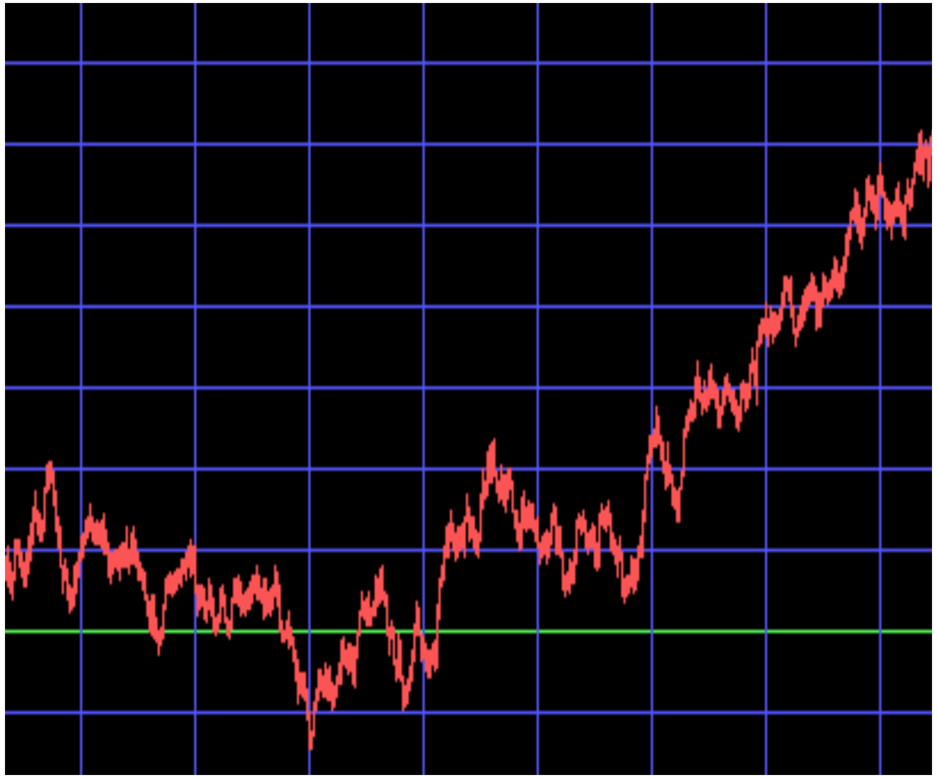
2.160 Identification, Estimation, and Learning

Part 1 Regression



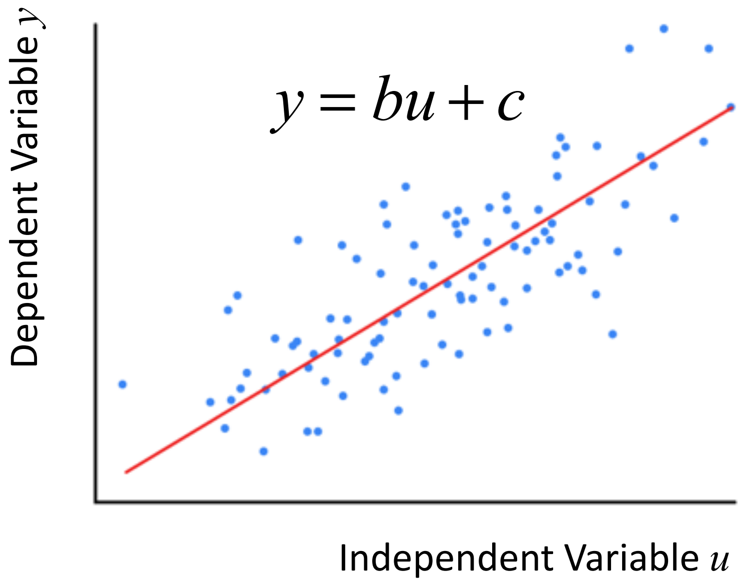
Lecture 4

Random Variables and Random Processes

H. Harry Asada
Department of Mechanical Engineering
MIT

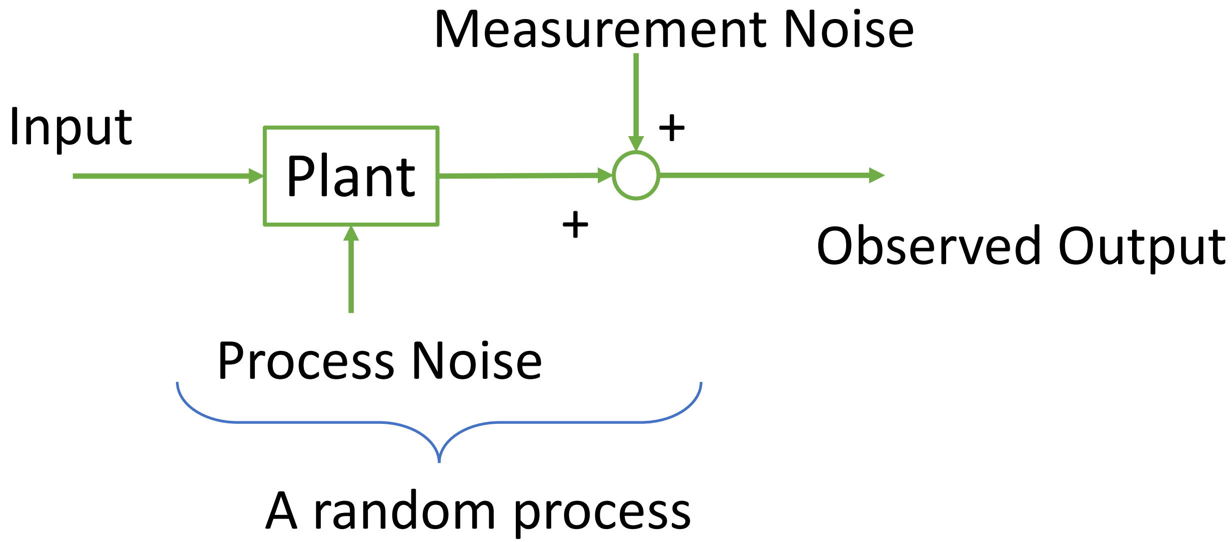
Quantifying Uncertainty

Estimation for deterministic systems



- Signals are corrupted with noise, which is random, but we ignored it;
- We did not quantify noise characteristics.

A dynamical system is perturbed by noise.



Quantification of Uncertainty (QU)

- Use stochastic properties (statistics) of the process for better estimating the parameters and the state of the process;
- Better understand, analyze, and evaluate estimation methods.

Quick Review of Probability and Random Variables

More rigorous studies:

- Set theory
- σ – field
- Lebesgue Measure

3.1 Probability and Random Variable

1). Random Variable

Random variable is a function that maps a random event to a numerical value.

$\longrightarrow$ 6+6 = 12

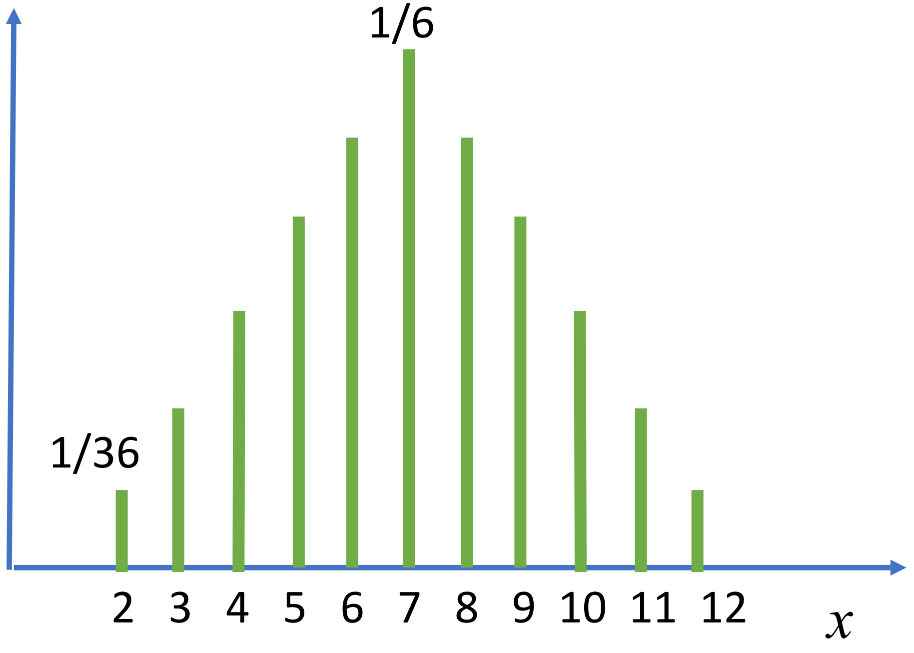
Event

Random event: $\omega \xrightarrow{\quad X \quad}$ Number

Random Variable Capital $\rightarrow$ $X = x_i$ $\leftarrow$ Instantiation Lower Case

Probability Mass Function

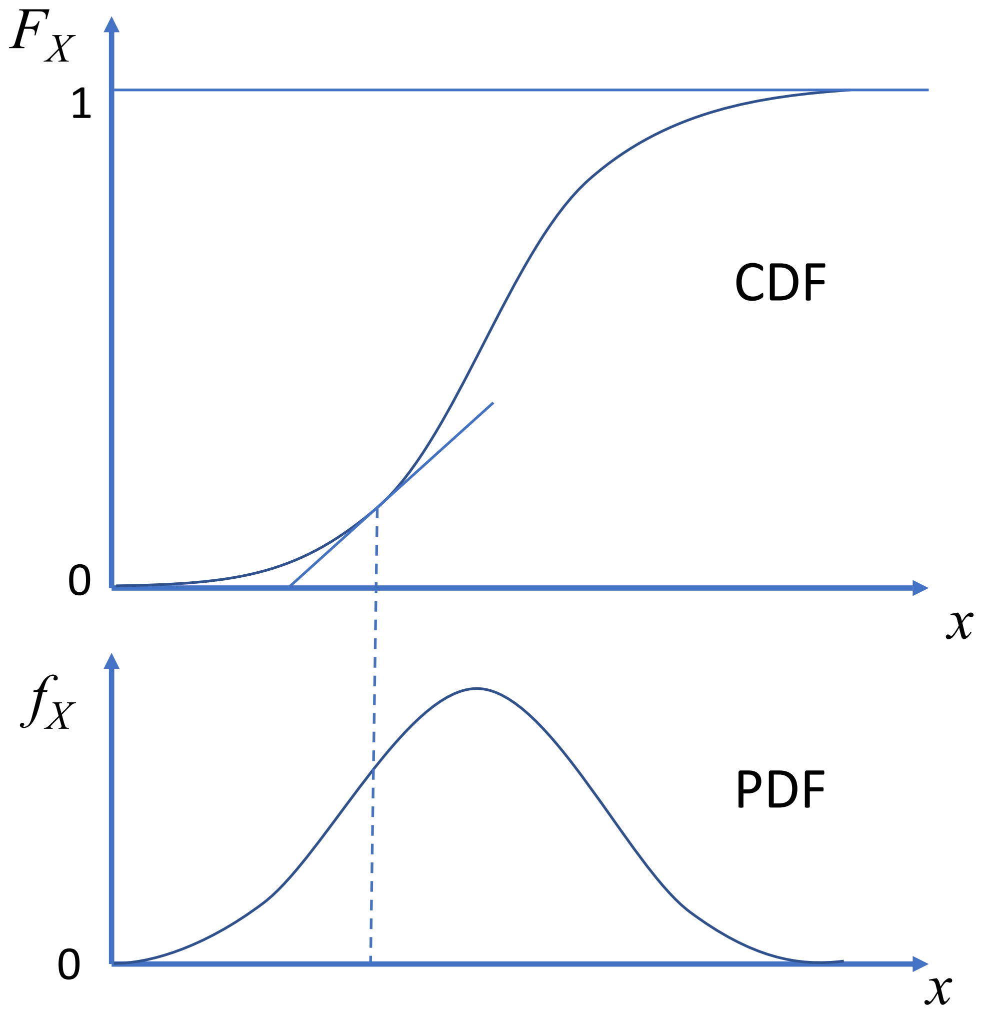
$$P_X(x_i) = \Pr(X = x_i)$$



2). Cumulative Distribution Function (CDF)

$$F_X(x) = \Pr(X \le x)$$

Note that for continuous X, $\Pr(X = x) = 0$



2-A). Probability Density Function (PDF)

$$f_X(x) = \frac{dF_X(x)}{dx}$$

$$F_X(x) = \int_{-\infty}^{x} f_X(z)\,dz$$

$$\int_{-\infty}^{\infty} f_X(z)\,dz = 1 \qquad\qquad f_X(x) \ge 0$$

3). Joint Probability

Discrete $p_{XY}(x_i, y_j) = \Pr(X = x_i \,\&\, Y = y_j)$ simultaneously

Continuous $f_{XY}(x, y)\Delta x \Delta y = \Pr(x \le X \le x + \Delta x \,\&\, y \le Y \le y + \Delta y)$

4). Independent Random Variables

$$f_{XY}(x, y) = f_X(x) f_Y(y), \quad \forall x, \forall y$$

For all x and y.

5). Conditional Probability Density

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

Conditional probability of X, given Y
(under the probability Y)

If X and Y are independent,

$$\frac{f_{XY}(x, y)}{f_Y(y)} = \frac{f_X(x) f_Y(y)}{f_Y(y)} = f_X(x)$$

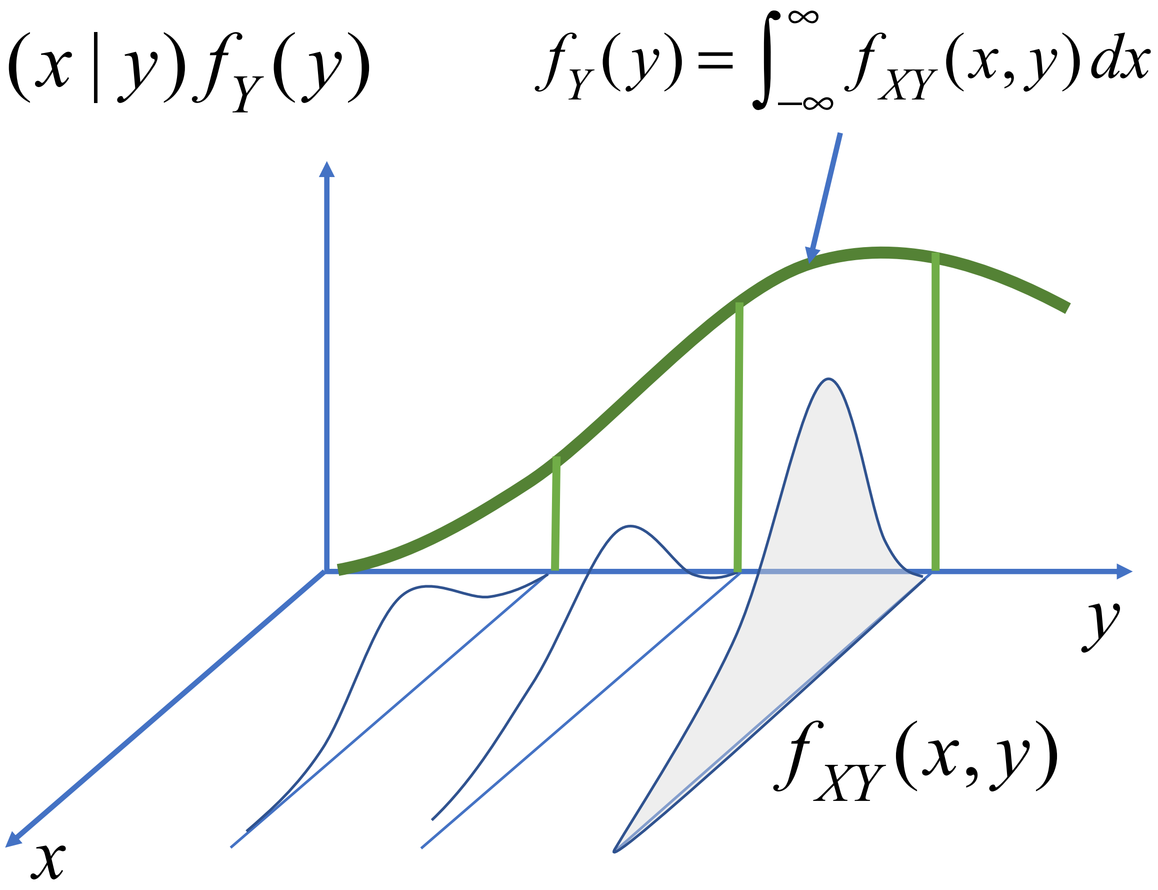
Does not depend on Y.

6). Bayes Rule

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

$$f_{XY}(x,y) = f_{Y|X}(y|x) f_X(x) = f_{X|Y}(x|y) f_Y(y)$$

$$\therefore f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)}$$



7). Marginal Probability

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$$

8). Expectation

Expected value of X , Synonym: mean, average

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

Discrete: $E[X] = \sum_i x_i p_i$

9). Variance

$$\mathrm{var}[X] = E[(X - E[X])^2] = E[X^2 - 2X \cdot E[X] + (E[X])^2]$$

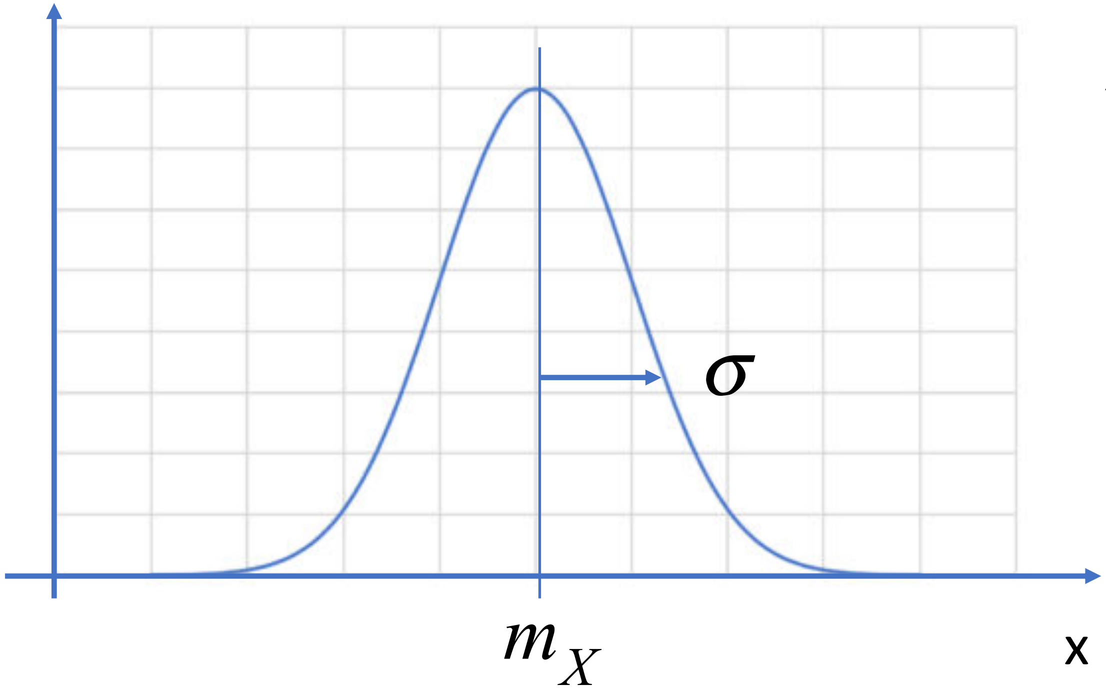
$$= E[X^2] - 2E[X] \cdot E[X] + (E[X])^2$$

$$\therefore \mathrm{var}[X] = E[X^2] - (E[X])^2$$

10). Moment

k-th moment $E[X^k] = \int_{-\infty}^{\infty} x^k f_X(x)\, dx$ k = 1, 2, 3, ….

11). Normal (Gaussian) Distribution



$$X \sim N(m_X, \sigma^2)$$

Random variable X has a normal distribution with mean m_X and variance σ^2

$$\sigma^2 = \int_{-\infty}^{\infty} (x - m_X)^2 f_X(x)\,dx$$

PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x - m_X)^2 \right]$$

Mean and variance completely characterize the distribution.

12). Correlation

The expectation of the product of two random variables, X and Y, is called "Correlation".

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, y\, \underline{f_{XY}(x,y)}\, dx\, dy$$

Joint probability density

If X and Y are independent,

$$E[XY] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = E[X]E[Y]$$

Note:

Although Correlation is zero, the two random variables are not necessarily independent.

13). Orthogonality

If correlation is zero, $E[XY] = 0$, X and Y are said to be "Orthogonal".

14). Covariance

$$\text{cov}[X,Y] = E[(X - E[X])(Y - E[Y])]$$

15). Correlation Coefficient

$$\rho_{XY} = \frac{\text{cov}[X,Y]}{\sqrt{\text{var}[X]}\sqrt{\text{var}[Y]}} = E[\frac{X - m_X}{\sigma_X}\frac{Y - m_Y}{\sigma_Y}]$$

$$-1 \le \rho_{XY} \le 1$$

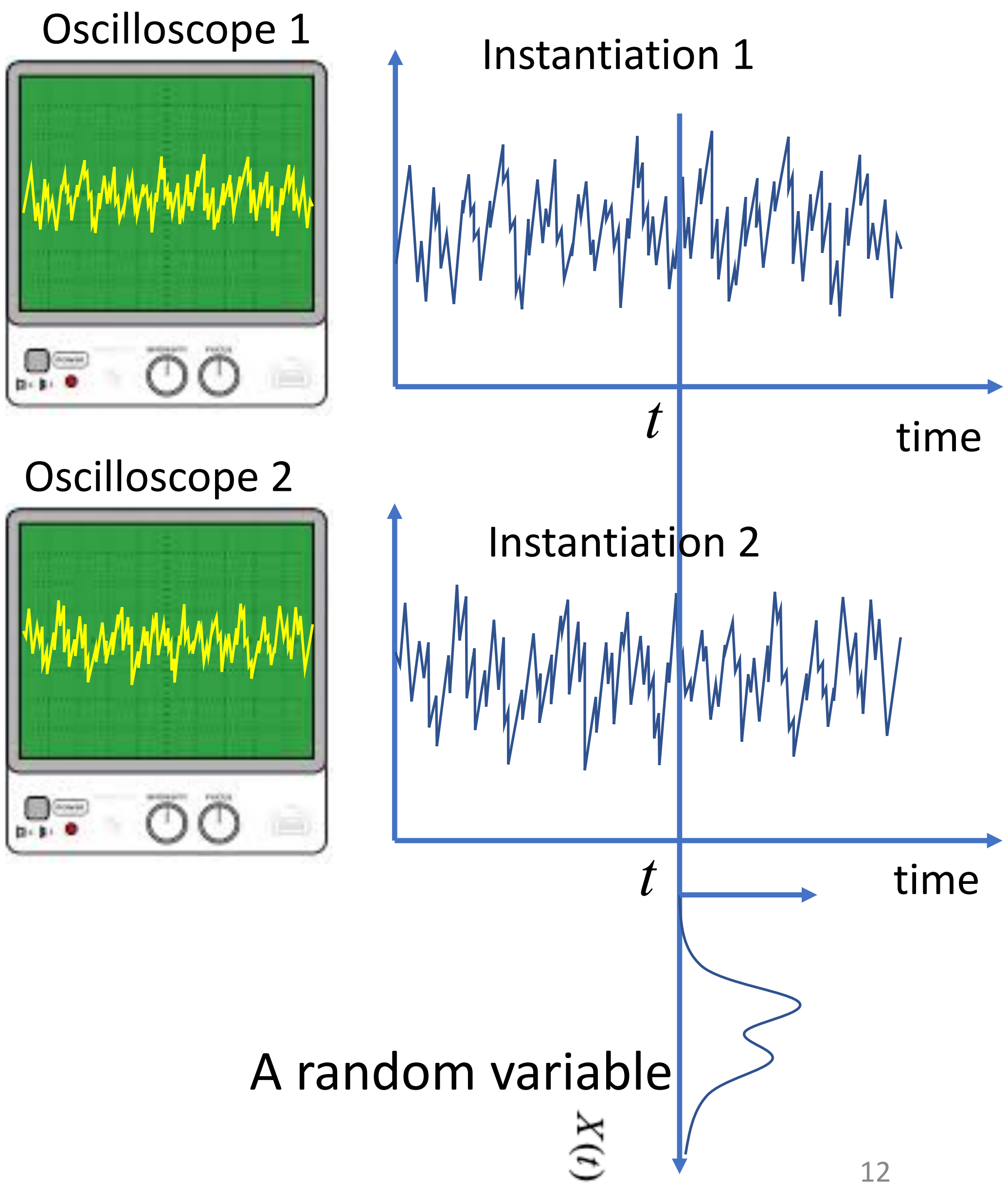
3.2 Random Processes

A random process is a family (ensemble) of time functions having a probability measure.

Consider a set of identical oscilloscopes measuring Ground Noise.
Each oscilloscope shows a particular noise waveform coming from the same source of random process, that is to say, each waveform is an instantiation.
Collecting voltage values displayed in all the oscilloscopes, we can construct a probability distribution, as shown right.
At each time the output of an oscilloscope is therefore a random variable, $X(t)$.
The random variables, $X(t), -\infty < t < \infty$ are <u>collectively</u> called a Random Process.

<u>Quantification of a random process:</u>

- First-Order Density $f_{X(t)}(x)$



Auto-Correlation

Quantification of a random process with correlation

Let us quantify how two random variables at two different times of a random process are correlated to each other.

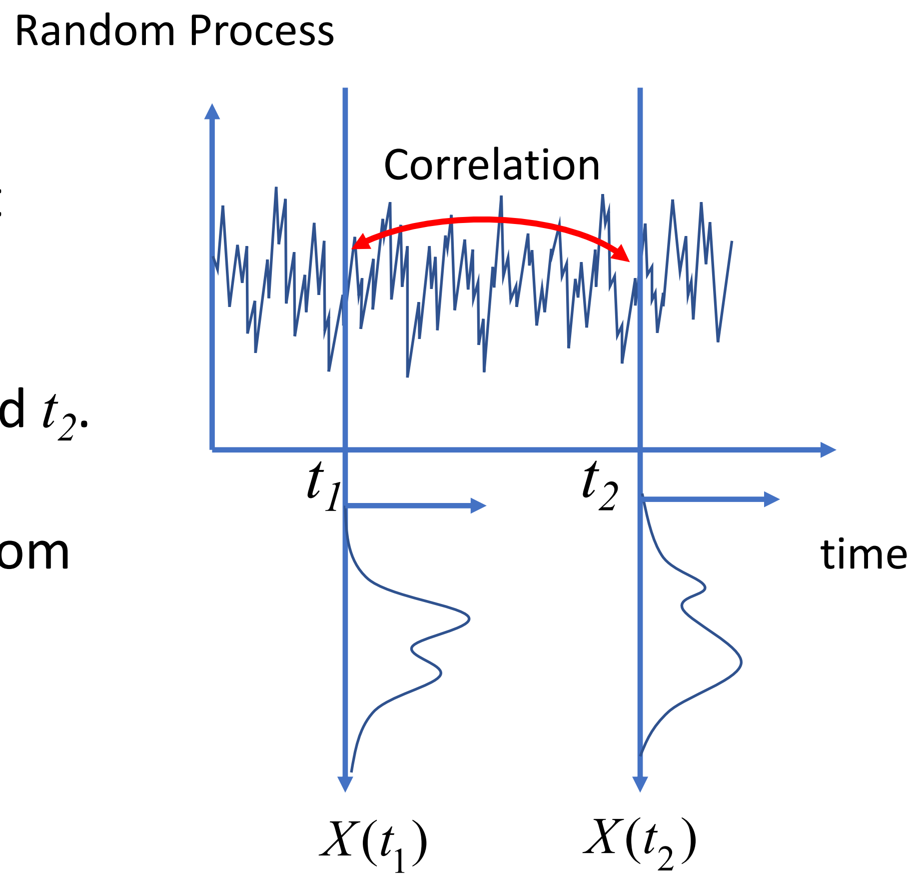
Let $X(t_1)$ and $X(t_2)$ be two random variables at time t_1 and t_2.

Let $f_{X_1X_2}(x_1,x_2)$ be the joint probability of the two random variables.

The correlation between the two is given by

$$R_{XX}(t_1,t_2) = E[X(t_1)X(t_2)]$$

$$= \int\int_{-\infty}^{\infty} x_1 x_2 f_{X_1X_2}(x_1,x_2)\,dx_1\,dx_2$$



Second-Order Densities

This is called ***Auto-Correlation***, since it is from the same random process.

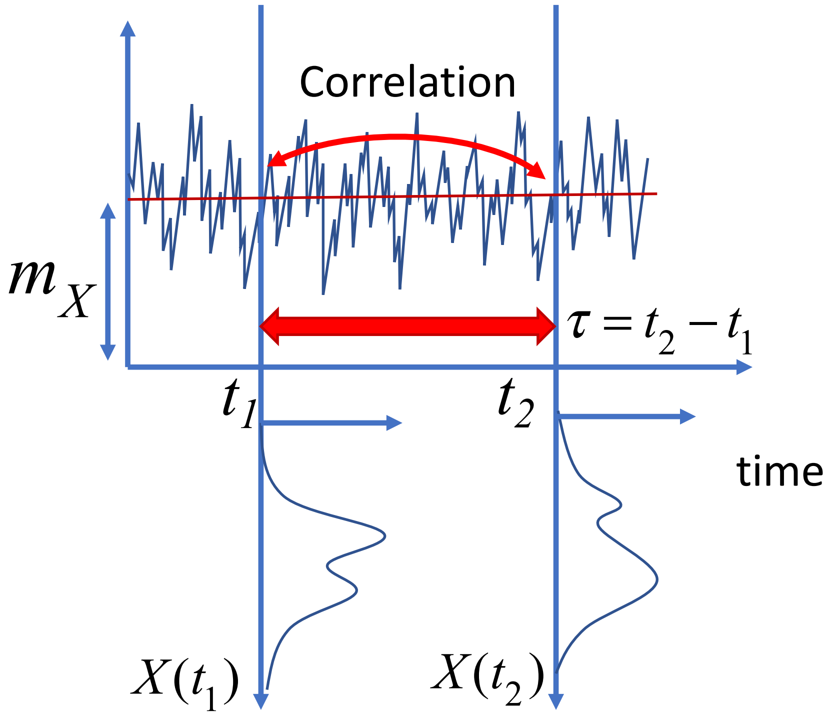
Wide-Sense Stationary Process

A random process is called a Wide-Sense Stationary Process, if the following two conditions are met:

- The mean value does not depend on time; the mean is uniform and constant throughout the process; $E[X(t)] = m_X; \quad \forall t$

- The Auto-Correlation depends only on the time interval, $\tau = t_2 - t_1$, rather than the specific times

$$R_{XX}(t_1, t_2) = R_{XX}(\tau) = E[X(t+\tau)X(t)]; \quad \forall t$$



Note that the auto-correlation of a wide-sense stationary process is an even function.

$$R_{XX}(-\tau) = R_{XX}(\tau)$$

Auto-Covariance: $C_{XX}(t_1, t_2) = E[(X(t_1) - m_X(t_1))(X(t_2) - m_X(t_2))]$

$$= R_{XX}(t_1, t_2) - m_X(t_1)m_X(t_2)$$

Independent and Identically Distributed (*i.i.d*. IID, iid) Random Variables and Processes

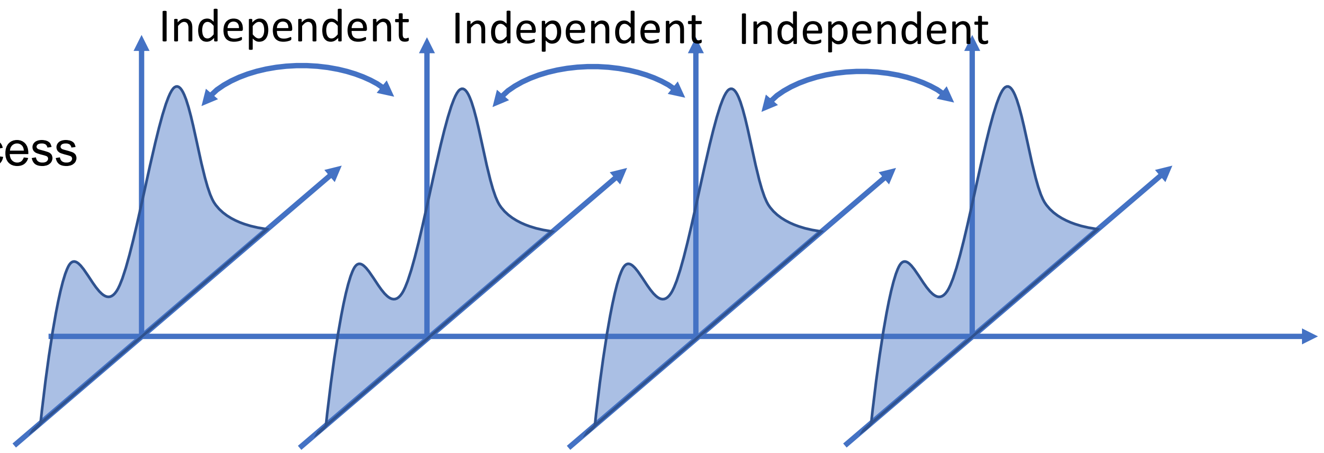
- Consider two random variables X and Y, with probability densities $f_X(x)$ and $f_Y(y)$.
- Random variables X and Y are called Independent and Identically Distributed (i.i.d.) if the following two conditions are met:

$$f_X(x) = f_Y(x), \quad \forall x$$

$$f_{XY}(x, y) = f_X(x) f_Y(y), \quad \forall x, \forall y$$

- i.i.d. is a convenient assumption widely used in statistics. Often it reflects samples taken from the same source.
- This concept can be easily extended to more than two random variables and random processes.

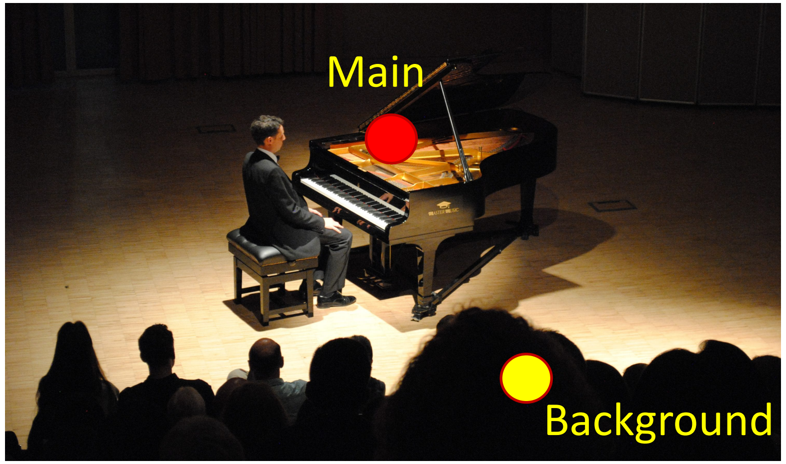
i.i.d. random process



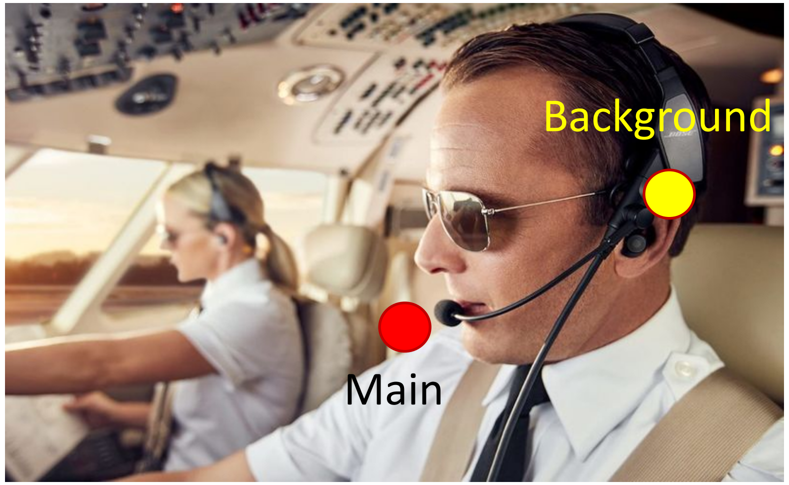
Note the difference between independence and orthogonality.

3.4 Adaptive Noise Cancellation: An Application

Before concluding this chapter, we apply the theory of random variables and random processes to a practical problem.



Active noise cancellation is a technique to measure the background acoustic noise and subtract it from the main signal so that the latter is not corrupted with the background noise.



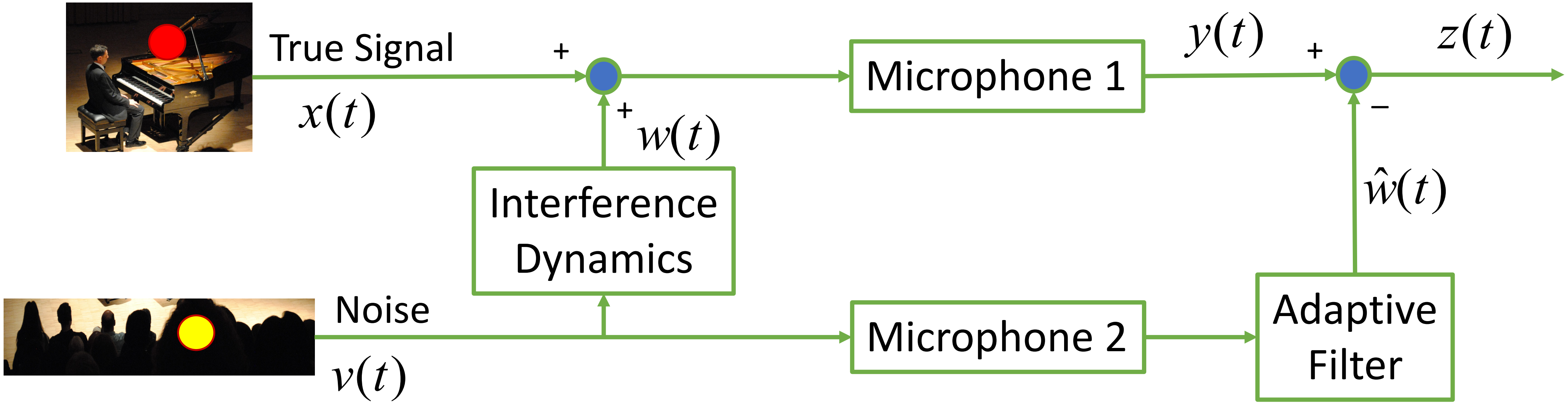
It has been applied to live music recording, aviation communication devices, and your music headset.

Two microphones are used; Microphone 1 for the main signal, and Microphone 2 for background noise.

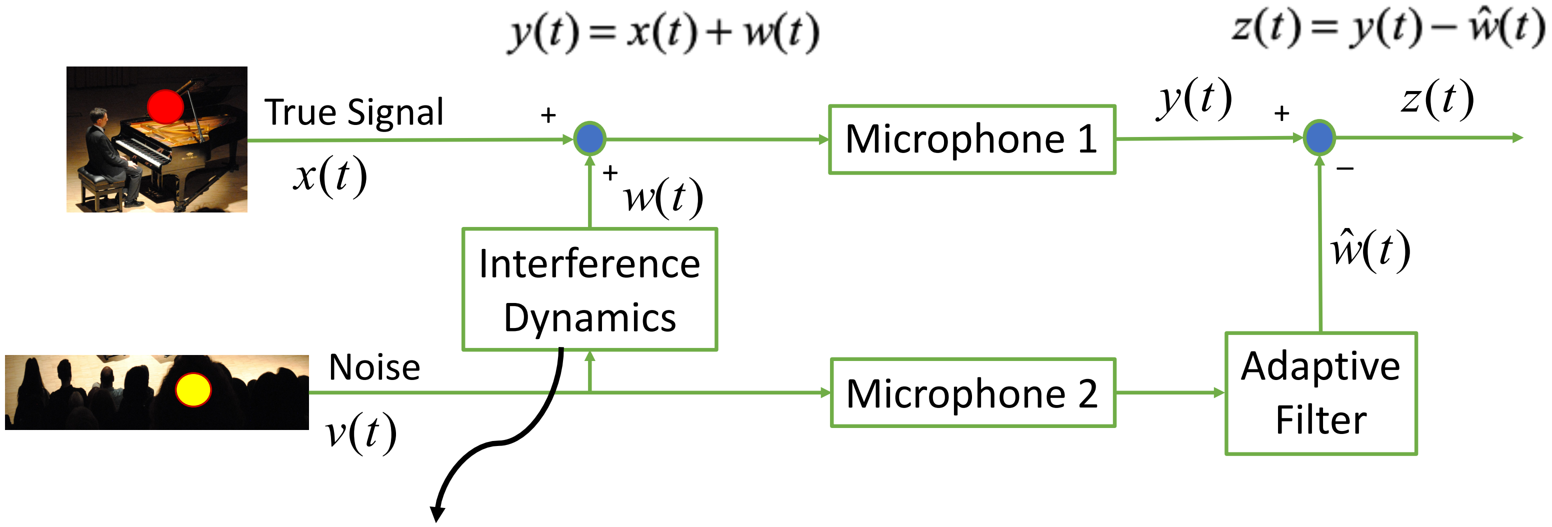
Active Noise Cancellation

- The background noise $v(t)$ propagates through an unknown dynamic process and arrives at Microphone 1, where the true signal $x(t)$ is mixed with the interfering noise $w(t)$.
- The background noise picked up with Microphone 2 is fed to Adaptive Filter, where the interfering noise is recovered.
- The predicted interfering noise $\hat{w}(t)$ is subtracted from the output of Microphone 1, $y(t)$, to recover the true signal.

$$y(t) = x(t) + w(t) \qquad z(t) = y(t) - \hat{w}(t)$$



Active Noise Cancellation



We assume a Finite Impulse Model for the Interfering Dynamics.

$$w(t) = b_1 v(t-1) + b_2 v(t-2) + \cdots + b_m v(t-m) \quad \Longrightarrow \quad w(t) = \theta^T \varphi(t)$$

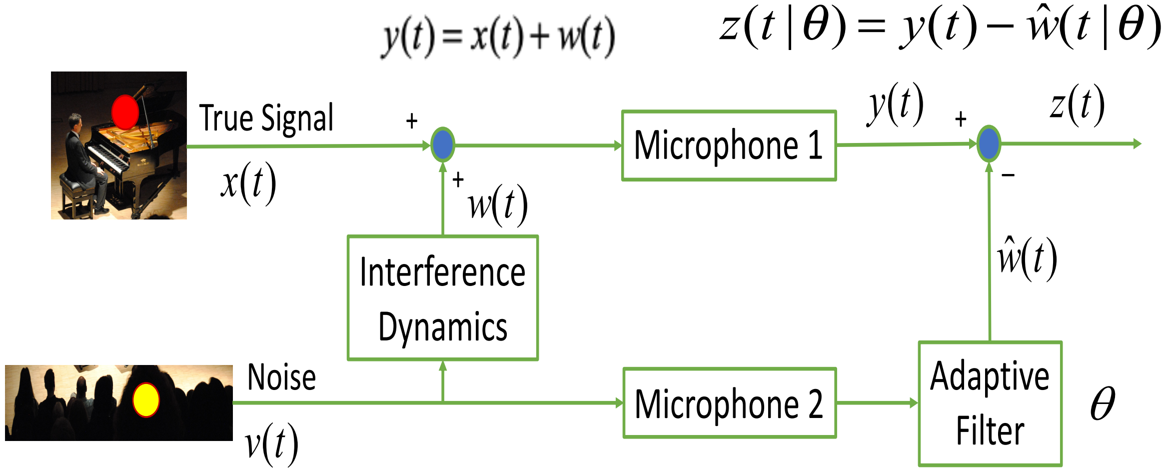
A linear regression

Parameters to estimate: $\theta = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$ Regressor: $\varphi(t) = \begin{pmatrix} v(t-1) \\ \vdots \\ v(t-m) \end{pmatrix}$

Assumption: True signal $x(t)$ and background noise $v(t)$ are uncorrelated random processes,

$$E[x(t)v(t-\tau)] = 0, \quad \forall t, \forall \tau$$

Active Noise Cancellation



Problem: Estimate parameter vector θ in real-time, so that the noise may be most suppressed.

Solution: Consider the recovered output $z(t)$, which depends on the parameter values :

$$z(t \mid \hat{\theta}) = y(t) - \hat{w}(t \mid \hat{\theta})$$

The mean squared signal strength of $z(t \mid \hat{\theta})$

$$E[(z(t \mid \hat{\theta}))^2] = E[(y(t) - \hat{w}(t \mid \hat{\theta}))^2] = E[(x(t) + w(t) - \hat{w}(t \mid \hat{\theta}))^2]$$

$$= E[(x(t))^2] + 2E[x(t)\{w(t) - \hat{w}(t \mid \hat{\theta})\}] + E[(w(t) - \hat{w}(t \mid \hat{\theta}))^2]$$

Examine

$$E[x(t)w(t)] = E[x(t)\{b_1 v(t-1) + \cdots + b_m v(t-m)\}]$$

$$= b_1 E[x(t)v(t-1)] + \cdots + b_m E[x(t)v(t-m)] = 0$$

since the true signal and background noise are uncorrelated: $E[x(t)v(t-\tau)] = 0, \quad \forall t, \forall \tau$

Similarly, $$E[x(t)\hat{w}(t)] = E[x(t)\{\hat{b}_1 v(t-1) + \cdots + \hat{b}_m v(t-m)\}]$$

$$= \hat{b}_1 E[x(t)v(t-1)] + \cdots + \hat{b}_m E[x(t)v(t-m)] = 0$$

Active Noise Cancellation

$$E[(z(t\,|\,\hat{\theta}))^2] = E[(x(t))^2] + \cancel{2E[x(t)w(t)]} - \cancel{2E[x(t)\hat{w}(t\,|\,\hat{\theta})]} + E[(w(t) - \hat{w}(t\,|\,\hat{\theta}))^2]$$

This term does not depend on parameter θ.

$$E[x(t)w(t)] = 0$$

$$E[x(t)\hat{w}(t)] = 0$$

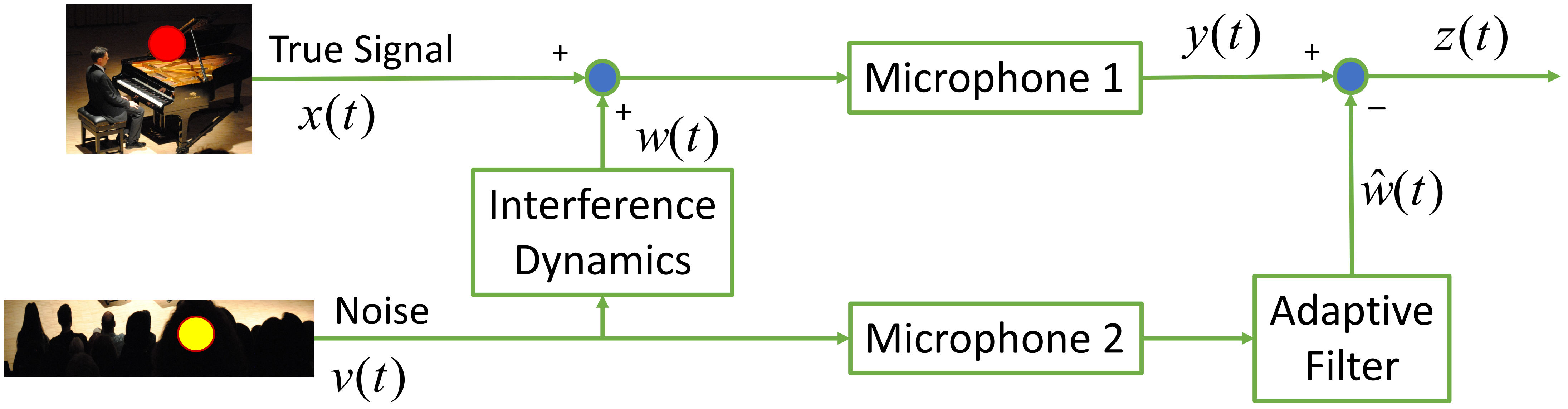
This can be treated as a prediction error.

$$\theta^o = \arg\min_{\hat{\theta}} E[(w(t) - \hat{w}(t\,|\,\hat{\theta}))^2]$$

We want to minimize this squared error in predicting the interfering noise.

$$y(t) = x(t) + w(t)$$

$$z(t) = y(t) - \hat{w}(t)$$



Active Noise Cancellation

$$E[(z(t\,|\,\hat{\theta}))^2] = E[(x(t))^2] + \cancel{2E[x(t)w(t)]} - \cancel{2E[x(t)\hat{w}(t\,|\,\hat{\theta})]} + E[(w(t)-\hat{w}(t\,|\,\hat{\theta}))^2]$$

This term does not depend on parameter θ.

$$E[x(t)w(t)] = 0 \qquad E[x(t)\hat{w}(t)] = 0$$

This can be treated as a prediction error.

$$\theta^o = \arg\min_{\hat{\theta}} E[(w(t)-\hat{w}(t\,|\,\hat{\theta}))^2]$$

We want to minimize this squared error in predicting the interfering noise.

$$= \arg\min \{E[(x(t))^2] + E[(w(t)-\hat{w}(t\,|\,\hat{\theta}))^2]\}$$

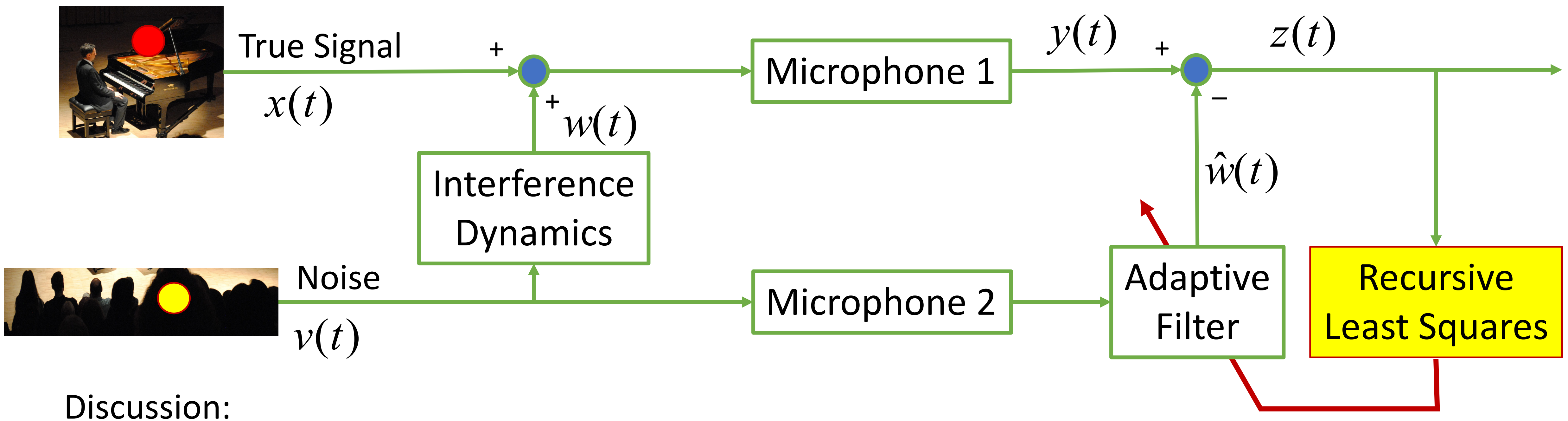
$$= \arg\min_{\hat{\theta}} E[(z(t\,|\,\hat{\theta}))^2]$$

Equivalent to minimize this, which is computable.

Treating $w(t)-\hat{w}(t\,|\,\hat{\theta})$ as a prediction error and replacing it by $z(t\,|\,\hat{\theta})$, we can apply the Recursive Least Squares algorithm to estimate the parameters involved in the interfering dynamics (FIR model) in real-time. Using forgetting factor α,

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{P_{t-1}\varphi(t)}{\alpha + \varphi^T(t)P_{t-1}\varphi(t)}[\underbrace{y(t) - \hat{y}(t\,|\,\hat{\theta}(t-1))}_{z(t\,|\,\hat{\theta})}] \qquad P_t = \frac{1}{\alpha}\left[P_{t-1} - \frac{P_{t-1}\varphi(t)\varphi^T(t)P_{t-1}}{\alpha + \varphi^T(t)P_{t-1}\varphi(t)}\right]$$

Active Noise Cancellation using RLS and Orthogonality



Discussion:

1. What if the true signal x(t) is picked up by Microphone 2, too?
2. Does $\hat{b}_i$ correlate with x(t)? If so, it cannot be factored out.

Answer:

Context-Oriented Project #1 is on a related topic.
You should discuss it in your study group.