Revision: Linear Algebra

(CIV6540 - Probabilistic Machine Learning for Civil Engineers)

Professor: James-A. Goulet

Département des génies civil, géologique et des mines
Polytechnique Montréal

Chapter 2 – Goulet (2020)
Probabilistic Machine Learning for Civil Engineers
MIT Press

Linear Algebra → Foundation

[Google images]

Module #1a Outline

Notation

Operations

Norms

Transformations

Eigen decomposition

Topics organization

Background	1	Revision probability & linear algebra
	2	Probability distributions
Machine Learning Basics	0	Introduction
	3	Bayesian Estimation $p(A\|B) = \frac{p(B\|A)p(A)}{p(B)}$
	4	MCMC sampling & Newton
Supervised learning	5	Regression
	6	Classification
	7	LSTM networks for time series
Unsupervised learning	7	State-space model for time-series
Decision Making & RL	8	Decision Theory
	9	AI & Sequential decision problems

Section Outline

Scalars

Scalar $\rightarrow$ Single number

$$
\begin{array}{lll}
x & \in & \mathbb{R} & \equiv (-\infty, \infty) \\
 & \in & \mathbb{R}^{+} & \equiv (0, \infty) \\
 & \in & \mathbb{Z} & \equiv \{-\infty, \cdots, -1, 0, 1, 2, \cdots, \infty\} \\
 & \in & \ldots & (0, 1) \\
 & \in & \ldots & (0, 1]
\end{array}
$$

Vectors

1-D array containing numbers or scalars

$$\mathbf{x} \equiv \underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

If e.g. each $[\mathbf{x}]_i \in \mathbb{R}, \forall i = \{1:n\} \rightarrow \mathbf{x} \in \mathbb{R}^n$

Matrices

2-D array containing numbers or scalars

$$\mathbf{X} \equiv \underline{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

If e.g. each $[\mathbf{X}]_{ij} \in \mathbb{R}, \forall i = \{1:m\}, j = \{1:n\} \rightarrow \mathbf{X} \in \mathbb{R}^{m \times n}$

Square Matrix

2-D array containing numbers or scalars

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix}$$

If e.g. each $[\mathbf{X}]_{ij} \in \mathbb{R}, \forall i, j = \{1 : n\} \rightarrow \mathbf{X} \in \mathbb{R}^{n \times n}$

Diagonal Matrix

2-D array containing numbers or scalars along its main diagonal and zero elsewhere.

$$\mathbf{Y} = \text{diag}(\mathbf{x}) = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_n \end{bmatrix}_{n\times n}$$

Identity Matrix – $\mathbf{I}$

An identity matrix is a square matrix where elements on the main diagonal are 1 and 0 everywhere else.

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{n\times n}$$

$$\begin{aligned} \mathbf{I}\mathbf{x} &= \mathbf{x} \\ \mathbf{I}\mathbf{X} &= \mathbf{X} \end{aligned}$$

Block diagonal matrix

A block diagonal matrix concatenates several matrices on the diagonal of a global matrix.

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}$$

$$\text{blkdiag}(\mathbf{A}, \mathbf{B}) = \begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 & 0 \\ 0 & 0 & 4 & 5 & 6 \\ 0 & 0 & 7 & 8 & 9 \\ 0 & 0 & 10 & 11 & 12 \end{bmatrix}$$

: Matrices

```
import numpy as np
import scipy as sp

A = np.ones((3,2))                    #3x2 matrix filled with ones
B = np.eye(5)                         #5x5 identity matrix
C = np.array([[1, 2, 3], [4, 5, 6]])  #2x3 matrix
D = C[1,2]                            #D=6
E = C[0,:]                            #E=[1 2 3]
F = np.diag(E)                        #F=3x3 diagonal matrix
G = sp.linalg.block_diag(A,B)         #Block diagonal matrix
```

Section Outline

Operations

Matrix transposition

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix} \rightarrow \mathbf{X}^{\intercal} = \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ x_{13} & x_{23} \end{bmatrix}$$

$$[\mathbf{X}^{\intercal}]_{ij} = [\mathbf{X}]_{ji}$$

Matrix multiplication

$$
\begin{aligned}
\mathbf{C} &= \mathbf{AB}\\
&= \mathbf{A}\times\mathbf{B}\\
[\mathbf{C}]_{ij} &= \textstyle\sum_k [\mathbf{A}]_{ik}\cdot[\mathbf{B}]_{kj}\\
\mathbf{C}_{m\times n} &= \mathbf{A}_{m\times k}\mathbf{B}_{k\times n}
\end{aligned}
$$

$$
\mathbf{B}\quad \begin{bmatrix} b_{11} & \boxed{b_{12}} & \boxed{b_{13}} \\ b_{21} & \boxed{b_{22}} & \boxed{b_{23}} \end{bmatrix}
$$

$$
\mathbf{A}\ \begin{bmatrix} \boxed{a_{11} \quad a_{12}} \\ a_{21} \quad a_{22} \\ a_{31} \quad a_{32} \\ \boxed{a_{41} \quad a_{42}} \end{bmatrix}
\begin{bmatrix} & \boxed{c_{12}} & \\ & & \\ & & \\ & & \boxed{c_{43}} \end{bmatrix}
$$

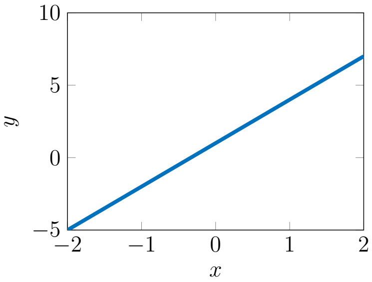
$$\boxed{c_{12}} = a_{11}b_{12} + a_{12}b_{22}$$

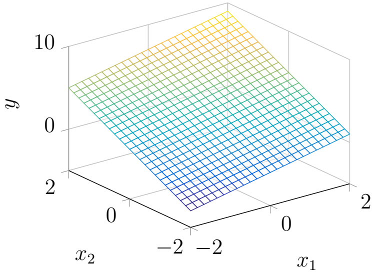
$$\boxed{c_{43}} = a_{41}b_{13} + a_{42}b_{23}$$

Multiplication rules

$$
\begin{aligned}
\mathbf{A}(\mathbf{B}+\mathbf{C}) &= \mathbf{AB}+\mathbf{AC} && \text{(Distributivity)}\\
\mathbf{A}(\mathbf{BC}) &= (\mathbf{AB})\mathbf{C} && \text{(Associativity)}\\
\mathbf{AB} &\neq \mathbf{BA} && \text{(Not commutative)}\\
(\mathbf{AB})^{\intercal} &= \mathbf{B}^{\intercal}\mathbf{A}^{\intercal} && \text{(Conjugate transposability)}\\
\mathbf{x}^{\intercal}\mathbf{y} &= \mathbf{x}\cdot\mathbf{y} && \text{(Inner product)}
\end{aligned}
$$

Linear systems





$$\begin{aligned} y &= 3x + 1 \\ &= ax + b \end{aligned}$$

$$a = 3, \;\; b = 1$$

$$\begin{aligned} y &= x_1 + 2x_2 + 3 \\ &= \mathbf{a}^\intercal \mathbf{x} + b \end{aligned}$$

$$\mathbf{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \;\; \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \;\; b = 3$$

Hadamar product – Element-wise product

$$\begin{aligned}\mathbf{C} &= \mathbf{A} \odot \mathbf{B}\\ [\mathbf{C}]_{ij} &= [\mathbf{A}]_{ij} \cdot [\mathbf{B}]_{ij}\\ \mathbf{C}_{m\times n} &= \mathbf{A}_{m\times n} \odot \mathbf{B}_{m\times n}\end{aligned}$$

Matrix inversion

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

To be invertible, a matrix **must be square** and **must not have linearly dependent rows or columns**.

For a system of equation

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{b} \\ \mathbf{A}^{-1}\mathbf{A}\mathbf{x} &= \mathbf{A}^{-1}\mathbf{b} \\ \mathbf{I}\mathbf{x} &= \mathbf{A}^{-1}\mathbf{b} \\ \mathbf{x} &= \mathbf{A}^{-1}\mathbf{b} \end{aligned}$$

: Operations

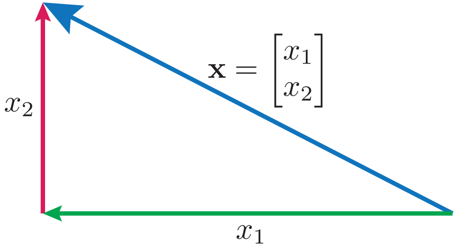
```
import numpy as np

A = np.ones((3,2))                #3x2 matrix filled with ones
B = np.array([[1, 2, 3],[4, 5, 6]])#2x3 matrix
C = B[0:1,0:1]
D = np.linalg.inv(C)              #D=inverse(C)
E = A @ B                         #Matrix multiplication
F = A * B.transpose()             #Element-wise multiplication
G = np.linalg.lstsq(F,B.transpose()[:,0][:,None])[0] #G=F^-1
                                  *B(1,:)
```

Definition

Norms measure how large a vector is. The L^p norm is

$$
\begin{aligned}
||\mathbf{x}||_p &= \left(\sum_i |[\mathbf{x}]_i|^p\right)^{\frac{1}{p}} && (L^p\text{-norm})\\
||\mathbf{x}||_2 &= \sqrt{\sum_i [\mathbf{x}]_i^2} \equiv \sqrt{\mathbf{x}^\intercal\mathbf{x}} && (\text{Euclidian norm})\\
||\mathbf{x}||_1 &= \sum_i |[\mathbf{x}]_i| && (\text{Manhattan norm})\\
||\mathbf{x}||_\infty &= \max_i |[\mathbf{x}]_i| && (\text{Max norm})
\end{aligned}
$$

Norms



$$\begin{aligned} ||\mathbf{x}||_2 &= \sqrt{x_1^2 + x_2^2} \\ ||\mathbf{x}||_1 &= |x_1| + |x_2| \\ ||\mathbf{x}||_\infty &= \max |x_1, x_2| = x_1 \end{aligned}$$

The L^2-norm (euclidian distance) is the most common

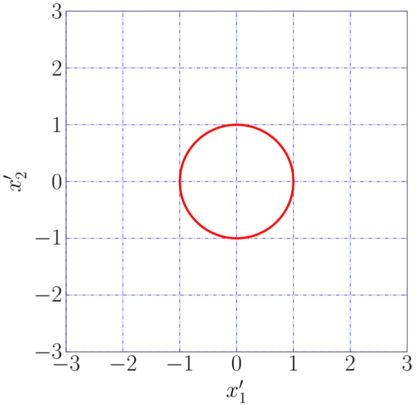
Section Outline

Transformations

Linear transformation [.m]

$$
\begin{aligned}
\mathbf{x}_1 &= [-1, -0.99, \cdots, 1]^{\intercal} \\
\mathbf{x}_2 : [\mathbf{x}_2]_i &= \sqrt{1 - [\mathbf{x}_1]_i^2} \\
\mathbf{X} &= [\mathbf{x}_1, \mathbf{x}_2] \\
\mathbf{X}' &= \mathbf{A}\mathbf{X}^{\intercal}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{A} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
\det(\mathbf{A}) &= 1
\end{aligned}
$$

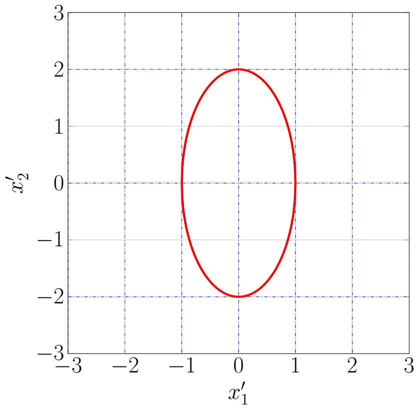


[CIVML/linear_transformation.m]

Linear transformation [.m]

$$
\begin{aligned}
\mathbf{x}_1 &= [-1, -0.99, \cdots, 1]^\intercal \\
\mathbf{x}_2 : [\mathbf{x}_2]_i &= \sqrt{1 - [\mathbf{x}_1]_i^2} \\
\mathbf{X} &= [\mathbf{x}_1, \mathbf{x}_2] \\
\mathbf{X}' &= \mathbf{A}\mathbf{X}^\intercal
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{A} &= \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \\
\det(\mathbf{A}) &= 2
\end{aligned}
$$

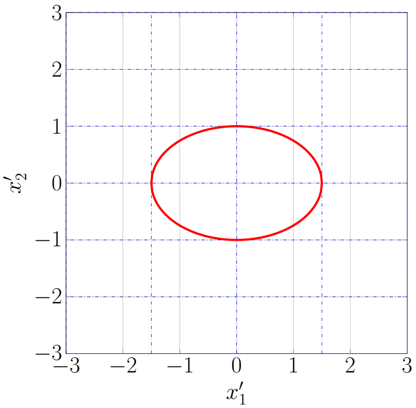


[CIVML/linear_transformation.m]

Linear transformation [.m]

$$
\begin{aligned}
\mathbf{x}_1 &= [-1, -0.99, \cdots, 1]^\intercal \\
\mathbf{x}_2 : [\mathbf{x}_2]_i &= \sqrt{1 - [\mathbf{x}_1]_i^2} \\
\mathbf{X} &= [\mathbf{x}_1, \mathbf{x}_2] \\
\mathbf{X}' &= \mathbf{A}\mathbf{X}^\intercal
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{A} &= \begin{bmatrix} 1.5 & 0 \\ 0 & 1 \end{bmatrix} \\
\det(\mathbf{A}) &= 1.5
\end{aligned}
$$

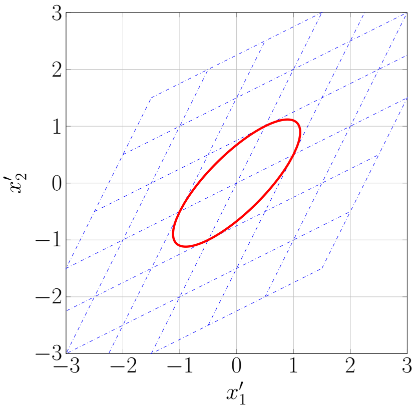


[CIVML/linear_transformation.m]

Linear transformation [.m]

$$
\begin{aligned}
\mathbf{x}_1 &= [-1, -0.99, \cdots, 1]^\intercal \\
\mathbf{x}_2 : [\mathbf{x}_2]_i &= \sqrt{1 - [\mathbf{x}_1]_i^2} \\
\mathbf{X} &= [\mathbf{x}_1, \mathbf{x}_2] \\
\mathbf{X}' &= \mathbf{A}\mathbf{X}^\intercal
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{A} &= \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \\
\det(\mathbf{A}) &= 0.75
\end{aligned}
$$

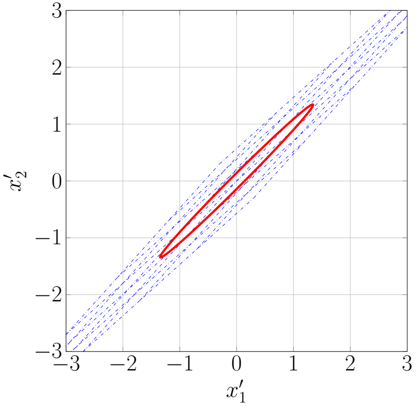


[CIVML/linear_transformation.m]

Linear transformation [.m]

$$\mathbf{x}_1 = [-1, -0.99, \cdots, 1]^\intercal$$
$$\mathbf{x}_2 : [\mathbf{x}_2]_i = \sqrt{1 - [\mathbf{x}_1]_i^2}$$
$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2]$$
$$\mathbf{X}' = \mathbf{A}\mathbf{X}^\intercal$$

$$\mathbf{A} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$
$$\det(\mathbf{A}) = 0.19$$



[CIVML/linear_transformation.m]

Determinant – $\det(\mathbf{A})$

For a square matrix $[\mathbf{A}]_{n\times n}$, $\det(\mathbf{A}) : \mathbb{R}^{n\times n} \rightarrow \mathbb{R}$.

The determinant **measures how much the matrix contracts or expands the space**.

- $\det(\mathbf{A}) = 1$: preserves the space/volume
- $\det(\mathbf{A}) = 0$: contracts the space/volume along 1-D

The determinant is the product of the eigenvalues of a matrix

Section Outline

Eigen decomposition

Eigen decomposition

A square matrix $[\mathbf{A}]_{n\times n}$ can be decomposed in **eigenvectors** $\{\boldsymbol{\nu}_1,\cdots,\boldsymbol{\nu}_n\}$ and **eigenvalues** $\{\lambda_1,\cdots,\lambda_n\}$.

In their matrix form,

$$\begin{array}{rcl}\mathbf{V} & = & [\boldsymbol{\nu}_1,\cdots,\boldsymbol{\nu}_n]\\ \boldsymbol{\lambda} & = & [\lambda_1,\cdots,\lambda_n]^\intercal\end{array}$$

$$\mathbf{A}=\mathbf{V}\mathrm{diag}(\boldsymbol{\lambda})\mathbf{V}^{-1}$$

A matrix is **positive definite** if all eigenvalues >0
A matrix is **positive semidefinite** if all eigenvalues ≥ 0

For positive semidefinite (PSD) matrices

$$\boxed{\forall\mathbf{x},\mathbf{x}^\intercal\mathbf{A}\mathbf{x}\geq 0}$$

Example – Eigendecomposition [.m]

$$\mathbf{x}_1 = [-1, -0.99, \cdots, 1]^\intercal$$

$$\mathbf{x}_2 : [\mathbf{x}_2]_i = \sqrt{1 - [\mathbf{x}_1]_i^2}$$

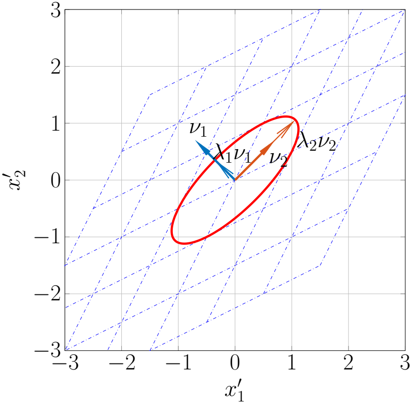
$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2]$$

$$\mathbf{X}' = \mathbf{A}\mathbf{X}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

$$\mathbf{V} = \begin{bmatrix} -0.71 & 0.71 \\ 0.71 & 0.71 \end{bmatrix}$$

$$\boldsymbol{\lambda} = [0.5, 1.5]^\intercal$$



[CIVML/linear_transformation_eig.m]

: Eigen decomposition

```
import numpy as np

A=[[1, 0.5],[0.5, 1]] #2x2 matrix
V = np.linalg.eig(A)  #Eigen decomposition of A
V[1]                  #matrix of eigenvectors
lambda_ = V[0]        #vector of eigenvalues
```

Summary

Matrix transposition

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix} \rightarrow \mathbf{X}^{\intercal} = \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ x_{13} & x_{23} \end{bmatrix}$$

Matrix multiplication

$$\begin{aligned} \mathbf{C} &= \mathbf{AB} \\ &= \mathbf{A} \times \mathbf{B} \\ [\mathbf{C}]_{ij} &= \textstyle\sum_k [\mathbf{A}]_{ik} \cdot [\mathbf{B}]_{kj} \\ \mathbf{C}_{m\times n} &= \mathbf{A}_{m\times k}\mathbf{B}_{k\times n} \end{aligned}$$

Element-wise product

$$\begin{aligned} \mathbf{C} &= \mathbf{A} \odot \mathbf{B} \\ [\mathbf{C}]_{ij} &= [\mathbf{A}]_{ij} \cdot [\mathbf{B}]_{ij} \\ \mathbf{C}_{m\times n} &= \mathbf{A}_{m\times n} \odot \mathbf{B}_{m\times n} \end{aligned}$$

Matrix inversion

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

Norm of a vector

$$||\mathbf{x}||_2 = \sqrt{\sum_i [\mathbf{x}]_i^2} \equiv \sqrt{\mathbf{x}^{\intercal}\mathbf{x}}$$

Determinant

$\det(\mathbf{A}) : \mathbb{R}^{n\times n} \rightarrow \mathbb{R}$ measures how much the matrix **contracts or expands the space**

Eigendecomposition

A matrix is **positive semidefinite** if all eigenvalues ≥ 0. For positive semidefinite (PSD) matrices

$$\boxed{\forall \mathbf{x}, \mathbf{x}^{\intercal}\mathbf{A}\mathbf{x} \geq 0}$$