Optimal Estimation Methods

(Lecture 12 – Kalman Filtering: Part I)

Dr. John L. Crassidis

University at Buffalo – State University of New York
Department of Mechanical & Aerospace Engineering
Amherst, NY 14260-4400
johnc@buffalo.edu
http://www.buffalo.edu/~johnc

University at Buffalo The State University of New York

Introduction (i)

- Least Squares Estimation
  - Applied to systems whose measured variables are related to the estimated parameters by *algebraic* equations
- We now extend these results to allow estimation of parameters embedded in the model of a dynamical system
  - The model usually includes both *algebraic* and *differential* equations
  - We will find that the least squares sequential estimation results developed for estimation of *algebraic* systems remain valid for estimation of *dynamical* systems upon making the appropriate new interpretations of the matrices involved in the estimation algorithms
    - Must be extended to properly account for “motion” of the dynamical system between measurement and estimation epochs

Introduction (ii)

- Suppose that we have a set of measurements that are a function of some desired states to be estimated
  - For example, say the measurements are from a three-axis magnetometer (TAM)
  - These measurements are both corrupted with noise and do not give full information about the states (e.g., roll, pitch and yaw)
    - A TAM only provides two out of the three pieces of the state (its usually a combination of the state quantities, i.e. one usually does not obtain inertial pitch directly from a TAM in general) Magnetic field not perfect dipole
  - A dynamic motion model may exist but it is inaccurate
    - For example, a "mission mode" spacecraft may have rates that are nearly constant; we use a constant rate kinematics model to approximate this motion
  - A proper combination of the model and measurements can be used to estimate the states and "filter" the measurements
    - Over time, a TAM with a dynamic model can provide full three-axis information using an estimator (oftentimes called a "filter")

First-Order Example (i)

- Suppose that a “truth” model is generated using

$$\dot{x}(t) = F\,x(t), \quad x(t_0) = 1$$
$$\tilde{y}(t) = H\,x(t) + v(t)$$

  - Synthetic measurements are created for a 10-second time interval with $F = -1$ and $H = 1$, and assume $v(t)$ is zero-mean Gaussian noise process with the standard deviation given by 0.05
  - Suppose now that we wish to estimate $x(t)$ using the available measurements and some dynamic model
    - In practice the actual “truth” model is unknown (if it were known exactly then we wouldn’t need an estimator!)
  - For this example, we will assume that the initial condition is known exactly, but the “modeled” value for F is given by $\bar{F} = -1.5$
    - Clearly, if we replace F with $\bar{F}$ and integrate this equation to find an estimate for $x(t)$, we would find that the estimated $x(t)$ is far from the truth

First-Order Example (ii)

- In order to produce better results, we shall use the age-old adage commonly spoken in control of dynamic systems: “when in doubt, use feedback!”
  - Consider the following linear feedback system for the state and output estimates

Model

Model

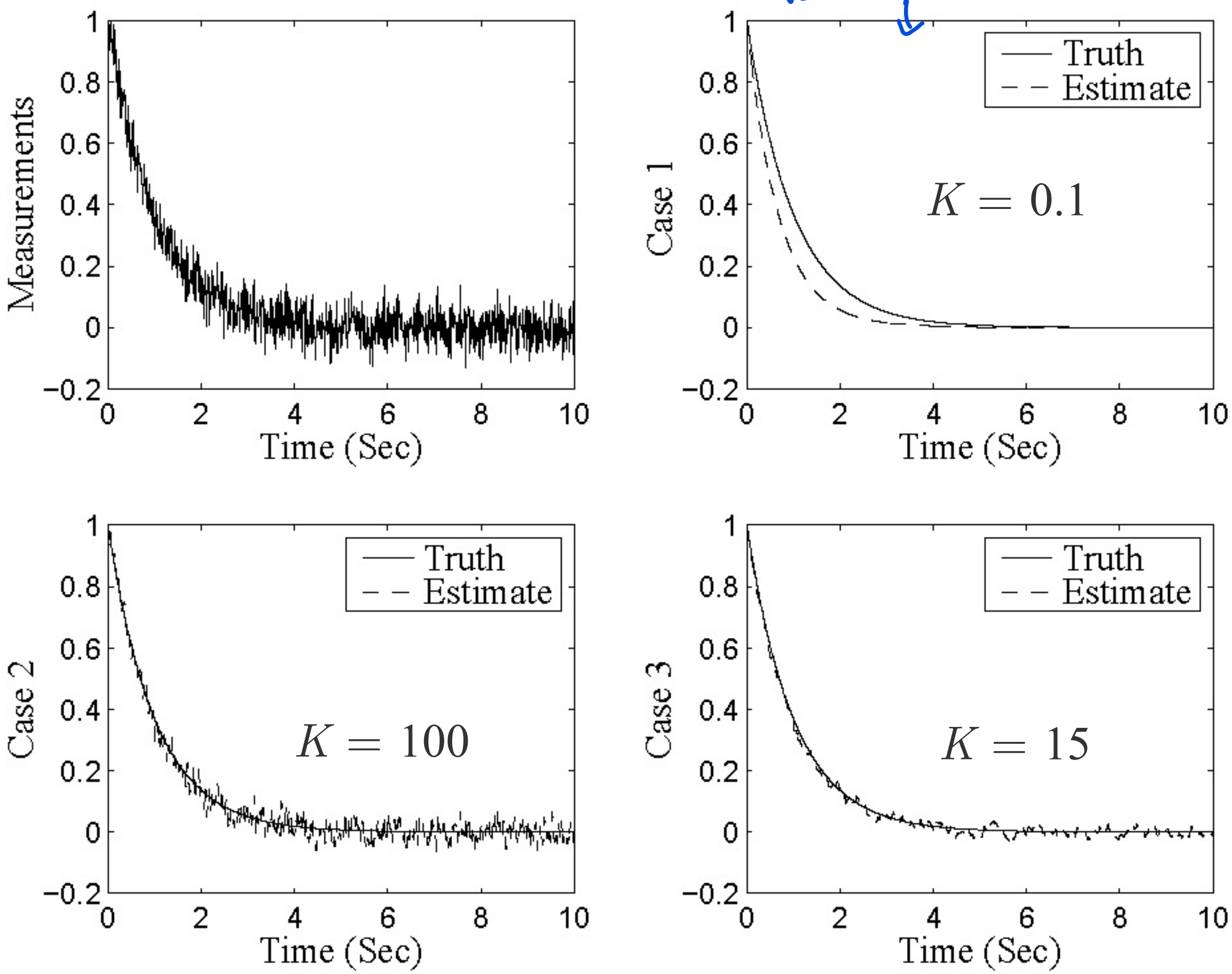
$$\dot{\hat{x}}(t) = \bar{F}\,\hat{x}(t) + K[\tilde{y}(t) - \bar{H}\,\hat{x}(t)], \quad \hat{x}(t_0) = 1$$
$$\hat{y}(t) = \bar{H}\,\hat{x}(t)$$

where $\hat{x}(t)$ denotes the estimate of $x(t)$, K is a constant gain, and $\bar{H} = H = 1$

  - At this point we do not consider how to determine the value of K, but instead (since we know the truth) we will pick various values and compare the resulting estimates with the truth
  - Three cases: Case 1 ($K = 0.1$), Case 2 ($K = 100$), and Case 3 ($K = 15$)

First-Order Example (iii)

relys more on model $K\tilde{y}$ = small



For small gains (such as Case 1) the estimates are far from the truth

Also, for large gains (such as Case 2) the estimates are very noisy

Case 3 depicts a gain that closely follows the truth, while at the same time providing filtered estimates

First-Order Example (iv)

- This simple example illustrates the basic concepts used in state estimation and filtering
- Go back to $\dot{\hat{x}}(t) = \bar{F}\,\hat{x}(t) + K[\tilde{y}(t) - \bar{H}\,\hat{x}(t)]$
  - As the gain decreases, measurements tend to be ignored and the system relies more heavily on the model (which in this case is incorrect leading to erroneous estimates)
  - As the gain increases the estimates rely more on the measurements; however, if the gain is too large then the model tends to be ignored all together, as shown by Case 2
  - Can be related to frequency domain filtering (bandwidth of the filter)

bad way to look at it

    - Large bandwidth admits a lot of noise, while too small bandwidth makes the response too slow
    - Try to pick a bandwidth that gives the best of both worlds (or least a good compromise)

Full-Order Observers (i)

- The full-order observer is given by

$$\dot{\hat{\mathbf{x}}} = F\,\hat{\mathbf{x}} + B\,\mathbf{u} + K[\tilde{\mathbf{y}} - H\,\hat{\mathbf{x}}]$$
$$\hat{\mathbf{y}} = H\,\hat{\mathbf{x}}$$

n×m

F-KH.
must be stable, all eigenvalues in left hand plane.

  - As mentioned previously we would like to find a method to determine K
  - Look at another form of the observer

Assume time invariant

$$\dot{\hat{\mathbf{x}}} = (F - KH)\,\hat{\mathbf{x}} + B\,\mathbf{u} + K\tilde{\mathbf{y}}$$

  - The dynamics of the observer are governed by $E = F - KH$
  - Say for now we are given some desired eigenvalues for E; using these eigenvectors we can form the characteristic equation given by

$$d(s) = s^n + \delta_{n-1}\,s^{n-1} + \cdots + \delta_1\,s + \delta_0 = 0$$

  - Goal is to determine K to match the desired characteristic equation

Full-Order Observers (ii)

- Solution is found using Ackermann's formula

$$\boxed{K = d(F)\begin{bmatrix} H \\ HF \\ HF^2 \\ \vdots \\ HF^{n-1} \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \equiv d(F)\mathcal{O}^{-1}\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}}$$

where

$$d(F) = F^n + \delta_{n-1}\,F^{n-1} + \cdots + \delta_1\,F + \delta_0 I$$

- Note that the system must be fully observable because the inverse of the observability matrix is required

Ackermann's Formula (i)

- Proof of Ackermann's formula
  - Consider the third-order observer canonical form

$$F_o = \begin{bmatrix} 0 & 0 & -a_0 \\ 1 & 0 & -a_1 \\ 0 & 1 & -a_2 \end{bmatrix}, \quad H_o = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$$

  - The matrix $(F_o - KH_o)$, with $K \equiv [k_1 \ k_2 \ k_3]^T$, is given by

$$F_o - KH_o = \begin{bmatrix} 0 & 0 & -(a_0 + k_1) \\ 1 & 0 & -(a_1 + k_2) \\ 0 & 1 & -(a_2 + k_3) \end{bmatrix}$$

  - The characteristic equation is given by

$$s^3 + (a_2 + k_3)\, s^2 + (a_1 + k_2)\, s + (a_0 + k_1) = 0$$

  - Suppose that we have a desired characteristic equation formed from a set of desired eigenvalues in the estimator, given by

$$d(s) = s^3 + \delta_2\, s^2 + \delta_1\, s + \delta_0 = 0$$

Ackermann's Formula (ii)

- Then, the gain matrix can be obtained by comparing the corresponding coefficients

$$k_1 = \delta_0 - a_0, \quad k_2 = \delta_1 - a_1, \quad k_3 = \delta_2 - a_2$$

- This approach can easily be expanded to higher-order systems
  - However, this can become quite tedious and numerically inefficient
  - Also, must convert gains back to form for general F
- It would be useful if the gain K can be derived using the matrix F directly, without having to convert F into observer canonical form

must be nonsingular

- Applying the Cayley-Hamilton theorem to the matrix $E = F - KH$ leads to

$$d(E) = E^3 + \delta_2 E^2 + \delta_1 E + \delta_0 I = 0$$

- Performing the multiplications for E^3 and E^2 yields

$$E^2 = F^2 - KHF - EKH$$

$$E^3 = F^3 - KHF^2 - EKHF - E^2 KH$$

Ackermann's Formula (iii)

- Substituting these in $d(E)$ gives

$$F^3 + \delta_2 F^2 + \delta_1 F + \delta_0 I$$
$$- \delta_1 KH - \delta_2 KHF - \delta_2 EKH - KHF^2 - EKHF - E^2KH = 0$$

- Since the first four terms are defined as $d(F)$, we can rewrite this as

$$d(F) = \begin{bmatrix} (\delta_1 K + \delta_2 EK + E^2 K) & (\delta_2 K + EK) & K \end{bmatrix} \begin{bmatrix} H \\ HF \\ HF^2 \end{bmatrix}$$

- Therefore, the gain K can be found from

$$K = d(F) \begin{bmatrix} H \\ HF \\ HF^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

- This can easily be extended to higher-order systems to give Ackermann's formula

Example (i)

- Consider the following example

$$F = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}, \quad H = \begin{bmatrix} h_1 & h_2 \end{bmatrix}$$

  - Let the gain be given by $K = [k_1 \quad k_2]^T$
  - The desired characteristic equation of the estimator is given by

$$d(s) = s^2 + \delta_1 s + \delta_0 = 0$$

  - Computing $\det(sI - F + KH)$ allows us to solve for the gain K by comparing coefficients to the desired characteristic equation, which gives (Highly nonlinear)

$$\delta_0 = (k_1 h_1 - f_{11})(k_2 h_2 - f_{22}) - (k_1 h_2 - f_{12})(k_2 h_1 - f_{21})$$
$$\delta_1 = k_1 h_1 + k_2 h_2 - f_{11} - f_{22}$$

  - Finding solutions for k_1 and k_2 is not easy from these two equations (try it!)

Example (ii)

- Ackermann's formula directly gives

$$k_1 = \frac{1}{b\,h_1 - a\,h_2}\left[d\,h_1 - c\,h_2 + \delta_1\,(h_1\,f_{12} - h_2\,f_{11}) - \delta_0\,h_2\right]$$

$$k_2 = \frac{1}{b\,h_1 - a\,h_2}\left[g\,h_1 - e\,h_2 + \delta_1\,(h_1\,f_{22} - h_2\,f_{21}) + \delta_0\,h_1\right]$$

$$a = h_1\,f_{11} + h_2\,f_{21}$$

$$b = h_1\,f_{12} + h_2\,f_{22}$$

$$c = f_{11}^2 + f_{12}\,f_{21}$$

$$d = f_{11}\,f_{12} + f_{12}\,f_{22}$$

$$e = f_{11}\,f_{21} + f_{21}\,f_{22}$$

$$g = f_{22}^2 + f_{12}\,f_{21}$$

As $b\,h_1 - a\,h_2 \to 0$ the gains k_1 and k_2 approach infinity. This is due to the fact that $b\,h_1 - a\,h_2$ is the determinant of the observability matrix.

Therefore, as observability slips away the gains must increase in order to "see" the states. This can have a negative effect for noisy systems.

Discrete-Time Estimator (i)

- Discrete-time estimator is usually given by two coupled equations

+ : you have measurement

− : No measurement

$$\hat{\mathbf{x}}_{k+1}^{-} = \Phi\,\hat{\mathbf{x}}_k^{+} + \Gamma\,\mathbf{u}_k$$
$$\hat{\mathbf{x}}_k^{+} = \hat{\mathbf{x}}_k^{-} + K[\tilde{\mathbf{y}}_k - H\,\hat{\mathbf{x}}_k^{-}]$$

  - The first is known as the *prediction* or *propagation* equation, and the second is known as the *update* equation
  - The truth model is given by

$$\mathbf{x}_{k+1} = \Phi\,\mathbf{x}_k + \Gamma\,\mathbf{u}_k$$
$$\mathbf{y}_k = H\,\mathbf{x}_k$$

  - Error states defined by $\tilde{\mathbf{x}}_k^{-} \equiv \hat{\mathbf{x}}_k^{-} - \mathbf{x}_k$ and $\tilde{\mathbf{x}}_k^{+} \equiv \hat{\mathbf{x}}_k^{+} - \mathbf{x}_k$
  - Substituting terms leads to

errors

$$\tilde{\mathbf{x}}_{k+1}^{-} = \Phi[I - K\,H]\tilde{\mathbf{x}}_k^{-}$$
$$\tilde{\mathbf{x}}_{k+1}^{+} = [I - K\,H]\Phi\,\tilde{\mathbf{x}}_k^{+}$$

Predicted

updated

  - Note that $\Phi[I - K\,H]$ and $[I - K\,H]\Phi$ have the same eigenvalues

Discrete-Time Estimator (ii)

- Say we have some desired characteristic equation of the form

$$d(z) = z^n + \delta_{n-1}\, z^{n-1} + \cdots + \delta_1\, z + \delta_0 = 0$$

  - Ackermann's formula for the discrete-time case is given by

$$K = d(\Phi) \begin{bmatrix} H\Phi \\ H\Phi^2 \\ H\Phi^3 \\ \vdots \\ H\Phi^n \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \equiv d(\Phi)\Phi^{-1}\mathcal{O}_d^{-1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

where $\mathcal{O}_d$ is the discrete-time observability matrix

Kalman Filter

- Introduction
  - The estimators derived previously require a desired characteristic equation in the filter dynamics
  - The answer to the obvious question "How do we choose the poles of the estimator?" is not trivial
  - The *Kalman filter* provides a rigorous theoretical approach to "place" the poles of the estimator, based upon stochastic processes for the measurement error and model error
  - We do not know the exact values for these errors; however, we do make some assumptions on the nature of the errors (e.g., a zero-mean Gaussian noise process)
  - Note: "sequential state estimation" and "filtering" are used synonymously throughout
    - Sequential state estimation is often used to not only reconstruct state variables but also "filter" noisy measurement processes

Discrete-Time Kalman Filter

- The "truth" model is given by

Process Noise

$$\mathbf{x}_{k+1} = \Phi_k \mathbf{x}_k + \Gamma_k \mathbf{u}_k + \Upsilon_k \mathbf{w}_k$$
$$\tilde{\mathbf{y}}_k = H_k \mathbf{x}_k + \mathbf{v}_k$$

Note "errors" are present in both the state and measurement models

where $\mathbf{v}_k$ and $\mathbf{w}_k$ are zero-mean Gaussian noise processes

$$E\left\{\mathbf{v}_k \mathbf{v}_j^T\right\} = \begin{cases} 0 & k \neq j \\ R_k & k = j \end{cases}, \qquad E\left\{\mathbf{w}_k \mathbf{w}_j^T\right\} = \begin{cases} 0 & k \neq j \\ Q_k & k = j \end{cases}$$

measurement noise covariance

Process Noise covariance. Tunable

- Estimator form for propagation and update are given by

$$\hat{\mathbf{x}}_{k+1}^- = \Phi_k \hat{\mathbf{x}}_k^+ + \Gamma_k \mathbf{u}_k$$
$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + K_k[\tilde{\mathbf{y}}_k - H_k \hat{\mathbf{x}}_k^-]$$

Note that this is the same as before!

- Error covariances

Bound error between estimate & truth

$$P_k^- \equiv E\left\{\tilde{\mathbf{x}}_k^- \tilde{\mathbf{x}}_k^{-T}\right\}, \quad \tilde{\mathbf{x}}_k^- \equiv \hat{\mathbf{x}}_k^- - \mathbf{x}_k$$
$$P_k^+ \equiv E\left\{\tilde{\mathbf{x}}_k^+ \tilde{\mathbf{x}}_k^{+T}\right\}, \quad \tilde{\mathbf{x}}_k^+ \equiv \hat{\mathbf{x}}_k^+ - \mathbf{x}_k$$

Error Equations

- Note on covariance P (Error Covariance)
  - Provides information on the state estimation errors
  - Get 3σ bounds from diagonal elements $3\sigma = 3\sqrt{\text{diag}(P)}$
- Propagation error equation

$$
\begin{aligned}
\tilde{\mathbf{x}}_{k+1}^- &\equiv \hat{\mathbf{x}}_{k+1}^- - \mathbf{x}_{k+1} \\
&= \Phi_k \hat{\mathbf{x}}_k^+ + \cancel{\Gamma_k \mathbf{u}_k} - \Phi_k \mathbf{x}_k - \cancel{\Gamma_k \mathbf{u}_k} - \Upsilon_k \mathbf{w}_k \\
&= \Phi_k \tilde{\mathbf{x}}_k^+ - \Upsilon_k \mathbf{w}_k
\end{aligned}
$$

- Update error equation

$$
\begin{aligned}
\tilde{\mathbf{x}}_k^+ &\equiv \hat{\mathbf{x}}_k^+ - \mathbf{x}_k \\
&= \hat{\mathbf{x}}_k^- + K_k [\tilde{\mathbf{y}}_k - H_k \hat{\mathbf{x}}_k^-] - \mathbf{x}_k \\
&= \hat{\mathbf{x}}_k^- - \mathbf{x}_k + K_k [H_k \mathbf{x}_k + \mathbf{v}_k - H_k \hat{\mathbf{x}}_k^-] \\
&= \tilde{\mathbf{x}}_k^- - K_k H_k \tilde{\mathbf{x}}_k^- + K_k \mathbf{v}_k \\
&= (I - K_k H_k) \tilde{\mathbf{x}}_k^- + K_k \mathbf{v}_k
\end{aligned}
$$

Discrete Kalman Filter (i)

- Propagation error equation

$$\tilde{\mathbf{x}}_{k+1}^{-} = \Phi_k \tilde{\mathbf{x}}_k^{+} - \Upsilon_k \mathbf{w}_k$$

  - We see that $\mathbf{w}_k$ and $\tilde{\mathbf{x}}_k^{+}$ are uncorrelated since $\tilde{\mathbf{x}}_{k+1}^{-}$ (not $\tilde{\mathbf{x}}_k^{+}$) directly depends on $\mathbf{w}_k$
  - So $E\left\{\tilde{\mathbf{x}}_k^{+}\mathbf{w}_k^T\right\} = E\left\{\mathbf{w}_k\tilde{\mathbf{x}}_k^{+T}\right\} = 0$
  - Propagation covariance

$$P_{k+1}^{-} = E\left\{\Phi_k \underbrace{\tilde{\mathbf{x}}_k^{+}\tilde{\mathbf{x}}_k^{+T}}_{P_k^+}\Phi_k^T\right\} - \cancelto{0}{E\left\{\Phi_k \tilde{\mathbf{x}}_k^{+}\mathbf{w}_k^T\Upsilon_k^T\right\}}$$

$$- \cancelto{0}{E\left\{\Upsilon_k \mathbf{w}_k \tilde{\mathbf{x}}_k^{+T}\Phi_k^T\right\}} + E\left\{\Upsilon_k \underbrace{\mathbf{w}_k\mathbf{w}_k^T}_{Q_k}\Upsilon_k^T\right\} = E\{\tilde{x}_{k+1}\tilde{x}_{k+1}^T\}$$

  - This gives

$$\boxed{P_{k+1}^{-} = \Phi_k P_k^{+}\Phi_k^T + \Upsilon_k Q_k \Upsilon_k^T}$$

Discrete Kalman Filter (ii)

- Update error equation

$$\tilde{\mathbf{x}}_k^+ = (I - K_k H_k)\tilde{\mathbf{x}}_k^- + K_k \mathbf{v}_k$$

  - We see that $\mathbf{v}_k$ and $\tilde{\mathbf{x}}_k^-$ are uncorrelated since $\tilde{\mathbf{x}}_k^+$ (not $\tilde{\mathbf{x}}_k^-$) directly depends on $\mathbf{v}_k$
  - So $E\left\{\tilde{\mathbf{x}}_k^- \mathbf{v}_k^T\right\} = E\left\{\mathbf{v}_k \tilde{\mathbf{x}}_k^{-T}\right\} = 0$
  - Update covariance

$$\begin{aligned} P_k^+ &\equiv E\left\{\tilde{\mathbf{x}}_k^+ \tilde{\mathbf{x}}_k^{+T}\right\} \\ &= E\left\{(I - K_k H_k)\, \overbrace{\tilde{\mathbf{x}}_k^- \tilde{\mathbf{x}}_k^{-T}}^{P_k^-} (I - K_k H_k)^T\right\} \\ &+ E\left\{(I - K_k H_k)\, \tilde{\mathbf{x}}_k^- \mathbf{v}_k^T K_k^T\right\} \\ &+ E\left\{K_k \mathbf{v}_k \tilde{\mathbf{x}}_k^{-T} (I - K_k H_k)^T\right\} + E\left\{K_k \overbrace{\mathbf{v}_k \mathbf{v}_k^T}^{R_k} K_k^T\right\} \end{aligned}$$

$$\boxed{P_k^+ = [I - K_k H_k] P_k^- [I - K_k H_k]^T + K_k R_k K_k^T}$$

Want to use this form, Better conditioned

  - Valid for any K_k $\rightarrow$ Joseph's Stabilized Form

Kalman Gain

- Minimizing trace of covariance seems logical

$$\text{minimize} \quad J(K_k) = \text{Tr}(P_k^+)$$

  - Solution found by taking partial

$$\frac{\partial J}{\partial K_k} = 0 = -2(I - K_k H_k)P_k^- H_k^T + 2K_k R_k$$

Acheives Cramer Bound

  - Solving for K_k gives

$$\boxed{K_k = P_k^- H_k^T [H_k P_k^- H_k^T + R_k]^{-1}}$$

Have to use this K_k

  - Then we have

$$P_k^+ = P_k^- - K_k H_k P_k^- - P_k^- H_k^T K_k^T + K_k [H_k P_k^- H_k^T + R_k] K_k^T$$
$$= P_k^- - K_k H_k P_k^-$$

$$\boxed{P_k^+ = [I - K_k H_k] P_k^-}$$

standard form

  - Only valid for the specific K_k shown above

Alternative Kalman Forms (i)

- The gain can also be written as

$$\boxed{K_k = P_k^+ H_k^T R_k^{-1}}$$

  - Useful for analysis purposes (not discussed here though)
- Rewrite the covariance update as

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

  - Also, state update can be written as

$$\hat{\mathbf{x}}_k^+ = [I - K_k H_k]\hat{\mathbf{x}}_k^- + K_k \tilde{\mathbf{y}}_k$$

  - Substituting gain and $I - K_k H_k$ into the above yields

$$\hat{\mathbf{x}}_k^+ = P_k^+ \left[\left(P_k^-\right)^{-1} \hat{\mathbf{x}}_k^- + H_k^T R_k^{-1} \tilde{\mathbf{y}}_k\right]$$

  - This not particularly useful since the inverse of P_k^- is required, but its helpfulness can be shown in the derivation of the discrete-time fixed-interval smoother

Alternative Kalman Forms (ii)

- Proof of the first form
  - Rewrite the gain as

$$
\begin{aligned}
K_k &= P_k^- H_k^T [H_k P_k^- H_k^T + R_k]^{-1} \\
&= P_k^- H_k^T R_k^{-1} R_k [H_k P_k^- H_k^T + R_k]^{-1} \\
&= P_k^- H_k^T R_k^{-1} [I + H_k P_k^- H_k^T R_k^{-1}]^{-1}
\end{aligned}
$$

  - This can be rewritten

$$
K_k [I + H_k P_k^- H_k^T R_k^{-1}] = P_k^- H_k^T R_k^{-1}
$$

  - Collecting terms now gives

$$
\begin{aligned}
K_k &= P_k^- H_k^T R_k^{-1} - K_k H_k P_k^- H_k^T R_k^{-1} \\
&= \underbrace{[I - K_k H_k] P_k^-}_{P_k^+} H_k^T R_k^{-1}
\end{aligned}
$$

Discrete-Time KF Summary

usually Scalar · I

Tuned

Model	$\mathbf{x}_{k+1} = \Phi_k \mathbf{x}_k + \Gamma_k \mathbf{u}_k + \Upsilon_k \mathbf{w}_k, \quad \mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, Q_k)$ $\tilde{\mathbf{y}}_k = H_k \mathbf{x}_k + \mathbf{v}_k, \quad \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, R_k)$ Fixed in calibration
Initialize	$\hat{\mathbf{x}}(t_0) = \hat{\mathbf{x}}_0$ $P_0 = E\left\{\tilde{\mathbf{x}}(t_0)\, \tilde{\mathbf{x}}^T(t_0)\right\}$ make it big
Gain	$K_k = P_k^- H_k^T [H_k P_k^- H_k^T + R_k]^{-1}$
Update	$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + K_k[\tilde{\mathbf{y}}_k - H_k \hat{\mathbf{x}}_k^-]$ $P_k^+ = [I - K_k H_k] P_k^-$
Propagation	$\hat{\mathbf{x}}_{k+1}^- = \Phi_k \hat{\mathbf{x}}_k^+ + \Gamma_k \mathbf{u}_k$ $P_{k+1}^- = \Phi_k P_k^+ \Phi_k^T + \Upsilon_k Q_k \Upsilon_k^T$

Some Thoughts

- The vector $\mathbf{v}_k$ is known as the *measurement noise*
  - Its covariance R_k is the measurement noise covariance
  - Usually R_k is well known from sensor calibration
  - Must be careful though because measurement noise process may not be Gaussian in nature
    - Can employ a colored-noise filter to help mitigate this though
- The vector $\mathbf{w}_k$ is known as the *process noise*
  - Its covariance Q_k is the process noise covariance
  - The larger the process noise the less we *trust* the model
  - The matrix Q_k is not easy to find in general
    - How do we know how much to trust the model???
  - Finding this matrix is often referred to as *tuning* the filter
    - Takes some experience about the problem at hand and usually is done with lots of simulations since we know the truth in this case
    - Adaptive methods can be employed to do this in realtime though

Predictor-Corrector Form

- Combining equations gives *predictor-corrector* form

$$\boxed{\begin{aligned}\hat{\mathbf{x}}_{k+1} &= \Phi_k \hat{\mathbf{x}}_k + \Gamma_k \mathbf{u}_k + \Phi_k K_k [\tilde{\mathbf{y}}_k - H_k \hat{\mathbf{x}}_k] \\ K_k &= P_k H_k^T [H_k P_k H_k^T + R_k]^{-1} \\ P_{k+1} &= \Phi_k P_k \Phi_k^T - \Phi_k K_k H_k P_k \Phi_k^T + \Upsilon_k Q_k \Upsilon_k^T\end{aligned}}$$

with $P_k = P_k^-$ and $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^-$

- Although “nice” in the sense that only three equations are required for the Kalman filter, this form is typically not used in practice because it is often desired to store both the updated and propagated values

Joseph Stabilized Form

- Standard form is given by

$$P_k^+ = [I - K_k H_k] P_k^-$$

  - Investigate numerical sensitivity by using $K_k \to K_k + \delta K_k$ and $P_k^+ \to P_k^+ + \delta P_k^+$, so $P_k^+ + \delta P_k^+ = [I - K_k H_k - \delta K_k H_k] P_k^-$

$$\boxed{\delta P_k^+ = -\delta K_k H_k P_k^-}$$

  - Errors are clearly first-order in nature
  - This will be highly sensitive to roundoff errors
- Joseph stabilized form

$$P_k^+ = [I - K_k H_k] P_k^- [I - K_k H_k]^T + K_k R_k K_k^T$$

  - Performing the same substitutions gives

$$\boxed{\delta P_k^+ = \delta K_k [H_k P_k^- H_k^T + R_k]\, \delta K_k^T}$$

  - Errors are clearly second-order in nature
  - Better than standard form but requires more computations

Example (i)

- Consider single-axis attitude estimation problem
  - Update both the attitude-angle estimates and gyro drift rate
  - Angle measurements are corrupted with noise, which can be filtered by using rate information
  - However, all gyros inherently drift over time, which degrades the rate information over time
  - Two error sources are generally present in gyros (there are others but we'll keep it simple)
    - The first is a short-term component of instability referred to as *random drift*
    - The second is a random walk component referred to as *drift rate ramp*
  - The effects of both of these noise sources on the uncertainty of the gyro outputs can be compensated using a Kalman filter with attitude measurements

Example (ii)

- The attitude rate $\dot{\theta}$ is assumed to be related to the gyro output $\tilde{\omega}$ by

$$\dot{\theta} = \tilde{\omega} - \overset{\text{drift}}{\beta} - \overset{\text{noise}}{\eta_v}$$

where β is the gyro drift rate, and η_v is a zero-mean Gaussian white-noise process with spectral density given by σ_v^2 Variance in continuous time

- The drift rate is modeled by a random walk process, given by

$$\dot{\beta} = \eta_u$$

where η_u is a zero-mean Gaussian white-noise process with spectral density given by σ_u^2 ← From manufactur same with σ_v

- The estimated states clearly follow

$$\dot{\hat{\theta}} = \tilde{\omega} - \hat{\beta}$$

$$\dot{\hat{\beta}} = 0$$

- Need to convert this to discrete-time since we'd like to use the discrete-time Kalman filter

Example (iii)

- Assuming a constant sampling interval in the gyro output, the discrete-time error propagation is given by

$$\begin{bmatrix}\theta_{k+1}-\hat{\theta}_{k+1}\\ \beta_{k+1}-\hat{\beta}_{k+1}\end{bmatrix}=\Phi\begin{bmatrix}\theta_k-\hat{\theta}_k\\ \beta_k-\hat{\beta}_k\end{bmatrix}+\begin{bmatrix}p_k\\ q_k\end{bmatrix}$$

where the state transition matrix is given by

$$\Phi=\begin{bmatrix}1 & -\Delta t\\ 0 & 1\end{bmatrix}$$

where $\Delta t = t_{k+1} - t_k$ is the sampling interval, and

$$p_k=\int_{t_k}^{t_{k+1}}\left[-\eta_v(\tau)-(t_{k+1}-\tau)\eta_u(\tau)\right]d\tau$$

$$q_k=\int_{t_k}^{t_{k+1}}\eta_u(\tau)\,d\tau$$

Example (iv)

- The process noise covariance matrix can be computed as

$$Q = \begin{bmatrix} E\left\{p_k^2\right\} & E\left\{p_k q_k\right\} \\ E\left\{q_k p_k\right\} & E\left\{q_k^2\right\} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_v^2 \Delta t + \frac{1}{3}\sigma_u^2 \Delta t^3 & -\frac{1}{2}\sigma_u^2 \Delta t^2 \\ -\frac{1}{2}\sigma_u^2 \Delta t^2 & \sigma_u^2 \Delta t \end{bmatrix}$$

  - Note that this is a constant matrix
  - Also note off-diagonal terms are nonzero
  - If the sampling interval is "small" then this matrix is well approximated using (diagonal matrix) $\Delta t^2 \simeq \Delta t^3 \approx 0$

$$Q = \Delta t \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_u^2 \end{bmatrix}$$

$= \Delta t \times$ Spectral Density

Example (v)

From external star tracker

- The attitude-angle measurement is modeled by

$$\tilde{y}_k = \theta_k + v_k$$

where v_k is a zero-mean Gaussian white-noise process with variance given by $R = \sigma_n^2$

- The discrete-time system used in the Kalman filter can now be written as

$$\mathbf{x}_{k+1} = \Phi\,\mathbf{x}_k + \Gamma\,\tilde{\omega}_k + \mathbf{w}_k$$
$$\tilde{y}_k = H\,\mathbf{x}_k + v_k$$

where $\mathbf{x} = \begin{bmatrix}\theta & \beta\end{bmatrix}^T$, $\Gamma = \begin{bmatrix}\Delta t & 0\end{bmatrix}^T$, $H = \begin{bmatrix}1 & 0\end{bmatrix}$, and $E\left\{\mathbf{w}_k\mathbf{w}_k^T\right\} = Q$

- We should note that the input to this system involves a measurement $(\tilde{\omega}_k)$, which is counterintuitive but valid in the Kalman filter form and poses no problems in the estimation process
  - Gyro measurements are given in discrete-time, which is why we converted the system to a fully discrete-time one

- Simulation Parameters
  - True rate given by $\dot{\theta} = 0.0011 \text{ rad/sec}$
  - Sampling interval is 1 second
  - The noise parameters are given by

Used in R

$$\sigma_n = 17 \times 10^{-6} \text{ rad}$$

$$\sigma_u = \sqrt{10} \times 10^{-10} \text{ rad/sec}^{3/2}$$

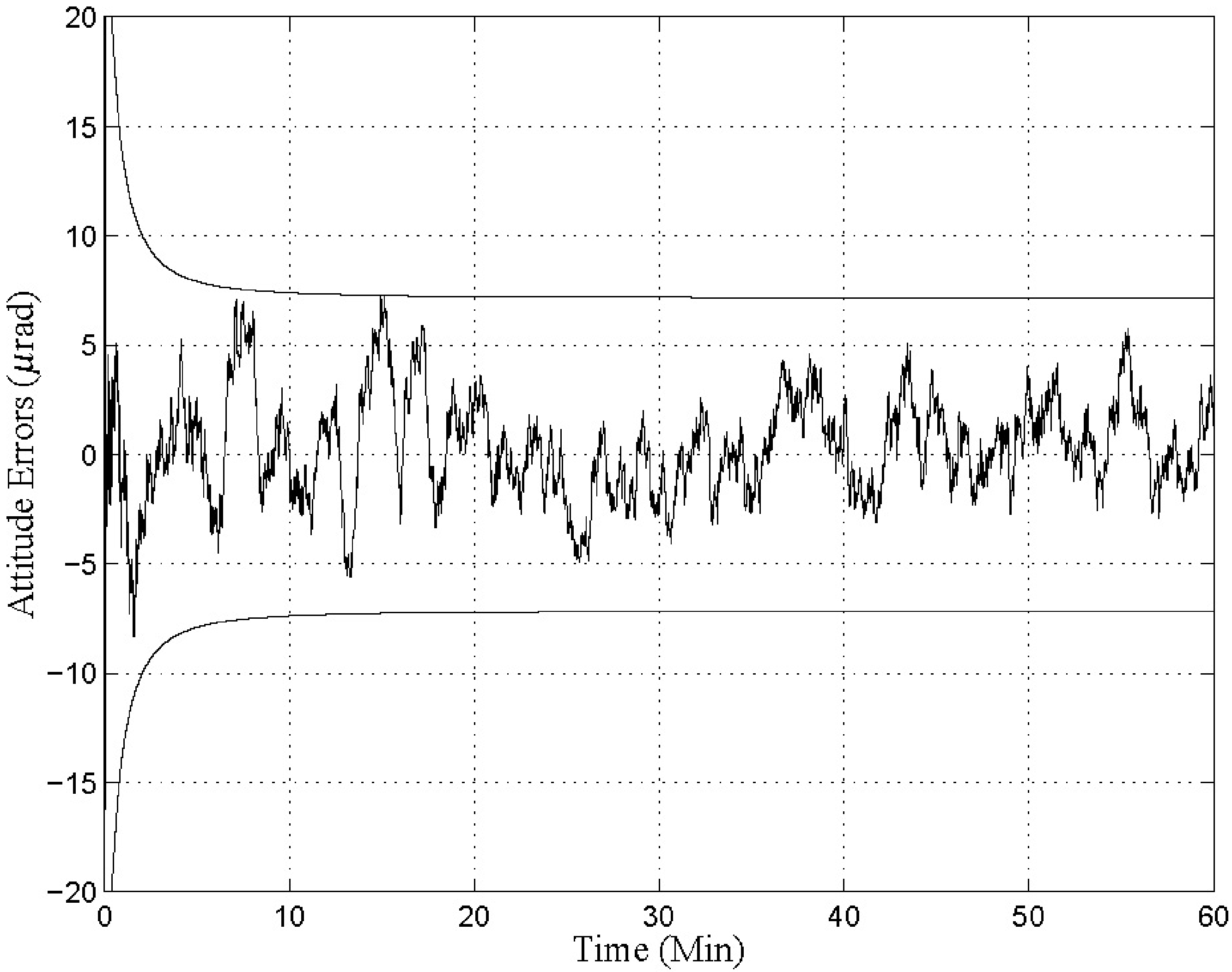
$$\sigma_v = \sqrt{10} \times 10^{-7} \text{ rad/sec}^{1/2}$$

Used in Q

  - Note this represents a good sensor (e.g. a star tracker) and a very good gyro
  - The initial drift β_0 is given as 0.1 deg/hr, and the initial covariance matrix is set to

$$P_0 = \text{diag}\left[1 \times 10^{-4} \quad 1 \times 10^{-12}\right]$$

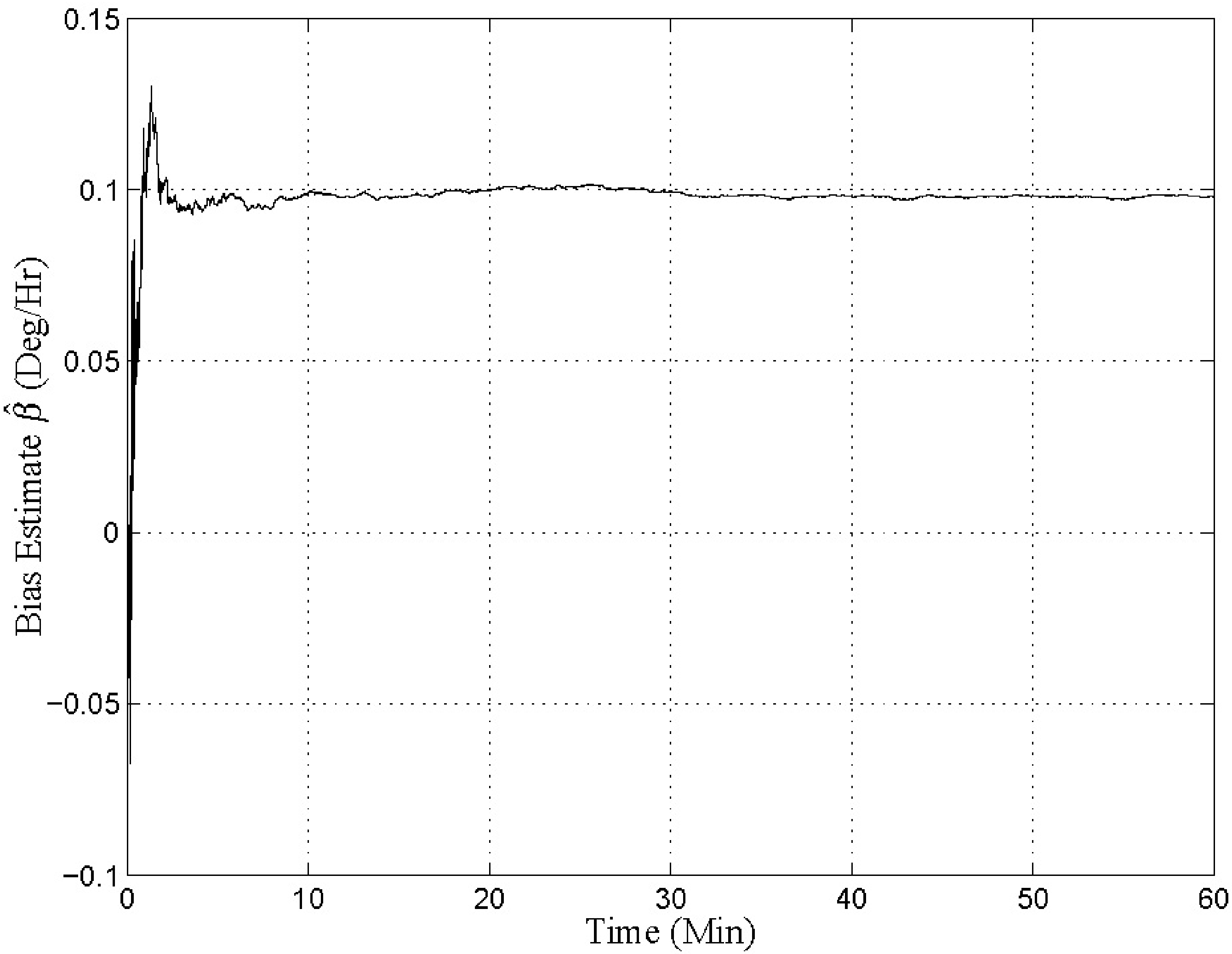
Example (vii)



3σ bounds are found from diagonal elements of P, taking the square root of each element and multiplying by 3

Note that 3σ bounds do bound the errors

Example (viii)



Drift is estimated well

Note that it's fairly constant. This is due to the fact that a high quality gyro is used

Drift is often referred as the "bias"

Example (ix)

```
% True Rate
dt=1;tf=3600;t=[0:dt:tf]';m=length(t);
wtrue=0.0011;

% Gyro and Attitude Parameters
sigu=sqrt(10)*1e-10;
sigv=sqrt(10)*1e-7;
sign=17*1e-6;

% Measurements with Bias (note: this is in discrete-time)
% Reynolds, R., "Maximum Likelihood Estimation of Stability Parameters for the Standard
% Gyroscopic Error Model," Flight Mechanics Symposium,
% Greenbelt, MD, Oct. 2003, NASA CP-2003-212246, paper #42.
ym=t*wtrue+sign*randn(m,1);
num_g=dt*[1 1];den_g=2*[1 -1];
[phi_g,gam_g,c_g,d_g]=tf2ss(num_g,den_g);
bias=dlsim(phi_g,gam_g,c_g,d_g,sigu/sqrt(dt)*randn(m,1),0.1*pi/180/3600/dt);
wm=wtrue+sqrt(sigv^2/dt+1/12*sigu^2*dt)*randn(m,1)+bias;
```

Example (x)

```
% Discrete-Time Process Noise Covriance
q=[sigv^2*dt+1/3*sigu^2*dt^3 -1/2*sigu^2*dt^2;-1/2*sigu^2*dt^2 sigu^2*dt];
phi=[1 -dt;0 1];gam=[dt;0];

% Initial Covariance
poa=1e-4;
pog=1e-12;
p=[poa 0;0 pog];
pcov=zeros(m,2);pcov(1,:)=[poa pog];

% Initial Condition and H Matrix (constant)
x0=[ym(1);0];xe=zeros(m,2);xe(1,:)=x0';x=x0;
h=[1 0];
```

Example (xi)

```
% Main Loop
for i = 1:m-1

% Kalman Gain
gain=p*h'*inv(h*p*h'+sign^2);

% Update
x=x+gain*(ym(i)-h*x);
p=[eye(2)-gain*h]*p;

% Propagate
x=phi*x+gam*wm(i);
p=phi*p*phi'+q;

% Store Variables
xe(i+1,:)=x';
pcov(i+1,:)=diag(p)';

end
```

Example (xii)

```
% 3-Sigma Outlier
sig3=pcov.^(0.5)*3;

% Plot Results
plot(t/60,[sig3(:,1) xe(:,1)-t*wtrue -sig3(:,1)]*1e6)
axis([0 60 -20 20]);grid
set(gca,'Fontsize',12);set(gca,'Fontname','Helvetica');
xlabel('Time (Min)');
hh=get(gca,'Ylabel');
set(hh,'String','\fontsize{12} {Attitude Error ({\mu}rad)}');

disp(' Press any key to continue')
pause

plot(t/60,xe(:,2)*180*3600/pi);grid
set(gca,'Fontsize',12);set(gca,'Fontname','Helvetica');
xlabel('Time (Min)');
ylabel('Bias Estimate (Deg/Hr)')
```