Skolkovo Institute of Science and Technology

CAIT
Mobile Robotics Lab
Perception in Robotics Course

L02: Gaussians

Gonzalo Ferrer

31 January 2023

1 Uni-variate Gaussian

The uni-variate (1 dimension) *probability density function* (PDF) of a Gaussian is defined as:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} = \mathcal{N}(x;\mu,\sigma^2),$$

where $\frac{1}{\sqrt{2\pi\sigma^2}}$ is a normalization factor such that $\int \mathcal{N}(x;\mu,\sigma^2)dx = 1$

If we plot this PDF, we can observe interesting facts:

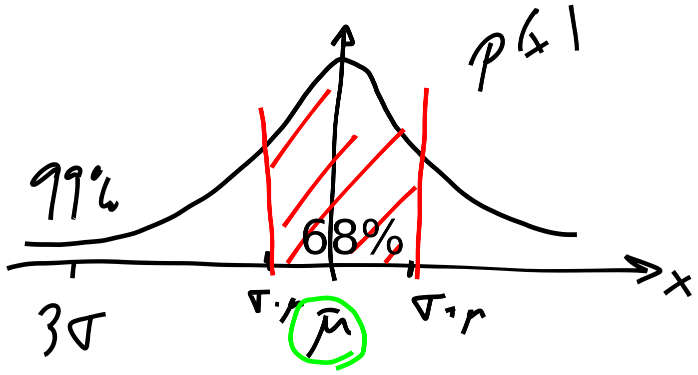

Figure 1: Uni-variate Gaussian probability density function PDF.

- 68% of the probability mass in contained within the sigma interval:

$$\int_{\mu-\sigma}^{\mu+\sigma} p(x)dx = 0.68$$

- 99.7% of its probability mass is within the 3-σ interval
- the mean μ, median and mode are one and the same value.

The cumulative distribution function CDF in figure 2

$$F_X(x) = \int_{-\infty}^{x} p(x)dx$$

Convention on notation: We will refer to a random variable x distributed by a particular PDF as $x \sim p(x)$.

All required parameters to describe a Gaussian PDF are its mean and covariance:

$$\mu = \mathbb{E}\{x\} = \int_{-\infty}^{\infty} x\, p(x)dx$$

$$\sigma^2 = \mathrm{cov}(x) = \mathbb{E}\{(x-\mu)^2\}$$

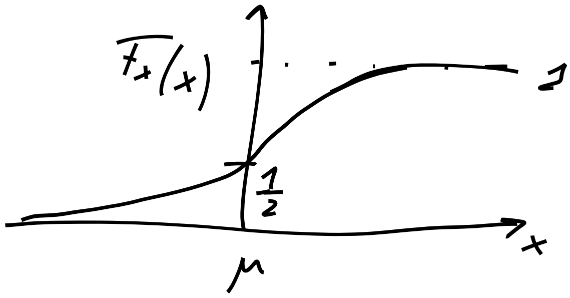

Figure 2: Univariate Gaussian cumulative distribution function CDF.

2 Multivariate Gaussian

The Gaussian probability density function of a multivariate random variable $x \in \mathbb{R}^N$ is:

$$p(x) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)} = \mathcal{N}(x; \mu, \Sigma).$$

Example: Intuition on a 2D Gaussian PDF

$$\Sigma_x = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

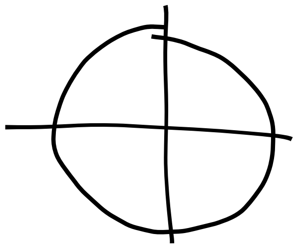

Figure 3: Independent variables x_1 and x_2, and hence, centered iso-contour

$$\Sigma_x = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

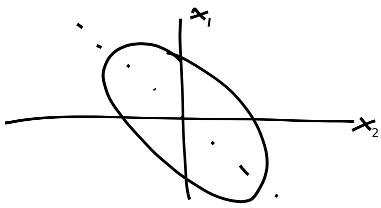

Figure 4: Negative correlation between variables x_1 and x_2.

3 Covariance Propagation

Given the r.v. $x \sim \mathcal{N}(\mu_x, \Sigma_x)$, the new r.v. $y = f(x)$.

3.1 Affine transformation f

We will start studying covariance propagation with the simplest example of an affine function f:

$$y = f(x) = Ax + b,$$

where $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^M$.

The mean and covariance of this new random variable can be obtained by simply applying the definition of the expectation operator:

$$\Sigma_x = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

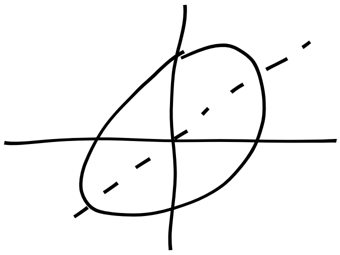

Figure 5: Positive correlation between variables x_1 and x_2.

$$\mu_y = \mathbb{E}\{y\} = \mathbb{E}\{Ax + b\} = A \cdot \mathbb{E}\{x\} + b = A\mu_x + b.$$

$$\begin{aligned}\Sigma_y &= \mathbb{E}\{(y - \mu_y)(y - \mu_y)^\top\} \\ &= \mathbb{E}\{(Ax + b - A\mu_x - b)(Ax + b - A\mu_x - b)^\top\} \\ &= \mathbb{E}\{A \underbrace{(x - \mu_x)(x - \mu_x)^\top}_{\Sigma_x} A^\top\} = A \cdot \Sigma_x \cdot A^\top.\end{aligned}$$

In order to obtain the previous expression, we have made use of the matrix property $(AB)^\top = B^\top A^\top$. This kind of manipulations will be a repeated topic during this course so it is recommended that the student practices manipulation of matrices and vectors when calculating the expectation.

An important result from this subsection is that we have obtained a mean and covariance after an affine transformation from the original parameters (mean and covariance) of the r.v. x. Since x follows a Gaussian PDF, the prob. density function of the r.v. y will be Gaussian as well, totally determined by the new mean μ_y and covariance Σ_y.

3.2 Non-Linear covariance propagation

Now, we will consider a non-linear transformation of the form

$$y = f(x).$$

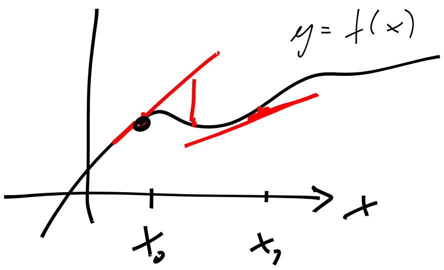

Figure 6: Example of a non-linear function and different linearization points, and their corresponding derivatives in red line.

As it can be seen in the graphic, one way of overcoming this problem is by linearizing around the point x_0. This approach is not exempt of problems, such as an implicit error in this approximation.

1D case: In this example, we can linearize and obtain the following expression:

$$y = f(x_0) + \underbrace{\left.\frac{d\,f(x)}{d\,x}\right|_{x_0}} (x - x_0) + O\big((x - x_0)^2\big),$$

where the first order derivative is computed analytically or numerically and the error in the approximation, denoted by $O(\cdot)$, is proportional to $(x - x_0)^2$.

One can improve the accuracy of this approximation, up to second order, third, etc. resulting in the limit in an infinite series, the Taylor expansion.

N-dimensions case: Now, the variables are multi-dimensional, $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^M$.

$$y = f(x_0) + \underbrace{\sum_{i=1}^{N} \frac{\partial f(x)}{\partial x^i}\bigg|_{x_0}} (x^i - x_0^i) + O\big(||x - x_0||^2\big)$$

$$\begin{aligned}
&\text{Jacobian:} \quad \left[\tfrac{\partial f}{\partial x^1} \quad \tfrac{\partial f}{\partial x^2} \quad \dots \quad \tfrac{\partial f}{\partial x^N}\right] = J \\
&= f(x_0) + J \cdot (x - x_0) + O\big(||x - x_0||^2\big) \\
&\simeq f(x_0) + Jx - Jx_0 = \underbrace{J}_{A} \cdot x + \underbrace{f(x_0) - J \cdot x_0}_{b}
\end{aligned}$$

This result highly resembles the affine transformed discussed above. One can choose the linearization point as $x_0 = \mu_x$, since it is around the mean of a Gaussian where most of its probability mass lies (Sec.2). Then,

$$\implies y \sim \mathcal{N}(f(\mu_x), J \cdot \Sigma_x \cdot J^\top).$$

We have finally obtained, after linearizing around a chosen point (mean) that a non-linear transformation of a Gaussian PDF is approximated as a Gaussian as well.

This approach however is not exempt of problems, the first one (and this will be recursive in this course) is the error introduced by linearizing, which could be problematic for highly non-linear functions and disperse (high covariance) density functions.

We will discuss possible alternatives to first order covariance propagation in lecture 7 and lecture 8.

4 Visualizing Gaussians, 2D case

Given a Gaussian PDF

$$p(x) = \alpha \cdot e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)}$$

the problem is to find the lines or contours of constant $p(x)$:

$$k - \text{sigma isocontour} = \left\{x \,:\, k^2 = (x-\mu)^\top \Sigma^{-1}(x-\mu)\right\}, \quad k = 1, 2, 3, \dots$$

Usually, we will use the $1 - \sigma$ (1-sigma) or $3 - \sigma$ iso-contours, for being the most significant.

As it can be seen in the sketch, iso-contours are the equations of a quadratic curve, en ellipse.

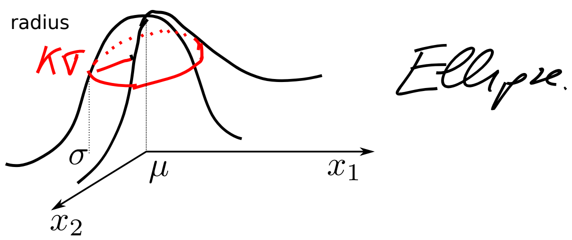

Figure 7: Iso-contour of $p(x)$ in two dimensions x_1 and x_2.

Mahalanobis distance

The Mahalanobis distance is tightly related to the Gaussian PDF and effectively what it measures is the distance, weighted by the uncertainty of the r.v.'s:

$$||x-w||_\Sigma \doteq \sqrt{(x-w)^\top \Sigma^{-1}(x-w)}$$

4.1 How to visualize an iso-contour

For the case that Σ is the identity matrix, the solution of the iso-contour equation is a circle in 2D with radius 1. One can obtain analytically this solution. Even for diagonal covariances, we know just by inspection (variables are independent and thus uncorrelated) that the solution will be an ellipsoid aligned with the axis and its solution requires only calculating one point for each axis.

Unfortunately, we are interested in the more general case when Σ is non-diagonal. For that, we will make use of the *covariance projection* technique and Linear Algebra.

The idea for plotting iso-contours is simple: we will transform an easy-to-obtain density, such as a Gaussian with covariace the identity, and then we will transform this circle into an ellipsoid. More formally:

Problem: Find an affine transformation for a given $x \sim \mathcal{N}(0, I)$, such that it propagates into the desired mean μ_y and covariance Σ_y:

$$\begin{aligned} &\text{if} \quad y = Ax + b, \quad x \sim \mathcal{N}(0, I) \\ &\text{then} \quad \Sigma_y = A\Sigma_x A^\top = AIA^\top = AA^\top. \end{aligned}$$

There are two alternatives for solving this problem:

1. SVD decomposition. Remember from L.A., that SVD and eigen decomposition provide the same result if the input matrix is p.s.d. and symmetric (all covariances are):

$$\Sigma_y = UDV^\top = UDU^\top = \underbrace{UD^{1/2}}_{A}\underbrace{D^{T/2}U^\top}_{A^\top}.$$

2. Cholesky decomposition (More efficient): For symmetric and p.d. matrices, Cholesky results in a product of a lower triangular matrix and its transpose:

$$\Sigma_y = L \cdot L^\top.$$

Example: Find the $1-\sigma$ iso-contour of $\Sigma = \begin{bmatrix} 4 & -2 \\ -2 & 10 \end{bmatrix}$:

$$\Sigma = \underbrace{\begin{bmatrix} a & 0 \\ b & c \end{bmatrix}}_{L}\underbrace{\begin{bmatrix} a & b \\ 0 & c \end{bmatrix}}_{L^\top} = \begin{bmatrix} 4 & -2 \\ -2 & 10 \end{bmatrix}$$

$$\begin{aligned} a^2 = 4 &\implies a = \pm 2 \\ ba = ab = -2 &\implies b = -1 \\ b^2 + c^2 = 10 &\implies c = \sqrt{10 - b^2} = \pm 3. \end{aligned}$$

If we choose positive values for diagonal $\implies \exists! L$ (exists a unique solution L).
Then, we project point from the circumference $r = 1$:

$$\begin{bmatrix} 2 & 0 \\ -1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

$$L \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

Note: Solution is centered at μ, so we need to translate accordingly.

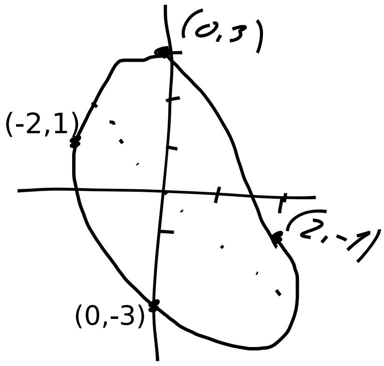

Figure 8: Example of the sketched iso-contour. With 2 points from the circle (and its negatives) we can drawn an approximate solution of the 1-sigma iso-contour. It also helps thinking on the existing correlation between variables.

5 Sampling from Gaussians

Most libraries have an implementation of the uni-dimensional Gaussian (or normal) PDF $\mathcal{N}(0,1)$. For instance, `numpy.random.randn()`.

Problem: how to sample $y \sim \mathcal{N}(\mu_y, \Sigma_y)$?

We will use covariance propagation of an affine transformation $y = Ax + b$

1. Sample from the standard $x \sim \mathcal{N}(0, I)$. This is equivalent to sample individually N times on the function, since the variables are independent and identically distributed (iid)

$$x = \begin{bmatrix} \mathcal{N}(0,1) \\ \mathcal{N}(0,1) \\ \dots \\ \mathcal{N}(0,1) \end{bmatrix}$$

2. Find the mean such that

$$y \sim \mathcal{N}(A\mu_x + b, A\Sigma_x A^\top) \implies b = \mu_y.$$

3. Find A such that $A \cdot A^\top = \Sigma_y$ (Cholesky)

4. Transform each of the samples from $x \sim \mathcal{N}(0, I)$ according to the affine function $y = Ax + b$, whose corresponding PDF will be equivalent to $y \sim \mathcal{N}(\mu_y, \Sigma_y)$.