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,

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

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,

\[
\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}
\]

**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 \( y = f(x) = a \cdot x + b \)
  \[
  \mu_y = E[a \cdot X + b] = a \cdot \mu_x + b \\
  \sigma_y^2 = E[(a \cdot X + b - \mu_y)^2] \\
  = a^2 \cdot \sigma_x^2
  \]

- What if the function involves some linear combination of random variables?
  e.g. Let \( y = f(x_1, x_2) = a \cdot x_1 + b \cdot x_2 \)
  \[
  \mu_y = a \cdot \mu_{x_1} + b \cdot \mu_{x_2} \\
  \sigma_y^2 = a^2 \cdot \sigma_{x_1}^2 + b^2 \cdot \sigma_{x_2}^2 + 2ab \cdot \sigma_{x_1 x_2}
  \]

- 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,

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

It can easily be shown that,

\[
\mu_y = A \mu_x
\]

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,

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

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

- 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 \( n \times n \)

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,
\[
ax_1^2 + 2bx_1y_1 + dy_1^2 = c^2
\]

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 \( \alpha \) 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.