Chapter 5
Optimal Estimation

Part 2

5.2 Covariance Propagation and Optimal Estimation
Outline

• 5.2 Random Variables, Processes and Transformation
  – 5.2.1 Variance of Continuous Integration and Averaging Processes
  – 5.2.2 Stochastic Integration
  – 5.2.3 Optimal Estimation
  – Summary
Outline

• 5.2 Random Variables, Processes and Transformation
  – 5.2.1 Variance of Continuous Integration and Averaging Processes
  – 5.2.2 Stochastic Integration
  – 5.2.3 Optimal Estimation
  – Summary
State Estimation

• Henceforth, reinterpret our “transformations” of uncertainty to cover recursive relationships.

• Our goal is a set of recursive algorithms to track the state $x$, and its uncertainty $P$, of a dynamical system.

• Define:
  
  – $x_k$: state estimate at time $k$
  
  – $P_k$: (state) covariance estimate at time $k$
  
  – $z_k$: measurement at time $k$
  
  – $R_k$: (measurement) covariance estimate at time $k$
5.2.1.2 Recursive Integration

- Recall the result for a sum of iid RVs and reinterpret the “summing” as integration as it occurs in dead reckoning. In our new notation:

- The summing process can be written:

\[ x_k = \sum_{i=1}^{k} z_i = \sum_{i=1}^{k-1} z_i + z_k = x_{k-1} + z_k \]

- Its covariance can be written recursively as:

\[ \sigma_k^2 = \sum_{i=1}^{k} \sigma_i^2 = \sum_{i=1}^{k-1} \sigma_i^2 + \sigma_i^2 = \sigma_{k-1}^2 + \sigma_z^2 \]

\[ \sigma_x^2 = n \sigma_z^2 \]

Variance grows linearly wrt time.
5.2.1.3 Variance of a Continuous Summing Process

Standard Deviation grows linearly wrt square root of time.
5.2.1.4 Recursive Averaging

- The averaging process can be written as:

\[ \hat{x}_k = \frac{1}{k} \sum_{i=1}^{k} z_i = \frac{1}{k} \left( \sum_{i=1}^{k-1} z_i + z_k \right) = \frac{1}{k} \left( \sum_{i=1}^{k-1} z_i + z_k \right) \]

- Isolate the last estimate:

\[ \hat{x}_k = \frac{1}{k} \left[ (k-1)\hat{x}_{k-1} + z_k \right] = \frac{1}{k} \left[ (k-1)\hat{x}_{k-1} + \hat{x}_{k-1} + z_k - \hat{x}_{k-1} \right] \]

- Simplifies to the recursive form:

\[ \hat{x}_k = \hat{x}_{k-1} + \frac{1}{k} (z_k - \hat{x}_{k-1}) \]

- Define the “Kalman Gain” K=1/k:

\[ \hat{x}_k = \hat{x}_{k-1} + K (z_k - \hat{x}_{k-1}) \]
5.2.1.4 Recursive Averaging

- Recall the result for an average of iid RVs. For $k$ measurements:

$$\frac{\sigma_{k}^2}{k} = \frac{1}{k^2} \sum_{i=1}^{k} \sigma_{z}^2 = \frac{1}{k} \sigma_{z}^2$$

- Note that:

$$\sigma_{k}^2 = \frac{\sigma_{z}^2}{k} \Rightarrow \frac{1}{\sigma_{k}^2} = \frac{k}{\sigma_{z}^2} = \frac{k-1}{\sigma_{z}^2} + \frac{1}{\sigma_{z}^2}$$

- Which means:

$$\frac{1}{\sigma_{k}^2} = \frac{1}{\sigma_{k-1}^2} + \frac{1}{\sigma_{z}^2}$$

- So, variances add by reciprocals, just like conductances in electric circuits.
5.2.1.4 Recursive Averaging

- Now because:
  \[ \sigma_k^2 = \frac{\sigma_z^2}{k} \]
  \[ \sigma_{k-1}^2 = \frac{\sigma_z^2}{k-1} \]

- Substitute to get:
  \[ \frac{\sigma_k^2}{\sigma_{k-1}^2} = \frac{k-1}{k} \Rightarrow \sigma_k^2 = \left(\frac{k-1}{k}\right) \sigma_{k-1}^2 = \left(1 - \frac{1}{k}\right) \sigma_{k-1}^2 \]

- Substituting the Kalman gain (and adding 1 to k):
  \[ \sigma_{k+1}^2 = (1 - K)\sigma_k^2 \]
5.2.1.3 Variance of a Continuous Averaging Process

Standard Deviation Decreases linearly wrt square root of time.
5.2.1.5 Measuring “Stability”

• Refers to changes in effective (average) bias and scale errors.
  – Often quoted as change in bias or scale as a function of temperature or time.

  Somehow, bias instability means the same as bias stability in this context.

Bias is unrelated to noise amplitude

This gyro has about 0.4 deg/sec peak-to-peak variation.

Average bias is < 20 deg/hr.

![Figure 1. VG700CA rate output](http://www.xbow.com)
Allan Variance (Measure of Bias Stability)

- **Average** all measurements over some time period $\Delta t$.
- Asks how much the average (over $\Delta t$) can change over a period of time $\Delta t$.
- Take difference in average in successive bins. Square it.
- Add up at least 9 of these and divide by $2(n-1)$

$$AVAR^2(\tau) = \frac{1}{2 \cdot (n-1)} \sum_i \left( y(\tau)_{i+1} - y(\tau)_i \right)^2,$$

Intuitively:
- Variance in the Bias
- For given level of averaging

**What is the 2 for?**
Allan Variance

- Compute the difference in average of successive bins.

\[ y(t)_i \quad y(t)_{i+1} \]

\[ \Delta t \]
Allan Variance

- Variance drops initially as $\Delta t$ increases (effect of averaging).
  - Sensor noise dominates for small $\Delta t$ (does not average out).
  - Rate random walk dominates for large $\Delta t$ (bias really is changing).

Inertial sensor manufacturers quote the minimum (equals best achievable result with active bias estimation and fully modeled sensor)
Allan Deviation Graph

- Angle Random Walk
- Rate Random Walk
- Bias
- Stability

Allan Deviation (deg/sec.) vs. Tau (sec.)
Outline

• 5.2 Random Variables, Processes and Transformation
  – 5.2.1 Variance of Continuous Integration and Averaging Processes
  – 5.2.2 Stochastic Integration
  – 5.2.3 Optimal Estimation
  – Summary
So What?: Recursive Form

- Recursive processes are also of the form $y = f(x_1, x_2)$.
- Let $y$ mean the new value $x_{i+1}$ of some state variables that we are trying to estimate.
- Let $x_1$ mean the last estimate of state $x_i$.
- Let $x_2$ mean the inputs $u_i$ that are required to compute the state.

$$y = f(x_1, x_2)$$

$$x_{i+1} = f(x_i, u_i)$$
5.2.2.1 Discrete Stochastic Integration

• Recall our result for covariance of a partitioned state vector:

  \[ \Sigma_y = J_1 \Sigma_{11} J_1^T + J_2 \Sigma_{22} J_2^T \]

• In our new notation, this becomes:

  \[ \Sigma_{i+1} = J_x \Sigma_i J_x^T + J_u \Sigma_u J_u^T \]

• In a more standard notation:

  \[ P_{i+1} = \Phi P_i \Phi^T + \Gamma Q_i \Gamma^T \]

This is one of the Equations of the Kalman Filter

Covariance Propagation in any Decorrelated Estimation Process.
5.2.2.2 Example: Dead Reckoning
(With Odometer Error Only)

\[
\begin{align*}
\text{State:} \quad & x_i = [x_i, y_i]^T \\
\text{Measurements:} \quad & u_i = [l_i, \theta_i]^T \\
\text{Update:} \quad & x_{i+1} = f(x_i, u_i) = \\
& \begin{bmatrix}
  x_i + l_i \cos(\theta_i) \\
  y_i + l_i \sin(\theta_i)
\end{bmatrix}
\end{align*}
\]
5.2.2.2 Example: Dead Reckoning

(Jacobian)

Linearize: \[ x_{i+1} = f(x_i, u_i) = \begin{bmatrix} x_i + l_i \cos(\psi_i) \\ y_i + l_i \sin(\psi_i) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} l_c_i \\ l_s_i \end{bmatrix} \]

\[ \Phi_i = \frac{\partial x_i}{\partial x_i} + 1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \]

\[ \Gamma_i = \frac{\partial x_i}{\partial u_i} + 1 = \begin{bmatrix} c_i & -l_i s_i \\ s_i & l_i c_i \end{bmatrix} \]

Jacobians are functions of the present estimate and the present measurements.
5.2.2.2 Example: Dead Reckoning

(Input Uncertainty)

• The uncertainty in the current position and measurements is:

\[ P_i = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} \]

\[ Q_i = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & 0 \end{bmatrix} \]

Assume a perfect compass.

Assume Decorrelated errors.
5.2.2.2 Example: Dead Reckoning

\( P_{i+1} = \Phi P_i \Phi^T + \Gamma_i Q_i \Gamma_i^T \rightarrow P_{i+1} = P_i + \begin{bmatrix} c_i^2 \sigma_1^2 & c_i s_i \sigma_1^2 \\ c_i s_i \sigma_1^2 & s_i^2 \sigma_1^2 \end{bmatrix} \)

Note: Trace of \( P_{i+1} \) increases monotonically
5.2.2.3.1 Variance of a Continuous Random Walk

- Recall that for $n$ summed iid random variables:
  \[ \sigma_y^2 = n \sigma_x^2 \]
- Suppose the $x$'es were velocities at time 1, 2, ..., $n$. Then:
  \[ n = \frac{t}{\Delta t} \quad \sigma_y^2 = \frac{\sigma_x^2 t}{\Delta t} \]
- But this means that $\sigma_y^2 \to \infty$ as $\Delta t \to 0$ !!!
- That cannot be right; it would require infinite power.
- What is more realistic is variance that grows linearly wrt time:
  \[ \sigma_x^2(t) = \sigma_z^2 \cdot t \]
5.2.2.3.2 Integrating Stochastic Differential Equations

- Lets reinterpret our perturbative differential equation so mean a DE driven by random noise.

\[ \delta\dot{x}(t) = F(t)\delta\dot{x}(t) + G(t)\delta u(t) \]

- Define the covariances:

\[ \text{Exp}(\delta x(t)\delta x(t)^T) = P(t) \]
\[ \text{Exp}(\delta u(t)\delta u(\tau)^T) = Q(t)\delta(t - \tau) \]

- We might be tempted to solve this using the vector convolution integral:

\[ \delta x(t) = \Phi(t, t_0)\delta x(t_0) + \int_{t_0}^{t} \Gamma(t, \tau)\delta u(\tau)d\tau \]

\[ \Gamma(t, \tau) = \Phi(t, \tau)G(\tau) \]

White noise is uncorrelated in time.

Reimann says this integral does not converge.
Motivation for Stochastic Calculus

• An integral is a limit of a sum of products.
• The limit exists when the wiggles go away when you zoom in on a function:

For a white random signal, autocorrelation is zero, and the wiggles never go away at any zoom level.
• The integral or derivative of a white signal is meaningless.
  – So what is “stochastic calculus”? 
Deterministic Statistics

- The statistics of a distribution of a random variable are deterministic quantities.
- i.e. $s$ has a time derivative because $s$ is not random.

We will write differential equations for the statistics, not the random signals.

- Individual random walk signal
- Variance of a zillion random walk signals
5.2.2.3.2 Integrating Stochastic Differential Equations

• **Recall**: we cannot integrate the following because it fails the Reimann condition.

\[ \delta x(t) = \Phi(t, t_0)\delta x(t_0) + \int_{t_0}^{t} \Gamma(t, \tau)\delta y(\tau) d\tau \]

• **Trick**: Introduce a differential random walk process:

\[ d\beta(\tau) = \delta y(\tau) d\tau \]

• **Now, integrate the following**: 

\[ \delta x(t) = \Phi(t, t_0)\delta x(t_0) + \int_{t_0}^{t} \Gamma(t, \tau)d\beta(\tau) \]
5.2.2.3.2 Integrating Stochastic Differential Equations

• The integral of (squared expectation) of the last result is:

\[
P(t) = \Phi(t, t_0)P(t_0)\Phi(t, t_0)^T + \int_{t_0}^{t} \Gamma(t, \tau)Q(\tau)\Gamma(t, \tau)^T d\tau
\]
5.2.2.3.4 Linear Variance Equation

• We can differentiate the last result to find the differential equation that is satisfied by:
  – The covariance matrix of a dynamical system
  – Driven by white noise

\[
\dot{P}(t) = F(t)P(t) + P(t)F(t)^T + G(t)Q(t)G(t)^T
\]

This term usually leads to unbounded growth
Outline

• 5.2 Random Variables, Processes and Transformation
  – 5.2.1 Variance of Continuous Integration and Averaging Processes
  – 5.2.2 Stochastic Integration
  – 5.2.3 Optimal Estimation
  – Summary
5.2.3.1 Maximum Likelihood Estimation

• Consider the problem of **optimally estimating state from a series of measurements**:
  
  – Let \( x \in \mathbb{R}^n \) denote the state and \( z \in \mathbb{R}^m \) denote the measurements.
  
  – Measurements relate to the state by a measurement matrix:

\[
  z = Hx + v \quad \text{where} \quad v \sim N(0, R)
\]

  – The measurements are assumed to be corrupted by a noise vector of covariance:

\[
  R = \text{Exp}(v v^T)
\]
5.2.3.1 Maximum Likelihood Estimation

- The innovation $\overline{z} - H\overline{x}$ ($= v$) is Gaussian by assumption, so...
- The probability of getting a measurement $\overline{z}$ when the true state is $\overline{x}$ is:
  \[
  p(\overline{z}|\overline{x}) = \frac{1}{(2\pi)^{m/2}|R|^{1/2}} \exp\left[-\frac{1}{2}(\overline{z} - H\overline{x})R^{-1}(\overline{z} - H\overline{x})^T\right]
  \]
- This exponential will be maximized when the form in the exponent (without negative sign) is minimized:
  \[
  \hat{x}^* = \text{argmin}_{\overline{x}} \left(\frac{1}{2}(\overline{z} - H\overline{x})R^{-1}(\overline{z} - H\overline{x})^T\right)
  \]
- If the system is overdetermined, the solution is simply the weighted left pseudoinverse:
  \[
  \hat{x}^* = \left(H^T R^{-1} H\right)^{-1}H^T R^{-1} \overline{z}
  \]
5.2.3.1.1 Covariance of the MLE Estimate

• The weighted left pseudoinverse is just a function that maps $\underline{z}$ onto $\hat{\underline{x}}^*$, so let's define its Jacobian:

$$J_z = (H^T R^{-1} H)^{-1} H^T R^{-1}$$

• Therefore the covariance of the MLE result is:

$$\Sigma_{xx} = J_z \Sigma_{zz} J_z^T = (H^T R^{-1} H)^{-1} H^T R^{-1} RR^{-1} H(H^T R^{-1} H)^{-1}$$

$$\Sigma_{xx} = J_z \Sigma_{zz} J_z^T = (H^T R^{-1} H)^{-1} H^T R^{-1} H(H^T R^{-1} H)^{-1}$$

• Which simplifies to:

$$\Sigma_{xx} = (H^T R^{-1} H)^{-1}$$

Equation 5.80

• Note that this expression appears in the pseudoinverse:

$$\hat{\underline{x}}^* = (H^T R^{-1} H)^{-1} H^T R^{-1} \underline{z} = \Sigma_{xx} H^T R^{-1} \underline{z}$$
5.2.3.2 Recursive Estimation of a Random Scalar

• Suppose:
  – Present state estimate $x$ has variance $\sigma_x^2$
  – Measurement $z$ has variance $\sigma_z^2$
  – Want to get new state estimate $x'$ and its variance $\sigma_x'^2$

• The trick to derive a Kalman filter is to pretend the present estimate comes in as a measurement with the same covariance.

• The measurement relationship for (both) measurements is:

$$
\begin{bmatrix}
  z \\
  x
\end{bmatrix} =
\begin{bmatrix}
  1 \\
  1
\end{bmatrix} x'
$$
5.2.3.2 Recursive Estimation of a Random Scalar

- That means the associated measurement and covariance matrices are:

\[ H' = \begin{bmatrix} 1 & 1 \end{bmatrix}^T \quad R' = \text{diag} \begin{bmatrix} \sigma_z^2 & \sigma_x^2 \end{bmatrix} \]

- So, the weighted least squares solution is:

\[
x' = (H'^T R'^{-1} H')^{-1} H'^T R'^{-1} z = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_z^2} & 0 \\ 0 & \frac{1}{\sigma_x^2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_z^2} & 0 \\ 0 & \frac{1}{\sigma_x^2} \end{bmatrix} z
\]
5.2.3.2 Recursive Estimation of a Random Scalar

- This simplifies to:

\[ x' = \left( \frac{1}{\sigma_z^2} + \frac{1}{\sigma_x^2} \right)^{-1} \left( \frac{1}{\sigma_z^2} + \frac{1}{\sigma_x^2} \right) \]

Equation 5.8.1

- And, the uncertainty in the new estimate is (from Equation 5.80):

\[ \sigma_{x'}^2 = (H' R^{-1} H')^{-1} = \left( \frac{1}{\sigma_z^2} + \frac{1}{\sigma_x^2} \right)^{-1} \]

- Which is the same as saying the new information is the sum of that of the measurement and state:

\[ \frac{1}{\sigma_{x'}^2} = \frac{1}{\sigma_z^2} + \frac{1}{\sigma_x^2} \]

Equation 5.8.3
5.2.3.3 Example: Estimating Temperature from Two Sensors

• An ocean-going robot has to measure water temperature using two sensors.

• One of the measurements is $z_1 = 4$ with variance $\sigma_{z_1}^2 = 2^2$. Therefore $p(x|z_1)$ is as shown:
5.2.3.3 Example: Estimating Temperature from Two Sensors

• Suppose the other measurement is $z_2 = 6$ with variance $\sigma_{z_2}^2 = (1.5)^2$. Therefore $p(x|z_2)$ is as shown:

![Temperature Distribution Graph](image.png)
5.2.3.3 Example: Estimating Temperature from Two Sensors

- Application of Eqns 5.81 and 5.83 gives:

\[ x' = 5.28 \quad \sigma_{x'}^2 = (1.2)^2 \]

- The result is denoted graphically as follows:
5.2.3.2 Recursive Estimation of a Random Vector

• Suppose:
  – Present state estimate $x$ has variance $P$
  – Measurement $z$ has variance $R$
  – Want to get new state estimate $x'$ and its variance $P'$

• The measurement relationship for (both) measurements is:

\[
\begin{bmatrix}
  z \\
  x
\end{bmatrix} = \begin{bmatrix}
  H \\
  I
\end{bmatrix} \begin{bmatrix}
  x' \\
  x'
\end{bmatrix} = H'x'
\]
5.2.3.2 Recursive Estimation of a Random Vector

• That means the associated measurement and covariance matrices are:

\[ H' = [H \ I]^T \quad R' = \text{diag}(\begin{bmatrix} R & P \end{bmatrix}) \]

• Having any measurement means the system is overdetermined.

• So, the weighted least squares solution is:

\[
x' = (H'^T R'^{-1} H')^{-1} H'^T R'^{-1} \tilde{z}
\]

\[
x' = \left( \begin{bmatrix} H \ I \\ I \end{bmatrix}^T \begin{bmatrix} R & 0 \\ 0 & P \end{bmatrix}^{-1} \begin{bmatrix} H \\ I \end{bmatrix} \right)^{-1} \begin{bmatrix} H \ I \\ I \end{bmatrix}^T \begin{bmatrix} R & 0 \\ 0 & P \end{bmatrix}^{-1} \tilde{z}
\]
5.2.3.2 Recursive Estimation of a Random Vector

• Invert the covariance matrix on the right:

\[ \bar{x}' = \begin{pmatrix} H^T & R^{-1} & 0 & H \end{pmatrix}^{-1} \begin{pmatrix} H \end{pmatrix}^T \begin{pmatrix} R^{-1} & 0 \\ 0 & P^{-1} \end{pmatrix} \begin{pmatrix} z \end{pmatrix} \]

• Simplify the quadratic form on left:

\[ \bar{x}' = (H^T R^{-1} H + P^{-1})^{-1} \begin{pmatrix} H^T R^{-1} & P^{-1} \end{pmatrix} \begin{pmatrix} z \\ \bar{x} \end{pmatrix} \]

• Multiply out the product on right:

\[ \bar{x}' = (H^T R^{-1} H + P^{-1})^{-1} (P^{-1} \bar{x} + H^T R^{-1} z) \]

• This looks like an inverse covariance weighted average.
5.2.3.4.1 Efficient State Update

- Apply the Matrix Inversion Lemma which states:
  \[
  [H^T R^{-1} H + P^{-1}]^{-1} = P - PH^T [HPH^T + R]^{-1} HP
  \]

- Substituting:
  \[
  \hat{x}' = (P - PH^T S^{-1} HP)(P^{-1} \hat{x} + H^T R^{-1} \hat{z})
  \]

- Where we define the innovation covariance:
  \[
  S = [HPH^T + R]
  \]

- Define the Kalman Gain:
  \[
  K = PH^T S^{-1} = PH^T [HPH^T + R]^{-1}
  \]

- Which gives the famous result:
  \[
  \hat{x}' = \hat{x} + K(\hat{z} - H\hat{x})
  \]
Information Weighted Average

• Once again, the result is:

\[
x' = \hat{x} + K(z - H\hat{x})
\]

• Multiply that by \(PP^{-1}\) to get:

\[
\hat{x}' = P[P^{-1}\hat{x} + H^T S^{-1}(z - H\hat{x})]
\]

• So, the Kalman Filter is computing an information weighted average of the prior state and the innovation.
5.2.3.4.2 Covariance Update

- Recall the MLE covariance:

\[
\Sigma_{xx} = (H^T R^{-1} H)^{-1}
\]

- Consider again:

\[
\begin{align*}
\begin{bmatrix}
    H^T & R^{-1} & 0 \\
    0 & P^{-1} & 0 \\
    H^T & R^{-1} & 0 \\
    0 & P^{-1} & 0
\end{bmatrix}
\]^{-1}
\]

- So, the first part in brackets is just:

\[
P' = (H^T R^{-1} H + P^{-1})^{-1}
\]

- Substitute the Kalman Gain into Equation A:

\[
[H^T R^{-1} H + P^{-1}]^{-1} = P - PH^T [HPH^T + R]^{-1} HP = P - KHP
\]

- To get, the final form of covariance update:

\[
P' = (I - KH)P
\]
5.2.3.4.3 Covariance Update for Direct Measurements

• When $H=I$ the sensor measures the state directly, so...

\[
    x' = (R^{-1} + P^{-1})^{-1} \left( P^{-1} x + R^{-1} z \right)
\]

\[
    P' = (R^{-1} + P^{-1})^{-1} \quad \Rightarrow \quad (P')^{-1} = (R^{-1} + P^{-1})
\]

• Suppose:

\[
    P_0 = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}
\]

• Consider the sequence of measurements:

\[
    R_1 = \text{diag} \begin{bmatrix} 10 & 1 \end{bmatrix} \quad R_2 = \text{diag} \begin{bmatrix} 1 & 10 \end{bmatrix}
\]

\[
    R_3 = \text{diag} \begin{bmatrix} 1 & 0.1 \end{bmatrix} \quad R_4 = \text{diag} \begin{bmatrix} 0.1 & 0.1 \end{bmatrix}
\]
5.2.3.4.3 Covariance Update for Direct Measurements

- Regardless of the measurements themselves (linear case), the covariance evolves as follows:
5.2.3.5 Nonlinear Optimal Estimation

• When the measurements are related to the state nonlinearly:
  \[ \tilde{z} = h(x) + \nu \]
  \[ R = \text{Exp}(\nu\nu^T) \]

• We simply use nonlinear weighted least squares. That means, we simply make one substitution:
  \[ H = \left. \frac{\partial}{\partial \tilde{x}} [h(\tilde{x})] \right|_{\tilde{x}} \]

• Whereupon the Kalman Filter becomes the Extended Kalman filter.
  – Which is no longer optimal, but is nonetheless super useful
  – Easily the estimation equivalent of PID control.
  – KF is just a special case of EKF.
Outline

• 5.2 Random Variables, Processes and Transformation
  – 5.2.1 Variance of Continuous Integration and Averaging Processes
  – 5.2.2 Stochastic Integration
  – 5.2.3 Optimal Estimation
  – Summary
Summary

• Compounding (adding) noisy measurements leads to a result with more noise.
• Merging (filtering) noisy redundant measurements leads to a result with less noise.
• Kalman Filters are just recursive weighted least squares estimators.
  – That and matrix inversion Lemma is all it takes to derive it.
  – We will shortly see that they are applicable to dynamical systems.