# VIP Refresher: Linear Algebra and Calculus

## General notations

r **Vector** – We note $x\in\mathbb{R}^n$ a vector with $n$ entries, where $x_i\in\mathbb{R}$ is the $i^{th}$ entry:

$$x=\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}\in\mathbb{R}^n$$

r **Matrix** – We note $A\in\mathbb{R}^{m\times n}$ a matrix with $m$ rows and $n$ columns, where $A_{i,j}\in\mathbb{R}$ is the entry located in the $i$th row and $j$th column:

$$A=\begin{pmatrix}A_{1,1}&\cdots&A_{1,n}\\\vdots&&\vdots\\A_{m,1}&\cdots&A_{m,n}\end{pmatrix}\in\mathbb{R}^{m\times n}$$

*Remark: the vector $x$ defined above can be viewed as a $n\times1$ matrix and is more particularly called a column-vector.*

r **Identity matrix** – The identity matrix $I\in\mathbb{R}^{n\times n}$ is a square matrix with ones in its diagonal and zero everywhere else:

$$I=\begin{pmatrix}1&0&\cdots&0\\0&\ddots&\ddots&\vdots\\\vdots&\ddots&\ddots&0\\0&\cdots&0&1\end{pmatrix}$$

*Remark: for all matrices $A\in\mathbb{R}^{n\times n}$, we have $A\times I=I\times A=A$.*

r **Diagonal matrix** – A diagonal matrix $D\in\mathbb{R}^{n\times n}$ is a square matrix with nonzero values in its diagonal and zero everywhere else:

$$D=\begin{pmatrix}d_1&0&\cdots&0\\0&\ddots&\ddots&\vdots\\\vdots&\ddots&\ddots&0\\0&\cdots&0&d_n\end{pmatrix}$$

*Remark: we also note $D$ as $\textrm{diag}(d_1,...,d_n)$.*

## Matrix operations

r **Vector-vector multiplication** – There are two types of vector-vector products:

• inner product: for $x,y\in\mathbb{R}^n$, we have:

$$x^Ty=\sum_{i=1}^nx_iy_i\in\mathbb{R}$$

• outer product: for $x\in\mathbb{R}^m,y\in\mathbb{R}^n$, we have:

$$xy^T=\begin{pmatrix}x_1y_1&\cdots&x_1y_n\\\vdots&&\vdots\\x_my_1&\cdots&x_my_n\end{pmatrix}\in\mathbb{R}^{m\times n}$$

r **Matrix-vector multiplication** – The product of matrix $A\in\mathbb{R}^{m\times n}$ and vector $x\in\mathbb{R}^n$ is a vector of size $\mathbb{R}^m$, such that:

$$Ax=\begin{pmatrix}a_{r,1}^Tx\\\vdots\\a_{r,m}^Tx\end{pmatrix}=\sum_{i=1}^na_{c,i}x_i\in\mathbb{R}^m$$

where $a_{r,i}^T$ are the vector rows and $a_{c,j}$ are the vector columns of $A$, and $x_i$ are the entries of $x$.

r **Matrix-matrix multiplication** – The product of matrices $A\in\mathbb{R}^{m\times n}$ and $B\in\mathbb{R}^{n\times p}$ is a matrix of size $\mathbb{R}^{n\times p}$, such that:

$$AB=\begin{pmatrix}a_{r,1}^Tb_{c,1}&\cdots&a_{r,1}^Tb_{c,p}\\\vdots&&\vdots\\a_{r,m}^Tb_{c,1}&\cdots&a_{r,m}^Tb_{c,p}\end{pmatrix}=\sum_{i=1}^na_{c,i}b_{r,i}^T\in\mathbb{R}^{n\times p}$$

where $a_{r,i}^T,b_{r,i}^T$ are the vector rows and $a_{c,j},b_{c,j}$ are the vector columns of $A$ and $B$ respectively.

r **Transpose** – The transpose of a matrix $A\in\mathbb{R}^{m\times n}$, noted $A^T$, is such that its entries are flipped:

$$\forall i,j,\quad A_{i,j}^T=A_{j,i}$$

*Remark: for matrices $A,B$, we have $(AB)^T=B^TA^T$.*

r **Inverse** – The inverse of an invertible square matrix $A$ is noted $A^{-1}$ and is the only matrix such that:

$$AA^{-1}=A^{-1}A=I$$

*Remark: not all square matrices are invertible. Also, for matrices $A,B$, we have $(AB)^{-1}=B^{-1}A^{-1}$*

r **Trace** – The trace of a square matrix $A$, noted $\textrm{tr}(A)$, is the sum of its diagonal entries:

$$\textrm{tr}(A)=\sum_{i=1}^nA_{i,i}$$

*Remark: for matrices $A,B$, we have $\textrm{tr}(A^T)=\textrm{tr}(A)$ and $\textrm{tr}(AB)=\textrm{tr}(BA)$*

r **Determinant** – The determinant of a square matrix $A\in\mathbb{R}^{n\times n}$, noted $|A|$ or $\textrm{det}(A)$ is expressed recursively in terms of $A_{\backslash i,\backslash j}$, which is the matrix $A$ without its $i$th row and $j$th column, as follows:

n
det(A) =
|A|= (−1)i+jAi,j|A\i,\j| ∑
j=1

*Remark: A is invertible if and only if*
*|A| 6= 0. Also, |AB| = |A||B| and |AT| = |A|.*

## Matrix properties

r **Symmetric decomposition** – A given matrix A can be expressed in terms of its symmetric and antisymmetric parts as follows:

AATA
−ATA=++
2 2

Symmetric Antisymmetric

r **Norm** – A norm is a function N : V
−→ [0, + ∞[ where V is a vector space, and such that for all x,y
∈ V , we have:

- N(x+y)6N(x)+N(y)
- N(ax)= |a|N(x) for a scalar
- if N(x) = 0, then x = 0

Forx
∈ V , the most commonly used norms are summed up in the table below:

**Norm Notation D**

∑**efinitionUsecase**nManhattan,L1||x||1|xi|LASSOregularization

√√i=1

√Euclidean,L2||x||2√∑nx2iRidgeregularization

(i=1∑)1npp-norm,Lp||x||xppiHölderinequality=1

Infinity,L∞
||x||∞max|xi|Uniformconvergencei

r **Linearly dependence** – A set of vectors is said to be linearly dependent if one of the vectors in the set can be defined as a linear combination of the others.
*Remark: if no vector can be written this way, then the vectors are said to be linearly independent.*

r **Matrix rank** – The rank of a given matrix A is noted rank(A) and is the dimension of the vector space generated by its columns. This is equivalent to the maximum number of linearly independent columns of A.

r **Positive semi-definite matrix** – A matrix A
∈ Rn×n is positive semi-definite (PSD)

A=AT and
∀x∈Rn, xTAx>0
*Remark: similarly, a matrix A is said to be positive definite, and is noted A*
*0, if it is a PSD*
*matrix which satisfies for all non-zero vector x, xTAx > 0.*

r **Eigenvalue, eigenvector** – Given a matrix A
∈ Rn×n, λ is said to be an eigenvalue of A if there exists a vector z
∈ Rn\{0}, called eigenvector, such that we have:

Az = λz

r **Spectral theorem** – Let A
∈ Rn×n. If A is symmetric, then A is diagonalizable by a real orthogonal matrix U
∈ Rn×n. By noting Λ = diag(λ1,...,λn), we have:

∃Λ diagonal, A = UΛUT

r **Singular-value decomposition** – For a given matrix A of dimensions m
× n, the singular-value decomposition (SVD) is a factorization technique that guarantees the existence of U m
× m
unitary,Σm
× n diagonal and V n × n unitary matrices, such that:

A=UΣVT

## Matrix calculus

r **Gradient** – Let f : Rm×n
→RbeafuncA∈Rm×n f
with respect to A is a m
× n matrix

(tio,noted∇)nandbeamatrix.ThegradientofAf(A),suchthat:)=∂f(A)∇Af(Ai,j∂Ai,j

*Remark: the gradient of f is only defined when f is a function that returns a scalar.*

r **Hessian** – Let f : Rn
→ R be a function and x ∈ Rn be a vector. The hessian of f with respect to x is a n
× n symmetric m

(atrix,no)ted∇2xf(x),suchthat:(∂2f∇2)=()xfxi,j∂xi∂xj

*Remark: the hessian of f is only defined when f is a function that returns a scalar.*

r **Gradient operations** – For matrices A,B,C, the following gradient properties are worth having in mind:

∇tr(AB)=BTA∇fATAT()=(∇Af(A))

∇tr(ABATC)=CAB+CTABTA∇−1TA|A|=|A|(A)