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,

\[
\sigma_{ij} = \text{Cov}(X_i, X_j) = E((X_i - \mu_{X_i})(X_j - \mu_{X_j})) = E(X_iX_j) - \mu_{X_i}\mu_{X_j}
\]

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,

\[
\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i\sigma_j}
\]
Variance can be considered a special case of covariance,

\[ \text{Var}(X) = \text{Cov}(X, X) \]

If two random variables are independent,

\[ \text{Cov}(X_1, X_2) = 0 \]

Enough about covariance for now... What about covariance matrices? They come up more naturally if we look at functions of random variables...
If a single random variable passes through a linear function, we can determine the expected value and variance of the output.

Let \( y = f(x) = a \cdot x + b \)

\[
\begin{align*}
\mu_y &= E[a \cdot X + b] = a \cdot \mu_x + b \\
\sigma^2_y &= E[(a \cdot X + b - \mu_y)^2] \\
&= a^2 \cdot \sigma^2_x
\end{align*}
\]

What if the function involves some linear combination of random variables?

E.g. Let \( y = f(x_1, x_2) = a \cdot x_1 + b \cdot x_2 \)

\[
\begin{align*}
\mu_y &= a \cdot \mu_{x_1} + b \cdot \mu_{x_2} \\
\sigma^2_y &= a^2 \cdot \sigma^2_{x_1} + b^2 \cdot \sigma^2_{x_2} + 2ab \cdot \sigma_{x_1x_2}
\end{align*}
\]

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,

\[
\begin{align*}
y_1 &= ax_1 + bx_2 \\
y_2 &= cx_1 + dx_2
\end{align*}
\]

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

\[
y = \begin{bmatrix} a & b \\ c & d \end{bmatrix} x = Ax
\]

It can easily be shown that,

\[
\mu_y = A\mu_x
\]
To talk about the uncertainty of the output, we must discuss uncertainty in the input—which is given by $\sigma^2_{x_1}$, $\sigma^2_{x_2}$, and $\sigma_{x_1x_2}$. First we organize these values into a covariance matrix,

$$
\Sigma_x = \begin{bmatrix}
\sigma^2_{x_1} & \sigma_{x_1x_2} \\
\sigma_{x_1x_2} & \sigma^2_{x_2}
\end{bmatrix}
$$

(Note that the symbol $\Sigma$ 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 $\Sigma_y$ for our system; It turns out that this is given by,

$$
\Sigma_y = A\Sigma_x A^T
$$

which is valid whenever $y$ is a linear function of $x$ (i.e. $y = Ax$).
For a larger system, the covariance matrix looks like this

$$\Sigma_x = \begin{bmatrix}
\sigma_{x_1}^2 & \sigma_{x_1,x_2} & \cdots & \sigma_{x_1,x_n} \\
\sigma_{x_2,x_1} & \sigma_{x_2}^2 & \cdots & \sigma_{x_2,x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{x_n,x_1} & \sigma_{x_n,x_2} & \cdots & \sigma_{x_n}^2
\end{bmatrix}$$

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

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

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.

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

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

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

- The mean \( \mu \) is now a vector of size \( n \)
- The variance \( \sigma^2 \) is now the covariance matrix \( \Sigma \), 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,

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

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),

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

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

\[
a x_1^2 + 2bx_1y_1 + dy_1^2 = c^2
\]

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 $\alpha$ comes from $\chi^2(\alpha) = c^2$ and $\chi^2(\alpha)$ is the upper $(100\alpha)$th percentile of a $\chi^2$ 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.