ASEN 5044, Fall 2018

Statistical Estimation for Dynamical Systems

Lecture 29: Chi-square NEES/NIS Consistency Tests; KF Tuning

Prof. Nisar Ahmed (Nisar.Ahmed@Colorado.edu)

Thurs 11/15/2018

University of Colorado Boulder

Aerospace Engineering

Announcements

- Midterm 2 due now

- **Final project groups now finalized (no changes possible from now on)**

- **Next steps for HW 8 and final project:**
  - **HW 8 [due 12/4]** = group assignment (do a KF on common system, then pick your final project system and do some initial deterministic stuff with it)
  - Final project report [**due by Tue 12/18**] = non-linear filtering and filter consistency analysis
  - Final description of candidate systems to be posted

Overview

Last Time:

- How to tell if your (linear) KF is actually working correctly?
- Dynamic consistency conditions
- Normalized estimation error squared (NEES) and normalized innovation squared (NIS)
- Chi-squared pdfs for linear KF NEES and NIS statistics

Today:

- **KF dynamic consistency analysis and "Truth Model Testing" (TMT)**
- **Chi-square tests for NEES & NIS** – check if KF's state errors/measurement residuals make sense for given system + measurement + noise models
  - **Do actual state errors/meas. residuals agree with KF's estimated error covariances?**
  - Formal statistical tests to examine this question
- **KF Tuning**: how to pick/adjust Q_{KF} (KF's guess of process noise)?

READ SIMON BOOK, CHAPTERS 13.1-13.2 (nonlinear KFs)

Optional: read Chapter 6.3-6.5, Chapter 7.1-7.6

<u>From Last Time</u>: Theoretical KF NEES and NIS PDFs

- So, combining Facts #1 and #2, we deduce the following must be true:

If the KF works properly as per our DT state space model and noise specs (i.e. if it meets the consistency criteria #1-#3 laid out earlier), <u>then we must have:</u>

$$\text{I. if } e_{x,k}(= x_k - \hat{x}_k^+) \sim \mathcal{N}(0, P_k^+) \text{ and } \epsilon_{x,k} = e_{x,k}^T (P_k^+)^{-1} e_{x,k} \text{ (NEES)}$$

$$\rightarrow \text{then } \epsilon_{x,k} \sim \chi_n^2 \ \forall k, \ \text{ where } E[\epsilon_{x,k}] = n, \ \text{var}(\epsilon_{x,k}) = 2n$$

$$\text{II. if } e_{y,k}(= y_k - \hat{y}_k^-) \sim \mathcal{N}(0, S_k) \text{ and } \epsilon_{y,k} = e_{y,k}^T (S_k)^{-1} e_{y,k} \text{ (NIS)}$$

$S_k = H P_k^- H^T + R$

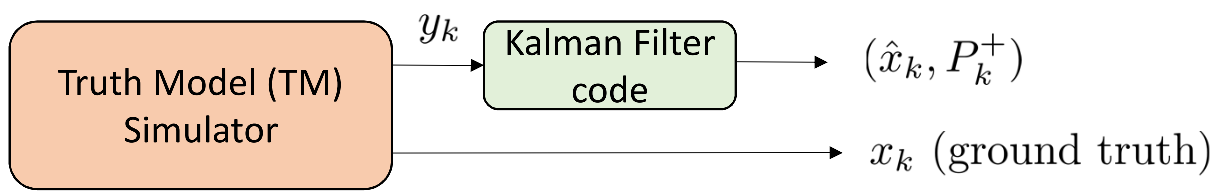
$$\rightarrow \text{then } \epsilon_{y,k} \sim \chi_p^2 \ \forall k, \ \text{ where } E[\epsilon_{y,k}] = p, \ \text{var}(\epsilon_{y,k}) = 2p$$

<u>Practical Upshot:</u> We can use "**truth model testing" (TMT) with NEES**
and use **real/simulated sensor data with NIS**
to see **if these pdfs actually show up with our KFs!**
→ if NOT, then we did something wrong!! (necessary but not sufficient conditions)

1st Statistical Test for KF Performance: NEES Chi-square

- Use "truth model test" (TMT) simulation to assess validity of NEES at every time step over N Monte Carlo runs

<u>TM = High-fidelity system dynamics + sensor model</u>: can include all kinds of non-linearities and other perversions of the actual physical system that we want to consider



Compute NEES $\epsilon_{x,k}$ and assess dynamical consistency conditions (1)-(3): <u>do results look right</u>?

Statistical consistency can be assessed formally via hypothesis testing :

"Null Hypothesis" : IF KF works properly, then $\epsilon_{x,k} \sim \chi^2_n \Rightarrow E[\epsilon_{x,k}] = n$ $\forall k$ time steps
[n = # states in x]

→ If we do N Monte Carlo sims of TM & KF (as depicted above):

(*) Does the sample average of $\epsilon_{x,k} = \epsilon_x(k)$ AT EVERY SINGLE TIME STEP K agree w/ $E[\epsilon_x(k)] = n$?

→ let $\bar{\epsilon}_{x,k} = \frac{1}{N}\sum_{i=1}^{N} \epsilon^i_{x,k}$, where $\epsilon^i_{x,k}$ is NEES for Monte Carlo sim i @ time step k
[empirical sample average]

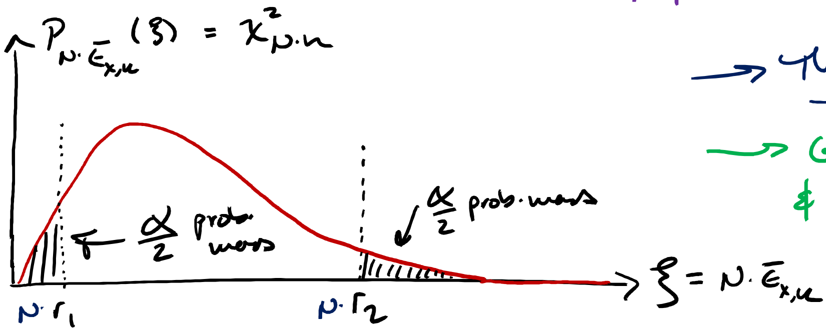
(*) Then (if Null Hyp true): $\bar{\epsilon}_{x,k} \to n$ as $N \to \infty$, if KF is working properly

NEES Chi-square Hypothesis Test

- If we have that the random variable $N \cdot \bar{\epsilon}_{x,k} \sim \chi^2_{N \cdot n}$, then it follows that

$N \cdot \bar{\epsilon}_{x,k}$ should be limited to range $N \cdot r_1 \leq N \cdot \bar{\epsilon}_{x,k} \leq N \cdot r_2$ $(1-\alpha) \cdot 100\%$ of the time

where α is determined by $\int_{N \cdot r_1}^{N \cdot r_2} p_{N \cdot \bar{\epsilon}_{x,k}}(\xi)\, d\xi \overset{=}{(1-\alpha)}$, for $p_{N \cdot \bar{\epsilon}_{x,k}} \sim \chi^2_{N \cdot n}$



$\rightarrow$ This is just a χ^2 hypothesis test!

$\rightarrow$ Given N sim runs of length T time steps & corresponding $\bar{\epsilon}_{x,k}$ values for each step $k = 1, \ldots, T$

we test the NEES statistics as follows:

- if $\bar{\epsilon}_{x,k} \in [r_1, r_2]$: declare KF consistent w/ sig. level α
- otherwise: declare KF inconsistent w/ significance level α

χ^2 hyp. test for some given α level, i.e. significance level, where α = "desired" prob. of declaring KF inconsistent (i.e. rejecting null hyp.) when, in fact, it is consistent ("false alarm prob.")

$\rightarrow$ usually: we choose r_1 & r_2 bounds s.t. $\alpha = 0.05$ or $\alpha = 0.01$

* in Matlab: r_1 = chi2inv($\frac{\alpha}{2}$, N·n)./N
r_2 = chi2inv(1-$\frac{\alpha}{2}$, N·n)./N

Example: 1D Robot Part 4: Piece de Consistance

- Same DT model as before:

$$x(k+1) = Fx(k) + Gu(k) + w(k)$$

$$x(k) = [\xi(k), \dot{\xi(k)}]^T$$

$$u(k) = 2\cos(0.75 t_k) \text{ (ZOH)}$$

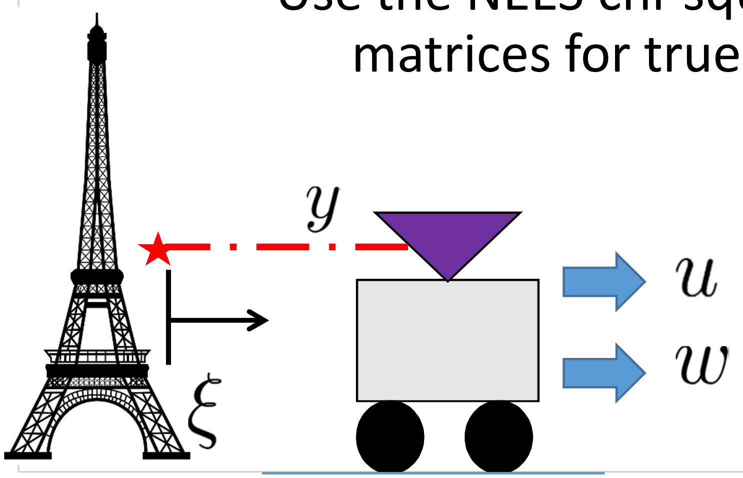
$$w(k) \sim \mathcal{N}(0, Q) \text{ (AWGN)}$$

$$x(0) \sim \mathcal{N}(\mu_0, P_0), \text{ where } \mu_0 = [0,0]^T,\ P_0 = 2I_{2\times 2}$$

$$F = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \quad G = \begin{bmatrix} 0.5\Delta t^2 \\ \Delta t \end{bmatrix} \quad \Delta t = 0.1 \text{ sec}$$

$$W = 1 \text{ (m/s)}^2,\ V = 0.5 \text{ m}^2$$

- Use the NEES chi-square test to compare effect of using different matrices for true process noise Q vs. Q_{KF} = KF's "guess" of Q

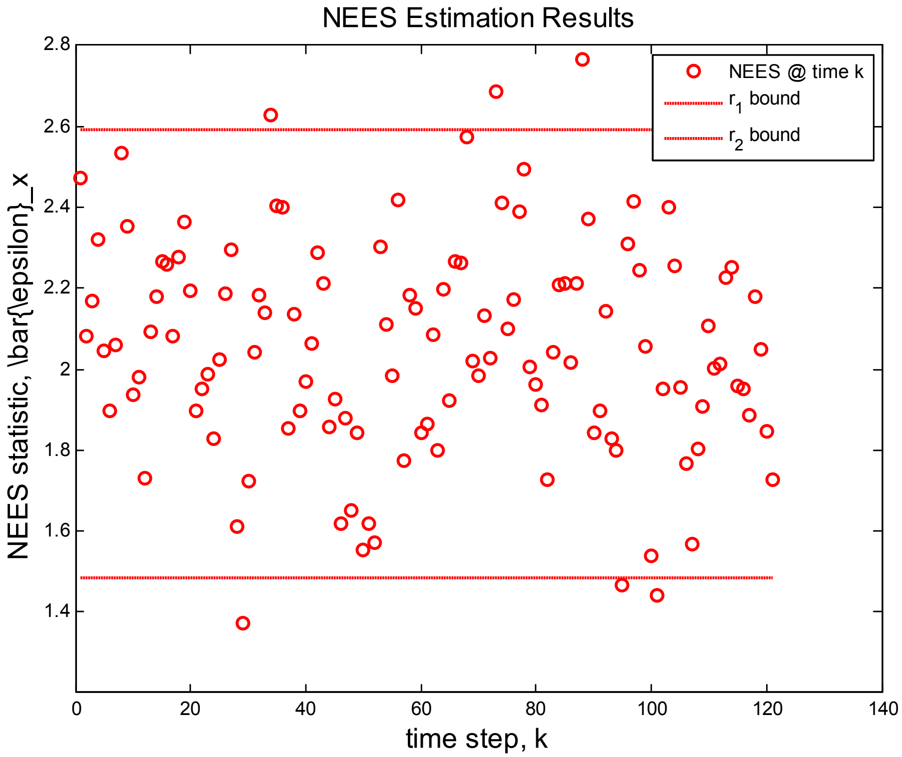


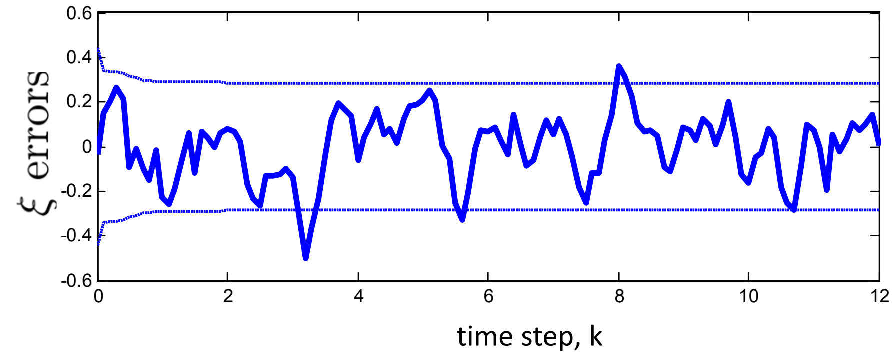
$$\text{Truth model}: \quad Q = \begin{bmatrix} 3\times 10^{-4} & 5\times 10^{-3} \\ 5\times 10^{-3} & 0.1 \end{bmatrix}$$

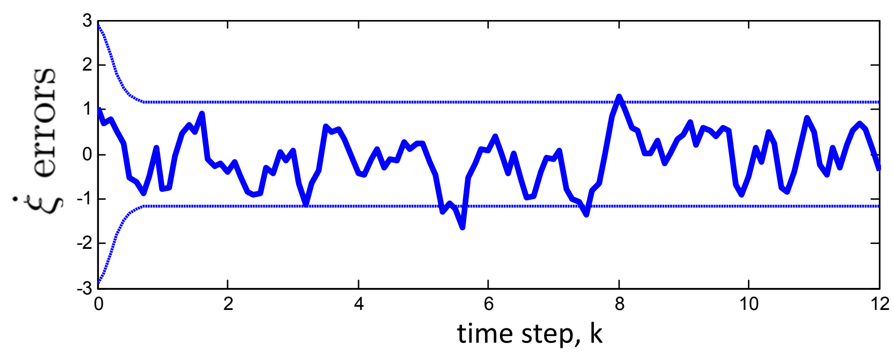
Case 1: Q_{KF} = true Q for system for W=1

- Should have following numbers for NEES test (N=50): $\alpha = 0.05$

$$\rightarrow r_1 = 1.4844$$
$$r_2 = 2.5912$$

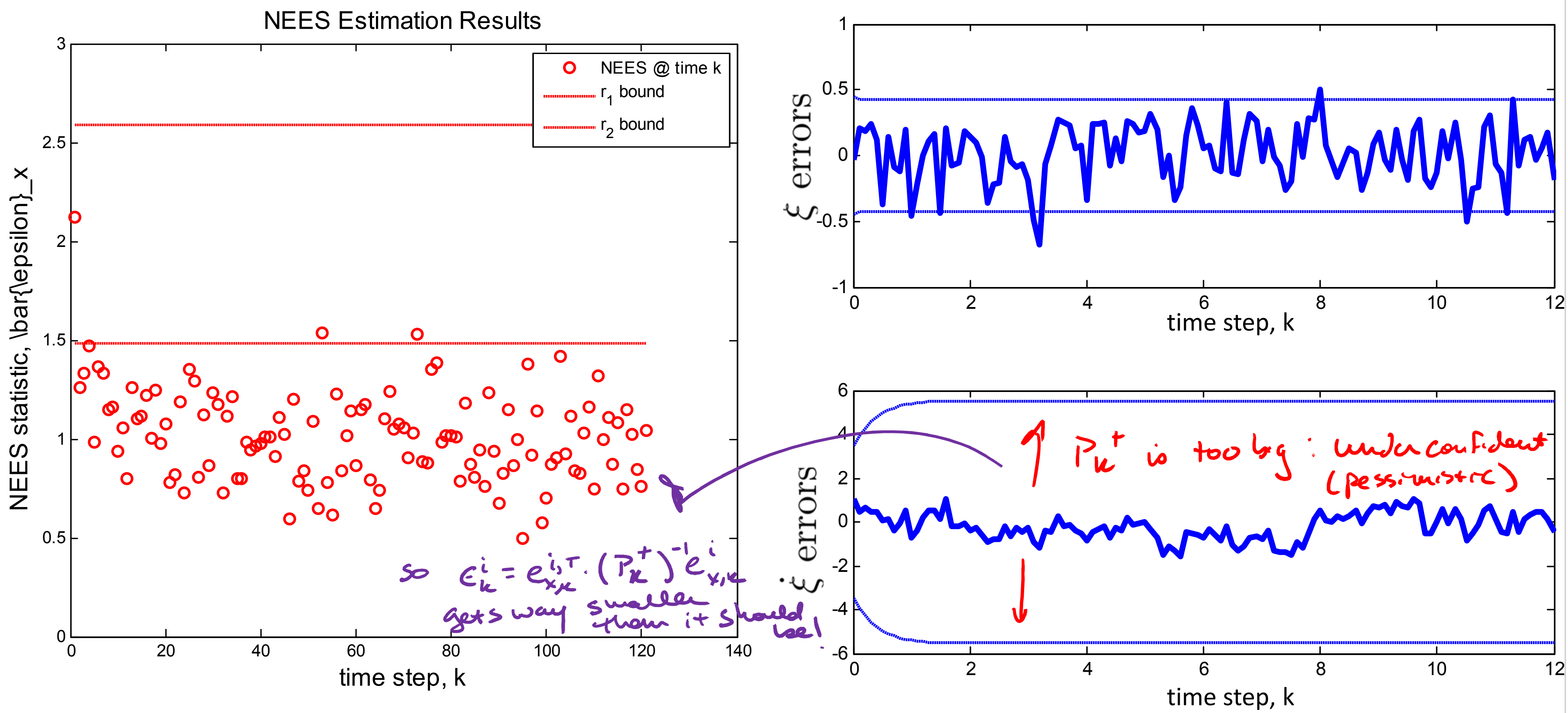






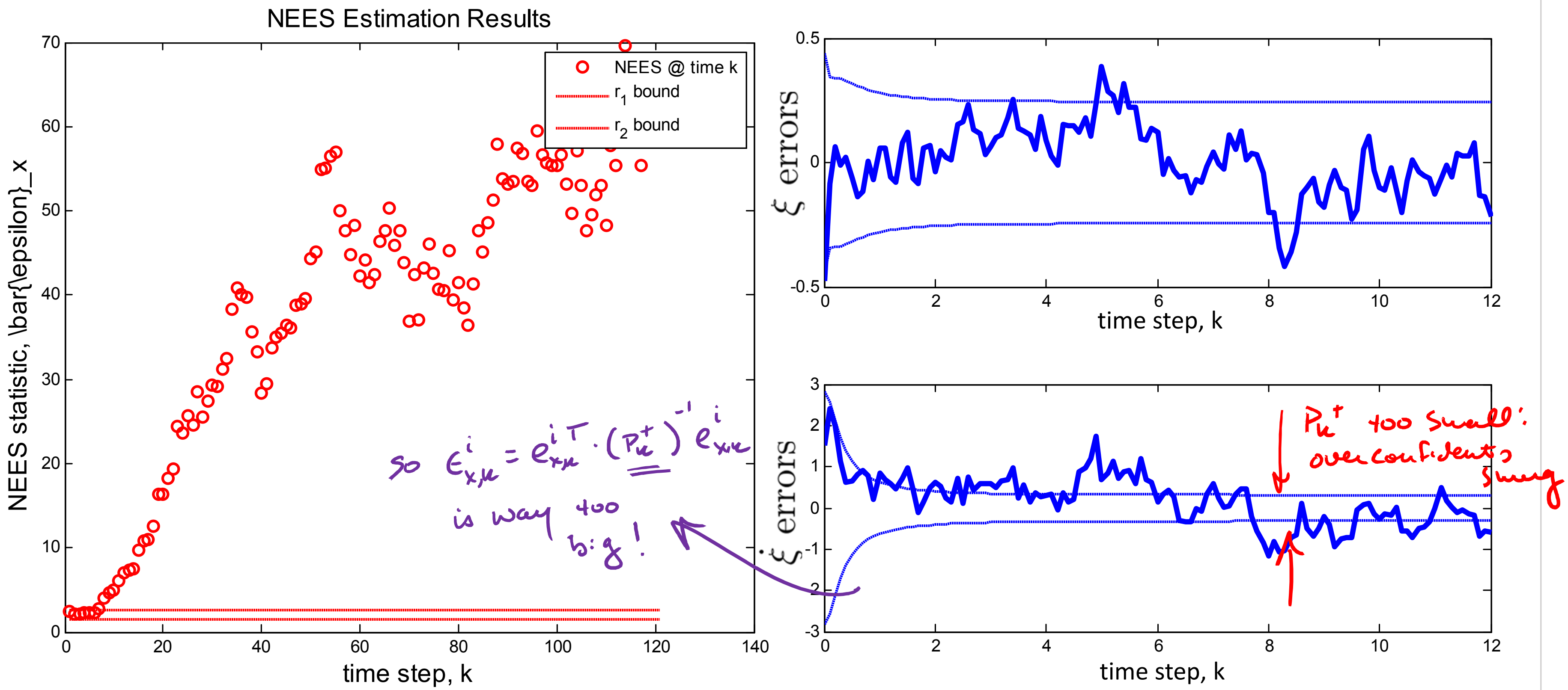
Case 2: Q_{KF} = diag([0.5, 1]) (much larger than Q_{true})

- Get same r_1 and r_2 for the same α and N, but now NEES results change



Case 3: Q_{KF} = diag([5e-3, 1e-3]) (smaller than Q_{true})

- Get same r_1 and r_2 for the same α and N, but now NEES changes



2^{nd} Statistical Test for KF Performance: NIS

- Similar test exists for looking at multiple runs of simulated or real sensor data with NIS – useful if truth model is unavailable (i.e. only real data at hand)

NIS: $\epsilon_{y,k} = e_{y,k}^T (S_k)^{-1} e_{y,k}$, where $e_{y,k} = y_k - \hat{y}_k = y_k - H\hat{x}_k^-$ (innovation vector) (meas. resid.)

$S_k = H P_k^- H^T + R$ (innovation error covar.)

Recall: $\epsilon_{y,k} \sim \chi^2_p$ $\forall$ time steps k if KF works properly (Null Hyp.)

$\longrightarrow$ can do N ⫫ runs of (real) experiment (ie get actual sensor data from real system on N field runs)

& then calculate: $\bar{\epsilon}_y(k) = \frac{1}{N}\sum_{i=1}^{N} \epsilon_y^i(k)$ [sample average of NIS @ time k across N expt. runs]

$\longrightarrow$ Analysis & χ^2 test similar to NEES: $N \cdot \bar{\epsilon}_y(k) \sim \chi^2_{N \cdot p}$, so again we have that $\bar{\epsilon}_{y,k}$ should be in the interval $r_1 \leq \bar{\epsilon}_{y,k} \leq r_2$ $(1-\alpha) \cdot 100\%$ of the time

(except: here we are testing the "zero mean-ness" & size of $e_{y,k}$, NOT $e_{x,k}$)

(*) Matlab: r_1 = chi2inv($\frac{\alpha}{2}$, N.p)./N

r_2 = chi2inv($1-\frac{\alpha}{2}$, N.p)./N

Kalman Filter Tuning Procedure

- What to do if your KF fails NEES/NIS chi-square consistency tests?
- Assume you've eliminated possibility of coding errors → then inconsistency must be explained by some error in formulation of stochastic state space model.

Strategy: Assume F, G, H, R, u (input vector) all correct (i.e. double-check these first).

Then we focus on tuning elements of Q_{KF} to pass consistency test.

(since we often know the least about Q in DT LTI model)

- If $\bar{\epsilon}_{y,k}$ 'too small' → Q_{KF} probably 'too big':

recall: $\text{NIS} = e_{y,k}^T (S_k)^{-1} e_{y,k}$, where $S_k = H P_k^- H^T + R$

→ if $e_{y,k}^T (S_k)^{-1} e_{y,k}$ too small → S_k too large → P_k^- **too large**

$K_k = P_k^- H^T S_k^{-1}$

(gain^{KF} too large or "SATURATED")

→ estimate fits data 'too well': KF not trusting dynamics model enough

→ KF adjusting $\hat{x}_k^-$ too much in response to meas. innovation $e_{y,k}$

also recall: $P_k^- = F P_{k-1}^+ F^T + Q_{KF}$ → if Q_{KF} too large, P_k^- will be too large

Tuning the KF: How to Select Q_{KF} (KF guess of Q)

- So, if $\bar{\epsilon}_{y,k}$ too small (i.e. KF too conservative/pessimistic), try decreasing Q_{KF}

  e.g $Q_{KF}^{NEW} = \frac{1}{10} \cdot Q_{KF}^{old}$ ("coarse adjust" first)

- What if $\bar{\epsilon}_{y,k}$ too big?

  $\rightarrow$ by similar logic, **KF is too optimistic/overconfident**

  $\rightarrow$ Q_{KF} **too small** (not enough prediction/model uncertainty)

  $\rightarrow$ KF does not update $\hat{x}_k^-$ enough in response to $e_{y,k}$

  $\rightarrow$ KF constantly 'too surprised' by new data y_k, but KF gain K_k too small

  $\Rightarrow$ In this case, try increasing Q_{KF} e.g. $Q_{KF}^{New} = 10 \cdot Q_{KF}^{old}$ ("coarse adjust" first)

- **In either case:** after adjusting Q_{KF} by overall factor, go in and tweak individual elements of Q_{KF} based on knowledge of system (e.g. which states you think are noisiest, most poorly modeled, etc.)