Numerical Linear Algebra Tools

Prof. Khalide Jbilou

Université du Littoral Côte d'Opale,
France

Ecole Centrale Casablanca 2022/2023

The Outline

- Basic linear algebra tools.
- The Singular Value Decomposition (SVD).
- QR decomposition (via Householder, Givens and Gram-Schmidt)
- Least-Squares problems via QR and SVD
- Fast Fourier Transform (DFT-FFT)
- Tensors
- Applications: Image processing (Compression, Restoration,...); Face recognition (PCA, LDA); Graphs in Network Analysis

Matrix operations and special matrices

Classical matrix operations A matrix is a rectangular array of complex or real numbers with m rows and n columns given as

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{1n} \end{pmatrix}.$$

The a_{ij}'s are called coefficients or elements of the matrix A. The coefficient is located a_{ij} is in the intersection of the row i and the column j of the matrix A. This matrix can also be denoted as $A = [a_{ij}]_{1 \leq i \leq n, 1 \leq j \leq m}$. The first mathematicians how used the term 'matrix' were J. Sylvester [1] and A. Cayley [2] in the middle of nineties.

[1]James Joseph Sylvester: 3 September 1814 ?- 15 March 1897, was an English mathematician. He made fundamental contributions to matrix theory, invariant theory, number theory, partition theory, and combinatorics.

[2]Arthur Cayley: 16 August 1821–26 January 1895, was a British mathematician who worked mostly on algebra. He helped found the modern British school of pure mathematics.

The classical matrix operations (addition and multiplication) are defined as follows.
The sum of two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ of the same sizes is the matrix $C = [c_{ij}]$ whose elements are

$$c_{ij} = a_{ij} + b_{ij}.$$

For the classical matrix product, let $A = [a_{ij}]$ be a matrix in $\mathbb{C}^{n\times m}$ (whose ij-th element is a_{ij}) and let $B = [b_{ij}]$ be a matrix in $\mathbb{C}^{m\times p}$, then the product $C = AB$ is defined by $C = [c_{ij}] \in \mathbb{C}^{n\times n\times p}$ such that

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj},\ i = 1, \ldots, n;\ j = 1, \ldots, m.$$

Notice that this matrix multiplication is associative and distributive

$$A(BC) = (AB)C,\ A(B + C) = AB + AC,$$

but not commutative; in general $AB \neq BA$ although if the two matrix products are possible.

In matrix computation, it is useful to define array operations obtained element by element on vectors and matrices. Using the convention in Matlab, we define the array multiplication by

$$C = A.*B,\ c_{ij} = a_{ij}b_{ij},$$

if A and B have the same dimensions. This product is also called the Hadamard product.

Consider now tow bloc-matrices

$$A = \begin{pmatrix} A_{11} & \dots & A_{1s} \\ \vdots & \ddots & \vdots \\ A_{q1} & \dots & A_{qs} \end{pmatrix} \text{ and } B = \begin{pmatrix} B_{11} & \dots & B_{1r} \\ \vdots & \ddots & \vdots \\ B_{s1} & \dots & B_{sr} \end{pmatrix},$$

where the block A_{ij} is of size $m_i \times p_j$ and the block B_{ij} is of size $p_i \times n_j$. Then we can also define the matrix product $C = AB$ can be done using the block structures of the matrices A and B as follows. The ij-block C_{ij} of the matrix C is of size $m_i \times n_j$ and defined by

$$C_{ij} = \sum_{k=1}^{s} A_{ik}B_{kj},\ i = 1,\dots,q;\ j = 1,\dots,r$$

Special matrices

The transpose of the matrix $A = [a_{ij}]$ of size $n \times m$ is the $m \times n$ matrix $A^T = [a_{ji}]$ obtained by transposing the rows and the columns of A. The transpose conjugate is defined by

$$A^H = \bar{A}^T$$

where the bar denotes the complex conjugate. We notice that

$$(A+B)^H = A^H + B^H, \;\; (AB)^H = B^H A^H$$

where A and B are complex matrices of appropriate sizes The ith row of the matrix A is defined by $a_{i,:} = (a_{i1}, a_{i2}, \ldots, a_{im})$ The jth column of A is defined by

$$a_{:,j} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix}.$$

Matrices are related to linear mapping between vector spaces. Let f be a linear mapping from two vector spaces E onto F with dimension m and n, Matrices are related to linear mapping between vector spaces. Let f be a linear mapping from two vector spaces E onto F with dimension m and n, respectively. Consider two bases $B_1 = \{v_1, \dots, v_m\}$ in E and $B_2 = \{w_1, \dots, w_n\}$ in F. Then

$$A = Mat_{B_1,B_2} f$$

where the j-th column of A is expressed as
$a_{:,j} = f(v_j) = a_{1j} w_1 + \ldots + a_{nj} w_n$, $j = 1, \dots, m$.

Special matrices

In matrix computation, we usually use the following particular matrices.

Definition

The matrix A is

1. Hermitian if $A^H = A$ and skew-Hermitian if $A^H = -A$.
2. Symmetric if $A^T = A$ and skew-symmetric if $A^T = -A$.
3. Normal if $A^H A = AA^H$.
4. Nonnegative if $a_{ij} \geq 0$ for all i, j.
5. Unitary if $A^H A = AA^H = I$ and A is said to be orthogonal if A is not necessary square and $A^H A = I$. For any integer n, the set of all $n \times n$ unitary matrices with matrix multiplication forms a group, called the unitary group $\mathcal{U}_n$.
6. Diagonal if and only if $a_{ij} = 0$ if $i \neq j$ and we denote $A = \mathrm{diag}(a_{11}, \ldots, a_{nn})$.
7. Upper triangular if and only if $a_{ij} = 0$ for $i > j$.
8. Lower triangular if $a_{ij} = 0$ for $i < j$.

Definition

1. Tridiagonal if $a_{ij} = 0$ for i, j such that $|i - j| > 1$.
2. Permutation if the columns (or rows) of A are a permutation of the identity matrix. When multiplying from the left a matrix A as PA results in permuting the rows of A and the product AP results in permuting the columns of A.
3. A matrix is sparse if a large fraction of its entries are zero. An important special case is the band matrix. We say that A has lower bandwidth p if $a_{ij} = 0$ whenever $i > j + p$ and upper bandwidth q if $j > i + q$ implies $a_{ij} = 0$. Such a matrix can be stored in a $(p + q + 1)$-by-n array.
4. A Topeplitz matrix T is a square matrix whose each ascending diagonal is constant.
5. A Hankel matrix H is a square matrix in which each ascending skew-diagonal from left to right is constant.

Range space, null space and matrix inversion

Let $\{u_1, u_2, \ldots, u_p\}$ be p vectors in $\mathbb{R}^n$, the subspace generated by these vectors is defined and denoted as follows

$$span\{u_1, \ldots, u_p\} = \{\alpha_1 u_1 + \ldots + \alpha_p u_p,\ \alpha_i \in \mathbb{R}\} \subset \mathbb{R}^n.$$

There are two important subspaces that are associated with the matrix A.

Definition

1. The **range** of $A \in \mathbb{R}^{n\times m}$ is defined by

$$\text{range}(A) = \{y \in \mathbb{R}^n :\ \exists x \in \mathbb{R}^m \text{ such that } y = Ax\}.$$

2. The **null** (or **Kernel**) space of A is defined by

$$\mathrm{null}(A) = \mathrm{Ker}(A) = \{x \in \mathbb{R}^m :\ Ax = 0\}.$$

3. The **rank** of the matrix A is defined by

$$\mathrm{rank}(A) = dim(\mathrm{range}(A)).$$

It is not difficult to show that

- $rank(A) = rank(A^T)$.
- The rank of the matrix A is equal to the maximum number of independent vector columns (or rows) of the matrix A.
- We have the classical relation:

$$dim(ker(A)) + rank(A) = m$$

- If $n = m$ and A nonsingular, then $rank(A) = n$ and $range(A) = \mathbb{R}^n$.
- A square matrix A of size $n \times n$ is singular iff $rank(A) < n$.

Let A the square matrix $A \in \mathbb{R}^{n \times n}$. Then A is said to be invertible (regular, nonsingular) iff there exists a matrix $X \in \mathbb{R}^{n \times n}$ such that

$$AX = XA = I_n.$$

In this case the matrix X is the inverse of A and is denoted by $X = A^{-1}$ and we have the following properties

Proposition

If A and B are two nonsingular $n \times n$ matrices, then

1. $(AB)^{-1} = B^{-1}A^{-1}$.
2. $(A^{-1})^T = (A^T)^{-1}$.
3. *The Sherman-Morrison-Woodbury formula:*

$$(A + UV^T)^{-1} = A^{-1} - A^{-1}U(I + V^T A^{-1}U)^{-1}V^T A^{-1},$$

where U and V are two $n \times k$ matrices.

If the matrix $A \in \mathbb{R}^{n\times m}$ is not square, then we can define a generalized inverse also called the **Pseudo-inverse** or the **Moore-Penrose** inverse which is the unique matrix $X \in \mathbb{R}^{m\times n}$ satisfying the following conditions

$$AXA = A, \;\; XAX = X, \;\; (AX)^T = AX, \;\; \text{and} \;\; (XA)^T = XA.$$

In this case the Pseudo-inverse is denoted by $X = A^+$.

Remark

- *If $m = n$ then $A^+ = A^{-1}$.*
- *If $\mathrm{rank}(A) = m$, then $A^+ = (A^TA)^{-1}A^T$.*
- $(AB)^+ = B^+A^+$
- $(A^+)^+ = A$
- *AA^+ and A^+A are orthogonal projections.*

Inner products, vector and matrix norms

Vector norms A norm in vector $\mathbb{R}$-space E is a function from E onto $\mathbb{R}^+$ satisfying the following properties

1. $\|x\| = 0$ if and only if $x = 0$.
2. $\|\lambda x\| = |\lambda|\|x\|$, $\forall \lambda \in \mathbb{R}$ and $\forall x \in E$.
3. $\|x + y\| \leq \|x\| + \|y\|$, $\forall x, y \in E$

In $\mathbb{R}^n$, a useful class of vector norms are the p-norms defined as follows. For a vector $x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$, we define $\|x\|_p$ as

$$\|x\|_p = (|x_1|^p + |x_2|^p + \ldots + |x_n|^p)^{\frac{1}{p}}.$$

In particular, we have the following well known 1,2 and ∞ norms

$$\begin{aligned} \|x\|_1 &= |x_1| + \ldots + |x_n| \\ \|x\|_2 &= (x^T x)^{1/2} = \sqrt{\sum_{i=1}^{n} |x_i|^2} \\ \|x\|_\infty &= \max_{1 \leq i \leq n} |x_i|. \end{aligned}$$

Inner products

The real case

Definition

An inner product on a real vector space E is a mapping

$$\begin{array}{rrcl} \phi & E \times E & \longrightarrow & \mathbb{R} \\ & (x, y) & \longrightarrow & \phi(x, y) \end{array}$$

from $E \times E$ into $\mathbb{R}$ satisfying the following properties

1. ϕ is linear with respect to x and with respect to y: ϕ is bilinear.
2. $\phi(x, y) = \phi(y, x)$: ϕ is symmetric.
3. $\phi(x, x) > 0$ if $x \neq 0$: ϕ is positive definite.

We have the well known Cauchy-Schwartz inequality

$$\phi(x, y)^2 \leq \phi(x, x)\phi(y, y).$$

Remark

For two vectors $x, y \in \mathbb{R}^n$, the Euclidean inner product on $\mathbb{R}^n$ is defined by

$$\langle x, y\rangle_2 = \sum_{i=0}^{n} x_i y_i = x^T y$$

where $x = (x_1, \ldots, x_n)^T$ and $y = (y_1, \ldots, y_n)^T$. We have the following properties

- *For all x, $y \in \mathbb{R}^n$, we have $\langle x, Ay\rangle_2 = \langle A^T x, y\rangle_2$.*
- *The Cauchy-Schwartz inequality*

$$|\langle x, y\rangle_2| \leq \|x\|_2 \|y\|_2.$$

- *All the norms in $\mathbb{R}^n$ are equivalent and we have (exercise)*

$$\begin{array}{ccccc} \|x\|_2 & \leq & \|x\|_1 & \leq & \sqrt{n}\|x\|_2 \\ \|x\|_\infty & \leq & \|x\|_2 & \leq & \sqrt{n}\|x\|_\infty \\ \|x\|_\infty & \leq & \|x\|_1 & \leq & n\|x\|_\infty \end{array}$$

For the complex case the inner product is defined as follows.

Definition

(The complex case) An inner product from a complex vector space E is a mapping

$$\begin{aligned} \psi \qquad E \times E &\longrightarrow \mathbb{C} \\ (x, y) &\longrightarrow \psi(x, y) \end{aligned}$$

from $E \times E$ into $\mathbb{C}$ satisfying the following properties

1. $\psi(\alpha x + \beta y, z) = \alpha\psi(x, z) + \beta\psi(y, z)$, $\forall x, y, z \in E$ and $\alpha, \beta \in \mathbb{C}$.
2. $\psi(x, y) = \overline{\psi(y, x)}$, $\forall x, y \in E$:ψ is Hermitian.
3. $\psi(x, x) > 0$ if $x \neq 0$: ψ is positive definite.

We notice that (2) implies that $\psi(x,x)$ is real. We also have $s(x,0) = s(0,y) = 0 \; \forall x, y \in E$.
The Cauchy-Schwartz inequality is given by

$$|\psi(x,y)|^2 \leq \psi(x,x)\,\psi(y,y).$$

For the complex case with $E = \mathbb{C}^n$, the canonical inner product called also the Euclidean inner product is defined as follows. For two vectors $x, y \in \mathbb{C}^n$, we have

$$\langle x, y\rangle = \sum_{i=0}^{n} x_i \overline{y_i} = x^H y$$

where $x = (x_1, \ldots, x_n)^T$ and $y = (y_1, \ldots, y_n)^T$. For all x, $y \in \mathbb{C}^n$, the adjoint matrix A^H of A satisfies the relation

$$\langle x, Ay\rangle = \langle A^H x, y\rangle.$$

Matrix norms

Usually in matrix computation, one needs the knowledge of the norm of a matrix . For a general matrix $A \in \mathbb{R}^{n \times m}$, we consider first the following induced norm

$$\|A\|_{pq} = \max_{x \in \mathbb{R}^n / \{0\}} \frac{\|Ax\|_p}{\|x\|_q}.$$

These norms satisfy the usual properties of the norm and If $p = q$, then we have the following property

$$\|AB\|_p \leq \|A\|_p \, \|B\|_p.$$

If the matrix A is square, then we have

$$\|A^k\|_p \leq \|A\|_p^k, \; k = 1, 2, \ldots .$$

Notice also that $\|I_n\|_p = 1$.

Another non-induced matrix norm is the well known Frobenius

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2}.$$

This is not an induced norm since $\|I_n\|_F = \sqrt{n}$. We notice that the Frobenius norm is consistent which means that for two matrices with appropriate sizes,

$$\|AB\|_F \leq \|A\|_F \|B\|_F.$$

The Frobenius norm is associated to the scalar product in the space of matrices in $\mathbb{R}^{n\times m}$ defined by

$$\begin{array}{llll} \langle .,. \rangle_F : & \mathbb{R}^{n\times m} \times \mathbb{R}^{n\times m} & \longrightarrow \mathbb{R} \\ & (X, Y) & \longrightarrow \langle X, Y \rangle_F = tr(X^T Y) \end{array}$$

where $tr(Z)$ denotes the sum of the elements on the diagonal of the square matrix Z.

For the classical 1, 2 and ∞ matrix norms, we have the following expressions

$$\begin{aligned} \|A\|_1 &= \max_{1\le j\le m} \sum_{i=1}^{n} |a_{ij}|, \\ \|A\|_\infty &= \max_{1\le i\le n} \sum_{i=1}^{m} |a_{ij}|, \\ \|A\|_2 &= \sqrt{\rho(AA^T)}, \end{aligned}$$

where $\rho(AA^T)$ denotes the spectral radius of the matrix AA^T.
The condition number of a square nonsingular matrix A is given by

$$\kappa(A) = \|A\| \, \|A^{-1}\|.$$

Different matrix norms give rise to different condition numbers such as $\kappa_p(A) = \|A\|_p \, \|A^{-1}\|_p$ for a matrix p-norm or $\kappa_F(A) = \|A\|_F \, \|A^{-1}\|_F$ for the Frobenius norm.

Orthogonal vectors

Two vectors $u = (u^1, \dots, u^n)^T$ and $v = (v^1, \dots, v^n)^T$ of $\mathbb{R}^n$ are orthogonal with respect to the classical Euclidean scalar product if and only if

$$\langle u, v \rangle_2 = \sum_{i=1}^{n} u^i v^i = 0.$$

A set of vectors $F = span\{u_1, u_2, \dots, u_p\}$ is orthogonal if

$$\langle u_i, u_j \rangle_2 = 0, \; i, j = 1, \dots, p, \; i \neq j.$$

Two subspaces $F = span\{u_1, u_2, \dots, u_p\}$ and $G = span\{v_1, v_2, \dots, v_q\}$ are orthogonal if

$$\langle u_i, v_j \rangle_2 = 0, \; i = 1, \dots, p, \; j = 1, \dots, q.$$

The vectors $\{u_1, u_2, \dots, u_p\}$ are orthonormal iff

$$\langle u_i, u_j \rangle_2 = \delta_{ij} \,, \; i, j = 1, \dots, p.$$

In this case, the matrix $U = [u_1, u_2, \dots, u_p]$ is orthogonal and we have

$$U^T U = I.$$

The orthogonal of the subspace $F = span\{u_1, u_2, \ldots, u_p\}$ is the orthogonal-subspace of F defined as

$$F^{\perp} = \{y \in \mathbb{R}^n \,/\langle y, u_i\rangle_2 = 0,\ i = 1, \ldots, p\}.$$

Remark

If Q and Z are orthogonal matrices ($Q^T Q = I$ and $Z^T Z = I$), then for any vector x, and for any matrix A, with appropriate sizes, we have

$$\|Qx\|_2 = \|x\|_2, \text{ and } \|QAZ\|_F = \|A\|_F.$$

Definition

(Invariant subspaces) The subspace F is an invariant subspace of A iff $AF \subset F$.

We have the following result

Theorem

If V is a matrix whose columns form a basis of the invariant subspace F, then there exists a unique matrix L such that

$$AV = VL,$$

and we also have (u, λ) is an eigenpair of L if and only if (Vu, λ) is an eigenpair of A.

Projectors

Oblique projectors In the real case, a projector P is a linear mapping from $\mathbb{R}^n$ onto $\mathbb{R}^n$ satisfying

$$P^2 = P.$$

If P is a projector so is $I - P$ and we have

$$Ker(P) = Range(I - P),$$

where $Ker(P) = \{x \in \mathbb{R}^n / \ Px = 0\}$ and $Range(P) = \{y = Px, \ x \in \mathbb{R}^n\}$. We also have the classical decomposition of $\mathbb{R}^n$ as

$$\mathbb{R}^n = Ker(P) \oplus Range(P).$$

This is due to the fact that for all $x \in \mathbb{R}^n$ we have $x = Px + (I - P)x$ and $Ker(P) \cap Range(P) = \{0\}$. Consider now two subspaces $\widetilde{F}$ and $\widetilde{G}$ of $\mathbb{R}^n$ and assume that

$$\mathbb{R}^n = \widetilde{F} \oplus \widetilde{G}^{\perp}.$$

We define the projection onto $\widetilde{F}$ and orthogonally to $\widetilde{G}$ as follows: for all $x \in \mathbb{R}^n$, $x = y + z$ where $y \in \widetilde{F}$ and $z \in \widetilde{G}^\perp$, the projector P onto $\widetilde{F}$ and parallel to $\widetilde{G}^\perp$ is given by

$$Px = y.$$

The oblique projector P is completely determined by the following relations

$$\begin{cases} Px \in \widetilde{F} \\ x - Px \perp \widetilde{G} \end{cases}$$

P is also called the projection onto $\widetilde{F}$ along $\widetilde{G}^\perp$. Notice that

$$Range(P) = \widetilde{F} \text{ and } Ker(P) = \widetilde{G}^\perp.$$

Let us give now a matrix representation of the projector P. Assume that m is the dimension of the two subspaces $\widetilde{F}$ and $\widetilde{G}$ and let $\{u_1, u_2, \ldots, u_m\}$ and $\{v_1, v_2, \ldots, v_m\}$ be two bases of $\widetilde{F}$ and $\widetilde{G}$, respectively and set

$$U = [u_1, u_2, \ldots, u_m], \quad V = [v_1, v_2, \ldots, v_m].$$

Then,

$$Px \in \widetilde{F} \iff Px = Uy \text{ for } y \in \mathbb{R}^m. \tag{1}$$

On the other hand, the orthogonality relation is expressed as

$$x - Px \perp \widetilde{G} \iff V^T(x - Px) = 0. \tag{2}$$

' From the relations (1) and (2) , we get the matrix representation of P as

$$P = U(V^TU)^{-1}V^T.$$

Notice that if the two bases are bi-orthonormal, i.e. $\langle u_i, w_j\rangle_2 = \delta_{ij}$ then $V^TU = I_m$ and

$$P = UV^T. \tag{3}$$

Orthogonal projectors

Assume now that $\widetilde{F} = \widetilde{G}$ and in this case P is an orthogonal projector. The matrix representation of the orthogonal projector P is given as

$$P = U(U^T U)^{-1} U^T. \tag{4}$$

If $\{u_1, u_2, \ldots, u_m\}$ is an orthonormal basis and $U = [u_1, \ldots, u_m]$, we get

$$P = UU^T. \tag{5}$$

Theorem

Let P be an orthogonal projector. Then

1. $\|P\|_2 = 1$
2. $\min\limits_{y \in \widetilde{F}} \|x - y\|_2 = \|x - Px\|_2, \; \forall x \in \mathbb{R}^n.$
3. $P^T = P$ *and* $Ker(P) = [(P)]^{\perp}.$

Proof.

For the item 1, we have

$$\|x\|_2^2 = \|Px\|_2^2 + \|x - Px\|_2^2$$

which shows that

$$\|Px\|_2^2 \leq \|x\|_2^2.$$

Then the maximum of $\|Px\|_2/\|x\|_2$ is less or equal to one and the value one is reached for any $x \in Range(P)$. Hence $\|P\|_2 = 1$.
For the item 2, let $x \in \mathbb{R}^n$ and $y \in \widetilde{F}$, then

$$\|x - y\|^2 = \|x - Px + Px - y\|^2$$

and since $x - Px \perp \widetilde{F}$ and $Px - y \in \widetilde{F}$, we obtain

$$\|x - y\|^2 = \|x - Px\|^2 + \|Px - y\|^2$$

and then the minimum is obtained for $y = Px$. For item 3, we can use the expression (5) of the orthogonal projector P. $\square$

Eigenvalues of a square matrix

Let A be a square matrix in $\mathbb{R}^{n\times n}$. Then a scalar $\lambda \in \mathbb{C}$ is called an eigenvalue of the matrix A if and only if there exists a nonzero vector $u \in \mathbb{C}^n$ such that

$$Au = \lambda u.$$

The vector u is called and eigenvector associated to the eigenvalue λ and the set of all eigenvalues of A will be denoted by $\Lambda(A)$. We notice that a scalar λ is an eigenvalue of A iff it is a zero of the characteristic polynomial P_A defined by

$$P_A(\lambda) = det(A - \lambda I_n),$$

where $det(Z)$ denotes for the determinant of the square matrix $Z = [z_{ij}]$ defined by

$$det(Z) = \sum_{j=1}^{n} (-1)^{j+1}\, z_{1j}\, det(Z_{1j}),$$

where $det(Z_{1j})$, is the $(n-1) \times (n-1)$ determinant obtained by deleting the first row and the j-th column.

Remark

1. *The square matrix A is singular iff $det(A) = 0$ which is equivalent to $\lambda = 0$ is an eigenvalue of A.*
2. *The $n \times n$ matrix A has exactly n complex eigenvalues (the n roots of the characteristic polynomial P_A).*
3. *The maximum modulus of the eigenvalues is called the spectral radius of A and denoted by $\rho(A)$.*
4. *The subspace $E_\lambda = Ker(A - \lambda I)$ (of all eigenvectors associated to λ + the null vector) is invariant under A which means that $AE_\lambda \subset E_\lambda$.*
5. *The characteristic polynomial P_A can be expressed as*

$$P_A(\lambda) = det(A - \lambda I) = \prod_{i=1}^{p} (\lambda_i - \lambda)^{m_i},$$

where m_i is the multiplicity of the eigenvalue λ_i with $\sum_{i=1}^{p} m_i = n$.

Remark

1. $dim(E_{\lambda_i}) \leq m_i$.
2. Cayley-Hamilton $P_A(A) = 0$.
3. $trace(A) = \sum_{i=1}^{n} \lambda_i$ and $det(A) = \prod_{i=1}^{n} \lambda_i$.
4. A and A^T have the same eigenvalues.
5. A unitary $\Longrightarrow |\lambda| = 1, \forall \lambda \in \Lambda(A)$.

Definition

Two matrices A and B are similar if and only if there exists a nonsingular matrix P such that $A = PBP^{-1}$ and then A and B have the same eigenvalues. The matrices A and B represent the same linear mapping but in different bases.

Definition

The $n \times n$ matrix A is diagonalizable if and only if there exists a nonsingular matrix P such that

$$A = P\ D\ P^{-1},$$

where $D = diag(\lambda_1, \dots, \lambda_n)$. This is also equivalent to the fact that the eigenvectors of A form a basis of $\mathbb{C}^n$.

It is shown that the matrix A is diagonalizable if and only if

$$dim(E_{\lambda_i}) = m_i$$

for $i = 1, \ldots, p$. The dimension of E_{λ_i} is called the geometric multiplicity. Another important invariant subspace is called the characteristic subspace F_{λ_i} associated to the eigenvalue λ_i and defined by

$$F_{\lambda_i} = (ker(A - \lambda_i I)^{m_i},\ i = 1, \ldots, p,$$

where m_i is the algebraic multiplicity of the eigenvalue λ_i.
We also have

$$dim(F_{\lambda_i}) = m_i,$$

$$AF_{\lambda_i} \subset F_{\lambda_i}, \tag{6}$$

and

$$F_{\lambda_1} \oplus F_{\lambda_2} \oplus \ldots \oplus F_{\lambda_p} = \mathbb{C}^n. \tag{7}$$

Choosing as a new basis of $\mathbb{C}^n$: the basis whose elements are the union of the elements in each basis in F_{λ_i},

$$\mathcal{B} = \{B_1, \ldots, B_p\},$$

where B_i is a basis of F_{λ_i}. The linear mapping corresponding to the matrix A can be represented in this basis $\mathcal{B}$ by a block-diagonal matrix. We summarize this result in the following theorem.

Theorem

Let A be a matrix in $\mathbb{C}^{n\times n}$. Then there exists a nonsingular matrix P such that

$$P^{-1}AP = \begin{pmatrix} A_1 & 0 & \ldots & 0 \\ 0 & A_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots 0 & A_p \end{pmatrix},$$

where the block A_i is of size $m_i \times m_i$ and m_i is the algebraic multiplicity of the eigenvalue λ_i, $i = 1, \ldots, p$ and p is the number of distinct eigenvalues of the matrix A. We notice that the sub-matrix A_i has λ_i as the unique eigenvalue with algebraic multiplicity equals m_i.

Next, we give the well known eigenvalue-localisation theorem.

Theorem

Let $A = [a_{ij}]$ be an $n \times n$ matrix and let D_i denotes the following complex disc

$$D_i = \{z \in \mathbb{C} : \ |z - a_{ii}| \leq \sum_{i \neq j = 1}^{n} |a_{ij}|\}.$$

Then the spectrum $\Lambda(A)$ satisfies

$$\Lambda(A) \subset \bigcup_{i=1}^{n} D_i.$$

We also have

$$\rho(A) \leq \min\{\max_i (\sum_{j=1}^{n} |a_{ij}|), \ \max_j (\sum_{i=1}^{n} |a_{ij}|)\}.$$

Schur decomposition

An important theme of matrix theory is the reduction of matrices to a simple form such as diagonal or triangular by similarity transformations. In particular, unitary transformations are particularly desirable. The following theorem shows that any matrix A can be reduced to an upper triangular matrix by a unitary similarity.

Theorem

Let $A \in \mathbb{C}^{n\times n}$, then there exists an $n \times n$ unitary matrix such that

$$T = U^T A U = \begin{pmatrix} \lambda_1 & t_{12} & t_{13} & \dots & t_{1n} \\ 0 & \lambda_2 & t_{23} & \dots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & \lambda_n \end{pmatrix}.$$

is upper triangular. The matrix U may be chosen such that the eigenvalues of A appear on the diagonal of T in any order.

Proof.

Assume that it is thrue for all complex matrices of size $(n-1)\times(n-1)$. Let λ be an eigenvector of A corresponding to a normalized eigenvector u_1. Then we can find a matrix $U_2 \in \mathbb{C}^{n\times(n-1)}$ such that $U=[u_1, U_2]$ is unitary. Now as $AU=[Au_1, AU_2]$, we get

$$U^H A U = \begin{pmatrix} \lambda u_1^H u_1 & u_1^H A U_2 \\ \lambda U_2^H u_1 & U_2^H A U_2 \end{pmatrix} = \begin{pmatrix} \lambda & w^H \\ O & B \end{pmatrix},$$

where $w^H = u_1^H A U_2$ and $B = U_2^H A U_2$. Since the matrix B is of order $n-1$ and by using the induction hypothesis there exists a unitary matrix V of order $n-1$ such that $V^H B V = \widetilde{T}$ where $\widetilde{T}$ is upper triangular. Setting

$$Q = U\begin{pmatrix} 1 & O \\ O & V \end{pmatrix},\ and\ Q^H A Q = \begin{pmatrix} \lambda & w^H V \\ O & V^H B V \end{pmatrix},$$

and since $V^H B V = \widetilde{T}$ is upper triangular and Q unitary, the result follows.
$\square$

Theorem

(The real Schur decomposition)
Let $A \in \mathbb{R}^{n\times n}$, *then there exists an orthonormal matrix* $Q \in \mathbb{R}^{n\times n}$ *such that*

$$Q^T A Q = \begin{pmatrix} R_{11} & R_{12} & \dots & R_{1n} \\ 0 & R_{22} & \dots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & R_{nn} \end{pmatrix}$$

is a real block upper triangular matrix where each diagonal square block R_{ii} *of size* 1×1 *or* 2×2 *and for the last case, the* 2×2 *blocks have complex conjugate eigenvalues.*

Jordan canonical form

Theorem

Let A be a square $n \times n$ matrix. Then there exists a nonsingular matrix P such that

$$P^{-1}AP = \begin{pmatrix} J_{k_1}(\lambda_1) & 0 & \dots & 0 \\ 0 & J_{k_2}(\lambda_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_{k_l}(\lambda_l) \end{pmatrix} \tag{8}$$

where $J_k(\lambda) \in \mathbb{C}^{k \times k}$ is a Jordan block of the form

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & \dots & 0 & \lambda \end{pmatrix},$$

with $k_1 + k_2 + \dots k_l = n$.

Let us partition P as $P = [u_1, \ldots, u_n]$. Then the first k_1 vectors associated with the Jordan block $J_{k_1}(\lambda_1)$ are such that

$$u_1 = \lambda_1 u_1 \text{ and } u_j = \lambda_1 u_j + u_{j-1},\ j = 2, \ldots, k_1.$$

Such a sequence is called a chain of principal vectors of A. The other columns of P corresponding to the other Jordan blocks verify a similar relation. Let us partition P conformally with the Jordan blocks in (8) as

$$P = [U_1, \ldots, U_l], \text{ and set } Y^H = P^{-1}.$$

Then

$$A = U_1 J_{k_1}(\lambda_1) Y_1^H + U_2 J_{k_2}(\lambda_2) Y_2^H + \ldots + U_l J_{k_1}(\lambda_l) Y_l^H.$$

It follows that

$$AU_i = U_i J_{k_i}(\lambda_i),\ i = 1, \ldots, l.$$

This shows that the operator A maps *range*(U_i) into itself.

Hermitian and Symmetric matrices

For symmetric (or Hermitian in the complex case) matrices, we have the following result.

Theorem

The eigenvalues of a complex Hermitian matrix are all real.

Theorem

Let A be an Hermitian complex matrix of size $n \times n$. Then A is unitary similar to a real diagonal matrix:

$$A = UDU^T,$$

where U is a unitary matrix and D is a diagonal matrix of the real eigenvalues of A.

This result shows that an Hermitian matrix has a set of orthonormal eigenvectors $\{u_1, u_2, \ldots, u_n\}$ that are the columns of the unitary matrix U.

Theorem

(Courant-Fisher min-max theorem) The eigenvalues of a real symmetric matrix are characterized by the following min-max principle also called the Courant-Fisher min-max theorem.

$$\lambda_k = \min_{S,\ dim(S)=k} \ \max_{x \in S, x \neq 0} \frac{\langle Ax, x\rangle}{\langle x, x\rangle}, \ k = 1, \ldots, n. \tag{9}$$

Another characterisation of eigenvalues is given in the next theorem.

Theorem

Let A be an $n \times n$ Hermitian matrix and let $\lambda_1, \ldots \lambda_n$ be the real eigenvalues of A ordred in decreasing order and corresponding to to the orthonormal eigenvectors $q_1, \ldots, q_n$. Then

$$\lambda_1 = \frac{\langle Aq_1, q_1 \rangle}{\langle q_1, q_1 \rangle} = \max_{x \neq 0} \frac{\langle Ax, x \rangle}{\langle x, x \rangle},$$

and for $k = 2, \ldots, n$, we have

$$\lambda_k = \frac{\langle Aq_k, q_k \rangle}{\langle q_k, q_k \rangle} = \max_{x \neq 0, x^h q_i = 0, i=1,\ldots,k-1} \frac{\langle Ax, x \rangle}{\langle x, x \rangle}.$$

This characterisation is known as the Courant characterisation and it show that the maximum of the Rayleigh quotient a subspace orthogonal to the first $k-1$ eigenvectors is equal to λ_k.

Theorem

Let A be an $n \times n$ real symmetric matrix and let Q be an orthogonal matrix

$$Q = [Q_1,\ Q_2]$$

where Q_1 is of size $n \times r$ and assume that range(Q_1) is invariant under A. Then

$$Q^T A Q = \begin{pmatrix} A_1 & O \\ O & A_2 \end{pmatrix}$$

and $\lambda(A) = \lambda(A_1) \cup \lambda(A_2)$ where A_1 and A_2 are square matrices of sizes r and $n - r$, respectively.

Theorem

Let A and E are $n \times n$ real symmetric matrices and let $Q = [Q_1,\ Q_2]$ be an orthogonal $n \times n$ matrix such that range(Q_1) is invariant under A. Partition the matrices $Q^T AQ$ and $Q^T EQ$ as follows

$$Q^T AQ = \begin{pmatrix} A_1 & O \\ O & A_2 \end{pmatrix},\ E^T AE = \begin{pmatrix} E_{11} & E_{21}^T \\ E_{21} & E_{22} \end{pmatrix}.$$

If $sep(A_1, A_2) > 0$ and $\|E\|_F \leq \dfrac{sep(A_1, A_2)}{5}$, then there exists a matrix $P \in \mathbb{R}^{(n-r)\times r}$ with

$$\|P\|_F \leq \frac{4}{sep(A_1, A_2)} \|E_{21}\|_F,$$

such that the column of $(Q_1 + Q_2 P)(I + P^T P)^{-1/2}$ define an orthonormal basis for a subspace that is invariant under $A + E$, where

$$sep(A_1, A_2) = \min_{\lambda \in \lambda(A_1), \mu \in \lambda(A_2)} |\lambda - \mu|.$$

Theorem

Let A and B be two real symmetric matrices of the same sizes $n \times n$ and set $C = A + B$. Let the eigenvalues of A be such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, and let the eigenvalues of C denotes by ψ_i such that $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_n$. If μ_n is the smallest eigenvalue of B, then

$$\gamma_i \geq \lambda_i + \mu_n, \; i = 1, \ldots, n.$$

Normal matrices

Definition

A square complex matrix A is normal if and only if

$$A^H A = AA^H.$$

Notice that if A is normal and triangular, then A is diagonal.

Theorem

Let A be a square complex matrix of size $n \times n$. Then A is normal if and only if there exists a unitary matrix Q, and a diagonal matrix D such that

$$A = QDQ^H,$$

which means that A is unitarily similar to a diagonal matrix.

In particular, any Hermitian matrix is unitarily similar to a diagonal matrix. Hermitian matrices form a subclass of normal matrices and we have the following result

Proposition

If the complex matrix A is normal and if all its eigenvalues are real, then A is Hermitian.

An important property of normal matrices is given in the next proposition.

Proposition

Let A be a complex matrix. Then A is normal if and only if

$$v \text{ eigenvector of } A \implies v \text{ eigenvector of } A^H.$$

Definition

Let A in $\mathbb{C}^{n\times n}$. Then

1. The field of values (also called numerical range) of the matrix A is defined by
$$\mathcal{F}(A) = \{x^H Ax,\ x \in \mathbb{C}^n,\ x^H x = 1\}.$$
2. The numerical radius of A is defined by
$$r(A) = sup\{|x^H Ax|,\ x^H x = 1\}.$$

In general, the field of values may contain complex values even if the eigenvalues of the matrix A are real.

Proposition

The field of values $\mathcal{F}(A)$ is a convex set

Since the field of values of A is convex , then $\mathcal{F}(A)$ should contain the smallest convex set containing all the eigenvalues. This set is the called the convex hull of the set of eigenvalues $\Lambda(A)$,

Proposition

We have the following properties:

1. $\Lambda(A) \subset \mathcal{F}(A)$.
2. *If U is unitary, then* $\mathcal{F}(U^H A U) = \mathcal{F}(A)$.
3. $\rho(A) \leq r(A)$ *and if A is normal* $r(A) = \rho(A) = \|A\|_2$.
4. $r(A) \leq \|A\|_2 \leq 2r(A)$.

Proposition

For normal matrices, the field of values is equal to the convex hull of its spectrum.

This is due to the fact that if A is normal, then A is unitarily similar to a diagonal matrix D such that $A = QDQ^H$. Hence

$$\mathcal{F}(A) = \mathcal{F}(Q^H D Q) = \mathcal{F}(D).$$

When the complex matrix A is far from normal, it is more interesting to consider pseudo-eigenvalues defined as follows. Let $\epsilon > 0$, then a complex number λ is called an ϵ-pseudo eigenvalue of A if for some matrix E with $\|E\|_2 \leq \epsilon$, λ is an eigenvalue of the matrix $A + E$. Next, we define the ϵ-pseudo-spectrum .

Definition

For a given $\epsilon > 0$, the ϵ-spectrum of the complex matrix A is defined by

$$\Lambda_\epsilon = \{z \in \mathbb{C} \text{ such that } \|(\mathrm{zI} - \mathrm{A})^{-1}\|_2 \geq \epsilon^{-1}\}. \tag{10}$$

Another equivalent definition is as follows

$$\Lambda_\epsilon = \{z \in \mathbb{C} \text{ such that } z \in \mathcal{L}(A + E),\ \|E\|_2 \leq \epsilon.\}. \tag{11}$$

The most useful definition for computation is as follows

$$\Lambda_\epsilon = \{z \in \mathbb{C} :\ \sigma_{min}(zI - A)\|_2 \leq \epsilon\}, \tag{12}$$

where σ_{min} is the smallest singular value of A to be seen in the next chapters.

Positive matrices

The matrix $A \in \mathbb{R}^{n\times n}$ is positive semi-definite ($A \geq 0$) if

$$x^T A x \geq 0,\ \forall x \in \mathbb{R}^n$$

and it is positive definite if $x^T A x > 0,\ \forall x \neq 0$. We have the following properties

Proposition

Let $A \in \mathbb{R}^{n\times n}$. Then if A is positive-definite, we have

1. *If $X \in \mathbb{R}^{n\times k}$ has rank k, then $X^T A X$ is also positive-definite.*
2. *All the principal sub-matrices of A are positive-definite and in particular, the diagonal entries are strictly positive.*
3. *A is positive definite if and only if the symmetric part $A_s = (1/2)(A + A^T)$ of A is positive definite.*
4. *If A is symmetric and positive-definite then all its eigenvalues are real strictly positive .*

Some particular matrices

Hamiltonian matrices

A matrix $\mathcal{H} \in \mathbb{R}^{2n \times 2n}$ is a Hamiltonian matrix if it has the form

$$\mathcal{H} = \begin{pmatrix} A & G \\ F & -A^T \end{pmatrix}$$

where A, F, and G are $n \times n$ matrices and F and G are symmetric. Hamiltonian matrices are very important in control and many other applications. Let J be the following matrix

$$J = \begin{pmatrix} O & I_n \\ -I_n & O \end{pmatrix}.$$

Notice that

$$J^{-1} = J^T = -J \text{ and } det(J) = 1.$$

A $2n \times 2n$ matrix $\mathcal{H}$ is a Hamiltonian matrix if and only if $J\mathcal{H}$ is symmetric

$$(J\mathcal{H})^T = J\mathcal{H}.$$

Symplectic matrices

Another important class of matrices that occur in control is the class of symplectic matrices. Let $\mathcal{S} \in \mathbb{R}^{2n\times 2n}$, then S is symplectic if

$$\mathcal{S}^T J\mathcal{S} = J.$$

Notice that any symplectic matrix has determinant 1 or -1, is nonsingular and the inverse is also symplectic and given by

$$\mathcal{S}^{-1} = J^{-1}\mathcal{S}^T J = -J\mathcal{S}^T J.$$

This comes from the fact that

$$J^{-1}\mathcal{S}^T J\,, \mathcal{S} = J^{-1}J = I_{2n}.$$

The product of two symplectic matrices is a symplectic matrix. The $2n \times 2n$ symplectic matrices with real entries form a subgroup of the general linear group $\mathrm{GL}(2n;\mathbb{R})$ under matrix multiplication.

M-matrices, Irreductible and Stochastic matrices

These class of matrices are very important in many applications such is in control, model reductions and others. First, we begin by the definition of a Z-matrix

Definition

A square matrix M is said to be a Z-Matrix if all its off-diagonal entries are less than or equal to zero; that is, the matrices of the form:

$$M = [m_{ij}] \text{ with } m_{ij} \leq 0, \quad i \neq j.$$

A second important class of Z-matrices are the M-Matrices defined as follows

Definition

A real square matrix M is said M-matrix if $M = sI - H$ with $H \geq 0$ and $s \geq \rho(H)$, where the notation $H \geq 0$ means that all the coefficient of the matrix H are positive and $\rho(H) = \max |\lambda_i(H)|$ where $\lambda_i(H)$ denotes the i-th eigenvalue of the matrix H.

Special matrix products

Let $A = [a_{ij}] \in \mathbb{R}^{n\times m}$and $B = [b_{ij}] \in \mathbb{R}^{p\times q}$, then the Kronecker [3] product of these two matrices in the $np \times mq$ matrix defined as follows

$$A \otimes B = [a_{ij}B] = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1m}B \\ a_{21}B & a_{22}B & \dots & a_{2m}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1}B & a_{n2}B & \dots & a_{nm}B \end{pmatrix}.$$

[3] Leopold Kronecker (German: 7 December 1823 –29 December 1891) was a German mathematician who worked on number theory, algebra and logic.

Proposition

The Kronecker product satisfies the following properties

1. $(A \otimes B)^T = A^T \otimes B^T$.
2. $(A \otimes B)(C \otimes D) = (AC \otimes BD)$ *and* $(A \otimes B)^k = A^k \otimes B^k$, *for an integer* k.
3. *If* A *and* B *are positive (semi)-definite, then* $A \otimes B$ *is also positive (semi) definite.*
4. $e^{A \otimes I + I \otimes B} = e^A \otimes e^B$.
5. *If* A *and* B *are nonsingular matrices of dimension* $n \times n$ *and* $p \times p$ *respectively, then*
$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$
6. *Let* A *and* B *be* $n \times n$ *and* $p \times p$ *matrices, then*
$$det(A \otimes B) = det(A)^p det(B)^n \ \text{ and } \ tr(A \otimes B) = tr(A)tr(B),$$

where A, B *and* C *are matrices of appropriate sizes.*

Proposition

Let A, B, C and X be matrices of appropriate sizes. Then

1. $\mathrm{vec}(AXB) = (B^T \otimes A)\,\mathrm{vec}(X)$,
2. $\mathrm{vec}(A)^T\,\mathrm{vec}(B) = \mathrm{trace}(A^T B) = \langle A, B \rangle_F$,
3. $vec(AX + XB) = (I \otimes A) + (B^T \otimes I))vec(X)$,
4. $\|X\|_F = \|vec(X)\|_2$.

where $vec(X) \in \mathbb{R}^{np}$ *is the long vector obtained by stacking the columns of the matrix* $X \in \mathbb{R}^{n \times p}$.

Proposition

If A and B are two $n \times n$ and $p \times p$ matrices, respectively, and if λ_i, $i = 1, \ldots, n$ are the eigenvalues of A and $\mu_1, \ldots, \mu_p$ are the eigenvalues of the matrix B, then

1. *The eigenvalues of the $np \times np$ matrix $A \otimes B$ are the np scalars $\lambda_i \mu_j$, $i = 1, \ldots, n$ and $j = 1, \ldots, p$.*
2. *The eigenvalues of $(I \otimes A) - (B^T \otimes I))$ are the scalars $\lambda_i - \mu_j$.*
3. *The eigenvalues of the operator $\mathcal{A}$ defined by $\mathcal{A}(X) = AX - XB$ are $\lambda_i - \mu_j$ where $\lambda_i(A)$ is an eigenvalue of A and μ_j is an eigenvalue of B.*

Proposition

Assume that the matrices $A \in \mathbb{R}^{n\times n}$ *and* $B \in \mathbb{R}^{p\times p}$ *are diagonalisable*

$$A = PDP^{-1},\ B = Q\Delta Q^{-1},$$

where D *and* δ *are diagonal matrices. Then, the Kronecker product* $A \otimes B$ *is also diagonalisable and we have*

$$\begin{aligned} A \otimes B &= (PDP^{-1}) \otimes (Q\Delta Q^{-1}) \\ &= (P \otimes Q)(D \otimes \Delta)(P \otimes Q)^{-1}. \end{aligned}$$

Hadamard product

The Hadamard [4] product of the matrices $A = [a_{ij}] \in \mathbb{R}^{n\times m}$and $B = [b_{ij}] \in \mathbb{R}^{n\times m}$ is the $n \times m$ matrix defined by

$$A \circ B = [a_{ij} b_{ij}] = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} & \dots & a_{1m}b_{1m} \\ a_{21}b_{21} & a_{22}b_{22} & \dots & a_{2m}b_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1}b_{n1} & a_{n2}b_{n2} & \dots & a_{nm}b_{nm} \end{pmatrix}.$$

The Hadamard product is carried out by multiplying the matrices entry by entry. We have the following properties

Proposition

Let A and B be two matrices of sizes $n \times m$. Then

$$rank(A \circ B) \leq rank(A)rank(B).$$

[4] Jacques Hadamard 8 December 1865 – 17 October 1963, was a French mathematician who made major contributions in number theory, complex analysis, differential geometry and partial differential equations

Khatri-Rao and Face-splitting products

Let the matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{p\times m}$. Then, the Khatri-Rao product product of the matrices A and B, denoted by $A \odot B$ is is the matrix of size $np \times m$ defined by

$$A \odot B = [a_{:,1} \otimes b_{:,1}, a_{:,2} \otimes b_{:,2}, \ldots, a_{:,m} \otimes b_{:,m}],$$

where $a_{:,j}$ is the j-th column of the matrix A. Another interesting product named 'Face-splitting' or 'transposed Khatri-Rao' product is based on row-by-row Kronecker products of two matrices and is defined as follows. Let A and B be two matrices of size $n \times m$ and $n \times p$, respectively. Then the Face-splitting product of A and B, denoted $A \bullet B$ is given by

$$A \bullet B = \begin{pmatrix} a_{1,:} \otimes b_{1,:} \\ a_{2,:} \otimes b_{2,:} \\ \vdots \\ a_{n,:} \otimes b_{n,:} \end{pmatrix},$$

where $A_{i,:}$ is the i-th row of the matrix A. It is clear that

$$(A \bullet B)^T = A^T \odot B^T.$$

The $\diamond$ product (Jbilou et al.)

In the following we consider the product denoted by $\diamond$ defined as follows [?]

Definition

Let $A = [A_1, A_2, \ldots, A_p]$ and $B = [B_1, B_2, \ldots, B_l]$ be matrices of dimension $n \times ps$ and $n \times ls$ respectively where A_i and B_j ($i = 1, \ldots, p$; $j = 1, \ldots, l$) are $n \times s$ matrices. Then the $p \times l$ matrix $A^T \diamond B$ is defined by:

$$A^T \diamond B = \begin{pmatrix} \langle A_1, B_1\rangle_F & \langle A_1, B_2\rangle_F & \ldots & \langle A_1, B_l\rangle_F \\ \langle A_2, B_1\rangle_F & \langle A_2, B_2\rangle_F & \ldots & \langle A_2, B_l\rangle_F \\ \vdots & \vdots & \vdots & \vdots \\ \langle A_p, B_1\rangle_F & \langle A_p, B_2\rangle_F & \ldots & \langle A_p, B_l\rangle_F \end{pmatrix}.$$

Proposition

Let $A, B, C \in \mathbb{R}^{n\times ps}$, $D \in \mathbb{R}^{n\times n}$, *and* $L \in \mathbb{R}^{p\times p}$. *Then we have*

1. $(A+B)^T \diamond C = A^T \diamond C + B^T \diamond C$ *and*
   $A^T \diamond (B+C) = A^T \diamond B + A^T \diamond C$.
2. $(A^T \diamond B)^T = B^T \diamond A$ *and* $(DA)^T \diamond B = A^T \diamond (D^T B)$.

We also have the following relations that could be useful to simplify some results in global methods when solving some matrix equations such as Lyapunov or Sylvester matrix equations or in some model reduction methods for large scale dynamical systems.

Proposition

Let A, B and L be the matrices as in the last proposition. Then

1. $A^T \diamond (B(L \otimes I_s)) = (A^T \diamond B)L.$
2. $||A^T \diamond B||_F \leq ||A||_F \, ||B||_F.$

The $\diamond$-product is related to the inner product $\langle \, ., \, . \rangle_F$ on matrix subspaces. In fact if $\mathcal{V} = [V_1, V_2, \ldots, V_m]$ where each $V_i \in \mathbb{R}^{n \times s}$, then the matrix $\mathcal{V}$ is F-orthonormal with respect to the inner product $\langle ., \, . \rangle_F$, that is

$$\langle V_i \, , \, V_j \rangle_F = \delta_{ij}$$

if and only if

$$\mathcal{V}^T \diamond \mathcal{V} = I.$$

Schur complement

We first recall the definition of the Schur [5] complement and give some of their properties.

Definition

Let M be a matrix partitioned in four blocks

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where the submatrix D is assumed to be square and nonsingular. The Schur complement of D in M, denoted by (M/D), is defined by

$$(M/D) = A - BD^{-1}C. \tag{13}$$

Moreover, since

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} (M/D) & B \\ O & D \end{pmatrix} \begin{pmatrix} I & 0 \\ D^{-1}C & I \end{pmatrix}, \tag{14}$$

[5] Issai Schur: January 10, 1875–January 10, 1941, was a Russian mathematician who

we get

$$det(M) = det(M/D) \times det(D).$$

If D is not a square matrix then a Pseudo-Schur complement of D in M can still be defined. Let us remark that having the nonsingular submatrix D in the lower right-hand corner of M is a matter of convention. We can similarly define the following Schur complements

$$(M/A) = D - CA^{-1}B, \tag{15}$$

$$(M/B) = C - DB^{-1}A, \tag{16}$$

and

$$(M/C) = B - AC^{-1}D. \tag{17}$$

If the two matrices A and D are square and nonsingular, we have the following relation

$$(M/D)^{-1} = A^{-1} + A^{-1}B(M/A)^{-1}CA^{-1}.$$

We also can show the Guttman rank additivity formula

$$rank(M) = rank(D) + rank(M/D).$$

Proposition

Let A be a square nonsingular matrix of size n and let U and V be two matrices of dimensions $n \times k$ such that $I + V^T A^{-1} U$ is nonsingular. Then

$$(A + UV^T)^{-1} = A^{-1} - A^{-1}U(I + V^T A^{-1} U)^{-1} V^T A^{-1}, \qquad (18)$$

which called the Sherman-Morrison-Woodbury formula.

In the vector case, the formula can be simplified as follows

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1} u v^T A^{-1}}{1 + v^T A^{-1} u} \qquad (19)$$

More generally,we have

$$(A + UFV^T)^{-1} = A^{-1} - A^{-1}U(F^{-1} + V^T A^{-1} U)^{-1} V^T A^{-1}, \qquad (20)$$

where U and V are $n \times k$ matrices, F is a square nonsingular $k \times k$ matrix and assuming that the matrix $F^{-1} + V^T A^{-1} U$ is nonsingular.

Theorem

Consider the real symmetric and positive definite matrix given by

$$M = \begin{pmatrix} A & B \\ B^T & D \end{pmatrix},$$

where A is square. Then A is nonsingular and the Shur complement

$$(M/A) = D - B^T A^{-1} B$$

is symmetric and positive definite.