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:

$$\sigma_x^2 = n\sigma_z^2$$

- The summing process can be written:

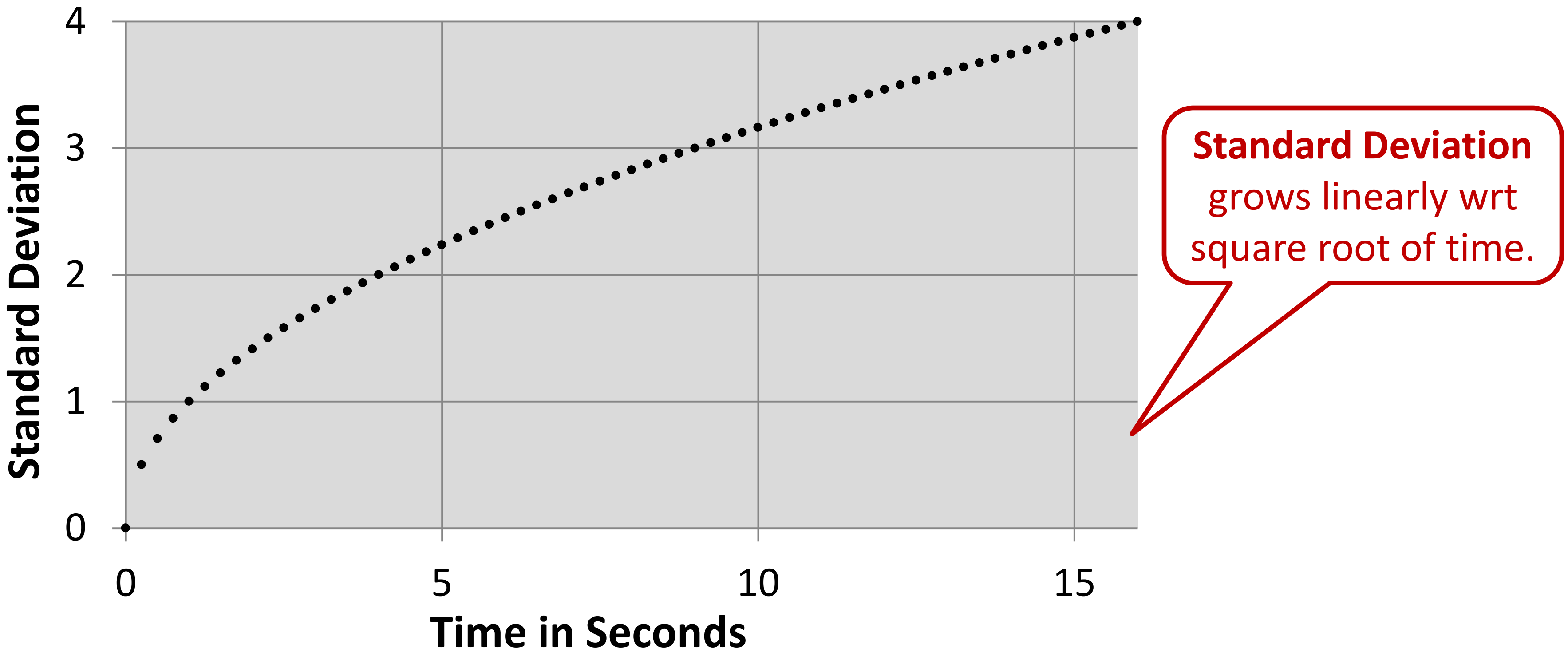
$$\hat{x}_k = \sum_{i=1}^{k} z_i = \sum_{i=1}^{k-1} z_i + z_k = \hat{x}_{k-1} + z_k$$

- Its covariance can be written recursively as:

$$\sigma_k^2 = \sum_{i=1}^{k} \sigma_z^2 = \sum_{i=1}^{k-1} \sigma_z^2 + \sigma_z^2 = \sigma_{k-1}^2 + \sigma_z^2$$

Variance grows linearly wrt time.

5.2.1.3 Variance of a Continuous Summing Process



5.2.1.4 Recursive Averaging

- The averaging process can be written as:

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

- Isolate the last estimate:

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

- 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:

$$\hat{x}_k = \hat{x}_{k-1} + K(z_k - \hat{x}_{k-1})$$

"Innovation"

5.2.1.4 Recursive Averaging

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

$$\sigma_k^2 = \frac{1}{k^2}\sum_{i=1}^{k}\sigma_z^2 = \frac{1}{k}\sigma_z^2$$

- 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:

$$\sigma_k^2 = \frac{\sigma_z^2}{k} \qquad \sigma_{k-1}^2 = \frac{\sigma_z^2}{k-1}$$

- Substitute to get:

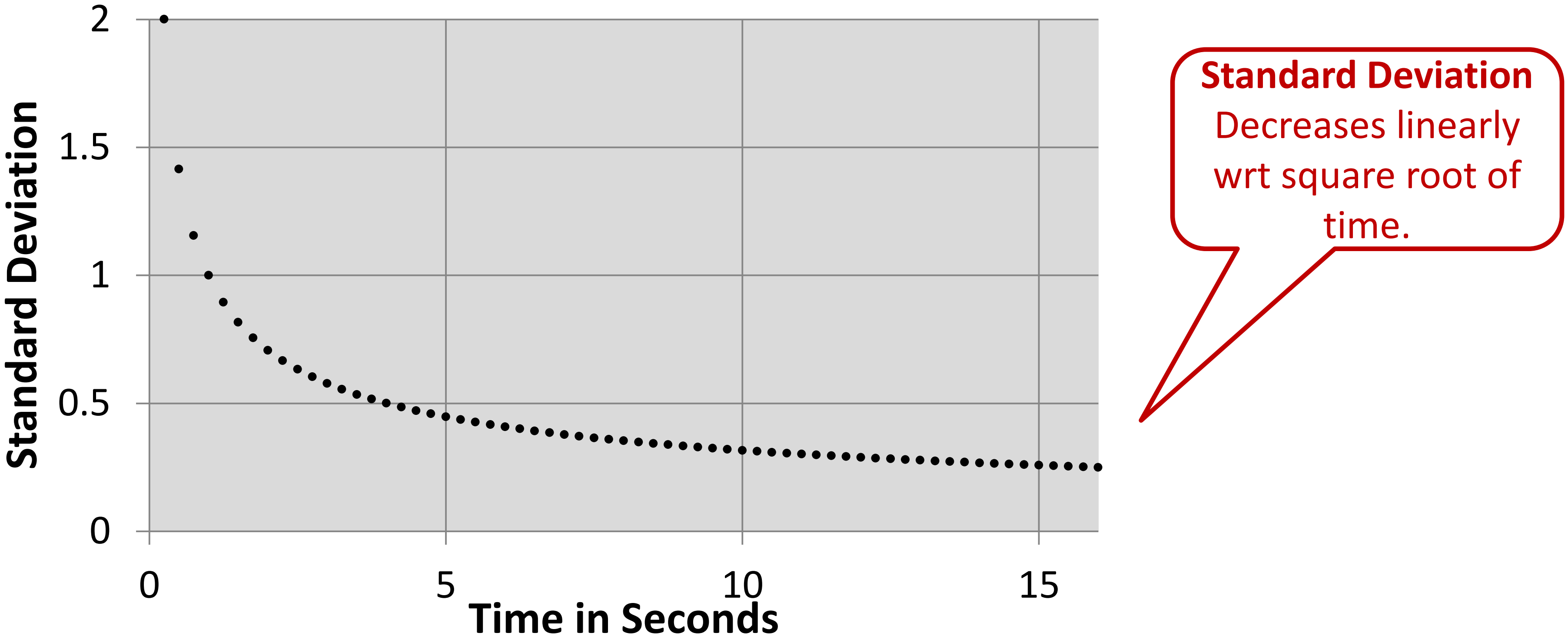
$$\frac{\sigma_k^2}{\sigma_{k-1}^2} = \frac{k-1}{k} \Rightarrow \sigma_k^2 = \left(\frac{k-1}{k}\right)\sigma_{k-1}^2 = \left(1 - \frac{1}{k}\right)\sigma_{k-1}^2$$

Typos In book Here

- 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



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.

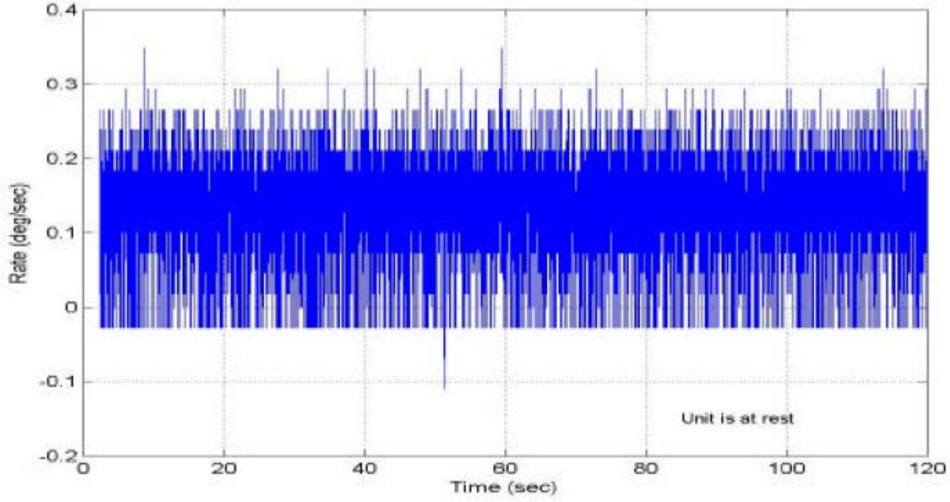


Figure 1. VG700CA rate output

Bias is unrelated to noise amplitude

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

Average bias is < 20 deg/hr.

From http://www.xbow.com

Allan Variance (Measure of Bias Stability)

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

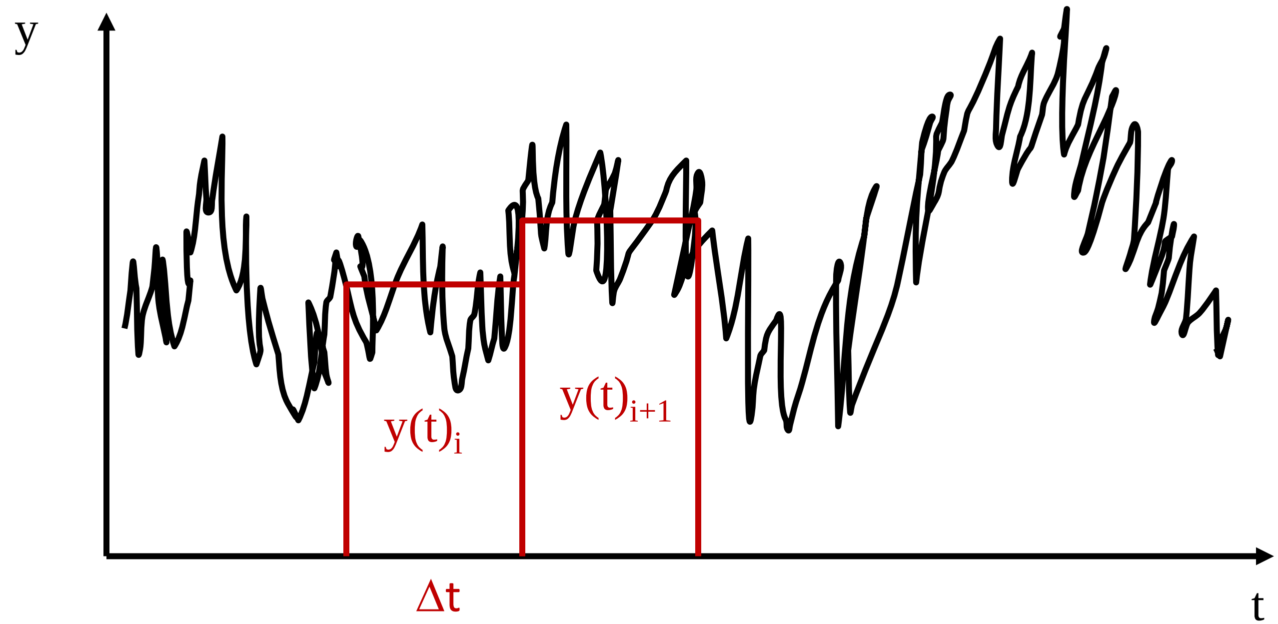
$$AVAR^2(\tau) = \frac{1}{2 \cdot (n-1)} \sum_i \left(y(\tau)_{i+1} - y(\tau)_i \right)^2 ,$$

Intuitively:
Variance in the Bias
For given level of averaging

Successive bins

What is the 2 for?

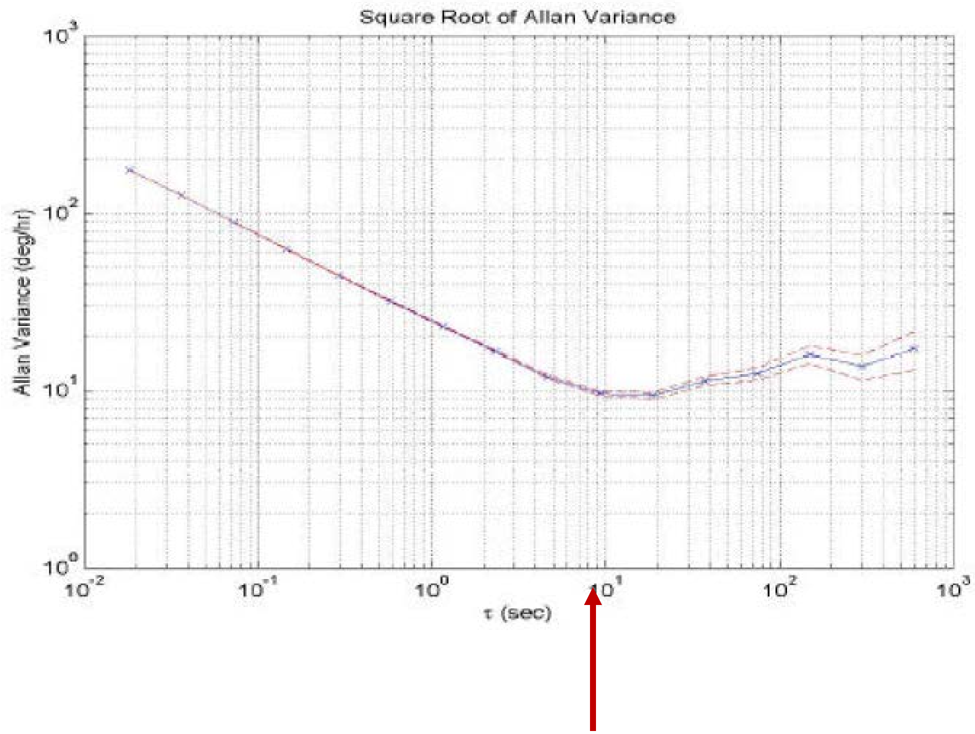
Allan Variance



- Compute the difference in average of successive bins.

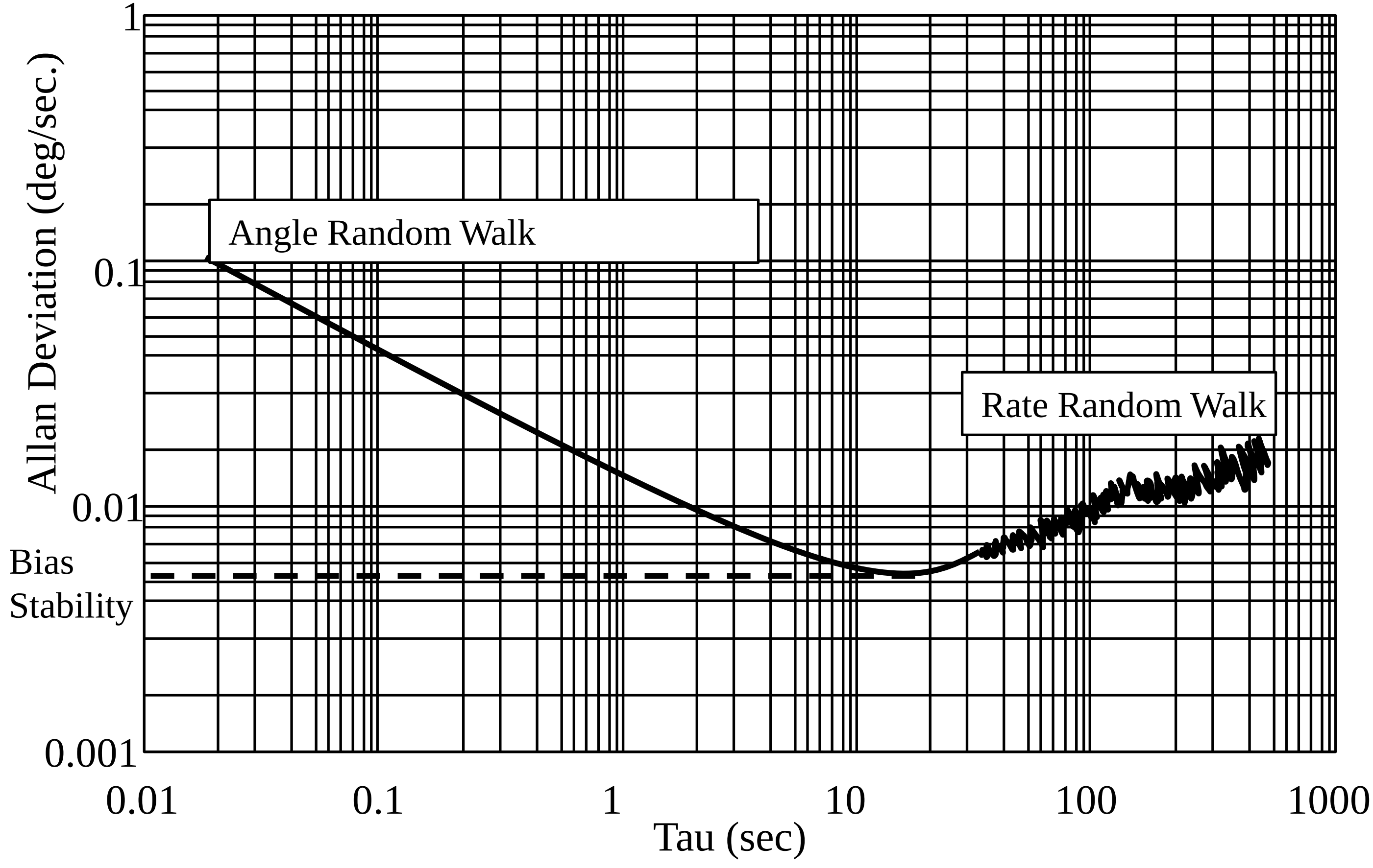
Allan Variance

- Variance drops initially as Δt increases (effect of averaging).
  - Sensor noise dominates for small Δt (does not average out).
  - Rate random walk dominates for large Δt (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



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 compute the state.

$$\underline{y} \longrightarrow \underline{x}_{i+1}$$

$$\underline{x}_1 \longrightarrow \underline{x}_i$$

$$\underline{x}_2 \longrightarrow u_i$$

$$\underline{y} = f(\underline{x}_1, \underline{x}_2) \longrightarrow \boxed{\underline{x}_{i+1} = f(\underline{x}_i, \underline{u}_i)}$$

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:

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

Covariance Propagation in any Decorrelated Estimation Process.

- In a more standard notation:

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

State Uncertainty

Transition Matrix

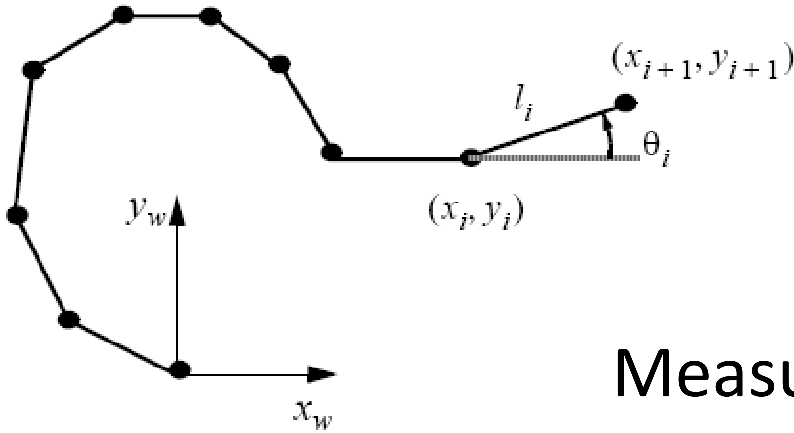
Input Uncertainty

Input Jacobian

This is one of the Equations of the Kalman Filter

5.2.2.2 Example: Dead Reckoning

(With Odometer Error Only)



State: $\underline{x}_i = \begin{bmatrix} x_i & y_i \end{bmatrix}^T$

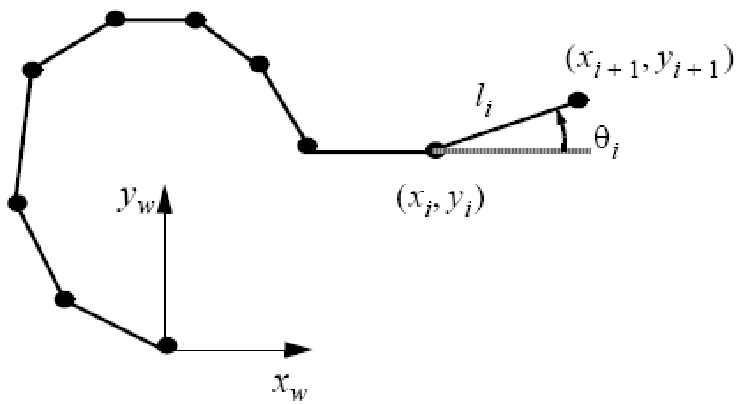
Measurements: $\underline{u}_i = \begin{bmatrix} l_i & \theta_i \end{bmatrix}^T$

Update: $\underline{x}_{i+1} = \underline{f}(\underline{x}_i, \underline{u}_i) = \begin{bmatrix} x_i + l_i \cos(\theta_i) \\ y_i + l_i \sin(\theta_i) \end{bmatrix}$

5.2.2.2 Example: Dead Reckoning

(Jacobian)

Linearize: $\underline{x}_{i+1} = f(\underline{x}_i, \underline{u}_i) = \begin{bmatrix} x_i + l_i cos(\psi_i) \\ y_i + l_i sin(\psi_i) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} lc_i \\ ls_i \end{bmatrix}$



$$\Phi_i = \frac{\partial \underline{x}_{i+1}}{\partial \underline{x}_i} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

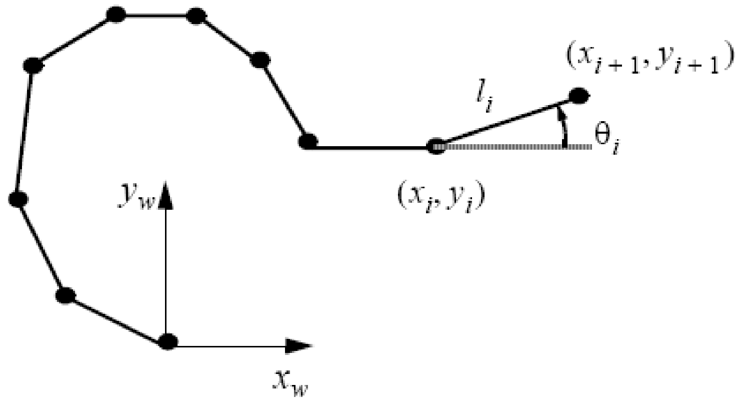
$$\Gamma_i = \frac{\partial \underline{x}_{i+1}}{\partial \underline{u}_i} = \begin{bmatrix} c_i & -l_i s_i \\ s_i & l_i c_i \end{bmatrix}$$

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:



$$P_i = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix}_i$$

$$Q_i = \begin{bmatrix} \sigma_l^2 & 0 \\ 0 & 0 \end{bmatrix}$$

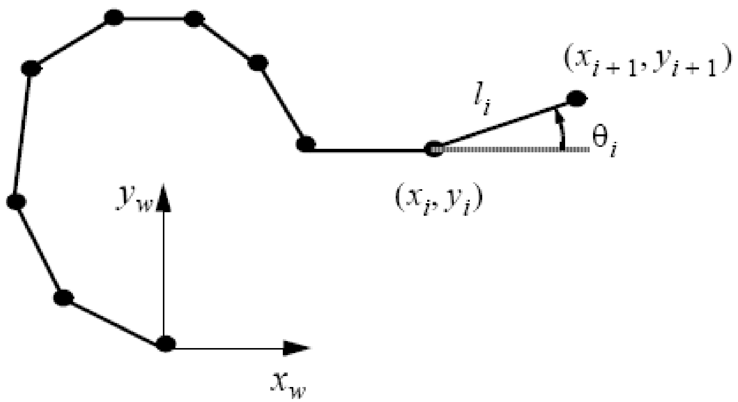
Assume Decorrelated errors.

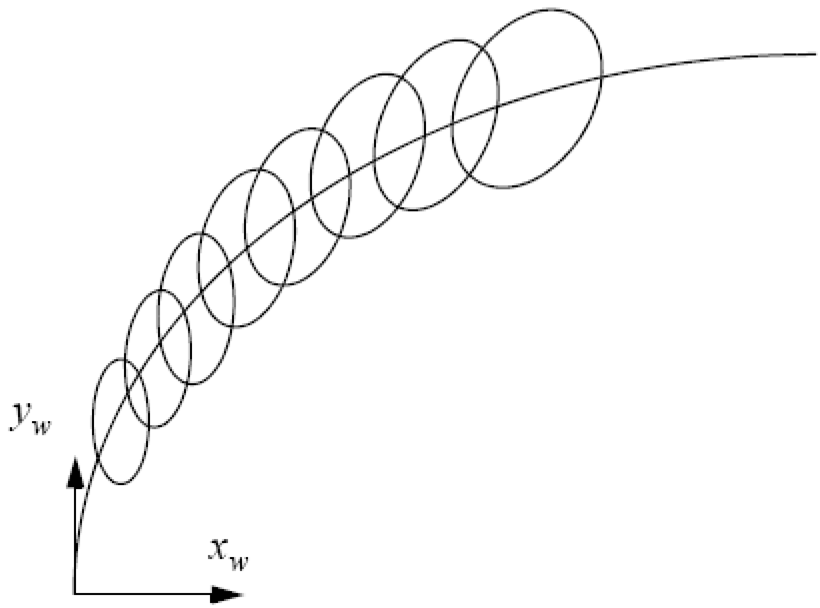
Assume a perfect compass.

5.2.2.2 Example: Dead Reckoning

(Answer)

$$P_{i+1} = \Phi P_i \Phi^T + \Gamma_i Q_i \Gamma_i^T \Rightarrow P_{i+1} = P_i + \begin{bmatrix} c_i^2 \sigma_l^2 & c_i s_i \sigma_l^2 \\ c_i s_i \sigma_l^2 & s_i^2 \sigma_l^2 \end{bmatrix}$$





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:

$$n = \frac{t}{\Delta t} \qquad \sigma_y^2 = \frac{\sigma_x^2 t}{\Delta t}$$

Constant

- But this means that $\sigma_y^2 \rightarrow \infty$ as $\Delta t \rightarrow 0$!!!
- That cannot be right; it would require infinite power.
- What is more realistic is variance that grows linearly wrt time:

$$\sigma_x^2(t) = \dot{\sigma}_z^2 \cdot t$$

5.2.2.3.2 Integrating Stochastic Differential Equations

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

$$\delta \dot{\underline{x}}(t) = F(t)\delta \underline{x}(t) + G(t)\delta \underline{u}(t)$$

Noise

- Define the covariances:

$$Exp(\delta \underline{x}(t)\delta \underline{x}(t)^T) = P(t)$$

$$Exp(\delta \underline{u}(t)\delta \underline{u}(\tau)^T) = Q(t)\delta(t-\tau)$$

White noise is uncorrelated in time.

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

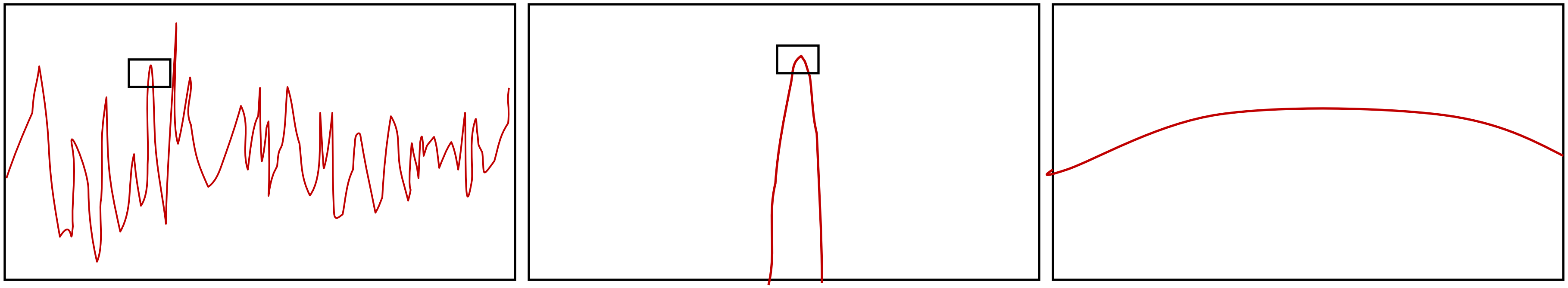
$$\delta \underline{x}(t) = \Phi(t, t_0)\delta \underline{x}(t_0) + \int_{t_0}^{t} \Gamma(t, \tau)\delta \underline{u}(\tau)d\tau$$

Reimann says this integral **does not converge**

$$\Gamma(t, \tau) = \Phi(t, \tau)G(\tau)$$

Motivation for Stochastic Calculus

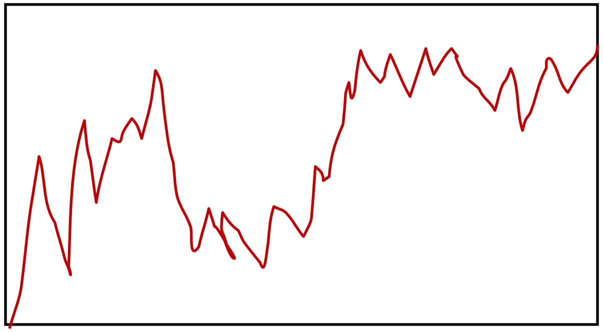
- An integral is a limit of a sum of products.
- The limit exists when the wiggles go away when you zoom in on a function:



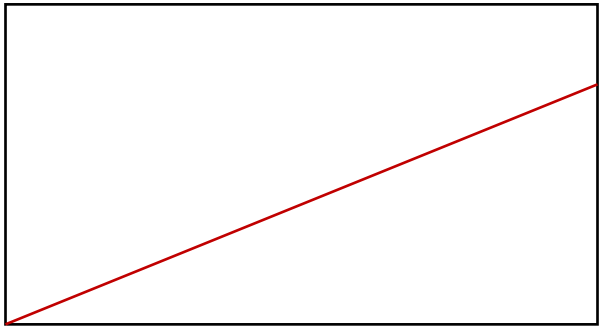
- For a white random signal, autocorrelation is zero, and the wiggles never go away at any zoom level.
- The integral or derivative of a white signal is meaningless.
  - So what is "stochastic calculus"?

Deterministic Statistics

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



Individual random walk signal



Variance of a zillion random walk signals

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

5.2.2.3.2 Integrating Stochastic Differential Equations

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

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

- Trick: Introduce a differential random walk process:

$$d\underline{\beta}(\tau) = \delta \underline{u}(\tau)d\tau$$

- Now, integrate the following:

$$\delta \underline{x}(t) = \Phi(t, t_0)\delta \underline{x}(t_0) + \int_{t_0}^{t} \Gamma(t, \tau)d\underline{\beta}(\tau)$$

5.2.2.3.2 Integrating Stochastic Differential Equations

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

$$P(t) = \Phi(t, t_0)P(t_0)\Phi(t, t_0)^T + \int_{t_0}^{t} \Gamma(t, \tau)Q(\tau)\Gamma(t, \tau)^T d\tau$$

State Covariance

Transition Matrix

Initial State Covariance

Transition Matrix

Input Transition Matrix

Input Covariance

Input Transition Matrix

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
  - <u>5.2.3 Optimal Estimation</u>
  - Summary

5.2.3.1 Maximum Likelihood Estimation

- Consider the problem of optimally estimating state from a series of measurements:
  - Let $\underline{x} \in \Re^n$ denote the state and $\underline{z} \in \Re^m$ denote the measurements.
  - Measurements relate to the state by a measurement matrix:

$$\underline{z} = H\underline{x} + \underline{v} \quad where \quad \underline{v} \sim N(0, R)$$

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

$$R = \text{Exp}(\underline{v}\,\underline{v}^T)$$

5.2.3.1 Maximum Likelihood Estimation

- The innovation $\underline{z} - H\underline{x} \quad (= \underline{v})$ is Gaussian by assumption, so…
- The probability of getting a measurement $\underline{z}$ when the true state is $\underline{x}$ is:

$$p(\underline{z}|\underline{x}) = \frac{1}{(2\pi)^{m/2}|R|^{1/2}} exp\left[-\frac{1}{2}(\underline{z} - H\underline{x})R^{-1}(\underline{z} - H\underline{x})^T\right]$$

- This exponential will be maximized when the form in the exponent (without negative sign) is minimized:

$$\hat{\underline{x}}^* = \underset{\underline{x}}{argmin}\left(\frac{1}{2}(\underline{z} - H\underline{x})R^{-1}(\underline{z} - H\underline{x})^T\right)$$

- If the system is overdetermined, the solution is simply the weighted left pseudoinverse:

$$\hat{\underline{x}}^* = \boxed{(H^T R^{-1} H)^{-1}} H^T R^{-1} \underline{z}$$

5.2.3.1.1 Covariance of the MLE Estimate

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

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

- Therefore the covariance of the MLE result is:

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

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

- Which simplifies to:

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

Equation 5.80

- Note that this expression appears in the pseudoinverse:

$$\underline{\hat{x}}^* = \boxed{(H^T R^{-1} H)^{-1}} H^T R^{-1} \underline{z} = \Sigma_{xx} H^T R^{-1} \underline{z}$$

5.2.3.2 Recursive Estimation of a Random Scalar

- Suppose:
  - Present state estimate x has variance σ_x^2
  - Measurement z has variance σ_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:

$$H' = \begin{bmatrix} 1 & 1 \end{bmatrix}^T \qquad R' = diag\begin{bmatrix} \sigma_z^2 & \sigma_x^2 \end{bmatrix}$$

- So, the weighted least squares solution is:

$$x' = (H'^T R'^{-1} H')^{-1} H'^T R'^{-1} \underline{z} = \left(\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_z^2} & 0 \\ 0 & \frac{1}{\sigma_x^2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_z^2} & 0 \\ 0 & \frac{1}{\sigma_x^2} \end{bmatrix} \underline{z}$$

5.2.3.2 Recursive Estimation of a Random Scalar

- This simplifies to:

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

Equation 5.8.1

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

$$\sigma_{x'}^2 = (H'^T R'^{-1} H')^{-1} = \left(\left[\frac{1}{\sigma_z^2} + \frac{1}{\sigma_x^2}\right]\right)^{-1}$$

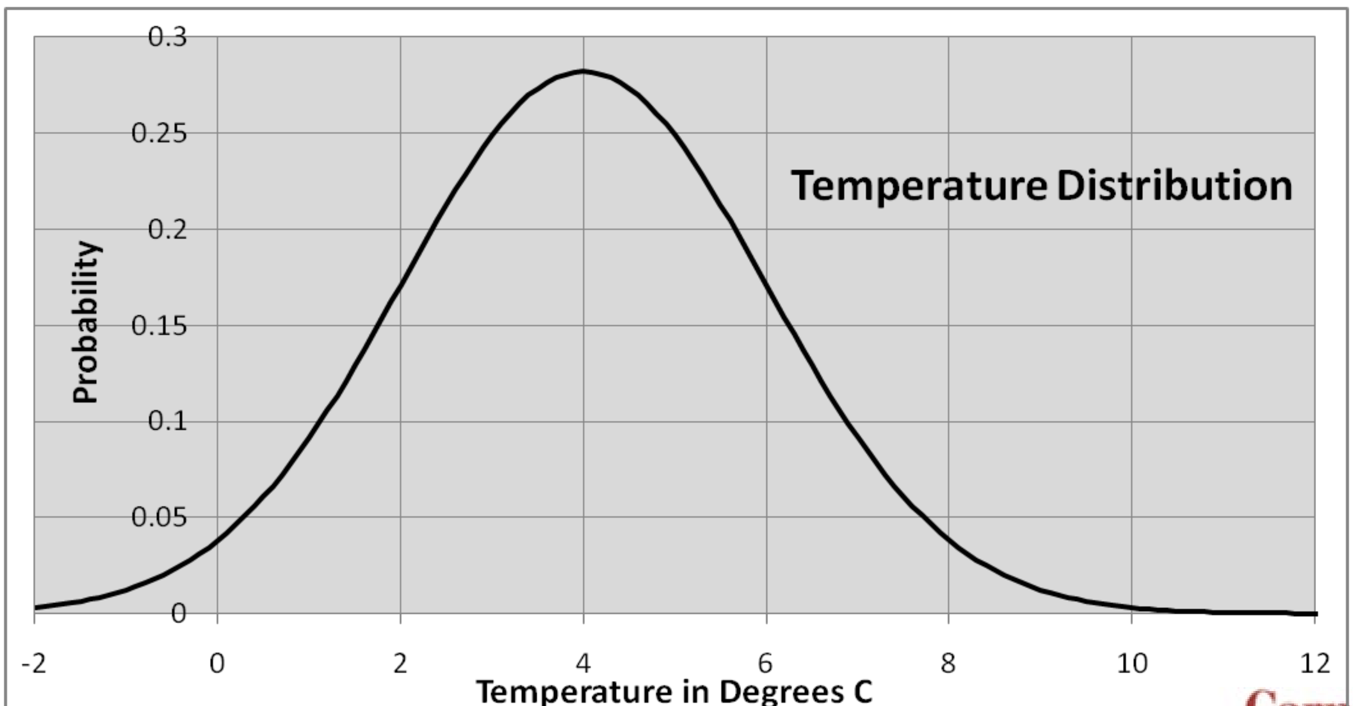
- Which is the same as saying the new information is the sum of that of the measurement and state:

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

Equation 5.8.3

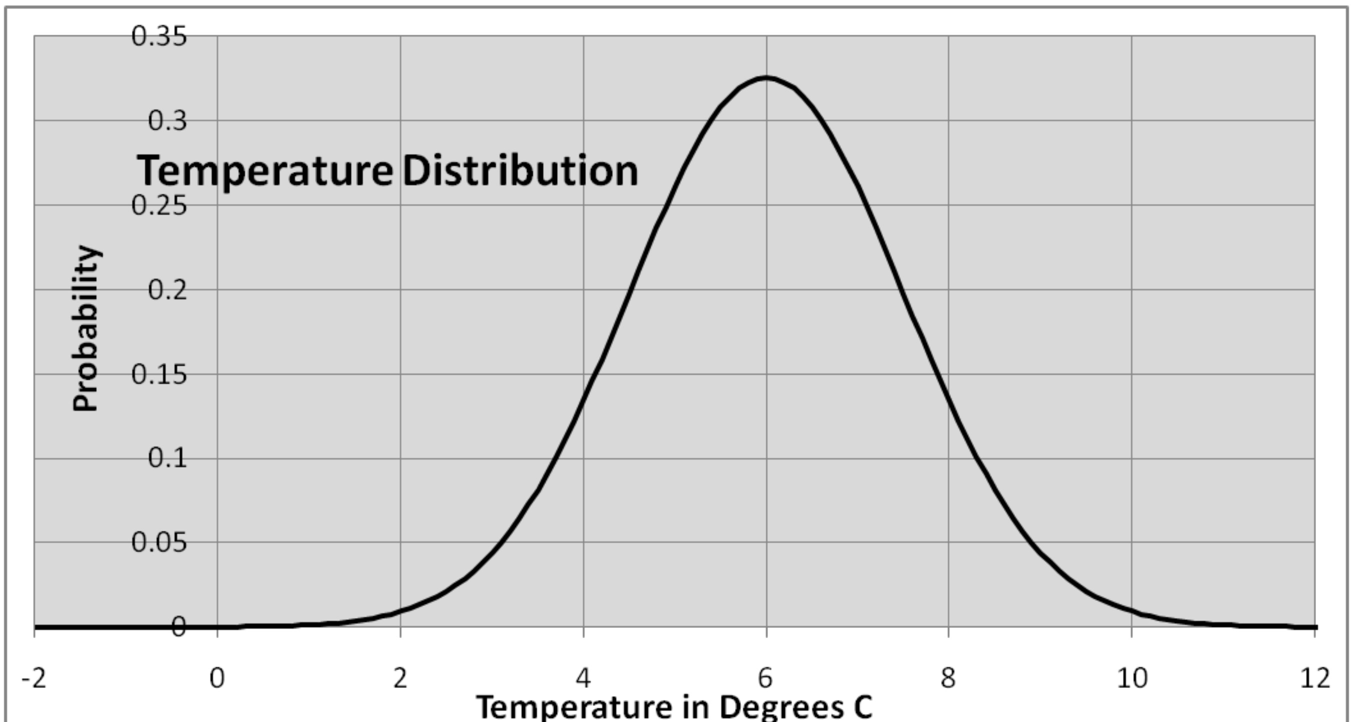
5.2.3.3 Example: Estimating Temperature from Two Sensors

- 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:

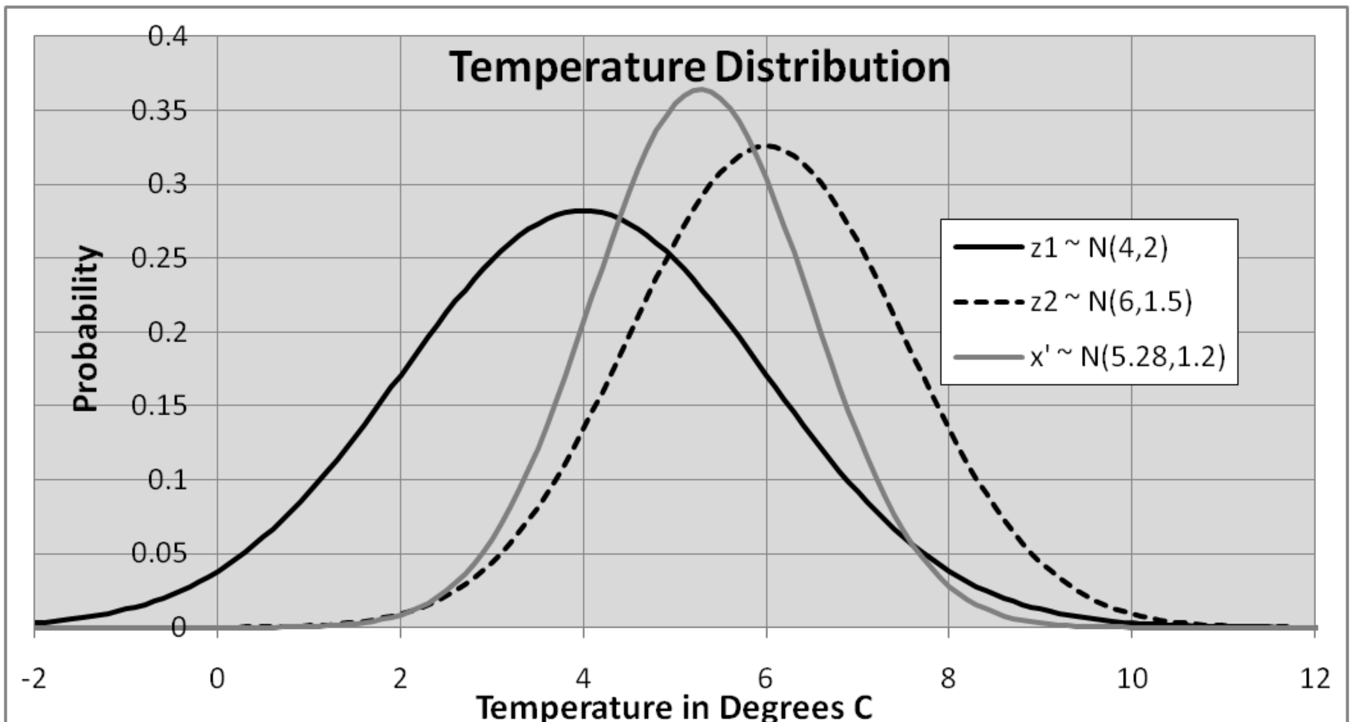


5.2.3.3 Example: Estimating Temperature from Two Sensors

- Application of Eqns 5.81 and 5.83 gives:

$$x' = 5.28 \qquad \sigma_{x'}^2 = (1.2)^2$$

- The result is denoted graphically as follows:



5.2.3.2 Recursive Estimation of a Random Vector

- Suppose:
  - Present state estimate $\underline{x}$ has variance P
  - Measurement $\underline{z}$ has variance R
  - Want to get new state estimate $\underline{x}'$ and its variance P'
- The measurement relationship for (both) measurements is:

$$\begin{bmatrix} \underline{z} \\ \underline{x} \end{bmatrix} = \begin{bmatrix} H \\ I \end{bmatrix} \underline{x}' = H' \underline{x}'$$

5.2.3.2 Recursive Estimation of a Random Vector

- That means the associated measurement and covariance matrices are:
  - $H' = [H \quad I]^T \qquad\qquad R' = diag([R \quad P])$
- Having any measurement means the system is overdetermined.
- So, the weighted least squares solution is:

$$x' = (H'^T R'^{-1} H')^{-1} H'^T R'^{-1} \underline{z}$$

$$\underline{x}' = \left(\begin{bmatrix} H \\ I \end{bmatrix}^T \left[\begin{bmatrix} R & 0 \\ 0 & P \end{bmatrix} \right]^{-1} \begin{bmatrix} H \\ I \end{bmatrix} \right)^{-1} \begin{bmatrix} H \\ I \end{bmatrix}^T \begin{bmatrix} R & 0 \\ 0 & P \end{bmatrix}^{-1} \underline{z}$$

5.2.3.2 Recursive Estimation of a Random Vector

- Invert the covariance matrix on the right:

$$\underline{x}' = \left(\begin{bmatrix} H \\ I \end{bmatrix}^T \begin{bmatrix} R^{-1} & 0 \\ 0 & P^{-1} \end{bmatrix} \begin{bmatrix} H \\ I \end{bmatrix} \right)^{-1} \begin{bmatrix} H \\ I \end{bmatrix}^T \begin{bmatrix} R^{-1} & 0 \\ 0 & P^{-1} \end{bmatrix} \underline{z}$$

- Simplify the quadratic form on left:

$$\underline{x}' = (H^T R^{-1} H + P^{-1})^{-1} \begin{bmatrix} H^T R^{-1} & P^{-1} \end{bmatrix} \begin{bmatrix} \underline{z} \\ \underline{x} \end{bmatrix}$$

- Multiply out the product on right:

$$\underline{x}' = \boxed{(H^T R^{-1} H + P^{-1})^{-1}} (P^{-1} \underline{x} + H^T R^{-1} \underline{z})$$

- This looks like an inverse covariance weighted average.

5.2.3.4.1 Efficient State Update

- Apply the Matrix Inversion Lemma which states:

$$[H^T R^{-1} H + P^{-1}]^{-1} = P - \boxed{PH^T[HPH^T + R]^{-1}} HP \quad \textbf{Equation A}$$

- Substituting:

$$\underline{x}' = (P - \boxed{PH^T S^{-1}} HP)(P^{-1}\underline{x} + H^T R^{-1}\underline{z})$$

- 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:

$$\boxed{\underline{x}' = \underline{x} + K(\underline{z} - H\underline{x})}$$

Information Weighted Average

- Once again, the result is:

$$\underline{x}' = \underline{x} + K(\underline{z} - H\underline{x})$$

- Multiply that by PP^{-1} to get:

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

- 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:

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

Equation 5.80

- Consider again:

$$\underline{x}' = \left(\begin{bmatrix} H \\ I \end{bmatrix}^T \begin{bmatrix} R^{-1} & 0 \\ 0 & P^{-1} \end{bmatrix} \begin{bmatrix} H \\ I \end{bmatrix} \right)^{-1} \begin{bmatrix} H \\ I \end{bmatrix}^T \begin{bmatrix} R^{-1} & 0 \\ 0 & P^{-1} \end{bmatrix} \underline{z}$$

Equation 5.85

$[\boldsymbol{H}^T \quad \boldsymbol{R}^{-1} \quad \boldsymbol{H}]^{-1}$

- 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...

$$\underline{x}' = (R^{-1} + P^{-1})^{-1}(P^{-1}\underline{x} + R^{-1}\underline{z})$$

$$P' = (R^{-1} + P^{-1})^{-1} \Rightarrow (P')^{-1} = (R^{-1} + P^{-1})$$

Information Adds Directly

- Suppose: $P_0 = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$

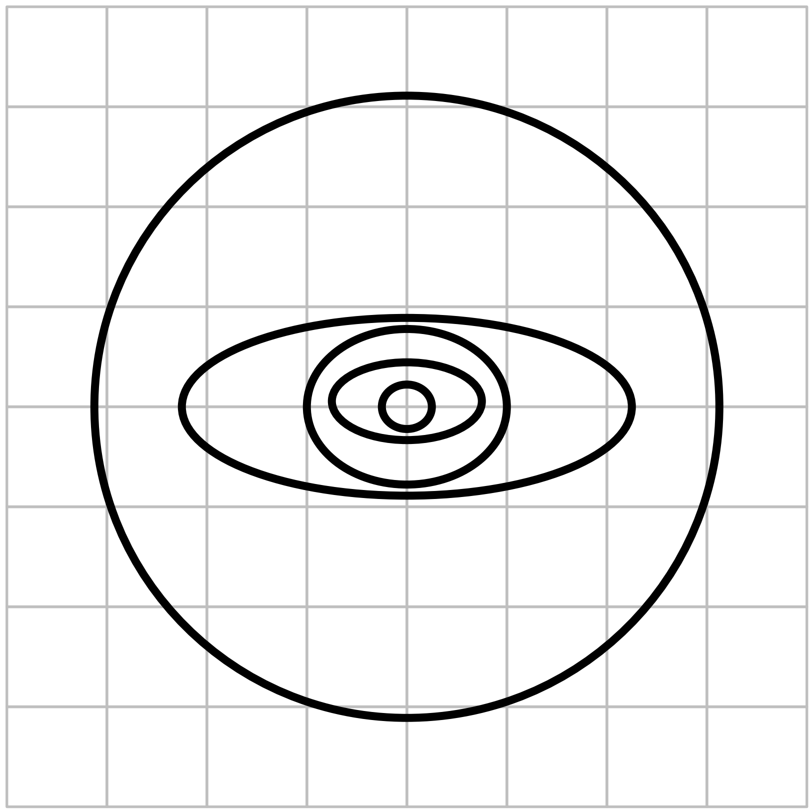
- Consider the sequence of measurements:

$$R_1 = diag\begin{bmatrix} 10 & 1 \end{bmatrix} \quad R_2 = diag\begin{bmatrix} 1 & 10 \end{bmatrix}$$

$$R_3 = diag\begin{bmatrix} 1 & 0.1 \end{bmatrix} \quad R_4 = diag\begin{bmatrix} 0.1 & 0.1 \end{bmatrix}$$

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:

$$\underline{z} = h(\underline{x}) + \underline{v} \qquad R = Exp(\underline{v}\underline{v}^T)$$

- We simply use nonlinear weighted least squares. That means, we simply make one substitution:

$$H = \frac{\partial}{\partial \underline{x}}[h(\underline{x})]\Big|_{\underline{x}}$$

- 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.