ASEN 5044, Fall 2024
Statistical Estimation for Dynamical Systems

Lecture 29:
Chi-square NEES/NIS Consistency Tests for KFs

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

Tuesday 11/12/2024

Ann and H.J. Smead
Aerospace Engineering Sciences
UNIVERSITY OF COLORADO BOULDER

Announcements

- **Midterm 2: due Thurs Nov 14 on Gradescope**
  - **HW 7 solutions posted**
- **Out Thurs 11/14: HW 8 [due 11/21]**
  - **last HW!**
- **Final project partner sign up sheet on "Assignments" tab in Canvas**
  - Google docs sheet (**Due: Mon 11/18**) – please read + follow all instructions!!
  - Folks with a partner: enter group names (**Groups of 2 <u>or</u> 3 students allowed**)
  - Folks without partner: start a new group or email each other to find a match
  - **Preview system descriptions posted on Canvas (each group must pick one)**
  - You may use Piazza to find partners (<u>but restrict messaging to final project only during Midterm 2!</u>)
  - **You must stay in your group for final assignment!!! <u>(2 interim progress reports + group gets same grade)</u>**

Overview

Last Time:

- KF gain properties/behavior
- Steady-state KF behavior: steady-state gain and error covariances
- KF error evaluation: how to tell if your (linear) KF is actually working correctly?

Today:

- Dynamic consistency conditions
- Normalized estimation error squared (NEES) and normalized innovation squared (NIS)
- Chi-squared pdfs for linear KF NEES and NIS statistics
- **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?**

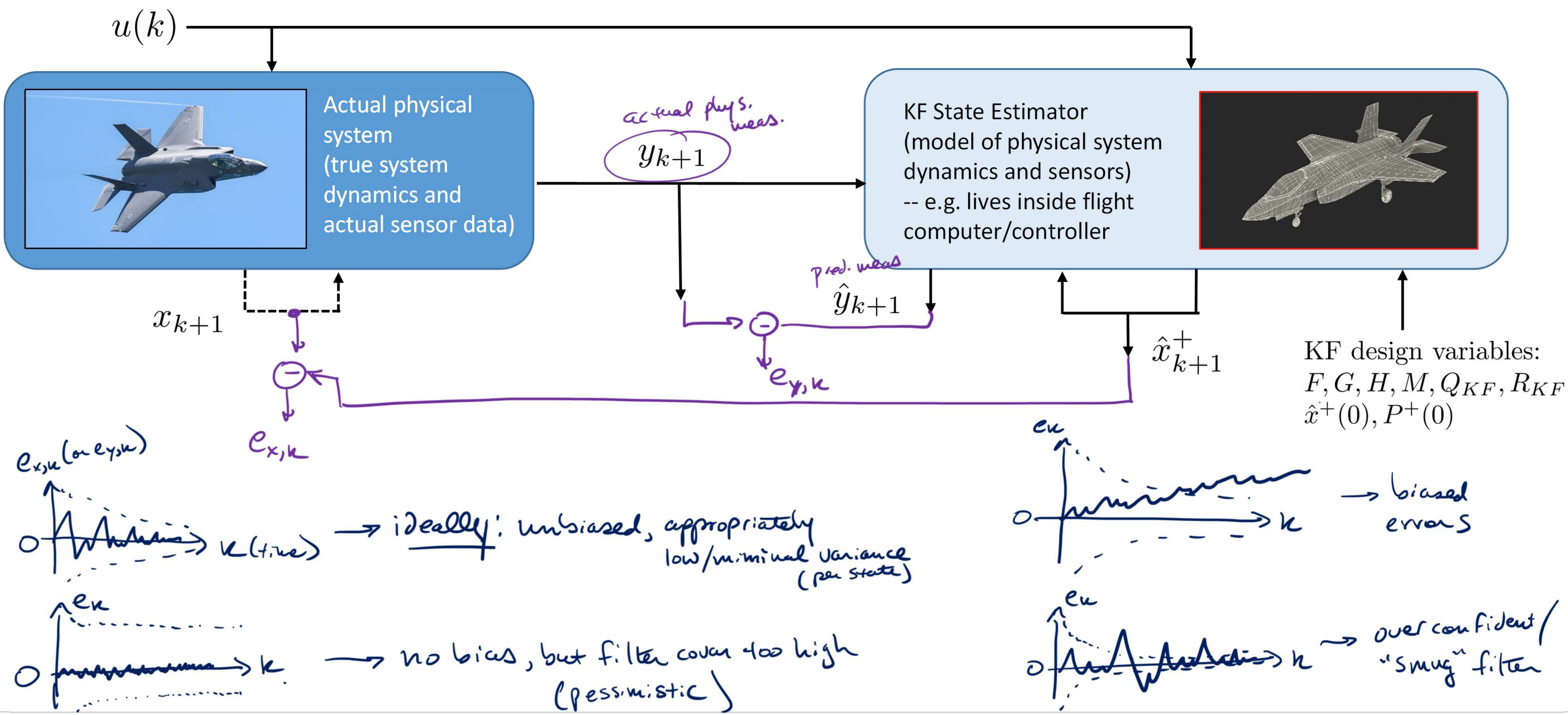
  → Derive formal statistical chi-square hypothesis tests to examine this question

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

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

(Last Time) Evaluating Errors in a KF State Estimator

- KF is itself a dynamical system (defined in software) that tries to track an actual physical system:



What Does "Minimal Error Estimation" Mean for the KF?

- In a perfect universe, we would like estimation errors to eventually vanish completely, i.e.

$$\text{if } e_{x,k} = x_k - \hat{x}_k^+, \text{ then } E[e_{x,k}] = 0 \text{ and } P_k^+ = E[e_{x,k}e_{x,k}^T] = 0 \text{ as } k \to \infty$$

(ideally: we would like perfect certainty in x_k as more data obtained)

- BUT, THIS DOES NOT HOLD AS EVIDENCED BY THE FACT THAT (IN MOST CASES) $P_\infty^- \neq 0$

(The Matrix & Algebraic Riccati Equations say so) $\longrightarrow$ so can't be perfectly certain! ($P_\infty^- > 0$ $\longrightarrow P_\infty^+ > 0$)

- What is responsible for this?: Random process noise incessently disturbs the true state x_k
$\longrightarrow \therefore e_{x,k} \not\to 0$ as $k \to \infty$! (generally speaking)

“Proper KF Error Characteristics”: Dynamic Filter Consistency

- Because some finite/non-zero error will always exist, we instead say that the KF is “working properly” (for a given DT state space model and noise specs) if it satisfies the following **3 demands for <u>dynamic filter consistency:</u>**

1) Unbiasedness: $E[e_{x,k}] = 0$ for all k

2) Efficiency: $E[e_{x,k}e_{x,k}^T] = P_k^+$ (true state errors match filter covariance)

3) KF measurement residuals/innovations are a white Gaussian sequence:

$$e_{y,k} \sim \mathcal{N}(0, S_k), \quad E[e_{y,k}e_{y,j}^T] = S_k \cdot \delta(k,j)$$

where $e_{y,k} = y_k - \hat{y}_k = y_k - H\hat{x}_k^-$,

$S_k = HP_k^- H^T + R$ (innovation covariance matrix $\in \mathbb{R}^{p \times p}$)

(Note: KF Gain $K = P_k^- H_k^T S_k^{-1}$)

KF Estimation Errors and Measurement Innovations

- How do we analyze the two types of random error vectors in a KF?
  - State estimation errors (w.r.t. ground truth x_k): $e_{x,k} = x_k - \hat{x}_k^+ \in \mathbb{R}^n \overset{?}{\sim} \mathcal{N}(0, P_k^+)$
  - Measurement innovations/residuals (w.r.t. observations y_k): $e_{y,k} = y_k - \hat{y}_k^- \in \mathbb{R}^p \overset{?}{\sim} \mathcal{N}(0, S_k)$

- Simplest check is to look at the <u>normalized magnitude</u> of these vectors over time:

$$\epsilon_{x,k} = e_{x,k}^T (P_k^+)^{-1} e_{x,k} \quad \rightarrow \textbf{Normalized estimation error squared (NEES)} \text{ at time k}$$

$$\epsilon_{y,k} = e_{y,k}^T (S_k)^{-1} e_{y,k} \quad \rightarrow \textbf{Normalized innovation squared (NIS)} \text{ at time k}$$

$\rightarrow$ NEES and NIS are both **positive random scalars:** (weighted 2-norms)^2 of $e_{x,k}$ and $e_{y,k}$!

i.e. what should $p(\epsilon_{x,k})$ & $p(\epsilon_{y,k})$ look like?

Key question: if dynamic consistency conditions (1)-(3) on previous slide hold, then <u>what pdfs ought to describe our *expected* NEES & NIS outcomes</u>?

Useful Fact #1: Squared 2-norms of Gaussian Random Vectors

- Suppose we are given some random vector $e \sim N_e(0, P_e)$, $e \in \mathbb{R}^n$

→ let $ZZ^T = (P_e)^{-1}$, where $Z \in \mathbb{R}^{n \times n}$ is cholesky factor of $(P_e)^{-1}$.
[matrix sq. root]

→ Also define: $g = Z^T e \in \mathbb{R}^n$ (linear transformation of e via Z^T)

→ Now since $e \sim N_e(0, P_e)$ & since $N_e(0, P_e) \underset{\text{pdf def.}}{=} (\text{const.}) \cdot \exp\left(-\frac{1}{2} e^T (P_e)^{-1} e\right)$

$\Rightarrow$ Apply cholesky factor.: $N_e(0, P_e) = (\text{const.}) \cdot \exp\left(-\frac{1}{2} e^T [Z Z^T] e\right)$

$\underset{\text{(apply def of } g \text{ above)}}{=} (\text{Const.}) \cdot \exp\left(-\frac{1}{2} g^T g\right) = N_g(0, I)$ [can easily show $E[gg^T] = I$]

→ it follows that $g \sim N_g(0, I)$ ie a standard multivariate normal pdf

"decorrelated"/"whitened" Gaussian random vector

→ $g_i \sim N(0,1)$ for $i = 1, \ldots, n$ (elements)

$\Rightarrow$ so $\epsilon = e^T (P_e)^{-1} e = e^T (Z Z^T) e = g^T g = \|g\|_2^2 = \sum_{i=1}^{n} g_i^2 =$ Sum of squares of n iid Gaussian random scalar variables!

Must have the same pdf! ie $p(\epsilon) = p\left(\sum_{i=1}^{n} g_i^2\right)$

Example:
Correlated and Decorrelated Gaussian Random Vectors

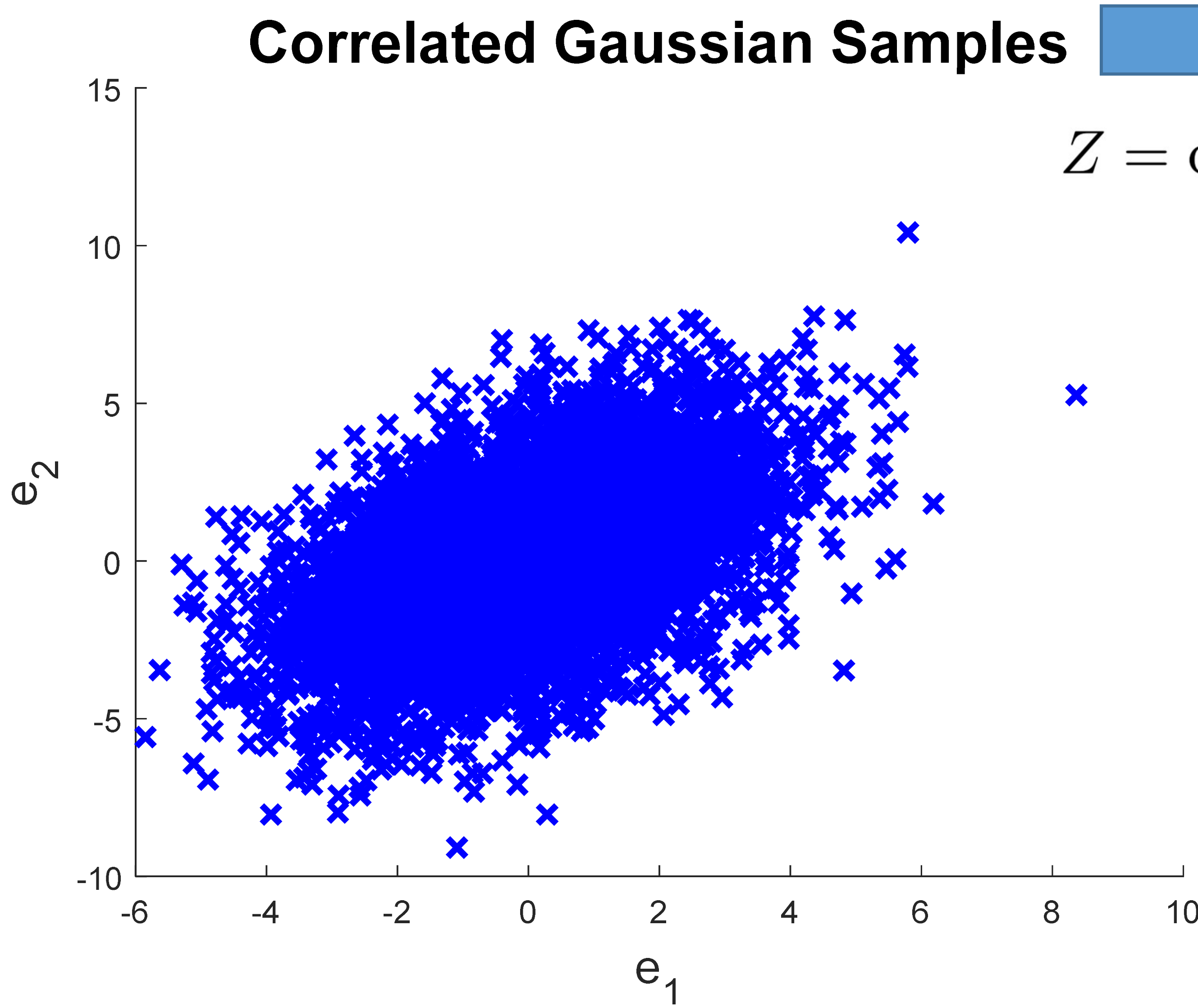
- Apply vector transform via Cholesky decomposition

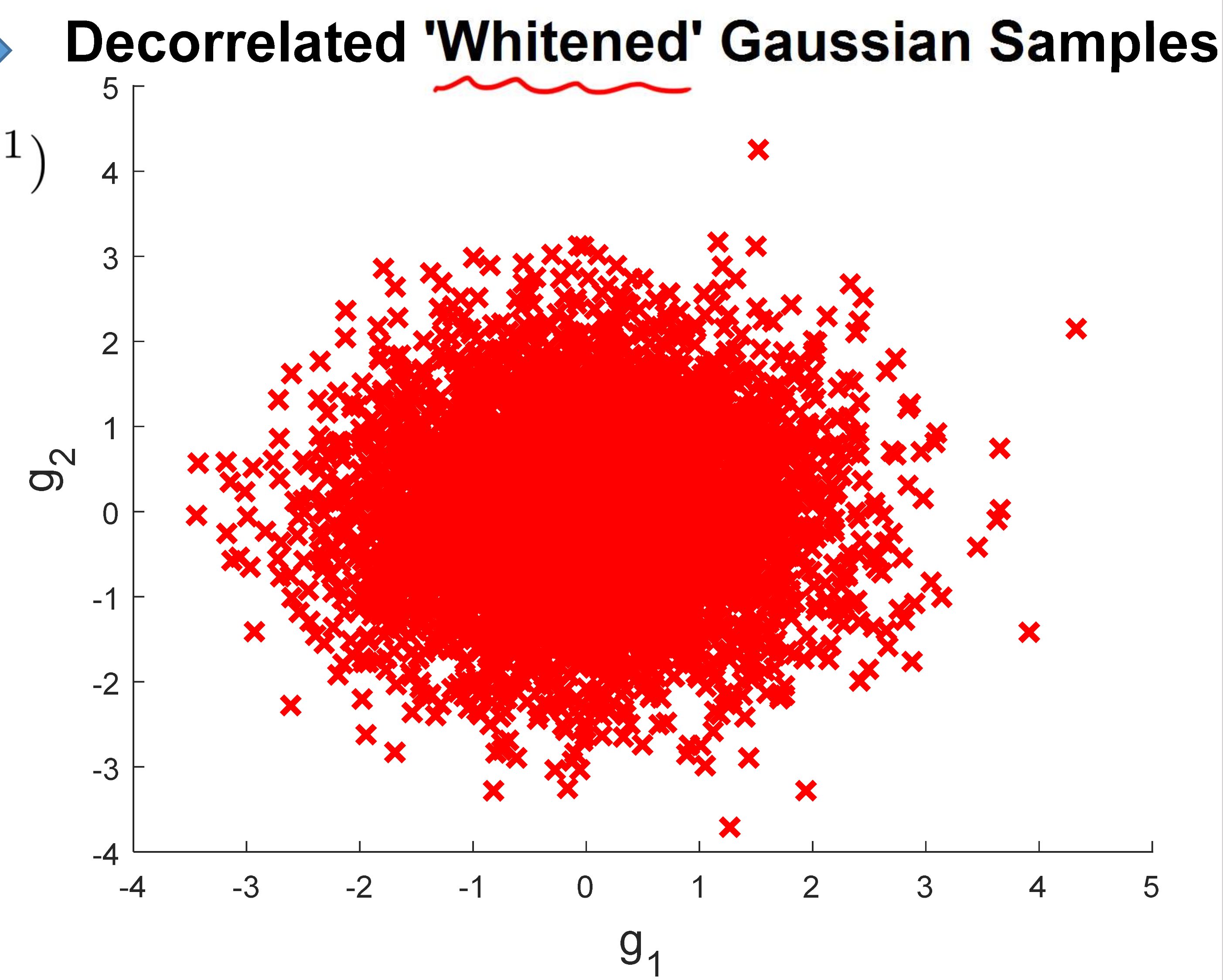
$$e \sim \mathcal{N}(0, P_e) \quad P_e = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$$

$$ZZ^T = P_e^{-1}$$

$$g = Z^T e, \text{ where } g \sim \mathcal{N}(0, I)$$

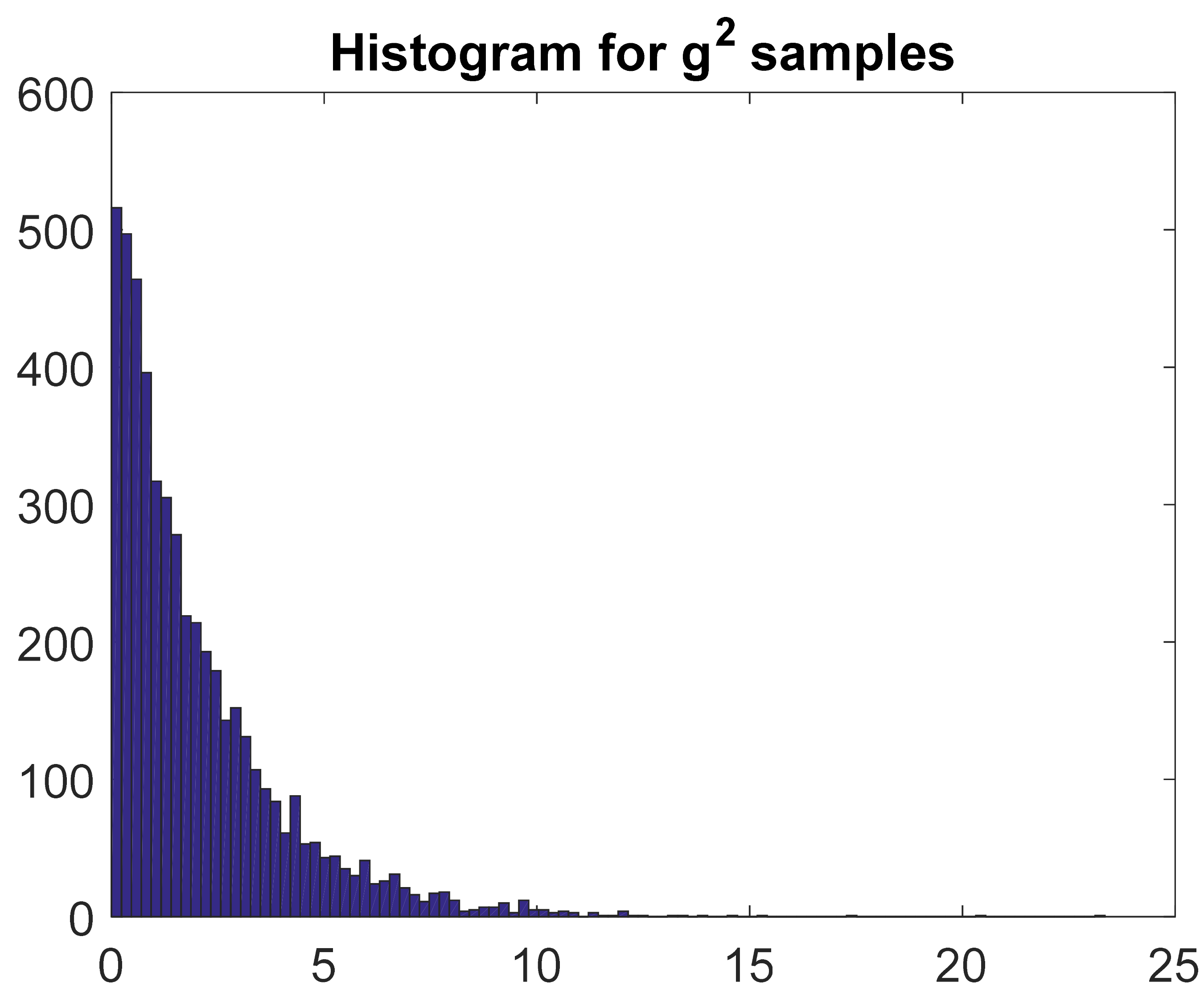
$$Z = \text{chol}(P_e^{-1})$$

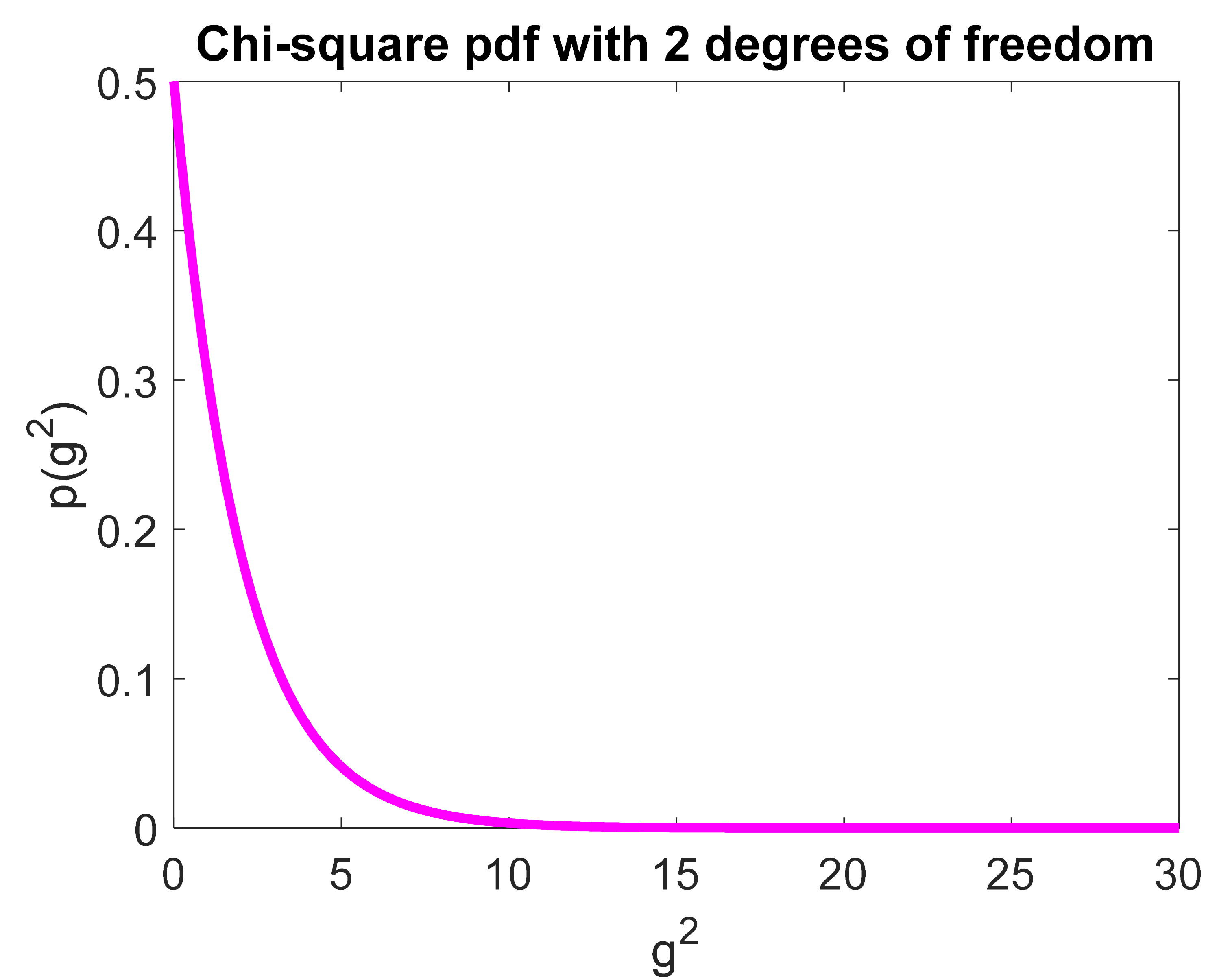




Example: Distribution of Gaussian RV Squared Magnitudes

- What does pdf look like for $g^T g = [\text{norm}(g)]^2 = g^2{}_1 + g^2{}_2$?





Useful Fact #2: The Chi-square Distribution

Suppose we have scalar i.i.d. random variables $g_1, ..., g_n$ where $g_i \sim \mathcal{N}(0,1)$ for $i = 1, ..., n$.

Define: random variable $q = \sum_{i=1}^{n} g_i^2 = \vec{g}^T \vec{g}$, where $\vec{g} = [g_1, \cdots, g_n]^T$ (note: $\vec{g} \sim \mathcal{N}(\vec{0}, I_{n \times n})$)

$\Rightarrow$ then the pdf $p(q)$ is a **chi-square** (χ^2) distribution with n degrees of freedom:

$$p(q) = \chi_n^2 = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} q^{\frac{n-2}{2}} \cdot \exp(-\frac{q}{2}), & \text{for } q \geq 0 \\ 0, & \text{for } q < 0 \end{cases}$$

where the 'gamma function' is

$\Gamma(\frac{1}{2}) = \sqrt{\pi}$

$\Gamma(1) = 1$

$\Gamma(m+1) = m \cdot \Gamma(m) = m!$ for integer m

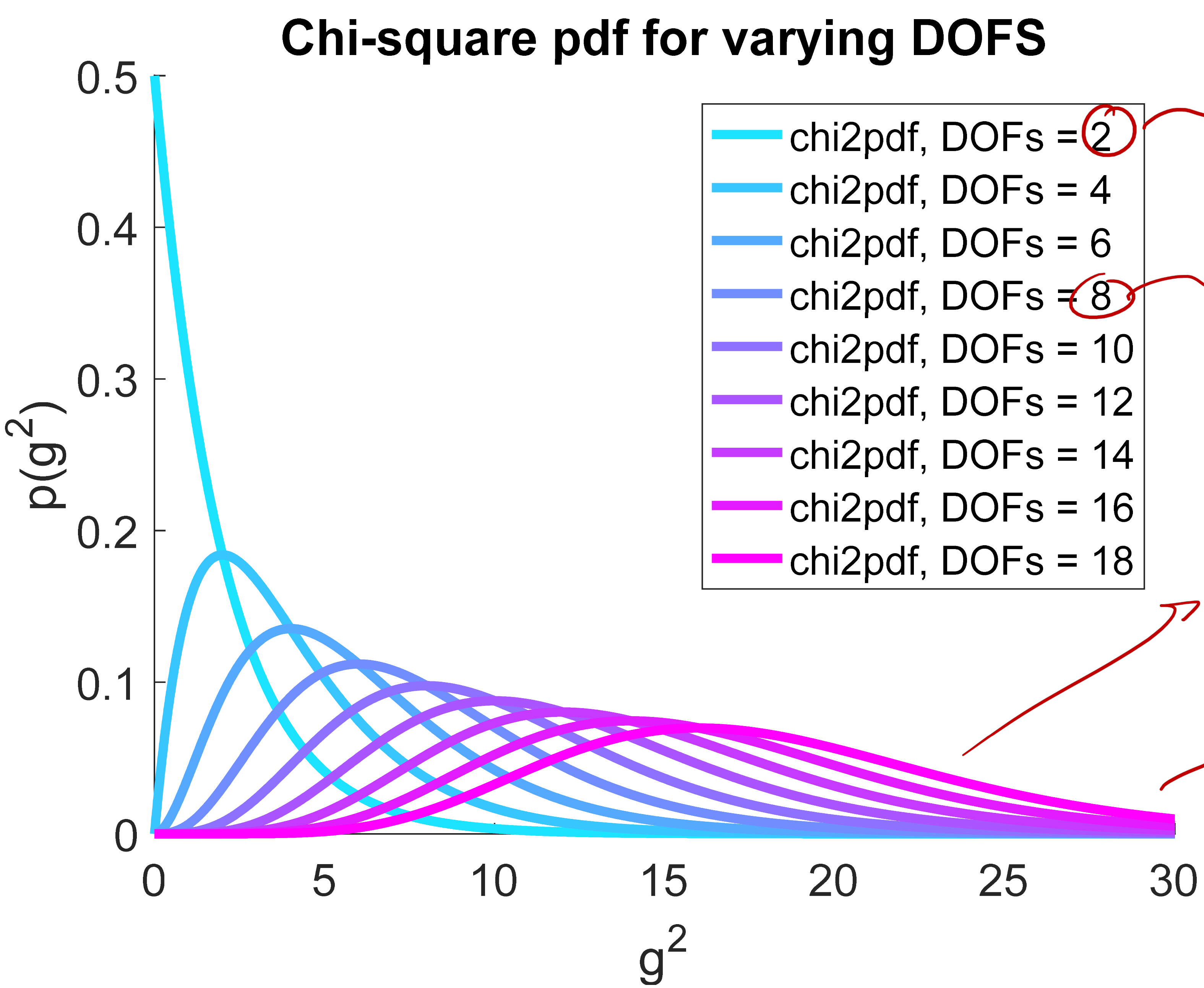
* easy to show that: $E[q] = n$, and $\mathrm{var}(q) = 2n$

* also can show: if $q_1 \sim \chi^2_{n_1}$ and $q_2 \sim \chi^2_{n_2}$, then $q_3 = q_1 + q_2 \Rightarrow q_3 \sim \chi^2_{n_1+n_2}$,

(i.e. $n_3 = n_1 + n_2$, so DOFs add!)

Example: Chi-square Distributions

- What does pdf look like for $[\text{norm}(g)]^2$ for different vector lengths n=length(e)?



$$g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \sim N(0, I_{2\times2})$$

$$g = \begin{bmatrix} g_1 \\ \vdots \\ g_8 \end{bmatrix} \sim N(0, I_{8\times8})$$

$$p(\|g\|^2) = p(g^T g) = p\left(\sum_{i=1}^{n} g_i^2\right) = p(\epsilon)$$

[Fact #1]

Note: Starts to look like a Gaussian pdf for large n DOFs! (by CLT!)

Upshot: Theoretical KF NEES and NIS Error Distributions

- 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), then we must have:

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

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

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

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

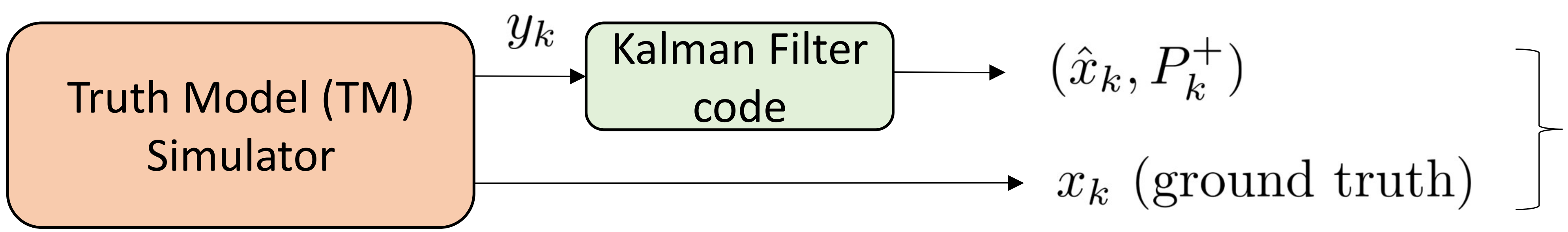
We can use "**truth model testing" (TMT)** with NEES and use real/simulated sensor data with NIS to see if these pdfs actually show up!

→ 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

TM = High-fidelity system dynamics + sensor model: 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): do results look right?

Statistical consistency can be formally assessed via **Hypothesis Testing:**

Null Hypothesis: If KF works properly, then $\epsilon_{x,k} \sim \chi^2_n \;\Rightarrow E[\epsilon_{x,k}] = n(\#\text{ states}),\; \forall k$

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

*Does the sample average of $\epsilon_x(k)$ *AT EVERY SINGLE TIME STEP* k agree with $E[\epsilon_x(k)] = n$?

N = # Monte Carlo sims

$\rightarrow$ 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 at time step k.

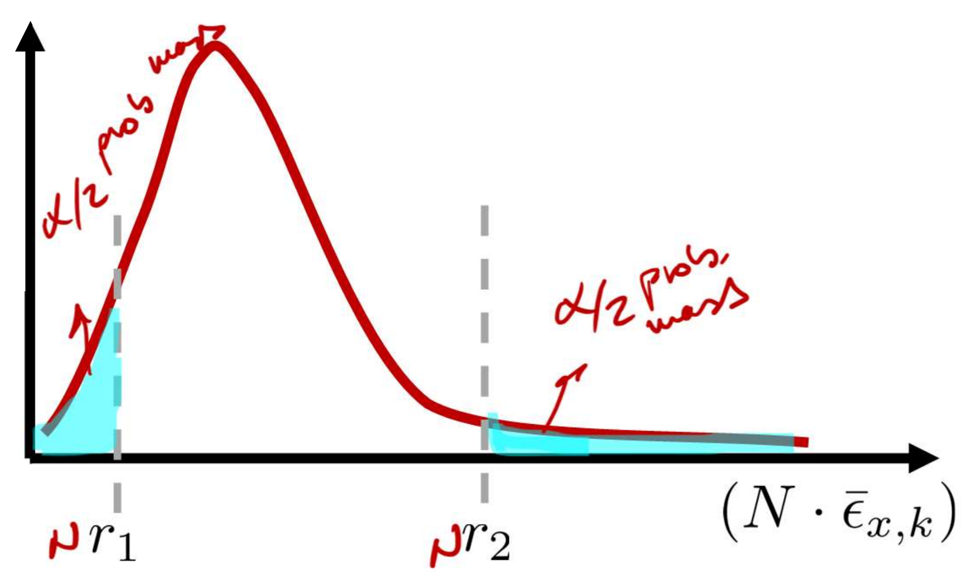
$\boxed{\text{* Then: } \bar{\epsilon}_{x,k} \rightarrow n \text{ as } N \rightarrow \infty, \text{ if KF is working properly (i.e. if our Null Hypothesis holds)}}$

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

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



χ^2 hyp. test for given sig. level α, where $\alpha =$ prob. of declaring KF inconsistent when it is in fact consistent, i.e. $\alpha =$ Type I error rate / 'False Positive' probability

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

$\rightarrow$ Given N simulation runs of length T steps *and* corresponding $\bar{\epsilon}_{x,k}$ values for *each* step $k = 1, 2, \cdots, T$, we test the NEES statistics as follows:

*if $\epsilon_{x,k} \in [r_1, r_2]$: declare KF consistent with significance level α

*otherwise: declare KF inconsistent with significance level α

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

$\rightarrow$ ** in Matlab: $r_1 = \texttt{chi2inv}(\frac{\alpha}{2}, N \cdot n)./N$

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

Example: 1D Robot Part 4: Piece de Consistance

- Same DT model as before:

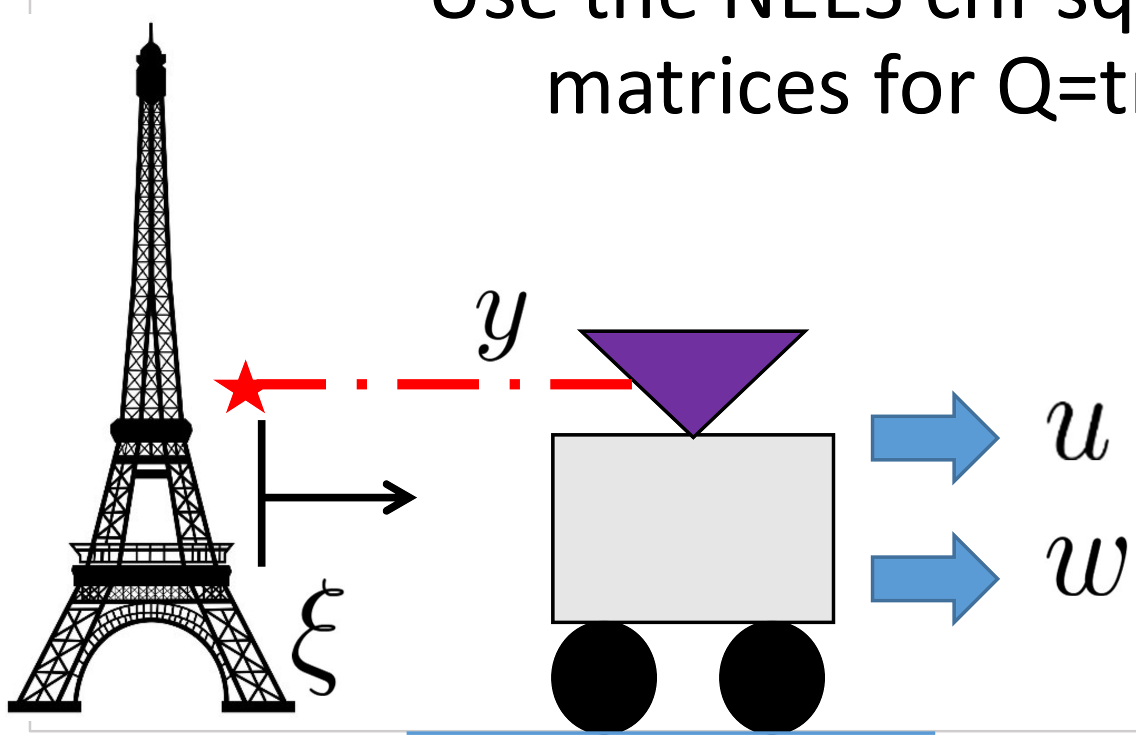
$$x(k+1) = Fx(k) + Gu(k) + w(k)$$

$$x(k) = [\xi(k), \dot{\xi}(k)]^T$$
$$u(k) = 2\cos(0.75 t_k) \ \text{(ZOH)}$$
$$w(k) \sim \mathcal{N}(0, Q) \ \text{(AWGN)}$$
$$x(0) \sim \mathcal{N}(\mu_0, P_0), \text{ where } \mu_0 = [0,0]^T, \ P_0 = 2I_{2\times 2}$$

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

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

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

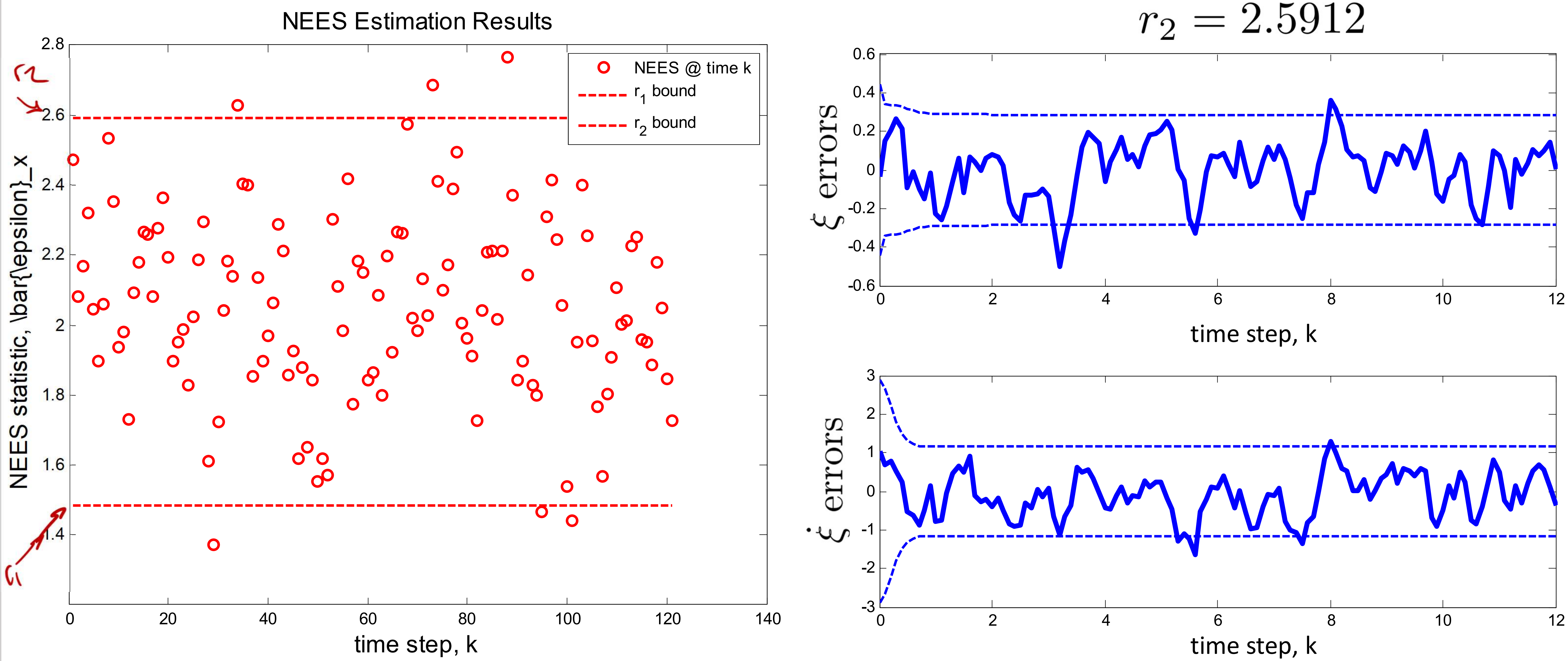


$$\text{Truth model}: \quad Q = \begin{bmatrix} 3\times 10^{-4} & 5\times 10^{-3} \\ 5\times 10^{-3} & 0.1 \end{bmatrix}$$

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

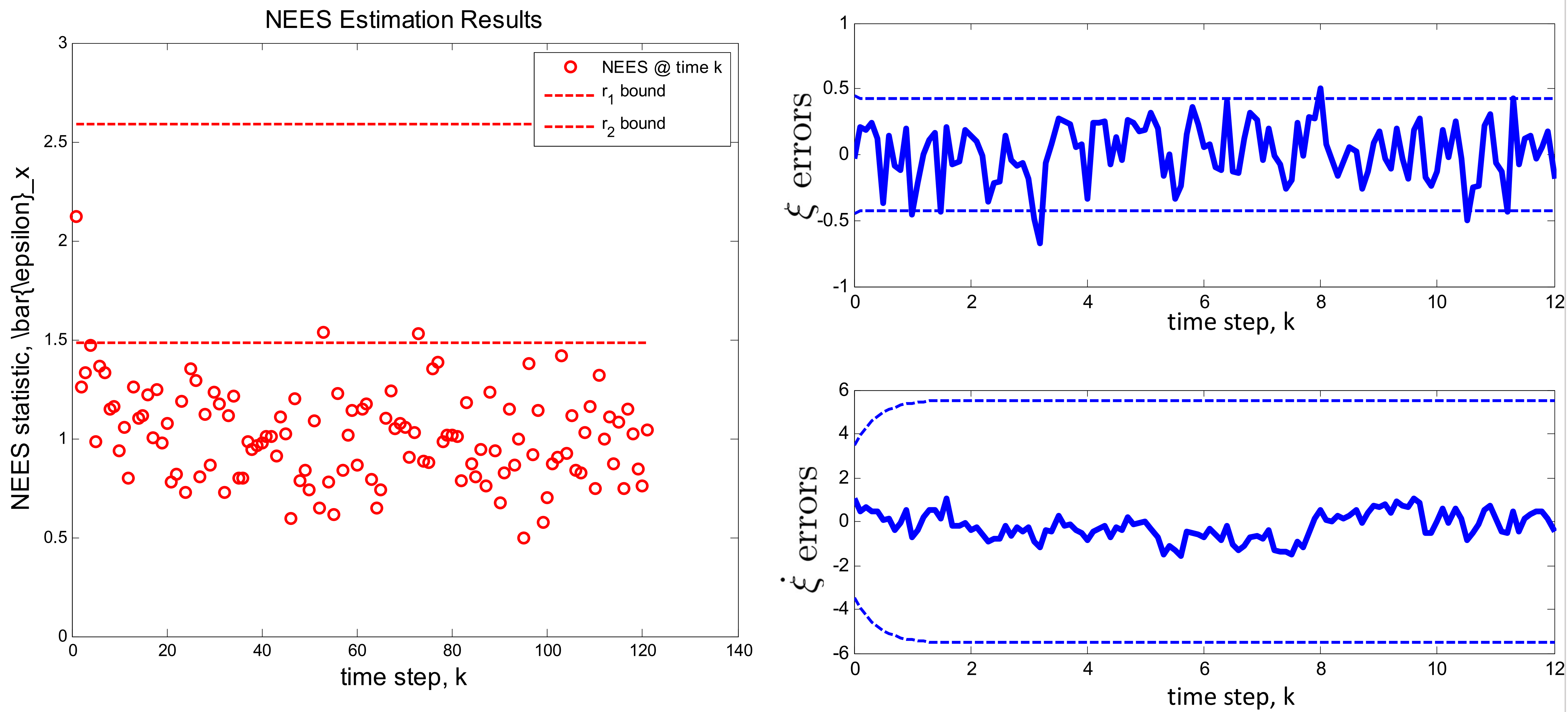
- Should have following numbers for NEES test (N=50):

$$\alpha = 0.05$$
$$\rightarrow r_1 = 1.4844$$
$$r_2 = 2.5912$$



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