# Lecture 10

## Gaussian random vectors and stochastic linear systems

Giancarlo Ferrari Trecate[1]

[1]Dependable Control and Decision Group
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
giancarlo.ferraritrecate@epfl.ch

# Gaussian random variable (RV)

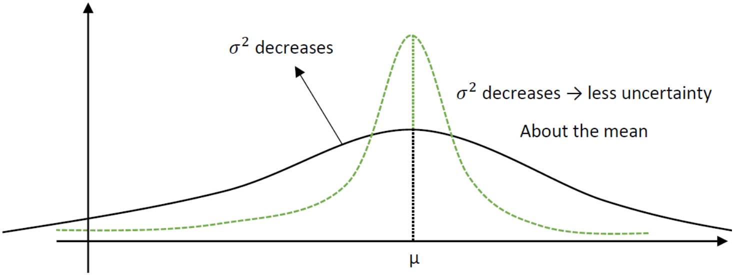
- $x \in \mathbb{R}$ is a Gaussian RV if its probability density is

$$f(q) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(q-\mu)^2}{2\sigma^2}}$$

- Notation: $x \sim N(\mu, \sigma^2)$

By construction

$$\mu = E[x] = \text{mean} = \int_{\mathbb{R}} x f(x)\, dx$$
$$\sigma^2 = E\left[(x-\mu)^2\right] > 0 = \text{variance}$$



# Gaussian random vector

- $x = [x_1, \dots, x_n]^T$ is a Gaussian random vector if its probability density is

$C \in \mathbb{R}^{n \times n}$

$$f(q) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{det(C)}} e^{-\frac{1}{2}(q-\mu)^T C^{-1}(q-\mu)} \quad , \quad q \in \mathbb{R}^n$$

$\begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$

where $\mu \in \mathbb{R}^n$ and $C = C^T \in \mathbb{R}^{n \times n}$ is positive-definite

- $x_1, \dots, x_n$ are also called jointly Gaussian

## Remark

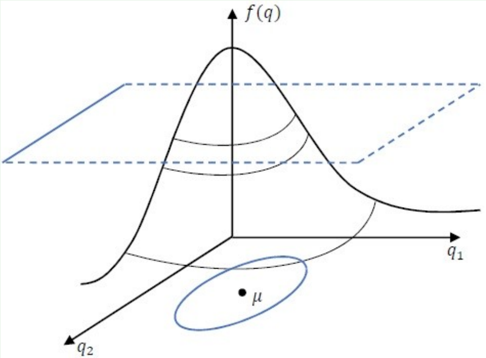
Possible to assume that C is positive-semidefinite by defining a Gaussian random vector using probability distributions instead of densities

By construction :

$$\mu = E[x] = \int_{\mathbb{R}^n} q f(q) dq_1 \dots dq_n \qquad \text{mean}$$

$$C = Var[x] = E[(x-\mu)(x-\mu)^T] \qquad \text{variance}$$

## Example




$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

New quantity: $C_{ij} = Cov[x_i, x_j] = E[(x_i - \mu_i)(x_j - \mu_j)]$ (for $i \neq j$)

- $\hookrightarrow$ $C_{ij} \neq 0$ means that the knowledge of $x_i$ brings information on the distribution of $x_j$ and vice-versa
- $\hookrightarrow$ $x_1, \ldots, x_n$ are uncorrelated if $C_{ij} = 0$ $\forall i \neq j$ (diagonal variance)
  - $\hookrightarrow$ For jointly Gaussian RVs incorrelation is the same as statistical independence

1h 19

# Marginal and conditional density

Let $X \in \mathbb{R}^n$, $Y \in \mathbb{R}^m$ and

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left( \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} C_{XX} & C_{XY} \\ C_{YX} & C_{YY} \end{bmatrix} \right) = f_{XY}(x, y)$$

- $X$ is a Gaussian RV, i.e. the marginal density

quantifies the uncertainty about $X$ if one does not measure $y$

$$f_X(x) = \int_{\mathbb{R}^m} f_{XY}(x, y)\, dy$$

is the Gaussian $N(\mu_x, C_{XX})$

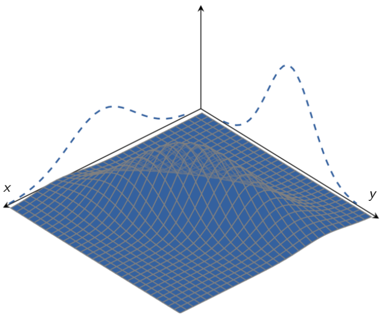
- The conditional density of $X$ for a measured value of Y is

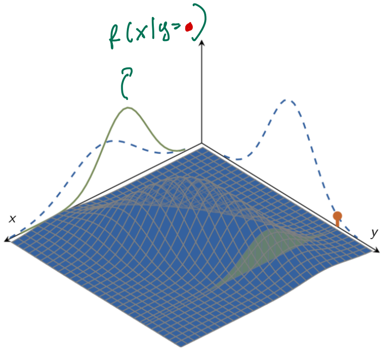
$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

→ quantifies how uncertainty on $X$ changes because $Y$ is no longer random

↪ Notation: $X|Y$ for denoting the random vector $X$ equipped with the conditional density $f_{X|Y}$

# Marginal and conditional density[1]




Two-dimensional Gaussian density and the marginal density for each component (dashed blue lines along each axis)
Remark: the marginal density do *not* contain all information about $f_{XY}(x, y)$, since the covariance information is lacking in that representation.

Conditional distribution of $X$ (green line), when $Y$ is observed (orange dot)
The conditional distribution of $x$, apart from a normalizing constant, is the green 'slice' of the joint distribution.

[1] F. Lindsten, T. Schön, A. Svensson and N. Wahlström. *Probabilistic modeling - linear regression and Gaussian processes*

## Proposition

$\begin{bmatrix} X \\ Y \end{bmatrix}$ is Gaussian $\Rightarrow f_{X|Y}$ is Gaussian with

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix} \right)$$

$$E[X|Y] = \overbrace{\mu_X + C_{XY} C_{YY}^{-1} (y - \mu_Y)}^{(a)} \quad \text{"a posteriori" mean}$$

$$Var[X|Y] = \underbrace{C_{XX} - C_{XY} C_{YY}^{-1} C_{YX}}_{(b)} \quad \text{"a posteriori" variance}$$

(a): Shift in the mean

(b): "reduction" of the original uncertainty $C_{XX}$

## Definition

$X$ and $Y$ are uncorrelated if $C_{XY} = 0$

$\hookrightarrow$ Then:

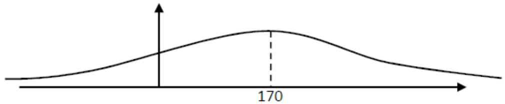
$$\left.\begin{aligned} Var[X|Y] &= Var[X] \\ E[X|Y] &= E[X] \\ f_{X|Y}(x|y) &= f_X(x) \end{aligned}\right\} \text{Knowing } Y \text{ does not bring any information on } X$$

## Example

$X$ = height, $Y$ = weight. Assume $\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left( \begin{bmatrix} 170 \\ 65 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right)$

$c_{12} = 0.5 \rightarrow$ makes sense: higher students weight more

Problem: which is the height density for a student weighting 110 Kg?




$f_X$ = prior height density $\sim N(170, 1)$

$f_{X|Y}$ = posterior height density when $Y = 110$

$$E[X|Y] = 170 + 0.5(110 - 65) = 192.5$$

$$Var(X|Y) = 1 - 0.5^2 = 0.75$$

# Affine transformation of a Gaussian random vector

## Proposition

If $x = [x_1, \ldots, x_n]^T \sim N(\mu_x, C_x)$ and

$$y = Ax + b$$

Example

$$\begin{bmatrix} A \end{bmatrix} \begin{bmatrix} x \end{bmatrix} + \begin{bmatrix} b \end{bmatrix}$$

where $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$, then

(a) $y \in \mathbb{R}^m$ is Gaussian with

$$E[y] = A\mu_x + b \qquad (*)$$

$$Var[y] = AC_xA^T \qquad (**)$$

(b) $z = \begin{bmatrix} x \\ y \end{bmatrix}$ is Gaussian with

$$E[z] = \begin{bmatrix} \mu_x \\ A\mu_x + b \end{bmatrix}$$

$$Var[z] = \begin{bmatrix} C_x & C_xA^T \\ AC_x & AC_xA^T \end{bmatrix}$$

# Affine transformation of a Gaussian random vector

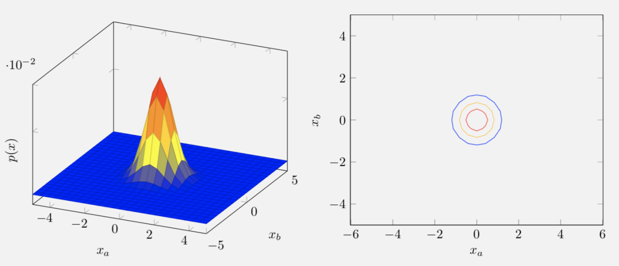
## Proof of (b)

$$z = \begin{bmatrix} I \\ A \end{bmatrix} x + \begin{bmatrix} 0 \\ b \end{bmatrix}$$

Apply $(*)$, $(**)$ with the substitutions: $A \to \begin{bmatrix} I \\ A \end{bmatrix}$, $b \to \begin{bmatrix} 0 \\ b \end{bmatrix}$

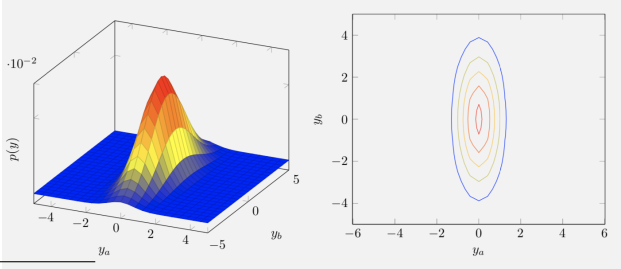
# Example[2]

Consider a two-dimensional Gaussian random vector $x = \begin{bmatrix} x_a \\ x_b \end{bmatrix} \sim N(\mu_x, C_x)$ where $\mu_x = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $C_x = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$




Do a linear transformation $y = A_1 x$ where $A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$. The random vector $y$ will have a Gaussian density with $y = \begin{bmatrix} y_a \\ y_b \end{bmatrix} \sim N(A_1 \mu_x, A_1 C_x A_1^T) = N(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 9 \end{bmatrix})$

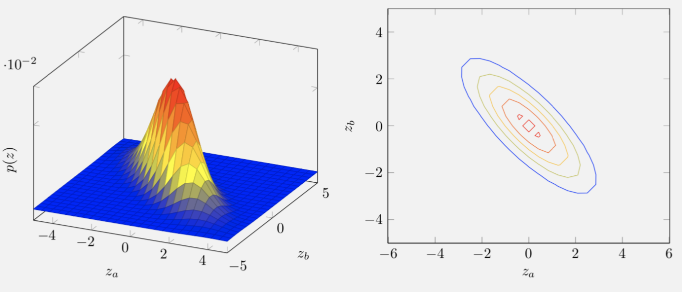
It can be seen that the distribution is scaled in the $y_b$ direction.




[2] F. Lindsten, T. Schön, A. Svensson and N. Wahlström. *Probabilistic modeling - linear regression and Gaussian processes*

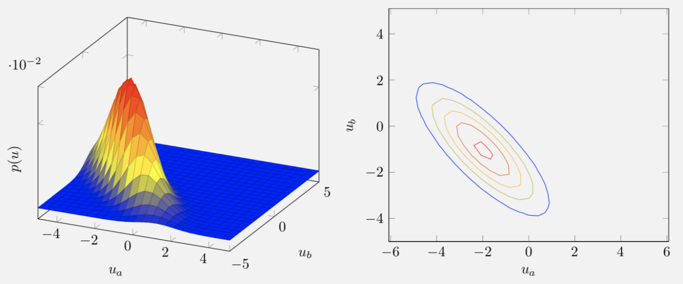
# Example

Do another linear transformation $z = A_2 y$, this time a rotation of $45^\circ$ where $A_2 = \begin{bmatrix} \cos 45^\circ & -\sin 45^\circ \\ \sin 45^\circ & \cos 45^\circ \end{bmatrix}$. The random variable $z$ will be distributed as $z \sim N(A_2 A_1 \mu_x, A_2 A_1 C_x A_1^T A_2^T) = N(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix})$. Consequently, also the density will be rotated.



# Example

Finally, consider a translation with $u = z + b$ where $b = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$. The final distribution will be
$u \sim N(A_2 A_1 \mu_x + b, A_2 A_1 C_x A_1^T A_2^T) = N(\begin{bmatrix} -2 \\ -1 \end{bmatrix}, \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix})$, i.e., the density will be shifted accordingly.



# Linear systems driven by Gaussian noise

$$x_{k+1} = Ax_k + w_k$$
$$y_k = Cx_k + v_k \quad \rightarrow \text{ observed output}$$
$$x_0 \sim N(\bar{x}_0, \Sigma_0)$$

- $w_k \in \mathbb{R}^n$: process noise (random vector)
- $v_k \in \mathbb{R}^p$: measurement noise (random vector)

## Standard statistical assumptions (from now on...)

1) $x_0, w_1, w_2, \ldots, v_1, v_2, \ldots$ are jointly Gaussian and independent
2) $w_k$ are iid (independent and identically distributed) with
$$E[w_k] = 0 \text{ and } Var[w_k] = W \geq 0$$
3) $v_k$ are iid with $E[v_k] = 0$ and $Var[v_k] = V > 0$

Example:

$$\xi = \begin{bmatrix} x_0 \\ v_2 \\ w_4 \end{bmatrix}$$

$$\xi \sim N\left(\begin{bmatrix} \bar{x}_0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_0 & 0 & 0 \\ 0 & V & 0 \\ 0 & 0 & W \end{bmatrix}\right)$$

## Definition

A stochastic process $w_k$ where $w_1, w_2, \ldots$ are jointly Gaussian and verify assumption 2 is called White Gaussian Noise (WGN) with variance $W$.
Notation $w \sim \text{WGN}\,(0, W)$

Define $X_k = \begin{bmatrix} x_0 \\ \vdots \\ x_k \end{bmatrix}$, $Y_k = \begin{bmatrix} y_0 \\ \vdots \\ y_k \end{bmatrix}$ etc...

$x_1 = Ax_0 + w_0$
$x_2 = A^2 x_0 + A w_0 + w_1$
$\vdots$

**Remark**

$X_k$ and $Y_k$ are linear combinations of $x_0, W_k, V_k$. Hence, $\begin{bmatrix} X_k \\ Y_k \end{bmatrix}$ is Gaussian

Statistical properties (under the standard assumptions)

- $v_k$ is independent of $x_j$, $\forall j \geq 0$ → $v_k$ acts only on $y_k$, not on $x_k$
- $w_k$ is independent of $X_k$ and $Y_k$ → $w_k$ influences $x_{k+1}$
- Markov property:

$$x_k | X_{k-1} = x_k | x_{k-1}$$

↪ follows from the state equation: if one knows $x_{k-1}$, the knowledge of $X_{k-2}$ does not bring additional information on $x_k$

# Mean and variance of $x_k$

The mean $\bar{x}_k = E[x_k]$ verifies

$$\bar{x}_{k+1} = A\bar{x}_k$$

i.e. the system dynamics

Proof:

$$E[x_{k+1}] = E[Ax_k + w_k] = AE[x_k] + \underbrace{E[w_k]}_{=0}$$

The variance $P_k = E[(x_k - \bar{x}_k)(x_k - \bar{x}_k)^T]$ verifies

$$P_{k+1} = AP_kA^T + W$$

Proof:

$$\begin{aligned} P_{k+1} &= E[\overbrace{(A(x_k - \bar{x}_k) + w_k)}^{x_{k+1} - E[x_{k+1}]}(A(x_k - \bar{x}_k) + w_k)^T] \\ &= AE[(x_k - \bar{x}_k)(x_k - \bar{x}_k)^T]A^T \\ &\qquad + 2 \underbrace{E[w_k(x_k - \bar{x}_k)^T]A^T}_{=0 \text{ as } w_k \text{ and } x_k \text{ are uncorrelated}} + \underbrace{E[w_k w_k^T]}_{W} \end{aligned}$$

## Lemma

If $A$ is Schur, $P_k$ converges, as $k \to +\infty$, to the solution P of the Lyapunov equation

$$P = APA^T + W$$

## Remark

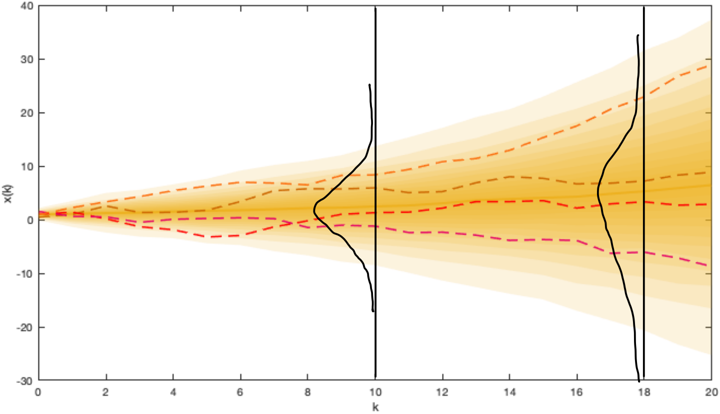
$P$ characterises the steady-state process.

# Example

Let consider the first order unstable system $x^+ = 1.2x + w$ where $x_0 \sim N(1, 0.5)$ and $w \sim \cancel{N}(0, 1)$. In the plot:

WGN(0, 1)

- for each $k$, the Gaussian density of $x(k)$ using color shades
- four samples of state trajectories in red dashed lines



# Example

Change for the stable system $\quad x^+ = 0.9x + w$

where $x_0 \sim N(9, 0.5)$ and $w \sim N(0, 1)$

- the variance $P_k$ converges since $A = 0.9$ is Schur
- solving the Lyapunov equation $\rightarrow P = 5.26$
- gray dashed lines: 95% confidence intervals associated with $P$