Overview of Kalman Filter Theory and Navigation Applications

Day 2

Rev. A

Michael L. Carroll

Feb 24, 2004

Day 2: The Basic Kalman Equations, Part II
Segments

- Probability and Statistics

- Random Processes

- Some Important Random Processes

- Kalman Statistics

Day 2, Segment 1
Probability and Statistics
Topics

- Sample Space and Probability Measures
- Probability Distributions and Densities
- Random Variables and Expectation
- Moments: Means and Variances

Probability: Sample Spaces

- Experiments

- Sample Space: Ω

  - consists of all possible outcomes ω.

- Events: Subsets of Ω forming a σ-algebra. Not necessarily *all* subsets.

Examples: Experiment and Sample Space

- 1. Experiment: Single coin toss
  - Sample Space: $\Omega = \{H, T\}$
- 2. Experiment: Toss coin twice
  - Sample Space: $\Omega = \{HH, HT, TH, TT\}$

Probability: σ-Algebras

- In probability, we are often less concerned with individual outcomes and more with sets of outcomes.

- Sets of outcomes, i.e., subsets of the sample space Ω, are called "events".

- Not every subset of the sample space Ω qualifies as an event.

- Qualified events form an "event algebra", known in mathematics as a σ-**algebra**

Probability: σ-Algebras

- A σ-algebra $\mathcal{M}$ is a collection of subsets of Ω that is closed under arbitrary unions and intersections, set complementation, and which also contains the whole space Ω as well.

- A σ-algebra can be expressed in structural notation as $< \mathcal{M}, \cup, \cap, ', \emptyset, \Omega >$

- Ω is the certain event.

Examples: σ-algebras

- Double coin toss:
  - Sample Space: $\Omega = \{HH, HT, TH, TT\}$
  - σ-algebra: $\mathcal{M} = \{\emptyset, \{HH\}, \{HT, TH, TT\}, \Omega\}$

Mutually Exclusive Events

- Events $A, B \in \mathcal{M}$ are called **mutually exclusive**, if they are disjoint as sets, i.e., if $A \cap B = \emptyset$.

- In last example, the events $\{HH\}$ and $\{HT, TH, TT\}$ are mutually exclusive.

Set Functions and Measures

- A measure is a way of assigning a number to a set.
- Intuitively, it measures the size of the set.
- In probability, a measure is a way of assigning a probability to an event.

Set Functions and Measures

- A set function μ on $\mathcal{M}$ assigns a real number $\mu(A)$ to each set $A \in \mathcal{M}$.

- A set function μ on $\mathcal{M}$ is said to be **additive**, if $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever A and B are disjoint.

- μ is called **countably additive** if $\mu(\bigcup_{i\in I} A_i) = \sum_{i\in I} \mu(A_i)$, if the family of sets $\{A_i : i \in I\}$ is disjoint.

Set Functions and Measures

- A measure μ on a σ-algebra $\mathcal{M}$ is a real-valued, non-negative, countably additive set function: $\mu : \mathcal{M} \to \mathbb{R}$

- Measures are used in mathematics to define abstract integration.

Set Functions and Measures

- Example: Counting Measure. Ω is finite and the measure of each subset is merely the number of elements in the set.

- Example: Dirac Measure. Let $\mathcal{M}$ be any σ-algebra and let $\omega \in \Omega$ be an arbitrary but fixed element of Ω. For each $A \in \mathcal{M}$ define

$$\mu_\omega(A) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{otherwise.} \end{cases}$$

Probability Measures

- A measure μ on $\mathcal{M}$ with the property that $\mu(\Omega) = 1$ is called a **probability measure** or **probability function**.

- Instead of μ we will use P, which is more customary.

Probability: Specialized Terminology

- Null Event: $P(A) = 0$. It is important to realize this does not imply that $A = \emptyset$. In measure theory this would be called a set of measure zero.

- Independent Events: $P(A \cap B) = P(A)P(B)$

- Mutually Exclusive Events:
  $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

Examples: Probability Measure

- Double coin toss:
  - Sample Space: $\Omega = \{HH, HT, TH, TT\}$
  - σ-algebra: $\mathcal{M} = \{\emptyset, \{HH\}, \{HT, TH, TT\}, \Omega\}$
  - Probability Measure: $P(A)$ =number of elements in A divided by total number of elements in Ω
  - E.g., $P(\{HH\}) = 1/4$

Examples: Probability Measure

- Let Ω be any finite set of cardinality n. Let the σ-algebra be the set of all subsets of Ω and let $|\cdot|$ denote the counting measure, i.e. for $A \subseteq \Omega$, $\|A|$ =number of elements in A. Define a probability measure by $P = \frac{1}{n}|\cdot|$.

- To illustrate the last example a little more, let $\Omega = \{\omega_1, \ldots, \omega_6\}$, and let $A = \{\omega_1, \omega_2\}$. Then $P(A) = \frac{1}{6} \cdot |A| = \frac{1}{6} \cdot 2 = \frac{1}{3}$

Examples: Probability Measure

- Let $\Omega = [0, 1]$ and let $\mathcal{B}$ be the Borel σ-algebra on the interval $[0, 1]$. Define $P = \lambda$, the Lebesgue measure on $\mathcal{B}$. Note that for any open subinterval (a, b) of $[a, b]$, we have $P(a, b) = b - a$.

Conditional Probability

- Suppose you have two events A and B and you know that the event B has in fact occured.

- How does that knowledge affect the probability that both A and B have occured?

- Intuitively, we would tend to think that the likelihood of $A \cap B$ increases with the knowledge that B has occured.

Conditional Probability

- The increased likelihood of A given that B has occured is called the **conditional probability** of A given B.

- If $P(B) \neq 0$, then the conditional probability of A, given that B has occurred, is: $P(A|B) := P(A \cap B)/P(B)$

- Sometimes we say the probability of A conditioned on B.

- Note that since $P(B) \leq 1$, $P(A \cap B)/P(B) \geq P(A \cap B)$

Conditional Probability: Bayes' Theorem

- Since $P(A|B) = P(A \cap B)/P(B)$ and $P(B|A) := P(B \cap A)/P(A)$, we can eliminate $P(A \cap B)$ and derive

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- Using this we obtain Bayes' Theorem (simple form):

(1) $$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A')P(B|A')}.$$

Conditional Probability

- Bayes' Theorem (complex form): Let $A_1, \ldots, A_n$ be a partition of Ω. Then,

(2)
$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}.$$

- Bayes' Theorem expresses posterior probability in terms of prior probabilities and conditional probabilities.

Conditional Probability

- In one derivation of the Kalman filter, the use of Bayes' Theorem is a crucial step. We'll see this later.

- Bayes' Theorem is used to derive the Kalman gain: to determine the best estimate of state **after** measurement has occurred.

Conditional Probability

- Bayes' Theorem allows us to reverse the conditional probability.

- We can derive information about the state conditioned on the measurement from information about the measurement conditioned on the state.

Random Variables

- In measure-theoretic terms, a ra**ndom variable** is simply a real-valued measurable function on the sample space Ω.

- A **measurable function** is one such that the inverse image of any open set in $\mathbb{R}$ is measurable (i.e. is a member of the σ-algebra).

- Some elementary probability texts are not that strict and instead consider any real-valued function of the sample space to be a random variable.

Random Variables

- We use capital letters at the end of the alphabet to denote random variables: $X, Y, Z, ...$ (This helps us distinguish them from ordinary real-valued variables.)

- Later we will use lowercase letters so as not to confuse them with certain transforms.

Random Variables

- Note that random variables are not really variables in any ordinary sense – they are functions. Thus, if $\omega \in \Omega$, and X is a random variable on Ω, then $X(\omega) \in \mathbb{R}$, i.e., $X(\omega)$ is a real number.

- Also note that $X^{-1}(A) = \{\omega \in \Omega | X(\omega) \in A\}$ is always an event in the event algebra by definition.

Random Variables

- Random variables provide a way to quantify the outcomes of experiments.

- Random variables are important because we are often more interested in the consequences of an experiment than in the experiment itself.

- For example, a gambler is usually more interested in his wins and losses than in the particular game he's playing.

Random Variables

- Moreover, random variables are not really random.
- They are deterministic functions from the space of all possible outcomes to the real numbers. If you give them a particular outcome to operate on, they return the same real value every time.
- What gives them the appearance of randomness is really the fact that you don't really know what those values are. Usually, you only know (or assume you know) certain statistical properties like distributions (to be discussed soon).

Random Variables

- The "randomness" of random variables stems from the probabilistic nature of the underlying sample space. Usually, we can't say much about the individual outcomes in a sample space. We can only measure the size of events. (The connection between the "randomness" of the underlying space and the random variables will be made by means of probability distributions.)

- It is because random variables are functions that we find ourselves having to deal with infinite dimensional vector spaces, i.e., function spaces.

Equality and Equivalence of Random Variables

- Since random variables are functions, strict equality of random variables is the same as the equality of functions. Two functions are equal if and only if they have equal values when evaluated at the same point: $X = Y \Leftrightarrow X(\omega) = Y(\omega)$ for all $\omega \in \Omega$.

- However, when you take probability into account, things can get a little fuzzier. We say that two random variables X and Y are equal **almost surely** (a.s.) if $P\{\omega \in \Omega : X(\omega) \neq Y(\omega)\} = 0$.

- In other words, $X = Y$ a.s., if the set of all points at which X and Y differ is insignificant, i.e., of measure 0.

Simple Random Variables

- If a random variable X takes on only finitely many values, then X is called a **simple random variable**.

- If X is a simple random variable and $\{a_1, \cdots, a_n\}$ are its values, let $A_i = \{\omega : X(\omega) = a_i\}$ for each $i = 1, \ldots, n$. Then X can bewritten as a sum:

  (3) $$X = \sum_{i=1}^{n} a_i \chi_{A_i}$$

  where $\chi_{A_i}(\omega) = \begin{cases} 1 & \text{if } \omega \in A_i, \\ 0 & \text{otherwise.} \end{cases}$ is the characteristic or indicator function on A_i.

Example: Simple Random Variable

- Double coin toss:
  - Sample Space: $\Omega = \{HH, HT, TH, TT\}$
  - σ-algebra: $\mathcal{M} = \{\emptyset, \{HH\}, \{HT, TH, TT\}, \Omega\}$
  - Probability Measure: $P(A)$ =number of elements in A divided by total number of elements in Ω

Example: Simple Random Variable

- Let $A = \{HT, TH, TT\}$ and define $X = \chi_A$. Then

$$X(HH) = 0$$
$$X(HT) = 1$$
$$X(TH) = 1$$
$$X(TT) = 1$$

Example: Simple Random Variable

- Let $B = \{HH\}$ and define $Y = -3\chi_A + 0.5\chi_B$
- Y is a simple random variable.
- We compute:

$$\begin{aligned} Y(HH) &= -3\chi_A(HH) + 0.5\chi_B(HH) \\ &= -3 \cdot 0 + 0.5 \cdot 1 \\ &= 0.5 \end{aligned}$$

Random Variables form a Vector Space

- Let $RV(\Omega)$ denote the set of all random variables on Ω. If $X, Y \in RV(\Omega)$ and $a, b \in \mathbb{R}$, then $aX + bY \in RV(\Omega)$.

- Recall that $(aX+bY)(\omega) = a \cdot X(\omega) + b \cdot Y(\omega)$, where $\cdot$ denotes ordinary real multiplication.

- For brevity's sake, we will drop the $\cdot$ and merely juxtapose. Likewise, we will drop Ω from $RV(\Omega)$ and merely write RV, assuming implicitly that the underlying sample space is Ω.

Probability Distributions and Densities

- Probability Distributions

- Probability Density Functions

Probability Distributions

- Every random variable X has an associated probability **distribution** function defined by

$$F_X(x) = P\{X^{-1}[-\infty, x)\} = P\{\omega \in \Omega : X(\omega) < x\}$$

- The value of the distribution function at a real number x is the probability that the value of X does not exceed x.

- We can simplify the notation somewhat by writing $\{X < x\}$ instead of $\{\omega \in \Omega : X(\omega) < x\}$.

- Thus, $F_X(x) = P\{X < x\}$.

Probability Distributions

- Distributions are nondecreasing: $x \leq y \Rightarrow F(x) \leq F(y)$.
- Distributions are always continuous from the left: $F(x) \to F(a)$ as $x \uparrow a$.
- The value of a distribution is always between 0 and 1 inclusively, and $\lim_{x \downarrow -\infty} F(x) = 0$ and $\lim_{x \uparrow \infty} F(x) = 1$

Probability Distributions
Three Flavors

- Continuous (no jumps)

- Piecewise continuous (some jumps)

- Discrete (only jumps and flat). X is discrete if and only if X takes on only finitely many, or at most countably infinite, distinct values. Note that a simple measurable function is a discrete random variable.

Some Discrete Distributions

- Discrete Uniform Distribution: Let $A = \{a_1, \ldots, a_n\}$ be a set of n distinct real numbers and assume that $a_1 < a_2 < \ldots < a_n$. Then a random variable X is said to have a discrete uniform distribution if
$F(x) = \frac{1}{n} \cdot |A \cap [-\infty, x)|$.

- Let Ω be any sample space and $A \subseteq \Omega$. Let $X(\omega) = \chi_A(\omega)$.
Recall that $\chi_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{otherwise.} \end{cases}$
Then $F(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - P(A) & \text{if } 0 \leq x < 1, \\ 1 & \text{if } x \geq 1. \end{cases}$

Continuous Distributions: Probability Density Functions

- X is said to have a continuous distribution if there exists a function $f : \mathbb{R} \to \mathbb{R}$ such that for each $x \in \mathbb{R}$

$$F(x) = \int_{-\infty}^{x} f(u)du. \tag{4}$$

- f is called a **probability density function**, or p.d.f, for short.

- The p.d.f of a random variable, if it exists, is not unique. Why?

Some Important Distributions

- The Gaussian or normal distribution with p.d.f:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{5}$$

- The uniform distribution on $[a, b]$ with distribution function:

$$F(x) = \begin{cases} 0 & \text{if } x < a, \\ \frac{x-a}{b-a} & \text{if } a \leq x < b, \\ 1 & \text{otherwise.} \end{cases} \tag{6}$$

Moments: Means and Variances of Random Variables

- 1st Order Moment: Expectation or Mean
- 2nd Order Moment: Variance
- Higher Order Moments

1st Order Moment: Expectation of a Random Variable

- Let X be a simple random variable and let $a_1, \ldots, a_n$ be the distinct values taken on by X. Then, the expectation operator $\mathsf{E} : SRV \to \mathbb{R}$ is defined as

$$\mathsf{E}[X] = \sum_{i=1}^{n} a_i P(A_i) \tag{7}$$

where $A_i = X^{-1}\{a_i\} = \{\omega \in \Omega : X(\omega) = a_i\}$.

- Thus, the expectation of X is the weighted sum of the values of X, the weights being the "size" or likelihood of the events.

- It is clear that $X = Y$ a.s. $\Rightarrow \mathsf{E}[X] = \mathsf{E}[Y]$.

Expectation

- Alternative Integral notation: $\mathsf{E}[X] = \int_\Omega X dP$, because the weighted sum given in 7 is just the definition of the integral of X with respect to the (probability) measure P.

- It can be shown that every random variable X can be written as the difference of two nonnegative random variables.

- It can also be shown that every nonnegative random variable is the limit of a monotone sequence of simple random variables. Therefore, integration can be extended from $SRV(\Omega)$ to $RV(\Omega)$

Expectation

- Note that we did not need the use of a probability distribution nor density in defining the expectation this way.
- We only needed the probability function.
- Note also that the integration we defined above is abstract in the sense that it works just as well for general measures, not just probability measures.

Expectation

- Expectation is a linear functional on the space RV of random variables over Ω:

$$\mathsf{E}[aX + bY] = a\mathsf{E}[X] + b\mathsf{E}[Y]$$

- Expectation is monotonic:

$$X \leq Y a.s. \Rightarrow \mathsf{E}[X] \leq \mathsf{E}[Y]$$

- We can compute $\mathsf{E}[X]$ by means of an ordinary Riemann-Stieltjes integral:

$$\mathsf{E}[X] = \int_{-\infty}^{\infty} x dF(x) \tag{8}$$

- If X also has a p.d.f, then equation (8) becomes:

$$\text{(9)} \qquad \mathsf{E}[X] = \int_{-\infty}^{\infty} x dF(x) = \int_{-\infty}^{\infty} x F'(x) dx = \int_{-\infty}^{\infty} x f(x) dx$$

- The expectation of $\exp(juX)$ for each $u \in \mathbb{R}$ is called the **characteristic function** of X:

$$\text{(10)} \qquad \phi_X(u) = \mathsf{E}[\exp(juX)] = \int_{-\infty}^{\infty} \exp(jux) dF(x)$$

where $j = \sqrt{-1}$.

2nd Order Moment: Variance

- The variance of a random variable is defined as the mean square deviation from the mean:

$$\sigma^2 = \mathsf{E}\left[(X-\mu)^2\right]$$

Higher Order Moments

- The value of the integral

$$m_i = \int_{-\infty}^{\infty} x^n f(x) dx$$

is called the n^{th} **moment** of the distribution. If instead we subtract out the mean $\mu = \mathsf{E}[X]$ prior to integrating, then we obtain the n^{th} **central moment**:

$$\mu_i = \int_{-\infty}^{\infty} (x - \mu)^n f(x) dx$$

- We note that $\mu = \mu_1$, the first moment.

Higher Order Moments

- ϕ_X is similar to the **moment generating function** defined as follows:

$$M_X(t) = \mathsf{E}\left[\exp(tX)\right]$$

- If X has a density function f, then

$$M_X(t) = \int_{-\infty}^{\infty} \exp(tx) f(x) dx$$

Higher Order Moments

- Note that

$$\mu = \mathsf{E}[X] = \frac{d}{dx} M_X(0)$$

  and, in general,

$$\mathsf{E}[X^n] = \frac{d^n}{dx^n} M_X(0).$$

Higher Order Moments

- The Gaussian distribution has only two fundamental moments $\mu = \mathsf{E}[X]$ and $\sigma^2 = \mathsf{E}\left[(X-\mu)^2\right]$. All higher-order moments can be expressed in terms of these two.

- Thus, the Gaussian distribution is said to be a second-order distribution.

Individual Exercises

- Show that $A \subseteq B \Rightarrow P(A) \leq P(B)$.
  (Hint: Use $B = A \cup (B \setminus A)$ and the fact that measures are additive.)

- Show that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- Let F be the distribution of a random variable X and suppose that $a \leq b$ are real numbers. Show that $P(x \in [a, b)) = F(b) - F(a)$.

Workgroup Exercises

- Prove the following: If two events A and B are independent, then the events A and B' are also independent.

- Show that for any distribution function F of a random variable X, $P(X > a) = \lim_{x \downarrow a} F(x)$.

Workgroup Exercises

- Let Ω be any sample space and $A \subseteq \Omega$. Let $X(\omega) = \chi_A(\omega)$.

  Recall that $\chi_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{otherwise.} \end{cases}$

  Prove that $F(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - P(A) & \text{if } 0 \leq x < 1, \\ 1 & \text{if } x \geq 1. \end{cases}$

Workgroup Exercises

- A gyro manufacturer produces defective gyros with probability p, where $0 < p < 1$. After a lot of n independent gyros have been produced, they are examined. Let X denote the number of defective gyros in the lot. What is the probability mass function of this discrete distribution?

Day 2, Segment 2
Random Processes
Topics

- Two Dimensions: Ensemble vs. Time

- Stationary and Ergodic

- Correlation and Covariance

- Power Spectral Density

Two Dimensions: Ensemble vs. Time

- There are two equivalent ways of defining a random process. One way is to view a random process as a time function whose value at each instant of time is a random variable.

- The other way is to define a random process as a function of two variables: the outcomes ω and time t.

- The outcomes ω belong to the sample space Ω while t is modeled as a real number, usually non-negative.

Two Dimensions: Ensemble vs. Time

- At each fixed t, the random process is simply a random variable, i.e., a function of the other variable ω alone

- For a fixed ω, we have an ordinary time function of t.

Random Processes

- A **Random Process** (also called a **stochastic process**) is a function $X : [0, T] \to RV(\Omega)$, where $[0, T]$ is a closed interval of the real line.

- We think of variables $t \in [0, T]$ as time.

- The image of $t \in [0, T]$ under X is a random variable:

$$X(t) : \Omega \to \mathbb{R}$$

.

Random Processes

- Thus a random process can be viewed as a function whose values are functions.

- We could also think of X as a two-variable function from the Cartesian product $[0, T] \times \Omega$ into the reals:

$$X(t, \omega) \in \mathbb{R}$$

Random Processes

- These two perspectives are interchangeable and we shall use both.

- We use all of the following notation for a random process

$$X_t$$

$$X_t(\omega)$$

$$X(t)$$

$$X(t, \omega)$$

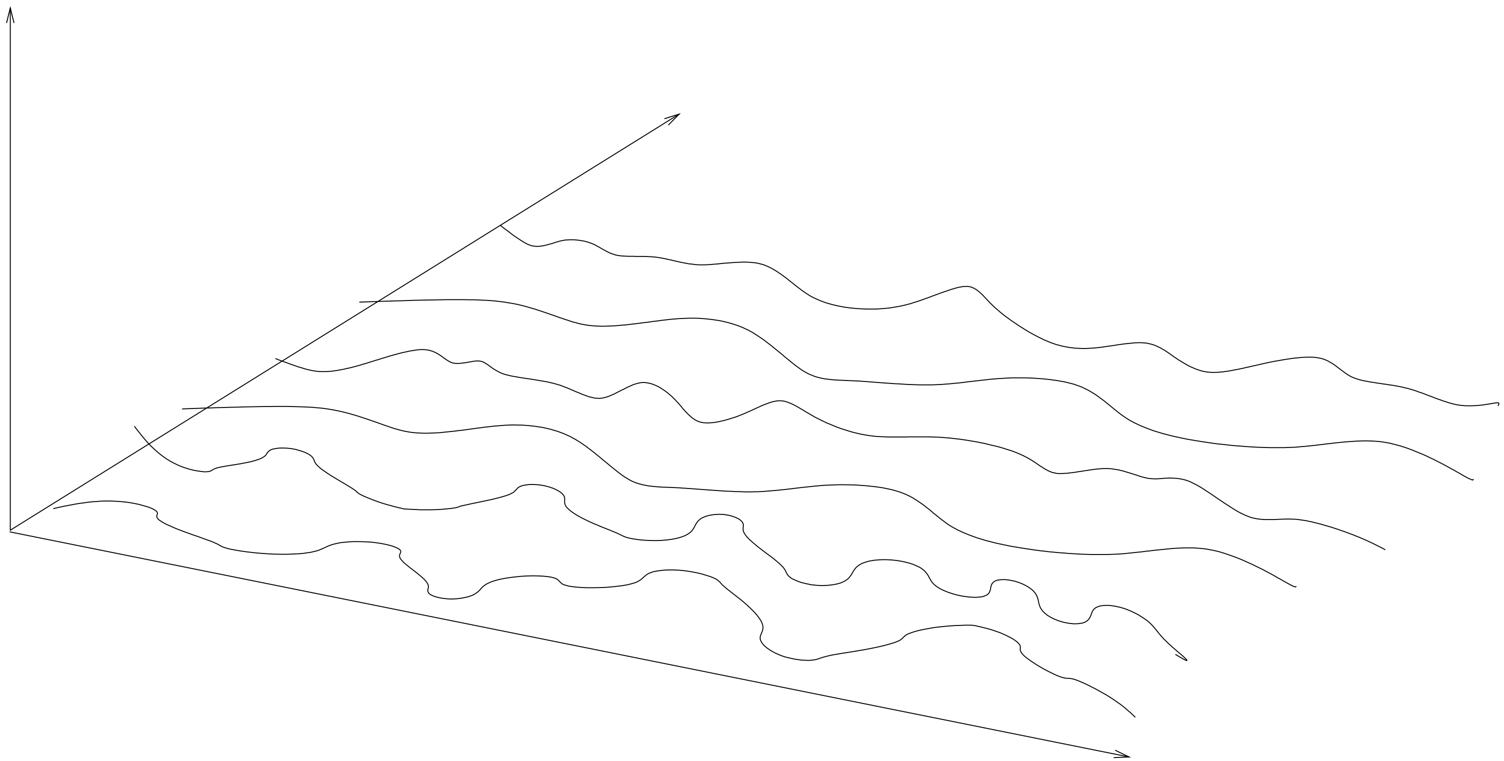
Random Processes

- If you fix ω and let t vary you get a **sample trajectory**.

- In this way, you can think of X as an ensemble of sample trajectories.

- Therefore, we will talk about both sample trajectory properties (continuity, differentiability, etc.) and ensemble properties (statistical).

A Random Process as an Ensemble of Sample Trajectories



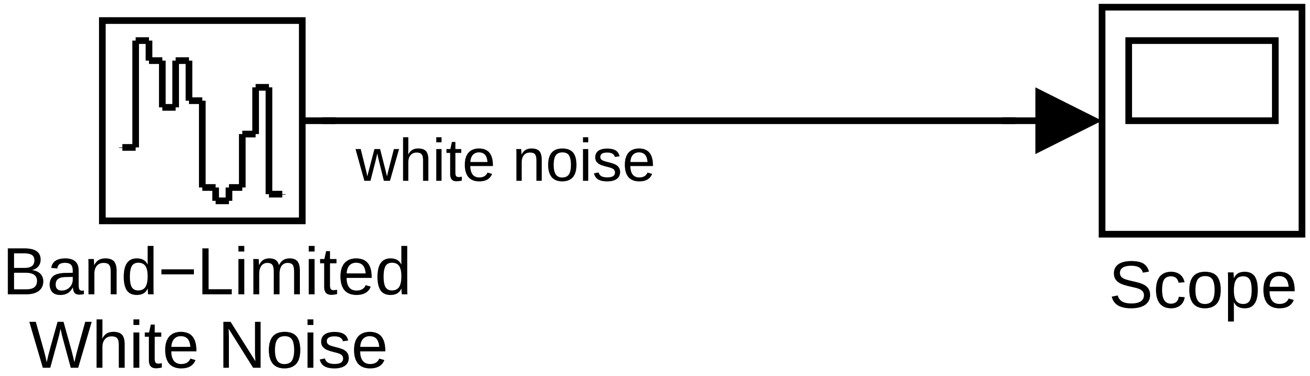
Examples of Random Processes

- A Discrete Process: Sampling with Replacement
  - A card player shuffles a deck of cards and then picks a card at random.
  - He enters the value of the card (1 for an ace through 13 for a king) in an Excel spreadsheet column.
  - He then replaces the card and repeats the process over and over again.
  - After 10 such experiments, he begins anew with a new Excel column. Each column represents a new sample trajectory.

Examples of Random Processes

- Matlab / Simulink Band-Limited White Noise
  - A single source block (band-limited white noise) and a single sink (the scope).
  - Each time the simulation is run, a new sample trajectory is generated.
  - Note that this discrete process is created by supposedly sampling a continuous process.

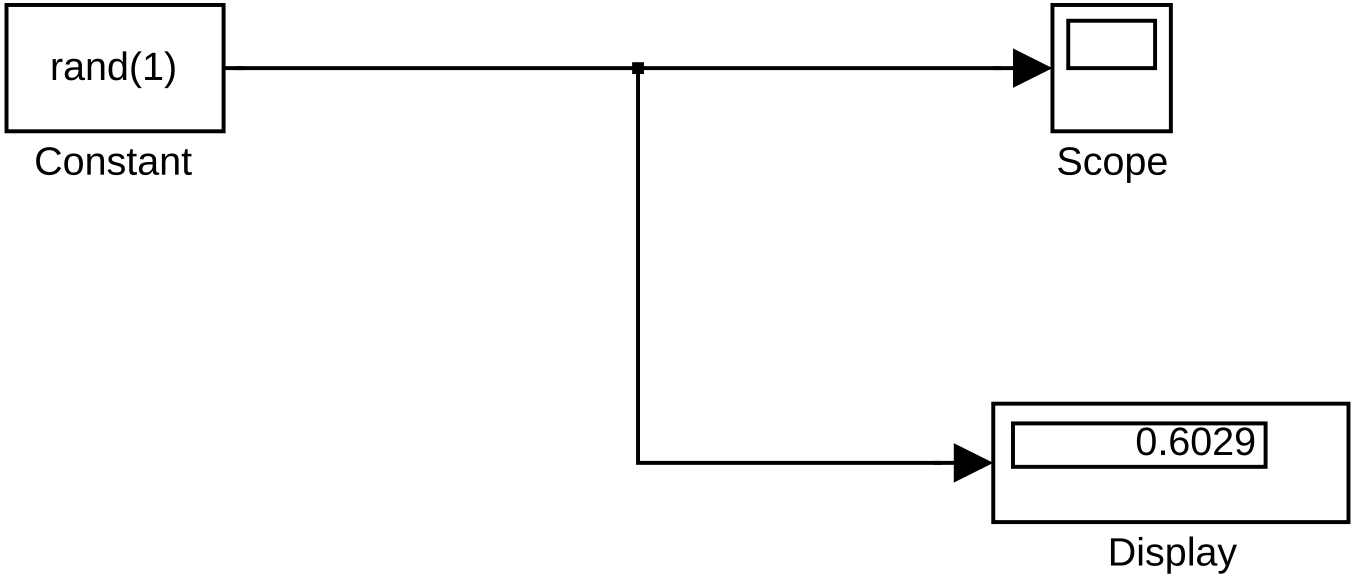
White Noise



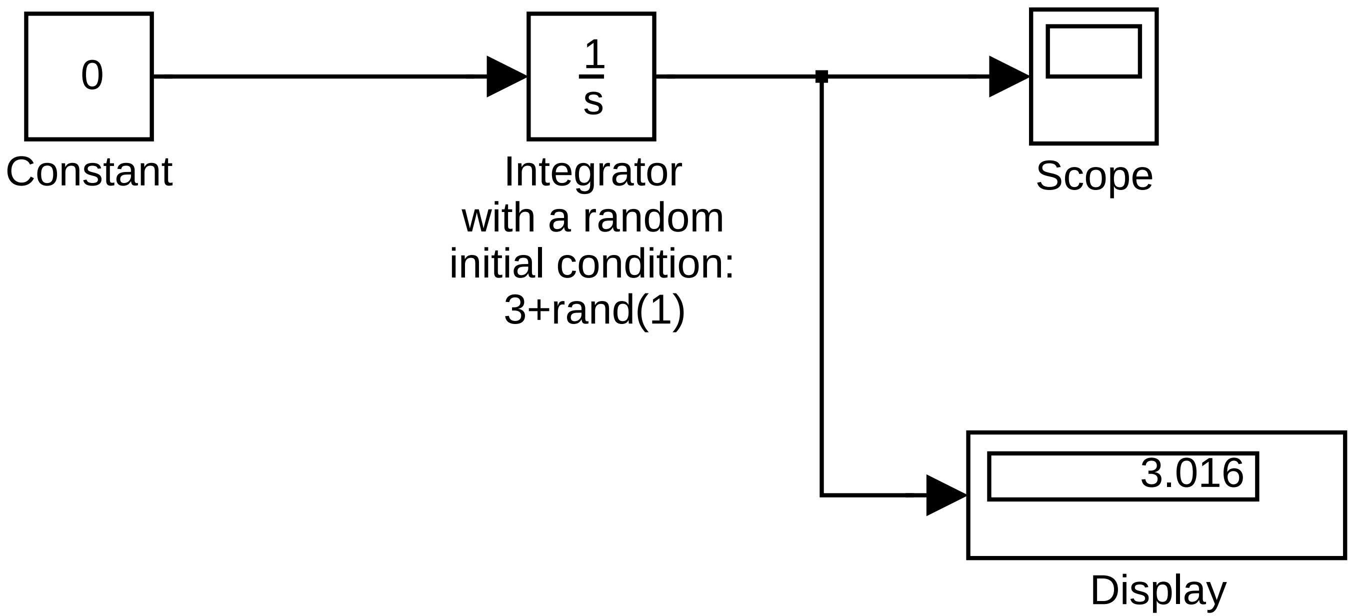
Examples of Random Processes

- Matlab / Simulink Random Constant
  - A single source block (constant block with value generated by random number generator) and a single sink (the scope).
  - Each time the simulation is run, a new sample trajectory is generated.
  - Note that this is a continuous process.

Random Constant



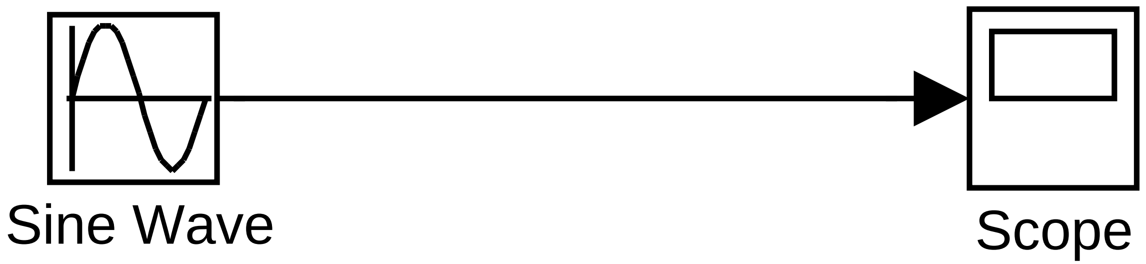
Random Constant



Examples of Random Processes

- Matlab / Simulink Sine Wave with Random Phase
  - A single source block (sine wave with random phase) and a single sink (the scope).
  - Each time the simulation is run, a new sample trajectory is generated.
  - Note that this is a continuous process.

Sine Wave with Random Phase



Stationary Random Processes

- A process X is called **strongly stationary** if for any finite set of times $\{t_1, \ldots, t_n\}$ and any $\tau > 0$ the two sets of random variables $\{X(t_1), \ldots, X(t_n)\}$ and $\{X(t_1+\tau), \ldots, X(t_n+\tau)\}$ have identical probability distributions.

- X is called **weakly stationary** if

$$\mathsf{E}\,[X(s)] = \mathsf{E}\,[X(t)]$$

  and

$$\mathsf{E}\,[X(s)X(t)] = \mathsf{E}\,[X(s+\tau)X(t+\tau)]$$

  for all appropriate $\tau > 0$ and all times s, t.

Correlation and Covariance

- Since at each t $X(t)$ is a random variable in its own right, it makes sense to talk about the correlation of two such random variables $X(t)$ and $X(s)$.

- Since both random variables come from the same underlying process, we refer to this as autocorrelation.

- Similarly, autocovariance involves first centering the process about its mean process μ_t.

- The function $\mathsf{E}[X(s)X(t)]$ is called the **autocorrelation function** of X and for a weakly stationary process X is denoted by $R_{XX}(\tau)$ since it is a function of τ alone.

- The **autocovariance** function for a weakly stationary process is formed by subtracting out the means: $C_{XX}(\tau) = \mathsf{E}\left[(X(t) - \mu_t)(X(t+\tau) - \mu_{t+\tau})\right]$.

- Cross correlation and cross covariance between two different random processes X and Y are defined in the obvious way.

Examples of Autocorrelation Functions

- Consider the random process $X(t,\omega) = A(\omega)\,\text{sin}(2\pi f t + \theta)$ in which $A : \Omega \to \mathbb{R}$ is a normally distributed random variable with variance σ^2.

- The autocorrelation function $R_{XX}(s,t)$ is therefore

$$
\begin{aligned}
\mathsf{E}\left[X(t,\omega)X(s,\omega)\right] &= \mathsf{E}\left[A(\omega)\,\text{sin}(2\pi f t + \theta)A(\omega)\,\text{sin}(2\pi f s + \theta)\right] \\
&= \mathsf{E}\left[A(\omega)^2\right]\text{sin}(2\pi f t + \theta)\,\text{sin}(2\pi f s + \theta) \\
&= \sigma^2\,\text{sin}(2\pi f t + \theta)\,\text{sin}(2\pi f s + \theta)
\end{aligned}
$$

Examples of Autocorrelation Functions

- Consider the white noise process $W(t)$.

- It is a hallmark of white noise to be uncorrelated in time.

- Thus, $\mathsf{E}[W(t)W(s)] = 0$ for every s and t such that $s \neq t$.

Ergodic Processes

- We shall be concerned almost exclusively with weakly stationary processes that are also **ergodic**, meaning that ensemble averaging at any one time gives the same result as time averaging one sample trajectory. Thus we may assume:

(11) $$\mu_X = \lim_{T\to\infty} \frac{1}{T} \int_0^T X(t)dt$$

(12) $$\psi_X^2 = \lim_{T\to\infty} \frac{1}{T} \int_0^T X^2(t)dt$$

(13) $$R_{XX}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T X(t)X(t+\tau)dt$$

Ergodic Processes

- We haven't yet defined want we mean by such time integrals, however, so we have to hold off dealing with them.

- Once we have defined the mean square limit of Riemann-Stieltjes sums, then we'll know what these integrals really mean.

- Nevertheless, if we fix ω and consider only sample trajectories $X(t, \omega)$, then we can use ordinary integrals provided that the sample trajectories are well-enough behaved.

Ergodic Processes

- Well-behaved means that they are of bounded variation, i.e., they don't wiggle around too much.

- Unfortunately, this doesn't really work for the usual conception of white noise, because its sample trajectories are not of bounded variation.

- If we restrict our attention to band-limited white noise (like Matlab / Simulink does), then we can stick to the ordinary integrals we learned in calculus.

Power Spectral Density

- The **power spectral density** (PSD) function is the Fourier transform of the autocorrelation function:

(14)
$$P_{XX}(\omega) = \int_{-\infty}^{\infty} R_{XX}(\tau)\,\text{exp}(-j\omega\tau)d\tau$$

where $j = \sqrt{-1}$ and ω is angular frequency in rad/sec (not an outcome of the sample space!).

- Units: If the process X has units of volts, say, then P_{XX} will have units of V^2/Hz or V^2 sec/rad. This is due to the fact that R_{XX} is a product of samples, each of which is in volts, and the transform kernel, exp$(-j\omega\tau)$ is the same as $\frac{1}{\text{exp}(j\omega\tau)}$.

Power Spectral Density

- Sometimes the definition in equation (14) is referred to as the autospectral density, since the autocorrelation function is involved.

- The cross-spectral density function is defined in the same way, but with R_{XY} substituted for R_{XX}.

Power Spectral Density

- Another way to obtain spectral density functions is through Fourier transforms of the original process data.

- However, we must change our notational convention for a while: We've been using uppercase letters such as X and Y to denote random processes and we've been using ω to denote a point of the sample space Ω.

- Now we are going to use lowercase letters such as $x(t)$ and $y(t)$ for the process time data and reserve the uppercase letters for their Fourier transforms. And ω will now stand for frequency in rad/sec.

Power Spectral Density

- The PSD can be expressed as follows:

$$P_{XY}(\omega) = \lim_{T\to\infty} \frac{1}{T} \mathsf{E}\left[X^*(\omega,T)Y(\omega,T)\right] \tag{15}$$

  where $X(\omega,T)$ and $Y(\omega,T)$ are the finite Fourier transforms of $x(t)$ and $y(t)$, respectively, and $X^*(\omega,t)$ is the complex conjugate of $X(\omega,t)$. ω is frequency in rad/sec.

- The finite Fourier transform of $x(t)$ is

$$X(\omega,T) = \int_0^T x(\tau)\,\mathrm{exp}(-j\omega\tau)d\tau$$

- Equation (15) Wiener-Khinchin equation.

Day 2, Segment 3
Some Important Random Processes
Topics

- White Noise
- Random Walk
- Random Constant
- Random Ramp
- Markov Processes

- State Space Augmentation

White Noise

- White noise can be defined as a process whose autocorrelation function is a Dirac delta function:

$$\delta(\tau) = \begin{cases} \infty & \text{if }, \tau = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

- There's only one mathematical problem: this is not a real-valued function.

- There's also a physical problem: such a process would have infinite mean-squared value and hence infinite energy.

White Noise

- There is a way to properly define the Dirac delta as a linear functional, not a function, and show that all of its desirable properties are preserved.

- However, we'll proceed as if this concept of white noise were well-defined and use the delta function's integration properties to define some other processes such as random walk.

- The hallmark of a white noise process W is that it is uncorrelated in time. This is the meaning of the statement that W has the Dirac function as its autocorrelation function.

White Noise Process

- The Matlab / Simulink approach to white noise is to use bandlimited white noise, which means that the bandwidth is not infinite as it is for pure white noise.

- This is valid as long as the bandwidth of the noise is sufficiently greater than that of the system you are dealing with.

- In Matlab, you can simulate a discrete white noise sequence by drawing a random number at each time step from a normal distribution.

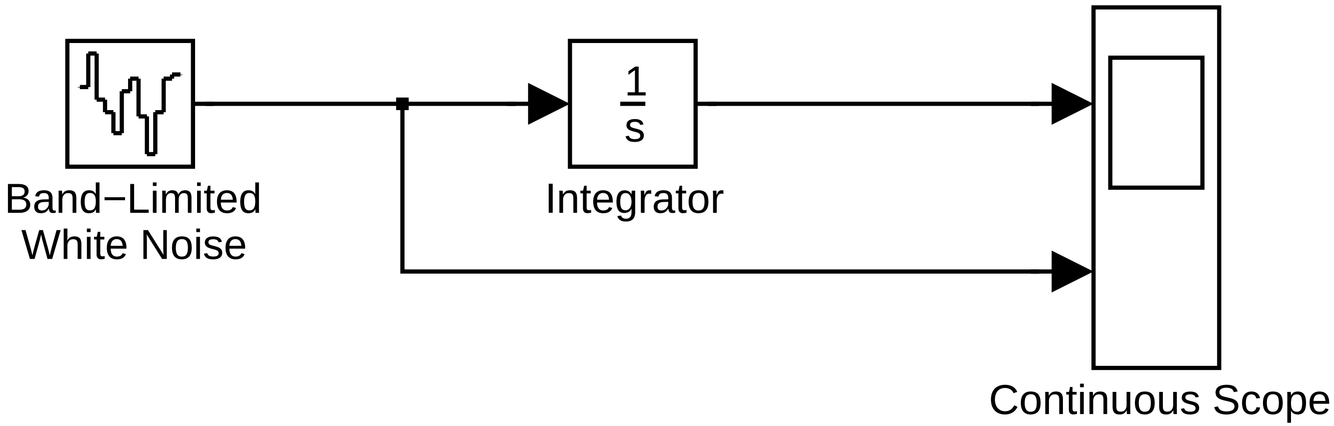
Random Walk

- In engineering applications, random walk is a discrete time process.

- Some authors (Gelb, for example) refer to random walk in the continuous case as the solution to a differential equation driven by white noise: $\dot{X} = W$. However, this would make W the derivative of the Wiener-Levy process, but that process is not differentiable.

- So we'll stick to the discrete case: $X_{k+1} = X_k + W_k$, where the W_k are normally distributed and statistically uncorrelated at distinct sample times.

Random Walk



Random Ramp and Random Constant

- The random ramp requires two state variables (to be discussed more later): a random constant (y-intercept) and the integral of a random constant (slope)

- A random constant is a process X such that the sample trajectories are constant with respect to time.

Random Ramp and Random Constant

- In principle, however, this constant value could vary with ω, i.e., across the ensemble, thus making the process non-Ergodic.

- Random constants are very useful in many applications to model turn-on to turn-on biases, for example.

Random Ramp and Random Constant

- See Matlab Simulink demos: rand_const.mdl and rand_ramp.mdl
- In modeling gyro biases, a combination of a random constant (bias stability) and random walk (bias drift) is common.
- Gyro scale factors are modeled similarly.
- Deterministic biases and scale factors are also part of the model, but these generally are calibrated out.

Second Order Processes

- We will generally restrict our attention to **second-order processes**, i.e., those such that $X(t) \in L^2(\Omega)$.

- Recall what this means:

$$\mathsf{E}\left[X^2(t)\right] < \infty$$

- This allows us to use Hilbert space techniques because, for every $s, t \in [0, T]$, $\mathsf{E}\left[X(s)Y(t)\right]$ is an inner product. Inner products, norms and metrics allow us to do calculus and geometry (orthogonal projection).

Gaussian Processes

- An important class of second-order processes are the Gaussian processes.
- Let $X : [0, T] \to L^2(\Omega)$ be a second-order process. X is said to be **Gaussian** if for every finite set of times $\{t_1, \ldots, t_n\}$ the vector of random variables $(X(t_1), \ldots, X(t_n))$ has a multivariate normal distribution.
- To fully appreciate what this means, we need to explain what a multivariate normal distribution is.

- A function $Q_C : \mathbb{R}^n \to \mathbb{R}$ of the form

$$Q_C(x) = x^\mathsf{T} C x$$

where $x \in \mathbb{R}^n$ is a column vector and C is a real symmetric matrix., is called a **quadratic form** associated with C. (I.e., $C = (c_{ij})$ and $c_{ij} = c_{ji}$).

- We will further assume that C is positive definite, which means that $Q_C(x) > 0$ for every non-zero vector x.

- Then $X_1, \ldots, X_n$ are said to have a **multivariate normal distribution** if their joint p.d.f. is of the form
(17)
$$f(x_1, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det(C)}} \exp\left[-\frac{1}{2}(x-\mu)^{\mathsf{T}} C^{-1}(x-\mu)\right]$$
for a real, positive definite symmetric matrix C and $\mu = (\mu_1, \ldots, \mu_n)$.

- C is called the **covariance matrix** and μ is called the **mean vector**.

- In terms of quadratic forms, this is

$$f(x_1, \ldots, x_n) = \sqrt{\frac{\det(C^{-1})}{(2\pi)^n}} \exp\left[-\frac{1}{2} Q_{C^{-1}}(x - \mu)\right]$$

- C^{-1} is sometimes called the information matrix, because its inverse, the covariance matrix, often expresses the uncertainty (or lack of information) associated with the random variables.

Gaussian Processes

- Thus a Gaussian process is a random process for which the joint distribution of every finite collection of range random variables is a Gaussian distribution.

- The Gaussian process is important in connection with the kinds of noise that drive the dynamic systems we are interested in.

- The linear Kalman filter is optimal only under assumptions that the process noise and the measurement noise are Gaussian.

Gauss-Markov Processes

- A Gauss-Markov process is for which the autocorrelation function has an exponential form:

(18) $$R_{XX}(\tau) = \sigma^2 \exp(-\beta\,|\tau|)$$

- In dynamic terms, this would correspond to passing white noise through first-order feedback system:

$$\dot{X} = -\beta X + W$$

- See Matlab Simulink file gauss_markov.mdl

Gauss-Markov Processes

- Discrete version of a Gauss-Markov process:

$$X_{k+1} = \exp\left(-\beta\Delta t\right) X_k + W_k$$

where $\Delta t = t_{k+1} - t_k$ and is assumed to be the same regardless of k.

State Space Augmentation

- Kalman filter theory assumes forcing function is white noise
- If this is not satisfied, then new states need to be added
- The most frequently used models can all be understood as differential equations (or difference equations) driven by white noise
- We'll see an example of state space augmentation later when we discuss filter maintenance

Exercises

1. Perform the following random walk experiment:
a. Create a graph with two axes marked off at integral (discrete) points.
b. Plot the first point at the origin.
c. Flip a coin. If the result is heads, move one step to the right and plot a point 1 unit above the previous y value. If the result is tails, move one step to the right and plot a point 1 unit below the previous y value.
d. Repeat step c ten times.

2. What is your estimate for how the variance of this process grows with time?

3. As a class exercise, we'll compute the ensemble averages across the class results at each point. Is this pocess ergodic?

Day 2, Segment 4
Kalman Statistics
Topics

- The Static Estimation Problem: Derivation of the Kalman Gain
- Covariance Matrices
- Process Noise
- Measurement Noise

- Kalman Assumptions
- Error Covariance Matrix (P)

The Static Estimation Problem: Derivation of the Kalman Gain

- Given: A known measurement vector $z = (z_1, z_2, \ldots, z_m)$ that is believed to be causally related to the state vector $x = (x_1, \ldots, x_n)$.

- Find: The "best" **estimate** of the state as a function of the measurements: $\widehat{x} = g(z) = g(z_1, \ldots, z_m)$.

The Static Estimation Problem

- Note that we are ignoring the state dynamics for now – hence the term "static."

- Since the problem for now is static, we do not have any stochastic differential equations to worry about. Hence, we don't have to use the mean square calculus.

- We are only concerned with the statistical problem of estimating a vector random variable x from a finite set of measurements.

The Static Estimation Problem: Why should we care?

- The solution to this problem will introduce concepts that are central to traditional Kalman filtering.

- The solution presented here is nearly identical to the one given by Kalman himself in his 1960 paper. Hence, the result is now classical.

- What we will be deriving is a matrix K, called the **Kalman gain**, that will be used to weight the measurement residual in order to update the prior state estimate.

The Static Estimation Problem (continued)

- What do we mean by "best" estimate?

- Need to define a measure of "goodness", i.e., a performance measure and maximize it.

- Actually, it turns out to be more convenient to define a measure of "badness" that we can minimize.
  - This is called a "loss function" or "penalty function".

The Static Estimation Problem (continued)

- Estimation Error: The difference between the truth and the estimate: $\tilde{x} = \widehat{x} - x$

- Estimation error is bad.
  - Our estimators should try to minimize this error.
  - At the very least, we should try to keep track of how well we're doing (i.e., should have an error covariance matrix).

The Static Estimation Problem (continued)

- Admissible Loss Functions:
  - nonnegative, real-valued function of the estimation error;
  - non-decreasing, i.e., the loss gets bigger as the estimation error gets bigger;
  - should be zero when there is no error, i.e., when we are certain;
  - should not be dependent on the sign of the error, i.e., should be a symmetric function.

- Note: As a function of a random variable, the loss function will yield a random variable. Hence, we'll want to minimize the mean of the loss function.

The Static Estimation Problem (continued)

- Estimation problem becomes: Find estimate $\hat{x}$ of x that minimizes the mean of the loss function L:
  $E\{L(\tilde{x})\} = E\{L(\hat{x} - x)\}$

- A common loss function is a quadratic function: the square of the norm of the error vector: $L(\tilde{x}) = \tilde{x}^\top \tilde{x}$
  Often referred to as the "square error."

- Therefore, we want to minimize the mean: $E\{L(\tilde{x})\} = E\{\tilde{x}^\top \tilde{x}\}$
  i.e., we want to minimize the "mean square error."

- Note that L is really a function of two variables: $L(\tilde{x}) = L(\widehat{x}, x)$.

The Static Estimation Problem (continued)

- Estimation problem becomes: Find estimate $\hat{x}$ of x that minimizes the integral:

(19) $$E\{\tilde{x}^\top \tilde{x}\} = \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{n+m} \tilde{x}^\top \tilde{x} f(x,z) dx_1 \cdots dx_n dz_1 \cdots dz_m$$

  where $f(x, z)$ is the joint probability density function for x and z.

- x is assumed to be an n-vector, while z is assumed to be an m-vector.

The Static Estimation Problem (continued) But we can factor the joint p.d.f, through the conditional p.d.f:

$$f(x, z) = f(x|z)f(z)$$

where

$f(z) = \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{n} f(x, z) dx_1 \cdots dx_n$ is the marginal p.d.f of z.

The Static Estimation Problem (continued) Therefore, we can modify Eq. (19), our integral expression for $E\{\tilde{x}^\top \tilde{x}\}$, as follows:

$$
\begin{aligned}
E\{\tilde{x}^\top \tilde{x}\} &= \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{n+m} \tilde{x}^\top \tilde{x} f(x,z) dx_1 \cdots dx_n dz_1 \cdots dz_m \\
&= \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{n+m} \tilde{x}^\top \tilde{x} f(x|z) f(z) dx_1 \cdots dx_n dz_1 \cdots dz_m \\
&= \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{m} f(z) \left[\underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{n} \tilde{x}^\top \tilde{x} f(x|z) dx_1 \cdots dx_n \right] dz_1 \cdots dz_m
\end{aligned}
$$

Note that we have pulled $f(z)$ out of the x-integral, because $f(z)$ does not depend on x.

The Static Estimation Problem (continued)

Notice that we now only need to minimize the integral within the brackets, i.e.,

$E\{\tilde{x}^\top \tilde{x}\}$ is minimized if and only if

$$\underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{n} \tilde{x}^\top \tilde{x} f(x|z) dx_1 \cdots dx_n \text{ is minimized.}$$

Remember that we are trying to find the value of $\hat{x}$ that will cause this loss function to be minimized.

The Static Estimation Problem (continued)

Let's expand $\tilde{x}^\top \tilde{x}$:

$$\begin{aligned}\tilde{x}^\top \tilde{x} &= (\hat{x} - x)^\top (\hat{x} - x) \\ &= \hat{x}^\top \hat{x} - \hat{x}^\top x - x^\top \hat{x} + x^\top x \\ &= \hat{x}^\top \hat{x} + x^\top x - 2\hat{x}^\top x\end{aligned}$$

The Static Estimation Problem (continued)

Thus, we expand the integral to be minimized accordingly:

$$
\begin{aligned}
\underbrace{\int_{-\infty}^{+\infty}\cdots}\tilde{x}^{\top}\tilde{x}f(x|z)dx_1\cdots dx_n = \hat{x}^{\top}\hat{x} \\
+ \underbrace{\int_{-\infty}^{+\infty}\cdots} x^{\top}xf(x|z)dx_1\cdots dx_n \\
- 2\hat{x}\underbrace{\int_{-\infty}^{+\infty}\cdots} xf(x|z)dx_1\cdots dx_n
\end{aligned}
$$

The Static Estimation Problem (continued)

This last expansion is justified because $\widehat{x}^\top\widehat{x}$ is a function of the z alone: $\widehat{x} = g(z_1, \ldots, z_m)$. The equation of interest has the form:

$$(20) \qquad \underbrace{\int_{-\infty}^{+\infty} \cdots}\tilde{x}^\top\tilde{x} f(x|z) dx_1 \cdots dx_n = f_1(\widehat{x}) + f_2(z) - 2f_3(\widehat{x}, z)$$

where $f_1(\widehat{x}) = \widehat{x}^\top\widehat{x}$, $f_2(z) = \underbrace{\int_{-\infty}^{+\infty} \cdots} x^\top x f(x|z) dx_1 \cdots dx_n$ and $f_3(\widehat{x}, z) = \widehat{x}^\top \underbrace{\int_{-\infty}^{+\infty} \cdots} f(x|z) dx_1 \cdots dx_n$.

The Static Estimation Problem (continued)

Now we need to minimize the right side of the last equation. To do so, we must take the gradient with respect to $\widehat{x}$, set it equal to zero, and solve for $\widehat{x}$ – a typical min/max problem. Recall that the gradient of the dot product is: $[\partial/\partial x]\,(y^\top x) = y$ and also $[\partial/\partial x]\,(x^\top x) = 2x$.

The Static Estimation Problem (continued)

Applying this to our problem, we note that the middle term of Eq. (20), $f_2(z)$ is a function of z alone and thus vanishes when the gradient with respect to $\widehat{x}$ is applied to it.
We are thus left with the first and third terms only:

$[\partial/\partial\widehat{x}]\,(\widehat{x}^\top\widehat{x}) = 2\widehat{x}$ and

$$[\partial/\partial\widehat{x}]\,(\widehat{x}^\top \underbrace{\int_{-\infty}^{+\infty}\cdots}\, x f(x|z)dx_1\cdots dx_n) = \underbrace{\int_{-\infty}^{+\infty}\cdots}\, x f(x|z)dx_1\cdots dx_n.$$

The Static Estimation Problem (continued)

Thus, taking the gradient with respect to $\widehat{x}$ of Eq. (20) we obtain

$$2\widehat{x} = 2 \underbrace{\int_{-\infty}^{+\infty} \cdots}_{} x f(x|z) dx_1 \cdots dx_n = 0 \qquad \text{or}$$

(21) $$\widehat{x} = \underbrace{\int_{-\infty}^{+\infty} \cdots}_{} x f(x|z) dx_1 \cdots dx_n$$

The Static Estimation Problem (continued)

This solution is just the conditional expectation for x, conditioned on the z:

$$\widehat{x} = \underbrace{\int_{-\infty}^{+\infty} \cdots}_{} x f(x|z) dx_1 \cdots dx_n = E\,[x|z].$$

The Static Estimation Problem (continued)

Now let's see what this looks like with Gaussian statistics for $f(x|z)$:

$$E[x|z] = m_x + P_{xz} P_{zz}^{-1} (z - m_z), \tag{22}$$

where m_x, m_z are the means of x and z, respectively, and P_{xz} is the cross-covariance matrix between x and z, while P_{zz} is the covariance matrix for z alone.

The Static Estimation Problem (continued)

In other words, we have for Eq. (22):

$$
\begin{aligned}
E\left[x|z\right] &= m_x + P_{xz}P_{zz}^{-1}(z - m_z) \\
m_x &= E[x] \\
m_z &= E[z] \\
P_{xz} &= E[(x - m_x)(x - m_z)^{\top}] \\
P_{zz} &= E[(z - m_z)(z - m_z)^{\top}]
\end{aligned}
$$

The Static Estimation Problem (continued)

Eq. (22) represents the essence of what will become the Kalman update algorithm. It will give us a new estimate of the state vector x conditioned on the actually observed measurements and all associated statistics.

Now consider again the relationship between x and z that we have previously assumed:

$$z = Hx + v$$

Remember that x and z are statistically independent of each other.

The Static Estimation Problem (continued)

Since $m_z = E[z] = E[Hx + v] = H \cdot E[x] + E[v] = Hm_x + m_v$, we have

$$\begin{aligned}
P_{xz} &= E[(x - m_x)(z - m_z)^\top] \\
&= E[(x - m_x)\{Hx + v - (Hm_x + m_v)\}^\top] \\
&= E[(x - m_x)\{(Hx - Hm_x) + (v - m_v)\}^\top] \\
&= E[(x - m_x)\{H(x - m_x) + (v - m_v)\}^\top] \\
&= E[(x - m_x)\{(x - m_x)^\top H^\top + (v - m_v)^\top\}] \\
&= E[(x - m_x)(x - m_x)^\top H^\top] + E[(x - m_x)(v - m_v)^\top] \\
&= E[(x - m_x)(x - m_x)^\top H^\top]
\end{aligned}$$

because x,z independent $\Rightarrow E[(x - m_x)(v - m_v)^\top] = 0$.

The Static Estimation Problem (continued)

But, $E[(x - m_x)(x - m_x)^\top] = P_{xx}$ by definition.

Thus

$$\begin{aligned} P_{xz} &= E[(x - m_x)(x - m_x)^\top H^\top] \\ &= P_{xx} H^\top \end{aligned}$$

A similar derivation shows that

$$P_{zz} = H P_{xx} H^\top + P_{vv}$$

Later we'll see this in the form of $H P_{xx} H^\top + R$ since $P_{vv} = R$ by definition.

The Static Estimation Problem (continued)

Inserting these last two results into Eq. (22) yields

$$\begin{aligned} E\left[x|z\right] &= m_x + P_{xz}P_{zz}^{-1}(z - m_z) \\ &= m_x + P_{xx}H^\top[HP_{xx}H^\top + P_{vv}]^{-1}(z - m_z) \\ &= m_x + P_{xx}H^\top[HP_{xx}H^\top + P_{vv}]^{-1}(z - Hm_x) \end{aligned}$$

This last step results because
$m_z = E[z] = E[Hx + v] = H \cdot E[x] + E[v] = Hm_x + 0 = Hm_x$
by definition.

The Static Estimation Problem (continued)
Kalman Gain

Eq. (22) now becomes the central result of classical Kalman Filter theory:

(23) $$\widehat{x} = E[x|z] = m_x + K(z - Hm_x)$$

where $K = P_{xx}H^{\top}[HPH^{\top} + P_{vv}]^{-1}$ is the **Kalman gain**.

Covariance Matrices

- There are several covariance matrices associated with the Kalman filter: P_k, Q_k, R_k

- In general, a covariance matrix is associated with a random vector X as in the following definition:

$$C_X = \mathsf{E}\left[XX^{\mathsf{T}}\right]$$

Covariance Matrices

- This is precisely how Q_k and R_k are defined:

$$Q_k = \mathsf{E}\left[w_k w_k^{\mathsf{T}}\right]$$

and

$$R_k = \mathsf{E}\left[v_k v_k^{\mathsf{T}}\right]$$

- $P_k(+)$, on the other hand is called the error covariance, because it is associated with $\tilde{x}_k$

$$P_k(+) = \mathsf{E}\left[\tilde{x}_k(+)\tilde{x}_k^{\mathsf{T}}(+)\right]$$

where $\tilde{x}_k = \widehat{x}_k(+) - x_k$.

Process Noise

- The process noise w_k is assumed to be a white, zero-mean, Gaussian sequence.

- Although the continuous form of white noise does not exist, it is sufficient in the discrete case to consider a random process sequence w_k that is uncorrelated from step to step: $\mathsf{E}\left[w_j w_k^{\mathsf{T}}\right] = 0$ if $j \neq k$.

- This can be achieved if w_k results from sampling a Gauss-Markov process whose exponential autocorrelation function rapidly approaches 0 before the next sample is taken.

Process Noise

- The process noise covariance affects the Kalman gain indirectly through the covariance propagation equation:

$$P_k(-) = \Phi_k P_{k-1}(+)\Phi_k^{\mathsf{T}} + Q_k$$

- This updated error covariance is then included in both the "numerator" and "denominator" of the Kalman gain:

$$K_k = P_k(-)H_k^{\mathsf{T}}\left[H_k P_k(-)H_k^{\mathsf{T}} + R_k\right]^{-1}$$

- If the process noise covariance is large relative to the measurement noise then the gain will be nearly unity, the extrapolated state estimate will cancel out and the updated state estimate will be nearly identical to the measurement itself.

Measurement Noise

- The measurement noise covariance matrix R_k indicates the amount of spread or dispersion and internal cross covariance within the measurement itself at time k.

- The diagonal elements of R_k give the expected variance in each of the individual measurements.

Measurement Noise

- The off-diagonal elements indicate the cross covariance between noise random variables associated with each individual measurement.

- If the noise random variables of individual measurements are uncorrelated, then R_k will be a diagonal matrix.

Comparison of Process Noise and Measurement Noise

- If the process noise covariance is large relative to the measurement noise then the gain will be nearly unity, the extrapolated state estimate will cancel out and the updated state estimate will consist essentially only of the measurement itself.

- This can be seen as follows in the special case that H_k is square and invertible:

$$
\begin{aligned}
\hat{x}_k(+) &= \hat{x}_k(-) + K_k\left[z_k - H_k\hat{x}_k(-)\right] \\
&\approx \hat{x}_k(-) + P_k(-)H_k^{\mathsf{T}}\left[H_kP_k(-)H_k^{\mathsf{T}}\right]^{-1}\left[z_k - H_k\hat{x}_k(-)\right] \\
&= \hat{x}_k(-) + P_k(-)H_k^{\mathsf{T}}\left[H_k^{\mathsf{T}}\right]^{-1}P_k(-)^{-1}H_k^{-1}\left[z_k - H_k\hat{x}_k(-)\right] \\
&= \hat{x}_k(-) + H_k^{-1}z_k - \hat{x}_k(-) \\
&= H_k^{-1}z_k
\end{aligned}
$$

Comparison of Process Noise and Measurement Noise

- On the other hand, when R_k dominates Q_k (and thus $P_k(-)$), the gain is nearly 0, and the new updated estimate is simply the previously extrapolated estimate:

$$\begin{aligned}
\widehat{x}_k(+) &= \widehat{x}_k(-) + K_k\left[z_k - H_k\widehat{x}_k(-)\right] \\
&\approx \widehat{x}_k(-) + 0\left[z_k - H_k\widehat{x}_k(-)\right] \\
&= \widehat{x}_k(-)
\end{aligned}$$

Error Covariance Matrix

- P is the **estimation error covariance**, although it is commonly referred to simply as THE covariance matrix.

- $P_k(+) = \mathsf{E}\left[\tilde{x}_k(+)\tilde{x}_k^{\mathsf{T}}(+)\right]$

- The report card for the Kalman filter

- Contains most of the information about how well, or poorly your filter is doing.

Error Covariance Matrix

- P brings Q and R together

- In continuous framework, P is solution to matrix Riccatti differential equation:

(24) $$\dot{P} = FP + PF^{\mathsf{T}} + Q - PH^{\mathsf{T}}R^{-1}HP$$

- Covariance analysis is primarily concerned with solving this equation or its discrete equivalent

Kalman Assumptions

- Your initial state estimate $\hat{x}_0$ is not too far off
- Your initial error covariance estimate P_0 is not too far off
- Process noise is truly white, i.e., uncorrelated in time
- Measurement noise is truly white, ie., uncorrelated in time
- Measurement noise and process noise are uncorrelated with each other

Kalman Assumptions: Measurement Noise and Process Noise Are Uncorrelated

- The Kalman assumptions require that the process noise and measurement noise be uncorrelated

  $$\mathsf{E}\left[w_j v_k^{\mathsf{T}}\right] = 0$$

  for all j and k.

- If this turns out not to be the case, then you have to re-model your filter, perhaps using state space augmentation. Otherwise, your filter could diverge.

Kalman Assumptions: Measurement Noise and Process Noise Are Uncorrelated: An Exception

- It is possible to allow for the case:

$$\mathsf{E}\left[w_j v_k^{\mathsf{T}}\right] = C_j$$

  when $j = k$ (and 0 otherwise).

- This leads to modification of the Kalman equations which incorporate the C_j cross covariance matrix.

Kalman Assumptions: Linearity

- System dynamics are linear
  - This means that to use Kalman filtering in non-linear situations, some sort of linearization has to introduced
- Relationship of measurement to state is linear
  - Again, if this is not the case, then some kind of linearization has to take place

Exercises

- Show that any covariance matrix C symmetric, i.e., that $C^{\mathsf{T}} = C$.

- Suppose that λ is an eigenvalue of P corresponding to the eigenvector v, i.e., $Pv = \lambda v$. Prove that $\lambda > 0$. Hint examine the quadratic form induced by P: $v^{\mathsf{T}} P v$.