Localization: Part 5
Covariance Matrices and the Multivariate Normal Distribution

Computer Science 4766/6912

Department of Computer Science
Memorial University of Newfoundland

July 11, 2018

To discuss the Kalman filter, we first need to discuss the **multivariate normal distribution**...

To discuss the multivariate normal distribution, we first need to discuss **covariance matrices**...

To discuss covariance matrices, we first need to *review* **covariance**...

Covariance

The effect of one random variable on another is expressed by the **covariance** between the two variables; This is defined as,

$$\begin{aligned} \sigma_{ij} &= Cov(X_i, X_j) = E((X_i - \mu_{x_i})(X_j - \mu_{x_j})) \\ &= E(X_i X_j) - \mu_{X_i}\mu_{X_j} \end{aligned}$$

Unlike variance, covariance can be positive or negative:

- Positive covariance implies that large values of X_i are associated with large values of X_j
- Negative covariance implies that large values of X_i are associated with small values of X_j, and vice versa

Related to covariance, is the **correlation coefficient**, which indicates a linear relationship between two random variables,

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

SPREADSHEET DEMO

Covariance: Other Properties

Variance can be considered a special case of covariance,

$$Var(X) = Cov(X, X)$$

If two random variables are independent,

$$Cov(X_1, X_2) = 0$$

Enough about covariance for now... What about covariance matrices? They come up more naturally if we look at functions of random variables...

Functions of Random Variables

- If a single random variable passes through a linear function, we can determine the expected value and variance of the output
  Let $y = f(x) = a \cdot x + b$

$$\begin{aligned} \mu_y &= E[a \cdot X + b] = a \cdot \mu_x + b \\ \sigma_y^2 &= E[(a \cdot X + b - \mu_y)^2] \\ &= a^2 \cdot \sigma_x^2 \end{aligned}$$

- What if the function involves some linear combination of random variables?
  e.g. Let $y = f(x_1, x_2) = a \cdot x_1 + b \cdot x_2$

$$\begin{aligned} \mu_y &= a \cdot \mu_{x_1} + b \cdot \mu_{x_2} \\ \sigma_y^2 &= a^2 \cdot \sigma_{x_1}^2 + b^2 \cdot \sigma_{x_2}^2 + 2ab \cdot \sigma_{x_1 x_2} \end{aligned}$$

- What if we have multiple outputs? Things get complicated and it becomes helpful to introduce the...

Covariance Matrix

Assume we have two linear functions, producing two outputs from two inputs,

$$\begin{aligned} y_1 &= a x_1 + b x_2 \\ y_2 &= c x_1 + d x_2 \end{aligned}$$

We can introduce the vectors $\boldsymbol{y} = [y_1, y_2]^T$ and $\boldsymbol{x} = [x_1, x_2]^T$; The system can now be represented in matrix notation,

$$\begin{aligned} \boldsymbol{y} &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} \boldsymbol{x} \\ &= A\boldsymbol{x} \end{aligned}$$

It can easily be shown that,

$$\mu_{\boldsymbol{y}} = A \mu_{\boldsymbol{x}}$$

To talk about the uncertainty of the output, we must discuss uncertainty in the input—which is given by $\sigma_{x_1}^2$, $\sigma_{x_2}^2$, and $\sigma_{x_1 x_2}$. First we organize these values into a **covariance matrix**,

$$\Sigma_{\boldsymbol{x}} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1 x_2} \\ \sigma_{x_1 x_2} & \sigma_{x_2}^2 \end{bmatrix}$$

(Note that the symbol Σ will be used to represent covariance matrices. It should be clear from the context whether a covariance matrix or a summation operator is intended.)

We now want to know $\Sigma_{\boldsymbol{y}}$ for our system; It turns out that this is given by,

$$\Sigma_{\boldsymbol{y}} = A \Sigma_{\boldsymbol{x}} A^T$$

which is valid whenever $\boldsymbol{y}$ is a linear function of $\boldsymbol{x}$ (i.e. $\boldsymbol{y} = A\boldsymbol{x}$)

For a larger system, the covariance matrix looks like this

$$\Sigma_{\boldsymbol{x}} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1,x_2} & \cdots & \sigma_{x_1,x_n} \\ \sigma_{x_2,x_1} & \sigma_{x_2}^2 & \cdots & \sigma_{x_2,x_n} \\ \vdots & \vdots & & \vdots \\ \sigma_{x_n,x_1} & \sigma_{x_n,x_2} & \cdots & \sigma_{x_n}^2 \end{bmatrix}$$

The formal definition of a covariance matrix is equivalent to the definition of covariance extended to vectors,

$$\Sigma = E\left[(\boldsymbol{x} - E[\boldsymbol{x}])(\boldsymbol{x} - E[\boldsymbol{x}])^T\right]$$

Covariance matrices have a number of special properties: square, symmetric, positive semidefinite

The Multivariate Normal Distribution

- We are all now familiar with the definition of a 1-D Normal p.d.f.

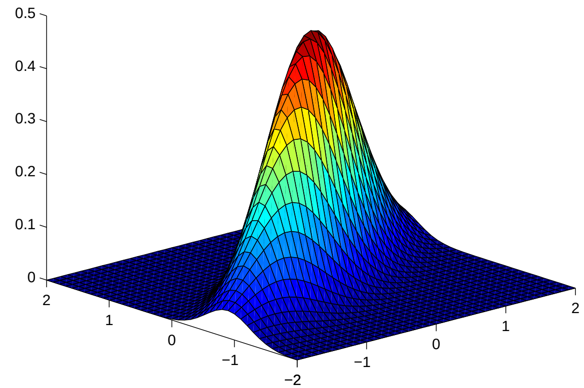
$$p(x) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}$$

- If x is a n-dimensional vector then the multivariate Normal p.d.f. is

$$p(x) = \det\left(2\pi\Sigma\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

- The mean μ is now a vector of size n
- The variance σ^2 is now the covariance matrix Σ, which is a square matrix of size $n \times n$

The following is a 2-D normal distribution with $\mu = [0, 0]^T$ and $\Sigma = \begin{bmatrix} 0.9 & 0.4 \\ 0.4 & 0.3 \end{bmatrix}$



Consider again the definition of the multivariate Normal,

$$p(x) = \det\left(2\pi\Sigma\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

For some values of x we get the same probability; In particular, when the expression $(x-\mu)^T\Sigma^{-1}(x-\mu)$ is constant we get a curve of constant probability; What curve? Define a constant c^2 as follows (shown in 2-D but works in higher dimensions as well),

$$[x_1, y_1]\begin{bmatrix} a & b \\ b & d \end{bmatrix}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = c^2$$

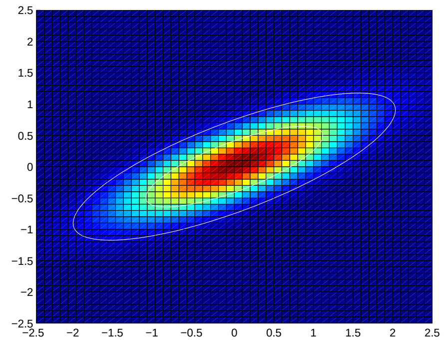
where $[x_1, y_1]^T = (x-\mu)$ and the matrix in the middle is Σ^{-1}. This can be expanded to obtain,

$$ax_1^2 + 2bx_1y_1 + dy_1^2 = c^2$$

which is the equation of an ellipse.

It can be shown that the probability of a value falling within the ellipse is $1-\alpha$, where α comes from $\chi^2(\alpha) = c^2$ and $\chi^2(\alpha)$ is the upper (100α)th percentile of a χ^2 distribution.

An ellipse with $\alpha = 0.1$ is known as a 90% *confidence ellipse*,



The smaller ellipse is the 50% confidence ellipse, while the larger is the 90% confidence ellipse