mathbf{A}*\mathbf{B})*\mathbf{C}=\mathbf{A}*(\mathbf{B}*\mathbf{C})=\mathbf{A}*\mathbf{B}*\mathbf{C}$$
$$(\mathbf{A}+\mathbf{B})*\mathbf{C}=(\mathbf{A}*\mathbf{C})+(\mathbf{B}*\mathbf{C})$$
$$(\mathbf{A}_1\cdot\mathbf{B}_1\cdot\mathbf{C}_1)*(\mathbf{A}_2\cdot\mathbf{B}_2\cdot\mathbf{C}_2)=(\mathbf{A}_1\odot\mathbf{A}_2)'\cdot(\mathbf{B}_1\otimes\mathbf{B}_2)\cdot(\mathbf{C}_1\odot\mathbf{C}_2)$$
$$(\mathbf{D}\cdot\mathbf{A})*(\mathbf{B}\cdot\mathbf{D})=\mathbf{D}\cdot(\mathbf{A}*\mathbf{B})\cdot\mathbf{D},\ \text{if}\ d_{ij}=0\ \text{for}\, i\neq j$$
$$(\mathbf{A}*\mathbf{B})'=\mathbf{A}'*\mathbf{B}'.$$

A2 Special Matrices

We will collect special matrices of symmetric, antisymmetric, diagonal, unity, zero, idempotent, normal, orthogonal, orthonormal, positive-definite and positive-semidefinite, special orthonormal matrices, for instance of type *Helmert* or of type *Hankel.*

Definitions (special matrices):

A quadratic matrix $\mathbf{A}=[a_{ij}]$ of the order $O(\mathbf{A})=n\times n$ is called

symmetric $\Leftrightarrow$ $a'_{ij}=a_{ji}\forall i,j\in\{1,...,n\}:\mathbf{A}=\mathbf{A}'$

antisymmetric $\Leftrightarrow$ $a_{ij}=-a_{ji}\forall i,j\in\{1,...,n\}:\mathbf{A}=-\mathbf{A}'$

diagonal $\Leftrightarrow$ $a_{ij}=0\ \ \forall i\neq j,$
$\mathbf{A}=\mathrm{Diag}[a_{11},...,a_{nn}]$

unity $\Leftrightarrow$ $\mathbf{I}_{n\times n}=\left[\begin{array}{l} a_{ij}=0\ \ \forall i\neq j \\ a_{ij}=1\ \ \forall i\neq j\end{array}\right.$

zero matrix $\mathbf{0}_{n\times n}:\ a_{ij}=0\ \forall i,j\in\{1,...,n\}$

$\left.\begin{array}{l}\textit{upper}\\ \\ \textit{lower}\end{array}\right]$ *triangular*: $\left[\begin{array}{l} a_{ij}=0\ \forall i>j \\ a_{ij}=0\ \forall i<j\end{array}\right.$

idempotent if and only if $\mathbf{A}\cdot\mathbf{A}=\mathbf{A}$
normal if and only if $\mathbf{A}\cdot\mathbf{A}'=\mathbf{A}'\cdot\mathbf{A}$.

Definition (orthogonal matrix) :

The matrix $\mathbf{A}$ is called *orthogonal* if $\mathbf{AA}'$ and $\mathbf{A}'\mathbf{A}$ are diagonal matrices. (The rows and columns of $\mathbf{A}$ are *orthogonal*.)

Definition (orthonormal matrix) :

The matrix $\mathbf{A}$ is called *orthonormal* if $\mathbf{AA}' = \mathbf{A}'\mathbf{A} = \mathbf{I}$.
(The rows and columns of $\mathbf{A}$ are *orthonormal.*)

Facts (representation of a 2×2 orthonormal matrix) $\mathbf{X} \in \mathrm{SO}(2)$:

A 2×2 orthonormal matrix $\mathbf{X} \in \mathrm{SO}(2)$ is an element of the special orthogonal group SO(2) defined by

$$\mathrm{SO}(2) := \{\mathbf{X} \in \mathbb{R}^{2\times 2} \mid \mathbf{X}'\mathbf{X} = \mathbf{I}_2 \text{ and } \det \mathbf{X} = +1\}$$

$$\{\mathbf{X} = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \in \mathbb{R}^{2\times 2} \left| \begin{array}{l} x_1^2 + x_2^2 = 1 \\ x_1x_3 + x_2x_4 = 0 \;, \quad x_1x_4 - x_2x_3 = +1\} \\ x_3^2 + x_4^2 = 1 \end{array} \right.$$

(i) $$\mathbf{X} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \in \mathbb{R}^{2\times 2} \,, \; \phi \in [0,\, 2\pi]$$

is a trigonometric representation of $\mathbf{X} \in \mathrm{SO}(2)$.

(ii) $$\mathbf{X} = \begin{bmatrix} x & \sqrt{1-x^2} \\ -\sqrt{1-x^2} & x \end{bmatrix} \in \mathbb{R}^{2\times 2} \quad , \quad x \in [+1, \;\; -1]$$

is an algebraic representation of $\mathbf{X} \in \mathrm{SO}(2)$

$(x_{11}^2 + x_{12}^2 = 1,\; x_{11}x_{21} + x_{12}x_{22} = -x\sqrt{1-x^2} + x\sqrt{1-x^2} = 0,\; x_{21}^2 + x_{22}^2 = 1)$.

(iii) $$\mathbf{X} = \begin{bmatrix} \dfrac{1-x^2}{1+x^2} & +\dfrac{2x}{1+x^2} \\ -\dfrac{2x}{1+x^2} & \dfrac{1-x^2}{1+x^2} \end{bmatrix} \in \mathbb{R}^{2\times 2}, \quad x \in \mathbb{R}$$

is called a stereographic projection of $\mathbf{X}$
(stereographic projection of SO(2) ~ $\mathbb{S}^1$ onto $\mathbb{L}^1$).

(iv) $$\mathbf{X} = (\mathbf{I}_2 + \mathbf{S})(\mathbf{I}_2 - \mathbf{S})^{-1},\; \mathbf{S} = \begin{bmatrix} 0 & x \\ -x & 0 \end{bmatrix},$$

where $\mathbf{S} = -\mathbf{S}'$ is a *skew matrix* (antisymmetric matrix), is called a *Cayley-Lipschitz representation* of $\mathbf{X} \in \mathrm{SO}(2)$.

(v) $\mathbf{X} \in \mathrm{SO}(2)$ is a commutative group (“*Abel*”)

(*Example*: $\mathbf{X}_1 \in \mathrm{SO}(2)$, $\mathbf{X}_2 \in \mathrm{SO}(2)$, then $\mathbf{X}_1\mathbf{X}_2 = \mathbf{X}_2\mathbf{X}_1$)

($\mathrm{SO}(n)$ for $n = 2$ is the only commutative group, $\mathrm{SO}(n \mid n \neq 2)$ is not “*Abel*”).

Facts (representation of an $n{\times}n$ orthonormal matrix) $\mathbf{X} \in \mathrm{SO}(n)$:

An $n{\times}n$ orthonormal matrix $\mathbf{X} \in \mathrm{SO}(n)$ is an element of the special orthogonal group SO(n) defined by

$$\mathrm{SO}(n) := \{\mathbf{X} \in \mathbb{R}^{n\times n} \mid \mathbf{X}'\mathbf{X} = \mathbf{I}_n \text{ and } \det\mathbf{X} = +1\} .$$

As a differentiable manifold SO(n) inherits a *Riemann structure* from the *ambient space* $\mathbb{R}^{n^2}$ with a *Euclidean metric* ($\operatorname{vec}\mathbf{X}' \in \mathbb{R}^{n^2}$, dim $\operatorname{vec}\mathbf{X}' = n^2$). Any *atlas* of the special orthogonal group SO(n) has at least *four distinct charts* and there is one with exactly four charts. (“minimal atlas”: *Lusternik – Schnirelmann category*)

(i) $\mathbf{X} = (\mathbf{I}_n + \mathbf{S})(\mathbf{I}_n - \mathbf{S})^{-1}$,

where $\mathbf{S} = -\mathbf{S}'$ is a *skew matrix* (antisymmetric matrix), is called a *Cayley-Lipschitz representation* of $\mathbf{X} \in \mathrm{SO}(n)$.
($n!/2(n-2)!$ is the number of independent parameters/coordinates of **X**)

(ii) If each of the matrices $\mathbf{R}_1, \cdots, \mathbf{R}_k$ is an $n{\times}n$ *orthonormal* matrix, then their product

$$\mathbf{R}_1\mathbf{R}_2\cdots\mathbf{R}_{k-1}\mathbf{R}_k \in \mathrm{SO}(n)$$

is an $n{\times}n$ *orthonormal* matrix.

Facts (*orthonormal matrix: Helmert representation*) :

Let $\mathbf{a}' = [\mathbf{a}_1, \cdots, \mathbf{a}_n]$ represent any row vector such that $\mathbf{a}_i \neq \mathbf{0}$ $(i \in \{1, \cdots, n\})$ is any row vector whose elements are all nonzero. Suppose that we require an $n{\times}n$ *orthonormal matrix,* one row which is proportional to $\mathbf{a}'$. In what follows *one such matrix* **R** *is derived*.

Let $[\mathbf{r}_1', \cdots, \mathbf{r}_n']$ represent the *rows* of **R** and take the *first row* $\mathbf{r}_1'$ to be the row of **R** that is proportional to $\mathbf{a}'$. Take the *second row* $\mathbf{r}_2'$ to be proportional to the n-dimensional row vector

$$[\mathbf{a}_1, -\mathbf{a}_1^2/\mathbf{a}_2, \ \mathbf{0}, \ \mathbf{0}, \ \cdots, \mathbf{0}], \tag{H2}$$

the *third row* $\mathbf{r}_3'$ proportional to

$$[\mathbf{a}_1, \mathbf{a}_2, -(\mathbf{a}_1^2 + \mathbf{a}_2^2)/\mathbf{a}_3, \mathbf{0}, \mathbf{0}, \cdots, \mathbf{0}] \tag{H3}$$

and more generally the first through nth rows $\mathbf{r}_1', \cdots, \mathbf{r}_n'$ proportional to

$$[\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_{k-1}, -\sum_{i=1}^{k-1}\mathbf{a}_i^2/\mathbf{a}_k, \mathbf{0}, \mathbf{0}, \cdots, \mathbf{0}] \tag{Hn-1}$$

$$\text{for } k \in \{2, \cdots, n\} ,$$

respectively confirm to yourself that the n-1 vectors ($\mathbf{H}_{n-1}$) are orthogonal to each other and to the vector $\mathbf{a}'$. In order to obtain explicit expressions for $\mathbf{r}'_1, \cdots,$ $\mathbf{r}'_n$ it remains to *normalize* $\mathbf{a}'$ and the vectors ($\mathbf{H}_{n-1}$). The *Euclidean norm* of the kth of the vectors ($\mathbf{H}_{n-1}$) is

$$\{\sum_{i=1}^{k-1}\mathbf{a}_i^2 + (\sum_{i=1}^{k-1}\mathbf{a}_i^2)^2/\mathbf{a}_k^2\}^{1/2} = \{(\sum_{i=1}^{k-1}\mathbf{a}_i^2)(\sum_{i=1}^{k}\mathbf{a}_i^2)/\mathbf{a}_k^2\}^{1/2}.$$

Accordingly for the *orthonormal vectors* $\mathbf{r}'_1, \cdots, \mathbf{r}'_n$ we finally find

(1st row) $$\mathbf{r}'_1 = [\sum_{i=1}^{n}\mathbf{a}_i^2]^{-1/2}(\mathbf{a}_1, \cdots, \mathbf{a}_n)$$

(kth row) $$\mathbf{r}'_k = [\frac{\mathbf{a}_k^2}{(\sum_{i=1}^{k-1}\mathbf{a}_i^2)(\sum_{i=1}^{k}\mathbf{a}_i^2).}]^{-1/2}(\mathbf{a}_1,\ \mathbf{a}_2, \cdots, \mathbf{a}_{k-1},\ -\sum_{i=1}^{k-1}\frac{\mathbf{a}_i^2}{\mathbf{a}_k},\ \mathbf{0},\ \mathbf{0},\ \cdots, \mathbf{0})$$

(nth row) $$\mathbf{r}'_n = [\frac{\mathbf{a}_n^2}{(\sum_{i=1}^{n-1}\mathbf{a}_i^2)(\sum_{i=1}^{n}\mathbf{a}_i^2).}]^{-1/2}[\mathbf{a}_1,\ \mathbf{a}_2, \cdots, \mathbf{a}_{n-1},\ -\sum_{i=1}^{n-1}\frac{\mathbf{a}_i^2}{\mathbf{a}_n}]\ .$$

The recipy is complicated: When $\mathbf{a}' = [1, 1, \cdots, 1, 1]$, the *Helmert factors* in the 1st row, …, kth row,…, nth row simplify to

$$\mathbf{r}'_1 = n^{-1/2}[1, 1, \cdots, 1, 1] \in \mathbb{R}^n$$

$$\mathbf{r}'_k = [k(k-1)]^{-1/2}[1, 1, \cdots, 1, 1-k,\ 0,\ 0, \cdots,\ 0,\ 0] \in \mathbb{R}^n$$

$$\mathbf{r}'_n = [n(n-1)]^{-1/2}[1, 1, \cdots, 1, 1-n] \in \mathbb{R}^n.$$

The orthonormal matrix

$$\begin{bmatrix} \mathbf{r}'_1 \\ \mathbf{r}'_2 \\ \cdots \\ \mathbf{r}'_{k-1} \\ \mathbf{r}'_k \\ \cdots \\ \mathbf{r}'_{n-1} \\ \mathbf{r}'_n \end{bmatrix} \in \mathrm{SO}(n)$$

is known as *the Helmert matrix* of order n. (Alternatively the transposes of such a matrix are called *the Helmert matrix*.)

Example *(Helmert matrix of order 3):*

$$\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{bmatrix} \in \mathrm{SO}(3).$$

Check that the rows are orthogonal and normalized.

Example *(Helmert matrix of order 4):*

$$\begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0 \\ 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & -3/\sqrt{12} \end{bmatrix} \in \mathrm{SO}(4).$$

Check that the rows are orthogonal and normalized.

Example *(Helmert matrix of order n)*:

$$\begin{bmatrix} 1/\sqrt{n} & 1/\sqrt{n} & 1/\sqrt{n} & 1/\sqrt{n} & \cdots & 1/\sqrt{n} & 1/\sqrt{n} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 & \cdots & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0 & \cdots & 0 & 0 \\ \cdots & & & & \cdots & & \\ \frac{1}{\sqrt{(n-1)(n-2)}} & \frac{1}{\sqrt{(n-1)(n-2)}} & \frac{1}{\sqrt{(n-1)(n-2)}} & \cdots & \cdots & \frac{1-(n-1)}{\sqrt{(n-1)(n-2)}} & 0 \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \cdots & \frac{1}{\sqrt{n(n-1)}} & \frac{1-n}{\sqrt{n(n-1)}} \end{bmatrix} \in \mathrm{SO}(n).$$

Check that the rows are orthogonal and normalized. An example is the nth row

$$\frac{1}{n(n-1)} + \cdots + \frac{1}{n(n-1)} + \frac{(1-n)^2}{n(n-1)} = \frac{n-1}{n(n-1)} + \frac{1-2n+n^2}{n(n-1)} =$$

$$= \frac{n^2-n}{n(n-1)} = \frac{n(n-1)}{n(n-1)} = 1,$$

where (n-1) terms 1/[n(n-1)] have to be summed.

Definition (orthogonal matrix) :

A rectangular matrix $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n \times m}$ is called

"a Hankel matrix" if the n+m-1 distinct elements of $\mathbf{A}$,

$$\begin{bmatrix} a_{11} & & & \\ a_{21} & & & \\ \cdots & & & \\ a_{n-11} & & & \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

only appear in the *first column* and *last row*.

Example: *Hankel matrix of power sums*

Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ be a $n \times m$ rectangular matrix ($n \le m$) whose entries are *power sums.*

$$\mathbf{A} := \begin{bmatrix} \sum_{i=1}^{n} \alpha_i x_i & \sum_{i=1}^{n} \alpha_i x_i^2 & \cdots & \sum_{i=1}^{n} \alpha_i x_i^m \\ \sum_{i=1}^{n} \alpha_i x_i^2 & \sum_{i=1}^{n} \alpha_i x_i^3 & \cdots & \sum_{i=1}^{n} \alpha_i x_i^{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{n} \alpha_i x_i^n & \sum_{i=1}^{n} \alpha_i x_i^{n+1} & \cdots & \sum_{i=1}^{n} \alpha_i x_i^{n+m-1} \end{bmatrix}$$

$\mathbf{A}$ is a *Hankel matrix.*

Definition *(Vandermonde matrix):*

Vandermonde matrix: $\mathbf{V} \in \mathbb{R}^{n \times n}$

$$\mathbf{V} := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix},$$

$$\det \mathbf{V} = \prod_{\substack{i,j \\ i>j}}^{n} (x_i - x_j).$$

Example**:** Vandermonde matrix $\mathbf{V} \in \mathbb{R}^{3 \times 3}$

$$\mathbf{V} := \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ x_1^2 & x_2^2 & x_3^2 \end{bmatrix}, \ \det \mathbf{V} = (x_2 - x_1)(x_3 - x_2)(x_3 - x_1).$$

Example**:** Submatrix of a Hankel matrix of power sums

Consider the submatrix $\mathbf{P} = [a_1, a_2, \cdots, a_n]$ of the *Hankel matrix* $\mathbf{A} \in \mathbb{R}^{n \times m}$ ($n \le m$) whose entries are *power sums*. The determinant of the power sums matrix $\mathbf{P}$ is

$$\det \mathbf{P} = (\prod_{i=1}^{n} \alpha_i)(\prod_{i=1}^{n} x_i)(\det \mathbf{V})^2 ,$$

where det **V** is the *Vandermonde determinant.*

***Example*: Submatrix $\mathbf{P} \in \mathbb{R}^{3\times 3}$ of a 3×4 *Hankel* matrix of power sums (*n=3,m=4*)**

$$\mathbf{A} =$$

$$\begin{bmatrix} \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 & \alpha_1 x_1^2 + \alpha_2 x_2^2 + \alpha_3 x_3^2 & \alpha_1 x_1^3 + \alpha_2 x_2^3 + \alpha_3 x_3^3 & \alpha_1 x_1^4 + \alpha_2 x_2^4 + \alpha_3 x_3^4 \\ \alpha_1 x_1^2 + \alpha_2 x_2^2 + \alpha_3 x_3^2 & \alpha_1 x_1^3 + \alpha_2 x_2^3 + \alpha_3 x_3^3 & \alpha_1 x_1^4 + \alpha_2 x_2^4 + \alpha_3 x_3^4 & \alpha_1 x_1^5 + \alpha_2 x_2^5 + \alpha_3 x_3^5 \\ \alpha_1 x_1^3 + \alpha_2 x_2^3 + \alpha_3 x_3^3 & \alpha_1 x_1^4 + \alpha_2 x_2^4 + \alpha_3 x_3^4 & \alpha_1 x_1^5 + \alpha_2 x_2^5 + \alpha_3 x_3^5 & \alpha_1 x_1^6 + \alpha_2 x_2^6 + \alpha_3 x_3^6 \end{bmatrix}$$

$$\mathbf{P} = [a_1, a_2, a_3]$$

$$\begin{bmatrix} \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 & \alpha_1 x_1^2 + \alpha_2 x_2^2 + \alpha_3 x_3^2 & \alpha_1 x_1^3 + \alpha_2 x_2^3 + \alpha_3 x_3^3 \\ \alpha_1 x_1^2 + \alpha_2 x_2^2 + \alpha_3 x_3^2 & \alpha_1 x_1^3 + \alpha_2 x_2^3 + \alpha_3 x_3^3 & \alpha_1 x_1^4 + \alpha_2 x_2^4 + \alpha_3 x_3^4 \\ \alpha_1 x_1^3 + \alpha_2 x_2^3 + \alpha_3 x_3^3 & \alpha_1 x_1^4 + \alpha_2 x_2^4 + \alpha_3 x_3^4 & \alpha_1 x_1^5 + \alpha_2 x_2^5 + \alpha_3 x_3^5 \end{bmatrix}.$$

Definitions (*positive definite and positive semidefinite matrices*)

A matrix $\mathbf{A}$ is called *positive definite*, if and only if

$$\mathbf{x}'\mathbf{A}\mathbf{x} > \mathbf{0} \quad \forall \mathbf{x} \in \mathbb{R}^n, \ \mathbf{x} \neq \mathbf{0}.$$

A matrix $\mathbf{A}$ is called *positive semidefinite*, if and only if

$$\mathbf{x}'\mathbf{A}\mathbf{x} \geq \mathbf{0} \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

An example follows.

Example (*idempotence*):

All *idempotent matrices* are *positive semidefinite*, at the time $\mathbf{B}'\mathbf{B}$ and $\mathbf{B}\mathbf{B}'$ for an arbitrary matrix $\mathbf{B}$.

What are *"permutation matrices"* or *"commutation matrices"*? After their definitions we will give some applications.

Definitions (*permutation matrix, commutation matrix*)

A matrix is called a *permutation matrix* if and only if *each column* of the matrix $\mathbf{A}$ and *each row* of $\mathbf{A}$ has only one element 1. All other elements are zero. There holds $\mathbf{A}\mathbf{A}' = \mathbf{I}$.

A matrix is called a *commutation matrix*, if and only if for a matrix of the order $n^2 \times n^2$ there holds

$$\mathbf{K} = \mathbf{K}' \text{ and } \mathbf{K}^2 = \mathbf{I}_{n^2}.$$

The *commutation matrix* is *symmetric* and *orthonormal.*

Example (commutation matrix)

$$n = 2 \Rightarrow \qquad \mathbf{K}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \mathbf{K}_4'.$$

A general definition of matrices $\mathbf{K}_{nm}$ of the order $nm \times nm$ with $n \neq m$ are to found in *J.R. Magnus and H. Neudecker* (1988 p.46-48). This definition does not lead to a symmetric matrix anymore. Nevertheless is the *transpose commutation matrix* again a *commutation matrix* since we have $\mathbf{K}_{nm}' = \mathbf{K}_{nm}$ and $\mathbf{K}_{nm}\mathbf{K}_{mn} = \mathbf{I}_{nm}$.

Example (commutation matrix)

$$\left.\begin{array}{r} n = 2 \\ m = 3 \end{array}\right] \Rightarrow \qquad \mathbf{K}_{2\cdot 3} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\left.\begin{array}{r} n = 3 \\ m = 2 \end{array}\right] \Rightarrow \qquad \mathbf{K}_{3\cdot 2} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{K}_{3\cdot 2}\,\mathbf{K}_{2\cdot 3} = \mathbf{I}_6 = \mathbf{K}_{2\cdot 3}\,\mathbf{K}_{3\cdot 2} \,.$$

A3 Scalar Measures and Inverse Matrices

We will refer to some *scalar measures*, also called *scalar functions*, *of matrices*. Beforehand we will introduce some classical definitions of type

- linear independence
- column and row rank
- rank identities.

Definitions (linear independence, column and row rank):

A set of vectors $\mathbf{x}_1, ..., \mathbf{x}_n$ is called *linear independent* if for an arbitrary linear combination $\sum_{i=1}^{n} \alpha_i \mathbf{x}_i = 0$ only holds if *all scalars* $\alpha_1, ..., \alpha_n$ disappear, that is if $\alpha_1 = \alpha_2 = ... = \alpha_{n-1} = \alpha_n = 0$ holds.

For all vectors which are characterized by $\mathbf{x}_1, ..., \mathbf{x}_n$ *unequal from zero* are called linear dependent.

Let $\mathbf{A}$ be a rectangular matrix of the order $O(\mathbf{A}) = n \times m$. The *column rank* of the matrix $\mathbf{A}$ is the largest number of linear *independent columns*, while the row rank is the largest number of *linear independent rows*. Actually the *column rank* of the matrix $\mathbf{A}$ is identical to its *row rank*. The *rank of a matrix* thus is called

$$\operatorname{rk} \mathbf{A} .$$

Obviously,

$$\operatorname{rk} \mathbf{A} \le \min\{n, m\}.$$

If $\operatorname{rk} \mathbf{A} = n$ holds, we say that the matrix $\mathbf{A}$ has full *row ranks*. In contrast if the *rank identity* $\operatorname{rk} \mathbf{A} = m$ holds, we say that the matrix $\mathbf{A}$ has *full column rank*.

We list the following important *rank identities*.

Facts (rank identities):

(i) $$\operatorname{rk} \mathbf{A} = \operatorname{rk} \mathbf{A}' = \operatorname{rk} \mathbf{A}'\mathbf{A} = \operatorname{rk} \mathbf{A}\mathbf{A}'$$

(ii) $$\operatorname{rk}(\mathbf{A} + \mathbf{B}) \le \operatorname{rk} \mathbf{A} + \operatorname{rk} \mathbf{B}$$

(iii) $$\operatorname{rk}(\mathbf{A} \cdot \mathbf{B}) \le \min\{\operatorname{rk} \mathbf{A}, \operatorname{rk} \mathbf{B}\}$$

(iv) $$\operatorname{rk}(\mathbf{A} \cdot \mathbf{B}) = \operatorname{rk} \mathbf{A} \text{ if } \mathbf{B} \text{ has full } \textit{row rank},$$

(v) $$\operatorname{rk}(\mathbf{A} \cdot \mathbf{B}) = \operatorname{rk} \mathbf{B} \text{ if } \mathbf{A} \text{ has full column rank.}$$

(vi) $$\operatorname{rk}(\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{C}) + \operatorname{rk} \mathbf{B} \ge \operatorname{rk}(\mathbf{A} \cdot \mathbf{B}) + \operatorname{rk}(\mathbf{B} \cdot \mathbf{C})$$

(vii) $$\operatorname{rk}(\mathbf{A} \otimes \mathbf{B}) = (\operatorname{rk} \mathbf{A}) \cdot (\operatorname{rk} \mathbf{B}).$$

If a rectangular matrix of the order $O(\mathbf{A}) = n \times m$ is fulfilled and, *in addition*, $\mathbf{A}\mathbf{x} = \mathbf{0}$ holds for a certain vector $\mathbf{x} \ne \mathbf{0}$, then

$$\operatorname{rk} \mathbf{A} \le m - 1 .$$

Let us define what is a *rank factorization, the column space*, a *singular matrix* and, especially, what is *division algebra*.

Facts (rank factorization)

We call a *rank factorization*

$$\mathbf{A} = \mathbf{G} \cdot \mathbf{F} ,$$

if $\operatorname{rk} \mathbf{A} = \operatorname{rk} \mathbf{G} = \operatorname{rk} \mathbf{F}$ holds for certain matrices $\mathbf{G}$ and $\mathbf{F}$ of the order

$$O(\mathbf{G}) = n \times \operatorname{rk} \mathbf{A} \quad \textit{and} \quad O(\mathbf{F}) = \operatorname{rk} \mathbf{A} \times m.$$

Facts

A matrix $\mathbf{A}$ has the *column space*

$$\mathcal{R}(\mathbf{A})$$

formed by the *column vectors*. The dimension of such a vector space is dim $\mathcal{R}(\mathbf{A}) = \operatorname{rk}\mathbf{A}$. In particular,

$$\mathcal{R}(\mathbf{A}) = \mathcal{R}(\mathbf{A}\mathbf{A}')$$

holds.

Definition (non-singular matrix versus singular matrix)

Let a quadratic matrix $\mathbf{A}$ of the order $O(\mathbf{A})$ be given. $\mathbf{A}$ is called *non-singular* or regular if $\operatorname{rk}\mathbf{A} = n$ holds. In case $\operatorname{rk}\mathbf{A} < n$, the matrix $\mathbf{A}$ is called *singular*.

Definition (division algebra):

Let the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ be quadratic and non-singular of the order $O(\mathbf{A}) = O(\mathbf{B}) = O(\mathbf{C}) = n \times n$. In terms of the *Cayley-product* an *inner relation* can be based on

$$\mathbf{A} = [a_{ij}],\ \mathbf{B} = [b_{ij}],\ \mathbf{C} = [c_{ij}],\ O(\mathbf{A}) = O(\mathbf{B}) = O(\mathbf{C}) = n \times n$$

(i) $(\mathbf{A} \cdot \mathbf{B}) \cdot \mathbf{C} = \mathbf{A} \cdot (\mathbf{B} \cdot \mathbf{C})$ (associativity)

(ii) $\mathbf{A} \cdot \mathbf{I} = \mathbf{A}$ (identity)

(iii) $\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{I}$ (inverse).

The non-singular matrix $\mathbf{A}^{-1} = \mathbf{B}$ is called *Cayley-inverse*. The conditions

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n \Leftrightarrow \mathbf{B} \cdot \mathbf{A} = \mathbf{I}_n$$

are *equivalent*. The *Cayley-inverse* $\mathbf{A}^{-1}$ is left and right identical. The Cayley-inverse is unique.

Fact: $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$: $\mathbf{A}$ is symmetric $\Leftrightarrow$ $\mathbf{A}^{-1}$ is symmetric.

Facts: (Inverse Partitional Matrix /IPM/ of a symmetric matrix):

Let the *symmetric* matrix $\mathbf{A}$ be partitioned as

$$\mathbf{A} := \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix},\ \mathbf{A}'_{11} = \mathbf{A}_{11},\ \mathbf{A}'_{22} = \mathbf{A}_{22}.$$

Then its *Cayley inverse* $\mathbf{A}^{-1}$ is symmetric and can be partitioned as well as

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix}^{-1} =$$

$$\begin{bmatrix} [\mathbf{I} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}'_{12}]\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} \\ -(\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}'_{12}\mathbf{A}_{11}^{-1} & (\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} \end{bmatrix},$$

$$\textit{if } \mathbf{A}_{11}^{-1} \textit{ exists},$$

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix}^{-1} =$$

$$\begin{bmatrix} (\mathbf{A}_{11} - \mathbf{A}'_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12})^{-1} & -(\mathbf{A}_{11} - \mathbf{A}'_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}'_{12}(\mathbf{A}_{11} - \mathbf{A}'_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12})^{-1} & [\mathbf{I} + \mathbf{A}_{22}^{-1}\mathbf{A}'_{12}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}'_{12})^{-1}\mathbf{A}_{12}]\mathbf{A}_{22}^{-1} \end{bmatrix},$$

$$\textit{if } \mathbf{A}_{22}^{-1} \textit{ exists}.$$

$$\mathbf{S}_{11} := \mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \quad \textit{and} \quad \mathbf{S}_{22} := \mathbf{A}_{11} - \mathbf{A}'_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12}$$

are the minors determined by properly chosen rows and columns of the matrix **A** called “*Schur complements*” such that

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix}^{-1} =$$

$$\begin{bmatrix} (\mathbf{I} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{S}_{11}^{-1}\mathbf{A}'_{12})\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{S}_{11}^{-1} \\ -\mathbf{S}_{11}^{-1}\mathbf{A}'_{12}\mathbf{A}_{11}^{-1} & \mathbf{S}_{11}^{-1} \end{bmatrix}$$

$$\textit{if } \mathbf{A}_{11}^{-1} \textit{ exists},$$

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix}^{-1} =$$

$$\begin{bmatrix} \mathbf{S}_{22}^{-1} & -\mathbf{S}_{22}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}'_{12}\mathbf{S}_{22}^{-1} & [\mathbf{I} + \mathbf{A}_{22}^{-1}\mathbf{A}'_{12}\mathbf{S}_{22}^{-1}\mathbf{A}_{12}]\mathbf{A}_{22}^{-1} \end{bmatrix}$$

$$\textit{if } \mathbf{A}_{22}^{-1} \textit{ exists},$$

are representations of the *Cayley inverse* partitioned matrix $\mathbf{A}^{-1}$ in terms of “*Schur complements*”.

The formulae $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$ were first used by *J. Schur* (1917). The term "*Schur complements*" was introduced by *E. Haynsworth* (1968). *A. Albert* (1969) replaced the *Cayley inverse* $\mathbf{A}^{-1}$ by the *Moore-Penrose inverse* $\mathbf{A}^{+}$. For a survey we recommend *R. W. Cottle* (1974), *D.V. Oullette* (1981) and *D. Carlson* (1986).

***:Proof*:**

For the proof of the "inverse partitioned matrix" $\mathbf{A}^{-1}$ (*Cayley inverse*) of the partitioned matrix $\mathbf{A}$ of full rank we apply *Gauss elimination* (without pivoting).

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix}, \ \mathbf{A}'_{11} = \mathbf{A}_{11}, \ \mathbf{A}'_{22} = \mathbf{A}_{22}$$

$$\left[\begin{array}{ll} \mathbf{A}_{11} \in \mathbb{R}^{m\times m}, & \mathbf{A}_{12} \in \mathbb{R}^{m\times l} \\ \mathbf{A}'_{12} \in \mathbb{R}^{l\times m}, & \mathbf{A}_{22} \in \mathbb{R}^{l\times l} \end{array}\right.$$

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}'_{12} & \mathbf{B}_{22} \end{bmatrix}, \ \mathbf{B}'_{11} = \mathbf{B}_{11}, \ \mathbf{B}'_{22} = \mathbf{B}_{22}$$

$$\left[\begin{array}{ll} \mathbf{B}_{11} \in \mathbb{R}^{m\times m}, & \mathbf{B}_{12} \in \mathbb{R}^{m\times l} \\ \mathbf{B}'_{12} \in \mathbb{R}^{l\times m}, & \mathbf{B}_{22} \in \mathbb{R}^{l\times l} \end{array}\right.$$

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \quad \Leftrightarrow$$

$$\left[\begin{array}{lr} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}'_{12} = \mathbf{B}_{11}\mathbf{A}_{11} + \mathbf{B}_{12}\mathbf{A}'_{12} = \mathbf{I}_{\mathrm{m}} & (1) \\ \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} = \mathbf{B}_{11}\mathbf{A}_{12} + \mathbf{B}_{12}\mathbf{A}_{22} = \mathbf{0} & (2) \\ \mathbf{A}'_{12}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}'_{12} = \mathbf{B}'_{12}\mathbf{A}_{11} + \mathbf{B}_{22}\mathbf{A}'_{12} = \mathbf{0} & (3) \\ \mathbf{A}'_{12}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} = \mathbf{B}'_{12}\mathbf{A}_{12} + \mathbf{B}_{22}\mathbf{A}_{22} = \mathbf{I}_{\mathrm{l}} & (4). \end{array}\right.$$

Case (i): $\mathbf{A}_{11}^{-1}$ *exists*

"forward step"

$$\left.\begin{array}{l} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}'_{12} = \mathbf{I}_m \ \text{(first left equation:} \\ \qquad\qquad \text{multiply by } -\mathbf{A}'_{12}\mathbf{A}_{11}^{-1}) \\ \mathbf{A}'_{12}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}'_{12} = \mathbf{0} \ \text{(second right equation)} \end{array}\right] \Leftrightarrow$$

$$\Leftrightarrow \left.\begin{array}{l} -\mathbf{A}'_{12}\mathbf{B}_{11} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}'_{12} = \mathbf{A}'_{12}\mathbf{A}_{11}^{-1} \\ \mathbf{A}'_{12}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}'_{12} = \mathbf{0} \end{array}\right] \Leftrightarrow$$

$$\Leftrightarrow \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}'_{12} = \mathbf{I}_m \\ (\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})\mathbf{B}'_{12} = -\mathbf{A}'_{12}\mathbf{A}_{11}^{-1} \end{bmatrix} \Rightarrow$$

$$\mathbf{B}'_{12} = -(\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}'_{12}\mathbf{A}_{11}^{-1}$$
$$\mathbf{B}'_{12} = -\mathbf{S}_{11}^{-1}\mathbf{A}'_{12}\mathbf{A}_{11}^{-1}$$

or

$$\begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ -\mathbf{A}'_{12}\mathbf{A}_{11}^{-1} & \mathbf{I}_l \end{bmatrix}\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{bmatrix}.$$

Note the "*Schur complement*" $\mathbf{S}_{11} := \mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$.

"backward step"

$$\begin{rmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}'_{12} = \mathbf{I}_m \\ \mathbf{B}'_{12} = -(\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}'_{12}\mathbf{A}_{11}^{-1} \end{rmatrix} \Rightarrow$$

$$\Rightarrow \mathbf{B}_{11} = \mathbf{A}_{11}^{-1}(\mathbf{I}_m - \mathbf{A}_{12}\mathbf{B}'_{12}) = (\mathbf{I}_m - \mathbf{B}_{12}\mathbf{A}'_{12})\mathbf{A}_{11}^{-1}$$

$$\mathbf{B}_{11} = [\mathbf{I}_m + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}'_{12}]\mathbf{A}_{11}^{-1}$$
$$\mathbf{B}_{11} = \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{S}_{11}^{-1}\mathbf{A}'_{12}\mathbf{A}_{11}^{-1}$$

$$\mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} = \mathbf{0} \text{ (second left equation)} \Rightarrow$$

$$\Rightarrow \mathbf{B}_{12} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}$$

$$\Leftrightarrow$$

$$\mathbf{B}_{22} = (\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}$$
$$\mathbf{B}_{22} = \mathbf{S}_{11}^{-1}.$$

Case (ii): $\mathbf{A}_{22}^{-1}$ *exists*

"forward step"

$$\left.\begin{array}{l} \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} = \mathbf{0} \quad \text{(third right equation)} \\ \mathbf{A}'_{12}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} = \mathbf{I}_l \quad \text{(fourth left equation:} \\ \qquad\qquad\qquad\qquad \text{multiply by } -\mathbf{A}_{12}\mathbf{A}_{22}^{-1}) \end{array}\right] \Leftrightarrow$$

$$\Leftrightarrow \left.\begin{array}{l} \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} = \mathbf{0} \\ -\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}'_{12}\mathbf{B}_{12} - \mathbf{A}_{12}\mathbf{B}_{22} = -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{array}\right] \Leftrightarrow$$

$$\Leftrightarrow \begin{bmatrix} \mathbf{A}'_{12}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} = \mathbf{I}_l \\ (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}'_{12})\mathbf{B}_{12} = -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{bmatrix} \Rightarrow$$

$$\mathbf{B}_{12} = -(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}'_{12})^{-1}\mathbf{A}'_{12}\mathbf{A}_{22}^{-1}$$
$$\mathbf{B}_{12} = -\mathbf{S}_{22}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}$$

or

$$\begin{bmatrix} \mathbf{I}_m & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_l \end{bmatrix}\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}'_{12} & \mathbf{0} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix}.$$

Note the "*Schur complement*" $\mathbf{S}_{22} := \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}'_{12}$.

"backward step"

$$\left.\begin{array}{l} \mathbf{A}'_{12}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} = \mathbf{I}_l \\ \mathbf{B}_{12} = -(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}'_{12})^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{array}\right] \Rightarrow$$

$$\Rightarrow \mathbf{B}_{22} = \mathbf{A}_{22}^{-1}(\mathbf{I}_l - \mathbf{A}'_{12}\mathbf{B}'_{12}) = (\mathbf{I}_l - \mathbf{B}'_{12}\mathbf{A}_{12})\mathbf{A}_{22}^{-1}$$

$$\mathbf{B}_{22} = [\mathbf{I}_l + \mathbf{A}_{22}^{-1}\mathbf{A}'_{12}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}'_{12})^{-1}\mathbf{A}_{12}]\mathbf{A}_{22}^{-1}$$
$$\mathbf{B}_{22} = \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}'_{12}\mathbf{S}_{22}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}$$

$$\mathbf{A}'_{12}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}'_{12} = \mathbf{0} \text{ (third left equation)} \Rightarrow$$

$$\Rightarrow \mathbf{B}'_{12} = -\mathbf{A}_{22}^{-1}\mathbf{A}'_{12}\mathbf{B}_{11} = -\mathbf{A}_{22}^{-1}\mathbf{A}'_{12}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}'_{12})^{-1}$$

$$\Leftrightarrow$$

$$\mathbf{B}_{11} = (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}'_{12})^{-1}$$
$$\mathbf{B}_{11} = \mathbf{S}_{22}^{-1}.$$

♣

The representations { $\mathbf{B}_{11}$, $\mathbf{B}_{12}$, $\mathbf{B}_{21} = \mathbf{B}'_{12}$, $\mathbf{B}_{22}$ } in terms of { $\mathbf{A}_{11}$, $\mathbf{A}_{12}$, $\mathbf{A}_{21} = \mathbf{A}'_{12}$, $\mathbf{A}_{22}$ } have been derived by *T. Banachiewicz* (1937). Generalizations are referred to *T. Ando* (1979), *R. A. Brunaldi and H. Schneider* (1963), *F. Burns, D. Carlson, E. Haynsworth and T. Markham* (1974), *D. Carlson* (1980), *C. D. Meyer* (1973) and *S. K. Mitra* (1982), *C. K. Li and R. Mathias* (2000).

We leave the proof of the following fact as an exercise.

Fact (Inverse Partitioned Matrix /IPM/ of a quadratic matrix):

Let the quadratic matrix $\mathbf{A}$ be partitioned as

$$\mathbf{A} := \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}.$$

Then its *Cayley inverse* $\mathbf{A}^{-1}$ can be partitioned as well as

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} =$$

$$\begin{bmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{S}_{11}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{S}_{11}^{-1} \\ -\mathbf{S}_{11}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{S}_{11}^{-1} \end{bmatrix},$$

if $\mathbf{A}_{11}^{-1}$ *exists*

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} =$$

$$\begin{bmatrix} \mathbf{S}_{22}^{-1} & -\mathbf{S}_{22}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{S}_{22}^{-1} & \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{S}_{22}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{bmatrix},$$

if $\mathbf{A}_{22}^{-1}$ *exists*

and the "*Schur complements*" are definded by

$$\mathbf{S}_{11} := \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \ \ \textit{and} \ \ \mathbf{S}_{22} := \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \ .$$

Facts: (*Cayley inverse*: sum of two matrices):

(s1) $(\mathbf{A}+\mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}\mathbf{A}^{-1}$

(s2) $(\mathbf{A}-\mathbf{B})^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}(\mathbf{A}^{-1}-\mathbf{B}^{-1})^{-1}\mathbf{A}^{-1}$

(s3) $(\mathbf{A}+\mathbf{CBD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}(\mathbf{I}+\mathbf{CBDA}^{-1})^{-1}\mathbf{CBDA}^{-1}$

(s4) $(\mathbf{A}+\mathbf{CBD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}(\mathbf{I}+\mathbf{BDA}^{-1}\mathbf{C})^{-1}\mathbf{BDA}^{-1}$

(s5) $(\mathbf{A}+\mathbf{CBD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{CB}(\mathbf{I}+\mathbf{DA}^{-1}\mathbf{CB})^{-1}\mathbf{DA}^{-1}$

(s6) $(\mathbf{A}+\mathbf{CBD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{CBD}(\mathbf{I}+\mathbf{A}^{-1}\mathbf{CBD})^{-1}\mathbf{A}^{-1}$

(s7) $(\mathbf{A}+\mathbf{CBD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{CBDA}^{-1}(\mathbf{I}+\mathbf{CBDA}^{-1})^{-1}$

(s8) $(\mathbf{A}+\mathbf{CBD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1}+\mathbf{DA}^{-1}\mathbf{C})^{-1}\mathbf{DA}^{-1}$

(Sherman-Morrison-Woodbury matrix identity)

(s9) $\mathbf{B}(\mathbf{AB}+\mathbf{C})^{-1} = (\mathbf{I}+\mathbf{BC}^{-1}\mathbf{A})^{-1}\mathbf{BC}^{-1}$

(s10) $\mathbf{BD}(\mathbf{A}+\mathbf{CBD})^{-1} = (\mathbf{B}^{-1}+\mathbf{DA}^{-1}\mathbf{C})^{-1}\mathbf{DA}^{-1}$

(Duncan-Guttman matrix identity).

W. J. Duncan (1944) calls (s8) the *Sherman-Morrison-Woodbury matrix identity*. If the matrix **A** is singular consult *H. V. Henderson and G. S. Searle* (1981), *D. V. Ouellette* (1981), *W. M. Hager* (1989), *G. W. Stewart* (1977) and *K. S. Riedel*

(1992). (s10) has been noted by *W. J. Duncan* (1944) and *L. Guttman* (1946): The result is directly derived from the identity

$$(\mathbf{A}+\mathbf{CBD})(\mathbf{A}+\mathbf{CBD})^{-1}=\mathbf{I} \quad \Rightarrow$$

$$\Rightarrow \ \mathbf{A}(\mathbf{A}+\mathbf{CBD})^{-1}+\mathbf{CBD}(\mathbf{A}+\mathbf{CBD})^{-1}=\mathbf{I}$$
$$(\mathbf{A}+\mathbf{CBD})^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{CBD}(\mathbf{A}+\mathbf{CBD})^{-1}$$

$$\mathbf{A}^{-1}=(\mathbf{A}+\mathbf{CBD})^{-1}+\mathbf{A}^{-1}\mathbf{CBD}(\mathbf{A}+\mathbf{CBD})^{-1}$$
$$\mathbf{DA}^{-1}=\mathbf{D}(\mathbf{A}+\mathbf{CBD})^{-1}+\mathbf{DA}^{-1}\mathbf{CBD}(\mathbf{A}+\mathbf{CBD})^{-1}$$
$$\mathbf{DA}^{-1}=(\mathbf{I}+\mathbf{DA}^{-1}\mathbf{CB})\mathbf{D}(\mathbf{A}+\mathbf{CBD})^{-1}$$
$$\mathbf{DA}^{-1}=(\mathbf{B}^{-1}+\mathbf{DA}^{-1}\mathbf{C})\mathbf{BD}(\mathbf{A}+\mathbf{CBD})^{-1}$$
$$(\mathbf{B}^{-1}+\mathbf{DA}^{-1}\mathbf{C})^{-1}\mathbf{DA}^{-1}=\mathbf{BD}(\mathbf{A}+\mathbf{CBD})^{-1}.$$

♣

Certain results follow directly from their definitions.

Facts (inverses):

(i) $(\mathbf{A}\cdot\mathbf{B})^{-1}=\mathbf{B}^{-1}\cdot\mathbf{A}^{-1}$

(ii) $(\mathbf{A}\otimes\mathbf{B})^{-1}=\mathbf{B}^{-1}\otimes\mathbf{A}^{-1}$

(iii) $\mathbf{A}$ *positive definite* $\Leftrightarrow$ $\mathbf{A}^{-1}$ *positive definite*

(iv) $(\mathbf{A}\otimes\mathbf{B})^{-1}$, $(\mathbf{A}*\mathbf{B})^{-1}$ and $(\mathbf{A}^{-1}*\mathbf{B}^{-1})$ are *positive definite*, then $(\mathbf{A}^{-1}*\mathbf{B}^{-1})-(\mathbf{A}*\mathbf{B})^{-1}$ is *positive semidefinite* as well as $(\mathbf{A}^{-1}*\mathbf{A})-\mathbf{I}$ and $\mathbf{I}-(\mathbf{A}^{-1}*\mathbf{A})^{-1}$.

Facts (rank factorization):

(i) If the $n\times n$ matrix is *symmetric* and *positive semidefinite*, then its *rank factorization* is

$$\mathbf{A}=\begin{bmatrix}\mathbf{G}_1\\ \mathbf{G}_2\end{bmatrix}\begin{bmatrix}\mathbf{G}_1' & \mathbf{G}_2'\end{bmatrix},$$

where $\mathbf{G}_1$ is a *lower triangular matrix* of the order $O(\mathbf{G}_1)=\operatorname{rk}\mathbf{A}\times\operatorname{rk}\mathbf{A}$ with

$$\operatorname{rk}\mathbf{G}_2=\operatorname{rk}\mathbf{A},$$

whereas $\mathbf{G}_2$ has the format $O(\mathbf{G}_2)=(n-\operatorname{rk}\mathbf{A})\times\operatorname{rk}\mathbf{A}$. In this case we speak of a *Choleski decomposition*.

(ii) In case that the matrix $\mathbf{A}$ is *positive definite, the matrix block* $\mathbf{G}_2$ is not needed anymore: $\mathbf{G}_1$ is *uniquely determined.* There holds

$$\mathbf{A}^{-1} = (\mathbf{G}_1^{-1})'\mathbf{G}_1^{-1}.$$

Beside the *rank* of a quadratic matrix $\mathbf{A}$ of the order $O(\mathbf{A}) = n \times n$ as the *first scalar measure* of a matrix, is its *determinant*

$$|\mathbf{A}| = \sum_{\substack{perm \\ (j_1,\ldots,j_n)}} (-1)^{\Phi(j_1,\ldots,j_n)} \prod_{i=1}^{n} a_{ij_i}$$

plays a similar role as a *second scalar measure*. Here the summation is extended as the summation perm $(j_1,\ldots,j_n)$ over all permutations $(j_1,\ldots,j_n)$ of the set of integer numbers $(1,\ldots,n)$. $\Phi(j_1,\ldots,j_n)$ is the number of permutations which transform $(1,\ldots,n)$ into $(j_1,\ldots,j_n)$.

Laws (determinant)

(i) $|\alpha \cdot \mathbf{A}| = \alpha^n \cdot |\mathbf{A}|$ for an arbitrary scalar $\alpha \in \mathbb{R}$

(ii) $|\mathbf{A} \cdot \mathbf{B}| = |\mathbf{A}| \cdot |\mathbf{B}|$

(iii) $|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^m \cdot |\mathbf{B}|^n$ for an arbitrary $m \times n$ matrix $\mathbf{B}$

(iv) $|\mathbf{A}'| = |\mathbf{A}|$

(v) $|\frac{1}{2}(\mathbf{A} + \mathbf{A}')| \leq |\mathbf{A}|$ if $\mathbf{A} + \mathbf{A}'$ is positive definite

(vi) $|\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}$ if $\mathbf{A}^{-1}$ exists

(vii) $|\mathbf{A}| = 0 \Leftrightarrow \mathbf{A}$ is singular ($\mathbf{A}^{-1}$ does *not* exist)

(viii) $|\mathbf{A}| = 0$ if $\mathbf{A}$ is idempotent, $\mathbf{A} \neq \mathbf{I}$

(ix) $|\mathbf{A}| = \prod_{i=1}^{n} a_{ii}$ if $\mathbf{A}$ is diagonal and a triangular matrix

(x) $0 \leq |\mathbf{A}| \leq \prod_{i=1}^{n} a_{ii} = |\mathbf{A} * \mathbf{I}|$ if $\mathbf{A}$ is positive definite

(xi) $|\mathbf{A}| \cdot |\mathbf{B}| \leq |\mathbf{A}| \cdot \prod_{i=1}^{n} b_{ii} \leq |\mathbf{A} * \mathbf{B}|$ if $\mathbf{A}$ and $\mathbf{B}$ are positive definite

(xii)
$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \begin{cases} \det \mathbf{A}_{11} \det(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}) \\ \quad \forall \mathbf{A}_{11} \in \mathbb{R}^{m_1 \times m_1}, \ \operatorname{rk} \mathbf{A}_{11} = m_1 \\ \det \mathbf{A}_{21} \det(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}) \\ \quad \forall \mathbf{A}_{22} \in \mathbb{R}^{m_2 \times m_2}, \ \operatorname{rk} \mathbf{A}_{22} = m_2 . \end{cases}$$

A *submatrix* of a rectangular matrix **A** is the result of a canceling procedure of *certain rows and columns* of the matrix **A**. A *minor* is the determinant of a *quadratic submatrix* of the matrix **A**. If the matrix **A** is a quadratic matrix, to any element a_{ij} there exists a *minor* being the determinant of a submatrix of the matrix **A** which is the result of *reducing the **i**-th row and the **j**-th column*. By multiplying with $(-1)^{i+j}$ we gain a new element c_{ij} of a matrix $\mathbf{C} = [\mathbf{c}_{ij}]$. The transpose matrix $\mathbf{C}'$ is called the *adjoint matrix* of the matrix **A**, written *adj***A**. Its order is the same as of the matrix **A**.

Laws (adjoint matrix)

(i) $|\mathbf{A}| = \sum_{j=1}^{n} a_{ij} c_{ij}, \quad \forall i = 1, \ldots, n$

(ii) $|\mathbf{A}| = \sum_{j=1}^{n} a_{jk} c_{jk}, \quad \forall k = 1, \ldots, n$

(iii) $\mathbf{A} \cdot (\mathrm{adj}\,\mathbf{A}) = (\mathrm{adj}\,\mathbf{A}) \cdot \mathbf{A} = |\mathbf{A}| \cdot \mathbf{I}$

(iv) $\mathrm{adj}(\mathbf{A} \cdot \mathbf{B}) = (\mathrm{adj}\,\mathbf{B}) \cdot (\mathrm{adj}\,\mathbf{A})$

(v) $\mathrm{adj}(\mathbf{A} \otimes \mathbf{B}) = (\mathrm{adj}\,\mathbf{A}) \otimes (\mathrm{adj}\,\mathbf{B})$

(vi) $\mathrm{adj}\,\mathbf{A} = |\mathbf{A}| \cdot \mathbf{A}^{-1}$ if $\mathbf{A}$ is nonsingular

(vii) $\mathrm{adj}\,\mathbf{A}$ positive definitive $\Leftrightarrow$ $\mathbf{A}$ positive definite.

As a *third scalar measure* of a quadratic matrix **A** of the order $O(\mathbf{A}) = n \times n$ we introduce the *trace* $\mathrm{tr}\,\mathbf{A}$ as the *sum of diagonal elements*,

$$\mathrm{tr}\,\mathbf{A} = \sum_{i=1}^{n} a_{ii}.$$

Laws (trace of a matrix)

(i) $\mathrm{tr}(\alpha \cdot \mathbf{A}) = \alpha \cdot \mathrm{tr}\,\mathbf{A}$ for an arbitrary scalar $\alpha \in \mathbb{R}$

(ii) $\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}\,\mathbf{A} + \mathrm{tr}\,\mathbf{B}$ for an arbitrary $n \times n$ matrix **B**

(iii) $\mathrm{tr}(\mathbf{A} \otimes \mathbf{B}) = (\mathrm{tr}\,\mathbf{A}) \cdot (\mathrm{tr}\,\mathbf{B})$ for an arbitrary $m \times m$ matrix **B**

iv) $\mathrm{tr}\,\mathbf{A} = \mathrm{tr}(\mathbf{B} \cdot \mathbf{C})$ for any factorization $\mathbf{A} = \mathbf{B} \cdot \mathbf{C}$

(v) $\mathrm{tr}\,\mathbf{A}'(\mathbf{B} * \mathbf{C}) = \mathrm{tr}(\mathbf{A}' * \mathbf{B}')\mathbf{C}$ for an arbitrary $n \times n$ matrix **B** *and* **C**

(vi) $\mathrm{tr}\,\mathbf{A}' = \mathrm{tr}\,\mathbf{A}$

(vii) $\mathrm{tr}\mathbf{A} = \mathrm{rk}\mathbf{A}$ if **A** is *idempotent*

(viii) $0 < \mathrm{tr}\,\mathbf{A} = tr(\mathbf{A} * \mathbf{I})$ if $\mathbf{A}$ is *positive definite*

(ix) $\mathrm{tr}(\mathbf{A} * \mathbf{B}) \le (\mathrm{tr}\mathbf{A}) \cdot (\mathrm{tr}\mathbf{B})$ if **A** und **B** are *positive semidefinite.*

In correspondence to the **W** – *weighted vector (semi) – norm.*

$$\| \mathbf{x} \|_W = (\mathbf{x}' \mathbf{W} \mathbf{x})^{1/2}$$

is the **W** – *weighted matrix (semi) norm*

$$\| \mathbf{A} \|_W = (\mathrm{tr}\mathbf{A}'\mathbf{W}\mathbf{A})^{1/2}$$

for a given positive – (semi) definite matrix **W** of proper order.

Laws (trace of matrices):

(i) $\mathrm{tr}\, \mathbf{A}'\mathbf{W}\mathbf{A} \geq 0$

(ii) $\mathrm{tr}\, \mathbf{A}'\mathbf{W}\mathbf{A} = 0 \Leftrightarrow \mathbf{W}\mathbf{A} = 0$

$\Leftrightarrow \mathbf{A} = 0$ if $\mathbf{W}$ is positive definite

A4 Vector-valued Matrix Forms

If **A** is a rectangular matrix of the order $O(\mathbf{A}) = n \times m$**,** a_j its j – th column, *then* $\mathrm{vec}\, \mathbf{A}$ is an $nm \times 1$ vector

$$\mathrm{vec}\, \mathbf{A} = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_{n-1} \\ a_n \end{bmatrix}.$$

In consequence, the *operator* “vec” of a matrix transforms a vector in such a way that the *columns are stapled* one after the other.

Definitions (vec, vech, veck):

(i) $$\mathrm{vec}\, \mathbf{A} = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_{n-1} \\ a_n \end{bmatrix}.$$

(ii) Let **A** be a *quadratic symmetric matrix,* $\mathbf{A} = \mathbf{A}'$**,** of order $O(\mathbf{A}) = n \times n$. Then $\mathrm{vech}\mathbf{A}$ (“vec - *koef*”) is the $[n(n+1)/2] \times 1$ vector which is the result of row (column) *stapels* of those matrix elements which are *upper and under* of its diagonal.

$$\mathbf{A} = [a_{ij}] = [a_{ji}] = \mathbf{A}' \Rightarrow \operatorname{vech}\mathbf{A} := \begin{bmatrix} a_{11} \\ \cdots \\ a_{n1} \\ a_{22} \\ \cdots \\ a_{n2} \\ \cdots \\ a_{nn} \end{bmatrix}.$$

(iii) Let **A** be a *quadratic, antisymmetric matrix*, $\mathbf{A} = \mathbf{A}'$, of order $O(\mathbf{A}) = n \times n$. Then $\operatorname{veck}\mathbf{A}$ ("vec - *skew*") is the $[n(n+1)/2] \times 1$ vector which is generated columnwise *stapels* of those matrix elements which are *under* its diagonal.

$$\mathbf{A} = [a_{ij}] = [-a_{ji}] = -\mathbf{A}' \Rightarrow \operatorname{veck}\mathbf{A} := \begin{bmatrix} a_{11} \\ \cdots \\ a_{n1} \\ a_{32} \\ \cdots \\ a_{n2} \\ \cdots \\ a_{n,n-1} \end{bmatrix}.$$

Examples

(i) $\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \Rightarrow \operatorname{vec}\mathbf{A} = [a, d, b, e, c, f]'$

(ii) $\mathbf{A} = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} = \mathbf{A}' \Rightarrow \operatorname{vech}\mathbf{A} = [a, b, c, d, e, f]'$

(iii) $\mathbf{A} = \begin{bmatrix} 0 & -a & -b & -c \\ a & 0 & -d & -e \\ b & d & 0 & -f \\ c & e & f & 0 \end{bmatrix} = -\mathbf{A}' \Rightarrow \operatorname{veck}\mathbf{A} = [a, b, c, d, e, f]'.$

Useful identities, relating to *scalar- and vector - valued measures* of matrices will be reported finally.

Facts (vec and trace forms):

(i) $\operatorname{vec}(\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{C}') = (\mathbf{C} \otimes \mathbf{A}) \operatorname{vec}\mathbf{B}$

(ii) $\operatorname{vec}(\mathbf{A} \cdot \mathbf{B}) = (\mathbf{B}' \otimes \mathbf{I}_n) \operatorname{vec}\mathbf{A} = (\mathbf{B}' \otimes \mathbf{A}) \operatorname{vec}\mathbf{I}_m =$

$$= (\mathbf{I}_1 \otimes \mathbf{A}) \operatorname{vec}\mathbf{B}, \quad \forall \mathbf{A} \in \mathbb{R}^{n \times m}, \ \mathbf{B} \in \mathbb{R}^{m \times q}$$

(iii) $\mathbf{A}\cdot\mathbf{B}\cdot\mathbf{c} = (\mathbf{c}'\otimes\mathbf{A})\text{vec}\mathbf{B} = (\mathbf{A}\otimes\mathbf{c}')\text{vec}\mathbf{B}', \forall\mathbf{c}\in\mathbb{R}^q$

(iv) $\text{tr}(\mathbf{A}'\cdot\mathbf{B}) = (\text{vec}\mathbf{A})'\text{vec}\mathbf{B} = (\text{vec}\mathbf{A}')\text{vec}\mathbf{B}' = \text{tr}(\mathbf{A}\cdot\mathbf{B}')$

(v) $\text{tr}(\mathbf{A}\cdot\mathbf{B}\cdot\mathbf{C}'\cdot\mathbf{D}') = (\text{vec}\,\mathbf{D})'(\mathbf{C}\otimes\mathbf{A})\,\text{vec}\,\mathbf{B} =$
$$= (\text{vec}\,\mathbf{D}')'(\mathbf{A}\otimes\mathbf{C})\,\text{vec}\,\mathbf{B}'$$

(vi) $\mathbf{K}_{nm}\cdot\text{vec}\mathbf{A} = \text{vec}\mathbf{A}', \forall\mathbf{A}\in\mathbb{R}^{n\times m}$

(vii) $\mathbf{K}_{qn}(\mathbf{A}\otimes\mathbf{B}) = (\mathbf{B}\otimes\mathbf{A})\mathbf{K}_{pm}$

(viii) $\mathbf{K}_{qn}(\mathbf{A}\otimes\mathbf{B})\mathbf{K}_{mp} = (\mathbf{B}\otimes\mathbf{A})$

(ix) $\mathbf{K}_{qn}(\mathbf{A}\otimes\mathbf{c}) = \mathbf{c}\otimes\mathbf{A}$

(x) $\mathbf{K}_{nq}(\mathbf{c}\otimes\mathbf{A}) = \mathbf{A}\otimes\mathbf{c}, \quad \forall\mathbf{A}\in\mathbb{R}^{n\times m},\ \mathbf{B}\in\mathbb{R}^{q\times p}, \quad \mathbf{c}\in\mathbb{R}^q$

(xi) $\text{vec}(\mathbf{A}\otimes\mathbf{B}) = (\mathbf{I}_m\otimes\mathbf{K}_{pn}\otimes\mathbf{I}_q)(\text{vec}\mathbf{A}\otimes\text{vec}\mathbf{B})$

(xii) $\mathbf{A} = (\mathbf{a}_1,\ldots,\mathbf{a}_m), \quad \mathbf{B} := \text{Diag}\mathbf{b}, \quad O(\mathbf{B}) = m\times m,$
$$\mathbf{C}' = [\mathbf{c}_1,\ldots,\mathbf{c}_m] \Rightarrow \text{vec}(\mathbf{A}\cdot\mathbf{B}\cdot\mathbf{C}') = \text{vec}[\sum_{j=1}^{m}(\mathbf{a}_j\mathbf{b}_j\mathbf{c}'_j)] =$$
$$= \sum_{j=1}^{m}(\mathbf{c}_j\otimes\mathbf{a}_j)\mathbf{b}_j = [\mathbf{c}_1\otimes\mathbf{a}_1,\ldots,\mathbf{c}_m\otimes\mathbf{a}_m)]\mathbf{b} = (\mathbf{C}\odot\mathbf{A})\mathbf{b}$$

(xiii) $\mathbf{A} = [\mathbf{a}_{ij}], \quad \mathbf{C} = [\mathbf{c}_{ij}], \quad \mathbf{B} := \text{Diag}\mathbf{b}, \quad \mathbf{b} = [\mathbf{b}_1,\ldots,\mathbf{b}_m]\in\mathbb{R}^m$
$$\Rightarrow \text{tr}(\mathbf{A}\cdot\mathbf{B}\cdot\mathbf{C}'\cdot\mathbf{B}) = (\text{vec}\,\mathbf{B})'\,\text{vec}(\mathbf{C}\cdot\mathbf{B}\cdot\mathbf{A}') =$$
$$= \mathbf{b}'(\mathbf{I}_m\odot\mathbf{I}_m)'\cdot(\mathbf{A}\odot\mathbf{C})\mathbf{b} = \mathbf{b}'(\mathbf{A}*\mathbf{C})\mathbf{b}$$

(xiv) $\mathbf{B} := \mathbf{I}_m \Rightarrow \text{tr}(\mathbf{A}\cdot\mathbf{C}') = \mathbf{r}'_m(\mathbf{A}*\mathbf{C})\mathbf{r}_m$
($\mathbf{r}_m$ is the $m\times 1$ summation vector: $\mathbf{r}_m := [1,\ldots,1]'\in\mathbb{R}^m$)

(xv) $\text{vec Diag}\mathbf{D} := (\mathbf{I}_m*\mathbf{D})\mathbf{r}_m = [\mathbf{I}_m*(\mathbf{A}'\cdot\mathbf{B}\cdot\mathbf{C})]\mathbf{r}_m =$
$$= (\mathbf{I}_m\odot\mathbf{I}_m)' = [\mathbf{I}_m\odot(\mathbf{A}'\cdot\mathbf{B}\cdot\mathbf{C})]\cdot\text{vec Diag}\mathbf{I}_m =$$
$$= (\mathbf{I}_m\odot\mathbf{I}_m)'\cdot\text{vec}(\mathbf{A}'\cdot\mathbf{B}\cdot\mathbf{C}) =$$
$$= (\mathbf{I}_m\odot\mathbf{I}_m)'\cdot(\mathbf{C}'\otimes\mathbf{A}')\text{vec}\mathbf{B} = (\mathbf{C}\odot\mathbf{A})'\text{vec}\mathbf{B}$$
when $\mathbf{D} = \mathbf{A}'\cdot\mathbf{B}\cdot\mathbf{C}$ is factorized.

Facts (Löwner partial ordering):

For any quadratic matrix $\mathbf{A}\in\mathbb{R}^{m\times m}$ there holds the *uncertainty*
$$\mathbf{I}_m*(\mathbf{A}'\cdot\mathbf{A}) \geq \mathbf{I}_m*\mathbf{A}*\mathbf{A} = \mathbf{I}_m*[(\mathbf{A}\odot\mathbf{I}_m)'\cdot(\mathbf{I}_m\odot\mathbf{A})]$$
in the Löwner *partial ordering* that is the *difference matrix*
$$\mathbf{I}_m*(\mathbf{A}'\cdot\mathbf{A}) - \mathbf{I}_m*\mathbf{A}*\mathbf{A}$$
is at least positive semidefinite.

A5 Eigenvalues and Eigenvectors

To any *quadratic matrix* **A** of the order $O(\mathbf{A}) = m \times m$ there exists an *eigenvalue* λ as a *scalar* which makes the matrix $\mathbf{A} - \lambda \mathbf{I}_m$ *singular*. As an equivalent statement, we say that the *characteristic equation* $|\lambda \mathbf{I}_m - \mathbf{A}| = 0$ *has a zero value which could be multiple of degrees*, if s is the dimension of the related *null space* $\mathcal{N}(\mathbf{A} - \lambda \mathbf{I})$. The non-vanishing element **x** of this *null space* for which $\mathbf{Ax} = \lambda \mathbf{x},\ \mathbf{x} \neq 0$ holds, is called *right eigenvector* of **A**. Related vectors **y** for which $\mathbf{y}'\mathbf{A} = \boldsymbol{\lambda}\mathbf{y},\ \ \mathbf{y} \neq 0$, holds, are called *left eigenvectors* of **A** and are representative of the *right eigenvectors* **A'.** *Eigenvectors* always belong to a certain *eigenvalue* and are usually *normed* in the sense of $\mathbf{x}'\mathbf{x} = 1,\ \ \mathbf{y}'\mathbf{y} = 1$ as long as they have *real components*. As the same time, the *eigenvectors* which belong to *different* eigenvalues are always *linear independent*: They obviously span a *subspace* of $\mathcal{R}(\mathbf{A})$.

In general, the eigenvalues of a matrix **A** are *complex*! There is an important exception: the *orthonormal matrices,* also called *rotation matrices* whose eigenvalues are +1 or, –1 *and idempotent matrices* which can only be 0 or 1 as a *multiple eigenvalue* generally, we call a *null eigenvalue a singular matrix.*

There is the special case of a symmetric matrix $\mathbf{A} = \mathbf{A}'$ of order $O(\mathbf{A}) = m \times m$. It can be shown that all *roots of the characteristic polynomial* are *real numbers* and accordingly *m* - not necessary different - *real eigenvalues exist*. In addition, the *different eigenvalues* λ and μ and their corresponding *eigenvectors* **x** and **y** are *orthogonal*, that is

$$(\lambda - \mu)\mathbf{x}' \cdot \mathbf{y} = (\mathbf{x}' \cdot \mathbf{A}') \cdot \mathbf{y} - \mathbf{x}'(\mathbf{A} \cdot \mathbf{y}) = 0,\ \forall \lambda - \mu \neq 0.$$

In case that the *eigenvalue* λ *of degrees s* appears *s*-times, the *eigenspace* $\mathcal{N}(\mathbf{A} - \lambda \cdot \mathbf{I}_m)$ is *s - dimensional*: we can choose *s orthonormal eigenvectors* which are orthonormal to all other! In total, we can organize *m* orthonormal eigenvectors which span the entire $\mathbb{R}^m$. If we restrict ourselves to eigenvectors and to eigenvalues $\lambda,\ \lambda \neq 0$, we receive the *column space* $\mathcal{R}(\mathbf{A})$. The *rank* of **A** coincides with the *number of non-vanishing eigenvalues* $\{\lambda_1, \ldots, \lambda_r\}$.

$$\mathbf{U} := [\mathbf{U}_1, \mathbf{U}_2],\ O(\mathbf{U}) = m \times m,\ \mathbf{U} \cdot \mathbf{U}' = \mathbf{U}'\mathbf{U} = \mathbf{I}_m$$

$$\mathbf{U}_1 := [\mathbf{u}_1, \ldots, \mathbf{u}_r],\ \ O(\mathbf{U}_1) = m \times r,\ r = \mathrm{rk}\mathbf{A}$$

$$\mathbf{U}_2 := [\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m],\ O(\mathbf{U}_2) = m \times (m - r),\ \mathbf{A} \cdot \mathbf{U}_2 = 0.$$

With the definition of the $r \times r$ *diagonal matrix* $\lambda := \mathrm{Diag}(\lambda_1, \ldots \lambda_r)$ of *non-vanishing eigenvalues* we gain

$$\mathbf{A} \cdot \mathbf{U} = \mathbf{A} \cdot [\mathbf{U}_1, \mathbf{U}_2] = [\mathbf{U}_1 \Lambda, 0] = [\mathbf{U}_1, \mathbf{U}_2] \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix}.$$

With permission from the author: Extract from "Erik W. Grafarend: Linear and nonlinear models. Fixed effects, random effects, and mixed models"

Due to the *orthonormality* of the matrix $\mathbf{U} := [\mathbf{U}_1, \mathbf{U}_2]$ we achieve the results about *eigenvalue – eigenvector analysis* and *eigenvalues – eigenvector synthesis.*

Lemma (*eigenvalue – eigenvector analysis*: decomposition):

Let $\mathbf{A} = \mathbf{A}'$ be a *symmetric matrix* of the order $O(\mathbf{A}) = m \times m$. Then there exists an *orthonormal matrix* ***U*** in such a way that

$$\mathbf{U}'\mathbf{A}\mathbf{U} = \text{Diag}(\lambda_1, \ldots \lambda_r, 0, \ldots, 0)$$

holds. $(\lambda_1, \ldots \lambda_r)$ denotes the set of *non – vanishing eigenvalues* of $\mathbf{A}$ with $r = \text{rk}\mathbf{A}$ *ordered decreasingly*.

Lemma (*eigenvalue – eigenvectorsynthesis*: decomposition):

Let $\mathbf{A} = \mathbf{A}'$ be a symmetric matrix of the order $O(\mathbf{A}) = m \times m$. Then there exists a *synthetic representation* of *eigenvalues and eigenvectors* of type

$$\mathbf{A} = \mathbf{U} \cdot \text{Diag}(\lambda_1, \ldots \lambda_r, 0, \ldots, 0)\mathbf{U}' = \mathbf{U}_1 \Lambda \mathbf{U}_1'.$$

In the class of *symmetric matrices* the *positive (semi)definite* matrices play a special role. Actually, they are just the *positive (nonnegative) eigenvalues* squarerooted.

$$\Lambda^{1/2} := \text{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_r}).$$

The matrix $\mathbf{A}$ is *positive semidefinite* if and only if there exists a *quadratic* $m \times m$ *matrix* $\mathbf{G}$ such that $\mathbf{A} = \mathbf{G}\mathbf{G}'$ holds, for instance, $\mathbf{G} := [\mathbf{u}_1 \Lambda^{1/2}, 0]$. The quadratic matrix is *positive definite if and only if* the $m \times m$ matrix $\mathbf{G}$ is *not singular*. Such a representation leads to the *rank fatorization* $\mathbf{A} = \mathbf{G}_1 \cdot \mathbf{G}_1'$ with $\mathbf{G}_1 := \mathbf{U}_1 \cdot \Lambda^{1/2}$. In general, we have

Lemma (representation of the matrix $\bar{\mathbf{U}}_1$):

If $\mathbf{A}$ is a positive semidefinite matrix of the order $O(\mathbf{A})$ with *non – vanishing eigenvalues* $\{\lambda_1, \ldots, \lambda_r\}$, then there exists an $m \times r$ matrix

$$\bar{\mathbf{U}}_1 := \mathbf{G}_1 \cdot \Lambda^{-1} = \mathbf{U}_1 \cdot \Lambda^{-1/2}$$

with

$$\mathbf{U}_1' \cdot \mathbf{U}_1 = \mathbf{I}_r,\ \mathcal{R}(\mathbf{U}_1) = \mathcal{R}(\bar{\mathbf{U}}_1) = \mathcal{R}(\mathbf{A}),$$

such that

$$\bar{\mathbf{U}}_1' \cdot \mathbf{A} \cdot \bar{\mathbf{U}}_1 = (\Lambda^{-1/2} \cdot \mathbf{U}_1') \cdot (\mathbf{U}_1 \cdot \Lambda \cdot \mathbf{U}_1') \cdot (\mathbf{U}_1 \cdot \Lambda^{-1/2}) = \mathbf{I}_r.$$

The *synthetic relation* of the matrix $\mathbf{A}$ is

$$\mathbf{A} = \mathbf{G}_1 \cdot \mathbf{G}_1' = \overline{\mathbf{U}}_1 \cdot \Lambda^{-1} \cdot \overline{\mathbf{U}}_1'.$$

The pseudoinverse has a peculiar representation if we introduce the matrices $\overline{\mathbf{U}}_1$, $\mathbf{U}_1$ and Λ^{-1}.

Definition (pseudoinverse):

If we use the representation of the matrix $\mathbf{A}$ of type $\mathbf{A} = \mathbf{G}_1 \cdot \mathbf{G}_1' =$ $\mathbf{U}_1\Lambda\mathbf{U}_1'$ *then*

$$\mathbf{A}^+ := \overline{\mathbf{U}}_1 \cdot \overline{\mathbf{U}}_1 = \mathbf{U}_1 \cdot \Lambda^{-1} \cdot \mathbf{U}_1'$$

is the representation of its *pseudoinverse* namely

(i) $\mathbf{A}\mathbf{A}^+\mathbf{A} = (\mathbf{U}_1\Lambda\mathbf{U}_1')(\mathbf{U}_1\Lambda^{-1}\mathbf{U}_1')(\mathbf{U}_1\Lambda\mathbf{U}_1') = \mathbf{U}_1\Lambda\mathbf{U}_1'$

(ii) $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = (\mathbf{U}_1\Lambda^{-1}\mathbf{U}_1')(\mathbf{U}_1\Lambda\mathbf{U}_1')(\mathbf{U}_1\Lambda^{-1}\mathbf{U}_1') = \mathbf{U}_1\Lambda^{-1}\mathbf{U}_1' = \mathbf{A}^+$

(iii) $\mathbf{A}\mathbf{A}^+ = (\mathbf{U}_1\Lambda\mathbf{U}_1')(\mathbf{U}_1\Lambda^{-1}\mathbf{U}_1') = \mathbf{U}_1\mathbf{U}_1' = (\mathbf{A}\mathbf{A}^+)'$

(iv) $\mathbf{A}^+\mathbf{A} = (\mathbf{U}_1\Lambda^{-1}\mathbf{U}_1')(\mathbf{U}_1\Lambda\mathbf{U}_1') = \mathbf{U}_1\mathbf{U}_1' = (\mathbf{A}^+\mathbf{A})'$.

The *pseudoinverse* $\mathbf{A}^+$ exists and is unique, even if $\mathbf{A}$ is *singular*. For a nonsingular matrix $\mathbf{A}$, the matrix $\mathbf{A}^+$ is identical with $\mathbf{A}^{-1}$. *Indeed, for the case of the pseudoinverse (or any other generalized inverse) the generalized inverse of a rectangular matrix exists.* The *singular value decomposition is an excellent tool which generalizes the classical eigenvalue – eigenvector decomposition of symmetric matrices.*

Lemma (Singular value decomposition):

(i) Let $\mathbf{A}$ be an $n \times m$ matrix of rank $r := \text{rk}\mathbf{A} \le \min(n, m)$. Then the matrices $\mathbf{A}'\mathbf{A}$ and $\mathbf{A}'\mathbf{A}$ are symmetric positive (semi) definite matrices whose nonvanishing eigenvalues $\{\lambda_1, \ldots \lambda_r\}$ are *positive*. Especially

$$r = \text{rk}(\mathbf{A}'\mathbf{A}) = \text{rk}(\mathbf{A}\mathbf{A}')$$

holds. $\mathbf{A}'\mathbf{A}$ contains 0 as a *multiple eigenvalue of degree* $m - r$, and $\mathbf{A}\mathbf{A}'$ has the *multiple eigenvalue of degree* $n - r$.

(ii) With the support of *orthonormal eigenvalues of* $\mathbf{A}'\mathbf{A}$ *and* $\mathbf{A}\mathbf{A}'$ we are able to introduce an $m \times m$ matrix $\mathbf{V}$ and an $n \times n$ matrix $\mathbf{U}$ such that $\mathbf{U}\mathbf{U}' = \mathbf{U}'\mathbf{U} = \mathbf{I}_n, \mathbf{V}\mathbf{V}' = \mathbf{V}'\mathbf{V} = \mathbf{I}_m$ holds *and*

$$\boxed{\begin{aligned}\mathbf{U}'\mathbf{A}\mathbf{A}'\mathbf{U} &= \text{Diag}(\lambda_1^2, \ldots, \lambda_r^2, 0, \ldots, 0),\\ \mathbf{V}'\mathbf{A}'\mathbf{A}\mathbf{V} &= \text{Diag}(\lambda_1^2, \ldots, \lambda_r^2, 0, \ldots, 0).\end{aligned}}$$

The diagonal matrices on the right side have different formats $m \times m$ and $m \times n$.

(iii) The original $n \times m$ matrix $\mathbf{A}$ can be decomposed according to

$$\boxed{\mathbf{U}'\mathbf{A}\mathbf{V} = \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix}, \; O(\mathbf{U}\mathbf{A}\mathbf{V}') = n \times m}$$

with the $r \times r$ diagonal matrix

$$\Lambda := \mathrm{Diag}(\lambda_1, \ldots, \lambda_r)$$

of *singular values* representing the positive roots of non-vanishing eigenvalues of $\mathbf{A}'\mathbf{A}$ and $\mathbf{A}\mathbf{A}'$.

(iv) A *synthetic form* of the $n \times m$ matrix **A is**

$$\boxed{\mathbf{A} = \mathbf{U}' \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \mathbf{V}'.}$$

We note here that all transformed matrices of type $\mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ of a *quadratic matrix* have the *same eigenvalues* as $\mathbf{A} = (\mathbf{A}\mathbf{T})\mathbf{T}^{-1}$ being used as often as an *invariance property*.

?what is the relation between *eigenvalues* and the *trace, the determinant, the rank*? The answer will be given *now*.

Lemma (relation between *eigenvalues and other scalar measures*):

Let $\mathbf{A}$ be a quadratic *matrix* of the order $O(\mathbf{A}) = m \times m$ with eigenvalues in decreasing order. Then we have

$$|\mathbf{A}| = \prod_{j=1}^{m} \lambda_j, \quad \mathrm{tr}\mathbf{A} = \sum_{j=1}^{m} \lambda_j, \quad \mathrm{rk}\mathbf{A} = \mathrm{tr}\mathbf{A},$$

if $\mathbf{A}$ is *idempotent*. If $\mathbf{A} = \mathbf{A}'$ is a *symmetric matrix with real eigenvalues*, then we gain

$$\lambda_1 \geq \max\{a_{jj} \mid j = 1, \ldots, m\},$$

$$\lambda_m \leq \min\{a_{jj} \mid j = 1, \ldots, m\}.$$

At the end we compute the *eigenvalues and eigenvectors* which relate the *variation problem* $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathrm{extr}$ subject to the *condition* $\mathbf{x}'\mathbf{x} = 1$, namely

$$\mathbf{x}'\mathbf{A}\mathbf{x} + \lambda(\mathbf{x}'\mathbf{x}) = \underset{\mathbf{x},\lambda}{\mathrm{extr}}.$$

The *eigenvalue* λ is the *Lagrange multiplicator* of the *optimization problem.*

A6 Generalized Inverses

Because the *inversion by Cayley inversion is* only possible for *quadratic nonsingular matrices*, we introduce a slightly more general *definition* in order to invert arbitrary matrices **A** of the *order* $O(\mathbf{A}) = n \times m$ by so – called *generalized inverses* or for short *g – inverses*.

An $m \times n$ matrix **G** is called *g – inverse of the* matrix **A** if it fulfils the equation

$$\mathbf{AGA} = \mathbf{A}$$

in the sense of *Cayley multiplication.* Such g – inverses always exist and are unique *if and only if* **A** is a *nonsingular quadratic matrix*. In this case

$$\mathbf{G} = \mathbf{A}^{-1} \text{ if } \mathbf{A} \text{ is } \textit{invertible},$$

in other cases we use the notation

$$\mathbf{G} = \mathbf{A}^{-} \quad \text{if} \quad \mathbf{A}^{-1} \text{ does } \textit{not} \text{ exist.}$$

For the *rank* of all g – inverses the *inequality*

$$r := \operatorname{rk}\mathbf{A} \leq \operatorname{rk}\mathbf{A}^{-} \leq \min\{n, m\}$$

holds. In reverse, for any even number *d* in this interval there exists a g – inverse $\mathbf{A}^{-}$ such that

$$d = \operatorname{rk}\mathbf{A}^{-} = \dim \mathcal{R}(\mathbf{A}^{-})$$

holds. Especially even for a singular quadratic matrix **A** of the order $O(\mathbf{A}) = n \times n$ there exist g-inverses $\mathbf{A}^{-}$ of full rank $\operatorname{rk}\mathbf{A}^{-} = n$. In particular, such g-inverses $\mathbf{A}_r^{-}$ are of interest which have the *same rank* compared to the matrix **A**, namely

$$\operatorname{rk}\mathbf{A}_r^{-} = r = \operatorname{rk}\mathbf{A}.$$

Those *reflexive g-inverse* $\mathbf{A}_r^{-}$ are equivalent due to the *additional condition*

$$\mathbf{A}_r^{-}\mathbf{A}\mathbf{A}_r^{-} = \mathbf{A}_r^{-}$$

but are *not* necessary symmetric for symmetric matrices **A**. In general,

$$\mathbf{A} = \mathbf{A}' \textit{ and } \mathbf{A}^{-} \text{ g-inverse of } \mathbf{A} \Rightarrow$$

$$\Rightarrow (\mathbf{A}^{-})' \ \textit{g-inverse of} \ \mathbf{A}$$

$$\Rightarrow \mathbf{A}_{rs}^{-} := \mathbf{A}^{-}\mathbf{A}(\mathbf{A}^{-})' \text{ is } \textit{reflexive symmetric g} - \textit{inverse of } \mathbf{A}.$$

For constructing $\mathbf{A}_{rs}^{-}$ we only need an arbitrary g-inverse of **A**. On the other side, $\mathbf{A}_{rs}^{-}$ does *not mean unique.* There exist certain matrix functions which are *independent of the choice of the g-inverse*. For instance,

---

With permission from the author: Extract from "Erik W. Grafarend: Linear and nonlinear models. Fixed effects, random effects, and mixed models"

$$\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A} \text{ and } \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}$$

can be used to generate *special g-inverses* of $\mathbf{A}'\mathbf{A}$ or $\mathbf{A}\mathbf{A}'$. For instance,

$$\mathbf{A}_{\ell}^{-} := (\mathbf{A}'\mathbf{A})^{-}\mathbf{A}' \text{ and } \mathbf{A}_{m}^{-} := \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-}$$

have the *special reproducing properties*

$$\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\mathbf{A} = \mathbf{A}\mathbf{A}_{\ell}^{-}\mathbf{A} = \mathbf{A}$$

and

$$\mathbf{A}\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-}\mathbf{A} = \mathbf{A}\mathbf{A}_{m}^{-}\mathbf{A} = \mathbf{A},$$

which can be *generalized* in case that **W** and **S** are *positive semidefinite matrices to*

$$\mathbf{W}\mathbf{A}(\mathbf{A}'\mathbf{W}\mathbf{A})^{-}\mathbf{A}'\mathbf{W}\mathbf{A} = \mathbf{W}\mathbf{A}$$

$$\mathbf{A}\mathbf{S}\mathbf{A}'(\mathbf{A}\mathbf{S}\mathbf{A}')^{-}\mathbf{A}\mathbf{S} = \mathbf{A}\mathbf{S},$$

where the matrices

$$\mathbf{W}\mathbf{A}(\mathbf{A}'\mathbf{W}\mathbf{A})^{-}\mathbf{A}'\mathbf{W} \text{ and } \mathbf{S}\mathbf{A}'(\mathbf{A}\mathbf{S}\mathbf{A}')^{-}\mathbf{A}\mathbf{S}$$

are *independent of the choice of the g-inverse* $(\mathbf{A}'\mathbf{W}\mathbf{A})^{-}$ *and* $(\mathbf{A}\mathbf{S}\mathbf{A}')^{-}$.

A *beautiful interpretation* of the various g-inverses is based on the *fact* that the matrices

$$(\mathbf{A}\mathbf{A}^{-})(\mathbf{A}\mathbf{A}^{-}) = (\mathbf{A}\mathbf{A}^{-}\mathbf{A})\mathbf{A}^{-} = \mathbf{A}\mathbf{A}^{-} \text{ and } (\mathbf{A}^{-}\mathbf{A})(\mathbf{A}^{-}\mathbf{A}) = \mathbf{A}^{-}(\mathbf{A}\mathbf{A}^{-}\mathbf{A}) = \mathbf{A}^{-}\mathbf{A}$$

are *idempotent* and can therefore be *geometrically* interpreted as projections. The *image* of $\mathbf{A}\mathbf{A}^{-}$, namely

$$\mathcal{R}(\mathbf{A}\mathbf{A}^{-}) = \mathcal{R}(\mathbf{A}) = \{\mathbf{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^{m}\} \subset \mathbb{R}^{n},$$

can be completed by the *projections* $\mathbf{A}^{-}\mathbf{A}$ along the *null space*

$$\mathcal{N}(\mathbf{A}^{-}\mathbf{A}) = \mathcal{N}(\mathbf{A}) = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} = 0\} \subset \mathbb{R}^{m}.$$

By the choice of the g – inverse we are able to choose the *projected direction* of $\mathbf{A}\mathbf{A}^{-}$ and the *image of the projections* $\mathbf{A}^{-}\mathbf{A}$ if we take advantage of the *complementary spaces of the subspaces*

$$\mathcal{R}(\mathbf{A}^{-}\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^{-}\mathbf{A}) = \mathbb{R}^{m} \text{ and } \mathcal{R}(\mathbf{A}\mathbf{A}^{-}) \oplus \mathcal{N}(\mathbf{A}\mathbf{A}^{-}) = \mathbb{R}^{n}$$

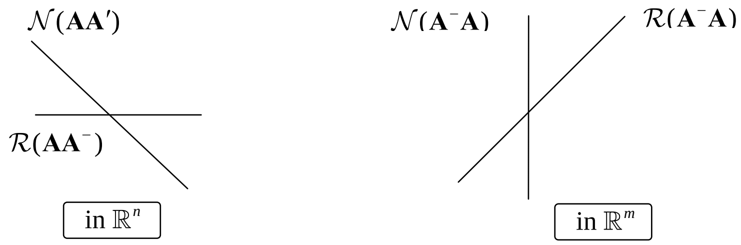
by using the symbol "$\oplus$" as the sign of “*direct sum*” of linear spaces which only have the zero element in common. Finally we have use the corresponding dimensions

$$\dim \mathcal{R}(\mathbf{A}^{-}\mathbf{A}) = r = \operatorname{rk}\mathbf{A} = \dim \mathcal{R}(\mathbf{A}\mathbf{A}^{-}) \Rightarrow$$

$$\Rightarrow \left[\begin{array}{l} \dim \mathcal{N}(\mathbf{A}^{-}\mathbf{A}) = m - \operatorname{rk}\mathbf{A} = m - r \\ \dim \mathcal{N}(\mathbf{A}\mathbf{A}^{-}) = n - \operatorname{rk}\mathbf{A} = n - r \end{array}\right.$$

---

independent of the special rank of the g-inverses $\mathbf{A}^-$ which are determined by the *subspaces* $\mathcal{R}(\mathbf{A}^-\mathbf{A})$ and $\mathcal{N}(\mathbf{A}\mathbf{A}^-)$, respectively.



Example (geodetic networks):

In a geodetic network, the projections $\mathbf{A}^-\mathbf{A}$ correspond to a *S – transformations* in the sense of *W. Baarda (1973).*

Example ($\mathbf{A}_\ell^-$ and $\mathbf{A}_m^-$ *g*-inverses):

The projections $\mathbf{AA}_\ell^- = \mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$ guarantee that the *subspaces* $\mathcal{R}(\mathbf{AA}^-)$ and $\mathcal{N}(\mathbf{AA}_\ell^-)$ *are orthogonal* to each other. The same holds for the *subspaces* $\mathcal{R}(\mathbf{A}_m^-\mathbf{A})$ and $\mathcal{N}(\mathbf{A}_m^-\mathbf{A})$ of the projections $\mathbf{A}_m^-\mathbf{A} = \mathbf{A}'(\mathbf{AA}')^-\mathbf{A}$.

In general, there exist more than one g-inverses which lead to *identical projections* $\mathbf{AA}^-$ *and* $\mathbf{A}^-\mathbf{A}$. For instance, following A. *Ben – Israel, T. N. E. Greville* (1974, p.59) we learn that the *reflexive g-inverse* which follows from

$$\mathbf{A}_r^- = (\mathbf{A}^-\mathbf{A})\mathbf{A}^-(\mathbf{AA}^-) = \mathbf{A}^-\mathbf{AA}^-$$

contains the *class of all reflexive g-inverses.* Therefore it is obvious that the *reflexive g-inverses* $\mathbf{A}_r^-$ contain exact by *one pair of projections* $\mathbf{AA}^-$ *and* $\mathbf{A}^-\mathbf{A}$ *and conversely*. In the *special case of a symmetric matrix* $\mathbf{A}$, $\mathbf{A} = \mathbf{A}'$, *and* $n = m$ we know due to

$$\mathcal{R}(\mathbf{AA}^-) = \mathcal{R}(\mathbf{A}) \perp \mathcal{N}(\mathbf{A}') = \mathcal{N}(\mathbf{A}) = \mathcal{N}(\mathbf{A}^-\mathbf{A})$$

that the *column spaces* $\mathcal{R}(\mathbf{AA}^-)$ are *orthogonal* to the *null space* $\mathcal{N}(\mathbf{A}^-\mathbf{A})$ *illustrated by the sign* "$\perp$". If these complementary subspaces $\mathcal{R}(\mathbf{A}^-\mathbf{A})$ and $\mathcal{N}(\mathbf{AA}^-)$ are *orthogonal to each other*, the postulate of a *symmetric reflexive g-inverse* agrees to

$$\mathbf{A}_{rs}^- := (\mathbf{A}^-\mathbf{A})\mathbf{A}^-(\mathbf{A}^-\mathbf{A})' = \mathbf{A}^-\mathbf{A}(\mathbf{A}^-)',$$

if $\mathbf{A}^-$ is a suited g-inverse.

There is *no insurance* that the complementary subspaces $\mathcal{R}(\mathbf{A}^-\mathbf{A})$ and $\mathcal{N}(\mathbf{A}^-\mathbf{A})$ and $\mathcal{R}(\mathbf{A}\mathbf{A}^-)$ and $\mathcal{N}(\mathbf{A}\mathbf{A}^-)$ are *orthogonal*. If such a result should be reached, we should use

the uniquely defined pseudoinverse $\mathbf{A}^+$,
also called Moore-Penrose inverse

for which holds

$$\mathcal{R}(\mathbf{A}^+\mathbf{A}) \perp \mathcal{N}(\mathbf{A}^+\mathbf{A}),\ \mathcal{R}(\mathbf{A}\mathbf{A}^+) \perp \mathcal{N}(\mathbf{A}\mathbf{A}^+)$$

or equivalent

$$\mathbf{A}\mathbf{A}^+ = (\mathbf{A}\mathbf{A}^+)',\quad \mathbf{A}^+\mathbf{A} = (\mathbf{A}^+\mathbf{A})'.$$

If we depart from an arbitrary g-inverse $(\mathbf{A}\mathbf{A}^-\mathbf{A})^-$, the *pseudoinverse* $\mathbf{A}^+$ can be build on

$$\mathbf{A}^+ := \mathbf{A}'(\mathbf{A}\mathbf{A}'\mathbf{A})^-\mathbf{A}' \text{ (Zlobec formula)}$$

or

$$\mathbf{A}^+ := \mathbf{A}'(\mathbf{A}\mathbf{A}')^-\mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}' \text{ (Bjerhammar formula)},$$

if both the g-inverses $(\mathbf{A}\mathbf{A}')^-$ and $(\mathbf{A}'\mathbf{A})^-$ exist. The *Moore-Penrose* inverse fulfils the *Penrose equations:*

(i) $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$ (g-inverse)

(ii) $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$ (reflexivity)

$$\left.\begin{array}{ll}\text{(iii)} & \mathbf{A}\mathbf{A}^+ = (\mathbf{A}\mathbf{A}^+)' \\ \text{(iv)} & \mathbf{A}^+\mathbf{A} = (\mathbf{A}^+\mathbf{A})'\end{array}\right] \text{Symmetry due to orthogonal projection.}$$

Lemma (*Penrose equations)*

Let $\mathbf{A}$ be a *rectangular matrix* $\mathbf{A}$ of the order $O(\mathbf{A})$ be given. A g-generalized matrix inverse which is rank preserving $\operatorname{rk}(\mathbf{A}) = \operatorname{rk}(\mathbf{A}^+)$ fulfils the axioms of the *Penrose equations* (i) - (iv).

For the special case of a *symmetric matrix* $\mathbf{A}$ also the *pseudoinverse* $\mathbf{A}^+$ is symmetric, fulfilling

$$\mathcal{R}(\mathbf{A}^+\mathbf{A}) = \mathcal{R}(\mathbf{A}\mathbf{A}^+) \perp \mathcal{N}(\mathbf{A}\mathbf{A}^+) = \mathcal{N}(\mathbf{A}^+\mathbf{A}),$$

in addition

$$\mathbf{A}^+ = \mathbf{A}(\mathbf{A}^2)^-\mathbf{A} = \mathbf{A}(\mathbf{A}^2)^-\mathbf{A}(\mathbf{A}^2)^-\mathbf{A}.$$

Various formulas of computing *certain g-inverses,* for instance by the *method of rank factorization,* exist. Let **A** be an $n \times m$ matrix **A** of rank $r := \operatorname{rk}\mathbf{A}$ such that

$$\mathbf{A} = \mathbf{GF},\ O(\mathbf{G}) = n \times r,\ O(\mathbf{F}) = r \times m .$$

Due to the *inequality* $r \le \operatorname{rk}\mathbf{G}^- \le \min\{r, n\} = r$ only **G** *posesses reflexive g-inverses* $\mathbf{G}_r^-$, because of

$$\mathbf{I}_{r\times r} = [(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}']\mathbf{G} = [(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'](\mathbf{G}\mathbf{G}_r'\mathbf{G}) = \mathbf{G}_r^-\mathbf{G}$$

represented by left inverses in the sense of $\mathbf{G}_L^-\mathbf{G} = \mathbf{I}$. In a similar way, all g-inverses of **F** are reflexive and right inverses subject to $\mathbf{F}_r^- := \mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}$.

The whole class of reflexive g-inverses of **A** can be represented by

$$\mathbf{A}_r^- := \mathbf{F}_r^-\mathbf{G}_r^- = \mathbf{F}_r^-\mathbf{G}_L^- .$$

In this case we also find the pseudoinverse, namely

$$\mathbf{A}^+ := \mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'$$

because of

$$\mathcal{R}(\mathbf{A}^+\mathbf{A}) = \mathcal{R}(\mathbf{F}') \perp \mathcal{N}(\mathbf{F}) = \mathcal{N}(\mathbf{A}^+\mathbf{A}) = \mathcal{N}(\mathbf{A})$$

$$\mathcal{R}(\mathbf{A}\mathbf{A}^+) = \mathcal{R}(\mathbf{G}) \perp \mathcal{N}(\mathbf{G}') = \mathcal{N}(\mathbf{A}\mathbf{A}^+) = \mathcal{N}(\mathbf{A}').$$

If we want to give up the orthogonality conditions, in case of a *quadratic matrix* $\mathbf{A} = \mathbf{GF}$, we could take advantage of the projections

$$\mathbf{A}_r^-\mathbf{A} = \mathbf{A}\mathbf{A}_r^-$$

we could postulate

$$\mathcal{R}(\mathbf{A}_p^-\mathbf{A}) = \mathcal{R}(\mathbf{A}\mathbf{A}_r^-) = \mathcal{R}(\mathbf{G}),$$

$$\mathcal{N}(\mathbf{A}'\mathbf{A}_r^-) = \mathcal{N}(\mathbf{A}_r^-\mathbf{A}) = \mathcal{N}(\mathbf{F}).$$

In consequence, if **FG** is a *nonsingular matrix,* we enjoy the representation

$$\mathbf{A}_r^- := \mathbf{G}(\mathbf{F}\mathbf{G})^{-1}\mathbf{F},$$

which reduces in case that **A** is a *symmetric matrix* to the *pseudoinverse* $\mathbf{A}^+$.

Dual methods of computing g-inverses $\mathbf{A}^-$ are based on the *basis of the null space, both* for **F** *and* **G**, or for **A** and $\mathbf{A}'$. On the *first side* we need the matrix $\mathbf{E}_\mathbf{F}$ by

$$\mathbf{F}\mathbf{E}_\mathbf{F}' = 0,\ \operatorname{rk}\mathbf{E}_\mathbf{F} = m - r \ \ \textit{versus}\ \ \mathbf{G}'\mathbf{E}_{\mathbf{G}'} = 0,\ \operatorname{rk}\mathbf{E}_{\mathbf{G}'} = n - r$$

on the *other side*. The enlarged matrix of the order $(n + r - r) \times (n + m - r)$ is automatically *nonsingular* and has the *Cayley inverse*

With permission from the author: Extract from "Erik W. Grafarend: Linear and nonlinear models. Fixed effects, random effects, and mixed models"

$$\begin{bmatrix} \mathbf{A} & \mathbf{E}_{\mathbf{G}'} \\ \mathbf{E}_{\mathbf{F}} & 0 \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{+} & \mathbf{E}_{\mathbf{F}}^{+} \\ \mathbf{E}_{\mathbf{G}'}^{+} & 0 \end{bmatrix}$$

with the pseudoinverse $\mathbf{A}^{+}$ on the upper left side. Details can be derived from *A. Ben – Israel and T. N. E. Greville (1974 p. 228).*

If the null spaces are always normalized in the sense of

$$< \mathbf{E}_{\mathbf{F}} \mid \mathbf{E}'_{\mathbf{F}} >= \mathbf{I}_{m-r}, < \mathbf{E}'_{\mathbf{G}'} \mid \mathbf{E}_{\mathbf{G}'} >= \mathbf{I}_{n-r}$$

because of

$$\mathbf{E}_{\mathbf{F}}^{+} = \mathbf{E}'_{\mathbf{F}} < \mathbf{E}_{\mathbf{F}} \mid \mathbf{E}'_{\mathbf{F}} >^{-1} = \mathbf{E}'_{\mathbf{F}}$$

and

$$\mathbf{E}_{\mathbf{G}'}^{+} =< \mathbf{E}'_{\mathbf{G}'} \mid \mathbf{E}_{\mathbf{G}'} >^{-1} \mathbf{E}'_{\mathbf{G}'} = \mathbf{E}'_{\mathbf{G}'}$$

$$\begin{bmatrix} \mathbf{A} & \mathbf{E}_{G'} \\ \mathbf{E}_{F} & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{+} & \mathbf{E}_{G'} \\ \mathbf{E}'_{F} & \mathbf{0} \end{bmatrix}.$$

These formulas gain a special structure if the matrix $\mathbf{A}$ is *symmetric* to the order $O(\mathbf{A})$. In this case

$$\mathbf{E}_{G'} = \mathbf{E}'_{F} =: \mathbf{E}', \ \ O(\mathbf{E}) = (m-r)\times m, \ \ \mathrm{rk}\,\mathbf{E} = m-r$$

and

$$\begin{bmatrix} \mathbf{A} & \mathbf{E}' \\ \mathbf{E} & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{+} & \mathbf{E}' < \mathbf{E} \mid \mathbf{E}' >^{-1} \\ < \mathbf{E} \mid \mathbf{E}' >^{-1} \mathbf{E} & \mathbf{0} \end{bmatrix}$$

on the basis of such a relation, namely $\mathbf{E}\mathbf{A}^{+} = \mathbf{0}$ there follows

$$\mathbf{I}_{m} = \mathbf{A}\mathbf{A}^{+} + \mathbf{E}' < \mathbf{E} \mid \mathbf{E}' >^{-1} \mathbf{E} =$$
$$= (\mathbf{A} + \mathbf{E}'\mathbf{E})[\mathbf{A}^{+} + \mathbf{E}'(\mathbf{E}\mathbf{E}'\mathbf{E}\mathbf{E}')^{-1}\mathbf{E}]$$

and with the *projection* (*S - transformation*)

$$\mathbf{A}^{+}\mathbf{A} = \mathbf{I}_{m} - \mathbf{E}' < \mathbf{E} \mid \mathbf{E}' >^{-1} \mathbf{E} = (\mathbf{A} + \mathbf{E}'\mathbf{E})^{-1}\mathbf{A}$$

and

$$\mathbf{A}^{+} = (\mathbf{A} + \mathbf{E}'\mathbf{E})^{-1} - \mathbf{E}'(\mathbf{E}\mathbf{E}'\mathbf{E}\mathbf{E}')^{-1}\mathbf{E}$$

pseudoinverse of $\mathbf{A}$

$$\mathcal{R}(\mathbf{A}^{+}\mathbf{A}) = \mathcal{R}(\mathbf{A}\mathbf{A}^{+}) = \mathcal{R}(\mathbf{A}) \perp \mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{E}').$$

In case of a *symmetric, reflexive g-inverse* $\mathbf{A}_{rs}^{-}$ there holds the *orthogonality* or *complementary*

With permission from the author: Extract from "Erik W. Grafarend: Linear and nonlinear models. Fixed effects, random effects, and mixed models"

$$\mathcal{R}(\mathbf{A}_{rs}^{-}\mathbf{A}) \perp \mathcal{N}(\mathbf{A}\mathbf{A}_{rs}^{-})$$

$$\mathcal{N}(\mathbf{A}\mathbf{A}_{rs}^{-}) \text{ complementary to } \mathcal{R}(\mathbf{A}\mathbf{A}_{rs}^{-}),$$

which is guaranteed by a matrix $\mathbf{K}$, $\operatorname{rk}\mathbf{K} = m-r$, $O(\mathbf{K}) = (m-r)\times m$ such that

$\mathbf{KE}'$ is a *non-singular matrix*.

At the same time, we take advantage of the *bordering* of the matrix $\mathbf{A}$ by $\mathbf{K}$ and $\mathbf{K}'$, by a *non-singular matrix* of the order $(2m-r)\times(2m-r)$.

$$\begin{bmatrix} \mathbf{A} & \mathbf{K}' \\ \mathbf{K} & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{rs}^{-} & \mathbf{K}_{R}^{-} \\ (\mathbf{K}_{R}^{-})' & \mathbf{0} \end{bmatrix}.$$

$\mathbf{K}_{R}^{-} := \mathbf{E}'(\mathbf{KE}')^{-1}$ is the *right inverse* of $\mathbf{A}$. Obviously, we gain the *symmetric reflexive g-inverse* $\mathbf{A}_{rs}^{-}$ *whose columns are orthogonal to* $\mathbf{K}'$:

$$\mathrm{R}(\mathbf{A}_{rs}^{-}\mathbf{A}) \perp \mathrm{R}(\mathbf{K}') = \mathcal{N}(\mathbf{A}\mathbf{A}_{rs}^{-})$$

$$\mathbf{K}\mathbf{A}_{rs}^{-} = \mathbf{0} \quad \Rightarrow$$

$$\Rightarrow \quad \mathbf{I}_m = \mathbf{A}\mathbf{A}_{rs}^{-} + \mathbf{K}'(\mathbf{EK}')^{-1}\mathbf{E} =$$

$$= (\mathbf{A} + \mathbf{K}'\mathbf{K})[\mathbf{A}_{rs}^{-} + \mathbf{E}'(\mathbf{EK}'\mathbf{EK}')^{-1}\mathbf{E}]$$

and *projection (S - transformation)*

$$\mathbf{A}_{rs}^{-}\mathbf{A} = \mathbf{I}_m - \mathbf{E}'(\mathbf{KE}')^{-1}\mathbf{K} = (\mathbf{A} + \mathbf{K}'\mathbf{K})^{-1}\mathbf{A}',$$

$$\mathbf{A}_{rs}^{-} = (\mathbf{A} + \mathbf{K}'\mathbf{K})^{-1} - \mathbf{E}'(\mathbf{EK}'\mathbf{EK}')^{-1}\mathbf{E}.$$

symmetric reflexive g-inverse

For the special case of a *symmetric and positive semidefinite* $m\times m$ matrix $\mathbf{A}$ the matrix set $\mathbf{U}$ and $\mathbf{V}$ are reduced to one. Based on the various matrix decompositions

$$\mathbf{A} = \begin{bmatrix} \mathbf{U}_1, & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U}_1' \\ \mathbf{U}_2' \end{bmatrix} = \mathbf{U}_1\mathbf{\Lambda}\mathbf{U}_1',$$

we find the different g - inverses listed as following.

$$\mathbf{A} = \begin{bmatrix} \mathbf{U}_1, & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}^{-1} & \mathbf{L}_{12} \\ \mathbf{L}_{21} & \mathbf{L}_{21}\mathbf{\Lambda}\mathbf{L}_{12} \end{bmatrix} \begin{bmatrix} \mathbf{U}_1' \\ \mathbf{U}_2' \end{bmatrix}.$$

Lemma (*g-inverses of symmetric and positive semidefinite matrices*):

(i) $$\mathbf{A}^{-} = \begin{bmatrix} \mathbf{U}_1, & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}^{-1} & \mathbf{L}_{12} \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{U}_1' \\ \mathbf{U}_2' \end{bmatrix},$$

(ii) reflexive g-inverse

$$\mathbf{A}_r^- = [\mathbf{U}_1, \quad \mathbf{U}_2]\begin{bmatrix}\mathbf{\Lambda}^{-1} & \mathbf{L}_{12} \\ \mathbf{L}_{21} & \mathbf{L}_{21}\mathbf{\Lambda}\mathbf{L}_{12}\end{bmatrix}\begin{bmatrix}\mathbf{U}_1' \\ \mathbf{U}_2'\end{bmatrix}$$

(iii) reflexive and symmetric g-inverse

$$\mathbf{A}_{rs}^- = [\mathbf{U}_1, \quad \mathbf{U}_2]\begin{bmatrix}\mathbf{\Lambda}^{-1} & \mathbf{L}_{12} \\ \mathbf{L}_{12} & \mathbf{L}_{12}\mathbf{\Lambda}\mathbf{L}_{12}\end{bmatrix}\begin{bmatrix}\mathbf{U}_1' \\ \mathbf{U}_2'\end{bmatrix}$$

(iv) pseudoinverse

$$\mathbf{A}^+ = [\mathbf{U}_1, \quad \mathbf{U}_2]\begin{bmatrix}\mathbf{\Lambda}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{U}_1' \\ \mathbf{U}_2'\end{bmatrix} = \mathbf{U}_1\mathbf{\Lambda}^{-1}\mathbf{U}_1.$$

We look at a representation of the *Moore-Penrose inverse* in terms of U_2, the basis of the *null space* $\mathcal{N}(\mathbf{A}^-\mathbf{A})$. In these terms we find

$$\mathbf{E} := \mathbf{U}_1 \qquad \Rightarrow \qquad \begin{bmatrix}\mathbf{A} & \mathbf{U}_2 \\ \mathbf{U}_2' & \mathbf{0}\end{bmatrix}^{-1} = \begin{bmatrix}\mathbf{A}^+ & \mathbf{U}_2 \\ \mathbf{U}_2' & \mathbf{0}\end{bmatrix},$$

by means of the *fundamental relation* of $\mathbf{A}^+\mathbf{A}$

$$\mathbf{A}^+\mathbf{A} = \lim_{\delta\to 0}(\mathbf{A}+\delta\mathbf{I}_m)^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^+ = \mathbf{I}_m - \mathbf{U}_2\mathbf{U}_2' = \mathbf{U}_1\mathbf{U}_1',$$

we generate the *fundamental relation of the pseudo inverse*

$$\mathbf{A}^+ = (\mathbf{A}+\mathbf{U}_2\mathbf{U}_2')^{-1} - \mathbf{U}_2\mathbf{U}_2'.$$

The main target of our discussion of various g-inverses is the easy handling of representations of solutions of arbitrary linear equations and their characterizations.

We depart from the solution of a *consistent system of linear equations*,

$$\mathbf{A}\mathbf{x} = \mathbf{c}, \quad O(\mathbf{A}) = n\times m, \quad \mathbf{c}\in\mathcal{R}(\mathbf{A}) \ \Rightarrow \ \mathbf{x} = \mathbf{A}^-\mathbf{c} \quad \text{for any g-inverse } \mathbf{A}^-.$$

$\mathbf{x} = \mathbf{A}^-\mathbf{c}$ is the *general solution* of such a linear system of equations. *If we want to generate a special g - inverse*, we can represent *the general solution* by

$$\mathbf{x} = \mathbf{A}^-\mathbf{c} + (\mathbf{I}_m - \mathbf{A}^-\mathbf{A})\mathbf{z} \quad \textit{for all} \ \ \mathbf{z}\in\mathbb{R}^m,$$

since the *subspaces* $\mathcal{N}(\mathbf{A})$ and $\mathcal{R}(\mathbf{I}_m - \mathbf{A}^-\mathbf{A})$ are *identical*. We test the *consistency* of our system by means of the *identity*

$$\mathbf{A}\mathbf{A}^-\mathbf{c} = \mathbf{c}.$$

$\mathbf{c}$ is mapped by the projection $\mathbf{A}\mathbf{A}^-$ *to itself*.

Similary we solve the matrix equation $\mathbf{AXB} = \mathbf{C}$ by the consistency test: the existence of the solution is granted by the identity

---

$$\mathbf{A}\mathbf{A}^{-}\mathbf{C}\mathbf{B}^{-}\mathbf{B} = \mathbf{C} \quad \textit{for any g-inverse } \mathbf{A}^{-} \text{ and } \mathbf{B}^{-}.$$

If this condition is fulfilled, we are able to generate the *general solution* by

$$\mathbf{X} = \mathbf{A}^{-}\mathbf{C}\mathbf{B} + \mathbf{Z} - \mathbf{A}^{-}\mathbf{A}\mathbf{Z}\mathbf{B}\mathbf{B}^{-},$$

where $\mathbf{Z}$ is an arbitrary matrix of suitable order. We can use an arbitrary g-inverse $\mathbf{A}^{-}$ and $\mathbf{B}^{-}$, for instance *the pseudoinverse* $\mathbf{A}^{+}$ and $\mathbf{B}^{+}$ which would be for $\mathbf{Z} = \mathbf{0}$ coincide with *two-sided orthogonal projections*.

How can we reduce the *matrix equation* $\mathbf{AXB} = \mathbf{C}$ to a *vector equation*?

The vec-operator is the door opener.

$$\mathbf{AXB} = \mathbf{C} \quad \Leftrightarrow \quad (\mathbf{B}' \otimes \mathbf{A}) \operatorname{vec} \mathbf{X} = \operatorname{vec} \mathbf{C}.$$

The *general solution* of our matrix equation reads

$$\operatorname{vec} \mathbf{X} = (\mathbf{B}' \otimes \mathbf{A})^{-} \operatorname{vec} \mathbf{C} + [\mathbf{I} - (\mathbf{B}' \otimes \mathbf{A})^{-} (\mathbf{B}' \otimes \mathbf{A})] \operatorname{vec} \mathbf{Z}.$$

Here we can use the identity

$$(\mathbf{A} \otimes \mathbf{B})^{-} = \mathbf{B}^{-} \otimes \mathbf{A}^{-},$$

generated by two *g-inverses of the Kronecker-Zehfuss product*.

At this end we solve the more general equation $\mathbf{Ax} = \mathbf{By}$ of *consistent type* $\mathcal{R}(\mathbf{A}) \subset \mathcal{R}(\mathbf{B})$ by

Lemma (*consistent system of homogenous equations* $\mathbf{Ax} = \mathbf{By}$):

Given the homogenous system of linear equations $\mathbf{Ax} = \mathbf{By}$ for $\mathbf{y} \in \mathbb{R}^{\ell}$ constraint by $\mathbf{By} \in \mathcal{R}(\mathbf{A})$. Then the solution $\mathbf{x} = \mathbf{Ly}$ can be given under the condition

$$\mathcal{R}(\mathbf{A}) \subset \mathcal{R}(\mathbf{B}).$$

In this case the matrix $\mathbf{L}$ may be decomposed by

$$\mathbf{L} = \mathbf{A}^{-}\mathbf{B} \quad \text{for a certain g-inverse } \mathbf{A}^{-}.$$

Appendix B: Matrix Analysis

A short version on *matrix analysis* is presented. Arbitrary *derivations of scalar-valued, vector-valued and matrix-valued vector – and matrix functions for functionally independent variables* are defined. Extensions for differenting *symmetric and antisymmetric matrices* are given. Special examples for *functionally dependent matrix variables* are reviewed.

B1 Derivatives of Scalar valued and Vector valued Vector Functions

Here we present the analysis of differentiating scalar-valued and vector-valued vector functions enriched by examples.

Definition: *(derivative of scalar valued vector function):*

Let a *scalar valued function* $f(\mathbf{x})$ of a vector $\mathbf{x}$ of the order $O(\mathbf{x}) = 1 \times m$ (row vector) be given, then we call

$$\mathbf{D}f(\mathbf{x}) = [\mathbf{D}_1 f(\mathbf{x}), \ldots, \mathbf{D}_m f(\mathbf{x})] := \frac{\partial f}{\partial \mathbf{x}'}$$

first derivative of $f(\mathbf{x})$ *with respect to* $\mathbf{x}'$.

Vector differentiation is based on the following definition.

Definition*: (*derivative of a matrix valued matrix function*):*

Let a $n \times q$ matrix-valued function $\mathbf{F(X)}$ of a $m \times p$ matrix of *functional independent variables* $\mathbf{X}$ be given. Then the $nq \times mp$ *Jacobi matrix of first derivates of* $\mathbf{F}$ is defined by

$$\mathbf{J_F} = \mathbf{DF(X)} := \frac{\partial \mathrm{vec}\mathbf{F(X)}}{\partial (\mathrm{vec}\mathbf{X})'} .$$

The definition of first derivatives of matrix-functions can be motivated as following. The matrices $\mathbf{F} = [f_{ij}] \in \mathbb{R}^{n \times q}$ and $\mathbf{X} = [x_{k\ell}] \in \mathbb{R}^{m \times p}$ are based *on two-dimensional arrays. In contrast*, the array of first derivatives

$$\left[\frac{\partial f_{ij}}{\partial \mathbf{x}_{k\ell}}\right] = \left[J_{ijk\ell}\right] \in \mathbb{R}^{n \times q \times m \times p}$$

is *four-dimensional* and automatic outside the usual frame of *matrix algebra of two-dimensional arrays.* By means of the operations $\mathrm{vec}\mathbf{F}$ and $\mathrm{vec}\mathbf{X}$ we will *vectorize* the matrices $\mathbf{F}$ and $\mathbf{X}$. Accordingly we will take advantage of $\mathrm{vec}\mathbf{F(X)}$ of the vector $\mathrm{vec}\mathbf{X}$ derived with respect to the matrix $\mathbf{J_F}$, a *two-dimensional array*.

With permission from the author: Extract from "Erik W. Grafarend: Linear and nonlinear models. Fixed effects, random effects, and mixed models"

Examples

(i) $f(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = a_{11}x_1^2 + (a_{12}+a_{21})x_1x_2 + a_{22}x_2^2$

$$\mathbf{D}f(\mathbf{x}) = [\mathbf{D}_1 f(\mathbf{x}), \mathbf{D}_2 f(\mathbf{x})] = \frac{\partial f}{\partial \mathbf{x}'} =$$

$$= [2a_{11}x_1 + (a_{12}+a_{21})x_2 \mid (a_{12}+a_{21})x_1 + 2a_{22}x_2] = \mathbf{x}'(\mathbf{A}+\mathbf{A}')$$

(ii) $f(x) = \mathbf{A}\mathbf{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}$

$$\mathbf{J_F} = \mathbf{D}f(\mathbf{x}) = \frac{\partial f}{\partial \mathbf{x}'} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \mathbf{A}$$

(iii) $\mathbf{F(X)} = \mathbf{X}^2 = \begin{bmatrix} x_{11}^2 + x_{12}x_{21} & x_{11}x_{12} + x_{12}x_{22} \\ x_{21}x_{11} + x_{22}x_{21} & x_{21}x_{12} + x_{22}^2 \end{bmatrix}$

$$\text{vec}\mathbf{F(X)} = \begin{bmatrix} x_{11}^2 + x_{12}x_{21} \\ x_{21}x_{11} + x_{22}x_{21} \\ x_{11}x_{12} + x_{12}x_{22} \\ x_{21}x_{12} + x_{22}^2 \end{bmatrix}$$

$$(\text{vec}\mathbf{X})' = [x_{11}, x_{21}, x_{12}, x_{22}]$$

$$\mathbf{J_F} = \mathbf{DF(X)} = \frac{\partial \text{vec}\mathbf{F(X)}}{\partial (\text{vec}\mathbf{X})'} = \begin{bmatrix} 2x_{11} & x_{12} & x_{21} & 0 \\ x_{21} & x_{11}+x_{22} & 0 & x_{21} \\ x_{12} & 0 & x_{11}+x_{22} & x_{12} \\ 0 & x_{12} & x_{21} & 2x_{22} \end{bmatrix}$$

$$O(\mathbf{J_F}) = 4 \times 4\ .$$

B2 Derivatives of Trace Forms

Up to now we have assumed that the vector $\mathbf{x}$ or the matrix $\mathbf{X}$ are *functionally idempotent*. For instance, the matrix $\mathbf{X}$ *cannot be a symmetric matrix* $\mathbf{X} = [x_{ij}] = [x_{ji}] = \mathbf{X}'$ or an *antisymmetric matrix* $\mathbf{X} = [x_{ij}] = [-x_{ji}] = -\mathbf{X}'$. In case of a *functional dependent variables*, for instance $x_{ij} = x_{ji}$ *or* $x_{ij} = -x_{ji}$ we can take advantage of the *chain rule* in order to derive the differential procedure.

$$\frac{\partial}{\partial \mathbf{X}} \text{tr}(\mathbf{AX}) = \begin{cases} \mathbf{A}', \text{ if } \mathbf{X} \text{ consists of functional independent elements;} \\ \mathbf{A}' + \mathbf{A} - \text{Diag}[a_{11}, \ldots, a_{nn}], \text{ if the } n \times n \text{ matrix } \mathbf{X} \text{ is } \textit{symmetric}; \\ \mathbf{A}' - \mathbf{A}, \text{ if the } n \times n \text{ matrix } \mathbf{X} \text{ is } \textit{antisymmetric}. \end{cases}$$

$$\frac{\partial}{\partial(\mathrm{vec}\mathbf{X})}\mathrm{tr}(\mathbf{AX}) = \begin{cases} [\mathrm{vec}\mathbf{A}']', \text{ if } \mathbf{X} \text{ consists of functional independent elements;} \\ [\mathrm{vec}(\mathbf{A}' + \mathbf{A} \text{ - Diag}[a_{11}, \ldots, a_{nn}])]', \text{ if the } n \times n \text{ matrix } \mathbf{X} \text{ is} \\ \textit{symmetric}; \\ [\mathrm{vec}(\mathbf{A}' - \mathbf{A})]', \text{ if the } n \times n \text{ matrix } \mathbf{X} \text{ is } \textit{antisymmetric}. \end{cases}$$

for instance

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}.$$

Case # 1: "the matrix **X** *consists of functional independent elements"*

$$\frac{\partial}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial}{\partial x_{11}} & \frac{\partial}{\partial x_{12}} \\ \frac{\partial}{\partial x_{21}} & \frac{\partial}{\partial x_{22}} \end{bmatrix}, \quad \frac{\partial}{\partial \mathbf{X}}\mathrm{tr}(\mathbf{AX}) = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} = \mathbf{A}'.$$

Case # 2: "the $n \times n$ matrix **X** *is symmetric: $\mathbf{X} = \mathbf{X}'$ "*

$$x_{12} = x_{21} \Rightarrow$$

$$\mathrm{tr}(\mathbf{AX}) = a_{11}x_{11} + (a_{12} + a_{21})x_{21} + a_{22}x_{22}$$

$$\frac{\partial}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial}{\partial x_{11}} & \frac{dx_{21}}{dx_{12}}\frac{\partial}{\partial x_{21}} \\ \frac{\partial}{\partial x_{21}} & \frac{\partial}{\partial x_{22}} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_{11}} & \frac{\partial}{\partial x_{21}} \\ \frac{\partial}{\partial x_{21}} & \frac{\partial}{\partial x_{22}} \end{bmatrix}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{tr}(\mathbf{AX}) = \begin{bmatrix} a_{11} & a_{12} + a_{21} \\ a_{12} + a_{21} & a_{22} \end{bmatrix} = \mathbf{A}' + \mathbf{A} - \mathrm{Diag}(a_{11}, \ldots, a_{nn}).$$

Case # 3: "the $n \times n$ matrix **X** *is antisymmetric: $\mathbf{X} = -\mathbf{X}'$"*

$$x_{11} = x_{22} = 0, \quad x_{12} = -x_{21} \Rightarrow \mathrm{tr}(\mathbf{AX}) = (a_{12} - a_{21})x_{21}$$

$$\frac{\partial}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial}{\partial x_{11}} & \frac{dx_{21}}{dx_{12}}\frac{\partial}{\partial x_{21}} \\ \frac{\partial}{\partial x_{21}} & \frac{\partial}{\partial x_{22}} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_{11}} & -\frac{\partial}{\partial x_{21}} \\ \frac{\partial}{\partial x_{21}} & \frac{\partial}{\partial x_{22}} \end{bmatrix}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{tr}(\mathbf{AX}) = \begin{bmatrix} 0 & -a_{12} + a_{21} \\ a_{12} - a_{21} & 0 \end{bmatrix} = \mathbf{A}' - \mathbf{A}.$$

Let us *now* assume that the matrix $\mathbf{X}$ of variables x_{ij} is *always* consisting of functionally independent elements. We note some useful identities o*f first derivatives.*

Scalar valued functions of vectors

$$\frac{\partial}{\partial \mathbf{x}'}(\mathbf{a}'\mathbf{x}) = \mathbf{a}' \tag{B1}$$

$$\frac{\partial}{\partial \mathbf{x}'}(\mathbf{x}'\mathbf{A}\mathbf{x}) = \mathbf{X}'(\mathbf{A}+\mathbf{A}'). \tag{B2}$$

Scalar-valued function of a matrix: trace

$$\frac{\partial\,\mathrm{tr}(\mathbf{AX})}{\partial \mathbf{X}} = \mathbf{A}'; \tag{B3}$$

especially:

$$\frac{\partial\,\mathbf{a}'\mathbf{X}\mathbf{b}}{\partial(\mathrm{vec}\mathbf{X})'} = \frac{\partial\,\mathrm{tr}(\mathbf{b}\mathbf{a}'\mathbf{X})}{\partial(\mathrm{vec}\mathbf{X})'} = \mathbf{b}'\otimes\mathbf{a}';$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{tr}(\mathbf{X}'\mathbf{A}\mathbf{X}) = (\mathbf{A}+\mathbf{A}')\mathbf{X}; \tag{B4}$$

especially:

$$\frac{\partial\,\mathrm{tr}(\mathbf{X}'\mathbf{X})}{\partial(\mathrm{vec}\mathbf{X})'} = 2(\mathrm{vec}\mathbf{X})'.$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{tr}(\mathbf{XAX}) = \mathbf{X}'\mathbf{A}' + \mathbf{A}'\mathbf{X}', \tag{B5}$$

especially:

$$\frac{\partial\,\mathrm{tr}\mathbf{X}^2}{\partial(\mathrm{vec}\mathbf{X})'} = 2(\mathrm{vec}\mathbf{X}')'.$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{tr}(\mathbf{A}\mathbf{X}^{-1}) = -(\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1}),\ \textit{if}\ \mathbf{X}\ \textit{is nonsingular}, \tag{B6}$$

especially:

$$\frac{\partial\,\mathrm{tr}(\mathbf{X}^{-1})}{\partial(\mathrm{vec}\mathbf{X})'} = -[\mathrm{vec}(\mathbf{X}^{-2})']';$$

$$\frac{\partial\,\mathbf{a}'\mathbf{X}^{-1}\mathbf{b}}{\partial(\mathrm{vec}\mathbf{X})'} = \frac{\partial\,\mathrm{tr}(\mathbf{b}\mathbf{a}'\mathbf{X}^{-1})}{\partial(\mathrm{vec}\mathbf{X})'} = -\mathbf{b}'(\mathbf{X}^{-1})'\otimes\mathbf{a}'\mathbf{X}^{-1}.$$

With permission from the author: Extract from "Erik W. Grafarend: Linear and nonlinear models. Fixed effects, random effects, and mixed models"

$$\frac{\partial}{\partial \mathbf{X}} \operatorname{tr}\mathbf{X}^{\alpha} = \alpha(\mathbf{X}')^{\alpha-1}, \text{ if } \mathbf{X} \text{ is quadratic}; \tag{B7}$$

especially:

$$\frac{\partial \operatorname{tr}\mathbf{X}}{\partial(\operatorname{vec}\mathbf{X})'} = (\operatorname{vec}\mathbf{I})'.$$

B3 Derivatives of Determinantal Forms

The scalarvalued forms of *matrix determinantal form* will be listed *now.*

$$\frac{\partial}{\partial \mathbf{X}} |\mathbf{AXB}'| = \mathbf{A}'(\operatorname{adj}\mathbf{AXB}')'\mathbf{B} = |\mathbf{AXB}'|\mathbf{A}'(\mathbf{BX}'\mathbf{A}')^{-1}\mathbf{B},$$

$$\text{if } \mathbf{AXB}' \text{ is nonsingular}; \tag{B8}$$

especially:

$$\frac{\partial \mathbf{a}'\mathbf{xb}}{\partial(\operatorname{vec}\mathbf{X})'} = \mathbf{b}' \otimes \mathbf{a}', \text{ where } \operatorname{adj}(\mathbf{a}'\mathbf{Xb}) = 1.$$

$$\frac{\partial}{\partial \mathbf{X}} |\mathbf{AXBX}'\mathbf{C}| = \mathbf{C}(\operatorname{adj}\mathbf{AXBX}'\mathbf{C})\mathbf{AXB} + \mathbf{A}'(\operatorname{adj}\mathbf{AXBX}'\mathbf{C})'\mathbf{CXB}'; \tag{B9}$$

especially:

$$\frac{\partial}{\partial \mathbf{X}} |\mathbf{XBX}'| = (\operatorname{adj}\mathbf{XBX}')\mathbf{XB} + (\operatorname{adj}\mathbf{XB}'\mathbf{X}')\mathbf{XB}';$$

$$\frac{\partial |\mathbf{XSX}'|}{\partial(\operatorname{vec}\mathbf{X})'} = 2(\operatorname{vec}\mathbf{X})'(\mathbf{S} \otimes \operatorname{adj}\mathbf{XSX}'), \text{ if } \mathbf{S} \text{ is } \textit{symmetric};$$

$$\frac{\partial |\mathbf{XX}'|}{\partial(\operatorname{vec}\mathbf{X})'} = 2(\operatorname{vec}\mathbf{X})'(\mathbf{I} \otimes \operatorname{adj}\mathbf{XX}').$$

$$\frac{\partial}{\partial \mathbf{X}} |\mathbf{AX}'\mathbf{BXC}| = \mathbf{BXC}(\operatorname{adj}\mathbf{AX}'\mathbf{BXC})\mathbf{A} + \mathbf{B}'\mathbf{XA}'(\operatorname{adj}\mathbf{AX}'\mathbf{BXC})'\mathbf{C}'; \tag{B10}$$

especially:

$$\frac{\partial}{\partial \mathbf{X}} |\mathbf{X}'\mathbf{BX}| = \mathbf{BX}(\operatorname{adj}\mathbf{X}'\mathbf{BX}) + \mathbf{B}'\mathbf{X}(\operatorname{adj}\mathbf{X}'\mathbf{B}'\mathbf{X});$$

$$\frac{\partial |\mathbf{X}'\mathbf{SX}|}{\partial(\operatorname{vec}\mathbf{X})'} = 2(\operatorname{vec}\mathbf{X})'(\operatorname{adj}\mathbf{X}'\mathbf{SX} \otimes \mathbf{S}), \text{ if } \mathbf{S} \text{ is } \textit{symmetric};$$

$$\frac{\partial |\mathbf{X}'\mathbf{X}|}{\partial(\operatorname{vec}\mathbf{X})'} = 2(\operatorname{vec}\mathbf{X})'(\operatorname{adj}\mathbf{X}'\mathbf{X} \otimes \mathbf{I}).$$

$$\frac{\partial}{\partial \mathbf{X}} |\mathbf{AXBXC}| = \mathbf{B}'\mathbf{X}'\mathbf{A}'(\text{adj}\mathbf{AXBXC})'\mathbf{C}' + \mathbf{A}'(\text{adj}\mathbf{AXBXC})'\mathbf{C}'\mathbf{X}'\mathbf{B}'; \quad \text{(B11)}$$

$$\frac{\partial}{\partial \mathbf{X}} |\mathbf{XBX}| = \mathbf{B}'\mathbf{X}'(\text{adj}\mathbf{XBX})' + (\text{adj}\mathbf{XBX})'\mathbf{XB}';$$

especially:

$$\frac{\partial |\mathbf{X}^2|}{\partial (\text{vec}\mathbf{X})'} = (\text{vec}[\mathbf{X}'\text{adj}(\mathbf{X}^2)' + \text{adj}(\mathbf{X}^2)'\mathbf{X}'])' =$$

$$= |\mathbf{X}|^2 \,(\text{vec}[\mathbf{X}'(\mathbf{X}')^{-2} + (\mathbf{X}')^{-2}\mathbf{X}'])' = 2\,|\mathbf{X}|^2\,[\text{vec}(\mathbf{X}^{-1})']', \text{ if } \mathbf{X} \text{ is non-singular .}$$

$$\frac{\partial}{\partial \mathbf{X}} |\mathbf{X}^\alpha| = \alpha\,|\mathbf{X}|^\alpha\,(\mathbf{X}^{-1})', \ \forall \alpha \in \mathbb{N} \text{ if } \mathbf{X} \text{ is non-singular ,} \quad \text{(B12)}$$

$$\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = |\mathbf{X}|\,(\mathbf{X}^{-1})' \text{ if } \mathbf{X} \text{ is non-singular;}$$

especially:

$$\frac{\partial |\mathbf{X}|}{\partial (\text{vec}\mathbf{X})'} = [\text{vec}(\text{adj}\mathbf{X}')]'.$$

B4 Derivatives of a Vector/Matrix Function of a Vector/Matrix

If we differentiate the vector or matrix valued function of a vector or matrix, we will find the results of type (B13) – (B20).

vector-valued function of a vector or a matrix

$$\frac{\partial}{\partial \mathbf{x}'} \mathbf{AX} = \mathbf{A} \quad \text{(B13)}$$

$$\frac{\partial}{\partial (\text{vec}\mathbf{X})'} \mathbf{AXa} = \frac{\partial (\mathbf{a}' \otimes \mathbf{A})\text{vec}\mathbf{X}}{\partial (\text{vec}\mathbf{X})'} = \mathbf{a}' \otimes \mathbf{A} \quad \text{(B14)}$$

matrix valued function of a matrix

$$\frac{\partial (\text{vec}\mathbf{X})}{\partial (\text{vec}\mathbf{X})'} = \mathbf{I}_{mp} \text{ for all } \mathbf{X} \in \mathbb{R}^{m \times p} \quad \text{(B15)}$$

$$\frac{\partial (\text{vec}\mathbf{X}')}{\partial (\text{vec}\mathbf{X})'} = \mathbf{K}_{m \cdot p} \text{ for all } \mathbf{X} \in \mathbb{R}^{m \times p} \quad \text{(B16)}$$

where $\mathbf{K}_{m \cdot p}$ is a suitable *commutation matrix*

$$\frac{\partial (\text{vec}\mathbf{XX}')}{\partial (\text{vec}\mathbf{X})'} = (\mathbf{I}_{m^2} + \mathbf{K}_{m \cdot m})(\mathbf{X} \otimes \mathbf{I}_m) \text{ for all } \mathbf{X} \in \mathbb{R}^{m \times p} \text{ ,}$$

where the matrix $\mathbf{I}_{m^2} + \mathbf{K}_{m \cdot m}$ is *symmetric and idempotent*,

$$\frac{\partial(\mathrm{vec}\mathbf{X}'\mathbf{X})}{\partial(\mathrm{vec}\mathbf{X})'} = (\mathbf{I}_{p^2} + \mathbf{K}_{p\cdot p})(\mathbf{I}_p \otimes \mathbf{X}') \text{ for all } \mathbf{X} \in \mathbb{R}^{m\times p}$$

$$\frac{\partial(\mathrm{vec}\mathbf{X}^{-1})}{\partial(\mathrm{vec}\mathbf{X})'} = -(\mathbf{X}^{-1})' \text{ if } \mathbf{X} \text{ is non-singular}$$

$$\frac{\partial(\mathrm{vec}\mathbf{X}^{\alpha})}{\partial(\mathrm{vec}\mathbf{X})'} = \sum_{j=1}^{\alpha} (\mathbf{X}')^{\alpha\text{-j}} \otimes \mathbf{X}^{j-1} \text{ for all } \alpha \in \mathbb{N}\text{, if } \mathbf{X} \text{ is a square matrix.}$$

B5 Derivatives of the Kronecker – Zehfuss product

Act a matrix-valued function of *two matrices* **X** and **Y** as variables be given. In particular, we assume the function

$$\mathbf{F}(\mathbf{X},\mathbf{Y}) = \mathbf{X} \otimes \mathbf{Y} \text{ for all } \mathbf{X} \in \mathbb{R}^{m\times p}, \quad \mathbf{Y} \in \mathbb{R}^{n\times q}$$

as the *Kronecker – Zehfuss product* of variables **X** and **Y** well defined. Then the identities of the *first differential and the first derivative follow:*

$$d\mathbf{F}(\mathbf{X},\mathbf{Y}) = (d\mathbf{X}) \otimes \mathbf{Y} + \mathbf{X} \otimes d\mathbf{Y},$$

$$d\mathrm{vec}\mathbf{F}(\mathbf{X},\mathbf{Y}) = \mathrm{vec}(d\mathbf{X} \otimes \mathbf{Y}) + \mathrm{vec}(\mathbf{X} \otimes d\mathbf{Y}),$$

$$\begin{aligned}\mathrm{vec}(d\mathbf{X} \otimes \mathbf{Y}) &= (\mathbf{I}_p \otimes \mathbf{K}_{qm} \otimes \mathbf{I}_n)\cdot(\mathrm{vec}d\mathbf{X} \otimes \mathrm{vec}\mathbf{Y}) = \\ &= (\mathbf{I}_p \otimes \mathbf{K}_{qm} \otimes \mathbf{I}_n)\cdot(\mathbf{I}_{mp} \otimes \mathrm{vec}\mathbf{Y})\cdot d(\mathrm{vec}\mathbf{X}) = \\ &= (\mathbf{I}_p \otimes [\mathbf{K}_{qm} \otimes \mathbf{I}_n)\cdot(\mathbf{I}_m \otimes \mathrm{vec}\mathbf{Y})])\cdot d(\mathrm{vec}\mathbf{X}),\end{aligned}$$

$$\begin{aligned}\mathrm{vec}(\mathbf{X} \otimes d\mathbf{Y}) &= (\mathbf{I}_p \otimes \mathbf{K}_{qm} \otimes \mathbf{I}_n)\cdot(\mathrm{vec}\mathbf{X} \otimes \mathrm{vecd}\mathbf{Y}) = \\ &= (\mathbf{I}_p \otimes \mathbf{K}_{qm} \otimes \mathbf{I}_n)\cdot(\mathrm{vec}\mathbf{X} \otimes \mathbf{I}_{nq})\cdot d(\mathrm{vec}\mathbf{Y}) = \\ &= ([(\mathbf{I}_p \otimes \mathbf{K}_{qm})\cdot(\mathrm{vec}\mathbf{X} \otimes \mathbf{I}_q)] \otimes \mathbf{I}_n)\cdot d(\mathrm{vec}\mathbf{Y}),\end{aligned}$$

$$\frac{\partial\,\mathrm{vec}(\mathbf{X} \otimes \mathbf{Y})}{\partial(\mathrm{vec}\mathbf{X})'} = \mathbf{I}_p \otimes [(\mathbf{K}_{qm} \otimes \mathbf{I} - n)\cdot(\mathbf{I}_m \otimes \mathrm{vec}\mathbf{Y})],$$

$$\frac{\partial\,\mathrm{vec}(\mathbf{X} \otimes \mathbf{Y})}{\partial(\mathrm{vec}\mathbf{Y})'} = (\mathbf{I}_p \otimes \mathbf{K}_{qm})\cdot(\mathrm{vec}\mathbf{X} \otimes \mathbf{I}_q)] \otimes \mathbf{I}_n \ .$$

B6 Matrix-valued Derivatives of Symmetric or Antisymmetric Matrix Functions

Many matrix functions $\mathbf{f}(\mathbf{X})$ or $\mathbf{F}(\mathbf{X})$ force us to pay attention to *dependencies within the variables.* As examples we treat here first derivatives of *symmetric or antisymmetric matrix functions* of **X**.

With permission from the author: Extract from "Erik W. Grafarend: Linear and nonlinear models. Fixed effects, random effects, and mixed models"

***Definition**: (derivative of a matrix-valued symmetric matrix function):*

Let $\mathbf{F(X)}$ be an $n \times q$ matrix-valued function of an $m \times m$ *symmetric* matrix $\mathbf{X = X'}$. The $nq \times m(m+1)/2$ *Jacobi matrix of first derivates of* $\mathbf{F}$ is defined by

$$\mathbf{J}_{\mathbf{F}}^{s} = \mathbf{DF(X = X')} := \frac{\partial \mathrm{vec}\mathbf{F(X)}}{\partial (\mathrm{vech}\mathbf{X})'}.$$

***Definition**: (derivative of matrix valued antisymmetric matrix function):*

Let $\mathbf{F(X)}$ be an $n \times q$ matrix-valued function of an $m \times m$ *anti-symmetric* matrix $\mathbf{X = -X'}$. The $nq \times m(m-1)/2$ *Jacobi matrix of first derivates of* $\mathbf{F}$ is defined by

$$\mathbf{J}_{\mathbf{F}}^{a} = \mathbf{DF(X = -X')} := \frac{\partial \mathrm{vec}\mathbf{F(X)}}{\partial (\mathrm{veck}\mathbf{X})'}.$$

Examples

(i) Given is a *scalar-valued matrixfunction* $\mathrm{tr}(\mathbf{AX})$ *of a symmetric variable matrix* $\mathbf{X = X'}$, for instance

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \ \mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}, \ \mathrm{vech}\mathbf{X} = \begin{bmatrix} x_{11} \\ x_{22} \\ x_{33} \end{bmatrix}$$

$$\mathrm{tr}(\mathbf{AX}) = a_{11}x_{11} + (\mathrm{a}_{12} + \mathrm{a}_{21})\mathrm{x}_{21} + a_{22}x_{22}$$

$$\frac{\partial}{\partial (\mathrm{vech}\mathbf{X})'} = [\frac{\partial}{\partial x_{11}}, \frac{\partial}{\partial x_{21}}, \frac{\partial}{\partial x_{22}}]$$

$$\frac{\partial \, \mathrm{tr}(\mathbf{AX})}{\partial (\mathrm{vech}\mathbf{X})'} = [a_{11}, a_{12} + a_{21}, a_{22}]$$

$$\frac{\partial \, \mathrm{tr}(\mathbf{AX})}{\partial (\mathrm{vech}\mathbf{X})'} = [\mathrm{vech}(\mathbf{A}' + \mathbf{A} - \mathrm{Diag}[a_{11}, \ldots, a_{nn}])]' = [\mathrm{vech} \frac{\partial \, \mathrm{tr}(\mathbf{AX})}{\partial \mathbf{X}}]'.$$

(ii) Given is *scalar-valued matrix function* $\mathrm{tr}(\mathbf{AX})$ of an *anti-symmetric variable matrix* $\mathbf{X} = -\mathbf{X}'$, for instance

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \ \mathbf{X} = \begin{bmatrix} 0 & -x_{21} \\ x_{21} & 0 \end{bmatrix}, \ \mathrm{veck}\mathbf{X} = x_{21},$$

$$\mathrm{tr}(\mathbf{AX}) = (\mathrm{a}_{12} - \mathrm{a}_{21})\mathrm{x}_{21}$$

$$\frac{\partial}{\partial(\text{veck}\mathbf{X})'} = \frac{\partial}{\partial x_{21}}, \quad \frac{\partial\, \text{tr}(\mathbf{AX})}{\partial(\text{veck}\mathbf{X})'} = a_{12} - a_{21},$$

$$\frac{\partial\, \text{tr}(\mathbf{AX})}{\partial(\text{veck}\mathbf{X})'} = [\text{veck}(\mathbf{A}' - \mathbf{A})]' = [\text{veck}\frac{\partial\, \text{tr}(\mathbf{AX})}{\partial \mathbf{X}}]'.$$

B7 Higher order derivatives

Up to now we computed only first derivatives of scalar-valued, vector-valued and matrix-valued functions. Second derivatives is our target now which will be needed for the classification of optimization problems of type minimum or maximum.

Definition: *(second derivatives of a scalar valued vector function):*

Let $f(\mathbf{x})$ a *scalar-valued function* of the $n \times 1$ vector $\mathbf{x}$. Then the $m \times m$ matrix

$$\mathbf{DD}'f(\mathbf{x}) = \mathbf{D}(\mathbf{D}f(\mathbf{x}))' := \frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}'}$$

denotes the *second derivatives* of $f(\mathbf{x})$ to $\mathbf{x}$ and $\mathbf{x}'$. Correspondingly

$$\mathbf{D}^2 f(\mathbf{x}) := \frac{\partial}{\partial \mathbf{x}'} \otimes \frac{\partial}{\partial \mathbf{x}} f(\mathbf{x}) = (\text{vec}\mathbf{DD}') f(\mathbf{x})$$

denotes the $\mathbf{1} \times m^2$ *vector of second derivatives*.

and

Definition: *(second derivative of a vector valued vector function):*

Let $\mathbf{f}(\mathbf{x})$ be an $n \times 1$ vector-valued function of the $m \times 1$ vector $\mathbf{x}$. Then the $n \times m^2$ *matrix of second derivatives*

$$\mathbf{H}_f = \mathbf{D}^2\mathbf{f}(\mathbf{x}) = \mathbf{D}(\mathbf{Df}(\mathbf{x})) =: \frac{\partial}{\partial \mathbf{x}'} \otimes \frac{\partial}{\partial \mathbf{x}} \mathbf{f}(x) = \frac{\partial^2 \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}' \partial \mathbf{x}}$$

is the *Hesse matrix of the function* $\mathbf{f}(\mathbf{x})$.

and

Definition: *(second derivatives of a matrix valued matrix function):*

Let $\mathbf{F(X)}$ be an $n \times q$ matrix valued function of an $m \times p$ matrix of *functional independent variables* $\mathbf{X}$. The $nq \times m^2 p^2$ *Hesse matrix of second derivatives of* ***F*** is defined by

$$\mathbf{H_F} = \mathbf{D}^2\mathbf{F(X)} = \mathbf{D}(\mathbf{DF(X)}) := \frac{\partial}{\partial(\text{vec}\mathbf{X})'} \otimes \frac{\partial}{\partial(\text{vec}\mathbf{X})'} \text{vec}\mathbf{F(X)} = \frac{\partial^2 \text{vec}\mathbf{F(X)}}{\partial(\text{vec}\mathbf{X})' \otimes \partial(\text{vec}\mathbf{X})'}.$$

The definition of *second derivatives of matrix functions* can be motivated as follows. The matrices $\mathbf{F} = [f_{ij}] \in \mathbb{R}^{n\times q}$ *and* $\mathbf{X} = [x_{k\ell}] \in \mathbb{R}^{m\times p}$ are the elements of a *two-dimensional* array. In contrast, the *array of second derivatives*

$$[\frac{\partial^2 f_{ij}}{\partial x_{k\ell}\partial x_{pq}}] = [k_{ijk\ell pq}] \in \mathbb{R}^{n\times q\times m\times p\times m\times p}$$

is six-dimensional and beyond the *common matrix algebra of two-dimensional arrays.* The following operations map a *six-dimensional array of second derivatives to a two-dimensional array*.

(i) vec$\mathbf{F(X)}$ is the *vectorized form* of the matrix valued function

(ii) vec$\mathbf{X}$ is the *vectorized form* of the variable matrix

(iii) the *Kronecker – Zehfuss product* $\frac{\partial}{\partial(\text{vec}\mathbf{X})'} \otimes \frac{\partial}{\partial(\text{vec}\mathbf{X})'}$ vectorizes the matrix of *second derivatives*

(iv) the formal product of the $\mathbf{1}\times m^2 p^2$ *row vector of second derivatives* with the $nq\times\mathbf{1}$ column vector vec$\mathbf{F(X)}$ leads to an $nq\times m^2 p^2$ *Hesse matrix of second derivatives.*

Again we assume the vector of variables $\mathbf{x}$ and the matrix of variables $\mathbf{X}$ consists of *functional independent elements*. If this is *not* the case we according to the *chain rule* must apply an *alternative differential calculus similary to the first deri-vative*, case studies of symmetric and antisymmetric variable matrices.

Examples:

$$\text{(i)} \quad f(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = a_{11}x_1^2 + (a_{12}+a_{21})x_1x_2 + a_{22}x_2^2$$

$$\mathbf{D}f(\mathbf{x}) = \frac{\partial f}{\partial \mathbf{x}'} = [2a_{11}x_1 + (a_{12}+a_{21})x_2 \,|\, (a_{12}+a_{21})x_1 + 2a_{22}x_2]$$

$$\mathbf{D}^2 f(\mathbf{x}) = \mathbf{D}(\mathbf{D}f(\mathbf{x}))' = \frac{\partial^2 f}{\partial\mathbf{x}\partial\mathbf{x}'} = \begin{bmatrix} 2a_{11} & a_{12}+a_{21} \\ a_{12}+a_{21} & 2a_{22} \end{bmatrix} = \mathbf{A}+\mathbf{A}'$$

$$\text{(ii)} \quad \mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}$$

$$\mathbf{Df}(\mathbf{x}) = \frac{\partial \mathbf{f}}{\partial \mathbf{x}'} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \mathbf{A}$$

$$\mathbf{DD}'\mathbf{f}(\mathbf{x}) = \frac{\partial^2 \mathbf{f}}{\partial\mathbf{x}\partial\mathbf{x}'} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad O(\mathbf{DD}'\mathbf{f}(\mathbf{x})) = 2\times 2$$

$$\mathbf{D}^2\mathbf{f}(\mathbf{x}) = [0\ 0\ 0\ 0], \quad O(\mathbf{D}^2\mathbf{f}(\mathbf{x})) = 1\times 4$$

(iii) $$\mathbf{F}(\mathbf{X}) = \mathbf{X}^2 = \begin{bmatrix} x_{11}^2 + x_{12}x_{21} & x_{11}x_{12} + x_{12}x_{22} \\ x_{21}x_{11} + x_{22}x_{21} & x_{21}x_{12} + x_{22}^2 \end{bmatrix}$$

$$\text{vec}\mathbf{F}(\mathbf{X}) = \begin{bmatrix} x_{11}^2 + x_{12}x_{21} \\ x_{21}x_{11} + x_{22}x_{21} \\ x_{11}x_{12} + x_{12}x_{22} \\ x_{21}x_{12} + x_{22}^2 \end{bmatrix}, \quad O(\mathbf{F}) = O(\mathbf{X}) = 2\times 2$$

$$(\text{vec}\mathbf{X})' = [x_{11}, x_{21}, x_{12}, x_{22}]$$

$$\mathbf{J}_\mathbf{F} = \frac{\partial\, \text{vec}\mathbf{F}(\mathbf{X})}{\partial (\text{vec}\mathbf{X})'} = \begin{bmatrix} 2x_{11} & x_{12} & x_{21} & 0 \\ x_{21} & x_{11}+x_{22} & 0 & x_{21} \\ x_{12} & 0 & x_{11}+x_{22} & x_{12} \\ 0 & x_{12} & x_{21} & 2x_{22} \end{bmatrix}$$

$$O(\mathbf{J}_\mathbf{F}) = 4\times 4$$

$$\mathbf{H}_\mathbf{F} = \frac{\partial}{\partial(\text{vec}\mathbf{X})'} \otimes \frac{\partial}{\partial(\text{vec}\mathbf{X})'} \text{vec}\mathbf{F}(\mathbf{X}) = [\frac{\partial}{\partial x_{11}}, \frac{\partial}{\partial x_{21}}, \frac{\partial}{\partial x_{12}}, \frac{\partial}{\partial x_{22}}] \otimes \mathbf{J}_\mathbf{F} =$$

$$= \left[\begin{array}{cccc|cccc|cccc|cccc} 2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 2 \end{array}\right]$$

$$O(\mathbf{H}_\mathbf{F}) = 4\times 16.$$

At the end, we want to define the derivative of order l of a matrix-valued matrix function whose structure is derived from the postulate of a suitable array.

Definition *(l-th derivative of a matrix-valued matrix function):*

Let $\mathbf{F}(\mathbf{X})$ be an $n\times q$ matrix valued function of an $m\times p$ matrix of functional independent variables $\mathbf{X}$. The $nq\times m^l p^l$ matrix of *l-th* derivative is defined by

$$\mathbf{D}^l\mathbf{F}(\mathbf{X}) := \frac{\partial}{\partial(\text{vec}\mathbf{X})'} \underbrace{\otimes \ldots \otimes}_{l\text{-times}} \frac{\partial}{\partial(\text{vec}\mathbf{X})'} \text{vec}\mathbf{F}(\mathbf{X}) =$$

$$= \frac{\partial^l}{\partial(\text{vec}\mathbf{X})' \underbrace{\otimes \ldots \otimes}_{l\text{-times}} (\text{vec}\mathbf{X})'} \text{vec}\mathbf{F}(\mathbf{X}) \text{ for all } l \in \mathbb{N}.$$

Appendix C: Lagrange Multipliers

?*How can we find extrema with side conditions*?

We generate solutions of such external problems first on the basis of *algebraic manipulations*, namely by the *lemma of implicit functions*, and secondly by a geometric tool box, by means of interpreting a *risk function and side conditions* as *level surfaces* (*specific normal images, Lagrange multipliers*).

C1 A first way to solve the problem

A first way to find extreme with side conditions will be based on a risk function

$$f(x_1, ..., x_m) = \text{extr} \tag{C1}$$

with unknowns $(x_1, ..., x_m) \in \mathbb{R}^m$, which are restricted by *side conditions* of type

$$\left[F_1(x_1, ..., x_m), F_2(x_1, ..., x_m), ..., F_r(x_1, ..., x_m)\right]' = 0 \tag{C2}$$

$$\text{rk}(\frac{\partial F_i}{\partial x_m}) = r < \text{m}. \tag{C3}$$

The *side conditions* $F_i(x_j)$ $(i = 1, ..., r,\ j = 1, ..., m)$ are reduced by the *lemma of the implicit function*: solve for

$$\begin{aligned} x_{m-r+1} &= G_1(x_1, ..., x_{m-r}) \\ x_{m-r+2} &= G_2(x_1, ..., x_{m-r}) \\ &... \\ x_{m-1} &= G_{r-1}(x_1, ..., x_{m-r}) \\ x_m &= G_r(x_1, ..., x_{m-r}) \end{aligned} \tag{C4}$$

and replace the result within the *risk function*

$$f(x_1, x_2, ..., x_{m-r}, G_1(x_1, ..., x_{m-r}), ..., G_r(x_1, ..., x_{m-1})) = \text{extr}. \tag{C5}$$

The "free" unknowns $(x_1, x_2, ..., x_{m-r-1}, x_{m-r}) \in \mathbb{R}^{m-r}$ can be found by taking the result of the *implicit function theorem* as follows.

Lemma C1 *("implicit function theorem"):*
Let $\mathbf{\Omega}$ be an open set of $\mathbb{R}^m = \mathbb{R}^{m-r} \times \mathbb{R}^r$ and $\mathbf{F}: \mathbf{\Omega} \to \mathbb{R}^r$ with vectors $\mathbf{x}_1 \in \mathbb{R}^{m-r}$ and $\mathbf{x}_2 \in \mathbb{R}^{m-r}$. The maps

---

$$(\mathbf{x}_1, \mathbf{x}_2) \mapsto \mathbf{F}(\mathbf{x}_1, \mathbf{x}_2) = \begin{bmatrix} F_1(x_1, ..., x_{m-r}; x_{m-r+1}, ..., x_m) \\ F_2(x_1, ..., x_{m-r}; x_{m-r+1}, ..., x_m) \\ ... \\ F_{r-1}(x_1, ..., x_{m-r}; x_{m-r+1}, ..., x_m) \\ F_r(x_1, ..., x_{m-r}; x_{m-r+1}, ..., x_m) \end{bmatrix} \tag{C6}$$

transform a continuously differential function with $\mathbf{F}(\mathbf{x}_1, \mathbf{x}_2) = 0$. In case of a *Jacobi determinant j not* zero or a *Jacobi matrix **J** of rank r*, or

$$j := \det \mathbf{J} \neq 0 \text{ or } \operatorname{rk} \mathbf{J} = r, \ \mathbf{J} := \frac{\partial(F_1, ..., F_r)}{\partial(x_{m-r+1}, ..., x_m)}, \tag{C7}$$

there exists a surrounding $\mathbf{U} := \mathbf{U}(\mathbf{x}_1) \subset \mathbb{R}^{m-r}$ and $\mathbf{V} := \mathbf{U}_\delta(\mathbf{x}_2) \subset \mathbb{R}^r$ such that the equation $\mathbf{F}(\mathbf{x}_1, \mathbf{x}_2) = 0$ for any $\mathbf{x}_1 \in \mathbf{U}$ *in* $\mathbf{V}'$ *has only one solution*

$$\mathbf{x}_2 = \mathbf{G}(\mathbf{x}_1) \ \textit{or} \begin{bmatrix} x_{m-r+1} \\ x_{m-r+2} \\ ... \\ x_{m-1} \\ x_m \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1(x_1, ..., x_{m-r}) \\ \mathbf{G}_2(x_1, ..., x_{m-r}) \\ ... \\ \mathbf{G}_{r-1}(x_1, ..., x_{m-r}) \\ \mathbf{G}_r(x_1, ..., x_{m-r}) \end{bmatrix}. \tag{C8}$$

The function $\mathbf{G} : \mathbf{U} \to \mathbf{V}$ is continuously differentiable.

A sample reference is any literature treating *analysis*, e.g. *C. Blotter* .

Lemma C1 is based on the Implicit Function Theorem whose result we insert within the *risk function* (C1) in order to gain (C5) in the *free* variables $(x_1, ..., x_{m-r}) \in \mathbb{R}^{m-r}$. Our example C1 explains the solution technique for finding *extreme with side conditions* within our *first approach*. *Lemma C1* illustrates that there exists a *local inverse of the side conditions* towards *r* unknowns $(x_{m-r+1}, x_{m-r+2}, ..., x_{m-1}, x_m) \in \mathbb{R}^r$ which in the case of *nonlinear side conditions* towards r unknowns $(x_{m-r+1}, x_{m-r+2}, ..., x_{m-1}, x_m) \in \mathbb{R}^r$ which in case of *nonlinear side conditions* is *not necessary unique*.

:Example C1:

Search for the *global extremum* of the function

$$f(x_1, x_2, x_3) = f(x, y, z) = x - y - z$$

subject to the side conditions

$$\begin{bmatrix} F_1(x_1, x_2, x_3) = Z(x, y, z) := x^2 + 2y^2 - 1 = 0 & \text{(elliptic cylinder)} \\ F_2(x_1, x_2, x_3) = E(x, y, z) := 3x - 4z = 0 & \text{(plane)} \end{bmatrix}$$

With permission from the author: Extract from "Erik W. Grafarend: Linear and nonlinear models. Fixed effects, random effects, and mixed models"

$$\mathbf{J} = (\frac{\partial F_i}{\partial x_j}) = \begin{bmatrix} 2x & 4y & 0 \\ 3 & 0 & -4 \end{bmatrix},\ \mathrm{rk}\,\mathbf{J}(x \neq 0 \text{ oder } y \neq 0) = r = 2$$

$$F_1(x_1, x_2, x_3) = Z(x, y, z) = 0 \Rightarrow \left[\begin{array}{l} {}_1 y = +\frac{1}{2}\sqrt{2}\sqrt{1-x^2} \\ {}_2 y = -\frac{1}{2}\sqrt{2}\sqrt{1-x^2} \end{array} \right.$$

$$F_2(x_1, x_2, x_3) = E(x, y, z) = 0 \Rightarrow z = \frac{3}{4}x$$

$${}_1 f(x_1, x_2, x_3) = {}_1 f(x, y, z) = f(x, +\frac{1}{2}\sqrt{2}\sqrt{1-x^2}, \frac{3}{4}) = \\ = \frac{x}{4} - \frac{1}{2}\sqrt{2}\sqrt{1-x^2}$$

$${}_2 f(x_1, x_2, x_3) = {}_2 f(x, y, z) = f(x, -\frac{1}{2}\sqrt{2}\sqrt{1-x^2}, \frac{3}{4}) \\ = \frac{x}{4} + \frac{1}{2}\sqrt{2}\sqrt{1-x^2}$$

$${}_1 f(x)' = 0 \Leftrightarrow \frac{1}{4} + \frac{1}{2}\sqrt{2}\frac{x}{\sqrt{1-x^2}} = 0 \Leftrightarrow {}_1 x = -\frac{1}{3}$$

$${}_2 f(x)' = 0 \Leftrightarrow \frac{1}{4} + \frac{1}{2}\sqrt{2}\frac{x}{\sqrt{1-x^2}} = 0 \Leftrightarrow {}_2 x = +\frac{1}{3}$$

$${}_1 f(-\frac{1}{3}) = -\frac{3}{4} \ (\textit{minimum}),\ {}_2 f(\frac{1}{3}) = +\frac{3}{4} \ (\textit{maximum}).$$

At the position $x = -1/3,\ y = 2/3,\ z = -1/4$ we find a *global minimum*, but at the position $x = +1/3,\ y = -2/3,\ z = -1/4$ a *global maximum*.

An *alternative path* to find *extreme with side conditions* is based on the *geometric interpretation* of *risk function and side conditions*. *First*, we form the conditions

$$\left. \begin{array}{c} F_1(x_1, \ldots, x_m) = 0 \\ F_2(x_1, \ldots, x_m) = 0 \\ \cdots \\ F_r(x_1, \ldots, x_m) = 0 \end{array} \right] \quad \mathrm{rk}(\frac{\partial F_i}{\partial x_j}) = r$$

by continuously differentiable *real functions on an open set* $\mathbf{\Omega} \subset \mathbb{R}^m$. *Then* we define r equations $F_i(x_1, ..., x_m) = 0$ for all $i = 1, ..., r$ with the rank conditions $\mathrm{rk}(\partial F_i / \partial x_j) = r$, *geometrically an* ($m$-1) *dimensional surface* $\mathbb{M}_F \subset \mathbf{\Omega}$ which can be seen as a *level surface*. See as an example our *Example C1* which describe as *side conditions*

$$F_1(x_1,x_2,x_3) = Z(x,y,z) = x^2 + 2y^2 - 1 = 0$$
$$F_2(x_1,x_2,x_3) = E(x,y,z) = 3x - 4z = 0$$

representing an *elliptical cylinder* and *a plane*. In this case is the (*m-r*) dimensional surface $\mathbb{M}_F$ the *intersection manifold of the elliptic cylinder and of the plane* as the *m-r* =1 dimensional manifold in $\mathbb{R}^3$, namely as "*spatial curve*". *Secondly*, the risk function $f(x_1,...,x_m) = \text{extr}$ generates an (*m*-1) dimensional surface $\mathbb{M}_f$ which is a *special level surface*. The *level parameter* of the (*m*-1) dimensional surface $\mathbb{M}_f$ should be external. In our *Example C1* one risk function can be interpreted as the *plane*

$$f(x_1,x_2,x_3) = f(x,y,z) = x - y - z\,.$$

We summarize our result within *Lemma C2*.

Lemma C2 *(extrema with side conditions)*
The side conditions $F_i(x_1,\dots,x_m) = 0$ for all $i \in \{1,\dots,r\}$ are built on continuously differentiable functions on an *open* set $\mathbf{\Omega} \subset \mathbb{R}^m$ which are subject to the side conditions $\text{rk}(\partial F_i / \partial x_j) = r$ generating an (*m-r*) dimensional *level surface* $\mathbb{M}_f$. The function $f(x_1,\dots,x_m)$ produces certain constants, namely an (*m*-1) dimensional *level surface* $\mathbb{M}_f$. $f(x_1,\dots,x_m)$ is geometrically as a *point* $p \in \mathbb{M}_F$ *conditionally extremal* (stationary) if and only if the (*m*-1) dimensional *level surface* $\mathbb{M}_f$ is *in contact* to the (*m-r*) dimensional *level surface* in *p*. That is there exist numbers $\lambda_1,\dots,\lambda_r$, the *Lagrange multipliers*, by

$$\mathbf{grad}\, f(p) = \sum\nolimits_{i=1}^{r} \lambda_i\, \mathbf{grad}\, F_i(p).$$

The unnormalized surface normal vector $\mathbf{grad}\, f(p)$ of the (*m*-1) dimensional level surface $\mathbb{M}_f$ in the normal space $\mathbb{N}\mathbb{M}_F$ of the level surface $\mathbb{M}_F$ *is in* the unnormalized surface normal vector $\mathbf{grad}\, F_i(p)$ in the point p. To this *equation* belongs the *variational problem*

$$\mathcal{L}(x_1,\dots,x_m;\lambda_1,\dots,\lambda_r) =$$
$$f(x_1,\dots,x_m) - \sum\nolimits_{i=1}^{r} \lambda_i F_i(x_1,\dots,x_m) = \text{extr}\,.$$

:proof:

First, the side conditions $F_i(x_j) = 0$, $\text{rk}(\partial F_i / \partial x_j) = r$ for all $i = 1,\dots,r$; $j = 1,\dots,m$ generate an (*m-r*) dimensional *level surface* $\mathbb{M}_F$ whose *normal vectors*

$$\mathbf{n}_i(p) := \mathbf{grad}\, F_i(p) \in \mathbb{N}_p\mathbb{M}_F \quad (i = 1,\dots,r)$$

span the *r* dimensional normal space $\mathbb{N}\mathbb{M}$ of the level surface $\mathbb{M}_F \subset \mathbf{\Omega}$. The r dimensional normal space $\mathbb{N}_p\mathbb{M}_F$ of the (*m-r*) dimensional level surface $\mathbb{M}_F$

---

is orthogonal complement $\mathbb{T}_p\mathbb{M}_p$ to the tangent space $\mathbb{T}_p\mathbb{M}_F \subset \mathbb{R}^{m-1}$ of $\mathbb{M}_F$ in the point *p* spanned by the *m-r* dimensional tangent vectors

$$\mathbf{t}_k(p) := \left.\frac{\partial \mathbf{x}}{\partial \mathbf{x}_k}\right|_{\mathbf{x}=p} \in \mathbb{T}_p\mathbb{M}_F \qquad (k = 1, \ldots, m-r).$$

:Example C2:

Let the $m-r=2$ dimensional *level surface* $\mathbb{M}_F$ of the *sphere* $\mathbb{S}_r^2 \in \mathbb{R}^3$ of radius r (“*level parameter* r^2 ”) be given by the side condition

$$F(x_1, x_2, x_3) = x_1^{\ 2} + x_2^{\ 2} + x_3^{\ 2} - r^2 = 0.$$

:Normal space:

$$\mathbf{n}(p) = \mathbf{grad}\, F(p) = \mathbf{e}_1 \frac{\partial F}{\partial x_1} + \mathbf{e}_2 \frac{\partial F}{\partial x_2} + \mathbf{e}_3 \frac{\partial F}{\partial x_3} = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3] \begin{bmatrix} 2x_1 \\ 2x_2 \\ 2x_3 \end{bmatrix}_p .$$

The *orthogonal vectors* $[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]$ span $\mathbb{R}^3$. The normal space will be generated locally by *a normal vector* $\mathbf{n}(p) = \mathbf{grad}\, F(p)$.

:Tangent space:

The *implicit representation* is the characteristic element of the *level surface*. In order to gain an *explicit representation*, we take advantage of the *Implicit Function Theorem* according to the following equations.

$$\left.\begin{matrix} F(x_1, x_2, x_3) = 0 \\ \mathrm{rk}(\dfrac{\partial F}{\partial x_j}) = r = 1 \end{matrix}\right] \Rightarrow x_3 = G(x_1, x_2)$$

$$x_1^{\ 2} + x_2^{\ 2} + x_3^{\ 2} - r = 0 \quad \textit{and} \quad (\frac{\partial F}{\partial x_j}) = [2x_1 + 2x_2 + 2x_3],\ \mathrm{rk}(\frac{\partial F}{\partial x_j}) = 1$$

$$\Rightarrow x_j = G(x_1, x_2) = +\sqrt{r^2 - (x_1^{\ 2} + x_2^{\ 2})}\,.$$

The *negation root* leads into another domain of the sphere: here holds the domain $0 < x_1 < r,\ 0 < x_2 < r,\ r^2 - (x_1^{\ 2} + x_2^{\ 2}) > 0.$

The *spherical position vector* $\mathbf{x}(p)$ allows the representation

$$\mathbf{x}(p) = \mathbf{e}_1 x_1 + \mathbf{e}_2 x_2 + \mathbf{e}_3 \sqrt{r^2 - (x_1^{\ 2} + x_2^{\ 2})}\ ,$$

which is the basis to produce

$$\left[\begin{array}{l} \mathbf{t}_1(p)=\dfrac{\partial \mathbf{x}}{\partial x_2}(p)=\mathbf{e}_1-\mathbf{e}_3\dfrac{x_1}{\sqrt{r^2-(x_1^2+x_2^2)}}=[\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3]\begin{bmatrix}1\\0\\-\frac{x_1}{\sqrt{r^2-(x_1^2+x_2^2)}}\end{bmatrix}\\ \mathbf{t}_1(p)=\dfrac{\partial \mathbf{x}}{\partial x_2}(p)=\mathbf{e}_2-\mathbf{e}_3\dfrac{x_2}{\sqrt{r^2-(x_1^2+x_2^2)}}=[\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3]\begin{bmatrix}0\\1\\-\frac{x_2}{\sqrt{r^2-(x_1^2+x_2^2)}}\end{bmatrix}\end{array}\right],$$

which span the *tangent space* $\mathbb{T}_p\mathbb{M}_F=\mathbb{R}^2$ at the point p.

:The general case:

In the general case of an $(m-r)$ dimensional *level surface* $\mathbb{M}_F$, *implicitly produced* by r side conditions of type

$$\left.\begin{array}{c} F_1(x_1,...,x_m)=0\\ F_2(x_1,...,x_m)=0\\ ...\\ F_{r-j}(x_1,...,x_m)=0\\ F_r(x_1,...,x_m)=0\end{array}\right] \mathrm{rk}\,(\frac{\partial F_i}{\partial x_j})=r,$$

the *explicit surface representation*, produced by the *Implicit Function Theorem*, reads

$$\mathbf{x}(p)=\mathbf{e}_1x_1+\mathbf{e}_2x_2+...+\mathbf{e}_{m-r}x_{m-r}+\mathbf{e}_{m-r+1}G_1(x_1,...,x_{m-r})+...+\mathbf{e}_mG_r(x_1,...,x_{m-r}).$$

The orthogonal vectors $[\mathbf{e}_1,...,\mathbf{e}_m]$ span $\mathbb{R}^m$.

Secondly, the at least once conditional differentiable risk function $f(x_1,...,x_m)$ for special constants describes an $(m-1)$ dimensional *level surface* $\mathbb{M}_F$ whose *normal vector*

$$\mathbf{n}_f := \mathbf{grad}\, f(p)\subset \mathbb{N}_p\mathbb{M}_f$$

spans an one-dimensional normal space $\mathbb{N}_p\mathbb{M}_f$ of the level surface $\mathbb{M}_f\subset\mathbf{\Omega}$ in the point p. The *level parameter* of the *level surface* is chosen in the extremal case that *it touches the level surface* $\mathbb{M}_f$ *the other level surface* $\mathbb{M}_F$ *in the point* p. That means that the *normal vector* $\mathbf{n}_f(p)$ in the point *p* is an element of the normal space $\mathbb{N}_p\mathbb{M}_f$. Or we may say the normal vector $\mathbf{grad}\, f(p)$ is a *linear combination of the normal vectors* $\mathbf{grad}\, F_i(p)$ *in the point p*,

$$\mathbf{grad}\, f(p)=\sum_{i=1}^r\lambda_i\,\mathbf{grad}\, F_i(p)\ \text{ for all }\ i=1,...,r,$$

where the *Lagrange multipliers* λ_i are the coordinates of the vector $\mathbf{grad}\, f(p)$ in the basis $\mathbf{grad}\, F_i(p)$.

With permission from the author: Extract from "Erik W. Grafarend: Linear and nonlinear models. Fixed effects, random effects, and mixed models"

:Example C3:

Let us assume that there will be given the point $\mathbf{X} \in \mathbb{R}^3$. Unknown is the point in the $m - r = 2$ dimensional *level surface* $\mathbb{M}_F$ of *type sphere* $\mathbb{S}_r^{\ 2} = \mathbb{R}^3$ which is from the point $\mathbf{X} \in \mathbb{R}^3$ at *extremal distance*, *either* minimal *or* maximal.

The distance function $\|\mathbf{X} - \mathbf{x}\|^2$ for $\mathbf{X} \in \mathbb{R}^3$ *and* $\mathbf{X} \in \mathbb{S}_r^{\ 2}$ describes the *risk function*

$$f(x_1, x_2, x_3) = (X_1 - x_1)^2 + (X_2 - x_2)^2 + (X_3 - x_3)^2 = \mathbb{R}^2 = \underset{x_1, x_2, x_3}{\operatorname{extr}} ,$$

which represents an $m - 1 = 2$ dimensional *level surface* $\mathbb{M}_f$ of *type sphere* $\mathbb{S}_r^{\ 2} \in \mathbb{R}^3$ at the origin (X_1, X_2, X_3) and *level parameter* R^2. The *conditional extremal problem is solved if the sphere* $\mathbb{S}_R^{\ 2}$ *touches the other sphere* $\mathbb{S}_r^{\ 2}$. This result is expressed in the *language of the normal vector*.

$$\mathbf{n}(p) := \mathbf{grad}\ f(p) = \mathbf{e}_1 \frac{\partial f}{\partial x_1} + \mathbf{e}_2 \frac{\partial f}{\partial x_2} + \mathbf{e}_3 \frac{\partial f}{\partial x_3} =$$

$$= [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3] \begin{bmatrix} -2(X_1 - x_1) \\ -2(X_2 - x_2) \\ -2(X_3 - x_3) \end{bmatrix}_p \in \mathbb{N}_p \mathbb{M}_f$$

$$\mathbf{n}(p) := \mathbf{grad}\, F(p) = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3] \begin{bmatrix} 2x_1 \\ 2x_2 \\ 2x_3 \end{bmatrix}$$

is an element of the normal space $\mathbb{N}_p \mathbb{M}_f$.
The normal equation

$$\mathbf{grad}\ f(p) = \lambda\ \mathbf{grad}\ F(p)$$

leads directly to *three equations*

$$x_i - X_0 = \lambda x_i \Leftrightarrow x_i(1 - \lambda) = X_i \quad (i = 1, 2, 3) ,$$

which are completed by the *fourth equation*

$$F(x_1, x_2, x_3) = x_1^{\ 2} + x_2^{\ 2} + x_3^{\ 2} - r^2 = 0.$$

Lateron we solve the 4 equations.

Third, we interpret the differential equations

$$\mathbf{grad}\ f(p) = \sum_{i=1}^{r} \lambda_i\ \mathbf{grad}\ F_i(p)$$

by the *variational problem*, *by direct differentiation* namely

---

With permission from the author: Extract from "Erik W. Grafarend: Linear and nonlinear models. Fixed effects, random effects, and mixed models"

$$\mathcal{L}(x_1,...,x_m;\lambda_1,...,\lambda_r) =$$
$$= f(x_1,...,x_m) - \sum_{i=1}^{r} \lambda_i F_i(x_1,...,x_m) = \underset{x_1,...,x_m;\lambda_1,...,\lambda_r}{\text{extr}}$$
$$\left[\begin{array}{ll} \dfrac{\partial \mathcal{L}}{\partial x_i} = \dfrac{\partial f}{\partial x_j} - \sum_{i=1}^{r} \lambda_i \dfrac{\partial F_i}{\partial x_j} = 0 & (j = 1,...,m) \\ -\dfrac{\partial \mathcal{L}}{\partial x_k} = F_i(x_j) = 0 & (i = 1,...,r). \end{array}\right.$$

:Example C4:

We continue our *third example* by solving the alternative system of equations.

$$\mathcal{L}(x_1,x_2,x_3;\lambda) = (X_1 - x_1)^2 + (X_2 - x_2)^2 + (X_3 - x_3)$$
$$-\lambda({x_1}^2 + {x_2}^2 + {x_3}^2 - r^2) = \underset{x_1,x_2,x_3;\lambda}{\text{extr}}$$

$$\left.\begin{array}{l} \dfrac{\partial \mathcal{L}}{\partial x_j} = -2(X_j - x_j) - 2\lambda x_j = 0 \\ -\dfrac{\partial \mathcal{L}}{\partial \lambda} = {x_1}^2 + {x_2}^2 + {x_3}^2 - r^2 = 0 \end{array}\right] \Rightarrow$$

$$\left.\begin{array}{l} x_1 = \dfrac{X_1}{1-\lambda}\ ;\ x_2 = \dfrac{X_2}{1-\lambda} \\ {x_1}^2 + {x_2}^2 + {x_3}^2 - r^2 = 0 \end{array}\right] \Rightarrow$$

$$\frac{{X_1}^2 + {X_2}^2 + {X_3}^2}{(1-\lambda)^2} - r^2 = 0 \Rightarrow -(1-\lambda)^2 r^2 + {X_1}^2 + {X_2}^2 + {X_3}^2 = 0 \Leftrightarrow$$

$$\Leftrightarrow (1-\lambda)^2 = \frac{{X_1}^2 + {X_2}^2 + {X_3}^2}{r^2} \Leftrightarrow 1 - \lambda_{1,2} = \pm\frac{1}{r}\sqrt{{X_1}^2 + {X_2}^2 + {X_3}^2} \Leftrightarrow$$

$$\lambda_{1,2} = 1 \pm \frac{1}{r}\sqrt{{X_1}^2 + {X_2}^2 + {X_3}^2} = \frac{r \pm \sqrt{{X_1}^2 + {X_2}^2 + {X_3}^2}}{r}$$
$$(x_1)_{1,2} = \pm\frac{rX_1}{\sqrt{{X_1}^2 + {X_2}^2 + {X_3}^2}},$$
$$(x_2)_{1,2} = \pm\frac{rX_2}{\sqrt{{X_1}^2 + {X_2}^2 + {X_3}^2}},$$
$$(x_3)_{1,2} = \pm\frac{rX_3}{\sqrt{{X_1}^2 + {X_2}^2 + {X_3}^2}}.$$

The matrix of second *derivatives* **H** decides upon whether at the point $(x_1, x_2, x_3, \lambda)_{1,2}$ we enjoy a *maximum or minimum*.

$$\mathbf{H} = (\frac{\partial^2 \mathcal{L}}{\partial x_j x_k}) = (\delta_{jk}(1-\lambda)) = (1-\lambda)\mathbf{I}_3$$

$$\left.\begin{matrix} \mathbf{H}(1-\lambda>0)>0\,(minimum) \\ (x_1,x_2,x_3)\text{ is the point of } minimum \end{matrix}\right] \left[\begin{matrix} \mathbf{H}(1-\lambda<0)<0\,(maximum) \\ (x_1,x_2,x_3)\text{ is the point of } maximum. \end{matrix}\right.$$

Our example illustrates how we can find the global optimum under side conditions by means of the technique of Lagrange multipliers.

:Example C5:

Search for the *global extremum* of the function $f(x_1,x_2,x_3)$ subject to *two side conditions* $F_1(x_1,x_2,x_3)$ and $F_2(x_1,x_2,x_3)$, namely

$$f(x_1,x_2,x_3) = f(x,y,z) = x-y-z \quad \text{(plane)}$$

$$\left[\begin{matrix} F_1(x_1,x_2,x_3) = Z(x,y,z) := x^2+2y^2-1=0 & \text{(elliptic cylinder)} \\ F_2(x_1,x_2,x_3) = E(x,y,z) := 3x-4z=0 & \text{(plane)} \end{matrix}\right.$$

$$\mathbf{J} = (\frac{\partial F_i}{\partial x_j}) = \begin{bmatrix} 2x & 4y & 0 \\ 3 & 0 & -4 \end{bmatrix}, \ \text{rk}\,\mathbf{J}(x \neq 0 \ \text{ oder } \ y \neq 0) = r = 2.$$

:Variational Problem:

$$\mathcal{L}(x_1,x_2,x_3;\lambda_1,\lambda_2) = \mathcal{L}(x,y,z;\lambda,\mu)$$
$$= x-y-z-\lambda(x^2+2y^2-1)-\mu(3x-4z) = \underset{x_1,x_2,x_3;\lambda,\mu}{\text{extr}}$$

$$\left.\begin{matrix} \frac{\partial \mathcal{L}}{\partial x} = 1-2\lambda x-3\mu=0 \\ \frac{\partial \mathcal{L}}{\partial y} = -1-4\lambda y=0 \Rightarrow \lambda = -\frac{1}{4y} \\ \frac{\partial \mathcal{L}}{\partial z} = -1-4\mu=0 \Rightarrow \lambda = -\frac{1}{4} \\ -\frac{\partial \mathcal{L}}{\partial \lambda} = x^2+2y^2-1=0 \\ -\frac{\partial \mathcal{L}}{\partial \mu} = 3x-4z=0. \end{matrix}\right] \Rightarrow$$

We multiply the *first equation* $\partial\mathcal{L}/\partial x$ by $4y$, the *second equation* $\partial\mathcal{L}/\partial y$ by $(-2x)$ and the *third equation* $\partial\mathcal{L}/\partial z$ by 3 and *add* !

$$4y-8\lambda xy-12\mu y+2x+8\lambda xy-3y+12\mu y = y+2x=0\,.$$

Replace in the *cylinder equation* (first side condition) $Z(x, y, z) = x^2 + 2y^2 - 1 = 0$, that is $x_{1,2} = \pm 1/3$. From the *second condition of the plane* (second side condition) $E(x, y, z) = 3x - 4z = 0$ we gain $z_{1,2} = \pm 1/4$. As a result we find $x_{1,2}, z_{1,2}$ and finally $y_{1,2} = \mp 2/3$.

The matrix of *second derivatives* **H** decides upon whether at the point $\lambda_{1,2} = \mp 3/8$ we find a *maximum or minimum.*

$$\mathbf{H} = \left(\frac{\partial^2 \mathcal{L}}{\partial x_j x_k}\right) = \begin{bmatrix} -2\lambda & 0 & 0 \\ 0 & -4\lambda & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{H}(\lambda_1 = -\frac{3}{8}) = \begin{bmatrix} \frac{3}{4} & 0 & 0 \\ 0 & \frac{3}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \geq 0$$

(*minimum*)

$(x, y, z; \lambda, \mu)_1 = (\frac{1}{3}, -\frac{2}{3}, \frac{1}{4}; -\frac{3}{8}, \frac{1}{4})$

is the restricted *minmal* solution point.

$$\mathbf{H}(\lambda_2 = \frac{3}{8}) = \begin{bmatrix} -\frac{3}{4} & 0 & 0 \\ 0 & -\frac{3}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \leq 0$$

(*maximum*)

$(x, y, z; \lambda, \mu)_2 = (-\frac{1}{3}, \frac{2}{3}, -\frac{1}{4}; \frac{3}{8}, \frac{1}{4})$

is the restricted *maximal* solution point.

The geometric interpretation of the *Hesse matrix* follows from *E. Grafarend and P. Lohle* (1991).

The matrix of second derivatives **H** decides upon whether at the point $(x_1, x_2, x_3, \lambda)_{1,2}$ we enjoy a *maximum or minimum*.