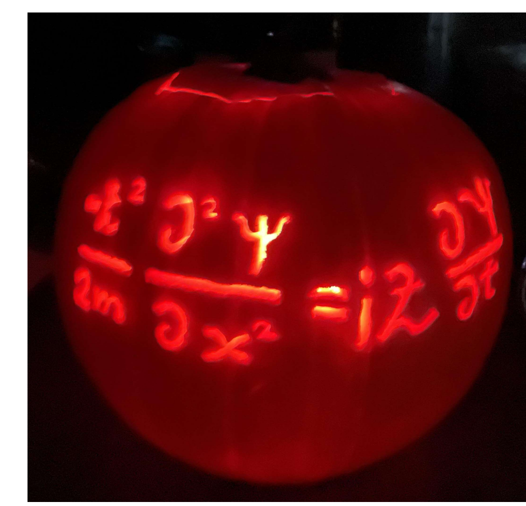
ASEN 5044, Fall 2024

Statistical Estimation for Dynamical Systems

Lecture 24:
Dynamic Non-deterministic State Estimation;
Pure State Prediction



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

Thursday 10/31/2024

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

Announcements

HW 6 due this **Fri 11/01** via Gradescope

Midterm 2: will be released Thurs Nov 7, due Thurs Nov 14

- **Cover HWs 5-7, Quizzes 5-7**

Quiz 7 out this Fri 11/01, Due Tues 11/05

HW 7 posted TODAY (Thurs 10/31), due Thurs Nov 7 (1 week!)

- **Solutions to be posted at start of Midterm 2**

Overview

Last Time
- Estimation error covariance (cont'd)
- Numerical considerations for batch LLS
- Recursive LLS (RLLS) definition, and cost function
- RLLS optimal gain derivation
- Initialization of RLLS

Today
- Analysis of Optimal RLLS gain behavior
- Start dynamic non-deterministic estimation (x_k with dynamics + process noise)
  - Pure state prediction problem: propagation of state mean + covariance without any sensor measurements

READ SIMON TEXT Chapters 4.1-4.2, 5.1-5.4

Quick Recap: Recursive LLS Formula and Cost Fxn

- Recursive Linear Least Squares (RLLS) estimator: for static unknown state x

$$\hat{x}_k = \hat{x}_{k-1} + K_k \underbrace{(y_k - H_k \hat{x}_{k-1})}_{\text{"surprise factor" = "innovation vector" = "meas. residual"}}, \quad k = 1, 2, ..., T$$

updated state estimate via RLLS @ time k

prior result of RLLS @ time k-1

n x p gain matrix: "blending factor"

$(x \in \mathbb{R}^n, y_k \in \mathbb{R}^p, \ v_k \in \mathbb{R}^p, \ H_k \in \mathbb{R}^{p\times n})$

$$\text{RLLS cost function:} \quad J(k) = J_k = \text{tr}(E[(x - \hat{x}_k)(x - \hat{x}_k)^T]) = \text{tr}(P_k)$$

$$\hat{x}_k = \arg \min_{x\in\mathbb{R}^n} J(k) = \arg \min_{x\in\mathbb{R}^n} \text{tr}(E[(x - \hat{x}_k)(x - \hat{x}_k)^T]) = \arg \min_{x\in\mathbb{R}^n} \text{tr}(P_k)$$

Showed last time that:

$$E[x - \hat{x}_k] = 0, \ \text{ if } E[x - \hat{x}_{k-1}] = 0$$ (for *any gain* K_k: $\hat{x}_k$ is unbiased if $\hat{x}_{k-1}$ is unbiased)

$$E[(x - \hat{x}_k)(x - \hat{x}_k)^T] = P_k = (I - K_k H_k) P_{k-1} (I - K_k H_k)^T + K_k R_k K_k^T$$

where optimal gain $K_k = \arg\min \text{tr}(P_k) = P_{k-1} H_k^T (R_k + H_k P_{k-1} H_k^T)^{-1}$

RLLS Initialization and Optimal Gain K_k Behavior

- What happens to Optimal RLLS Gain K_1 at k=1 if we have $P_0 = `\infty' \cdot I$?

$$K_1 = P_0 H_1^T \left[H_1 P_0 H_1^T + R_1\right]^{-1}, \text{ where } P_0 `>>' R \quad (\text{e.g. } Z = (H_1 P_0 H_1^T - R_1) \text{ is posdef})$$

$\rightarrow$ Consider scalar x and y_k case, such that K_k, H_k, and R_k are all scalars:

$$K_1 = \frac{P_0 H_1}{H_1^2 P_0 + R_1}, \text{ where } H_1^2 P_0 >> R_1 \text{ because } P_0 = `\infty' \cdot I$$

Then the denominator of K_1 is dominated by P_0 :

$$\lim_{P_0 \to \infty} K_1 = \frac{P_0 H_1}{H_1^2 P_0} = \frac{1}{H_1}$$

So the (scalar) RLLS update becomes: $\hat{x}_1 = \hat{x}_0 + \frac{1}{H_1}(y_1 - H_1 \hat{x}_0)$

$$= \hat{x}_0 + \frac{y_1}{H_1} - \frac{H_1}{H_1}\hat{x}_0 = (\hat{x}_0 - \hat{x}_0) + \frac{y_1}{H_1}$$

Optimal RLLS gain K_1 says to just completely trust new measurement y_1 when $P_0 = `\infty'$…

$$\boxed{\Rightarrow \hat{x}_1 = \frac{y_1}{H_1}}$$

…Exactly what we intuitively expect should happen if no prior info at all about x!

Non-deterministic Estimation of <u>Dynamic</u> States

- Now let's finally relax the assumption that state x of linear system is static, so we have

$$x(k+1) = Fx(k) + Gu(k) + w(k), \qquad w(k) \sim \mathcal{N}(0, Q) \ (\text{AWGN})$$
$$y(k) = Hx(k) + v(k), \qquad v(k) \sim \mathcal{N}(0, R) \ (\text{AWGN})$$

→ *<u>Dynamic</u> x_k with <u>process noise</u> w_k and <u>known inputs</u> u_k*

- Now question is: how to find "optimal estimate" (best guess) of x(k) <u>at each time k</u>?
- Note: ever-increasing number of x(k)'s to estimate from corresponding set of y(k)'s !!!
- Question: why can't we just use the following LS cost at each time step k=1,…,T?

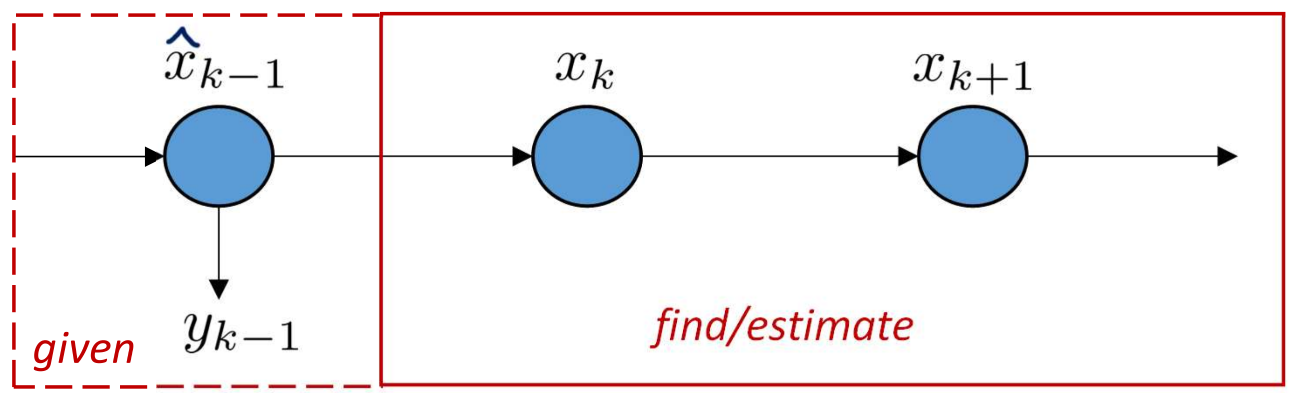
$$J(T) = \sum_{k=1}^{T} (y_k - Hx_k)^T R^{-1} (y_k - Hx_k)$$

<u>3 Big Reasons:</u>

1) State may not be completely sensed: i.e. H generally not square/invertible (recall: H alone generally not enough for observability – need to consider F also!)
2) States x_k & x_{k-1} *not independent of each other*: recall states encode system's <u>memory</u> ("knowledge of past constrains the future, and vice-versa" – x_k has <u>prior info</u> about x_{k+1})
3) Influence of process noise is NOT captured in either batch LLS or RLLS cost functions!

Non-deterministic Estimation of Dynamic States

- How to deal with fact that **states (and hence state estimates) have dynamical dependence on one another across time? (i.e. states and estimates MUST be constrained by system dynamics)**
- Note: existence of time dependencies via states implies **three possible types of (non-deterministic) dynamical state estimation problems**:

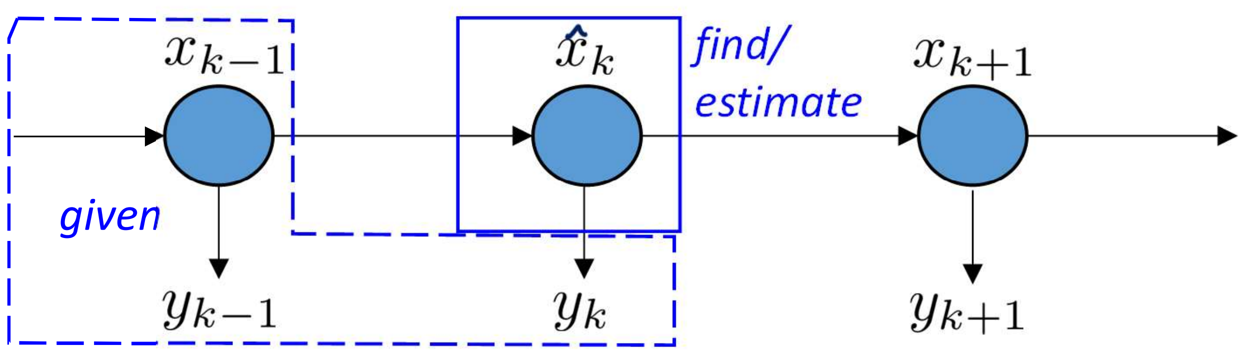


(I) 'Pure Prediction' problem:

Find $\hat{x}^-(k|j)$ = propagated/predicted state estimate

$\hat{x}(k)$ *given* info up to time $j \in [0, k-1]$

[est. error uncertainty / covariance: $P^-(k|j)$]

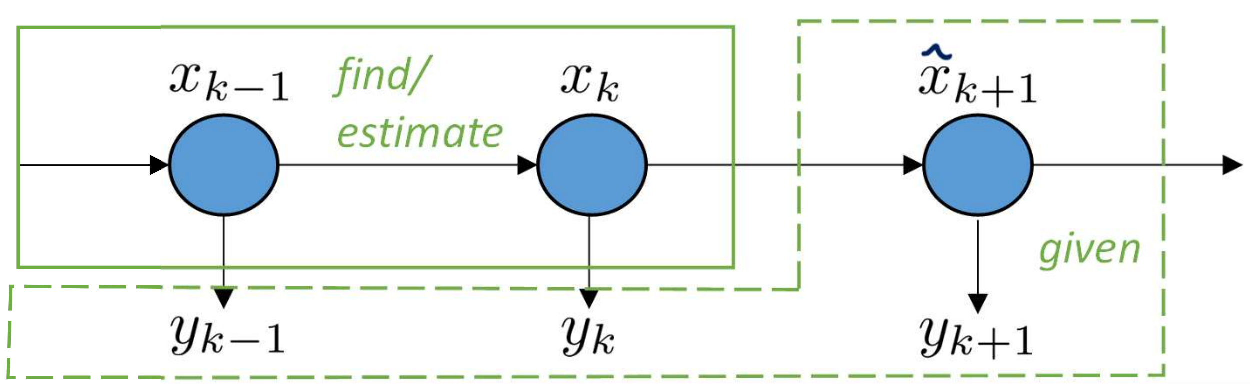


(II) 'Filtering' problem: ('clean up' noisy data $y_1, \ldots, y_k$)

Find $\hat{x}^+(k|k)$ = *filtered* state estimate ("current" state est.)

$\hat{x}(k)$ *given* info up to and including time k

[est. error uncertainty / covariance: $P^+(k|k)$]



(III) 'Smoothing' problem: (revise past given info from present)

Find $\hat{x}^S(k|m)$ = *smoothed* state estimate (retrodicted state est.)

$\hat{x}(k)$ *given* info up to time $m > k$

[est. error uncertainty / covariance: $P^S(k|m)$]

Dynamic Propagation/ "Pure Prediction" of Uncertain States

- How to **push uncertain states/estimates forward in time through dynamics models?**
- Look at DT linear models driven by **AWGN process noise only (i.e. NO MEASUREMENTS)**:

$$x(k+1) = Fx(k) + Gu(k) + w(k), \quad w(k) = \mathcal{N}(0, Q) \text{ (AWGN)}$$

- Suppose we are given Gaussian pdf for initial state statistics: $x(0) \sim N(m_0, P_0)$
- We know for fact that since the state dynamics for x(k) → x(k+1) are linear, [see lecs 16 & 15 & Q in HW5]

Then $x(1)$ must also have a Gaussian pdf! ie $x(1) \sim \mathcal{N}(m(1), P^-(1)) = p(x_1)$ [marginal pdf of x_1 given no other info besides $m(0)$ & $P(0)$ & Dynamics]

→ what should be the "best guess" (optimal estimate) $\hat{x}_k^-$ for x_k @ time $k=1$?

→ Can define cost fxn: $J(k) = tr(P_k^-) = tr(E[(x_k - \hat{x}_k^-)(x_k - \hat{x}_k^-)^T])$ ↔ minimize uncertainty in predicted state for each time k

→ Note! $J(k) = tr(P_k^-) = tr(E[(x_k - \hat{x}_k^-)(\ldots)^T]) = E[(x_k - \hat{x}_k^-)^T(x_k - \hat{x}_k^-)]$
$= E[x_k^T x_k - \hat{x}_k^{-T} x_k - x_k^T \hat{x}_k^- + \hat{x}_k^{-T}\hat{x}_k^-]$

→ So $\frac{\partial J(k)}{\partial \hat{x}_k^-} = \frac{\partial}{\partial \hat{x}_k^-}[E[x_k^T x_k - \hat{x}_k^{-T} x_k - x_k^T \hat{x}_k^- + \hat{x}_k^{-T}\hat{x}_k^-]] = 0$ → $\boxed{\hat{x}_k^- = E[x_k] = m_k = \text{mean of } p(x_k) \text{ @ time } k}$

ie ($\hat{x}_1^- = E[x_1]$ given m_0 & P_0 & Dynamics)

(*) optimal estimate for pure state prediction!

"Pure Prediction" of Uncertain States

- To find the **predicted state mean** at time k+1, given the mean at time k:

Start w/ $k+1 = 1$, given mean $m_0 = m(0) = \underbrace{\hat{x}_0^-}_{\text{(given)}}$ @ time $k = 0$:

$$m_1 = \hat{x}_1^- = E[\underline{x(1)}] = E[\,F x(0) + G u(0) + w(0)\,]$$

linearity of $E[\cdot]$

$$= F \cdot \underbrace{E[x(0)]}_{= m_0} + \underbrace{G u(0)}_{\text{not random}} + \cancel{E[w(0)]}$$

$= 0$, b/c AWGN for w_k

$$\boxed{\begin{aligned} m_1 = \hat{x}_1^- &= F \cdot m_0 + G u(0) \\ &= F \cdot \hat{x}_0^- + G u(0) \end{aligned}}$$

→ More generally! for $k = 1, 2, 3, \ldots$ the recursive formula to predict the mean of a linear-Gaussian Dynamical System into the future is:

$$(*) \quad \boxed{m(k+1) = \hat{x}^-(k+1) = F\, m(k) + G u(k)} \quad (*)$$

where $p(x_{k+1}) = N(m_{k+1}, \underline{\underline{P_{k+1}^-}})$ → uncertainty in predicted state @ time $k+1$

→ how to get this?

"Pure Prediction" of Uncertain States

- Likewise, we can find the **predicted state covariance** at time k+1:

given P_0^-, then for P_1^-, we have:

$$P_1^- = E\left[\{x(1) - m(1)\}\{x(1) - m(1)\}^T\right] = E\left[\{x(1) - \hat{x}_1^-\}\{\dots\}^T\right]$$

→ sub in $x(1)$ & $m(1)$ expressions!

$$P_1^- = E\left[\{\underbrace{Fx(0) + \cancel{Gu(0)} + w(0)}_{x(1)} - \underbrace{Fm(0) - \cancel{Gu(0)}}_{m(1)}\}\{\dots\}^T\right]$$

$$= E\left[\{Fx(0) - Fm(0) + w(0)\}\{\dots\}^T\right]$$

Group terms:

$$= E\left[\{F[x(0) - m(0)] + w(0)\}\{\dots\}^T\right]$$

FOIL:

$$= E\left[F(x(0) - m(0))(x(0) - m(0))^T F^T + F(x(0) - m(0))w^T(0) + w(0)(x(0) - m(0))^T F^T + w(0)w(0)^T\right]$$

“Pure Prediction” of Uncertain States

- So we get for propagated state covariance: using linearity of $E[\cdot]$:

$$P_1^- = F \cdot E[\underbrace{\{x(0)-m(0)\}\{\cdots\}^T}] \cdot F^T$$
$$+ F \cdot E[(x(0)-m(0))\,w^T(0)] + E[w(0)\,(x(0)-m(0))^T] \cdot F^T$$
$$+ E[w(0)\,w^T(0)]$$

→ Now use the facts that:

(i) $E[\{x(0)-m(0)\}\{x(0)-m(0)\}^T] = P_0^-$ (which is given!)

(ii) $E[(x(0)-m(0))\,w^T(0)] = 0$ (b/c $w(k) \sim$ AWGN → $x(k) \perp w(k)$ & $w(k)$ has 0 mean)

(iii) $E[w(0)\,w^T(0)] = Q$

$$\rightarrow \boxed{P_1^- = F P_0^- F^T + Q}$$

(note: $\perp$ of $w(k)$)

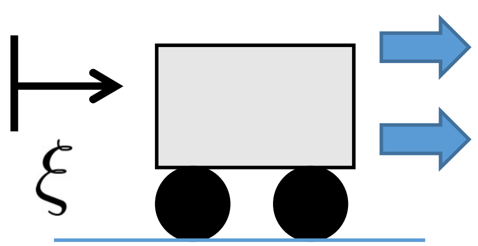
(*) More generally: Same reasoning gives a recursion for pure prediction:

$$(*) \quad \boxed{P_{k+1}^- = F P_k^- F^T + Q, \quad \text{for } k = 1, 2, 3, \ldots} \quad (*)$$

$$[\, P_0^- \rightarrow P_1^- \rightarrow P_2^- \rightarrow P_3^- \rightarrow \cdots \,]$$

Example: Pure Statistical State Prediction for 1D Robot Car

- **DT dynamics model:** $$x(k+1) = Fx(k) + Gu(k) + w(k)$$



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

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

$$\Delta t = 0.1 \text{ sec}$$

$$x(0) \sim \mathcal{N}(\mu_0, P_0), \text{ where } \mu_0 = [0,0]^T,\ P_0 = \begin{bmatrix} \sigma_p^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}$$

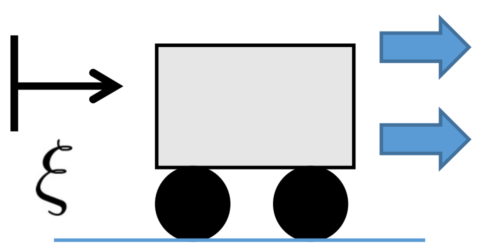
To calculate $m_k = \hat{x}_k^-$ and P_k^-
at $k = 1, 2, 3, \ldots$ and known $u(k)$:
initialize $m_0 = \hat{x}_0^- = \mu_0$ and $P_0^- = P_0$
for $k = 1, 2, 3, \ldots$
compute $m_k = F m_{k-1} + G u_{k-1}$
$P_k^- = F P_{k-1} F^T + Q$
end for

$\rightarrow$ If $P_0 = \begin{bmatrix} \sigma_p^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}$, then

$$P_1^- = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_p^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \Delta t & 1 \end{bmatrix} + Q$$
$$= \begin{bmatrix} \sigma_p^2 & \sigma_v^2 \Delta t \\ 0 & \sigma_v^2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \Delta t & 1 \end{bmatrix} + Q$$
$$= \underbrace{\begin{bmatrix} \sigma_p^2 + \sigma_v^2 \Delta t^2 & \sigma_v^2 \Delta t \\ \sigma_v^2 \Delta t & \sigma_v^2 \end{bmatrix}}_{\text{Uncertainty in state at time k due to (linear) dynamics only}} + Q \leftarrow \text{Uncertainty due to process noise only}$$

Example: Pure Statistical State Prediction for 1D Robot Car

- **DT dynamics model:** $x(k+1) = Fx(k) + Gu(k) + w(k)$

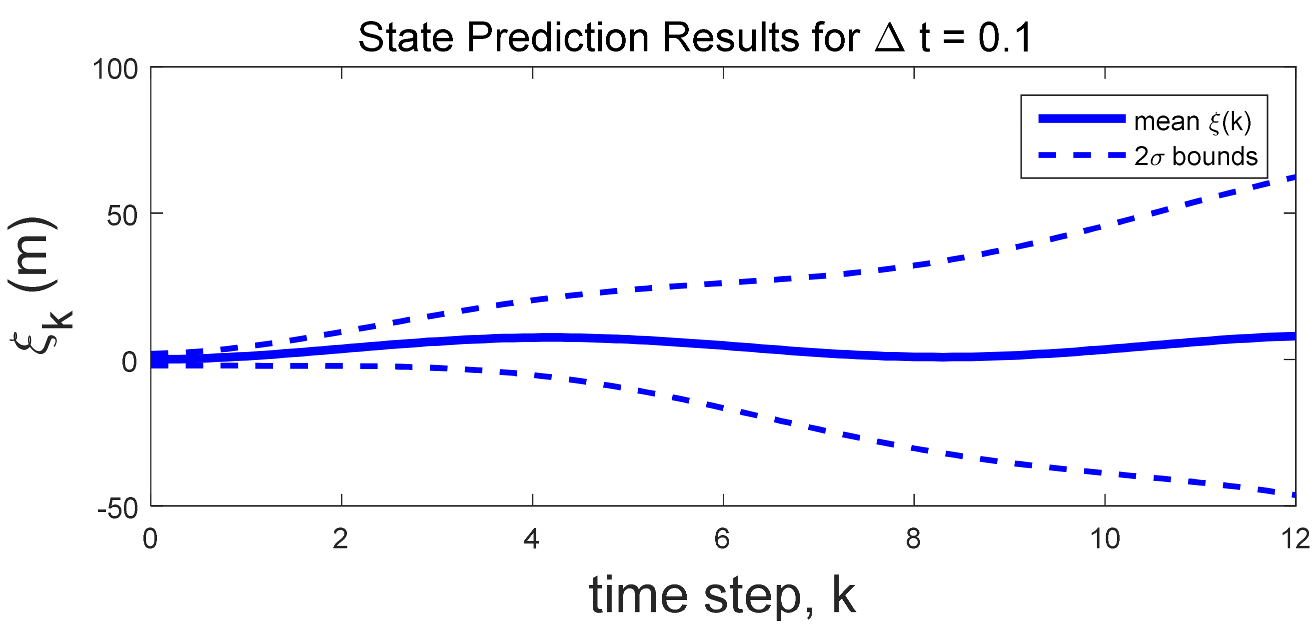


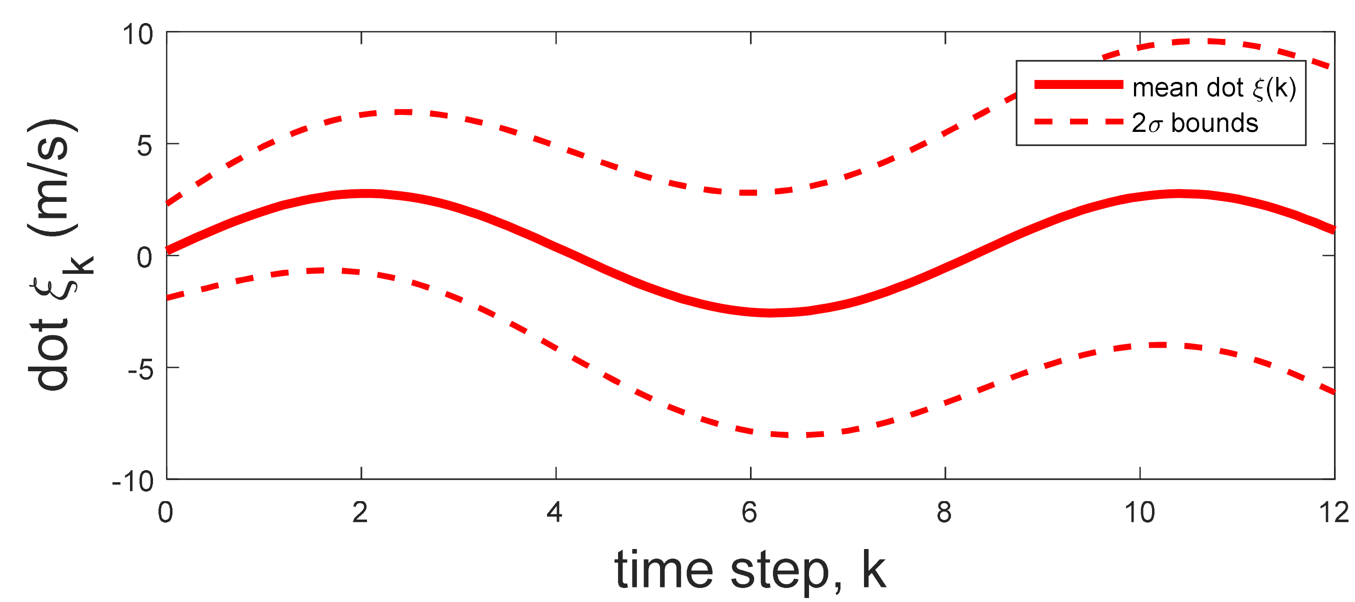
$$x(k) = [\xi(k), \dot{\xi(k)}]^T$$
$$u(k) = 2\cos(0.75 t_k) \text{ (ZOH)}$$
$$w(k) \sim \mathcal{N}(0, Q) \text{ (AWGN)}$$

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

$$\Delta t = 0.1 \text{ sec}$$

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

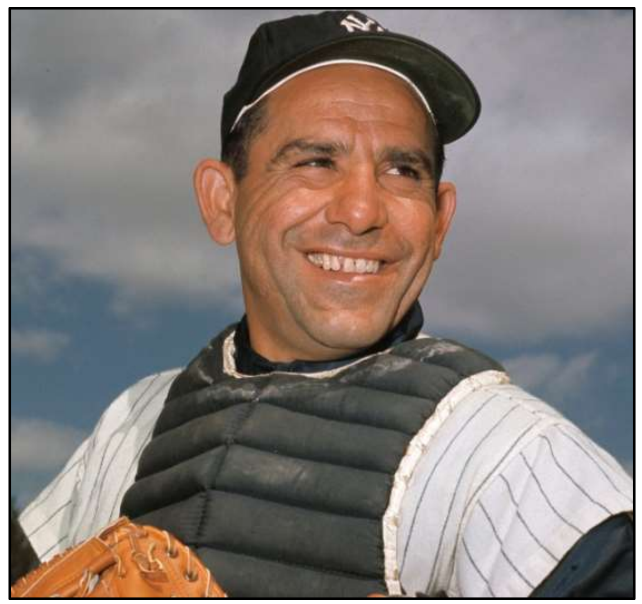


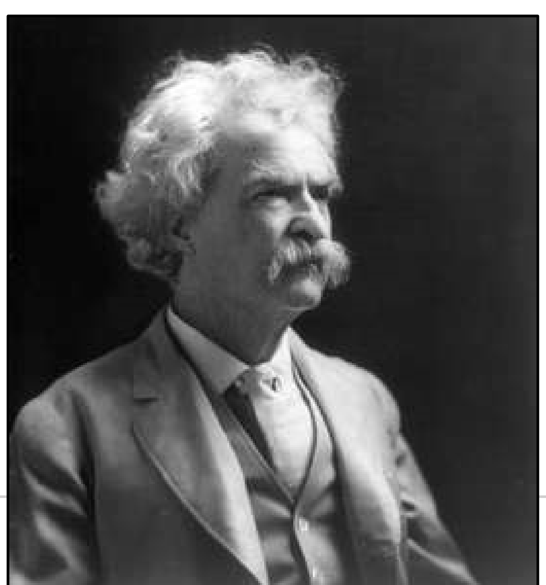


Always Remember:
Unless You Have a Linear-Gaussian System…

- **“It’s difficult to make predictions, especially about the future.”**

– attributed to Niels Bohr, Yogi Berra, Mark Twain, Nostradamus, and many others…





It’s Difficult to Make Predictions, Especially About the Future

Niels Bohr? Samuel Goldwyn? K. K. Steincke? Robert Storm Petersen? Yogi Berra? Mark Twain? Nostradamus? Anonymous?

predict01

Dear Quote Investigator: There is a family of popular humorous sayings about the formidable task of successful prognostication. Here are five examples:

1. It is difficult to make predictions, especially about the future.
2. Predictions are hazardous, especially about the future.
3. It is hard to prophecy, particularly about the future.
4. It’s dangerous to prophesy, particularly about the future.
5. Never make forecasts, especially about the future.