Lecture 8
The Kalman filter

- Linear system driven by stochastic process
- Statistical steady-state
- Linear Gauss-Markov model
- Kalman filter
- Steady-state Kalman filter

Linear system driven by stochastic process

we consider linear dynamical system $x_{t+1} = Ax_t + Bu_t$, with x_0 and $u_0,\ u_1, \ldots$ random variables

we'll use notation

$$\bar{x}_t = \mathbf{E}\, x_t, \qquad \Sigma_x(t) = \mathbf{E}(x_t - \bar{x}_t)(x_t - \bar{x}_t)^T$$

and similarly for $\bar{u}_t$, $\Sigma_u(t)$

taking expectation of $x_{t+1} = Ax_t + Bu_t$ we have

$$\bar{x}_{t+1} = A\bar{x}_t + B\bar{u}_t$$

i.e., the means propagate by the same linear dynamical system now let's consider the covariance

$$x_{t+1} - \bar{x}_{t+1} = A(x_t - \bar{x}_t) + B(u_t - \bar{u}_t)$$

and so

$$\begin{aligned} \Sigma_x(t+1) &= \mathbf{E}\left(A(x_t - \bar{x}_t) + B(u_t - \bar{u}_t)\right)\left(A(x_t - \bar{x}_t) + B(u_t - \bar{u}_t)\right)^T \\ &= A\Sigma_x(t)A^T + B\Sigma_u(t)B^T + A\Sigma_{xu}(t)B^T + B\Sigma_{ux}(t)A^T \end{aligned}$$

where

$$\Sigma_{xu}(t) = \Sigma_{ux}(t)^T = \mathbf{E}(x_t - \bar{x}_t)(u_t - \bar{u}_t)^T$$

thus, the covariance $\Sigma_x(t)$ satisfies another, Lyapunov-like linear dynamical system, driven by Σ_{xu} and Σ_u

consider special case $\Sigma_{xu}(t) = 0$, $i.e.$, x and u are uncorrelated, so we have Lyapunov iteration

$$\Sigma_x(t+1) = A\Sigma_x(t)A^T + B\Sigma_u(t)B^T,$$

which is stable if and only if A is stable

if A is stable and $\Sigma_u(t)$ is constant, $\Sigma_x(t)$ converges to Σ_x, called the *steady-state covariance*, which satisfies Lyapunov equation

$$\Sigma_x = A\Sigma_x A^T + B\Sigma_u B^T$$

thus, we can calculate the steady-state covariance of x exactly, by solving a Lyapunov equation

(useful for starting simulations in statistical steady-state)

Example

we consider $x_{t+1} = Ax_t + w_t$, with

$$A = \begin{bmatrix} 0.6 & -0.8 \\ 0.7 & 0.6 \end{bmatrix},$$

where w_t are IID $\mathcal{N}(0, I)$

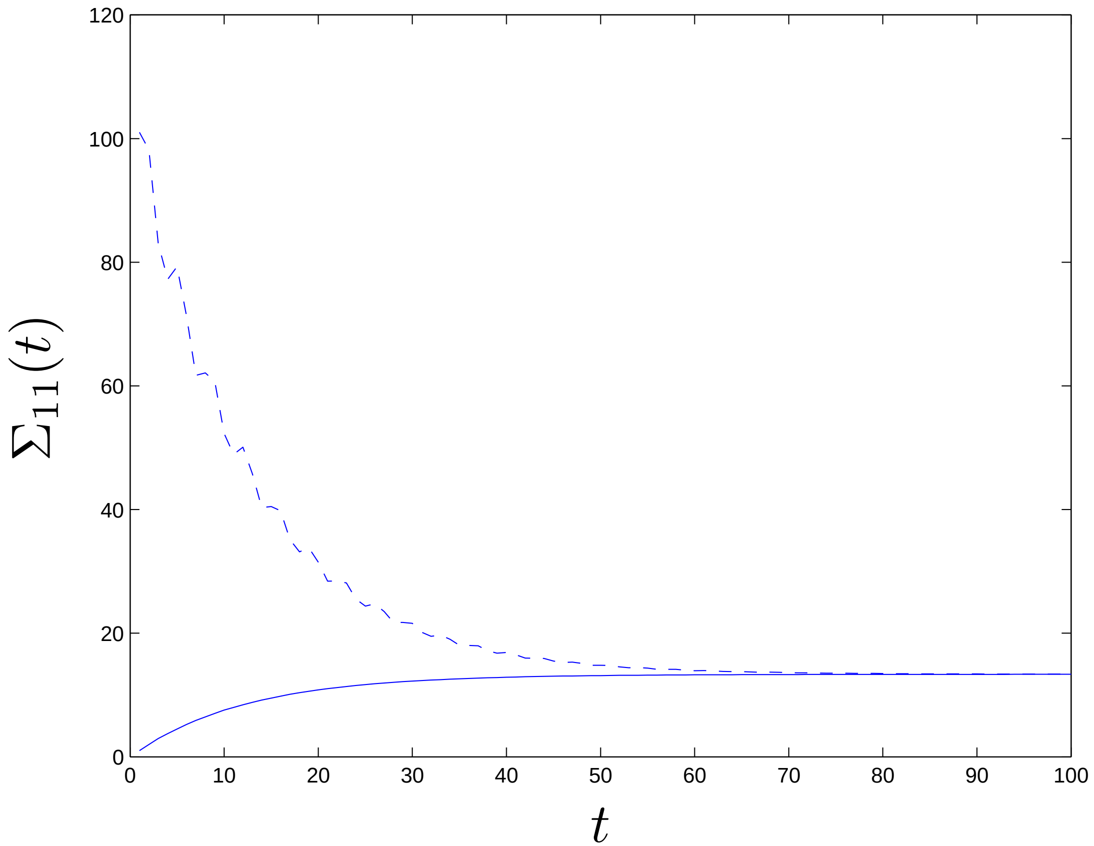
eigenvalues of A are $0.6 \pm 0.75j$, with magnitude 0.96, so A is stable

we solve Lyapunov equation to find steady-state covariance

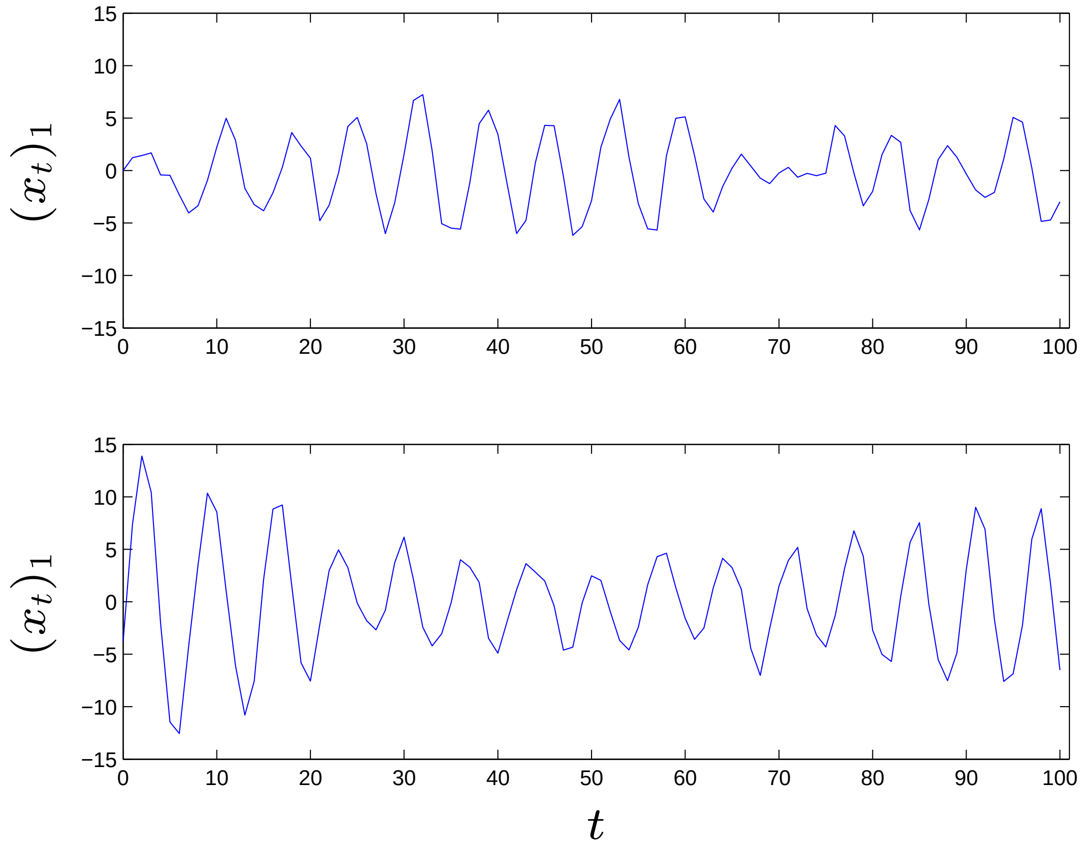
$$\Sigma_x = \begin{bmatrix} 13.35 & -0.03 \\ -0.03 & 11.75 \end{bmatrix}$$

covariance of x_t converges to Σ_x no matter its initial value two initial state distributions: $\Sigma_x(0) = 0$, $\Sigma_x(0) = 10^2 I$

plot shows $\Sigma_{11}(t)$ for the two cases



$(x_t)_1$ for one realization from each case:

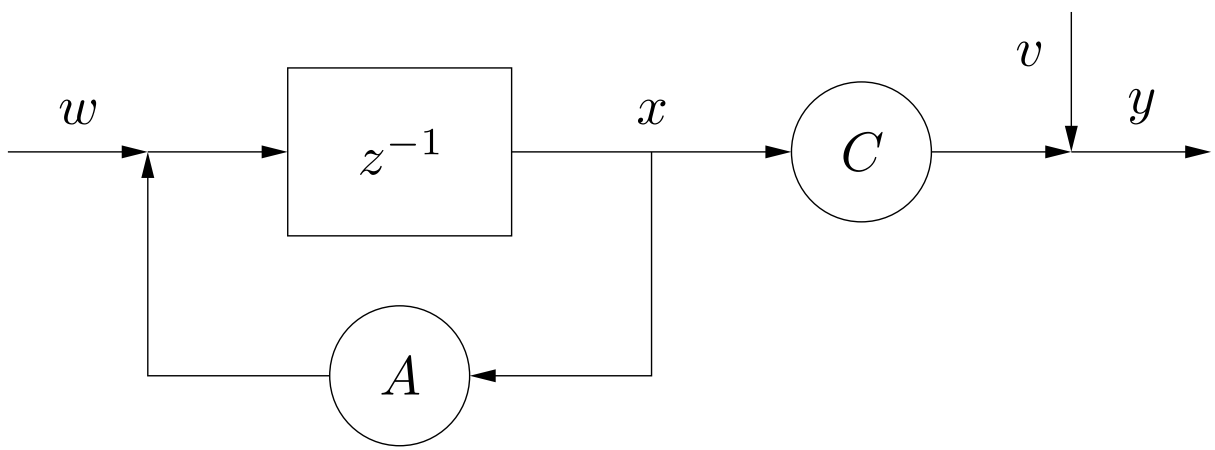


Linear Gauss-Markov model

we consider linear dynamical system

$$x_{t+1} = Ax_t + w_t, \qquad y_t = Cx_t + v_t$$

- $x_t \in \mathbf{R}^n$ is the state; $y_t \in \mathbf{R}^p$ is the observed output
- $w_t \in \mathbf{R}^n$ is called *process noise* or *state noise*
- $v_t \in \mathbf{R}^p$ is called *measurement noise*



Statistical assumptions

- x_0, $w_0, w_1, \ldots$, $v_0, v_1, \ldots$ are jointly Gaussian and independent
- w_t are IID with $\mathbf{E}\, w_t = 0$, $\mathbf{E}\, w_t w_t^T = W$
- v_t are IID with $\mathbf{E}\, v_t = 0$, $\mathbf{E}\, v_t v_t^T = V$
- $\mathbf{E}\, x_0 = \bar{x}_0$, $\mathbf{E}(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T = \Sigma_0$

(it's not hard to extend to case where w_t, v_t are not zero mean)

we'll denote $X_t = (x_0, \ldots, x_t)$, etc.

since X_t and Y_t are linear functions of x_0, W_t, and V_t, we conclude they are all jointly Gaussian (*i.e.*, the process $x,\ w,\ v,\ y$ is Gaussian)

Statistical properties

- sensor noise v independent of x
- w_t is independent of $x_0, \ldots, x_t$ and $y_0, \ldots, y_t$
- *Markov property:* the process x is Markov, $i.e.$,

$$x_t | x_0, \ldots, x_{t-1} = x_t | x_{t-1}$$

  roughly speaking: if you know x_{t-1}, then knowledge of $x_{t-2}, \ldots, x_0$ doesn't give any more information about x_t

Mean and covariance of Gauss-Markov process

mean satisfies $\bar{x}_{t+1} = A\bar{x}_t$, $\mathbf{E}\, x_0 = \bar{x}_0$, so $\bar{x}_t = A^t \bar{x}_0$

covariance satisfies

$$\Sigma_x(t+1) = A\Sigma_x(t)A^T + W$$

if A is stable, $\Sigma_x(t)$ converges to steady-state covariance Σ_x, which satisfies Lyapunov equation

$$\Sigma_x = A\Sigma_x A^T + W$$

Conditioning on observed output

we use the notation

$$\hat{x}_{t|s} = \mathbf{E}(x_t|y_0, \ldots y_s),$$
$$\Sigma_{t|s} = \mathbf{E}(x_t - \hat{x}_{t|s})(x_t - \hat{x}_{t|s})^T$$

- the random variable $x_t|y_0, \ldots, y_s$ is Gaussian, with mean $\hat{x}_{t|s}$ and covariance $\Sigma_{t|s}$
- $\hat{x}_{t|s}$ is the minimum mean-square error estimate of x_t, based on $y_0, \ldots, y_s$
- $\Sigma_{t|s}$ is the covariance of the error of the estimate $\hat{x}_{t|s}$

State estimation

we focus on two state estimation problems:

- finding $\hat{x}_{t|t}$, *i.e.*, estimating the current state, based on the current and past observed outputs
- finding $\hat{x}_{t+1|t}$, *i.e.*, predicting the next state, based on the current and past observed outputs

since x_t, Y_t are jointly Gaussian, we can use the standard formula to find $\hat{x}_{t|t}$ (and similarly for $\hat{x}_{t+1|t}$)

$$\hat{x}_{t|t} = \bar{x}_t + \Sigma_{x_t Y_t} \Sigma_{Y_t}^{-1} (Y_t - \bar{Y}_t)$$

the inverse in the formula, $\Sigma_{Y_t}^{-1}$, is size $pt \times pt$, which grows with t

the *Kalman filter* is a clever method for computing $\hat{x}_{t|t}$ and $\hat{x}_{t+1|t}$ recursively

Measurement update

let's find $\hat{x}_{t|t}$ and $\Sigma_{t|t}$ in terms of $\hat{x}_{t|t-1}$ and $\Sigma_{t|t-1}$

start with $y_t = Cx_t + v_t$, and condition on Y_{t-1}:

$$y_t|Y_{t-1} = Cx_t|Y_{t-1} + v_t|Y_{t-1} = Cx_t|Y_{t-1} + v_t$$

since v_t and Y_{t-1} are independent

so $x_t|Y_{t-1}$ and $y_t|Y_{t-1}$ are jointly Gaussian with mean and covariance

$$\begin{bmatrix} \hat{x}_{t|t-1} \\ C\hat{x}_{t|t-1} \end{bmatrix}, \qquad \begin{bmatrix} \Sigma_{t|t-1} & \Sigma_{t|t-1}C^T \\ C\Sigma_{t|t-1} & C\Sigma_{t|t-1}C^T + V \end{bmatrix}$$

now use standard formula to get mean and covariance of

$$(x_t|Y_{t-1})|(y_t|Y_{t-1}),$$

which is exactly the same as $x_t|Y_t$:

$$\begin{aligned}
\hat{x}_{t|t} &= \hat{x}_{t|t-1} + \Sigma_{t|t-1}C^T\left(C\Sigma_{t|t-1}C^T + V\right)^{-1}(y_t - C\hat{x}_{t|t-1}) \\
\Sigma_{t|t} &= \Sigma_{t|t-1} - \Sigma_{t|t-1}C^T\left(C\Sigma_{t|t-1}C^T + V\right)^{-1}C\Sigma_{t|t-1}
\end{aligned}$$

this gives us $\hat{x}_{t|t}$ and $\Sigma_{t|t}$ in terms of $\hat{x}_{t|t-1}$ and $\Sigma_{t|t-1}$

this is called the *measurement update* since it gives our updated estimate of x_t based on the measurement y_t becoming available

Time update

now let's increment time, using $x_{t+1} = Ax_t + w_t$

condition on Y_t to get

$$\begin{aligned} x_{t+1}|Y_t &= Ax_t|Y_t + w_t|Y_t \\ &= Ax_t|Y_t + w_t \end{aligned}$$

since w_t is independent of Y_t

therefore we have $\hat{x}_{t+1|t} = A\hat{x}_{t|t}$ and

$$\begin{aligned} \Sigma_{t+1|t} &= \mathbf{E}(\hat{x}_{t+1|t} - x_{t+1})(\hat{x}_{t+1|t} - x_{t+1})^T \\ &= \mathbf{E}(A\hat{x}_{t|t} - Ax_t - w_t)(A\hat{x}_{t|t} - Ax_t - w_t)^T \\ &= A\Sigma_{t|t}A^T + W \end{aligned}$$

Kalman filter

measurement and time updates together give a recursive solution

start with prior mean and covariance, $\hat{x}_{0|-1} = \bar{x}_0$, $\Sigma_{0|-1} = \Sigma_0$

apply the measurement update

$$\begin{aligned}\hat{x}_{t|t} &= \hat{x}_{t|t-1} + \Sigma_{t|t-1}C^T\left(C\Sigma_{t|t-1}C^T + V\right)^{-1}(y_t - C\hat{x}_{t|t-1}) \\ \Sigma_{t|t} &= \Sigma_{t|t-1} - \Sigma_{t|t-1}C^T\left(C\Sigma_{t|t-1}C^T + V\right)^{-1}C\Sigma_{t|t-1}\end{aligned}$$

to get $\hat{x}_{0|0}$ and $\Sigma_{0|0}$; then apply time update

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t}, \qquad \Sigma_{t+1|t} = A\Sigma_{t|t}A^T + W$$

to get $\hat{x}_{1|0}$ and $\Sigma_{1|0}$

now, repeat measurement and time updates . . .

Riccati recursion

we can express measurement and time updates for Σ as

$$\Sigma_{t+1|t} = A\Sigma_{t|t-1}A^T + W - A\Sigma_{t|t-1}C^T(C\Sigma_{t|t-1}C^T + V)^{-1}C\Sigma_{t|t-1}A^T$$

which is a Riccati recursion, with initial condition $\Sigma_{0|-1} = \Sigma_0$

- $\Sigma_{t|t-1}$ can be computed *before any observations are made*
- thus, we can calculate the estimation error covariance *before* we get any observed data

Comparison with LQR

in LQR,

- Riccati recursion for P_t (which determines the minimum cost to go from a point at time t) runs *backward* in time
- we can compute cost-to-go before knowing x_t

in Kalman filter,

- Riccati recursion for $\Sigma_{t|t-1}$ (which is the state prediction error covariance at time t) runs *forward* in time
- we can compute $\Sigma_{t|t-1}$ before we actually get any observations

Observer form

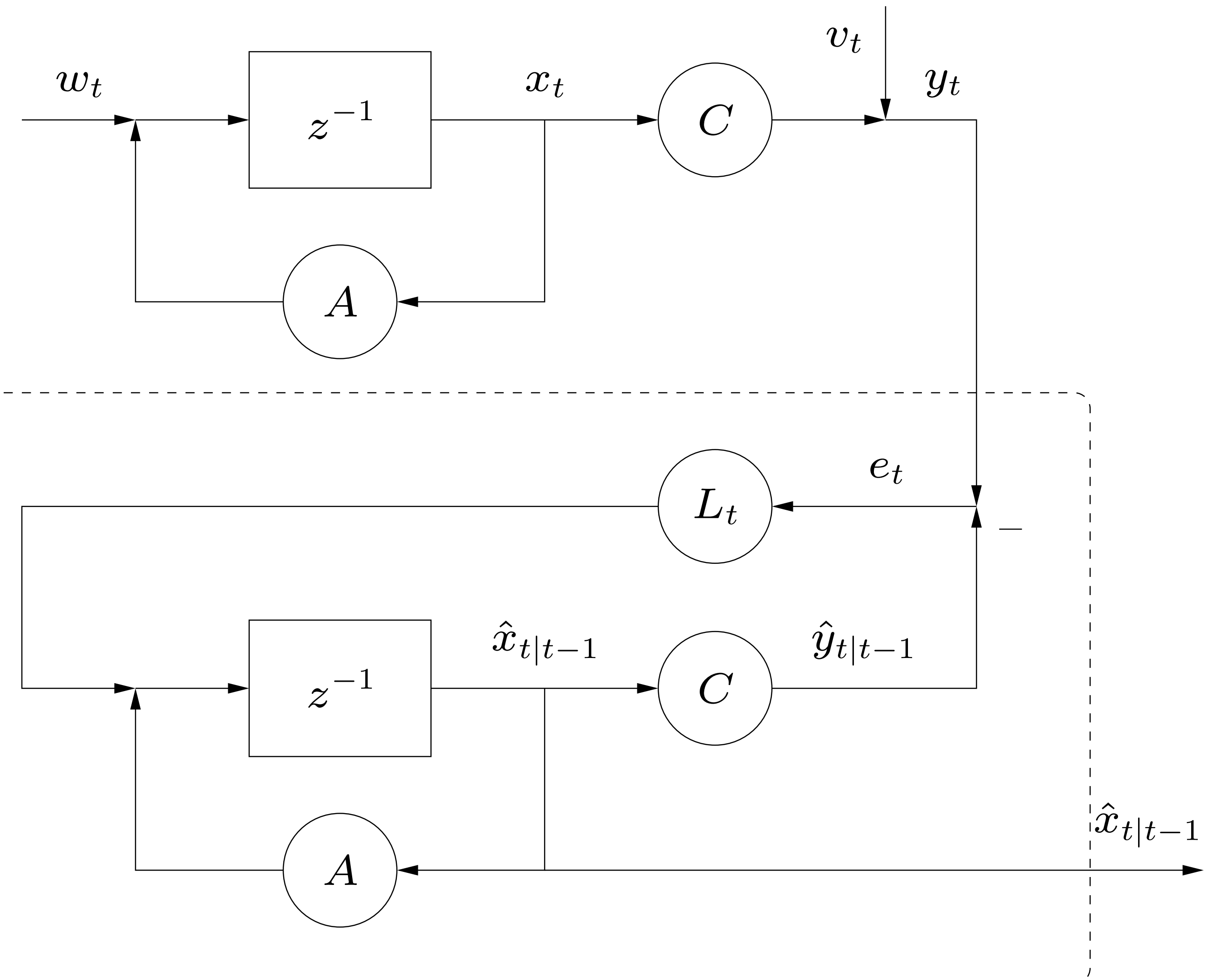
we can express KF as

$$\begin{aligned}\hat{x}_{t+1|t} &= A\hat{x}_{t|t-1} + A\Sigma_{t|t-1}C^T(C\Sigma_{t|t-1}C^T + V)^{-1}(y_t - C\hat{x}_{t|t-1}) \\ &= A\hat{x}_{t|t-1} + L_t(y_t - \hat{y}_{t|t-1})\end{aligned}$$

where $L_t = A\Sigma_{t|t-1}C^T(C\Sigma_{t|t-1}C^T + V)^{-1}$ is the *observer gain*

- $\hat{y}_{t|t-1}$ is our output prediction, $i.e.$, our estimate of y_t based on $y_0, \ldots, y_{t-1}$
- $e_t = y_t - \hat{y}_{t|t-1}$ is our output prediction error
- $A\hat{x}_{t|t-1}$ is our prediction of x_{t+1} based on $y_0, \ldots, y_{t-1}$
- our estimate of x_{t+1} is the prediction based on $y_0, \ldots, y_{t-1}$, plus a linear function of the output prediction error

Kalman filter block diagram



Steady-state Kalman filter

as in LQR, Riccati recursion for $\Sigma_{t|t-1}$ converges to steady-state value $\hat{\Sigma}$, provided (C, A) is observable and (A, W) is controllable

$\hat{\Sigma}$ gives steady-state error covariance for estimating x_{t+1} given $y_0, \ldots, y_t$

note that state prediction error covariance converges, even if system is unstable

$\hat{\Sigma}$ satisfies ARE

$$\hat{\Sigma} = A\hat{\Sigma}A^T + W - A\hat{\Sigma}C^T(C\hat{\Sigma}C^T + V)^{-1}C\hat{\Sigma}A^T$$

(which can be solved directly)

steady-state filter is a time-invariant observer:

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t-1} + L(y_t - \hat{y}_{t|t-1}), \qquad \hat{y}_{t|t-1} = C\hat{x}_{t|t-1}$$

where $L = A\hat{\Sigma}C^T(C\hat{\Sigma}C^T + V)^{-1}$

define state estimation error $\tilde{x}_{t|t-1} = x_t - \hat{x}_{t|t-1}$, so

$$y_t - \hat{y}_{t|t-1} = Cx_t + v_t - C\hat{x}_{t|t-1} = C\tilde{x}_{t|t-1} + v_t$$

and

$$\begin{aligned} \tilde{x}_{t+1|t} &= x_{t+1} - \hat{x}_{t+1|t} \\ &= Ax_t + w_t - A\hat{x}_{t|t-1} - L(C\tilde{x}_{t|t-1} + v_t) \\ &= (A - LC)\tilde{x}_{t|t-1} + w_t - Lv_t \end{aligned}$$

thus, the estimation error propagates according to a linear system, with closed-loop dynamics $A - LC$, driven by the process $w_t - LCv_t$, which is IID zero mean and covariance $W + LVL^T$

provided A, W is controllable and C, A is observable, $A - LC$ is stable

Example

system is

$$x_{t+1} = Ax_t + w_t, \qquad y_t = Cx_t + v_t$$

with $x_t \in \mathbf{R}^6$, $y_t \in \mathbf{R}$

we'll take $\mathbf{E}\, x_0 = 0$, $\mathbf{E}\, x_0 x_0^T = \Sigma_0 = 5^2 I$; $W = (1.5)^2 I$, $V = 1$

eigenvalues of A:

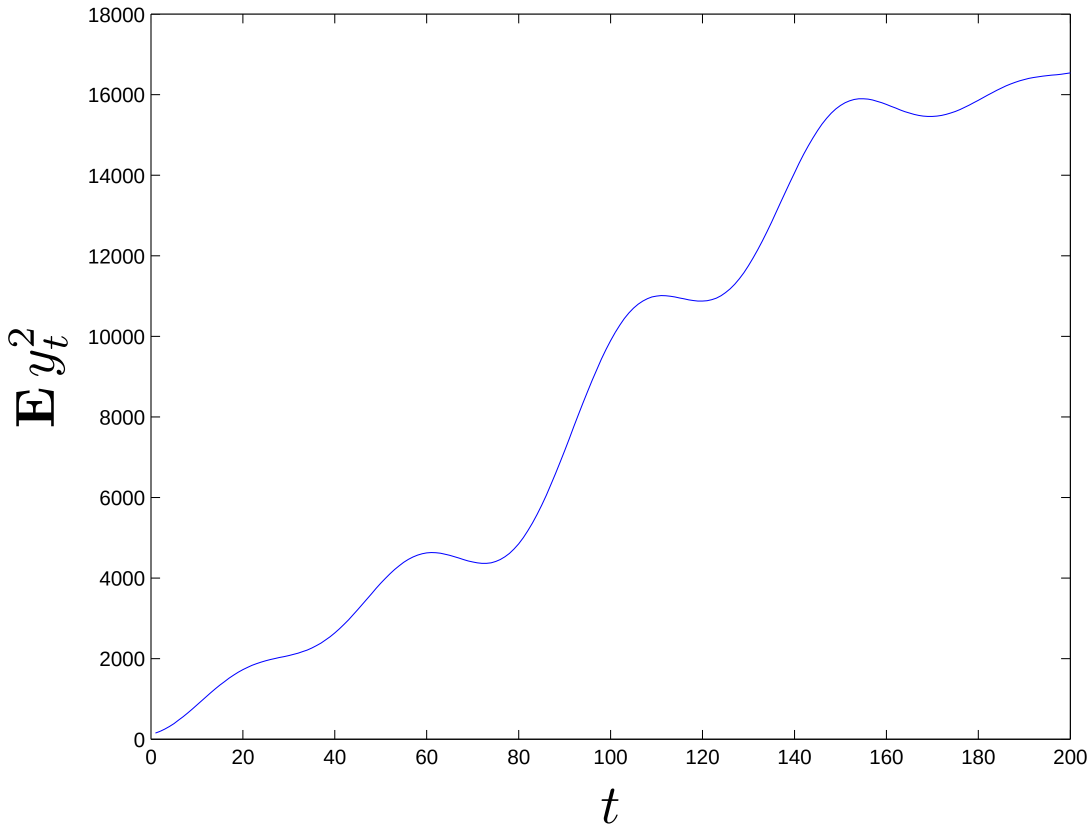
$$0.9973 \pm 0.0730j, \qquad 0.9995 \pm 0.0324j, \qquad 0.9941 \pm 0.1081j$$

(which have magnitude one)

goal: predict y_{t+1} based on $y_0, \ldots, y_t$

first let's find variance of y_t versus t, using Lyapunov recursion

$$\mathbf{E}\, y_t^2 = C\Sigma_x(t)C^T + V, \qquad \Sigma_x(t+1) = A\Sigma_x(t)A^T + W, \qquad \Sigma_x(0) = \Sigma_0$$

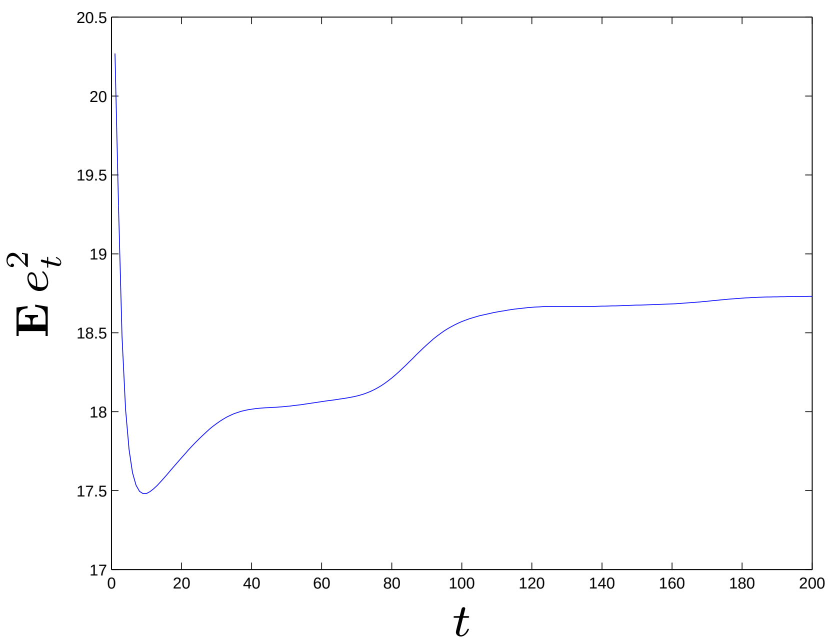


now, let's plot the prediction error variance versus t,

$$\mathbf{E}\, e_t^2 = \mathbf{E}(\hat{y}_{t|t-1} - y_t)^2 = C\Sigma_{t|t-1}C^T + V,$$

where $\Sigma_{t|t-1}$ satisfies Riccati recursion, initialized by $\Sigma_{-1|-2} = \Sigma_0$,

$$\Sigma_{t+1|t} = A\Sigma_{t|t-1}A^T + W - A\Sigma_{t|t-1}C^T(C\Sigma_{t|t-1}C^T + V)^{-1}C\Sigma_{t|t-1}A^T$$



prediction error variance converges to steady-state value 18.7

now let's try the Kalman filter on a realization y_t

top plot shows y_t; bottom plot shows e_t (on different vertical scale)