COMP0130: Robotic Vision and Navigation

Lecture 04: Covariance Matrices (Optional / Supplemental)

Simon Julier

Structure

- Motivation
- Scalar Case
- Multi-dimensional Case Interaction correlations
- Covariance Ellipses and Confidence Bounds
- Limitations of Covariances

The Need for Uncertainty

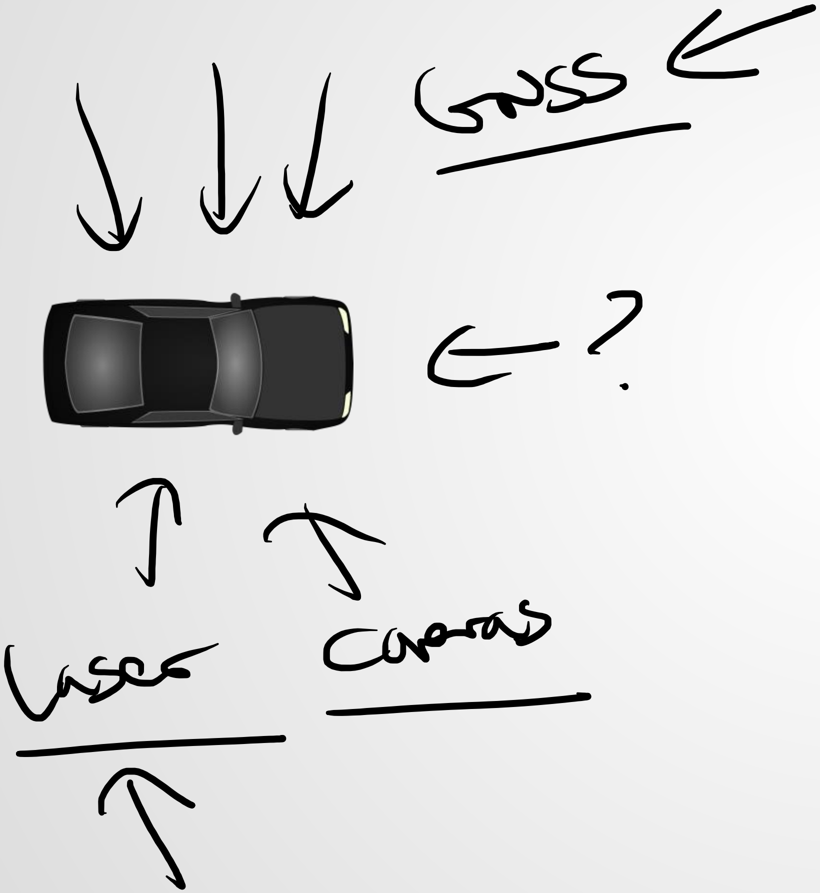
- Estimators often return a point-estimate of a quantity of interest

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

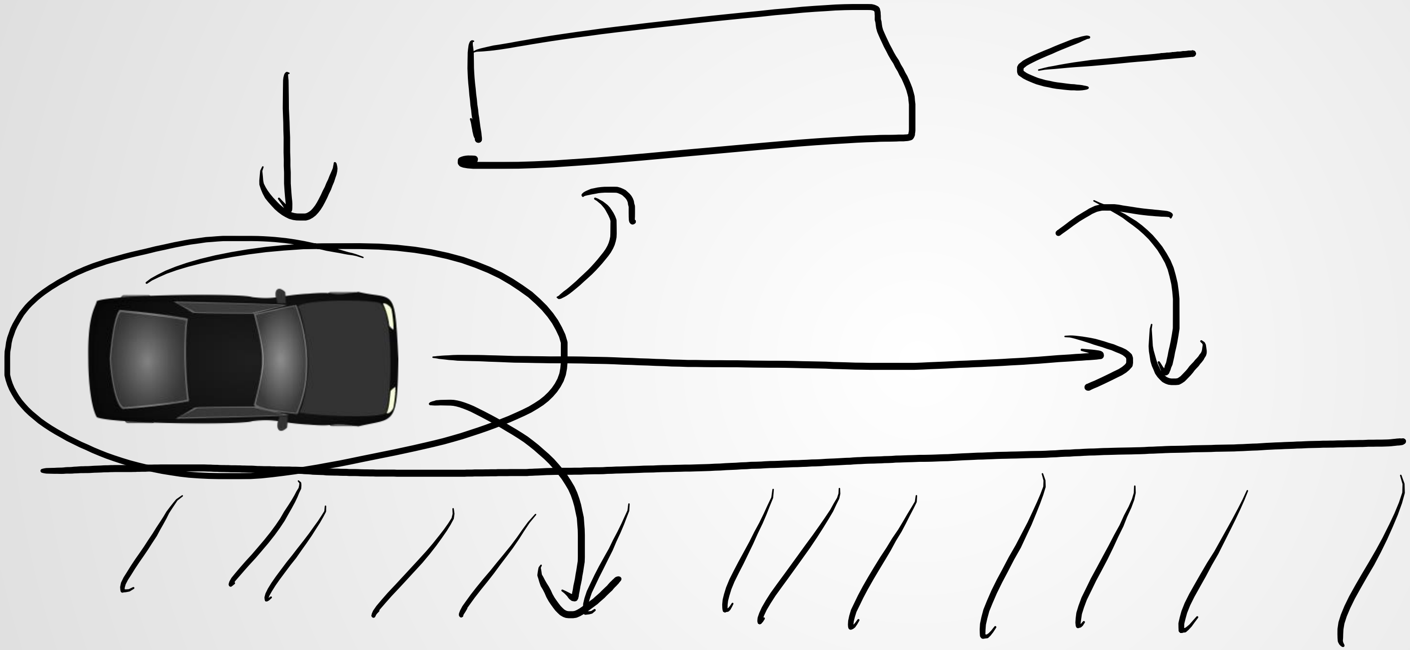
$\hat{x}$

- However, we often need to know how accurate this estimate is

The Need for Uncertainty



The Need for Uncertainty



Quantifying Uncertainty

- The way we quantify uncertainty is to accompany every estimated state with *a covariance matrix*

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \mathbf{P} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$

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

- But what does it actually mean and why do we use it?

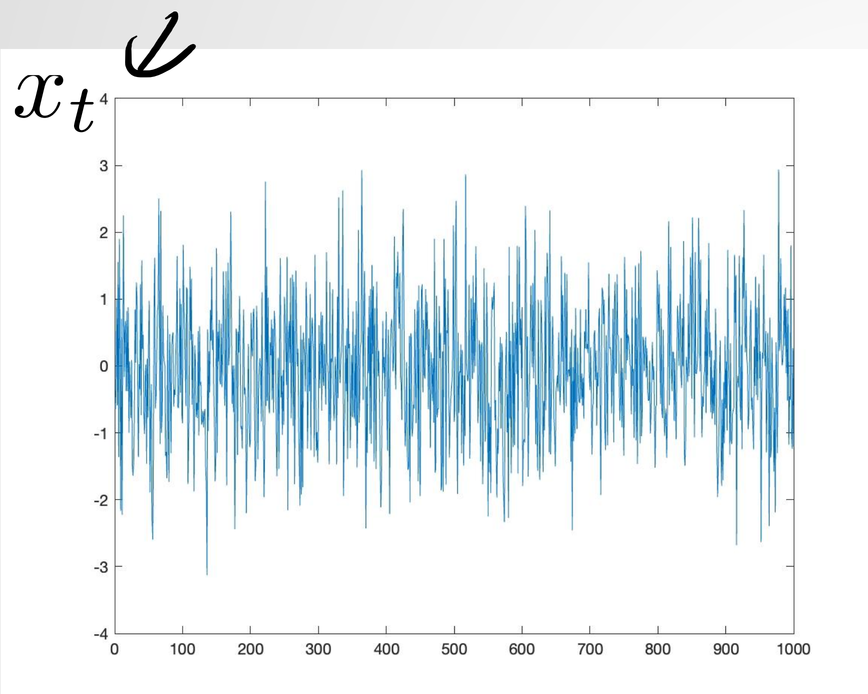
Quantifying Uncertainty in 1D

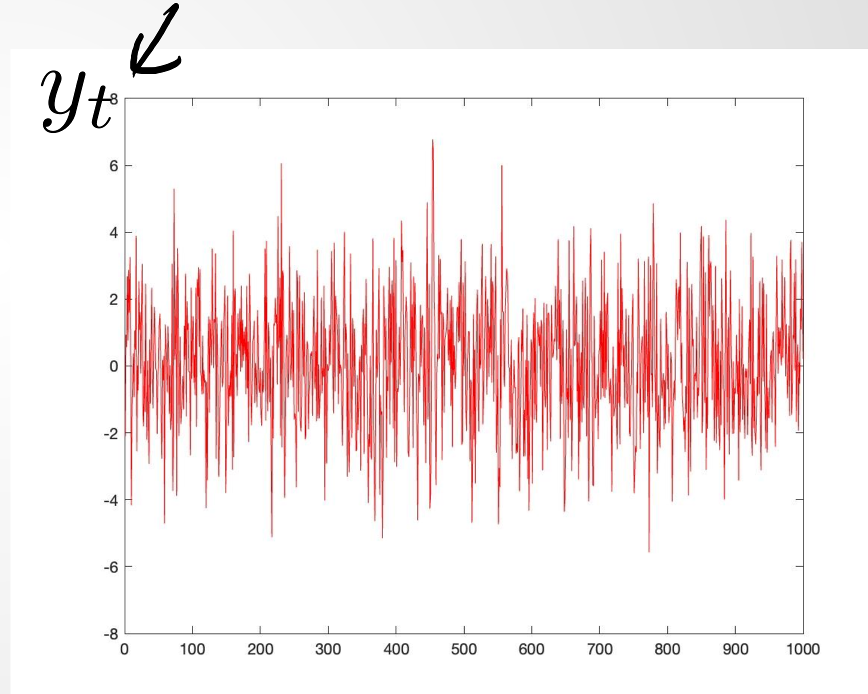
- Consider the noisy time-dependent series

$$x_t, y_t$$

- Which one is more uncertain?

Quantifying Uncertainty in 1D



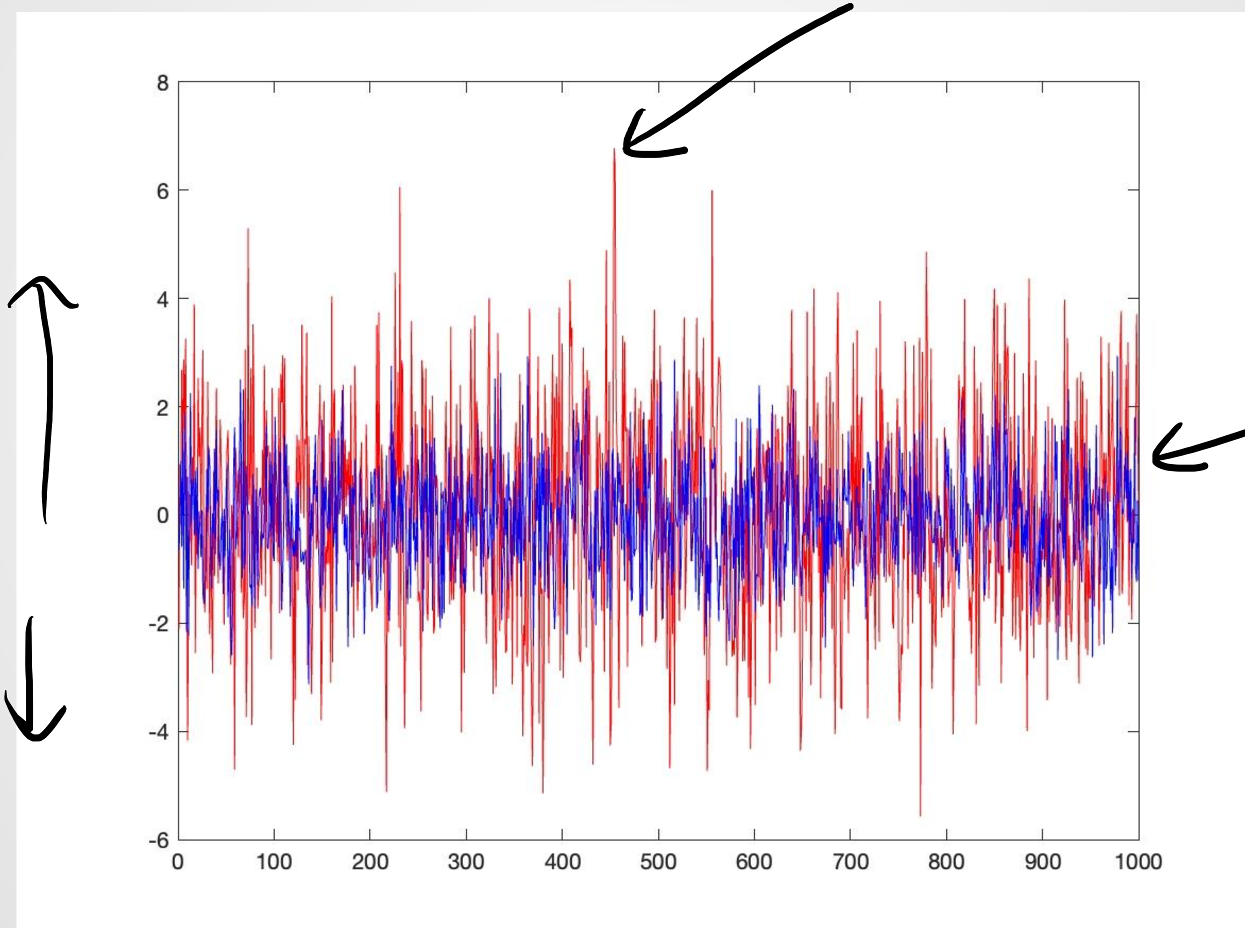


Quantifying Uncertainty in 1D

y_t more uncertain

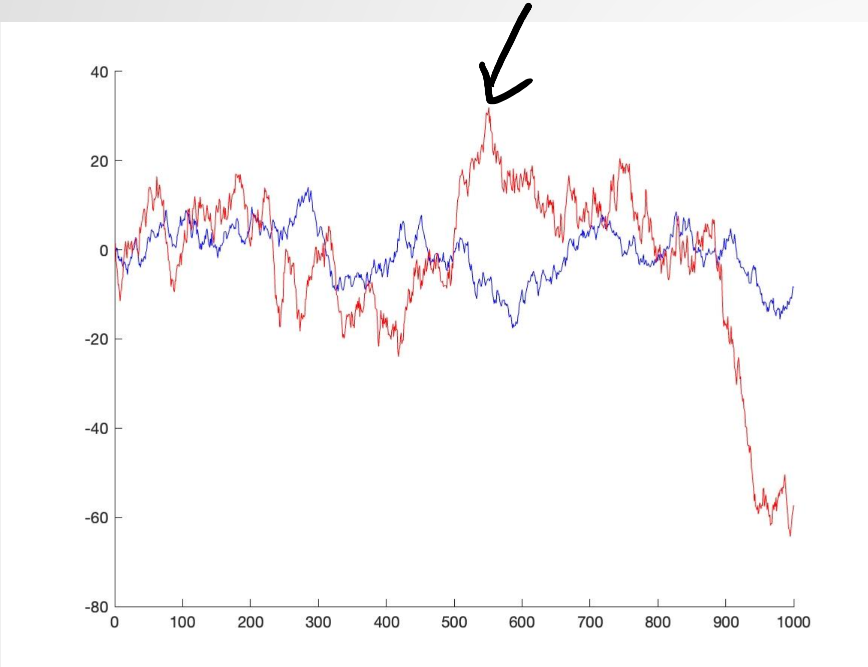
x_t

T

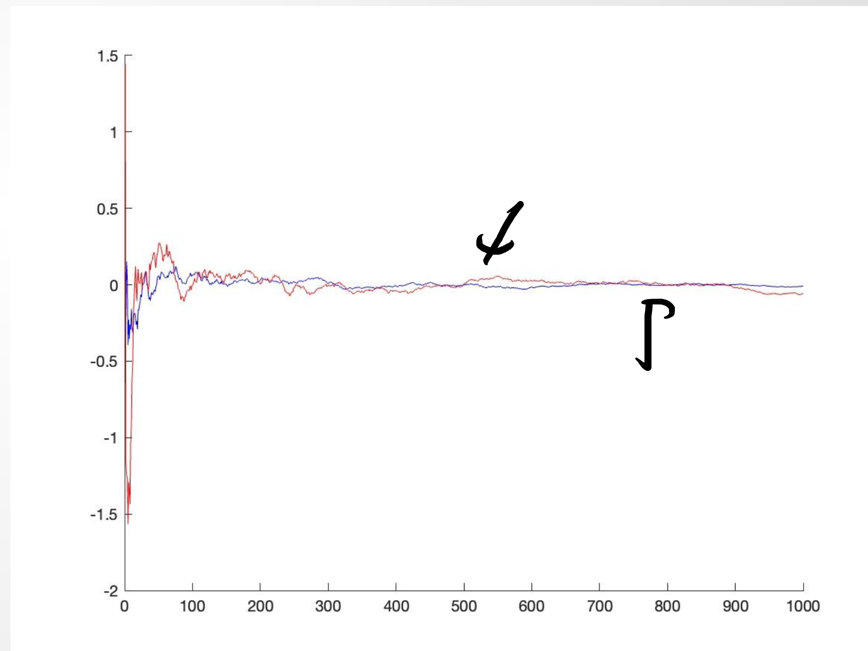


Quantifying Uncertainty in 1D

$\sum_{i}^{t} y_t$

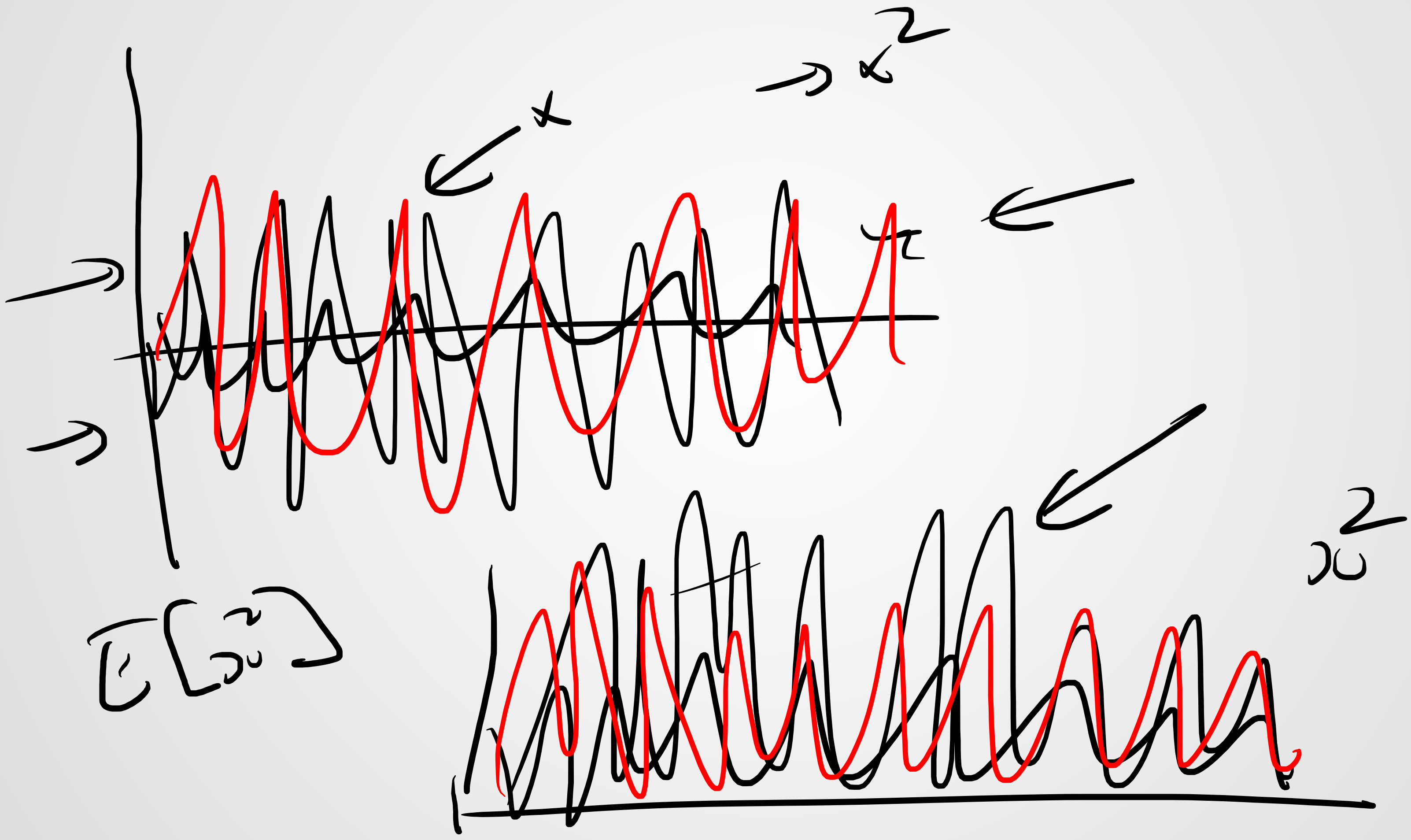


Cumulative sum

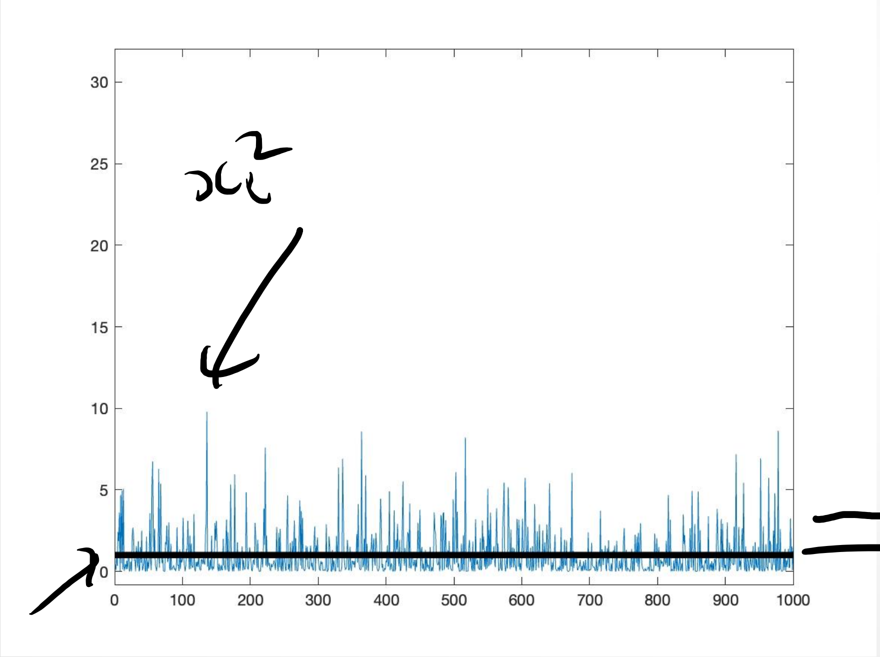


Cumulative average

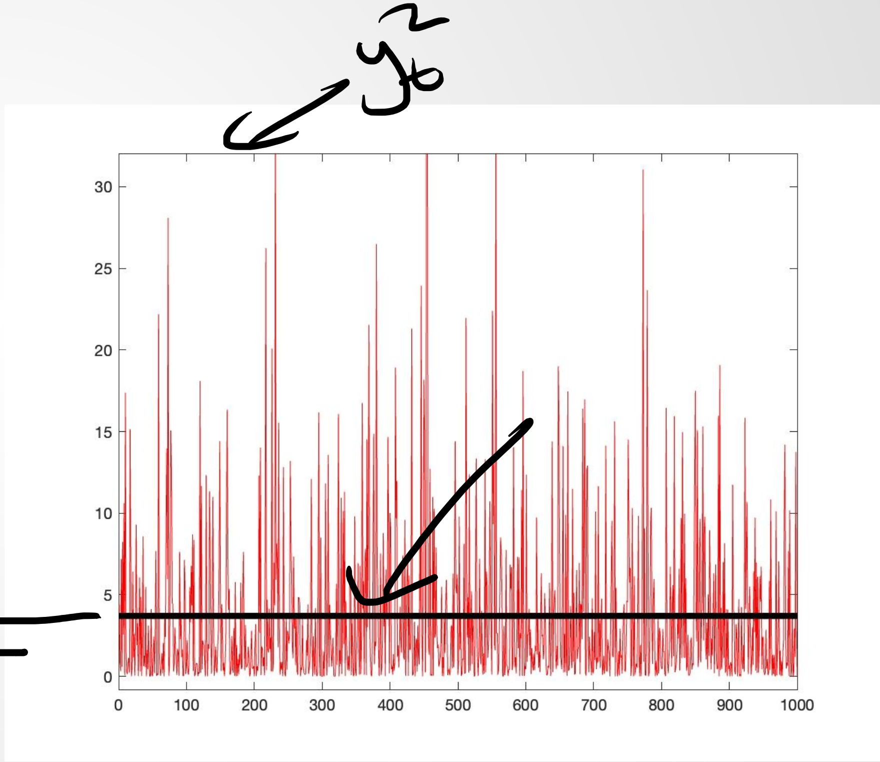
Squaring the Error



Quantifying Uncertainty in 1D



=



Covariance vs the Standard Deviation

- Formally, the covariance is given by

$$\mathrm{cov}\left(x_t\right) = \mathbb{E}\left[\left(x_t - \mathbb{E}\left[x_t\right]\right)^2\right]$$

- The standard deviation is the square root of this,

$$\mathrm{std}\left(x_t\right) = \sqrt{\mathrm{cov}\left(x_t\right)} = \sqrt{\mathbb{E}\left[\left(x_t - \mathbb{E}\left[x_t\right]\right)^2\right]}$$

Units and Standard Deviation and Covariance

- Standard deviation and covariance both have units
- Therefore, any operation with them has to preserve unit consistency

$x_t \to$ pos m

$\mathrm{cov}(x_i)$ m^2

$\mathrm{std}(x_i)$ m

$m + m^2$

$x_t + \cancel{\mathrm{cov}(x_i)}$

$\underset{m}{x_t} + \underset{m}{\mathrm{std}(x_i)}$ ✓

Other Things about Mean Squared Error

- It relates to actual physical quantities such as energy consumed
- It can be related to maximum entropy estimators
- It can be related to Central Limit Theorem
- It is differentiable $x^2 \rightarrow 2x \leftarrow$
- It does *not* depend upon or assume Gaussians in any way, shape or form

Multi-Dimensional Uncertainty

- So far we've seen the case in 1D but all interesting cases are multi-dimensional
- We can compute the covariances on each state separately

$$\operatorname{cov}(x_t) = P_{xx} = \mathbb{E}\left[(x_t - \mathbb{E}[x_t])^2\right]$$

$$\operatorname{cov}(y_t) = P_{yy} = \mathbb{E}\left[(y_t - \mathbb{E}[y_t])^2\right]$$

Multi-Dimensional Uncertainty

x_t

y_t

$$P_{xx} = \mathbb{E}\left[\left(x_t - \mathbb{E}\left[x_t\right]\right)^2\right]$$

$$P_{yy} = \mathbb{E}\left[\left(y_t - \mathbb{E}\left[y_t\right]\right)^2\right]$$

Interacting States

- However, different states can *interact* with one another
- For example, suppose that

$a_n \Delta$

$$x_{k+1} = x_k + \Delta T y_k + \frac{\Delta T^2}{2} a_k$$

$$y_{k+1} = y_k + \Delta T a_k$$

Example of Interacting States

$$x_{k+1} = x_k + \Delta T y_k + \frac{\Delta T^2}{2} a_k$$

$$y_{k+1} = y_k + \Delta T a_k$$

Cross-Correlation

- The cross correlation is computed from

$$P_{xy} = \mathbb{E}\left[\left(x_t - \mathbb{E}\left[x_t\right]\right)\left(y_t - \mathbb{E}\left[y_t\right]\right)\right]$$

$P_{xyz} \Leftarrow$ Skew

$P_{xx} = \mathbb{E}\left[\left(x_t - \mathbb{E}\left[x_t\right]\right)^2\right]$

What do Cross Correlations Mean?

- We'll consider two cases:
  - When the random variables are independent of one another
  - When the random variables are not independent of one another

Independent Random Variables

- When variables are independent of one another, the joint probability functions become the product of functions,

$$f(x, y) = \boxed{f(x)f(y)}$$

Independence and Cross Correlation

- Therefore, when they are independent,

$$P_{xy} = \mathbb{E}\left[\left(x_t - \mathbb{E}\left[x_t\right]\right)\left(y_t - \mathbb{E}\left[y_t\right]\right)\right]$$
$$= \mathbb{E}\left[\left(x_t - \mathbb{E}\left[x_t\right]\right)\right] \times \mathbb{E}\left[\left(y_t - \mathbb{E}\left[y_t\right]\right)\right]$$

$$0 \times 0$$

$$E\left[\left(x_t - E\left[x_t\right]\right)\right] = 0$$
$$= E\left[x_t\right] - E\left[x_t\right] = 0$$

Interpreting the Cross-Correlation

- There are several ways we can interpret what it is representing for us

- Here we'll just look at it as a linear model which describes the relationship between two random variables

Interpretation 1: Linear Models

$y_t = \alpha x_t + \beta$

- Suppose that we can write one random variable as a linear function of the other,

$$\boxed{y_t} = \alpha \boxed{x_t} + \beta$$

$\beta \sim L(\cdot, -)$

- The covariance and cross correlation are

$$P_{yy} = \alpha^2 P_{xx} + P_{\beta\beta}$$

$$P_{xy} = \alpha P_{xx}$$

Deriving the Quantities

$$\underline{P_{yy}} = \alpha^2 \underline{P_{xx}} + P_{\beta\beta} \quad \checkmark$$

$$P_{yy} = E[(y - E[y])^2] \qquad \alpha^2 E[(x - E[x])^2]$$

$$+ \underline{E[(\beta - E[\beta])^2]}$$

$$y = \alpha x + \beta$$

$$E[y] = \alpha E[x] + E[\beta] \qquad = \alpha^2 P_{xx} + P_{\beta\beta}$$

$$y - E[y]$$

$$= \underline{\alpha (x - E[x])} + \underline{(\beta - E[\beta])}$$

Deriving the Quantities

$$P_{xy} = \alpha P_{xx}$$

$$(x - E[x])(y - E[y])$$

$$= (x - E[x])\left(\alpha(x - E[x]) + \beta - E[\beta]\right)$$

$$= \alpha \underbrace{(x - E[x])^2}_{P_{xx}} + \cancel{(x - E[x])(\beta - E[\beta])}$$

$$= \alpha P_{xx}$$

Example Covariance Matrix

$$\begin{bmatrix} 1 & 0 \\ 0 & 9 \end{bmatrix}$$

$\alpha = 0, \; P_{\beta\beta} = 9$

$$P_{yy} = \alpha^2 P_{xx} + P_{\beta\beta}$$
$$P_{xy} = \alpha P_{xx}$$

Example Covariance Matrix

$$\begin{array}{c} x \\ y \end{array} \begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix}$$

$\alpha = -3$

$y = \alpha x$
$= -3x$

$P_{yy} = 9.1$

$P_{\beta\beta} = 0$

$$P_{yy} = \alpha^2 P_{xx} + P_{\beta\beta}$$
$$P_{xy} = \alpha P_{xx}$$

Example Covariance Matrix

$$\begin{bmatrix} 1 & 3 \\ 3 & 10 \end{bmatrix}$$

$\alpha = 3$

$y = 3x + \beta$

$P_{yy} = 3^2 \cdot 1 + P_{\beta\beta}$
$= 9 + P_{\beta\beta}$

$P_{\beta\beta} = 1$

$$P_{yy} = \alpha^2 P_{xx} + P_{\beta\beta}$$
$$P_{xy} = \alpha P_{xx}$$

Quantifying the Linear Dependency

- We often like to have a sense of how independent the two random variables are from one another
- However, simply looking at the cross correlation directly doesn't tell us what the situation is

$$\begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix} \qquad \begin{bmatrix} 1 & 3 \\ 3 & 10 \end{bmatrix} \leftarrow \times 100 \qquad \begin{bmatrix} 100 & 300 \\ 300 & 1000 \end{bmatrix}$$

$y = 3x$

$y = 3x + \beta$

Simply Comparing Covariance Matrices

Normalized Cross Correlations

- A way to eliminate these scaling effects is to use the normalized cross correlation or *correlation coefficient*

$E[(x-E[x])(y-E[y])]$

$$\frac{c_{xy}}{T} = \frac{P_{xy}}{\sqrt{P_{xx}P_{yy}}}$$

unitless quantity

$x_t + c_{xy}$

Normalized Cross Correlations

- Substituting for the values,

$$c_{xy} = \alpha \frac{P_{xx}}{\sqrt{\alpha^2 P_{xx}^2 + P_{xx} P_{\beta\beta}}}$$

Case when $\alpha = 0$

$y = \alpha x + \beta$

$= \beta$

$C_{xy} = \alpha \frac{P_{xx}}{\sqrt{\quad}}$

$\alpha = 0$

$\boxed{C_{xy} = 0}$

$$c_{xy} = \alpha \frac{P_{xx}}{\sqrt{\alpha^2 P_{xx}^2 + P_{xx} P_{\beta\beta}}}$$

Case when $P_{\beta\beta} = 0$

$\alpha = \sqrt{\alpha^2}$

$\sqrt{\alpha^2} = |\alpha|$

$$c_{xy} = \alpha \frac{P_{xx}}{\sqrt{\alpha^2 P_{xx}^2 + P_{xx} \cdot 0}}$$

$y = 3x \leftarrow$

$y = -3x \leftarrow$

$$= \frac{\alpha \cancel{P_{xx}}}{\sqrt{\alpha^2 \cancel{P_{xx}^2}}} \quad > 1$$

$$= \frac{\alpha}{|\alpha|}$$

$\boxed{c_{xy} = 1, -1}$

$P_{\beta\beta} = 0$

$$c_{xy} = \alpha \frac{P_{xx}}{\sqrt{\alpha^2 P_{xx}^2 + P_{xx} P_{\beta\beta}}}$$

Covariance Ellipses

- We are often interested in providing a compact way to draw covariance information
- A common way to do this is to draw the *covariance ellipse*
- This is the locus of points given by

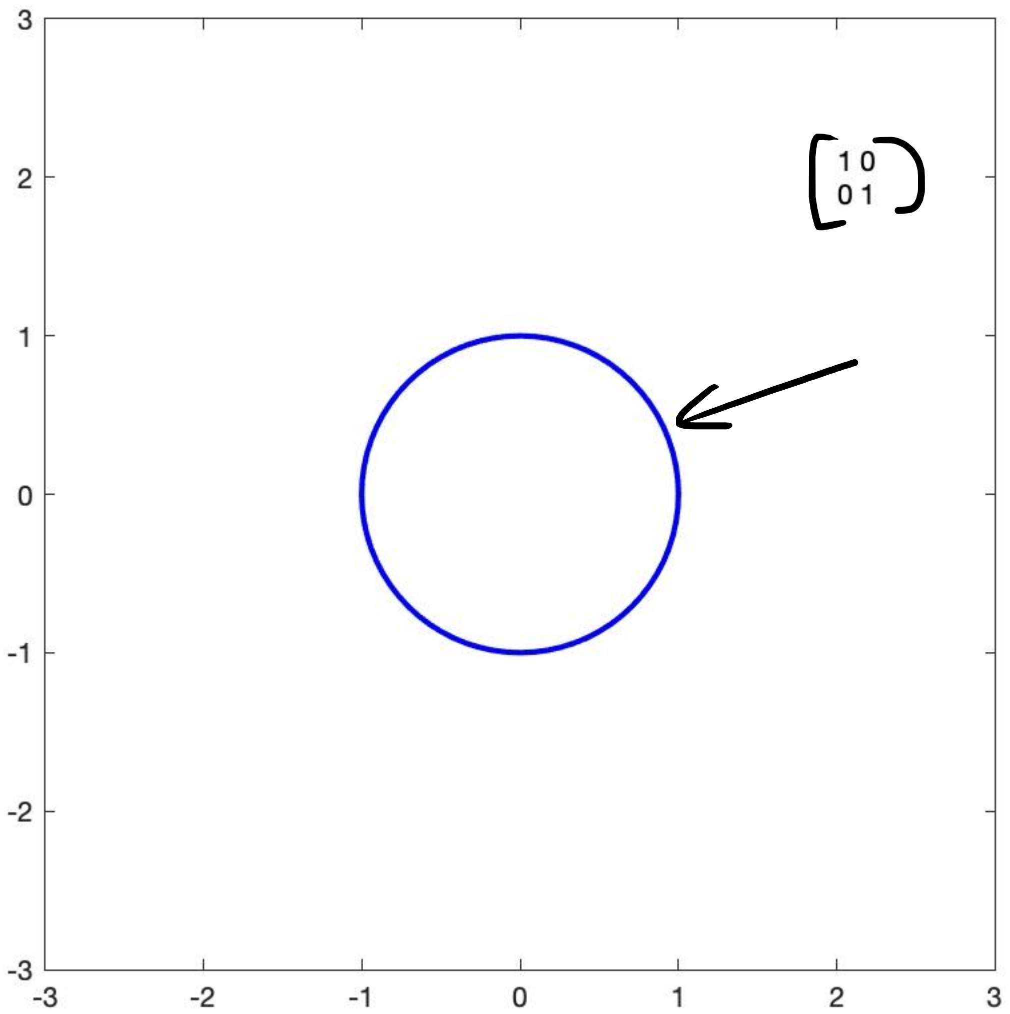
$$\mathbf{x}^\top \mathbf{P}^{-1} \mathbf{x} = c$$

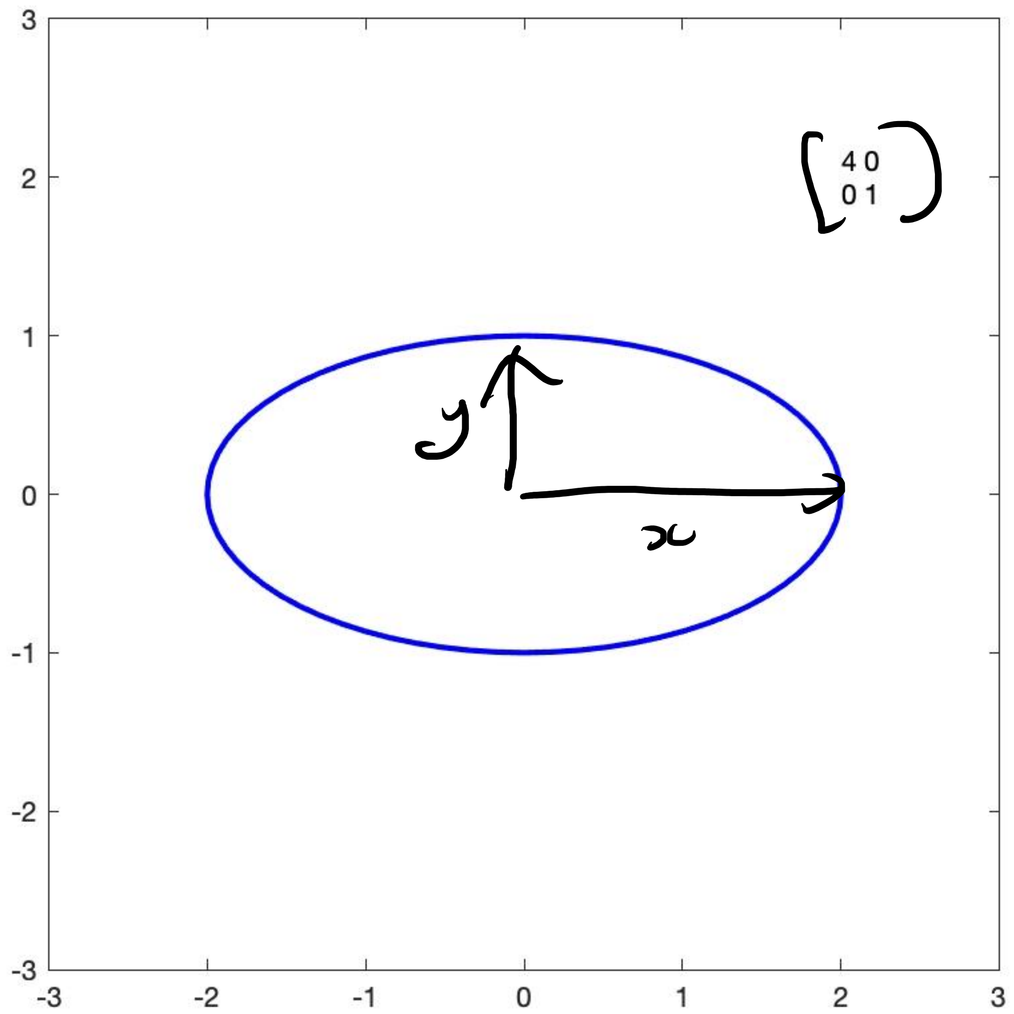
$$(x - E(x))^T P^{-1} (x - E[x]) = c$$

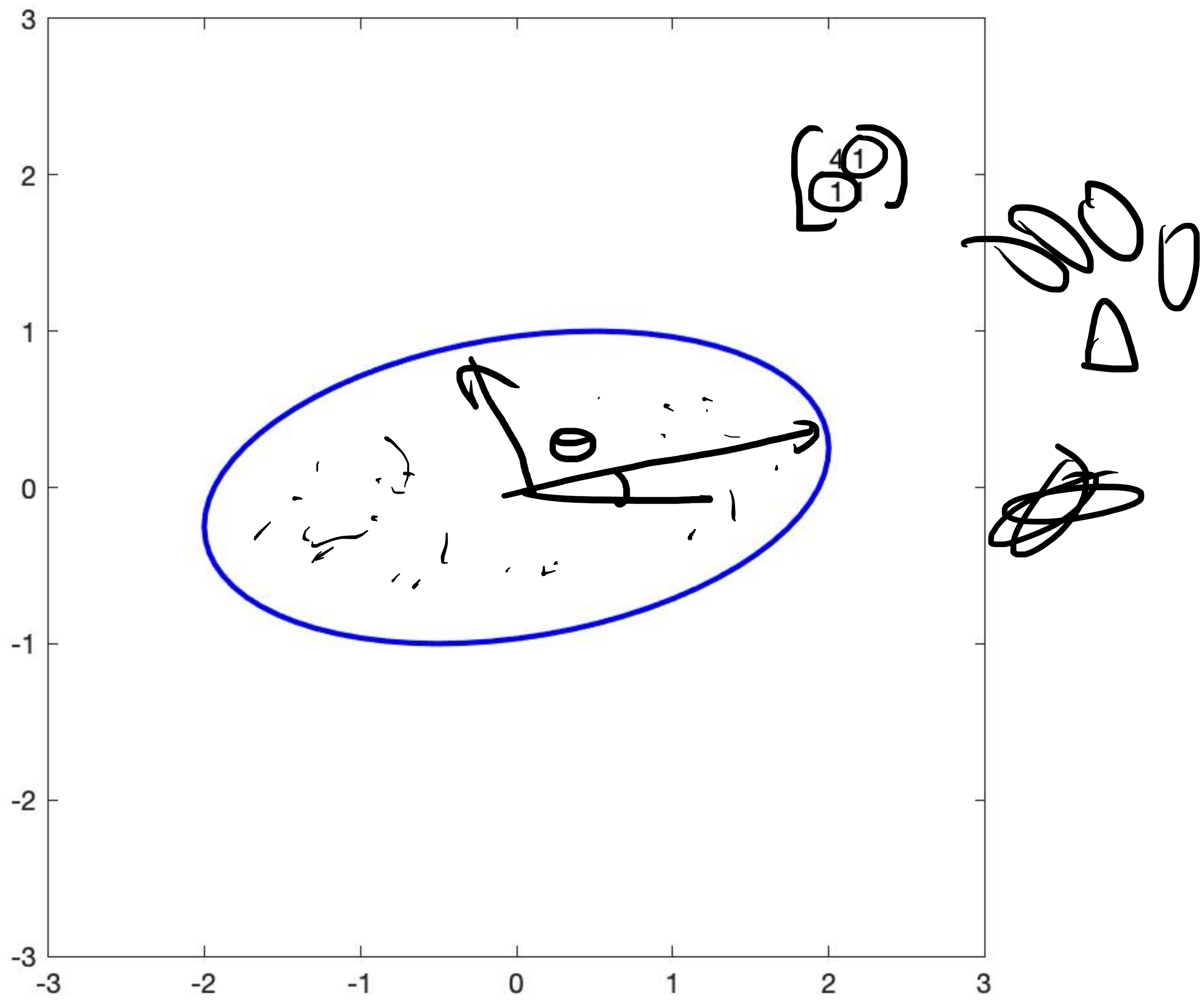
Covariance Ellipses

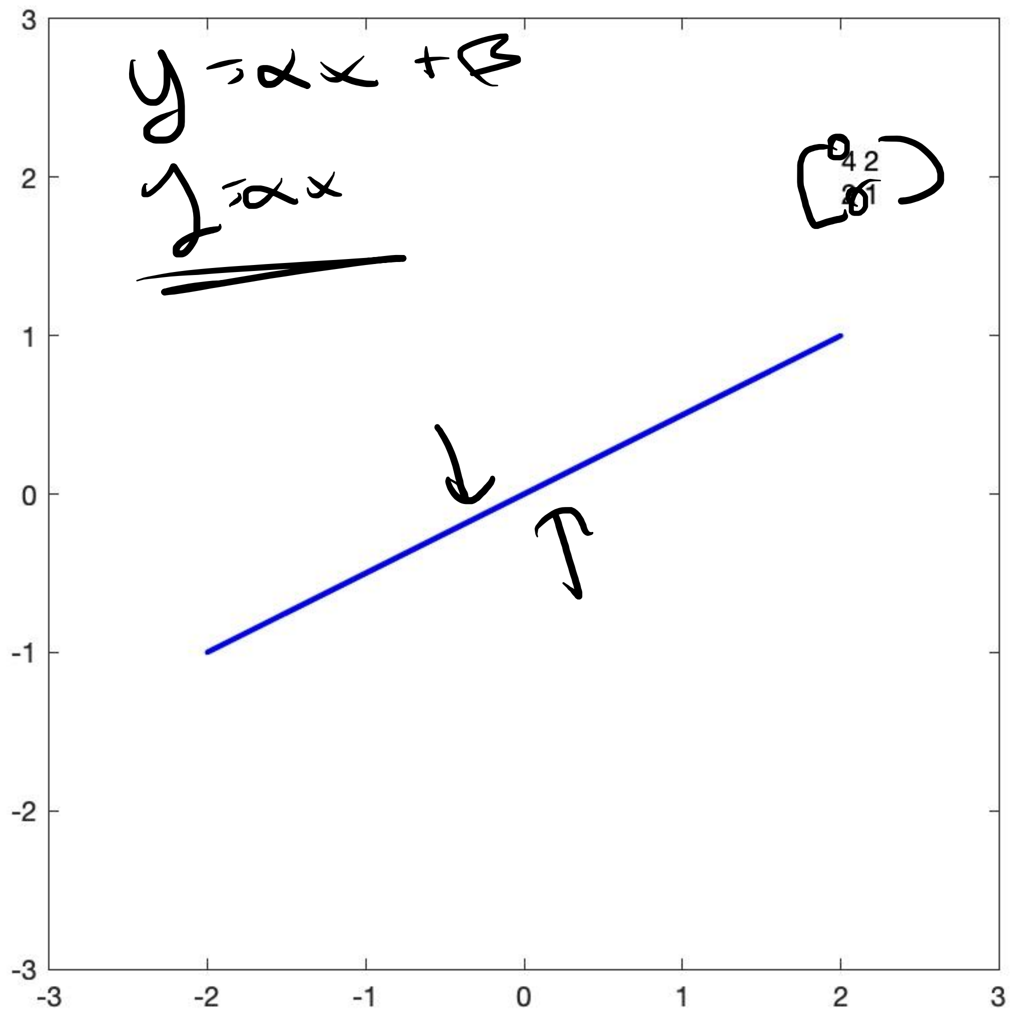
$$\boxed{p(x)} \propto \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

- There are two interpretations of this:
  - If your distribution is Gaussian, it is a contour of constant probability
  - If your distribution isn't Gaussian, it's a level set of points whose Mahalanobis Distance is the same value


3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
1 0
0 1


4 0
0 1
y
x


3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
4 1
1 1



A picture containing man, table, room

Description automatically generated

Limitations of Covariance Matrices

- People often use the word correlations to denote any form of dependency
- However, correlations only store linear relationships

$y = \alpha x + \beta$

Counter Example for Correlations

- Consider the system

$$x_t = \cos\theta_t$$

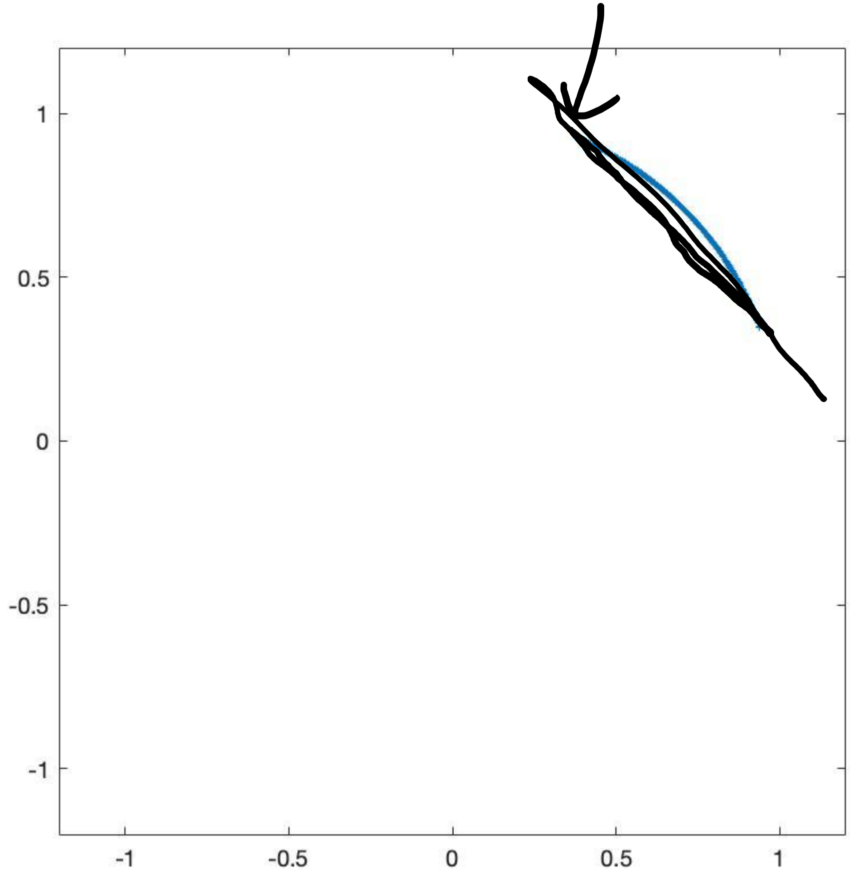
$$y_t = \sin\theta_t$$

$$\theta_t \sim \mathcal{G}\left(\theta; \hat{\theta}, P_{\theta\theta}\right)$$

$\hat{\theta} = 45°$

Small Error Case

$P_{00} = (5^\circ)^2$

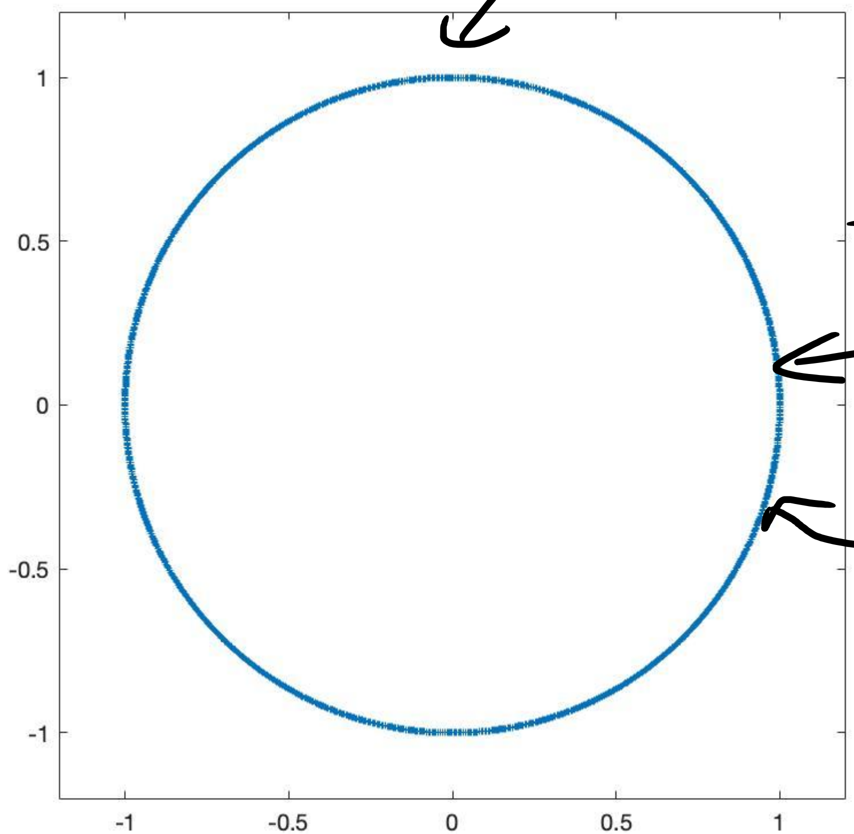


$$P_{xy} = 10^{-3} \times \begin{bmatrix} 3.795 & -3.766 \\ -3.766 & 3.795 \end{bmatrix}$$

$$c_{xy} = \underline{-0.992}$$

Large Error Case

$P_{\theta\theta} = (180°)^2$



$$P_{xy} = 10^{-3} \times \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

$$\boxed{c_{xy} = 0}$$

P_{xy}

Summary

$y_t = \alpha x_t + \beta$

- Mean squared error gives you a sense of how noisy your signal is
- Correlation coefficients give you a measure of dependency
- You can visualize them using covariance ellipses
- However, correlations are limited in the types of information they can capture