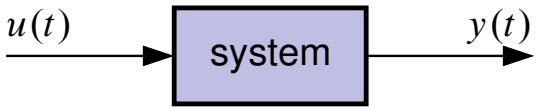
Unit-Pulse-Response Identification

2.1: Continuous-time linear time-invariant systems

- Consider a continuous-time system, with input $u(t)$ and output $y(t)$.



- In this course, we deal exclusively with linear time-invariant (LTI) systems, although the text has some sections on nonlinear and time-varying system identification.

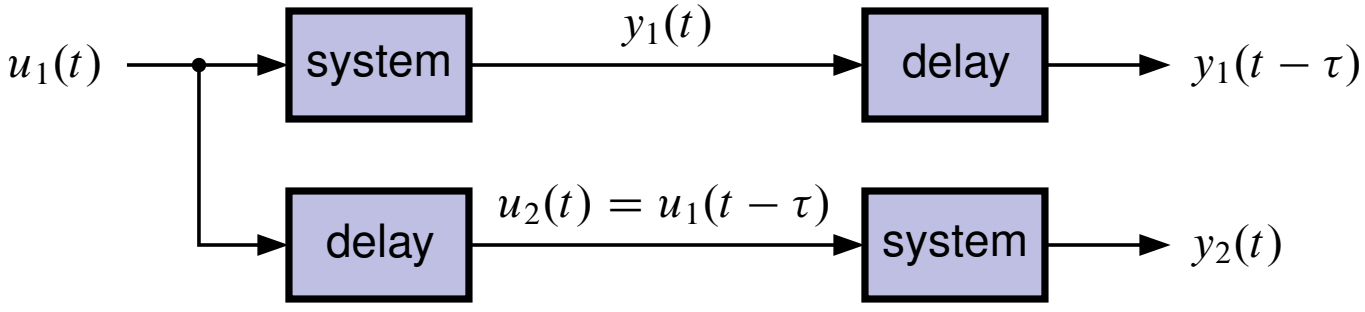
TIME INVARIANT: We first consider time invariance.

- A system is either time-varying or time-invariant, not both.
- A time-invariant system does not change its fundamental behavior over different periods of time. Its parameter values are constant.
- A time-invariant system satisfies the property (for any $u(t)$, τ)

$$u(t-\tau) \mapsto y(t-\tau)$$

  when $u(t) \mapsto y(t)$.

- We can test a system for this property using ideas from the figure.



- A time-invariant system will have $y_2(t) = y_1(t-\tau)$ for all $u_1(t)$ and τ.

> **TEST:** To test for time-invariance, we must
>
> - Input $u_1(t)$ to the system and measure the output $y_1(t)$.
> - Input $u_2(t) = u_1(t-\tau)$ to the system and measure $y_2(t)$.
> - If $y_2(t) = y_1(t-\tau)$ for all possible delays τ and signals $u_1(t)$, then the system is time-invariant.

EXAMPLE: For example, consider a square-law system $y(t) = (u(t))^2$.

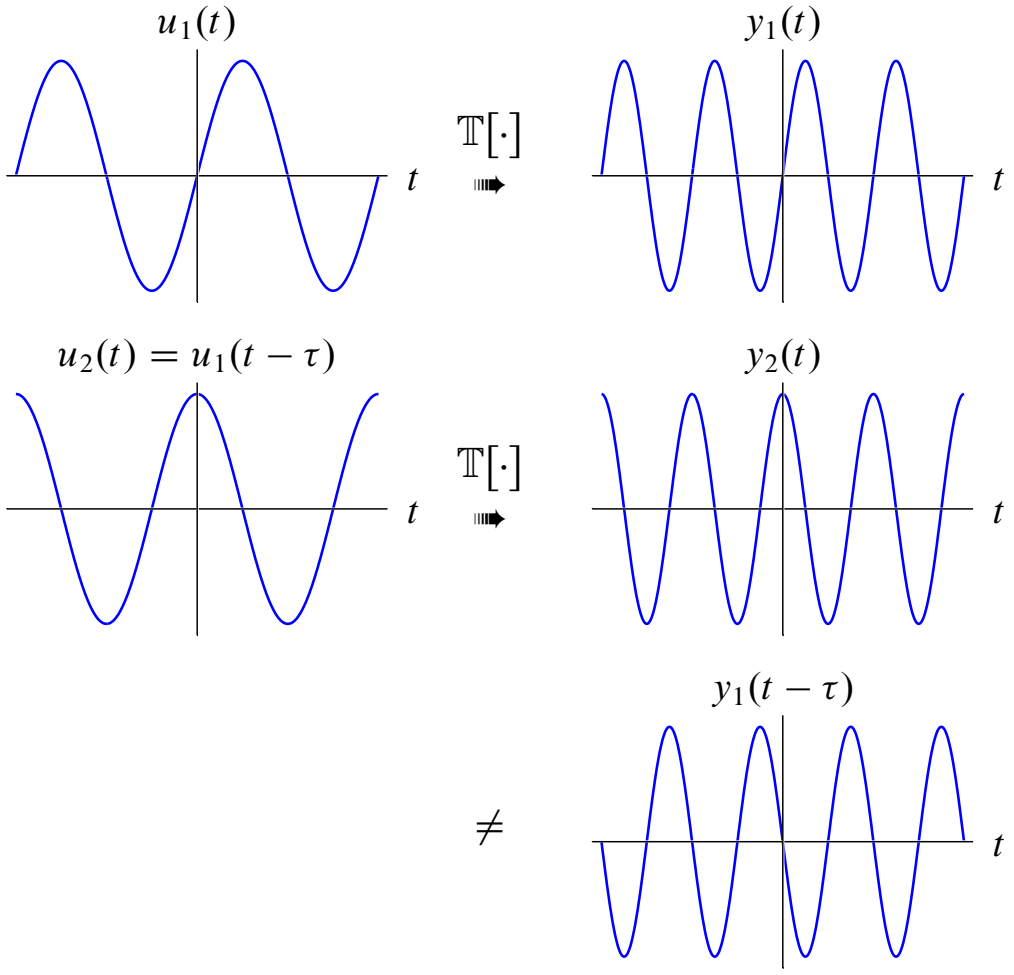
- Input $u_1(t)$ to the system and measure $y_1(t)$: $y_1(t) = (u_1(t))^2$.
- Input $u_2(t)$ to the system and measure $y_2(t)$: $y_2(t) = (u_2(t))^2$.
- But, $u_2(t) = u_1(t-\tau)$, so $y_2(t) = (u_1(t-\tau))^2 = y_1(t-\tau)$.
- Since this relationship holds for all τ and all $u_1(t)$, the square-law system *is* time-invariant.

EXAMPLE: Let us examine a "delay operator." (The delay operator is a fundamental building-block of digital-signal-processing systems and digital control systems).

- The output of a delay is equal to the input, but shifted a constant amount λ seconds: $y(t) = u(t-\lambda), \quad \lambda \geq 0$.
- Input $u_1(t)$ to the system and measure $y_1(t)$: $y_1(t) = u_1(t-\lambda)$.
- Input $u_2(t)$ to the system and measure $y_2(t)$: $y_2(t) = u_2(t-\lambda)$.
- But, $u_2(t) = u_1(t-\tau)$, so $y_2(t) = u_1(t-\tau-\lambda) = y_1(t-\tau)$.
- Since this relationship holds for all τ and all $u_1(t)$, the delay operator *is* time-invariant.

EXAMPLE: Let us examine a "time compressor" whose output is equal to its input, but "squashed" in time: $y(t) = u(kt)$, $k > 1$.

- Input $u_1(t)$ to the system and measure $y_1(t)$: $y_1(t) = u_1(kt)$.
- Input $u_2(t)$ to the system and measure $y_2(t)$: $y_2(t) = u_2(kt)$.
- But, $u_2(t) = u_1(t-\tau)$, so $y_2(t) = u_1(kt-\tau) \neq u_1(kt-k\tau) = y_1(t-\tau)$.
- Therefore, the compressor is time-varying.



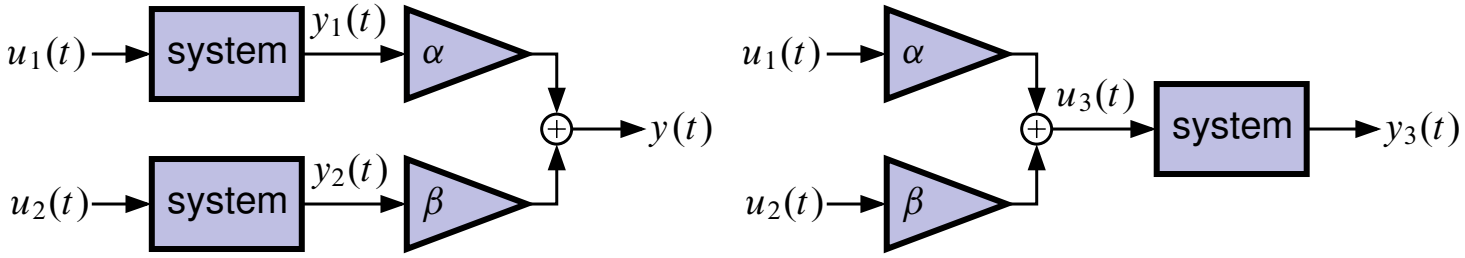
LINEAR: The second property that we look at is linearity.

- For linear systems, if $u_1(t) \mapsto y_1(t)$ and $u_2(t) \mapsto y_2(t)$, then

$$u_3(t) = \alpha u_1(t) + \beta u_2(t) \mapsto y_3(t) = \alpha y_1(t) + \beta y_2(t),$$

for any such $u_1(t)$, $u_2(t)$, α, β.

- We can test a system for this property using ideas from the figure.



TEST: To test for linearity, we must

- Input $u_1(t)$ to the system and measure the output $y_1(t)$.
- Input $u_2(t)$ to the system and measure $y_2(t)$.
- Input $u_3(t) = \alpha u_1(t) + \beta u_2(t)$ to the system and measure $y_3(t)$.
- If $y_3(t) = \alpha y_1(t) + \beta y_2(t)$ for all possible α and β values, and $u_1(t)$ and $u_2(t)$, then the system is linear.

EXAMPLE: Is the following system, described by the differential equation $\dot{y}(t) + ty(t) = u(t)$, linear?[1]

- Input $u_1(t)$ and output is $y_1(t)$: $\dot{y}_1(t) + ty_1(t) = u_1(t)$.
- Input $u_2(t)$ and output is $y_2(t)$: $\dot{y}_2(t) + ty_2(t) = u_2(t)$.
- Input $u_3(t) = \alpha u_1(t) + \beta u_2(t)$ and measure $y_3(t)$.

$$\dot{y}_3(t) + ty_3(t) = u_3(t);$$

but, $u_3(t) = \alpha u_1(t) + \beta u_2(t)$, so

$$\begin{aligned}\dot{y}_3(t) + ty_3(t) &= \alpha u_1(t) + \beta u_2(t) \\ &= \alpha\left(\dot{y}_1(t) + ty_1(t)\right) + \beta\left(\dot{y}_2(t) + ty_2(t)\right)\end{aligned}$$

[1] Note, the "dot" decoration on a variable indicates a time derivative. That is, $\dot{y}(t) = \mathrm{d}y(t)/\mathrm{d}t$, $\ddot{y}(t) = \mathrm{d}^2y(t)/\mathrm{d}t^2$, and so forth.

$$= \frac{\mathrm{d}}{\mathrm{d}t} \underbrace{\left(\alpha y_1(t) + \beta y_2(t)\right)}_{i.e.,\ y_3(t)} + t \underbrace{\left(\alpha y_1(t) + \beta y_2(t)\right)}_{i.e.,\ y_3(t)}.$$

By examining both sides of this equation, we realize that $y_3(t) = \alpha y_1(t) + \beta y_2(t)$. Therefore, the system is linear.

EXAMPLE: Trying this on the square-law system,

- Input $u_1(t)$ and output is $y_1(t)$: $y_1(t) = (u_1(t))^2$.
- Input $u_2(t)$ and output is $y_2(t)$: $y_2(t) = (u_2(t))^2$.
- Input $u_3(t) = \alpha u_1(t) + \beta u_2(t)$ and measure $y_3(t)$.

$$\begin{aligned} y_3(t) &= (u_3(t))^2 \\ &= (\alpha u_1(t) + \beta u_2(t))^2 \\ &= \alpha^2 (u_1(t))^2 + 2\alpha\beta u_1(t) u_2(t) + \beta^2 (u_2(t))^2 \\ &\neq \alpha \left(u_1(t)\right)^2 + \beta \left(u_2(t)\right)^2. \end{aligned}$$

- So, the square-law system is *not* linear (it is a *nonlinear system*).

KEY POINT: If a continuous-time system is LTI, then it has an impulse response. This entirely characterizes the system's dynamics.

- The Fourier transform of the impulse response is the frequency response, which also entirely characterizes the system.
- The Laplace transform of the impulse response is the transfer function, which also entirely characterizes the system. And,
- Lumped LTI systems can also be represented in state-space form.

- Similar concepts apply to discrete-time systems.
- This course examines system ID using these four different types of LTI system representation, for discrete-time systems.

2.2: The importance of the unit-pulse response

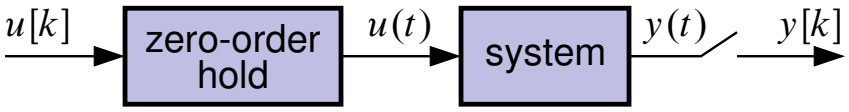
- Systems that are both linear and time invariant are called "linear time-invariant" (LTI) systems.
- LTI is really an abstraction: There is no such thing in practice.
- But, from an engineering point of view, the approximation is often "close enough."
- Continuous-time LTI systems have output completely defined by the input and their impulse response, $g(t)$.
  - Recall that the impulse function $\delta(t)$ is a strange "generalized" function with two properties:
    - Zero duration: $\delta(t) = 0, \quad t \neq 0.$
    - Unit area: $\int_{-\infty}^{\infty} \delta(t)\,\mathrm{d}t = 1.$
  - If the input to a system is $\delta(t)$, then we define the system's output (time response) to be the impulse response, $g(t)$.

KEY POINT: If an LTI system's impulse response $g(t)$ is known, we can find its time response to any input via a convolution integral

$$y(t) = \int_{-\infty}^{\infty} g(\tau)u(t-\tau)\,\mathrm{d}\tau.$$

- We typically assume that $g(t) = 0 \quad \forall \quad t < 0.$
  - $g(t)$ then defines a causal relationship between the continuous-time input signal $u(t)$ and the continuous-time output signal $y(t)$.
- We could model the system very well given a good way to find $g(t)$.
- Complicating factor: System identification almost always done in discrete time, which has a slightly different "impulse response."

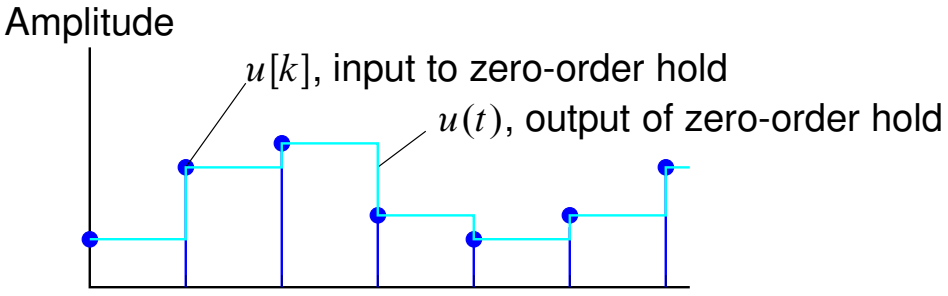
- That is, our data will typically be collected from real systems
  - Discrete data, which therefore requires discrete models.



- Assume that we sample the output of the system at discrete times $t = kT$, (and introduce notation $[\cdot]$ versus $(\cdot)$, explained below)

$$y[k] = y(kT) = \int_0^\infty g(\tau)u(kT-\tau)\,\mathrm{d}\tau.$$

- To simplify the analysis, assume that the input $u(t)$ is piecewise constant (output of $u[k]$ applied to a zero-order hold, ZOH).
  - That is, $u(t) = u[k] \quad \forall \quad kT < t < (k+1)T$.
  - Not always valid (*e.g.*, uncontrolled inputs such as process noise).



- We now get, $y(kT) = \sum_{m=1}^{\infty} \int_{(m-1)T}^{mT} g(\tau)u(kT-\tau)\,\mathrm{d}\tau.$
- As $u(kT-\tau)$ is constant over an integration interval, equal to $u[k-m]$,

$$y[k] = \sum_{m=1}^{\infty} g_T[m]u[k-m],$$

where

$$g_T[m] = \int_{(m-1)T}^{mT} g(\tau)\,\mathrm{d}\tau,$$

is the unit-pulse response (also commonly known, less precisely, as the discrete-time impulse response) of the sampled-data system.

- Note some differences in notation between these notes and Ljung:
  - I use square brackets to indicate functions of a discrete input variable, such as $u[k]$ versus $u(t)$.
  - Ljung uses $u(t)$ to refer to both continuous-time and discrete-time inputs t, which I find confusing, so I prefer $u[k]$ for discrete time.
  - $y(kT)$ is a bit of a hybrid, since it evaluates $y(t)$ at discrete points $t = kT$. Notionally, $y[k] = y(kT)$.
  - I don't believe we use $g_T[m]$ in the future to distinguish between the continuous-time impulse responses and discrete-time unit-pulse responses.
    - We discern the difference by context (*e.g.,* $(\cdot)$ versus $[\cdot]$) only.

2.3: Direct approach to finding unit-pulse response

- Worth considering: Can we measure a unit-pulse response directly?

  SPOILER: If the answer were an unqualified "yes," this would be a very short course!

- Let us assume for now that the actual system dynamics are given by

$$y[k] = \sum_{m=1}^{\infty} g_0[m]u[k-m] + v[k],$$

  where $g_0[m]$ is the discrete-time unit-pulse response of the "true" system, and $v[k]$ is a "noise term."

- If we apply a scaled discrete pulse,

$$u[k] = \alpha\delta[k] = \begin{cases} \alpha, & k = 0; \\ 0, & k \neq 0, \end{cases}$$

  the output will be $y[k] = \alpha g_0[k] + v[k]$.

- If the noise level is "low," we can then estimate the unit-pulse response coefficients from $\hat{g}[k] = y[k]/\alpha$.

PROBLEM: Most systems do not allow large enough input pulses such that the errors $v[k]/\alpha$ are small compared to $g_0[k]$.

ATTEMPT AT SOLUTION: A more interesting input is a scaled step function,

$$u[k] = \alpha \cdot 1[k] = \begin{cases} \alpha, & k \geq 0; \\ 0, & k < 0, \end{cases}$$

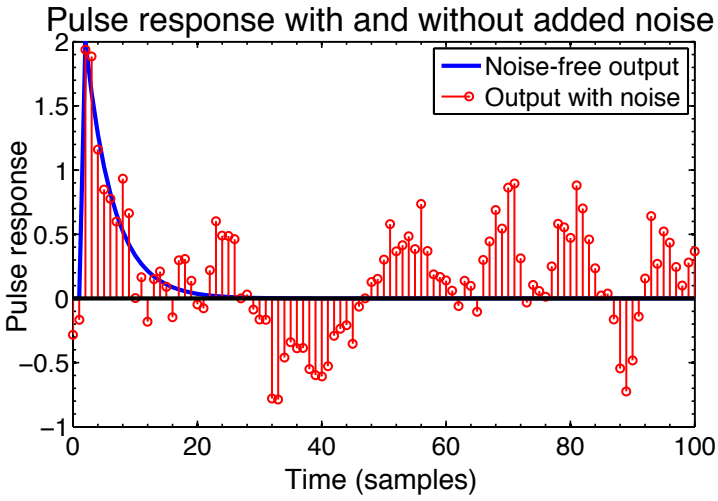
with output $y[k] = \alpha \sum_{m=1}^{k} g_0[m] + v[k]$, and $\hat{g}[k] = (y[k] - y[k-1])/\alpha$.

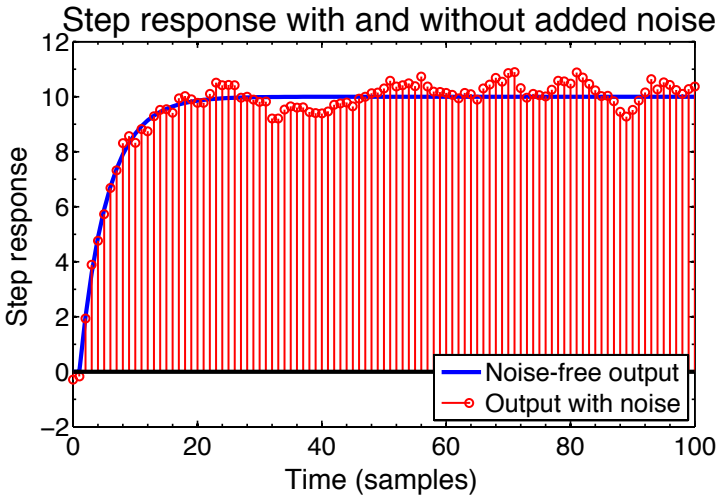
- Example of both approaches given below:
  - Used fairly large amplitude for noise in $v[k]$.
  - Performed step and pulse estimates of the unit-pulse response.
  - Estimates poor even for initial values; get worse for larger k.

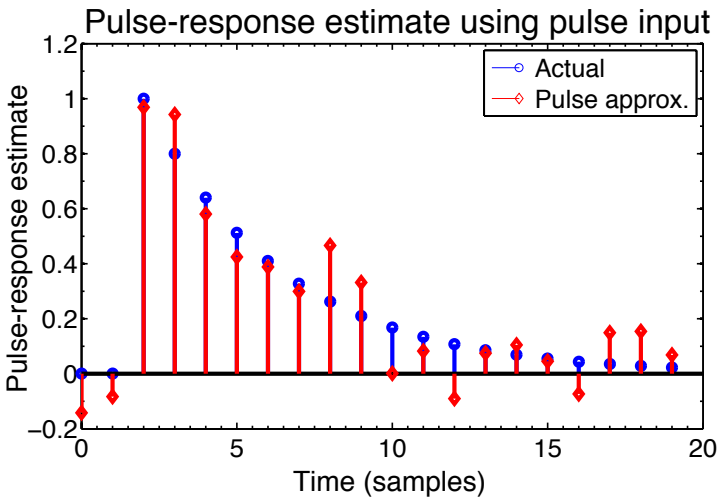
```
% Example from Jonathan How, E211, Stanford University, Fall 1999

% System model in polynomial form
a01 = -0.8; b01 = 1; c01 = 1;
lambda = 0.05; % variance of input noise
N = 512; % number of data points
alpha = 2; % input scaling

% various inputs
upulse = alpha * eye(N,1);
ustep  = alpha * ones(N,1);
err    = sqrt(lambda)*randn(N,1);
th1    = poly2th([1 a01],[0 0 b01],[c01],[1]); % 0 0 = 2 delays

% various responses
yact   = idsim([upulse 0*err],th1); % decays ~ (0.8)^(k-2) for k>=2
ypulse = idsim([upulse err],th1);
ystepact = idsim([ustep 0*err],th1);
ystep  = idsim([ustep err],th1);

% various unit-pulse response estimates
ghat_p = ypulse/alpha;
ghat_s = ([ystep;0]-[0;ystep])/alpha;

% and then some plotting...
k = 0:19; figure(1); clf;
stem(k,yact(k+1)/alpha); hold on; stem(k,ghat_p(k+1),'rd'); hold off
legend('Actual','Pulse approx');
title('Pulse response estimate using pulse input');
ylabel('Pulse resp. est.'); xlabel('Time (samples)');

figure(2); clf;
stem(k,yact(k+1)/alpha); hold on; stem(k,ghat_s(k+1),'r>'); hold off
legend('Actual','Step approx');
title('Pulse reponse estimate using step input');
```

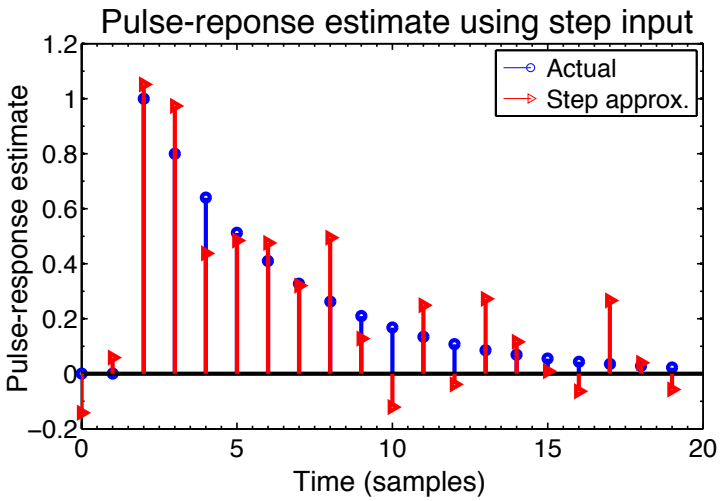


```
ylabel('Pulse resp. est.'); xlabel('Time (samples)');

k = 0:100; figure(3); clf;
plot(k,yact(k+1)); hold on; stem(k,ypulse(k+1),'ro'); hold off
legend('Noise-free output','Output with noise');
title('Pulse response with and without added noise');
ylabel('Pulse response'); xlabel('Time (samples)');

figure(4); clf;
plot(k,ystepact(k+1)); hold on; stem(k,ystep(k+1),'ro'); hold off
legend('Noise-free output','Output with noise','location','southeast');
title('Step response with and without added noise');
ylabel('Step response'); xlabel('Time (samples)');
```









- Neither of these works particularly well, but are a good way to get characteristics like: delay time, static gain, dominant time constant.

Understanding disturbance

- Without noise/disturbance $v[k]$ in the system, we could have determined the unit-pulse response exactly from the measured data.
  - But, that would be unrealistic. It is always present in some form.
- The term “noise” is most often used to denote a disturbance whose time samples are completely independent from one another.
- Disturbance includes “completely random” noise, but also correlated noise signals and deterministic unmeasured signals (like sinusoids).
- Disturbance arises from many causes, with two main categories

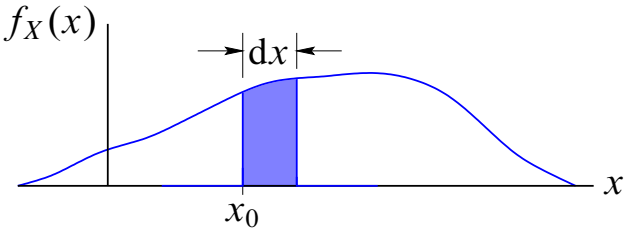
  MEASUREMENT NOISE: Sensor thermo-electric measurement noise, nonlinearity, and drift; quantization error of sampling.

  UNCONTROLLABLE INPUTS: System is subject to signals like input signals, not controlled by user (and, usually not measurable).

2.4: Scalar random variables

- The most characteristic feature of a disturbance *is that its value is neither known beforehand nor measured*. But, can often describe some of its properties (*e.g.*, its mean, variation, frequency content).
- To do so, we often treat $v[k]$ as a random process, or a sequence in time of random variables (RVs).
  - Cannot predict exactly what we will get each time we measure or sample the random variable, but
  - We can characterize the probability of each sample value by the "probability density function" (PDF).

Probability density functions (PDFs)



- $f_X(x_0)\,\mathrm{d}x$ is the probability that random variable X is between $[x_0, x_0 + \mathrm{d}x]$.

- Properties that are true for every PDF:

1. $f_X(x) \geq 0 \quad \forall \quad x.$
2. $\displaystyle\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1.$
3. $\Pr(X \leq x_0) = \displaystyle\int_{-\infty}^{x_0} f_X(x)\,\mathrm{d}x \triangleq F_X(x_0).$

- Problem: Apart from simple examples it is often difficult to determine $f_X(x)$ accurately. ➠ Use approximations to capture the key behavior.
- Need to define key characteristics of $f_X(x)$.

STATISTICAL AVERAGE (MEANS): Consider a (discrete) RV X which can assume n values $x_1, x_2 \ldots x_n$. Make many measurements, $N \to \infty$.

- Let m_i be the number of times the value of the measurement is x_i. Define a statistical average/mean as,

$$\bar{x} = \frac{1}{N}(m_1 x_1 + m_2 x_2 + \cdots + m_n x_n) = \frac{m_1}{N}x_1 + \frac{m_2}{N}x_2 + \cdots + \frac{m_n}{N}x_n.$$

- In the limit, $m_1/N \to \Pr(X = x_1)$ and so forth (assuming ergodicity, defined later), so we have

$$\bar{x} = \sum_{i=1}^{n} x_i \Pr(X = x_i).$$

- For continuous RVs, in the limiting ergodic case, we get

$$\bar{x} = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x.$$

EXPECTATION: A more rigorous average than the statistical average.

- Expectation describes average behavior of system:

$$\bar{x} = \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x,$$

  whether or not the process is ergodic.

- Note: Expectation is a linear operator. So, for example, the first moment about the mean: $\mathbb{E}[X - \bar{x}] = 0$.

VARIANCE: Second (central) moment about the mean:

$$\mathrm{var}(X) = \mathbb{E}[(X - \bar{x})^2] = \int_{-\infty}^{\infty} (x - \bar{x})^2 f_X(x)\,\mathrm{d}x = \mathbb{E}[X^2] - (\mathbb{E}[X])^2,$$

or is equal to the mean-square minus the square-mean.

STANDARD DEVIATION: Measure of dispersion about the mean of the samples of X, $\sigma_X = \sqrt{\mathrm{var}(X)}$.

- The expectation and variance capture key features of the actual PDF. Higher-order moments are available, but *we won't need them!*

KEY POINT #1: We can measure various samples X_i of X and develop approximations of σ_X, $\bar{x}$.

$$\bar{x} \approx \frac{1}{N}\sum_{i=1}^{N} X_i \qquad \sigma_X^2 \approx \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{x})^2.$$

KEY POINT #2: Understanding variance via Chebychev's inequality

- Chebychev's inequality states (for positive ε)

$$\Pr(|X - \bar{x}| \geq \varepsilon) \leq \frac{\sigma_X^2}{\varepsilon^2}.$$

- It may be proven as follows:

$$\Pr(|X - \bar{x}| \geq \varepsilon) = \int_{-\infty}^{\bar{x}-\varepsilon} f_X(x)\,\mathrm{d}x + \int_{\bar{x}+\varepsilon}^{\infty} f_X(x)\,\mathrm{d}x.$$

- For the two regions of integration $|x - \bar{x}|/\varepsilon \geq 1$ or $(x-\bar{x})^2/\varepsilon^2 \geq 1$. So,

$$\Pr(|X - \bar{x}| \geq \varepsilon) \leq \int_{-\infty}^{\bar{x}-\varepsilon} \frac{(x-\bar{x})^2}{\varepsilon^2} f_X(x)\,\mathrm{d}x + \int_{\bar{x}+\varepsilon}^{\infty} \frac{(x-\bar{x})^2}{\varepsilon^2} f_X(x)\,\mathrm{d}x.$$

- Since $f_X(x)$ is positive, then we also have

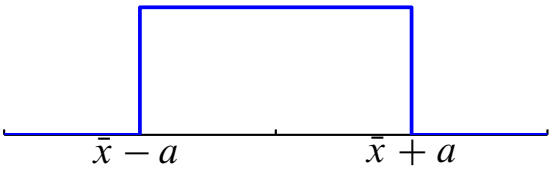
$$\Pr(|X - \bar{x}| \geq \varepsilon) \leq \int_{-\infty}^{\infty} \frac{(x-\bar{x})^2}{\varepsilon^2} f_X(x)\,\mathrm{d}x = \frac{\sigma_X^2}{\varepsilon^2}.$$

- This inequality shows that probability is clustered around the mean, and variance indicates the dispersion of the probability distribution.

Two important probability distributions

- Generally don't know true probability distribution of noise/disturbance.
- Will use the following two to approximate many others.

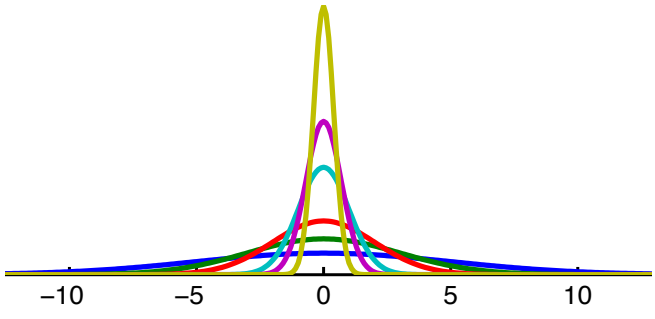
UNIFORM: All possible outcomes equally likely.



$$f_X(x) = \begin{cases} \dfrac{1}{2a}, & |x-\bar{x}| \le a; \\ 0, & \text{otherwise.} \end{cases}$$

- $\mathbb{E}[X] = \bar{x}$. Variance $\sigma_X^2 = a^2/3$. Notation: $X \sim U(\bar{x}-a, \bar{x}+a)$.
- Useful probabilistic model if
  - Entire set of values equally likely, or
  - If you have *no idea* what $f_X(x)$ is.

GAUSSIAN: (Normal)



$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma_X^2}\right).$$

- Symmetric about $\bar{x}$. Peak proportional to $1/\sigma_X$ at $\bar{x}$.
- Notation: $X \sim \mathcal{N}(\bar{x}, \sigma_X^2)$. Some important features:
  - Probability that X within $\pm\sigma_X$ of $\bar{x}$ is 68%.
  - Probability that X within $\pm 2\sigma_X$ of $\bar{x}$ is 96%.
  - Probability that X within $\pm 3\sigma_X$ of $\bar{x}$ is 99.7%.
  - A "$3\sigma_X$" value almost certainly covers observed samples.
- "Narrow" distribution ➠ Sharp peak. High confidence in predicting X.
- "Wide" distribution ➠ Poor knowledge in what to expect for X.

2.5: Vector random variables

- With very little change in the preceding, we can also handle vectors of random variables.

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}. \qquad \text{Let } x_o = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

- Behavior of X described by (scalar) joint PDF $f_X(x)$.
  - $f_X(x_o)$ means $f_X(X_1 = x_1, X_2 = x_2 \cdots X_n = x_n)$.
  - That is, $f_X(x_o)\,\mathrm{d}x_1\,\mathrm{d}x_2 \cdots \mathrm{d}x_n$ is the probability that X lies in the hypercube between vertexes x_o and $x_o + \mathrm{d}x$.
- Properties:
  1. $f_X(x) \geq 0 \quad \forall \quad x$. Same as before.
  2. $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_n = 1$. Basically the same.
  3. $\bar{x} = \mathbb{E}[X] = \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_n$. Basically same.
  4. *Co*variance: Different.

$$\begin{aligned}\Sigma_{\widetilde{X}} &= \mathbb{E}[(X-\bar{x})(X-\bar{x})^T] \qquad \text{(outer product)} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(x-\bar{x})(x-\bar{x})^T f_X(x)\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_n.\end{aligned}$$

$\Sigma_{\widetilde{X}}$ is symmetric and positive-semi-definite (psd). This means

$$y^T \Sigma_{\widetilde{X}} y \geq 0 \quad \forall \quad y.$$

PROOF: For all $y \neq 0$,

$$0 \leq \mathbb{E}[(y^T(X-\bar{x}))^2]$$

$$= y^T \mathbb{E}[(X - \bar{x})(X - \bar{x})^T] y$$

$$= y^T \Sigma_{\widetilde{X}} y.$$

5. Correlation: Different.

$$\Sigma_X = \mathbb{E}[XX^T]$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x x^T f_X(x) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \cdots \mathrm{d}x_n.$$

- Notice that correlation Σ_X and covariance $\Sigma_{\widetilde{X}}$ are the same for zero-mean random vectors.
- Also notice function of elements of covariance matrix:

$$(\Sigma_{\widetilde{X}})_{ii} = \sigma_{X_i}^2$$

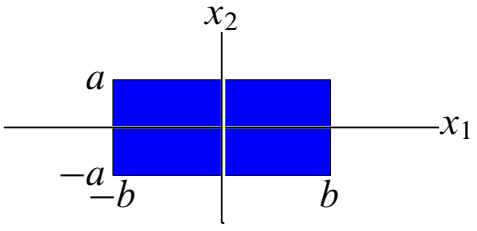
$$(\Sigma_{\widetilde{X}})_{ij} = \rho_{ij} \sigma_{X_i} \sigma_{X_j} = (\Sigma_{\widetilde{X}})_{ji}.$$

Then, ρ_{ij} is called a "correlation coefficient" and is a measure of linear dependence between X_i and X_j. $|\rho_{ij}| \leq 1$.

Two important multivariate distributions

- As before, we focus on two main distributions to approximate many others.

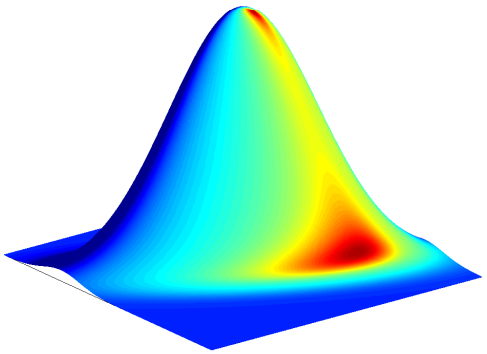
MULTIVARIABLE UNIFORM:



$$S = \{|x_1| \leq b \quad \text{and} \quad |x_2| \leq a\}$$

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} \dfrac{1}{4ab}, & X \in S; \\ 0, & \text{otherwise.} \end{cases}$$

MULTIVARIABLE GAUSSIAN:



- Notation: $X \sim \mathcal{N}(\bar{x}, \Sigma_{\widetilde{X}})$.
- Contours of constant $f_X(x)$ are hyper-ellipsoids, centered at $\bar{x}$, directions governed by $\Sigma_{\widetilde{X}}$. Principle axes decouple $\Sigma_{\widetilde{X}}$ (eigenvectors).

$$f_X(x) = \frac{1}{(2\pi)^{n/2}\,|\Sigma_{\widetilde{X}}|^{1/2}} \exp\left(-\frac{1}{2}(x-\bar{x})^T \Sigma_{\widetilde{X}}^{-1}(x-\bar{x})\right).$$

$$|\Sigma_{\widetilde{X}}| = \det(\Sigma_{\widetilde{X}}), \quad \Sigma_{\widetilde{X}}^{-1} \text{ requires positive-definite } \Sigma_{\widetilde{X}}.$$

- Two-dimensional zero-mean case: (Let $\sigma_1 = \sigma_{X_1}$ and $\sigma_2 = \sigma_{X_2}$)

$$\Sigma_{\widetilde{X}} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \qquad |\Sigma_{\widetilde{X}}| = \sigma_1^2\sigma_2^2(1-\rho_{12}^2).$$

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \exp\left(-\frac{\frac{x_1^2}{\sigma_1^2} - 2\rho_{12}\frac{x_1}{\sigma_1}\frac{x_2}{\sigma_2} + \frac{x_2^2}{\sigma_2^2}}{2(1-\rho_{12}^2)}\right).$$

- Constant-probability contours defined by

$$\frac{x_1^2}{\sigma_1^2} - 2\rho_{12}\frac{x_1}{\sigma_1}\frac{x_2}{\sigma_2} + \frac{x_2^2}{\sigma_2^2} = k^2.$$

2.6: Properties of jointly-distributed random variables

INDEPENDENCE: Iff jointly-distributed RVs are independent, then

$$f_X(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n).$$

- Joint distribution can be split up into the product of individual distributions for each RV.
- Equivalent condition: For *ALL* functions $f(\cdot)$ and $g(\cdot)$,

$$\mathbb{E}[f(X_1)g(X_2)] = \mathbb{E}[f(X_1)]\mathbb{E}[g(X_2)].$$

- "The particular value of the random variable X_1 has no impact on what value we would obtain for the random variable X_2."

UNCORRELATED: Two jointly-distributed RVs X_1 and X_2 are uncorrelated if their second moments are finite and

$$\text{cov}(X_1, X_2) = \mathbb{E}[(X_1 - \bar{x}_1)(X_2 - \bar{x}_2)] = 0,$$

which implies $\rho_{12} = 0$.

MAIN POINT #1: If jointly-distributed RV X_1 and X_2 are independent then they are uncorrelated. Independence implies uncorrelation.

- To see this, notice that independence means $\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2]$. Therefore,

$$\text{cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = 0,$$

therefore uncorrelated.

- Does uncorrelation imply independence? (No.) Consider the example

$$Y_1 = \sin(2\pi X)$$

$$Y_2 = \cos(2\pi X)$$

with X uniformly distributed on $[0, 1]$.

- We can show that $\mathbb{E}[Y_1] = \mathbb{E}[Y_2] = \mathbb{E}[Y_1 Y_2] = 0$.
- So, $\text{cov}(Y_1, Y_2) = 0$ and Y_1 and Y_2 are uncorrelated.
- But, $Y_1^2 + Y_2^2 = 1$ and therefore Y_1 and Y_2 are clearly not independent.
- Note: $\text{cov}(X_1, X_2) = 0$ (uncorrelated) implies that

$$\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2]$$

but independence requires

$$\mathbb{E}[f(X_1)g(X_2)] = \mathbb{E}[f(X_1)]\mathbb{E}[g(X_2)]$$

for all $f(\cdot)$ and $g(\cdot)$.

- Therefore, independence is *MUCH* stronger than uncorrelated.

COROLLARY: Consider RV X having independent components. The covariance matrix $\Sigma_{\widetilde{X}} = \mathbb{E}[(X - \bar{x})(X - \bar{x})^T]$ is diagonal.

PROOF: Notice that:

- The diagonal elements are $(\Sigma_{\widetilde{X}})_{ii} = \mathbb{E}[(X_i - \bar{x}_i)^2] = \sigma_i^2$.
- The off-diagonal elements are

$$\begin{aligned}(\Sigma_{\widetilde{X}})_{ij} &= \mathbb{E}[(X_i - \bar{x}_i)(X_j - \bar{x}_j)] \\ &= \mathbb{E}[X_i X_j - X_i \bar{x}_j - \bar{x}_i X_j + \bar{x}_i \bar{x}_j] \\ &= \mathbb{E}[X_i]\mathbb{E}[X_j] - 2\mathbb{E}[X_i]\mathbb{E}[X_j] + \mathbb{E}[X_i]\mathbb{E}[X_j] \\ &= 0.\end{aligned}$$

MAIN POINT #2: If jointly *NORMALLY* distributed RVs are uncorrelated, then they are independent. This is a special case.

MAIN POINT #3: Suppose we are given two random variables Y and X with $Y = g(X)$. Also assume that g^{-1} exists, g and g^{-1} are

continuously differentiable, then

$$f_Y(y) = f_X(g^{-1}(y)) \left\| \left| \frac{\partial g^{-1}(y)}{\partial y} \right| \right\|,$$

where $\|\|\cdot\|\|$ means to take the absolute value of the determinant.

EXAMPLE: $Y = AX + B$ where A is a constant (non-singular) matrix, B is a constant vector, and $X \sim \mathcal{N}(\bar{x}, \Sigma_{\widetilde{X}})$.

- $X = A^{-1}Y - A^{-1}B$ so $g^{-1}(y) = A^{-1}(y - B)$. Then, $\dfrac{\partial g^{-1}(y)}{\partial y} = A^{-1}$.
- Also: $f_X(x) = \dfrac{1}{(2\pi)^{n/2}|\Sigma_{\widetilde{X}}|^{1/2}} \exp\left[-\dfrac{1}{2}(x - \bar{x})^T \Sigma_{\widetilde{X}}^{-1}(x - \bar{x})\right]$. So,

$$\begin{aligned} f_Y(y) &= \frac{|A^{-1}|}{(2\pi)^{n/2}|\Sigma_{\widetilde{X}}|^{1/2}} \exp\left[-\frac{1}{2}(A^{-1}(y - B) - \bar{x})^T \Sigma_{\widetilde{X}}^{-1}(A^{-1}(y - B) - \bar{x})\right] \\ &= \frac{1}{(2\pi)^{n/2}(|A||\Sigma_{\widetilde{X}}||A^T|)^{1/2}} \exp\left[-\frac{1}{2}(y - \bar{y})^T (A^{-1})^T \Sigma_{\widetilde{X}}^{-1} A^{-1}(y - \bar{y})\right] \\ &= \frac{1}{(2\pi)^{n/2}|\Sigma_{\widetilde{Y}}|^{1/2}} \exp\left[-\frac{1}{2}(y - \bar{y})\Sigma_{\widetilde{Y}}^{-1}(y - \bar{y})\right], \end{aligned}$$

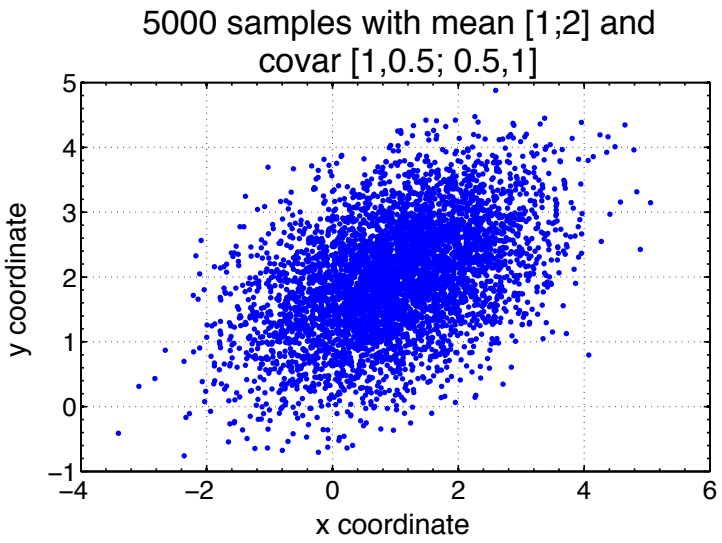
if $\Sigma_{\widetilde{Y}} = A\Sigma_{\widetilde{X}}A^T$ and $\bar{y} = A\bar{x} + B$.

CONCLUSION: Sum of Gaussians is Gaussian—very special case.

RANDOM NOTE: How to use `randn.m` to simulate non-zero mean Gaussian noise with covariance $\Sigma_{\widetilde{Y}}$?

- $Y \sim \mathcal{N}(\bar{y}, \Sigma_{\widetilde{Y}})$ but `randn.m` returns $X \sim \mathcal{N}(0, I)$.
- Try $y = \bar{y} + A^T x$ where A is square with the same dimension as $\Sigma_{\widetilde{Y}}$; $A^T A = \Sigma_{\widetilde{Y}}$. ($A$ is the Cholesky decomposition of positive-definite symmetric matrix $\Sigma_{\widetilde{Y}}$).

```
ybar = [1; 2];
covar = [1, 0.5; 0.5, 1];
A = chol(covar);
x = randn([2, 1]);
y = ybar + A'*x;
```



- When $\Sigma_{\widetilde{Y}}$ is non-positive definite (but also non-negative definite)

```
[L,D] = ldl(covar);
x = randn([2,5000]);
y = ybar(:,ones([1 5000]))
  +(L*sqrt(D))*x;
```

MAIN POINT #4: Conditional probabilities.

- Given jointly-distributed RVs, it is often of extreme interest to find the PDF of some of the RVs given known values for the rest.
- For example, given the joint PDF $f_{X,Y}(x,y)$ for RVs X and Y, we want the conditional PDF of $f_{X|Y}(x|y)$ which is the PDF of X for a known value $Y = y$. Can also think of it as $f_{X|Y}(X = x|Y = y)$.

DEFINE: Conditional PDF

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

is the probability that $X = x$ given that $Y = y$ has happened.

NOTE I: The marginal probability $f_Y(y)$ may be calculated as

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x.$$

For each y, integrate out the effect of X.

NOTE II: If X, Y independent, $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. Therefore

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

Knowing that $Y = y$ has occurred provides *NO* information about X. (Is this what you would expect?)

DIRECT EXTENSION:

$$\begin{aligned} f_{X,Y}(x,y) &= f_{X|Y}(x|y)\,f_Y(y) \\ &= f_{Y|X}(y|x)\,f_X(x), \end{aligned}$$

Therefore,

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

This is known as *Bayes' rule.* It relates the *a posteriori* probability to the *a priori* probability. Very important for analyzing communication systems, probably of limited use to us.

MAIN POINT #5: Conditional expectation.

- Now that we have a way of expressing a conditional probability, we can also express a conditional mean.
- What do we expect the value of X to be given that $Y = y$ has happened?

$$\mathbb{E}[X = x|Y = y] = \mathbb{E}[X|Y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|Y)\,\mathrm{d}x.$$

- Conditional expectation is not a constant (as expectation is) but a random variable. It is a function of the conditioning random variable (*i.e.,* Y).
- A sometimes-useful tool is iterated expectation, $\bar{x} = \mathbb{E}_Y[\mathbb{E}[X|Y]]$.

2.7: Vector random (stochastic) processes

- A stochastic random process is a family of random scalars or vectors indexed by a parameter set ("time" in our case).
  - For simplicity of notation, will now adopt lowercase names for random signals. For example, the random process $u[k]$.
  - The value of the random process at any time k_o is a random variable $u[k_o]$.

Properties

- Define correlation (or cross-correlation) between two signals to be,

$$R_{wx}[k,m] \triangleq \mathbb{E}\big[w[k]x^T[m]\big].$$

- Define the auto-correlation of a signal to be $R_u[k,m] \triangleq \mathbb{E}\big[u[k]u^T[m]\big]$.
  - Provides a measure of correlation between elements of the process two different time points.
- Key feature of a random process $x[k]$ is whether it is "wide-sense stationary" (WSS). Requires
  - Mean is time-invariant: $\mathbb{E}\big[x[k]\big] = \bar{x}[k] = \bar{x} < \infty \quad \forall\, k.$
  - Autocorrelation is also time-invariant:

$$\mathbb{E}\big[x[k]x[k-\tau]^T\big] = R_x[\tau] < \infty \quad \forall\, k, \tau.$$

  - $R_u[0]$ is always the maximum value of $R_u[\tau]$. Also, $R_u[0] = \sigma_u^2$ for zero-mean u.
- There are no truly WSS signals, since all signals "turn on" at some point. However, often good enough approximation.

  - If two signals are WSS, then the cross-correlation is also not a function of absolute time, $R_{wx}[\tau] \triangleq \mathbb{E}\big[w[k]x^T[k-\tau]\big]$.

- Also define the covariance (or cross-covariance) between two signals:

$$C_{wx}[k,m] \triangleq \mathbb{E}\big[(w[k]-\bar{w}[k])(x[m]-\bar{x}[m])^T\big].$$

  - If the signals are wide-sense stationary stationary, then

$$C_{wx}[\tau] \triangleq \mathbb{E}\big[(w[k]-\bar{w}[k])(x[k-\tau]-\bar{x}[k-\tau])^T\big].$$

- If both signals are zero mean, then $C_{wx}[\tau] = R_{wx}[\tau]$.
- Define the auto-covariance of a signal to be

$$C_u[k,m] \triangleq \mathbb{E}\big[(u[k]-\bar{u}[k])(u[m]-\bar{u}[m])^T\big].$$

  If the signal is wide-sense stationary then,

$$C_u[\tau] = \mathbb{E}[(u[k]-\bar{u})(u[k-\tau]-\bar{u})^T].$$

NOTE: The text is sloppy, treating correlation and covariance as identical quantities. This is because it assumes that all signals are zero mean, which makes sense in the context of system ID, but not in general.

"White" Noise

- Correlation from one time instant to the next is a key property of a random process. ➠ $R_x[\tau]$ and $C_x[\tau]$ very important.
- Some processes have a unique autocorrelation:
  1. Zero mean,
  2. $R_x[\tau] = \sigma^2\delta[\tau]$, where $\delta[\tau]$ is the Dirac delta, $\delta[\tau] = \begin{cases} 1, & \tau = 0; \\ 0, & \tau \neq 0. \end{cases}$
- Therefore, the process is uncorrelated in time.
- Clearly an *ABSTRACTION*, but proves to be a *VERY* useful one.

Power spectral density (PSD)

- Consider a wide-sense stationary stochastic process $x[k]$.
  - If the process varies slowly, then adjacent time samples will be highly correlated, and $R_x[\tau]$ will drop off slowly ➠ Mostly low-frequency content.
  - If the process varies quickly, then adjacent time samples will not be so highly correlated, and $R_x[\tau]$ will drop off quickly ➠ More high-frequency content.
- Therefore, the autocorrelation function tells us about the frequency content of the random signal.
- Consider scalar case. We define the power spectral density (text calls this the "spectrum")

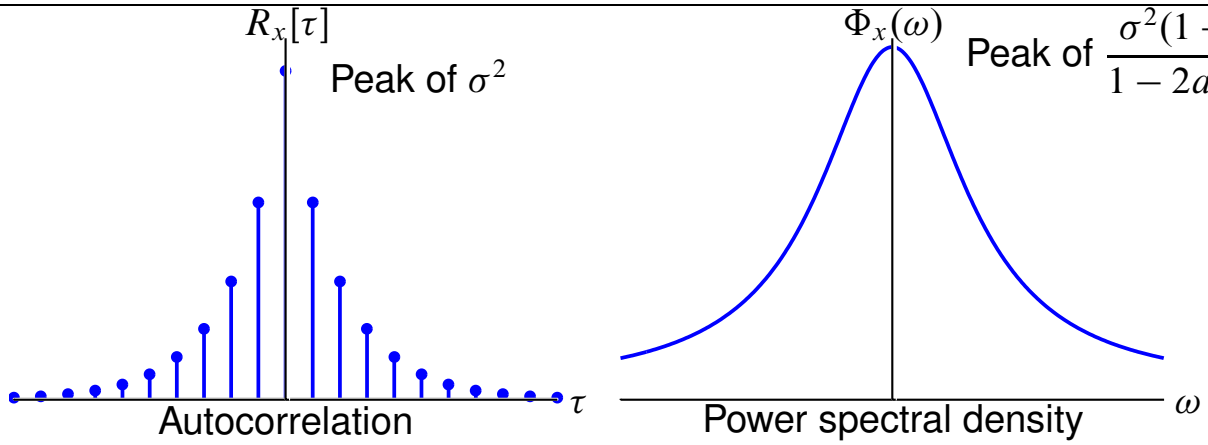
$$\Phi_x(\omega) = \sum_{\tau=-\infty}^{\infty} R_x[\tau] e^{-j\omega\tau}$$

$$R_x[\tau] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_x(\omega) e^{j\omega\tau} \, \mathrm{d}\omega.$$

- $\Phi_x(\omega)$ gives a frequency-domain interpretation of the "power" in the random process $x[k]$.
- $\Phi_x(\omega)$ is real, symmetric and positive definite for real random processes.

EXAMPLE: Consider a random process with $R_x[\tau] = \sigma^2 a^{|\tau|}$ for $0 < a < 1$.

- We can show, $\Phi_x(\omega) = \sum_{\tau=-\infty}^{\infty} \sigma^2 a^{|\tau|} e^{-j\omega\tau} = \dfrac{\sigma^2(1-a^2)}{1+a^2-2a\cos(\omega)}$.



- a is a bandwidth measure.

INTERPRETATION:

1. Area under PSD $\Phi_x(\omega)$ for $[\omega_1, \omega_2]$ provides a measure of energy in the signal in that frequency range.
2. White noise can be thought of as a limiting process as $a \to 0$.

- As $a \to 0$ $R_x[\tau] \to \sigma^2\delta[\tau]$; therefore white. $\Phi_x(\omega) = \sigma^2$ for all ω.

2.8: Back to system identification

- We will treat $v[k]$ as a stochastic process. Ideally, would like to specify exactly how the RVs at different times are related.
- Would require a joint probability density function between $v[k]$ and $v[k+m]$ $\forall$ $m \geq 1$.
  - Not really practical and/or feasible.
  - Will make several approximations instead.

APPROXIMATION I: Assume that we need only specify the first- and second-order properties of $\{v[k]\}$ (*i.e.*, its mean and variance).

- For a Gaussian process, this is all we need (completely specified by mean/covariance).
- Relationships between samples described by correlation and covariance functions.

APPROXIMATION II: $v[k]$ is the output of a (linear) filtering process

$$v[k] = \sum_{m=0}^{\infty} h[m]e[k-m],$$

where the sequence $\{e[k]\}$ is assumed to be white noise.

- Because the disturbance is modeled as a linear combination of white-noise inputs, it also has zero mean

$$\bar{v} = \mathbb{E}\big[v[k]\big] = \sum_{m=0}^{\infty} h[m]\mathbb{E}\big[e[k-m]\big] = 0.$$

- We can also find its autocovariance function,

$$R_v[\tau] = \mathbb{E}\big[v[k]v[k-\tau]\big] = \mathbb{E}\left[\sum_{n=0}^{\infty}\sum_{l=0}^{\infty} h[n]h[l]e[k-n]e[k-\tau-l]\right]$$

$$= \sigma^2 \sum_{n=0}^{\infty} h[n]h[n-\tau].$$

- Since $\bar{v}$ and $R_v[\tau]$ don't depend on absolute time k, $v[k]$ is WSS.
- Can use filter $h[k]$ to "shape" the autocovariance of v.

Correlation approach to system ID

- With this added knowledge, can try performing system ID with an even more "exciting" input than before (*i.e.*, one that has more frequency content, and "excites" different "modes" of the system).
- Assume $u[k]$ is white noise:
  - White noise is a "good" input signal for system ID of linear systems because it has (even distribution of) power at all input frequencies.
- The output from our system will be the result of the convolution

$$y[k] = \sum_{m=1}^{\infty} g[m]u[k-m] + v[k].$$

- We will use a correlation approach to de-tangle the convolution, and come up with an estimate $\hat{g}[m]$ of $g[m]$.
- The cross correlation between the (possibly vector) input and output signals is

$$\begin{aligned} R_{yu}[\tau] &= \mathbb{E}\left[y[k]u[k-\tau]^T\right] \\ &= \mathbb{E}\left[\sum_{m=1}^{\infty} g[m]u[k-m]u^T[k-\tau] + v[k]u^T[k-\tau]\right] \\ &= \sum_{m=1}^{\infty} g[m] \underbrace{\mathbb{E}\left[u[k-m]u^T[k-\tau]\right]}_{R_u[\tau-m]} + \underbrace{\mathbb{E}\left[v[k]u^T[k-\tau]\right]}_{0} \end{aligned}$$

$$= \sum_{m=1}^{\infty} g[m] R_u[\tau - m].$$

- When the input signal is white, this reduces to $R_{yu}[\tau] = \sigma^2 g[\tau]$.
- Correlation helps eliminate the impact of noise (it "averages it out").
- But, we need to find $R_{yu}[\tau]$ to use this method to find $g[\tau]$.

ISSUE: $R_{yu}[\tau]$ is based on statistical expectation (ensemble averaging). This is not something we can obtain from data, unless we simultaneously conduct an infinite number of experiments.

- Will assume that the underlying processes are ergodic.
  - *i.e,* all possible "ensemble" averages are equal to time averages.
  - Key property to assume since it allows us to average over time rather than having to perform many experiments and average (at a given time) over the ensemble (more work!)
  - Then, we can approximate auto- and cross-correlations as

$$\widehat{R}_v^N[\tau] = \frac{1}{N} \sum_{k=0}^{N-1} v[k] v[k - \tau]$$

$$\widehat{R}_{wv}^N[\tau] = \frac{1}{N} \sum_{k=0}^{N-1} w[k] v[k - \tau].$$

  - These are called "sample estimates" of the correlation functions.
- Back to our problem: We need to find $R_{yu}[\tau]$ to use the correlation method to find $g[\tau]$.
- We assume ergodicity. Then, we can approximate

$$\widehat{R}_u^N[\tau] = \frac{1}{N}\sum_{k=0}^{N-1} u[k]u[k-\tau]$$

$$\widehat{R}_{yu}^N[\tau] = \frac{1}{N}\sum_{k=0}^{N-1} y[k]u[k-\tau].$$

- If $u[k]$ is white, I need only to compute $\widehat{R}_{yu}^N[\tau]$. However, if $u[k]$ is not white, I can still use the approach by recognizing

$$R_{yu}[\tau] = \sum_{m=1}^{M} g[m]R_u[\tau-m]$$

$$\text{so, } \widehat{R}_{yu}^N[\tau] \approx \sum_{m=1}^{M} g[m]\widehat{R}_u^N[\tau-m].$$

- Writing these M summation terms as a matrix formula

$$\left[\widehat{R}_{yu}^N[\tau]\right] = \left[\widehat{R}_u^N[\tau-m]\right]\left[g[m]\right],$$

and solving for $g[m]$. (Reordering assumes scalar; can also solve if matrix.)

- For example, consider a scenario where we have $\widehat{R}_{yu}[\tau]$ for $0 \le \tau \le 3$ and $\widehat{R}_u[\tau]$ for $0 \le \tau \le 3$.

$$R_{yu}[0] = \sum_{m=0}^{3} g[m]R_u[0-m] = g[0]R_u[0] + g[1]\underbrace{R_u[-1]}_{=R_u[1]} + \cdots + g[3]\underbrace{R_u[-3]}_{=R_u[3]}$$

$$R_{yu}[1] = \sum_{m=0}^{3} g[m]R_u[1-m] = g[0]R_u[1] + g[1]R_u[0] + \cdots + g[3]\underbrace{R_u[-2]}_{=R_u[2]}$$

$$R_{yu}[2] = \sum_{m=0}^{3} g[m]R_u[2-m] = g[0]R_u[2] + g[1]R_u[1] + \cdots + g[3]\underbrace{R_u[-1]}_{=R_u[1]}$$

$$R_{yu}[3] = \sum_{m=0}^{3} g[m]R_u[3-m] = g[0]R_u[3] + g[1]R_u[2] + \cdots + g[3]R_u[0].$$

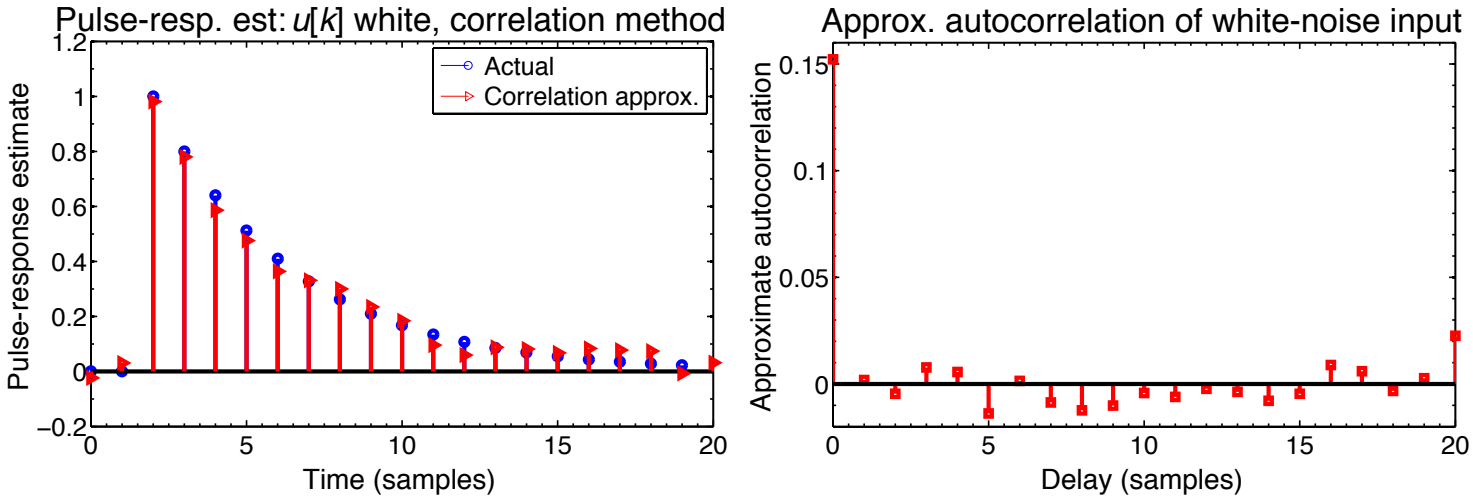
- We re-arrange this, substituting approximate values, to get

$$\begin{bmatrix} \widehat{R}_{yu}[0] \\ \widehat{R}_{yu}[1] \\ \widehat{R}_{yu}[2] \\ \widehat{R}_{yu}[3] \end{bmatrix} = \begin{bmatrix} \widehat{R}_u[0] & \widehat{R}_u[1] & \widehat{R}_u[2] & \widehat{R}_u[3] \\ \widehat{R}_u[1] & \widehat{R}_u[0] & \widehat{R}_u[1] & \widehat{R}_u[2] \\ \widehat{R}_u[2] & \widehat{R}_u[1] & \widehat{R}_u[0] & \widehat{R}_u[1] \\ \widehat{R}_u[3] & \widehat{R}_u[2] & \widehat{R}_u[1] & \widehat{R}_u[0] \end{bmatrix} \begin{bmatrix} g[0] \\ g[1] \\ g[2] \\ g[3] \end{bmatrix}$$

$$\begin{bmatrix} g[0] \\ g[1] \\ g[2] \\ g[3] \end{bmatrix} = \begin{bmatrix} \widehat{R}_u[0] & \widehat{R}_u[1] & \widehat{R}_u[2] & \widehat{R}_u[3] \\ \widehat{R}_u[1] & \widehat{R}_u[0] & \widehat{R}_u[1] & \widehat{R}_u[2] \\ \widehat{R}_u[2] & \widehat{R}_u[1] & \widehat{R}_u[0] & \widehat{R}_u[1] \\ \widehat{R}_u[3] & \widehat{R}_u[2] & \widehat{R}_u[1] & \widehat{R}_u[0] \end{bmatrix}^{-1} \begin{bmatrix} \widehat{R}_{yu}[0] \\ \widehat{R}_{yu}[1] \\ \widehat{R}_{yu}[2] \\ \widehat{R}_{yu}[3] \end{bmatrix}.$$

- Note that the autocorrelation matrix is a “Toeplitz” matrix, easily created in MATLAB using `toeplitz.m`.
- Matrix inversion is best solved as `g = Ru \ Ryu` in MATLAB.

```
% ...prior example continued... correlation analysis
uwhite = alpha/5*randn(N,1); % white noise input sequence
ywhite = idsim([uwhite err],th1);

[Ryu,lagsyu] = xcorr(ywhite,uwhite,'biased'); indyu = find(lagsyu == 0);
[Ruu,lagsuu] = xcorr(uwhite,uwhite,'biased'); induu = find(lagsuu == 0);
RuMat = toeplitz(Ruu(induu:end));
ghat = RuMat\Ryu(induu:end);

k = lagsyu(indyu:end); figure(5); clf;
stem(0:19,yact(1:20)/alpha); hold on; stem(k,ghat,'r>');
legend('Actual','Correlation approx');
title('Pulse resp. est: \itu\rm[\itk\rm] white, correlation method');
ylabel('Pulse-response estimate'); xlabel('Time (samples)');

figure(6); clf; stem(k,Ruu(induu:end),'rs');
title('Approx. autocorrelation of white-noise input');
h = ylabel('Approximate autocorrelation'); xlabel('Delay (samples)');
```

- Results shown for white $u[k]$ input
  - Tried to select $u[k]$ variance σ^2 so noise $v[k]$ was still a factor.
  - $\hat{g}[m]$ much better via this analysis.
- There are still problems with this approach, but the correlation analysis seems to be the best result of the three we've tried so far.

Where from here?

- We have found that it is possible to estimate a system's unit-pulse response, but the methods we've seen so far are pretty poor.
- We also know that an LTI system's frequency response contains all the dynamic information of the system.
  - Maybe we can gain insight by looking at the frequency domain.
  - Maybe it's valuable to see how to estimate a frequency response.
- So, that's our next step.