

MIT OpenCourseWare
<http://ocw.mit.edu>

HST.582J / 6.555J / 16.456J Biomedical Signal and Image Processing
Spring 2007

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Random Signals (statistics and signal processing)

Primary Concepts for Random Processes

You should understand...

- Random processes as a straightforward extension of random variables.
- What is meant by a *realization* and an *ensemble*.
- The importance of *stationarity* and *ergodicity*
 - Useful for estimating statistical properties of random processes.
- Some idea of how the autocorrelation/autocovariance functions describe the statistical structure of a random signal.

Random Processes

- A random process, X , is an infinite-dimensional random variable.
 - If the size (dimension) is *countably* infinite then we have a **discrete-time** random process.
 - If the size (dimension) is *uncountably* infinite then we have a **continuous-time** random process.
- As with any multi-dimensional random variable we can
 - compute marginal and conditional statistics/densities over subsets of the dimensions.
- Random processes are interesting/tractable when there is some **structure** to the statistical relationships between various dimensions (i.e. the structure lends itself to analysis).
- We'll focus primarily on second-order statistical properties.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

3

Some Notation

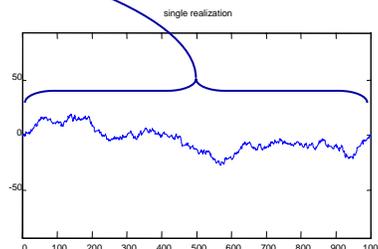
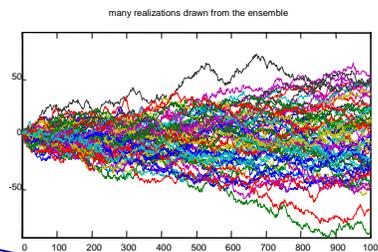
- Let X denote a random process which can be thought of as the *ensemble* (or set) of all *realizations*

$$X = \{x[n]\}$$

- A single *realization* (a sample from the ensemble) can be denoted by:

$$x[n]$$

- $x[45]$ - is a random variable, which represents the 45th sample over the ensemble
- $[x[2], x[11]]^T$ - is a two-dimensional random variable, comprised of the 2nd and 11th samples over the ensemble.
- When analyzing random processes, we are interested in the statistical properties of (potentially) all such combinations.



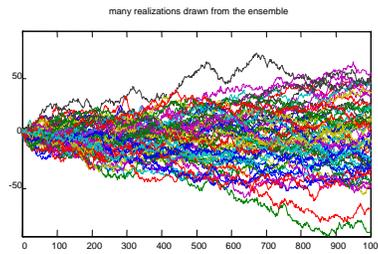
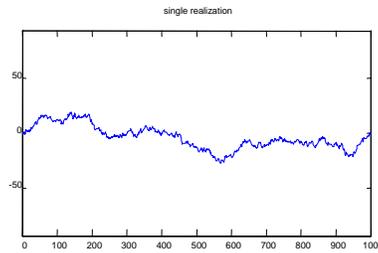
4/24/2007

HST 582 © John W. Fisher III, 2002-2007

4

Realization vs. Ensemble

- Top plot: a single *realization* drawn from the ensemble.
- Bottom plot: many realizations drawn from the *ensemble*.
- An *ensemble* is characterized by the set of all realizations.
- An important question:
When can one infer the statistical properties of the *ensemble* from the statistical properties of a *single* realization?

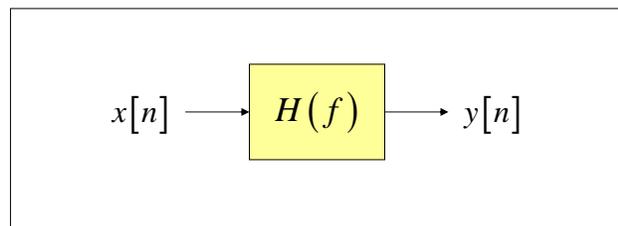


4/24/2007

HST 582 © John W. Fisher III, 2002-2007

5

Relationship to Earlier Material



What, if anything, can we say about $y[n]$ when:

- $H(f)$ is LTI/LSI
- $x[n]$ is a random process

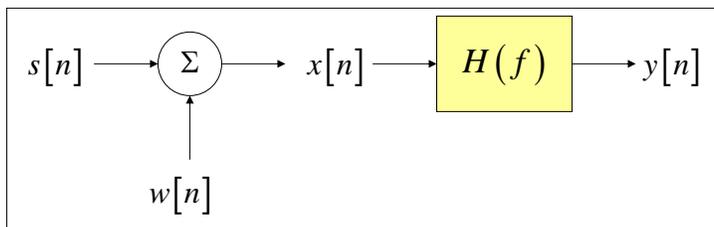
(Random Processes and Linear Systems)

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

6

Relationship to Earlier Material



What, if anything, can we say about $y[n]$ when:

- $H(f)$ is LTI/LSI
- $w[n]$ is a random process
- $s[n]$ is a deterministic signal or random process

(Wiener Filtering)

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

7

Noise reduction with Wiener Filtering



- Left channel
- Right channel
- Wiener (left)
- Wiener (right)

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

8

Important Concepts

Stationarity

Do the statistical properties of the random process vary or remain constant over time?

Ergodicity

Do time-averages over a single realization approach statistical expectations over the ensemble?

Autocorrelation Function

A function which describes the 2nd order statistical properties of a random process.

Crosscorrelation Function

A function which describes the 2nd order statistical relationships between two random processes.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

9

Three Example Random Processes

We'll consider three random processes which illustrate concepts we are interested in.

Notation Reminder

- Samples are drawn from a PDF.

$$x[n] \sim p_X(x)$$

- Samples are drawn from a PDF with parameters, α .

$$x[n] \sim p(x; \alpha)$$

- Samples are drawn from a Gaussian PDF with mean μ and variance σ^2 .

$$x[n] \sim N(x; \mu, \sigma^2)$$

- I.I.D. Gaussian

$$x[n] \sim N(x; 0, \sigma^2)$$

- Gaussian Random Walk

$$y[n] = y[n-1] + x[n]$$
$$x[n] \sim N(x; 0, \sigma^2)$$

- Smoothed Gaussian

$$y[n] = \frac{1}{M} \sum_{k=0}^{M-1} x[n-k]$$
$$x[n] \sim N(x; 0, \sigma^2)$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

10

Some tools we'll need...

- Law of Large Numbers
- Central Limit Theorem
- Chebyshev's Inequality
- Addition of Independent Variables
- Convolution of Gaussians
 - Used to derive properties of the Random Walk process and the smoothed Gaussian i.i.d. process
- Memoryless transformations of RVs
 - Used to derive statistical properties of smoothed Gaussian i.i.d. process
- Multi-Dimensional Gaussian Random Variables
- Correlation and Correlation Coefficient

Binomial Random Process

Consider the Bernoulli random process and its running sum:

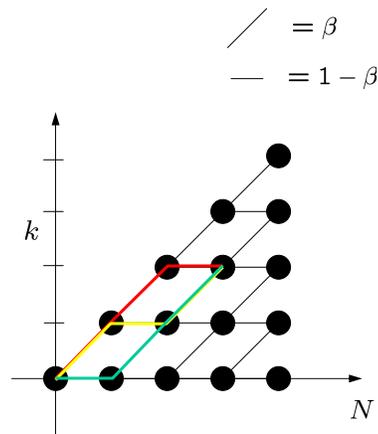
$$x[n] \sim \text{Ber}(\beta)$$

$$k[N] = \sum_{i=0}^N x[i]$$

We showed earlier that

$$p(k) = \binom{N}{k} \beta^k (1 - \beta)^{N-k}$$

- The binomial coefficient counts the number of ways we can observe k successes in N trials.
- The edge weights compute the probability of a given sequence with k successes in N trials.



Binomial Random Process

Consider the Bernoulli random process and its running sum:

$$x[n] \sim \text{Ber}(\beta)$$

$$k[N] = \sum_{i=0}^N x[i]$$

DeMoivre-Laplace Approximation

$$p(K) = \binom{N}{k} \beta^k (1-\beta)^{N-k}$$

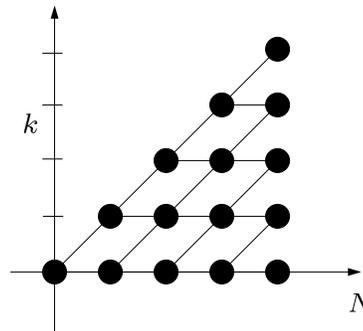
DeMoivre-Laplace Theorem

$$\approx \frac{1}{\sqrt{2\pi N\beta(1-\beta)}} \exp\left(-\frac{(k-N\beta)^2}{2N\beta(1-\beta)}\right)$$

Q. So why does this approximation work?

A. DeMoivre-Laplace is an example of the Central Limit Theorem.

/ = β
 — = $1 - \beta$



4/24/2007

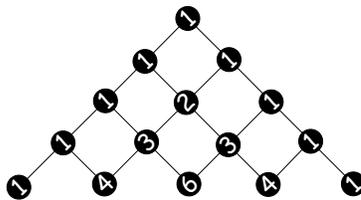
HST 582 © John W. Fisher III, 2002-2007

13

Binomial Random Process

Pascal's triangle

- A simple way to compute the binomial coefficient.
- Turn the graph on its side.



- So Pascal's triangle is just doing the bookkeeping for us.
- It is also an example of something else: namely what happens when we add independent random variables.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

14

Sum of Two Independent Random Variables

- Suppose we generate a new random variable as the sum of two random variables

$$z = x + y \Leftrightarrow p_z(z) = ?$$

- Use CDF relationship to derive the PDF of the new variable
- The PDF resulting from the sum of two independent variables is the convolution of the individual PDFs

$$p_z(z) = p_x(z) * p_y(z)$$

Derivation

$$\begin{aligned} p_z(z) &= \frac{\partial}{\partial z} \Pr \{Z \leq z\} \\ &= \frac{\partial}{\partial z} \Pr \{x + y \leq z\} \\ &= \frac{\partial}{\partial z} \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} p_{XY}(x, y) dx dy \\ &= \frac{\partial}{\partial z} \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} p_Y(y) p_X(x) dx dy \\ &= \int_{-\infty}^{\infty} p_Y(y) \frac{\partial}{\partial z} \int_{-\infty}^{z-y} p_X(x) dx dy \\ &= \int_{-\infty}^{\infty} p_Y(y) p_X(z-y) dy \\ &= p_Y(z) * p_X(z) \end{aligned}$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

15

Chebyshev's Inequality

- A simple *distribution free* bound which is a function of the variance of the random variable.
- It bounds the Probability that a random variable will deviate from it's mean by more than some threshold t .

$$\Pr \{|x - \mu_x| > t\} \leq \frac{\sigma_x^2}{t^2}$$

- Simple to prove and useful in proofs.
- It is a loose bound.

Derivation

$$\begin{aligned} \sigma_x^2 &= \int_{-\infty}^{\infty} (x - \mu_x)^2 p_X(x) dx \\ \text{(a)} &= \int_{-\infty}^{\mu_x-t} (x - \mu_x)^2 p_X(x) dx \\ &\quad + \int_{\mu_x-t}^{\mu_x+t} (x - \mu_x)^2 p_X(x) dx \\ &\quad + \int_{\mu_x+t}^{\infty} (x - \mu_x)^2 p_X(x) dx \\ \text{(b)} &\geq \int_{-\infty}^{\mu_x-t} (x - \mu_x)^2 p_X(x) dx \\ &\quad + \int_{\mu_x+t}^{\infty} (x - \mu_x)^2 p_X(x) dx \\ \text{(c)} &\geq \int_{-\infty}^{\mu_x-t} t^2 p_X(x) dx \\ &\quad + \int_{\mu_x+t}^{\infty} t^2 p_X(x) dx \end{aligned}$$

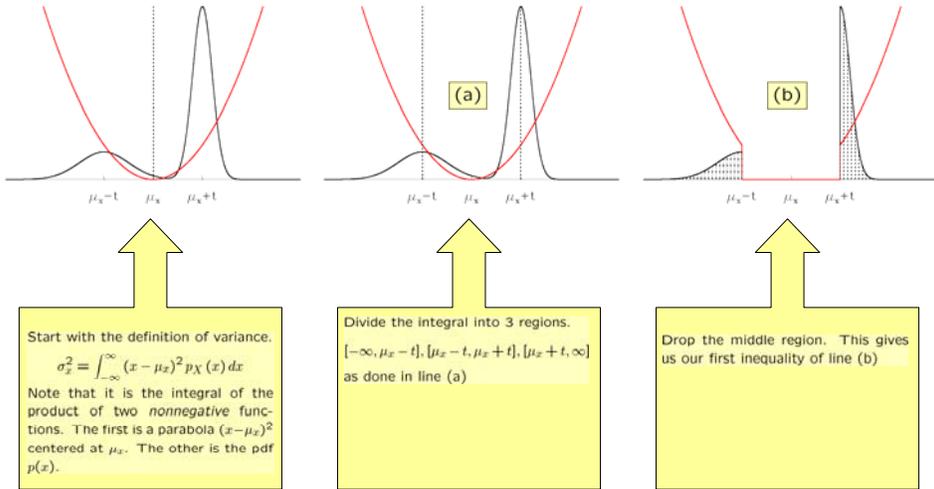
$$\begin{aligned} \frac{\sigma_x^2}{t^2} &\geq \int_{-\infty}^{\mu_x-t} p_X(x) dx \\ &\quad + \int_{\mu_x+t}^{\infty} p_X(x) dx \\ \text{(d)} &\geq \Pr \{|x - \mu_x| > t\} \end{aligned}$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

16

Chebyshev's Inequality (Graphically)

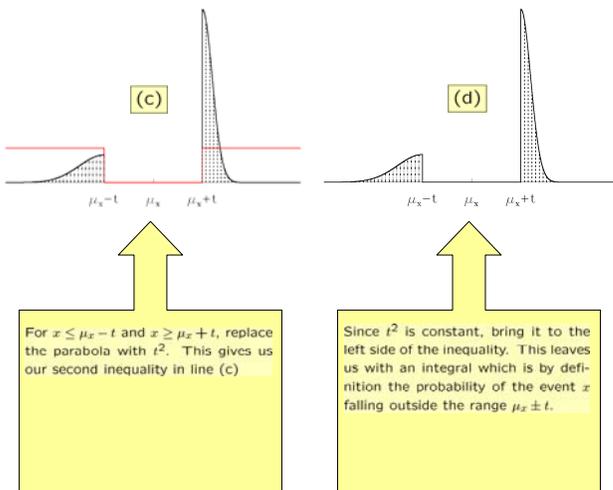


4/24/2007

HST 582 © John W. Fisher III, 2002-2007

17

Chebyshev's Inequality (Graphically)



4/24/2007

HST 582 © John W. Fisher III, 2002-2007

18

The Utility of 2nd Order Statistical Relationships

What do we mean by 2nd-order statistical relationships?

- For a single RV x this refers to expectations of functions of the form:

$$E\{(\alpha x + \beta)(ax + b)\} = \alpha a E\{x^2\} + (\alpha b + \beta a) E\{x\} + \beta b$$

- Similarly for two RVs x_1 and x_2 :

$$E\{(\alpha x_1 + \beta)(ax_2 + b)\} = \alpha a E\{x_1 x_2\} + \alpha b E\{x_1\} + \beta a E\{x_2\} + \beta b$$

This includes quantities such as

- mean: $\mu_i = E\{x_i\}$
- variance: $\sigma_x^2 = E\{(x - \mu_x)^2\}$
- covariance: $E\{(x_1 - \mu_1)(x_2 - \mu_2)\}$
- correlation: $E\{x_1 x_2\}$
- correlation coefficient: $\rho_{12} = \frac{1}{\sigma_1 \sigma_2} E\{(x_1 - \mu_1)(x_2 - \mu_2)\}$

We say two variables are *uncorrelated* if

$$E\{x_1 x_2\} = \mu_1 \mu_2 \Rightarrow E\{(x_1 - \mu_1)(x_2 - \mu_2)\} = 0 \Rightarrow \rho_{12} = 0$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

19

2nd Order Statistical Relationships (Gaussian Case)

- Given x, y jointly Gaussian with the following mean and covariance

$$\begin{aligned} \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} &= E\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \right\} \\ \Sigma_{xy} &= E\left\{ \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix} \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix}^T \right\} \\ &= E\left\{ \begin{bmatrix} (x - \mu_x)^2 & (x - \mu_x)(y - \mu_y) \\ (x - \mu_x)(y - \mu_y) & (y - \mu_y)^2 \end{bmatrix} \right\} \\ &= \begin{bmatrix} \sigma_x^2 & \rho_{xy} \sigma_x \sigma_y \\ \rho_{xy} \sigma_x \sigma_y & \sigma_y^2 \end{bmatrix} \end{aligned}$$

- For a general N-dimensional Gaussian random vector, x :

$$p(x; \Sigma, \mu) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

20

2nd Order Statistical Relationships (Gaussian Case)

- Given x, y jointly Gaussian with the following mean and covariance

$$\begin{aligned} \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} &= E \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \right\} \\ \Sigma_{xy} &= E \left\{ \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix} \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix}^T \right\} \\ &= E \left\{ \begin{bmatrix} (x - \mu_x)^2 & (x - \mu_x)(y - \mu_y) \\ (x - \mu_x)(y - \mu_y) & (y - \mu_y)^2 \end{bmatrix} \right\} \\ &= \begin{bmatrix} \sigma_x^2 & \rho_{xy}\sigma_x\sigma_y \\ \rho_{xy}\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} \end{aligned}$$

- Then $p(y|x)$ is also Gaussian with the following conditional mean and variance

$$\begin{aligned} \mu_{y|x} &= E \{ Y | X = x \} \\ &= \mu_y + \rho_{xy} \frac{\sigma_y}{\sigma_x} (x - \mu_x) \\ \sigma_{y|x}^2 &= \sigma_y^2 (1 - \rho_{xy}^2) \end{aligned}$$

- Note that these are specified in terms of the second order relationships between x and y .

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

21

2nd Order Statistical Relationships (Gaussian Case)

Depicted graphically

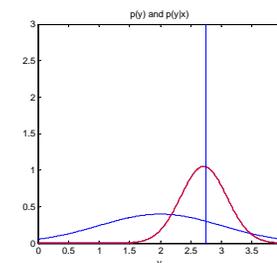
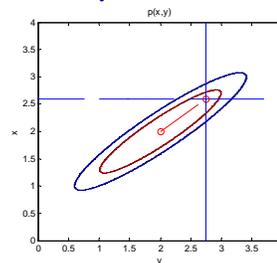
- μ_x and μ_y represent our prior expectation of the random variables
- Having observed a value of X our new expectation of Y is modified by

$$\mu_{y|x} = \mu_y + \rho_{xy} \frac{\sigma_y}{\sigma_x} (x - \mu_x)$$

- Since $0 < \rho^2 < 1$, our uncertainty is reduced.

$$\sigma_{y|x}^2 = \sigma_y^2 (1 - \rho_{xy}^2)$$

- Note that all parameters are 1st or 2nd order statistical properties.



4/24/2007

HST 582 © John W. Fisher III, 2002-2007

22

Linear Prediction (NonGaussian Case)

- Suppose x, y are **not** jointly Gaussian.
- We wish to find the "best" linear predictor of y as a function of x

$$\hat{y}(x) = ax + b$$

- Define "best" in terms of expected squared error:

$$J = E \left\{ (y - \hat{y}(x))^2 \right\}$$

expectation over
both x and y

- The (a, b) which minimize J are

$$a = \rho_{xy} \frac{\sigma_y}{\sigma_x} \quad b = \mu_y - \rho_{xy} \frac{\sigma_y}{\sigma_x} \mu_x$$

Derivation

Express J in terms of a, b , and x .

$$\begin{aligned} J &= E \left\{ (y - \hat{y}(x))^2 \right\} \\ &= E \left\{ (y - ax - b)^2 \right\} \end{aligned}$$

Take the derivatives with respect to a and b

$$\begin{aligned} \frac{\partial J}{\partial a} &= -2E \left\{ (y - ax - b)x \right\} \\ &= -2E \left\{ (yx - ax^2 - b)x \right\} \\ \frac{\partial J}{\partial b} &= -2E \left\{ (y - ax - b) \right\} \end{aligned}$$

Evaluate expectations

$$\begin{aligned} E \left\{ (yx - ax^2 - bx) \right\} &= \rho_{xy} \sigma_x \sigma_y + \mu_x \mu_y \\ E \left\{ (y - ax - b) \right\} &= \mu_y - a\mu_x - b \end{aligned}$$

Set equations to zero and solve for a and b

$$\begin{aligned} \rho_{xy} \sigma_x \sigma_y + \mu_x \mu_y &= 0 \\ \mu_y - a\mu_x - b &= 0 \end{aligned}$$

Convolution of Gaussians*

- If $x=y+z$ (y and z statistically independent) then $p_x(x)$ is the convolution of $p_z(z)$ and $p_y(y)$.

$$x = y + z \quad \leftrightarrow \quad p_x(x) = p_y(x) * p_z(x)$$

- A property of Gaussian functions is that when convolved with each other the result is another Gaussian:

$$\begin{aligned} N(x; \mu_1, \sigma_1^2) * N(x; \mu_2, \sigma_2^2) &= \int_{-\infty}^x N(u; \mu_1, \sigma_1^2) N(x-u; \mu_2, \sigma_2^2) du \\ &= N(x; \mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \end{aligned}$$

- Related to the Central Limit Theorem: when we repeatedly sum many RVs *with finite variance* (not just Gaussians) the repeated convolution tends to a Gaussian (in distribution).

while this is a useful property in general, we'll use it to specifically to analyze the statistics of random walk.

Memoryless Transformations of RVs

Linear Case

$$y = ax + b$$

- Assuming $a > 0$ then

$$p_y(y) = \frac{1}{a} p_x(x) \Big|_{x=\frac{y-b}{a}}$$

$$= \frac{1}{a} p_x\left(\frac{y-b}{a}\right)$$

- For *any* a (not equal to zero) we get

$$p_y(y) = \frac{1}{|a|} p_x\left(\frac{y-b}{a}\right)$$

Derivation

- Define the event over Y in terms of an event over X

$$P_y(y) = \Pr\{Y \leq y\}$$

$$= \Pr\{aX + b \leq y\}$$

$$= \begin{cases} \Pr\left\{X \leq \frac{y-b}{a}\right\} & ; a > 0 \\ \Pr\left\{X \geq \frac{y-b}{a}\right\} & ; a < 0 \end{cases}$$

$$= \begin{cases} \int_{-\infty}^{\frac{y-b}{a}} p_x(u) du & ; a > 0 \\ \int_{\frac{y-b}{a}}^{\infty} p_x(u) du & ; a < 0 \end{cases}$$

substitute $u = \frac{w-b}{a} \Rightarrow du = \frac{1}{a} dw$

$$= \begin{cases} \int_{-\infty}^y \frac{1}{a} p_x\left(\frac{w-b}{a}\right) dw & ; a > 0 \\ \int_y^{\infty} \frac{1}{a} p_x\left(\frac{w-b}{a}\right) dw & ; a < 0 \end{cases}$$

$$= \begin{cases} \int_{-\infty}^y \frac{1}{|a|} p_x\left(\frac{w-b}{a}\right) dw & ; a > 0 \\ \int_{-\infty}^y \frac{1}{|a|} p_x\left(\frac{w-b}{a}\right) dw & ; a < 0 \end{cases}$$

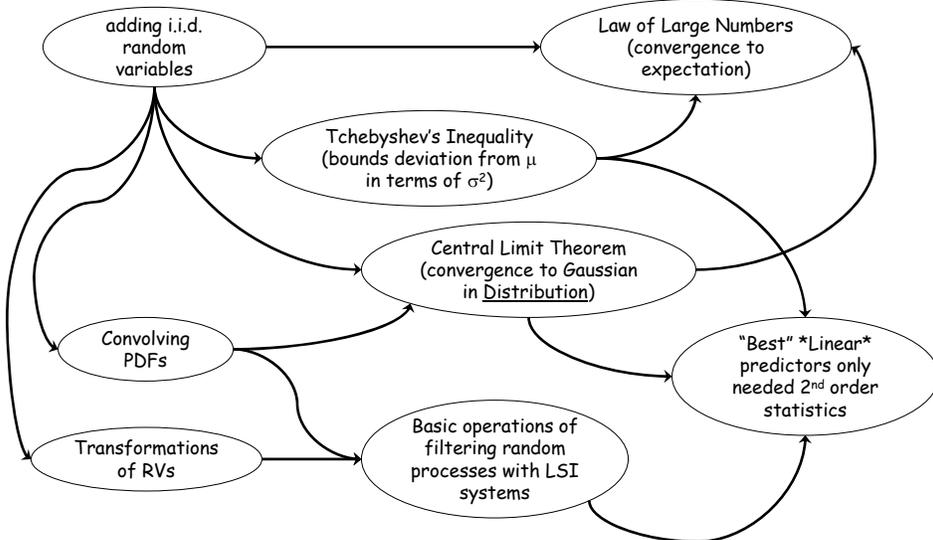
$$= \int_{-\infty}^y \frac{1}{|a|} p_x\left(\frac{w-b}{a}\right) dw$$

$$p_y(y) = \frac{\partial}{\partial y} P_y(y) = \frac{1}{|a|} p_x\left(\frac{y-b}{a}\right)$$

Random Signals Continued (statistics and signal processing)

Recap

at 6AM this figure seemed like a good idea, 9:30AM...not so much



4/24/2007

HST 582 © John W. Fisher III, 2002-2007

27

Gaussian Random Walk

- The evolution equation for (Gaussian) random walk is specified in terms of a *linear difference equation*:

$$\begin{aligned}y[0] &= 0 \\y[n] &= y[n-1] + x[n] \\x[n] &\sim N(x; 0, \sigma^2)\end{aligned}$$

- Given the specification we can derive statistical properties of the random process $y[n]$ in terms of the statistical properties of $x[n]$
- The tools for linear systems are *sometimes* useful for analyzing such equations (when?).
- We'll get at a statistical description of $y[n]$ in a somewhat cumbersome way as a means of motivating the concept of *stationarity*.

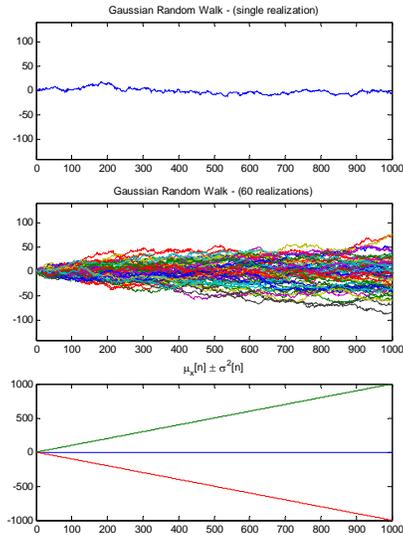
4/24/2007

HST 582 © John W. Fisher III, 2002-2007

28

Random Walk

- The top plot shows a single realization.
- The second plot shows many realizations.
- The bottom plot shows the mean \pm the variance as a function of time (index).
- The mean is constant ($=y[0]$) over all time, but the variance grows linearly with time.



4/24/2007

HST 582 © John W. Fisher III, 2002-2007

29

Gaussian Random Walk

Use the *linear difference equation* to derive the process statistics in closed form:

- Given the following:

$$\begin{aligned} y[0] &= 0 \\ y[n] &= y[n-1] + x[n] \\ x[n] &\sim N(x; 0, \sigma^2) \end{aligned}$$

- Express $y[1]$ in terms of $x[1]$ given $y[0]=0$ to get the PDF of $y[1]$:

$$\begin{aligned} y[1] &= x[1] \\ p_{y_1}(y[1]) &= p_{x_1}(y[1]) \\ &= N(y[1]; 0, \sigma^2) \end{aligned}$$

- Repeat for $y[2], y[3], \dots, y[N]$ (use the convolution property of Gaussians)

$$\begin{aligned} y[2] &= y[1] + x[2] \\ &= x[1] + x[2] \\ p_{y_2}(y[2]) &= p_{x_1}(y[2]) * p_{x_2}(y[2]) \\ &= N(y[2]; 0, 2\sigma^2) \end{aligned}$$

$$\begin{aligned} y[3] &= y[2] + x[3] \\ &= x[1] + x[2] + x[3] \\ p_{y_3}(y[3]) &= N(y[3]; 0, 3\sigma^2) \end{aligned}$$

$$\begin{aligned} y[n] &= \sum_{i=1}^n x[i] \\ p_{y_n}(y[n]) &= N(y[n]; 0, n\sigma^2) \end{aligned}$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

30

Gaussian Random Walk

- We can derive the joint PDF of $\{y[n], y[m]\}$ using conditional probability:

$$p(y[n]) = N(y[n]; 0, n\sigma^2)$$

$$p(y[n], y[m]) = p(y[n])p(y[m]|y[n])$$

- To get $p(y[m]|y[n])$ assume $m > n$ and write $y[m]$ in terms of $y[n]$

$$y[m] = \sum_{i=1}^m x[i]$$

$$= \sum_{i=1}^n x[i] + \sum_{i=n+1}^m x[i]$$

$$= y[n] + \sum_{i=n+1}^m x[i]$$

- When conditioned on $y[n]$, randomness comes only from the second term, which is Gaussian, yielding:

$$p(y[n], y[m]) = N(y[n]; 0, n\sigma^2) N(y[m]; y[n], (m-n)\sigma^2)$$

Conditional Mean

$$E\{y[m]|y[n]\} = E\left\{y[n] + \sum_{i=n+1}^m x[i]|y[n]\right\}$$

linearity of expectation

$$= y[n] + E\left\{\sum_{i=n+1}^m x[i]|y[n]\right\}$$

sum is independent of $y[n]$

$$= y[n] + E\left\{\sum_{i=n+1}^m x[i]\right\}$$

$$= y[n]$$

Conditional Variance

$$E\{(y[m] - E\{y[m]|y[n]\})^2 | y[n]\}$$

$$= E\{(y[m] - y[n])^2 | y[n]\}$$

$$= E\left\{\left(\sum_{i=n+1}^m x[i]\right)^2 | y[n]\right\}$$

x_i, x_j independent for $i \neq j$

$$= \sum_{i=n+1}^m E\{(x[i])^2 | y[n]\}$$

$$= (m-n)\sigma^2$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

31

Gaussian Random Walk

- Alternatively, since variables are jointly Gaussian (which we would have to prove), we could compute the mean and covariance over the vector $[y[n], y[m]]^T$.

$$p(y[n], y[m]) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \begin{bmatrix} y[n] \\ y[m] \end{bmatrix} - \mu \right)^T \Sigma^{-1} \begin{bmatrix} y[n] \\ y[m] \end{bmatrix} - \mu \right)$$

$$\mu = \begin{bmatrix} \mu_n \\ \mu_m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Sigma_{y_n, y_m} = \begin{bmatrix} \sigma_n^2 & \rho_{nm}\sigma_n\sigma_m \\ \rho_{nm}\sigma_n\sigma_m & \sigma_m^2 \end{bmatrix} = \begin{bmatrix} n\sigma_x^2 & \sqrt{\frac{n}{m}}\sqrt{n}\sigma_x\sqrt{m}\sigma_x \\ \sqrt{\frac{n}{m}}\sqrt{n}\sigma_x\sqrt{m}\sigma_x & m\sigma_x^2 \end{bmatrix}$$

$$= \begin{bmatrix} n\sigma_x^2 & n\sigma_x^2 \\ n\sigma_x^2 & m\sigma_x^2 \end{bmatrix}$$

- This result easily extends to all combinations of elements of the random walk process (i.e. we can derive covariances for collections of time samples).

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

32

Gaussian Random Walk

So what have we shown?

- Given a specification in terms of linear difference equations we can derive properties of a random process in closed form.
- However, the derivation for this very *simple* case was nontrivial.
- We will look for simpler ways of describing random processes.

Specific to random walk we see that the marginal and joint densities change over time.

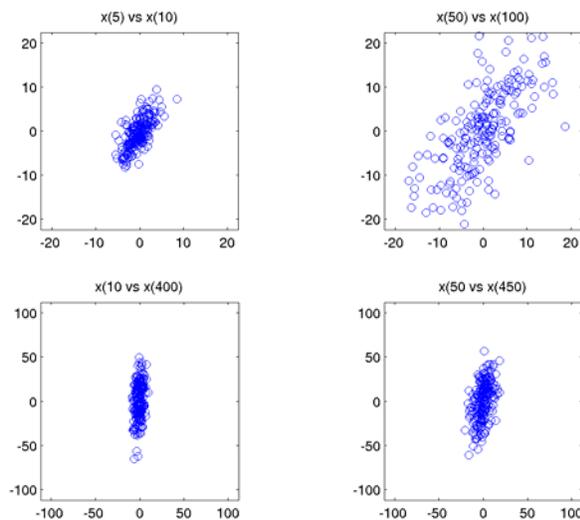
- This is an example of a **nonstationary** random process.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

33

Random Walk (scatter plots)



4/24/2007

HST 582 © John W. Fisher III, 2002-2007

34

Stationary Processes

Stationarity (or lack thereof) is an important statistical property of a random process.

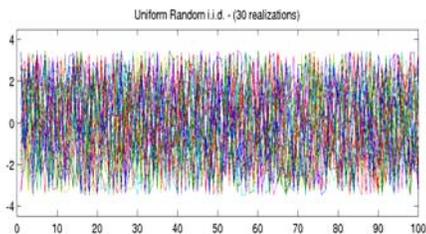
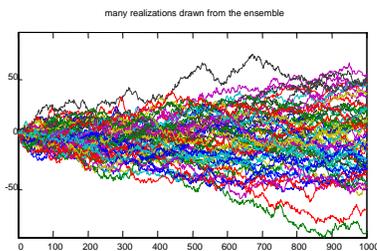
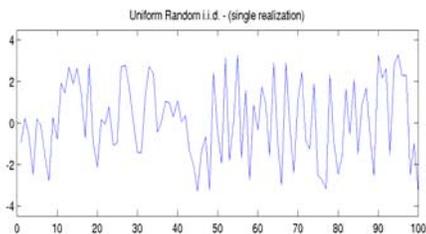
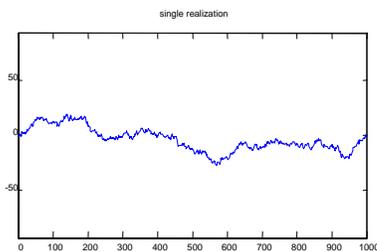
- There are two types of stationarity we are interested in:
 - Strict Sense Stationarity (SSS)
 - A very restrictive class of random processes
 - Wide Sense Stationarity (WSS)
 - A much broader class of random processes
- SSS implies WSS, but the converse is not true.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

35

Strict Sense Stationary



4/24/2007

HST 582 © John W. Fisher III, 2002-2007

36

Strict Sense Stationarity (SSS)

- A random process, $x[n]$, is strict sense stationary if the following property holds:

$$p(x[i] = x_0, \dots, x[i + N - 1] = x_{N-1}) = p(x[j] = x_0, \dots, x[j + N - 1] = x_{N-1})$$

for all i, j, N , and $\{x_0, \dots, x_{N-1}\}$.

- This is basically saying that the signal statistics don't inform us about index or time.
- The Gaussian Random Walk process is an example of a **non-Stationary** process (e.g. variance changes over time)
- The Gaussian I.I.D. process is perhaps the simplest example of a **Stationary** process (the statistics don't change)

Gaussian I.I.D. Process

This is a simple example of a stationary process:

$$y[n] \sim N(y; 0, \sigma^2)$$

- $y[n]$ are sampled **Independently** and **Identically** from the same marginal density
- All joint densities can be expressed as products:

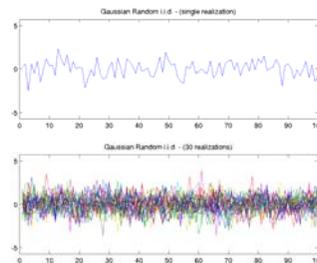
$$p(y[i], \dots, y[i + N - 1]) = \prod_{k=0}^{N-1} N(y[i + k]; 0, \sigma^2)$$

- The distribution is **independent** of i and the form is **identical** for all N .

- The statement of SSS holds trivially:

$$p(y[i] = y_0, \dots, y[i + N - 1] = y_{N-1}) = p(y[j] = y_0, \dots, y[j + N - 1] = y_{N-1})$$

- All i.i.d. (Independent and Identically Distributed) processes are SSS.



Smoothed Gaussian I.I.D. Process

- The smoothed i.i.d. random process (which is SSS) is also defined in terms of a linear difference equation:

$$y[n] = \frac{1}{M} \sum_{k=0}^{M-1} x[n-k]$$

$$x[n] \sim N(x; 0, \sigma^2)$$

- Use convolution property and transformation property to derive $p_y(y[n])$:

$$p_y(y[n]) = (Mp_x(My[n])) * \dots$$

$$\dots * (Mp_x(My[n]))$$

$$= N\left(y[n]; 0, \frac{\sigma^2}{N}\right)$$

- Write $y[n]$ and $y[n+M]$ in terms of $x[n]$ to see that they are **independent** of each other (i.e. there no terms in $y[n]$ that have any statistical dependence on any terms in $y[n+m]$)

$$y[n] = \frac{1}{M} \sum_{k=0}^{M-1} x[n-k]$$

$$= \frac{1}{M} (x[n] + \dots + x[n-M+1])$$

$$y[n+M] = \frac{1}{M} \sum_{k=0}^{M-1} x[n+M-k]$$

$$= \frac{1}{M} (x[n+M] + \dots + x[n+1])$$

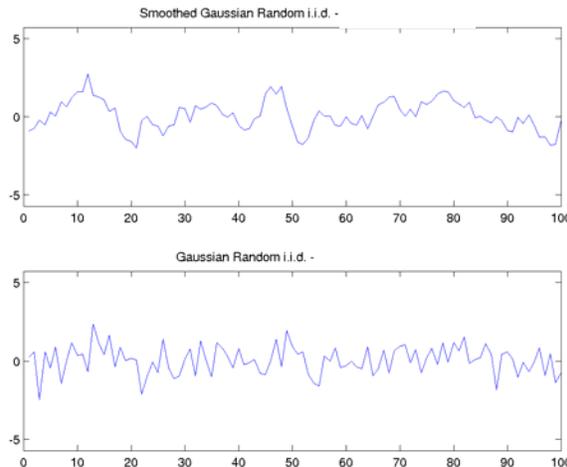
$$p(y[n], y[n+M]) = N\left(y[n]; 0, \frac{\sigma^2}{M}\right) N\left(y[n+M]; 0, \frac{\sigma^2}{M}\right)$$

- Consequently, in order to show SSS we are only left to show that

$$p_y(y[n], \dots, y[n+N]) \forall N = 1 \dots M$$

- Is independent of n .
- This would be done in the same way that we derived joint statistics for random walk

Smoothed Gaussian I.I.D. Process



Relaxing Strict Sense Stationarity -> WSS

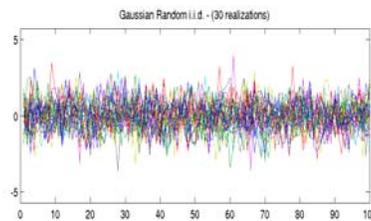
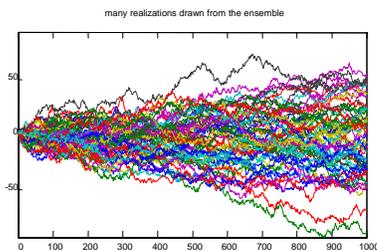
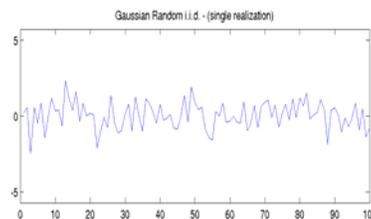
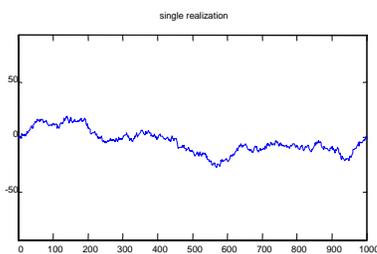
- Strict Sense Stationarity is a somewhat restricted class of random processes.
- This motivates us to look at milder statement of stationarity, specifically Wide Sense Stationarity.
- Wide Sense Stationarity only considers up to 2nd order statistics of a random process.
- Relating the 2nd order statistical properties of output of a linear system when the input is a WSS random process is much easier than in the SSS case.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

41

Wide Sense Stationary



4/24/2007

HST 582 © John W. Fisher III, 2002-2007

42

Wide Sense Stationarity (WSS)

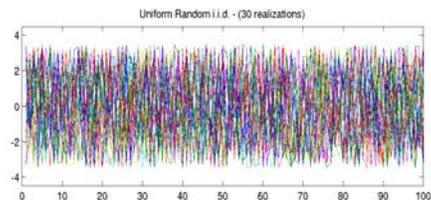
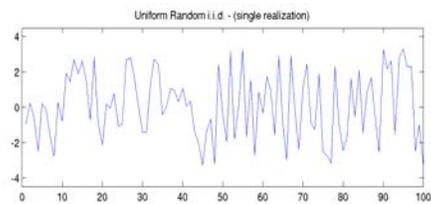
- A random process $x[n]$ is wide sense stationary if the following properties hold:

$$\begin{aligned} E\{x[n]\} &= \mu_x \quad \forall n \\ E\{x[n]x[n+m]\} &= R_x[m] \quad \text{depends only on } m \end{aligned}$$

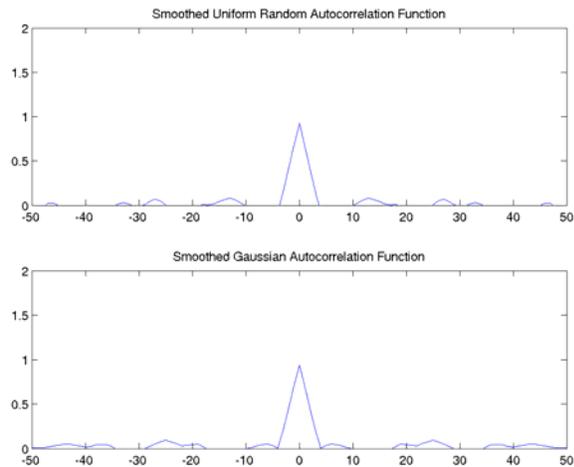
- $R_x[m]$ is the auto-correlation function (discussed later).
- Straightforward to show that SSS implies WSS.
- WSS implies SSS *only* when all of the joint densities of the random process are Gaussian.
- WSS implies constant 2nd order statistics.

Uniform I.I.D. (smoothing revisited)

- Suppose we replace the Gaussian i.i.d. random process in our smoothing example with an i.i.d. sequence of **Uniform** random variables with the same variance.
- The result is a random process whose 2nd order statistical properties are the same as the Gaussian case.
- Why?



Autocorrelation Functions (example)



4/24/2007

HST 582 © John W. Fisher III, 2002-2007

45

Time-Averages of Random Signals

- The *time-average* or *mean* of a random signal is defined as:

$$\langle x[n] \rangle = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x[n]$$

- when the limits *exist*.
- You should be able to show that the time-average operator is linear

$$\langle ax[n] + by[n] \rangle = a \langle x[n] \rangle + b \langle y[n] \rangle$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

46

Signal Statistics using Time-Averages

- The mean (average) of a signal is denoted:

$$\mu_x = \langle x[n] \rangle$$

- The mean power of a signal is defined in terms of its time-average:

$$P_x = \langle x[n]^2 \rangle$$

- The AC power refers to the power in the variational signal component:

$$\sigma_x^2 = \langle (x[n] - \mu_x)^2 \rangle$$

- Fairly straightforward to show that

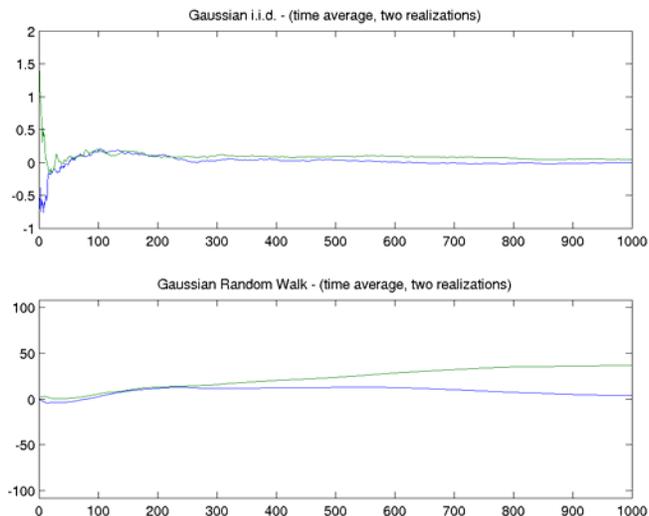
$$P_x = \sigma_x^2 + \mu_x^2$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

47

Time-Averages of Random Signals (Examples)



4/24/2007

HST 582 © John W. Fisher III, 2002-2007

48

Ergodicity

- Random processes which are ergodic have the property that their time averages approximate their statistical averages. That is,

$$\langle x[n]^p \rangle = E \{ x[n]^p \}$$

- for all p (actually for all functions of $x[n]$ or $x(t)$).
- The left hand side is an *average* over all indices (or time) for a specific realization.
- The right hand side is an *expectation* over all realizations at a specific index (or time).
- Generally, SSS implies Ergodicity (there are special cases for which this is not true)

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

49

Mean and Covariance Ergodicity

- Mean ergodicity is a milder form of ergodicity stating simply that the statistical average equals the time average:

$$\begin{aligned} \langle x[n] \rangle &= E \{ x[n] \} \\ &= \mu_x \end{aligned}$$

- Covariance ergodicity states that the variance over time is equal to the variance over realizations:

$$\langle (x[n] - \mu_x)^2 \rangle = E \{ (x[n] - \mu_x)^2 \}$$

- Covariance-ergodicity implies mean-ergodicity, but not the reverse.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

50

- Autocorrelation Functions, Crosscorretion Functions
 - Properties(these were presented on the board)

Autocorrelation and Autocovariance

It turns out that we could have made a choice in our definition of the auto-correlation function.

$$R_x[n, m] = E\{x[n]x[n - m]\}$$

OR

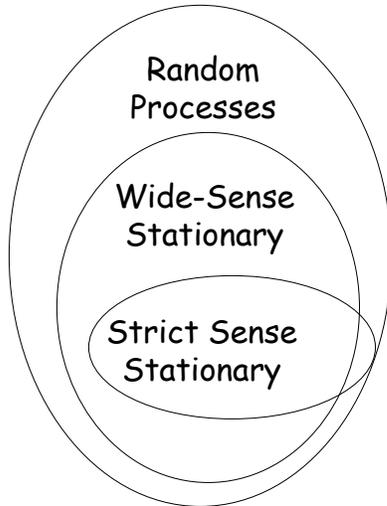
$$R_x[n, m] = E\{x[n]x[n + m]\}$$

Q: So why choose one over the other?

A1: By choosing the definition with $x[n+m]$, many of the spectral properties used later will work out in way that is consistent with processing of deterministic signals

A2: I like Gubner's book on probability and random processes

Stationarity



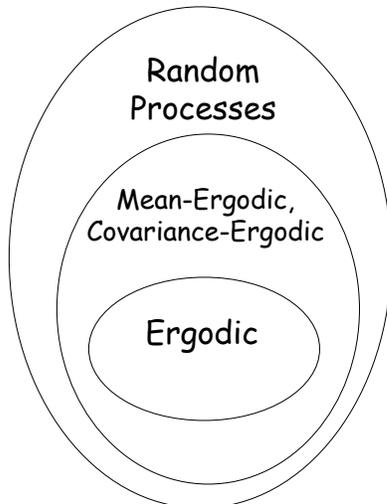
- Wide Sense Stationarity states that statistics up to *second-order* are constant over time.
- Strict Sense Stationarity states that *all* of the process statistics are constant over time.
- SSS implies WSS, but not vice versa (except for a few special cases)
 - e.g. unbounded variance

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

53

Ergodicity



- Strictly Sense Stationarity (almost always) implies Ergodicity.
- Wide Sense Stationarity (almost always) implies Mean and Covariance Ergodicity
- We care about Ergodicity because it means time-averages converge to statistical averages.
- Consequently, for Stationary-Ergodic processes we can estimate statistical properties from the time-average of a *single realization*.
- We can come up with uninteresting examples of SSS processes which are not Ergodic

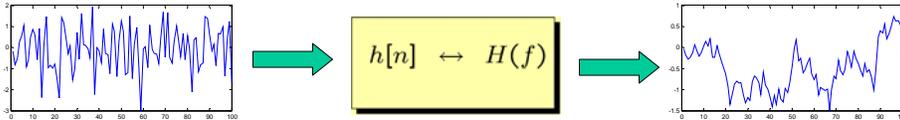
4/24/2007

HST 582 © John W. Fisher III, 2002-2007

54

Linear Systems and Random Processes

- It turns out that characterizing the 2nd order statistical properties of a random process which is input to a linear time/shift invariant system allows us to easily characterize the 2nd order statistical properties of the resulting output.



- Most (all?) of the properties we'll be discussing are geared towards understanding the nature of this relationship.

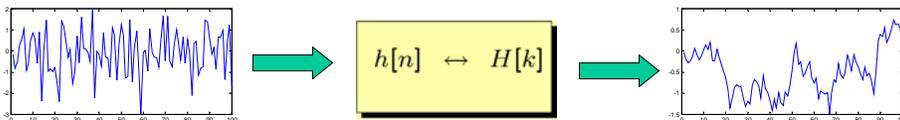
4/24/2007

HST 582 © John W. Fisher III, 2002-2007

55

Linear Systems and Random Processes

- It turns out that characterizing the 2nd order statistical properties of a random process which is input to a linear time/shift invariant system allows us to easily characterize the 2nd order statistical properties of the resulting output.



$$\begin{aligned}
 x[n] &\Rightarrow R_x[m] \\
 R_x[m], h[n] &\Rightarrow R_y[m] \\
 R_{xy}[m], R_y[M] &\Rightarrow g[n] \Rightarrow \hat{x}[n]
 \end{aligned}$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

56

Response of a Linear System

- Given a WSS (or SSS) random process, $x[n]$, and LSI system with impulse response $h[n]$ we can compute the time-average of the output $y[n]$.

$$\begin{aligned}\langle y[n] \rangle &= \langle x[n] * h[n] \rangle \\ &= \left\langle \sum_{k=-\infty}^{\infty} x[n-k] h[k] \right\rangle_n \\ (\text{linearity}) &= \sum_{k=-\infty}^{\infty} h[k] \langle x[n-k] \rangle_n \\ (\text{WSS}) &= \sum_{k=-\infty}^{\infty} h[k] \langle x[n] \rangle \\ &= \langle x[n] \rangle \sum_{k=-\infty}^{\infty} h[k] \\ (\text{ZVT}) &= \langle x[n] \rangle H[0]\end{aligned}$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

57

- Cross-correlation of input and output

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

58

White Noise Sequences/Estimating System Response

- White noise sequences are WSS processes whose autocorrelation function is equal to:

$$R_w[m] = \langle w[n]w[n+m] \rangle = \sigma_w^2 \delta[m]$$

where

$$\langle w[n] \rangle = 0 \\ \langle w[n]^2 \rangle = \sigma_w^2$$

- The process is uncorrelated from sample-to-sample.
- All i.i.d. sequences satisfy this constraint.
- However, we only require the samples to be uncorrelated, but not independent.

- So, if the input to our system is a white-noise sequence:

$$R_{wy}[m] = h[m] * R_w[m] \\ = h[m] * \sigma_w^2 \delta[m] \\ = \sigma_w^2 h[m]$$

then the 2nd order statistics of the output process can be used to estimate the system response function.

- Why is this called "white" noise?
- related to Power Spectrum.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

59

Power Spectral Density

- Note that in the previous result we stated:

$$R_h[m] = h[-m] * h[m]$$

so, the autocorrelation of the system response function is computed as the convolution of the $h[m]$ with a time-reversed version $h[-m]$.

- This convolution could be computed in the Fourier domain:

$$h[-m] * h[m] = \mathcal{F}^{-1} \{ H^*(f)H(f) \} \\ = \mathcal{F}^{-1} \{ |H(f)|^2 \}$$

$$R_h[m] \leftrightarrow |H(f)|^2$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

60

Power Spectrum (con't)

- Motivated by the previous convolution, we shall define the power spectrum of a random process to be the Fourier transform pair of the autocorrelation function:

$$R_x[m] \leftrightarrow S_x(f)$$
$$S_x(f) = \sum_{k=-\infty}^{\infty} R_x[m] e^{-j2\pi f m}$$

- Consequently, we can express the output power spectrum as a function of the system response in the frequency domain and the input power spectrum:

$$S_y(f) = |H(f)|^2 S_x(f)$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

61

Power Spectrum Estimation (nonparametric)

- The Periodogram estimate of the Power Spectrum is defined as

$$\hat{S}_x(f) = \frac{1}{N} |X_N(f)|^2$$
$$= \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j2\pi f n} \right|^2$$

Where the estimate is being computed from the first N observations of a realization, $x[n]$.

- While this seems like a logical estimate of $S_x(f)$, it turns out that it does not converge to $S_x(f)$ as N grows large.
- Why? Because as N grows large the number of degrees of freedom in our estimator grows as well.
- The consequence is that the variance of the estimate is of the same order as the value of $S_x(f)$ at every f.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

62

Power Spectrum Estimation (nonparametric)

- However, all is not lost. If we apply the estimator below to multiple sections of N observations of the realization, $x[n]$.

$$\hat{S}_x(f) = \frac{1}{NM} \sum_{m=0}^{M-1} |X_{N,M}(f)|^2$$

where $X_{N,M}(f)$ is the periodogram of the M th section, then the estimate converges to the true Power Spectrum as M and N grow large.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

63

Power Spectrum Estimation (parametric)

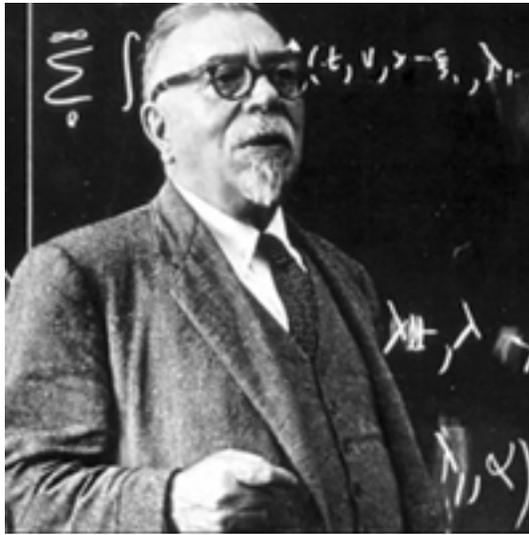
- Based on previous results, suppose we model an observed realization, $y[n]$, as the output of a filter, $h[n]$, whose input was a white-noise sequence.
 - Can you think of a convenient way to estimate the power spectrum?
 - How does it differ from the periodogram approach?
 - The estimate would simply be:
- $$\hat{S}_y(f) = |H(f)|^2$$
- Why?
 - The essential difference (and potential win) comes from how you specify $h[n]$.
 - If you leave the number of free parameters low **AND** your model (the function of the parameters) is reasonably close, the estimate can be quite good.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

64

Wiener Filtering



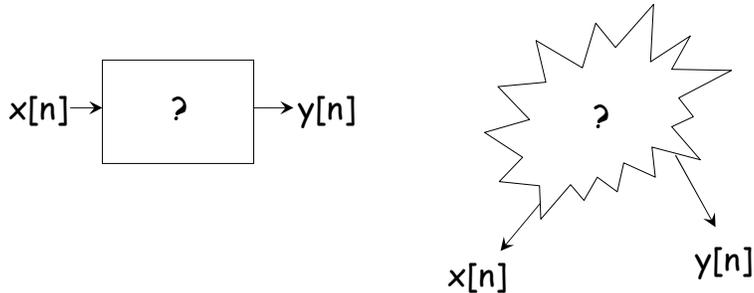
Courtesy of RLE at MIT. Source: Wikipedia.

Primary Concepts for the Wiener Filter

- The problem that a Wiener filter solves
- Applications of Wiener filters
- Wiener filtering implementation details

The Problem

- Suppose we have two random processes whose joint second-order statistics are known, and we want to estimate one from an observation of the other.
 - Disclaimer: We'll only deal with jointly WSS processes.



4/24/2007

HST 582 © John W. Fisher III, 2002-2007

67

Wiener Filter (one case)

- Suppose we have 2 realizations of random processes, $y[n]$ and $x[n]$ and we wish to construct a filter which predicts $y[n]$ from observations of $x[n]$.
- One criteria for choosing such a filter might be to minimize the mean square error (MMSE) between $y[n]$ and your prediction.

$$\begin{aligned}
 \langle e[n]^2 \rangle &= \langle (y[n] - \hat{y}[n])^2 \rangle \\
 &= \langle (y[n] - h[n] * x[n])^2 \rangle \\
 &= \left\langle \left(y[n] - \sum_k h[k] x[n-k] \right)^2 \right\rangle \\
 &= \sigma_y^2 - 2 \sum_k h[k] R_{xy}[k] + \sum_k \sum_l h[k] h[l] R_x[k-l]
 \end{aligned}$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

68

Wiener-Hopf Equations

- We wish to determine the $h[k]$ which minimize the criterion.
- Taking a gradient with respect to the coefficients yields the following system of linear equations

$$R_{xy}[k] = \sum_{l=-\infty}^{\infty} h[l] R_x[k-l]$$

- These equations are known as the Wiener-Hopf or Yule-Walker equations.
- They are easily solved provided there is enough data given the size of the filter.

Nth-order FIR filters

- If we restrict ourselves to Nth order FIR filters, the Wiener-Hopf equations become:

$$\begin{bmatrix} R_x[0] & R_x[1] & \dots & R_x[N-1] \\ R_x[1] & R_x[0] & \dots & R_x[N-1] \\ \vdots & \vdots & \ddots & \vdots \\ R_x[N-1] & R_x[N-1] & \dots & R_x[0] \end{bmatrix} \begin{bmatrix} h[0] \\ \vdots \\ h[N-1] \end{bmatrix} = \begin{bmatrix} R_{xy}[0] \\ \vdots \\ R_{xy}[N-1] \end{bmatrix}$$

- This is a system of linear equations which can be solved if N is not too large.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

69

Orthogonality Principle

- The error is uncorrelated with the estimate.

$$R_{\hat{y}e}[k] = 0$$

Editorial Comments of Kevin Wilson (former 6.555 TA)

- This is big with the 6.432 crowd.
- It can be useful to check this when you're debugging Wiener filter code.
- Why is this "the right thing"?

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

70

Non-causal Wiener Filter

$$R_{xy}[k] = \sum_{l=-\infty}^{\infty} h[l] R_x[k-l]$$

- The Wiener-Hopf equations are just a convolution in this case.
- Time-domain convolutions are usually easier to manipulate in the frequency domain.

yet another editorial comment from K. Wilson

- Let's go there together.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

71

Welcome to the frequency domain

- Power spectrum reminders:

$$S_x(f) = \sum_{k=-\infty}^{\infty} R_x[k] e^{-j2\pi fk}$$

$$\begin{aligned} R_{xy}[n] &= \sum_{l=-\infty}^{\infty} h[l] R_x[n-l] \\ &= h[n] * R_x[n] \end{aligned}$$

$$\begin{aligned} R_{xy}[n] &\leftrightarrow S_{xy}(f) \\ h[n] &\leftrightarrow H(f) \\ R_x[n] &\leftrightarrow S_x(f) \end{aligned}$$

$$H(f) = \frac{S_{xy}(f)}{S_x(f)}$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

72

Non-causal Wiener filtering in the frequency domain

- So Wiener-Hopf is:

$$S_{xy}(f) = S_x(f)H(f)$$
$$H(f) = \frac{S_{xy}(f)}{S_x(f)}$$

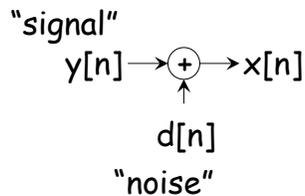
- Each frequency is independent of all other frequencies.
- Now we'll look at some special cases and examples.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

73

Uncorrelated noise case



- Uncorrelated noise: $R_{yd}[n] = 0 \leftrightarrow S_{yd}(f) = 0$

$$H(f) = \frac{S_{xy}(f)}{S_x(f)}$$
$$= \frac{S_y(f)}{S_y(f) + S_d(f)}$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

74

Signal-to-noise ratio interpretation

$$SNR(f) \triangleq \frac{S_y(f)}{S_d(f)}$$
$$H(f) = \frac{S_y(f)}{S_y(f) + S_d(f)}$$
$$= \frac{SNR(f)}{SNR(f) + 1}$$

$$SNR(f) \gg 1 : H(f) \approx 1$$
$$SNR(f) \ll 1 : H(f) \approx 0$$

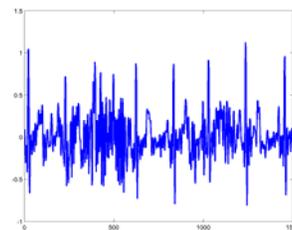
4/24/2007

HST 582 © John W. Fisher III, 2002-2007

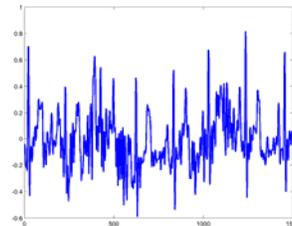
75

ECG example

- What you did in lab (band-pass filter):



- Wiener filtered result:

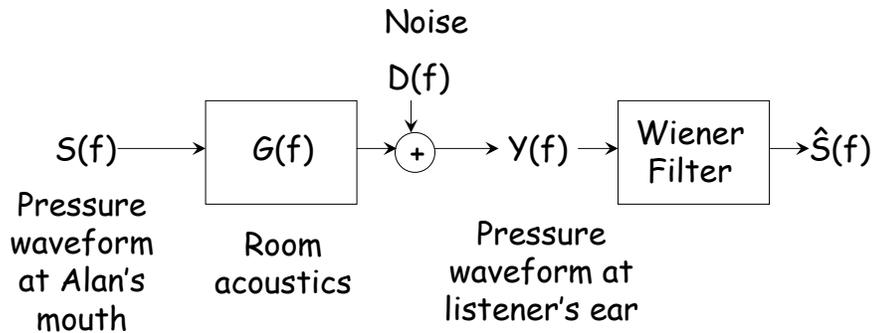


4/24/2007

HST 582 © John W. Fisher III, 2002-2007

76

(Synthetic) Audio Example



- $S(f)$: 
- $Y(f)$: 
- $\hat{S}(f)$: 

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

77

Wiener filter abuse, Part 1

- What if we use the filter derived from Alan's speech on Bob's speech?
- Bob's noisy speech: 
- Bob's Wiener-filtered speech: 

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

78

Wiener filter abuse, Part 2

- What if we use the filter derived from Alan's speech on Caroline's speech?
- Caroline's noisy speech: 
- Caroline's Wiener-filtered speech: 

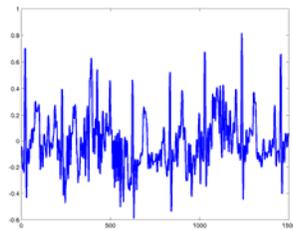
4/24/2007

HST 582 © John W. Fisher III, 2002-2007

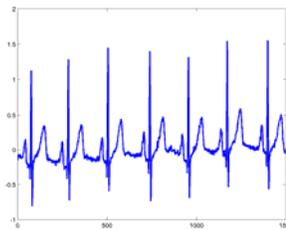
79

Wiener Filters are not the magic solution to all of your problems.

- Additive noise



Wiener filtered



Clean

- Frequency nulls

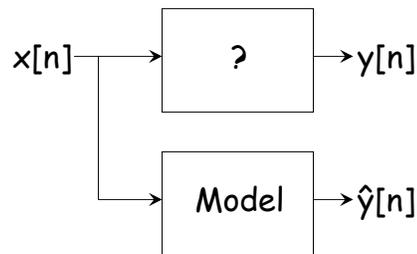
4/24/2007

HST 582 © John W. Fisher III, 2002-2007

80

System Identification - Problem Definition

- Estimate a model of an unknown system based on observations of inputs and outputs.



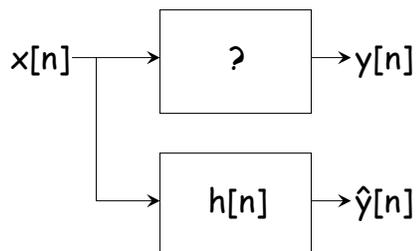
4/24/2007

HST 582 © John W. Fisher III, 2002-2007

81

Wiener Filters for System ID

- Let's restrict our model to be an LTI system.



- Now the goal is to find the LTI system that best predicts the output from the input. This is the Wiener filter (except that we now estimate output from input instead of input from output).
 - The result is the same, but we shift focus from $\hat{y}[n]$ to $H(f)$.

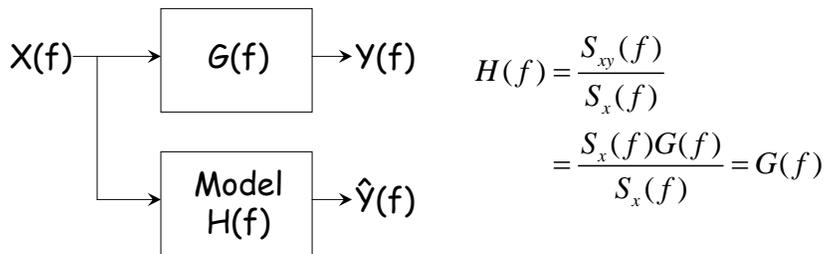
4/24/2007

HST 582 © John W. Fisher III, 2002-2007

82

Example - LTI System

- If the unknown system really is LTI, we hope that Wiener filtering recovers that LTI system exactly.

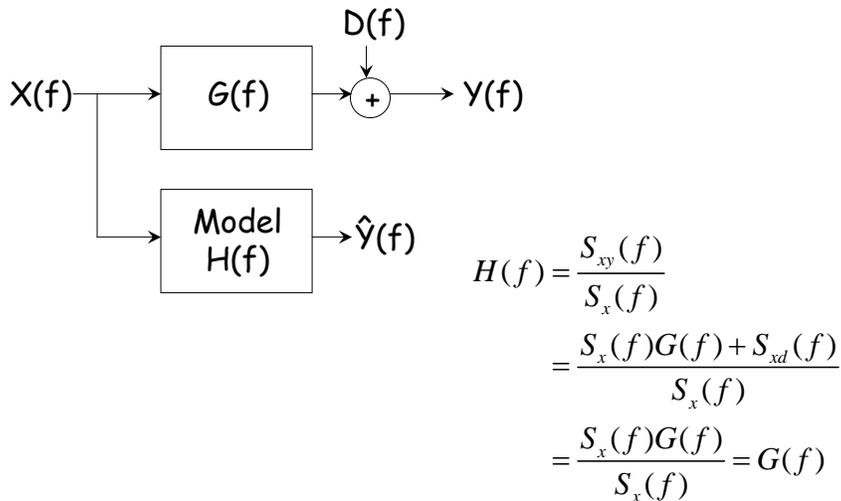


4/24/2007

HST 582 © John W. Fisher III, 2002-2007

83

Example - LTI system plus noise

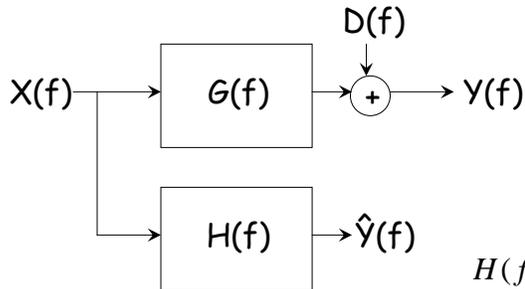


4/24/2007

HST 582 © John W. Fisher III, 2002-2007

84

Example - LTI system plus noise



$$\begin{aligned}
 H(f) &= \frac{S_{xy}(f)}{S_x(f)} \\
 &= \frac{S_x(f)G(f) + S_{xd}(f)}{S_x(f)} \\
 &= \frac{S_x(f)G(f)}{S_x(f)} = G(f)
 \end{aligned}$$

We can still recover the LTI system exactly!

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

85

Causal Wiener Filters

- The Wiener-Hopf equations had a simple form in the non-causal case. Is there an equally simple form for the causal Wiener filter?
- Short answer - not really, but we'll do our best.

$$\begin{pmatrix} R_x[0] & R_x[1] & \cdots & R_x[N-1] \\ R_x[1] & R_x[0] & \cdots & R_x[N-2] \\ \vdots & \vdots & \ddots & \vdots \\ R_x[N-1] & R_x[N-2] & \cdots & R_x[0] \end{pmatrix} \begin{pmatrix} h[0] \\ \vdots \\ \vdots \\ h[N-1] \end{pmatrix} = \begin{pmatrix} R_{xy}[0] \\ \vdots \\ \vdots \\ R_{xy}[N-1] \end{pmatrix}$$

- Let $N \rightarrow \infty$. How do we solve this equation?

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

86

Causal Wiener Filters

- The non-causal Wiener filter had a simple frequency-domain interpretation. Can we do something similar for the causal (IIR) Wiener filter?
- Short answer - sort of.

$$\begin{pmatrix} R_x[0] & R_x[1] & \cdots & R_x[N-1] \\ R_x[1] & R_x[0] & \cdots & R_x[N-2] \\ \vdots & \vdots & \ddots & \vdots \\ R_x[N-1] & R_x[N-2] & \cdots & R_x[0] \end{pmatrix} \begin{pmatrix} h[0] \\ \vdots \\ \vdots \\ h[N-1] \end{pmatrix} = \begin{pmatrix} R_{xy}[0] \\ \vdots \\ \vdots \\ R_{xy}[N-1] \end{pmatrix}$$

- Let $N \rightarrow \infty$. How do we solve this equation?
 - Answer: I have no idea for the general case; let's try a special case.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

87

Causal Wiener Filter w/ white noise observations

- If $R_x[n] = \delta[n]$, then the system simplifies to:

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} h[0] \\ \vdots \\ \vdots \\ h[N-1] \end{pmatrix} = \begin{pmatrix} R_{xy}[0] \\ \vdots \\ \vdots \\ R_{xy}[N-1] \end{pmatrix}$$

- I can solve that!

$$h[n] = \begin{cases} R_{xy}[n] & n \geq 0 \\ 0 & n < 0 \end{cases}$$

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

88

Causal Wiener filter with whitening filter

- If we can find a causal whitening filter $W(f)$ for the observation, we may be able to reduce the problem to the form on the previous slide.
- Hand-waving time... You'll also need the joint statistics between your whitened observation and the process that you're estimating. To get these joint statistics, $W(f)$ must have a causal inverse.
- It's not always possible to find $W(f)$ such that both $W(f)$ and $W^{-1}(f)$ are causal.

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

89

Implementation details

- Filter form restrictions:
 - FIR: solve linear system
 - Causal: whiten observation
 - Noncausal: nice frequency-domain expression
- Acquiring the statistics
 - Where do we get R_x and R_{xy} ?
 - Sample statistics for auto- and cross-correlations
 - Periodogram estimates of power spectral densities

4/24/2007

HST 582 © John W. Fisher III, 2002-2007

90

Summary

- A Wiener filter finds the MMSE estimate of one random process as a linear function of another random process.
- Applications includes noise removal and system identification.