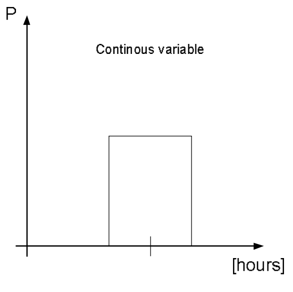
Basic statistics

- Needed for:
  - Kinematic models, error predictions
  - Sensors for mobile robots (precision and accuracy)
  - Different methods for sensor fusion:
    - Bayesian methods (general case),
    - Kalman filters (special case of Bayesian methods),
    - and others.
  - Research papers

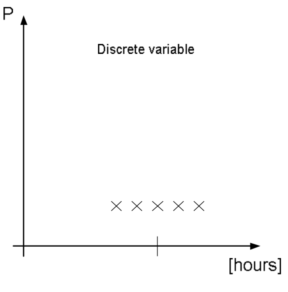
Basic statistics – Statistical representation – Stochastic variable

Battery lasting time, X = 5hours ±1hour

X can have many different values



Continous – The variable can have any value within the bounds



Discrete – The variable can have specific (discrete) values

Basic statistics – Statistical representation – Stochastic variable

Another way of describing the stochastic variable, i.e. by another form of bounds

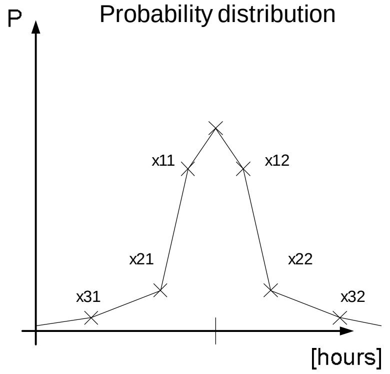
In 68%: x11 < X < x12

In 95%: x21 < X < x22

In 99%: x31 < X < x32

In 100%: -∞ < X < ∞

The value to expect is the mean value => Expected value

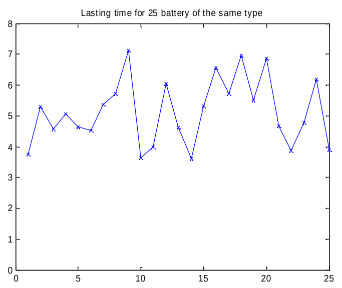


How much X varies from its expected value => Variance

Expected value and Variance

$$E[X] = \int_{-\infty}^{\infty} x.f_X(x)dx$$

$$E[X] = \sum_{k=-\infty}^{\infty} k.p_X(k)$$

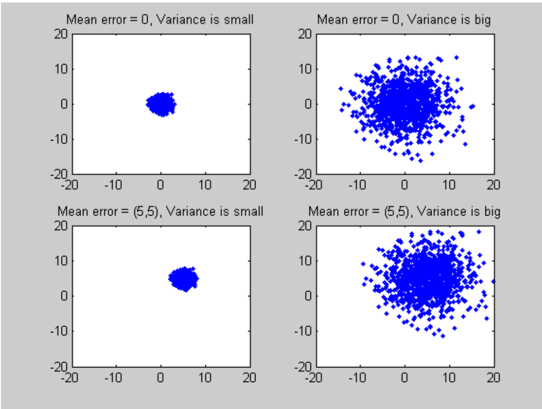


$$V[X] = \sigma_X^2 = \int_{-\infty}^{\infty} (x - E[X])^2 . f_X(x)dx$$

$$V[X] = \sigma_X^2 = \sum_{K=-\infty}^{\infty} (k - E[X])^2 . p_X(k)$$

The standard deviation σ_X is the square root of the variance

Important termonology



Mean error?

Bias?

Variance?

Symmetric?

What causes the errors?

Error =

$$e_i = \hat{y}_i - y_i$$

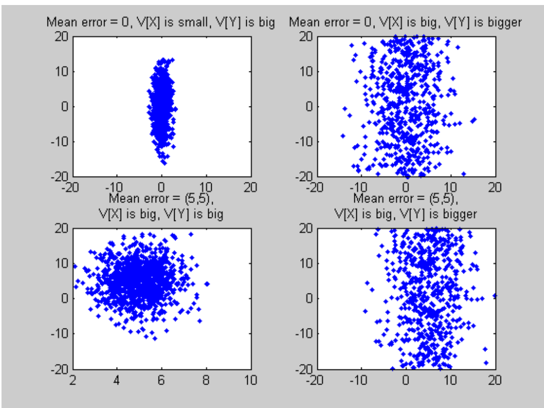
Estimated value – True value

What does it mean if we know that the error has **zero mean / not zero mean**?

What does it mean if we know that the error has **small / big variance**?

Two inpedendent measurment systems, both have errors with zero mean but they have different variances. **What would you do**?

Important termonology



Mean error?

Bias?

Variance?

Symmetric?

How do these plots differ compared to the plots in the previous slide?

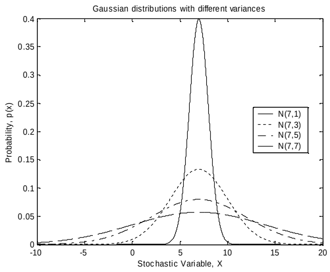
What does this difference means to our measurment system?

How come the mean errors are the same as in the previous examples?

Gaussian (Normal) distribution

By far the mostly used probability distribution because of its nice statistical and mathematical properties

$$X \sim N(m_X, \sigma_X)$$



$$p_X(x) = \frac{1}{\sigma_X . \sqrt{2\pi}} e^{-\frac{(x-E[X])^2}{2\sigma_X^2}}$$

What does it means if a specification tells that a sensor measures a distance [mm] and has an error that is normally distributed with zero mean and σ = 100mm?

Normal distribution:

~68.3% $[m_X - \sigma_X, m_X + \sigma_X]$

~95% $[m_X - 2\sigma_X, m_X + 2\sigma_X]$

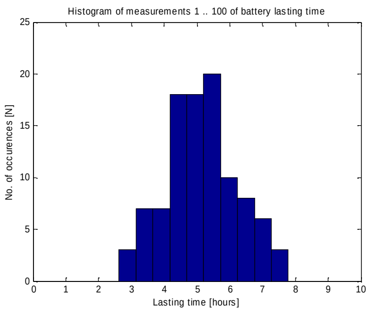
~99% $[m_X - 3\sigma_X, m_X + 3\sigma_X]$

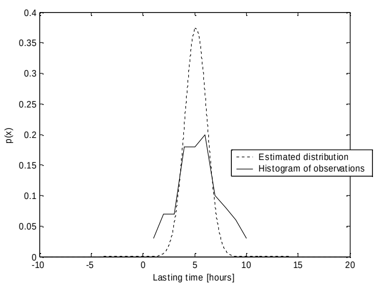
etc.

Estimate of the expected value and the variance from observations

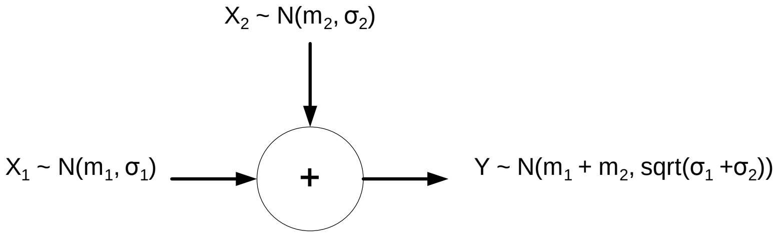
$$\hat{m}_X = \tfrac{1}{N} \sum_{k=1}^{N} X(k)$$

$$\hat{\sigma}_X^2 = \tfrac{1}{N-1} \sum_{k=1}^{N} (X(k) - \hat{m}_X)^2$$





Linear combinations (1)



$$E[aX+b]=aE[X]+b \qquad V[aX+b]=a^2V[X]$$

$$E[X_1+X_2]=E[X_1]+E[X_2] \qquad V[X_1+X_2]=V[X_1]+V[X_2]$$

This property that Y remains Gaussian if the s.v. are combined linearly is one of the great properties of the Gaussian distribution!

Linear combinations (2)

We measure a distance by a device that have normally distributed errors,

$$\hat{D} \sim N(D, \sigma_D)$$

Do we win something of making a lot of measurements and use the average value instead?

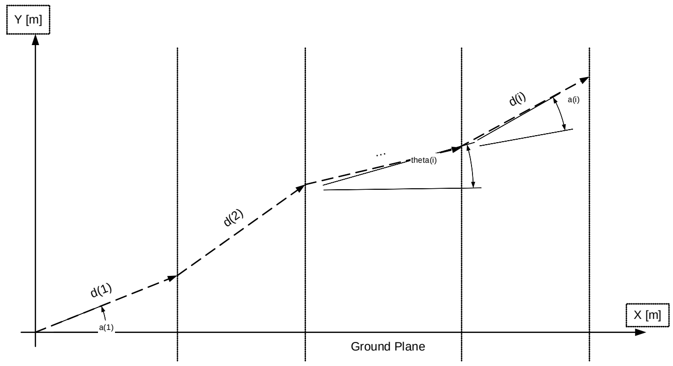
$$Y = \tfrac{1}{5}D + \tfrac{1}{5}D + \tfrac{1}{5}D + \tfrac{1}{5}D + \tfrac{1}{5}D = \tfrac{1}{N}\sum_{i=1}^{N=5} D_i$$

What will the expected value of Y be?

What will the variance (and standard deviation) of Y be?

If you are using a sensor that gives a large error, how would you best use it?

Linear combinations (3)



$$\hat{d}_i = d_i + \varepsilon_d$$

d_i is the mean value and $\varepsilon_d \sim N(0, \sigma_d)$

$$\hat{\alpha}_i = \alpha_i + \varepsilon_\alpha$$

α_i is the mean value and $\varepsilon_\alpha \sim N(0, \sigma_\alpha)$

With ε_d and ε_α un-correlated => $V[\varepsilon_d, \varepsilon_\alpha] = 0$ (Actually the co-variance, which is defined later)

Linear combinations (4)

D = {The total distance} is calculated as before as this is only the sum of all d's

$$D = d_1 + d_2 + \ldots + d_N = \sum_{k=1}^{N} d(k)$$

The expected value and the variance become:

$$E\left[\hat{D}\right] = E\left[\sum_{k=1}^{N}\{d(k) + \varepsilon_d\}\right] = \sum_{k=1}^{N} E\left[d(k)\right] + \sum_{k=1}^{N} E\left[\varepsilon_d\right] = \sum_{k=1}^{N} E\left[d(k)\right] + \sum_{i=1}^{N} 0 = \sum_{i=1}^{N} d(k)$$

$$V\left[\hat{D}\right] = V\left[\sum_{k=1}^{N}\{d(k) + \varepsilon_d\}\right] = \sum_{k=1}^{N} V\left[d(k)\right] + \sum_{k=1}^{N} V\left[\varepsilon_d\right] = \sum_{k=1}^{N} 0 + \sum_{k=1}^{N} V\left[\varepsilon_d\right] = N\sigma_d^2$$

Linear combinations (5)

θ = {The heading angle} is calculated as before as this is only the sum of all α's, i.e. as the sum of all changes in heading

$$\theta = \theta_0 + \alpha(0) + \alpha(1) + ... + \alpha(N-1) = \theta_0 + \sum_{k=0}^{N-1} \alpha(k)$$

The expected value and the variance become:

$$E\left[\hat{\theta}\right] = E\left[(\theta(0) + \varepsilon_\theta(0)) + \sum_{k=0}^{N-1} \{\alpha(k) + \varepsilon_\alpha(k)\} \right] = \theta(0) + \sum_{k=0}^{N-1} \alpha(k)$$

$$V\left[\hat{\theta}\right] = V\left[(\theta(0) + \varepsilon_\theta(0)) + \sum_{k=0}^{N-1} \{\alpha(k) + \varepsilon_\alpha(k)\} \right] = V\left[\varepsilon_\theta(0)\right] + \sum_{k=0}^{N-1} V\left[\varepsilon_\alpha(k)\right]$$

What if we want to predict X (distance parallel to the ground) (or Y (hight above ground)) from our measured d's and α's?

Non-linear combinations (1)

X(N) is the previous value of X plus the latest movement (in the X direction)

$$X_N = X_{N-1} + d_{N-1}\cos(\theta_{N-1} + \alpha_{N-1})$$

The estimate of X(N) becomes:

$$\hat{X}_N = \hat{X}_{N-1} + \hat{d}_{N-1}\cos(\hat{\theta}_{N-1} + \hat{\alpha}_{N-1}) = (X_{N-1} + \varepsilon_{X_{N-1}}) + (d_{N-1} + \varepsilon_d)\cos(\theta_{N-1} + \varepsilon_{\theta_{N-1}} + \alpha_{N-1} + \varepsilon_\alpha)$$

This equation is non-linear as it contains the term:

$$\cos(\theta_{N-1} + \varepsilon_{\theta_{N-1}} + \alpha_i + \varepsilon_\alpha)$$

and for X(N) to become Gaussian distributed, this equation must be replaced with a linear approximation around $\theta_{N-1} + \alpha_{N-1}$. To do this we can use the Taylor expansion of the first order. By this approximation we also assume that the error is rather small!

With perfectly known θ_{N-1} and α_{N-1} the equation would have been linear!

Non-linear combinations (2)

Use a first order Taylor expansion and linearize X(N) around $\theta_{N-1}+\alpha_{N-1}$.

$$\hat{X}_N = (X_{N-1}+\varepsilon_{X_{N-1}}) + (d_{N-1}+\varepsilon_d)\cos(\theta_{N-1}+\varepsilon_{\theta_{N-1}}+\alpha_{N-1}+\varepsilon_\alpha) \approx$$

$$\approx (X_{N-1}+\varepsilon_{X_{N-1}}) + (d_{N-1}+\varepsilon_d)\Big(\cos(\theta_{N-1}+\alpha_{N-1}) - (\varepsilon_{\theta_{N-1}}+\varepsilon_\alpha)\sin(\theta_{N-1}+\alpha_{N-1})\Big)$$

This equation is linear **as all error terms are multiplied by constants** and we can calculate the expected value and the variance as we did before.

$$E\big[\hat{X}_N\big] \approx E\Big[(X_{N-1}+\varepsilon_{X_{N-1}}) + (d_{N-1}+\varepsilon_d)\Big(\cos(\theta_{N-1}+\alpha_{N-1}) - (\varepsilon_{\theta_{N-1}}+\varepsilon_\alpha)\sin(\theta_{N-1}+\alpha_{N-1})\Big)\Big] =$$

$$= E[X_{N-1}] + E[d_{N-1}\cos(\theta_{N-1}+\alpha_{N-1})] = X_{N-1} + d_{N-1}\cos(\theta_{N-1}+\alpha_{N-1})$$

Non-linear combinations (3)

The variance becomes (calculated exactly as before):

$$V\big[\hat{X}_N\big] \approx V\Big[(X_{N-1}+\varepsilon_{X_{N-1}}) + (d_{N-1}+\varepsilon_d)\Big(\cos(\phi) - (\varepsilon_{\theta_{N-1}}+\varepsilon_\alpha)\sin(\phi)\Big)\Big] =$$

$$= V[X_{N-1}] + V\big[\varepsilon_{X_{N-1}}\big] + V\Big[(d_{N-1}+\varepsilon_d)\Big(\cos(\phi) - (\varepsilon_{\theta_{N-1}}+\varepsilon_\alpha)\sin(\phi)\Big)\Big] =$$

$$= \sigma^2_{X_{N-1}} + \big(-d_{N-1}\sin(\phi)\big)^2\sigma^2_\alpha + \big(\cos(\phi)\big)^2\sigma^2_d + \big(-d_{N-1}\sin(\phi)\big)^2\sigma^2_{\theta_{N-1}}$$

Two really important things should be noticed, first the linearization only affects the calculation of the variance and second (which is even more important) is that the above equation is ***the partial derivatives of***:

$$\hat{X}_N = \hat{X}_{N-1} + \hat{d}_{N-1}\cos(\hat{\theta}_{N-1}+\hat{\alpha}_{N-1})$$

with respect to our uncertain parameters squared multiplied with their variance!

Non-linear combinations (4)

This result is very good => an easy way of calculating the variance => the ***law of error propagation***

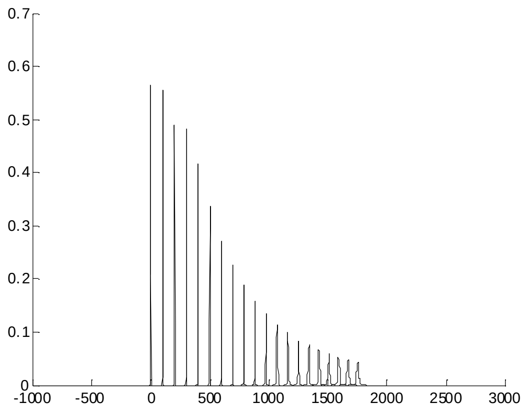
$$V\left[\hat{X}_N\right] = \left(\frac{\partial \hat{X}_N}{\partial X_{N-1}}\right)^2 \sigma^2_{X_{N-1}} + \left(\frac{\partial \hat{X}_N}{\partial d_{N-1}}\right)^2 \sigma^2_d + \left(\frac{\partial \hat{X}_N}{\partial \alpha_{N-1}}\right)^2 \sigma^2_\alpha + \left(\frac{\partial \hat{X}_N}{\partial \theta_{N-1}}\right)^2 \sigma^2_{\theta_{N-1}}$$

The partial derivatives of $\hat{X}_N = \hat{X}_{N-1} + \hat{d}_{N-1}\cos(\hat{\theta}_{N-1} + \hat{\alpha}_{N-1})$ become:

$$\frac{\partial \hat{X}_N}{\partial X_{N-1}} = 1 \qquad \frac{\partial \hat{X}_N}{\partial d_{N-1}} = \cos(\phi) \qquad \frac{\partial \hat{X}_N}{\partial \alpha_{N-1}} = -\hat{d}_{N-1}\sin(\phi) \qquad \frac{\partial \hat{X}_N}{\partial \theta_{N-1}} = -\hat{d}_{N-1}\sin(\phi)$$

Non-linear combinations (5)

The plot shows the variance of X for the time step 1, …, 20 and as can be noticed the variance (or standard deviation) is constantly increasing.



$\sigma^2_{X_0} = 0.5$

$\sigma_d = 1/10$

$\sigma_\alpha = 5/360$

Multidimensional Gaussian distributions (1)

The Gaussian distribution can easily be extended for several dimensions by: replacing the variance (σ) by a co-variance matrix (Σ) and the scalars (x and m_X) by column vectors.

$$p_X(x) = \left(\frac{1}{(2\pi)^N \left| \Sigma_X \right|} \right)^{\frac{1}{2}} e^{-\frac{1}{2}(x-m_X)^T \Sigma_X^{-1} (x-m_X)}$$

The CVM describes (consists of):

1) the variances of the individual dimensions => diagonal elements

2) the co-variances between the different dimensions => off-diagonal elements

$$\Sigma_X = \begin{pmatrix} \sigma_{x1}^2 & C(x1,x2) & C(x1,x3) \\ C(x2,x1) & \sigma_{x2}^2 & C(x2,x3) \\ C(x3,x1) & C(x3,x2) & \sigma_{x3}^2 \end{pmatrix}$$

! Symmetric

! Positive definite

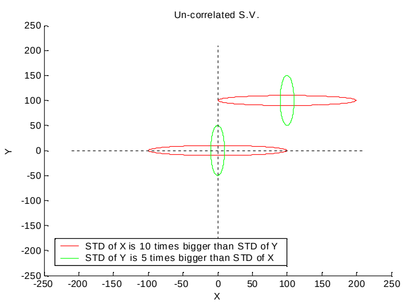
MGD (2)

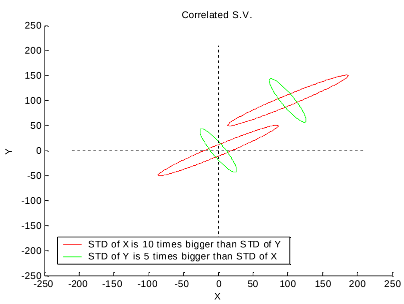
$$\Sigma(green) = \begin{pmatrix} 100 & 0 \\ 0 & 2500 \end{pmatrix}$$

$$\Sigma(red) = \begin{pmatrix} 10000 & 0 \\ 0 & 100 \end{pmatrix}$$

$$\Sigma(green) = \begin{pmatrix} 700 & -1039 \\ -1039 & 1900 \end{pmatrix}$$

$$\Sigma(red) = \begin{pmatrix} 7525 & 4287 \\ 4287 & 2575 \end{pmatrix}$$





Eigenvalues => standard deviations

Eigenvectors => rotation of the ellipses

Means?

Variances?

MGD (3)

The co-variance between two stochastic variables is calculated as:

$$C(X,Y) = E[(X - m_X)(Y - m_Y)]$$

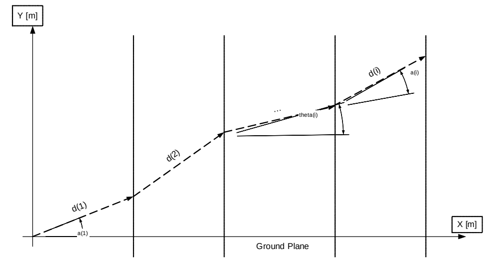
Which for a discrete variable becomes:

$$C(X,Y) = \sum_j \sum_k (j - m_X)(k - m_Y)\, p_{X,Y}(j,k) \qquad V[X] = \sigma_X^2 = \sum_{K=-\infty}^{\infty} (k - E[X])^2 . p_X(k)$$

And for a continuous variable becomes:

$$C(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - m_X)(y - m_Y)\, f_{X,Y}(x,y)\,dxdy$$

MGD (4) - Non-linear combinations



The state variables (x, y, θ) at time k+1 become:

$$Z(k+1) = \begin{pmatrix} x(k+1) \\ y(k+1) \\ \theta(k+1) \end{pmatrix} = f(Z(k), U(k)) = \begin{pmatrix} x(k) + d(k)\cos(\theta(k) + \alpha(k)) \\ y(k) + d(k)\sin(\theta(k) + \alpha(k)) \\ \theta(k) + \alpha(k) \end{pmatrix}$$

MGD (5) - Non-linear combinations

$$Z(k+1) = \begin{pmatrix} x(k+1) \\ y(k+1) \\ \theta(k+1) \end{pmatrix} = f(Z(k),U(k)) = \begin{pmatrix} x(k) + d(k)\cos(\theta(k)+\alpha(k)) \\ y(k) + d(k)\sin(\theta(k)+\alpha(k)) \\ \theta(k)+\alpha(k) \end{pmatrix}$$

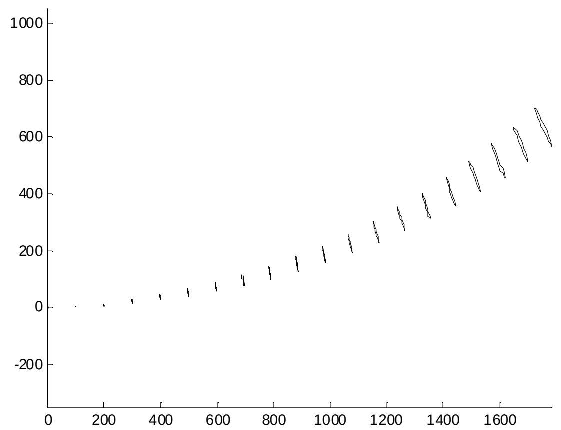
We know that to calculate the variance (or co-variance) at time step k+1 we must linearize Z(k+1) by e.g. a Taylor expansion - but we also know that this is done by the ***law of error propagation***, which for matrices becomes:

$$\Sigma(k+1|k) = \nabla f_X \Sigma(k|k) \nabla f_X^T + \nabla f_U U(k+1) \nabla f_U^T$$

With ∇f_X and ∇f_U are the Jacobian matrices (w.r.t. our uncertain variables) of the state transition matrix.

$$\nabla f_X = \begin{pmatrix} 1 & 0 & -d(k)\sin(\theta(k)+\alpha(k)) \\ 0 & 1 & d(k)\cos(\theta(k)+\alpha(k)) \\ 0 & 0 & 1 \end{pmatrix} \qquad \nabla f_U = \begin{pmatrix} \cos(\theta(k)+\alpha(k)) & -d(k)\sin(\theta(k)+\alpha(k)) \\ \sin(\theta(k)+\alpha(k)) & d(k)\cos(\theta(k)+\alpha(k)) \\ 0 & 1 \end{pmatrix}$$

MGD (6) - Non-linear combinations



The uncertainty ellipses for X and Y (for time step 1 .. 20) are shown in the figure.

The Error Propagation Law

- One-dimensional case of a nonlinear error propagation problem
- It can be shown, that the output covariance matrix C_Y is given by the error propagation law:
- where $C_Y = F_X C_X F_X^T$
  - C_X: covariance matrix representing the input uncertainties
  - C_Y: covariance matrix representing the propagated uncertainties for the outputs.
  - F_X: is the **Jacobian** matrix defined as:

$$F_X = \nabla f = \left[\nabla_X \cdot f(X)^T\right]^T = \begin{bmatrix} f_1 \\ : \\ f_m \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial X_1} & \cdots & \frac{\partial}{\partial X_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial X_1} & \cdots & \frac{\partial f_1}{\partial X_n} \\ : & \cdots & : \\ \frac{\partial f_m}{\partial X_1} & \cdots & \frac{\partial f_m}{\partial X_n} \end{bmatrix}$$

  - which is the transposed of the gradient of f(X).