GEL-4250 / GEL-7015
Commande multivariable

Extraits de GEL-7029 Observation et commande prédictive
Version 2.1
31 janvier 2020

André Desbiens, Ing. Ph.D.
Département de génie électrique et de génie informatique

Université Laval
Québec, Canada

UNIVERSITÉ LAVAL

Part 3

REVIEW OF MATRIX ALGEBRA AND STOCHASTIC SIGNALS

Matrix algebra

Homogeneous system of linear equations

A $n \times n$ homogeneous system of linear equations:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = 0$$

$$Ax = 0$$

- has a unique solution (the trivial solution $x = 0$) if and only if $\det[A] \neq 0$
- has a solution set with an infinite number of solutions if $\det[A] = 0$

Matrix algebra

Eigenvalues

If $Av = \lambda v$ where A: $n\times n$, v: $n\times 1$ and λ: 1×1 then

- v is an eigenvector of A (there are n eigenvectors)
- λ is the corresponding eigenvalue (there are n eigenvalues)
- Eigenvalues with Matlab: eig(A)

$$Av = \lambda v$$

$$\lambda v - Av = 0$$

$$[\lambda I - A]v = 0$$

- If we dismiss the trivial solution, we need $\det[\lambda I - A] = 0$. This is how the eigenvalues λ are calculated.

Matrix algebra

Quadratic form

Quadratic form: scalar $Q(x) = x^T A x$ where x: $n \times 1$ and A symmetric

- Positive (semi-) definite if
  - $Q(x)\ (\geq) > 0 \quad \forall x \neq 0$
  - All eigenvalues of A $(\geq) > 0$
- Negative (semi-) definite if
  - $Q(x)\ (\leq) < 0 \quad \forall x \neq 0$
  - All eigenvalues of A $(\leq) < 0$
- Indefinite if
  - $Q(x) > 0$ for some x and $Q(x) < 0$ for some other x
  - Some eigenvalues of $A > 0$ and some < 0
- $A = B^T B$ is positive semi-definite

Matrix algebra

Matrix differentiation

$$x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$$

If $f(x)$ is a scalar function of x then: $\dfrac{\partial f(x)}{\partial x} = \begin{bmatrix} \dfrac{\partial f(x)}{\partial x_1} & \dfrac{\partial f(x)}{\partial x_2} & \cdots & \dfrac{\partial f(x)}{\partial x_n} \end{bmatrix}^T$

If $f(x) = Bx$ where B is $1 \times n$ then: $\dfrac{\partial f(x)}{\partial x} = B^T$

If $f(x) = x^T A x$ where A is $n \times n$ then: $\dfrac{\partial f(x)}{\partial x} = Ax + A^T x$

If A is symmetric then: $\dfrac{\partial f(x)}{\partial x} = 2Ax$

If $f(x) = x^T A x - 2Bx + D$ where A is symmetric and D is a scalar then:

$$\frac{\partial f(x)}{\partial x} = 2Ax - 2B^T$$

Matrix algebra

Minimization of a quadratic function

A: positive definite and symmetric

$$\frac{\partial\left[x^T A x - 2Bx + D\right]}{\partial x} = 2Ax - 2B^T = 0$$

$$x = A^{-1}B^T$$

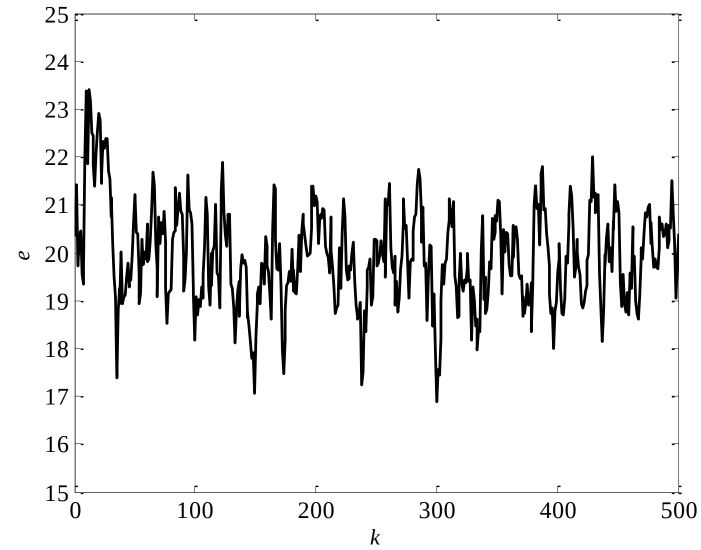
Minimum or maximum?

$$\frac{\partial^2\left[x^T A x - 2Bx + D\right]}{\partial x^2} = \frac{\partial\left[2Ax - 2B^T\right]}{\partial x} = 2A^T = 2A$$

Positive definite: minimum

Stochastic signals

Mathematical expectation



Time series: recording of N values of the variable e: $\{e(k)\}$ or $e(k)$ for $k = 1$ to N

If e is a signal whose value has some element of chance associated with it, then e is a stochastic (random) signal. Therefore it cannot be predicted exactly.

The probability density function (pdf) specifies the probability of the random variable falling within a particular range of values. Example: the well-known normal (or Gaussian) distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

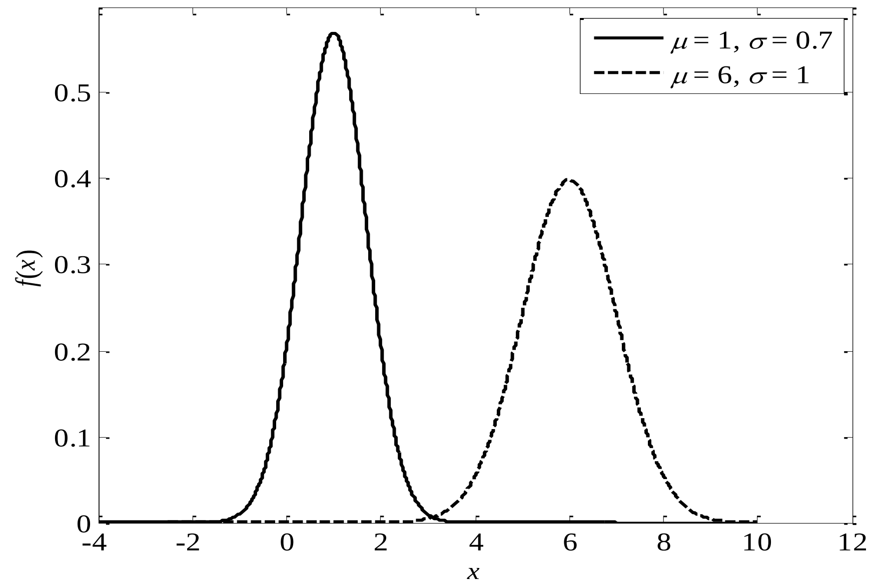
μ : mean

σ : standard deviation

Stochastic signals

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

γ	$P(\mu - \gamma\sigma \le e(k) \le \mu + \gamma\sigma)$
1	0.6826
2	0.9544
3	0.9974



If the pdf does not vary with time then the signal is stationary (for a Gaussian signal: its mean and variance do not vary with time)

The mathematical expectation (or expected value) of e is:

$$E\{e(k)\} = \int_{-\infty}^{\infty} x f(x) dx$$

It corresponds to its “true” mean:

$$\mu_e = E\{e(k)\}$$

Stochastic signals

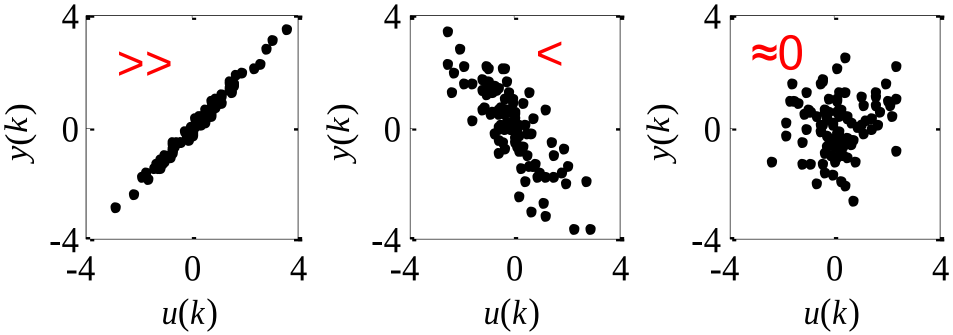
The “true” variance: $\sigma_e^2 = E\left\{\left(e(k)-\mu_e\right)^2\right\}$

Linear operation (a and b are constant):

$$E\{ae(k)+b\}=E\{ae(k)\}+E\{b\}=aE\{e(k)\}+b=a\mu_e+b$$

Covariance

Correlation between 2 random variables: how strongly are they related?



$y(k)$ vs $u(k)$, for $k = 1$ to N

The correlation is not necessarily instantaneous, i.e. $y(k)$ may be very lightly related to $u(k)$ but may be strongly related to $u(k - 2)$. We therefore look for different delays between the two variables.

Stochastic signals

Measure of how $u(k - i)$ and $y(k)$ are related: cross-covariance

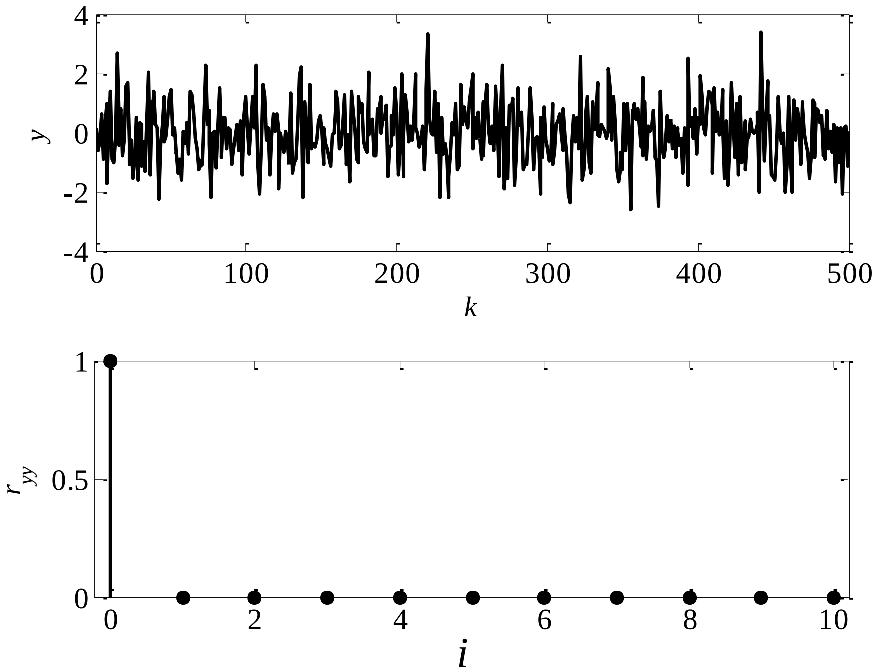
$$C_{yu}(i) = E\left\{\left(y(k) - \mu_y\right)\left(u(k-i) - \mu_u\right)\right\}$$

Can be scaled (normalized), between -1 and 1: cross-correlation

$$r_{yu}(i) = \frac{C_{yu}(i)}{\sqrt{C_{yy}(0)C_{uu}(0)}} = \frac{C_{yu}(i)}{\sigma_y \sigma_u}$$

- If $u(k - 3) = -y(k)$: $r_{yu}(3) = -1$
- If $u(k - 3)$ and $y(k)$ are independent: $C_{yu}(3) = 0$ (but $C_{yu}(3) = 0$ does not necessarily mean that $u(k - 3)$ and $y(k)$ are independent)
- If $u(k) = y(k)$: auto-covariance and auto-correlation
- $C_{yy}(0)$ = variance of $y(k)$
- $r_{yy}(-i) = r_{yy}(i)$ and $r_{yy}(0) = 1$

Stochastic signals

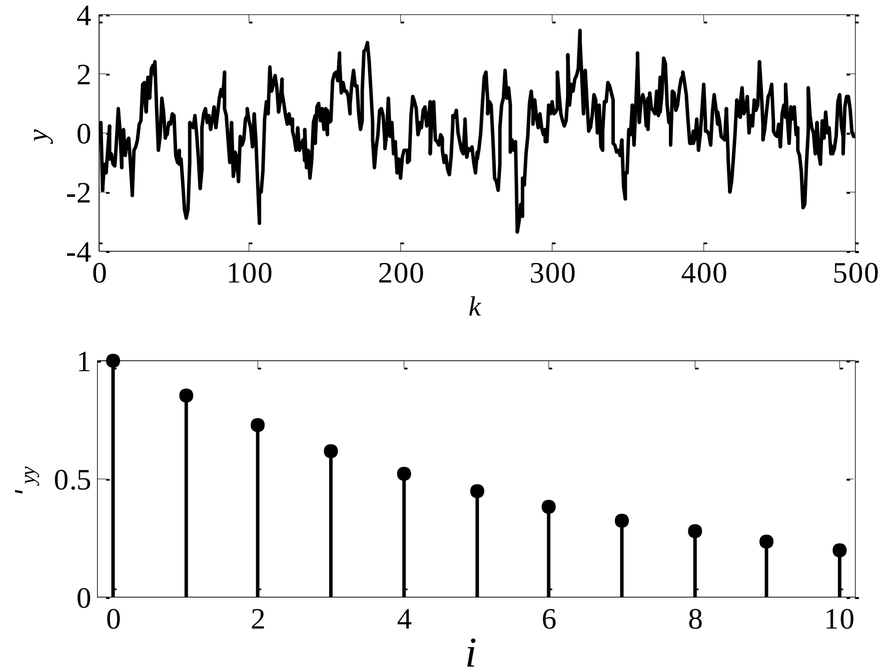


White noise

$$E\{y(k)\} = 0$$

$$E\left\{\left(y(k)-\mu_y\right)\left(y(i)-\mu_y\right)\right\} = E\{y(k)\,y(i)\}$$

$$= \begin{cases} = 0 & \text{if } k \neq i \\ = \sigma_y^2 & \text{if } k = i \end{cases}$$



Colored noise

$$E\{y(k)\} = 0$$

$$E\{y(k)\,y(i)\} \begin{cases} \neq 0 & \text{if } k \neq i \\ = \sigma_y^2 & \text{if } k = i \end{cases}$$

Stochastic signals

If n stochastic signals: $e_1(k)$, $e_2(k)$, …, $e_n(k)$ then the covariance matrix is:

$$Q = \begin{bmatrix} E\{(e_1(k)-\mu_1)(e_1(k)-\mu_1)\} & E\{(e_1(k)-\mu_1)(e_2(k)-\mu_2)\} & \cdots & E\{(e_1(k)-\mu_1)(e_n(k)-\mu_n)\} \\ E\{(e_2(k)-\mu_2)(e_1(k)-\mu_1)\} & E\{(e_2(k)-\mu_2)(e_2(k)-\mu_2)\} & \cdots & E\{(e_2(k)-\mu_2)(e_n(k)-\mu_n)\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{(e_n(k)-\mu_n)(e_1(k)-\mu_1)\} & E\{(e_n(k)-\mu_n)(e_2(k)-\mu_2)\} & \cdots & E\{(e_n(k)-\mu_n)(e_n(k)-\mu_n)\} \end{bmatrix}$$

- symmetric
- variance of $e_i(k)$ on the diagonal
- if $e_i(k)$ and $e_j(k)$ are independent for all $i \neq j$: diagonal matrix
- positive semi-definite

Part 11

OBSERVERS

Observers – What for?

Prediction equations for a state-space model

Model: $y(k) = Cx(k)$

$$x(k+1) = Ax(k) + Bu(k)$$

Predictions:

$$\hat{y}(k+1/k) = Cx(k+1)$$
$$= CAx(k) + CBu(k)$$
$$\hat{y}(k+2/k) = CAx(k+1) + CBu(k+1)$$
$$= CA^2x(k) + CABu(k) + CBu(k+1)$$
$$\dots$$
$$\hat{y}(k+H_p/k) = CA^{H_p}x(k) + CA^{H_p-1}Bu(k) + \dots + CABu(k+H_p-2) + CBu(k+H_p-1)$$

Observers – What for?

Predictions:

$$\begin{bmatrix}\hat{y}(k+1/k)\\ \hat{y}(k+2/k)\\ \vdots\\ \hat{y}(k+H_p/k)\end{bmatrix} = \begin{bmatrix}CA\\ CA^2\\ \vdots\\ CA^{H_p}\end{bmatrix} x(k) + \begin{bmatrix}CB & 0 & \dots & & 0\\ CAB & CB & 0 & \dots & 0\\ \vdots & & & & \\ CA^{H_p-1}B & \dots & & CAB & CB\end{bmatrix}\begin{bmatrix}u(k)\\ u(k+1)\\ \vdots\\ u(k+H_p-2)\\ u(k+H_p-1)\end{bmatrix}$$

$$= \begin{bmatrix}CA\\ CA^2\\ \vdots\\ CA^{H_p}\end{bmatrix} x(k) + \begin{bmatrix}CB & 0 \quad \dots & & 0\\ \vdots & & \dots & \\ CA^{H_p-2}B & \dots & CA^{H_p-H_c}B & \sum_{i=0}^{H_p-H_c-1} CA^iB\\ CA^{H_p-1}B & \dots & CA^{H_p-H_c+1}B & \sum_{i=0}^{H_p-H_c} CA^iB\end{bmatrix}\begin{bmatrix}u(k)\\ u(k+1)\\ \vdots\\ u(k+H_c-2)\\ u(k+H_c-1)\end{bmatrix}$$

To take into account the control horizon constraint, we keep the first H_c - 1 columns and we sum the remaining columns to form the last one, since $u(k + H_c - 1) = u(k + H_c) = \ldots = u(k + H_p - 1)$

Observers – What for?

- For the predictions to be functions of only $u(k)$, …, $u(k + H_c - 1)$, we need the value of the state vector $x(k)$

- Infinity of state-space representations for a same linear model: the states may or may not represent real signals

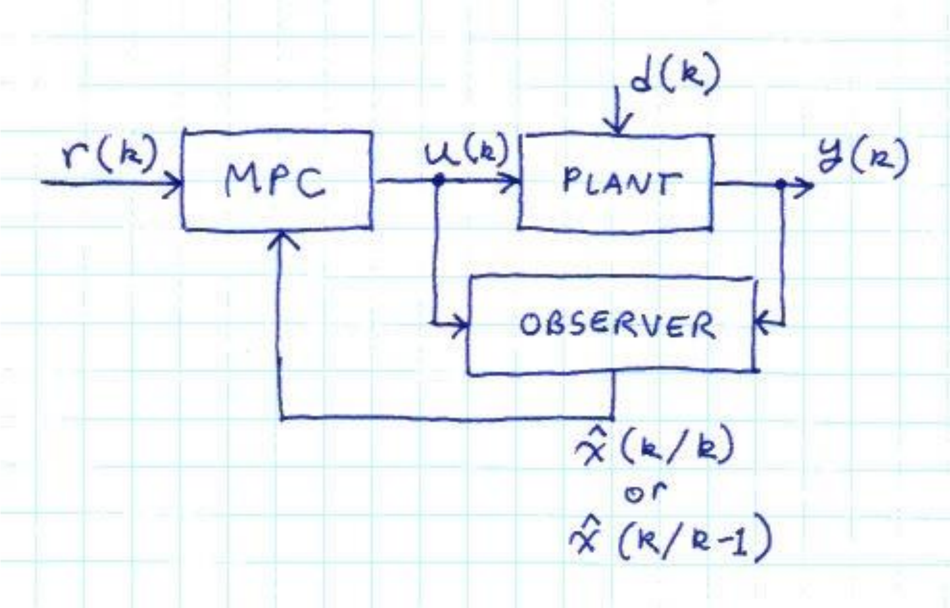
- Even if they represent real signals, the states are very seldom all measured

➢ An observer is an algorithm to estimate the states (state estimates $\hat{x}$):

  - Luenberger observer: deterministic observer, designed by pole-placement
  - Kalman filter: stochastic optimal observer

Observers – What for?

MPC based on a state-space model



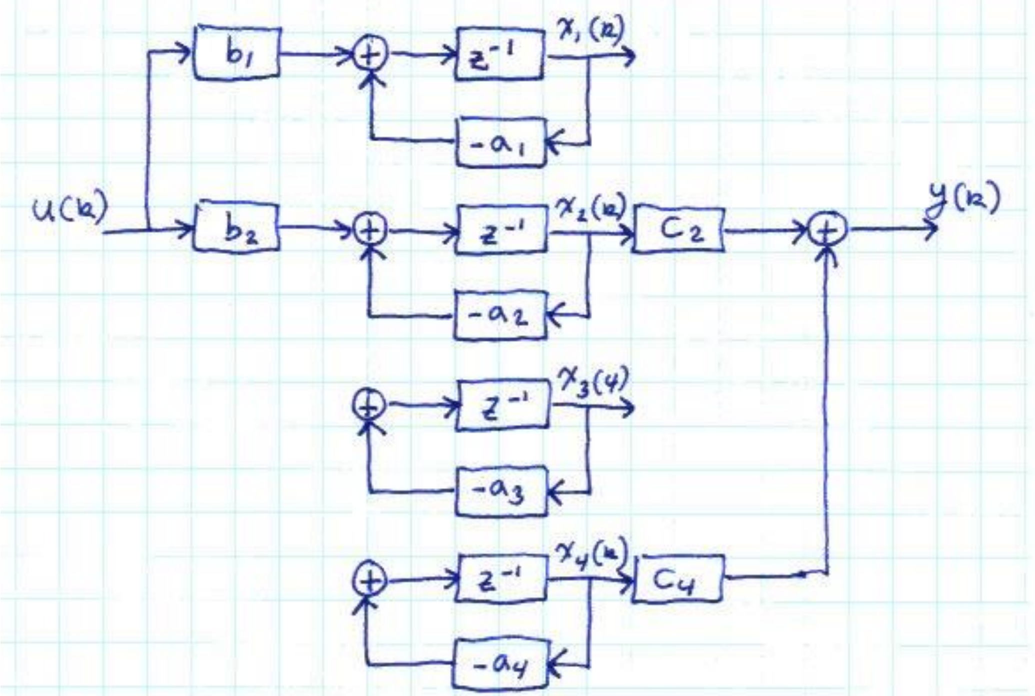
Depending of the observer, the estimate of $x(k)$ may have been calculated:

- at time k: $\hat{x}(k/k)$
- at time $k-1$: $\hat{x}(k/k-1)$

➢ The model must be controllable and observable

Observers – Controllability/Observability

- Controllability: Can we design control inputs to steer the model states to arbitrarily values?
- Observability: Without knowing the initial state, can we determine the state of a model given the input and the output?



x_1: controllable, not observable x_2: controllable, observable

x_3: not controllable, not observable x_4: not controllable, observable

Observers – Controllability

Controllability

For a discrete model with n states and n_u inputs, can we find $u(0)$ to $u(n - 1)$ to steer the states from its initial condition $x(0)$ to any $x(n)$?

$$x(1) = Ax(0) + Bu(0)$$

$$x(2) = Ax(1) + Bu(1)$$

$$= A^2x(0) + ABu(0) + Bu(1)$$

$$\ldots$$

$$x(n) = A^n x(0) + A^{n-1}Bu(0) + \ldots + ABu(n-2) + Bu(n-1)$$

$$\underbrace{\begin{bmatrix} B & AB & \ldots & A^{n-1}B \end{bmatrix}}_{\substack{Q_{con} \\ n \times n \cdot n_u}} \underbrace{\begin{bmatrix} u(n-1) \\ u(n-2) \\ \vdots \\ u(0) \end{bmatrix}}_{\substack{U_{con} \\ n \cdot n_u \times 1}} = \underbrace{x(n) - A^n x(0)}_{\substack{W_{con} \\ n \times 1}}$$

Q_{con} : controllability matrix, known
W_{con} : known

A unique solution exists, and thus the system if fully controllable, if:

$\mathrm{rank}(Q_{con}) = n$ (Q_{con} is full rank)

Observers – Observability

Observability

For a discrete model with n states and n_y outputs, can we find its initial state $x(1)$ from the knowledge of $y(i)$ and $u(i)$ for $i = 1$ to n?

$$y(1) = Cx(1)$$

$$y(2) = Cx(2) = CAx(1) + CBu(1)$$

$$\cdots$$

$$y(n) = CA^{n-1}x(1) + CA^{n-2}Bu(1) + \dots + CABu(n-2) + CBu(n-1)$$

$$\underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}}_{\substack{Q_{obs} \\ n \cdot n_y \times n}} \underbrace{x(1)}_{n \times 1} = \underbrace{\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(n) \end{bmatrix} - \begin{bmatrix} 0 \\ CBu(1) \\ \vdots \\ CA^{n-2}Bu(1) + \dots + CBu(n-1) \end{bmatrix}}_{\substack{W_{obs} \\ n \cdot n_y \times 1}}$$

Q_{obs} : observability matrix, known

W_{obs} : known

A unique solution exists, and thus the system if fully observable, if:

$\mathrm{rank}(Q_{obs}) = n$ (Q_{obs} is full rank)

Observers\Controllability_Observability.mlx

Observers – Luenberger

Structure of an observer

Model: $y(k) = Cx(k)$

$$x(k+1) = Ax(k) + Bu(k)$$

General equation of an observer:

$$\hat{x}(k+1/k) = \underbrace{A\hat{x}(k/k-1) + Bu(k)}_{\text{Model}} + \boxed{K\underbrace{\left[y(k) - \hat{y}(k/k-1)\right]}_{\text{Output prediction error}}}$$

Correction based on feedback

Output prediction: $\hat{y}(k/k-1) = C\hat{x}(k/k-1)$

Gain of correction: K $(n \times n_y)$

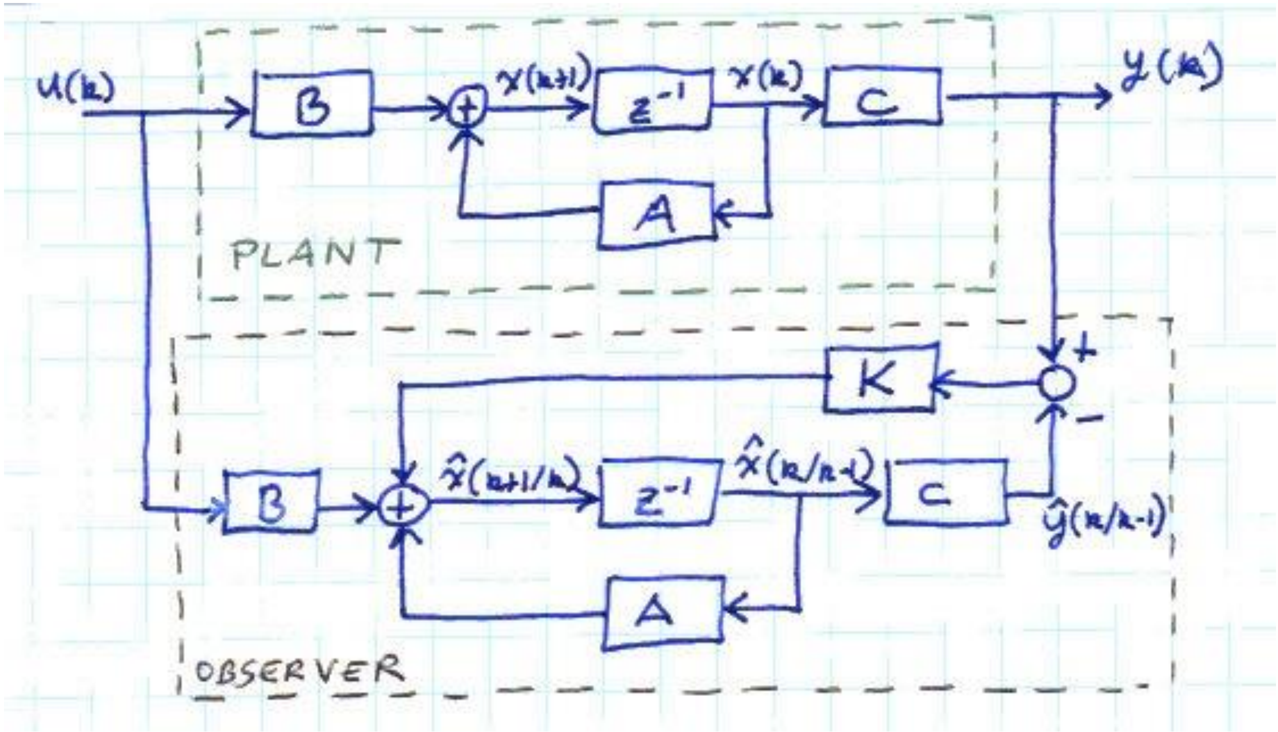
If $K = 0$: open-loop observer (like internal model control)

Observers – Luenberger

$$\hat{x}(k+1/k) = A\hat{x}(k/k-1) + Bu(k) + K\left[y(k) - \hat{y}(k/k-1)\right]$$

$$= A\hat{x}(k/k-1) + Bu(k) + Ky(k) - KC\hat{x}(k/k-1)$$

$$\boxed{\hat{x}(k+1/k) = (A-KC)\hat{x}(k/k-1) + Bu(k) + Ky(k)}$$



Observers – Luenberger

Estimation error

Time evolution of the estimation error, without disturbances and model mismatches:

$$
\begin{aligned}
e(k+1) &= x(k+1) - \hat{x}(k+1/k) \\
&= Ax(k) + Bu(k) - (A-KC)\hat{x}(k/k-1) - Bu(k) - Ky(k) \\
&= A\left[x(k) - \hat{x}(k/k-1)\right] - KC\left[x(k) - \hat{x}(k/k-1)\right] \quad \text{since } y(k) = Cx(k) \\
&= (A-KC)\left[x(k) - \hat{x}(k/k-1)\right]
\end{aligned}
$$

$$
\boxed{e(k+1) = (A-KC)e(k)}
$$

- The estimation error will go to zero (the observer is stable) if the eigenvalues of A - KC lie within the unit circle (just like the stability of the model $x(k+1) = Ax(k)$)
- The dynamics of the error are dictated by the eigenvalues of A - KC
- Selecting K to get the desired eigenvalues (poles): **Luenberger observer**

Observers – Luenberger

Example – Design of a Luenberger observer

Model: $y(k) = \begin{bmatrix} -0.5 & 1 \end{bmatrix} x(k)$

$$x(k+1) = \begin{bmatrix} 0.82 & 0 \\ 0 & 0.9 \end{bmatrix} x(k) + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u(k)$$

Observer:

$$K = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = ?$$

Desired dynamics of the estimation error: all poles equal to 0.3

The poles are obtained from the characteristic equation:

$$\det\left[zI - A + KC\right] = 0 = (z-0.3)(z-0.3)$$

$$zI - A + KC = \begin{bmatrix} z - 0.82 - 0.5k_1 & k_1 \\ -0.5k_2 & z - 0.9 + k_2 \end{bmatrix}$$

$$\det\left[zI - A + KC\right] = z^2 + (-1.72 + k_2 - 0.5k_1)z + \left((-0.82 - 0.5k_1)(-0.9 + k_2) + 0.5k_1k_2\right)$$

$$= z^2 + \qquad (-0.6)z \qquad + \qquad 0.09$$

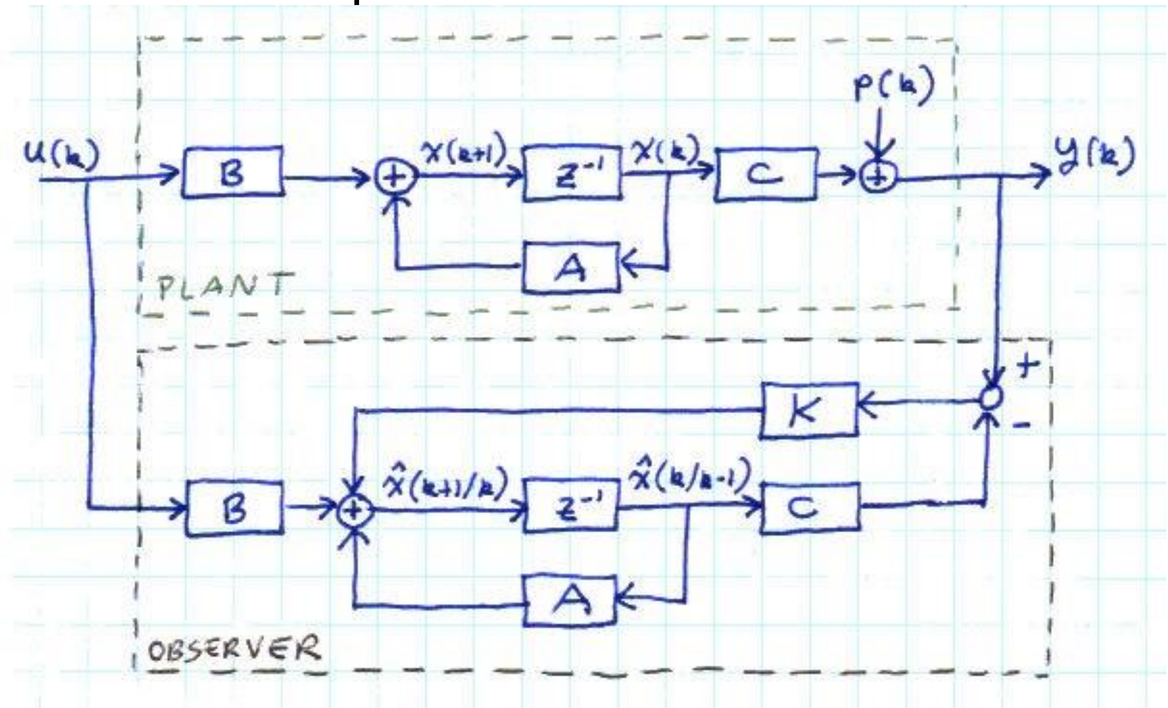
Solving term by term: $k_1 = 6.76 \quad k_2 = 4.5$

Observers\Luenb_no_integr.mlx

Observers – Luenberger with integration

Observation in presence of disturbances

- Both a model mismatch or a disturbance have a similar effect: they modify $y(k)$ unexpectedly and thus change the state estimates
- To simplify the analysis (without loosing the generality), we assume a perfect model and an output disturbance



- What is the estimation error?

Observers – Luenberger with integration

Estimation error in presence of disturbances

Plant: $y(k) = Cx(k) + p(k)$

$$x(k+1) = Ax(k) + Bu(k)$$

Observer:
$$\begin{aligned}\hat{x}(k+1/k) &= A\hat{x}(k/k-1) + Bu(k) + K\left[y(k) - C\hat{x}(k/k-1)\right] \\ &= A\hat{x}(k/k-1) + Bu(k) + K\left[Cx(k) + p(k) - C\hat{x}(k/k-1)\right] \\ &= (A-KC)\hat{x}(k/k-1) + Bu(k) + KCx(k) + Kp(k)\end{aligned}$$

Estimation error:

$$\begin{aligned}e(k+1) &= x(k+1) - \hat{x}(k+1/k) \\ &= Ax(k) + Bu(k) - (A-KC)\hat{x}(k/k-1) - Bu(k) - KCx(k) - Kp(k) \\ &= (A-KC)\left[x(k) - \hat{x}(k/k-1)\right] - Kp(k) \\ &= (A-KC)e(k) - Kp(k)\end{aligned}$$

Observers – Luenberger with integration

$$e(k+1) = (A - KC)e(k) - Kp(k)$$

- Assumptions:
  - The observer is stable (eigenvalues of $A - KC$ lie within the unit circle)
  - The disturbance is constant: $p(k) = p$
- Since the observation will converge, in steady-state we will have $e(k + 1) = e(k) = e(\infty)$ and thus

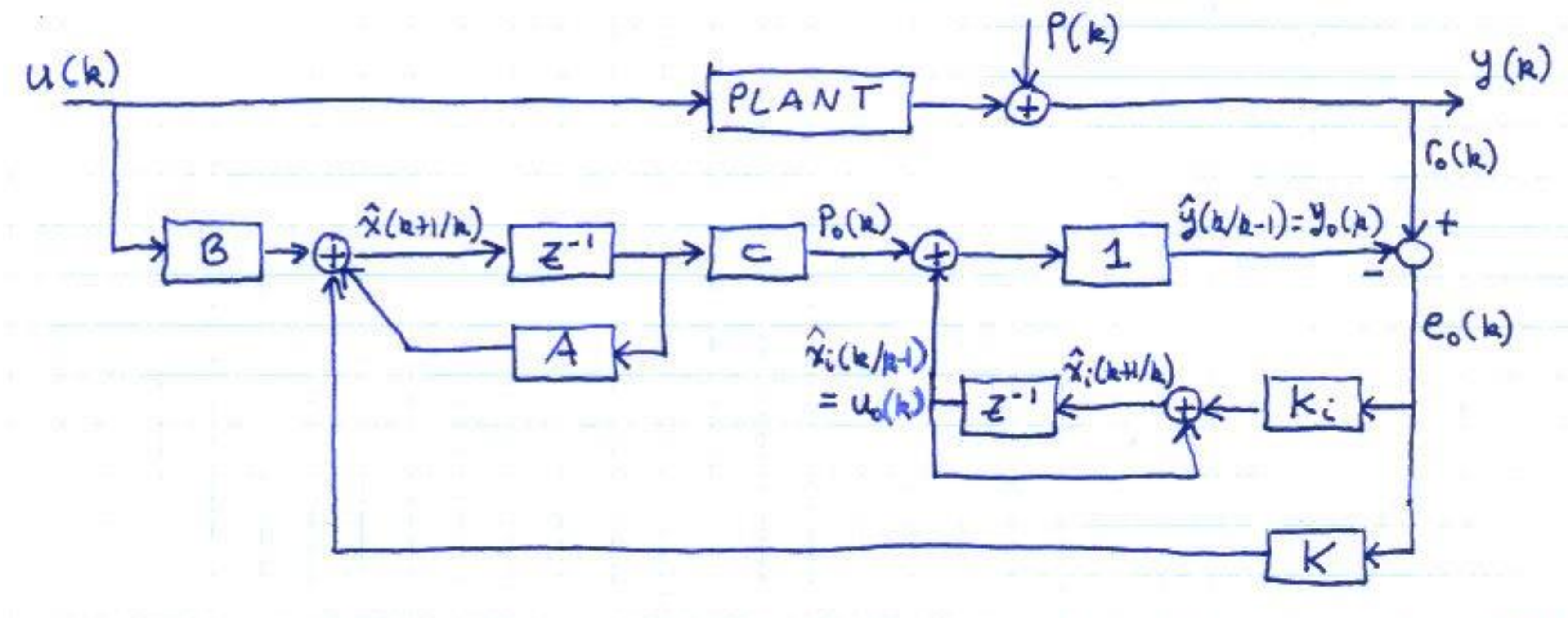
$$\begin{aligned} e(\infty) &= (A - KC)e(\infty) - Kp \\ &= (I - A + KC)^{-1}(-Kp) \end{aligned} \quad \Rightarrow \quad \begin{aligned} y(\infty) - \hat{y}(\infty) &= Cx(\infty) - C\hat{x}(\infty) = Ce(\infty) \\ &= C(I - A + KC)^{-1}(-Kp) \end{aligned}$$

- The estimation error will converge to zero only if $p = 0$
- Therefore, the predictions $\hat{y}$ (based on $\hat{x}$) and y will not converge to the same value in steady-state: **MPC will not provide a zero static error**
- Solution: add an integral action to the observer, or equivalently, estimate the disturbance

Observers – Luenberger with integration

Luenberger observer with integral action

The objective is to obtain $\hat{y}(k/k-1) = y(k)$ in steady-state despite an output disturbance

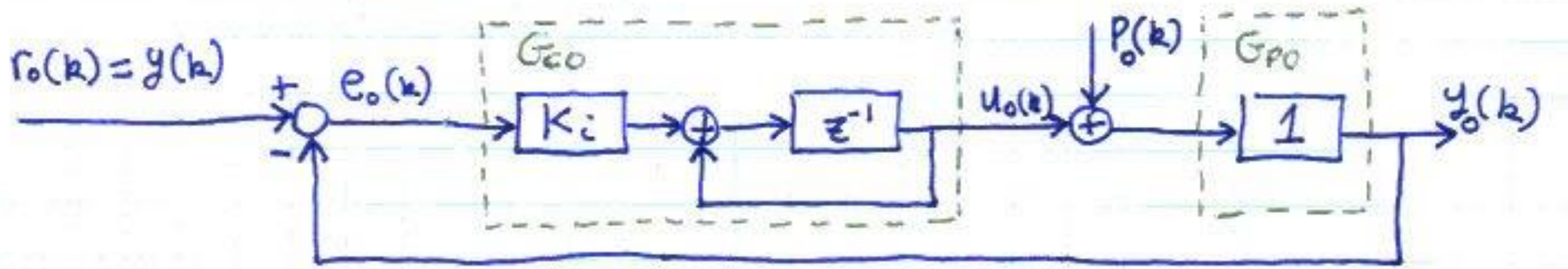


Observers – Luenberger with integration

Let`s analyse the right part of the observer as if it was a basic control loop with the controller G_{co}, the plant G_{po}, the set-point r_o, the error e_o, the manipulated variable u_o, the input disturbance p and the controlled variable y_o

Regulator with an integral action (pole $z = 1$)

$$u_o(k) = u_o(k-1) + K_i e_o(k-1) \Rightarrow G_{co}(z) = \frac{U(z)}{E_o(z)} = \frac{K_i}{z-1}$$



Observers – Luenberger with integration

Does that lead to $e_o(\infty)=0$ (i.e. to $\hat{y}(k/k-1)=y(k)$) in steady-state?

Suppose that G_{co} has m integrators (m is an integer > 0) and analyze the signal e_o:

$$G_{co}(z)=\frac{K_i}{(z-1)^m} \qquad E_o(z)=\frac{1}{1+G_{co}(z)}R_o(z)-\frac{1}{1+G_{co}(z)}P_o(z)$$

We analyze one input at the time (linear superposition)

a) Effect of $R_o(z)$ (with $P_o=0$)

$$R_o(z)=\frac{z}{(z-1)^n} \qquad n=1\text{: step, } n=2\text{: ramp; etc.}$$

$$E_o(z)=\frac{1}{1+G_{co}(z)}R_o(z)=\frac{z(z-1)^{m-n}}{(z-1)^m+K_i}$$

Observers – Luenberger with integration

$$E_o(z) = \frac{z(z-1)^{m-n}}{(z-1)^m + K_i}$$

Using the final value theorem:

$$e_o(\infty) = \lim_{z\to 1}(z-1)E_o(z) = \lim_{z\to 1}\frac{z(z-1)^{m-n+1}}{(z-1)^m + K_i}$$

$e_o(\infty) = 0$ if $m - n + 1 > 0$ or if $m > n - 1$

- r_o is a constant (step): at least one integrator is required in the observer
- $u(k)$ is a constant, plant with one integrator and $p(k)$ is a constant, than r_o is a ramp: at least two integrators are required in the observer
- $u(k)$ is a constant, plant without integrator and $p(k)$ is a ramp, than r_o is a ramp: at least two integrators are required in the observer
- etc.

Observers – Luenberger with integration

b) Effect of $P_o(z)$ (with $R_o = 0$)

Except for the sign, same as with R_o: same conclusion

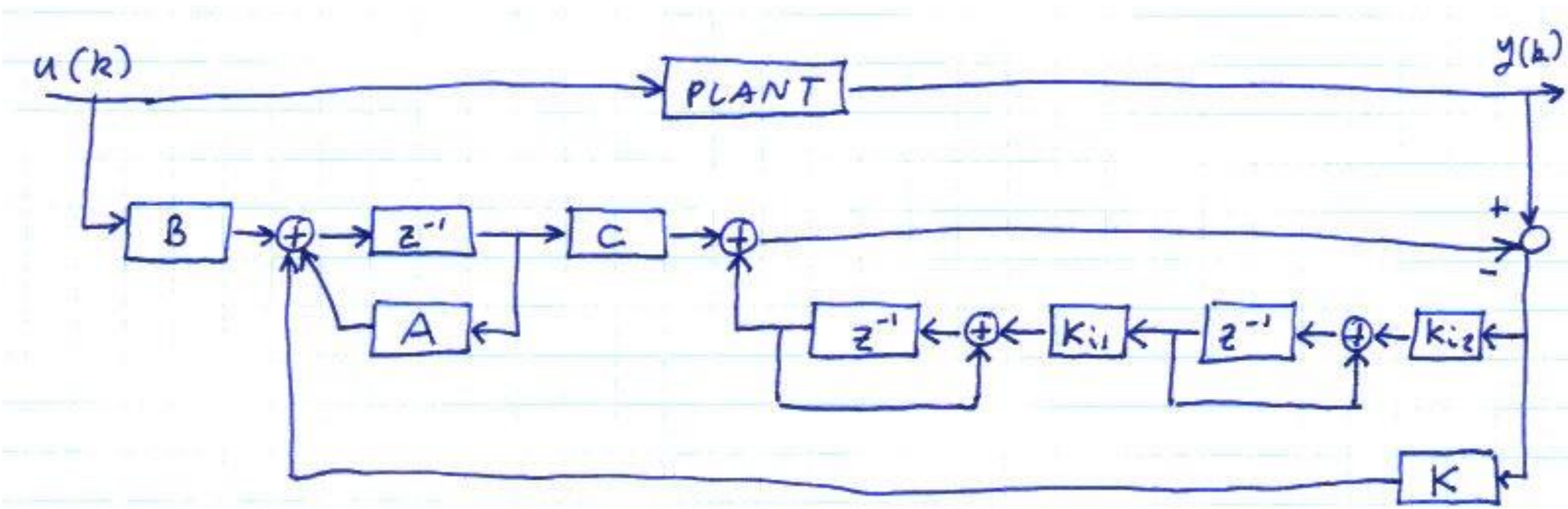
- $u(k)$ is a constant and model with one integrator than p_o is a ramp: two integrators are required in the observer
- etc.

We select the minimal value m leading to $e_o(\infty) = 0$ for both R_o and P_o

In practice, when coupling the observer to a MPC: the number of integrators in the observer for each plant output = 1 + number of integrators in the plant

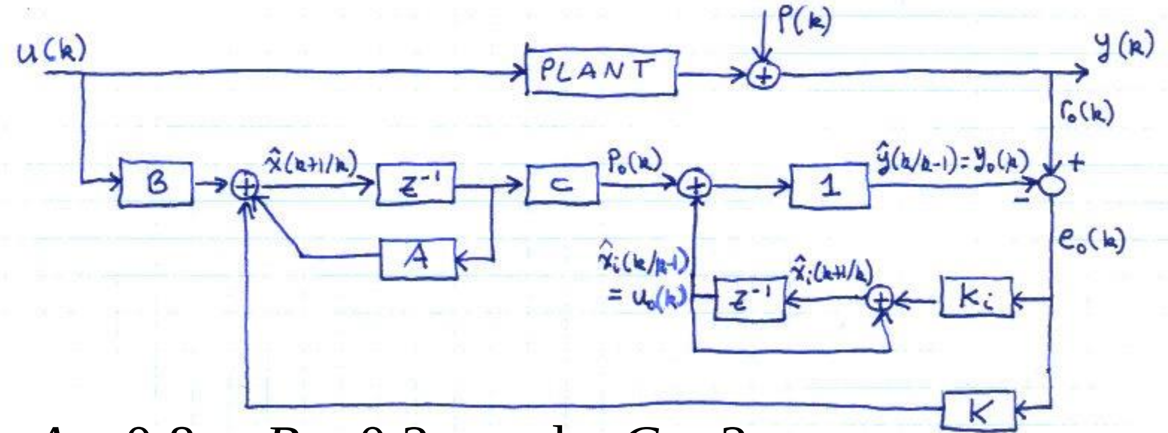
Observers – Luenberger with integration

Observer with two integrators



Observers – Luenberger with integration

Example – Design of a Luenberger observer with integration



$$A = 0.8, \quad B = 0.2 \quad \text{and} \quad C = 2$$

$$\hat{x}(k+1/k) = 0.8\hat{x}(k/k-1) + Ky(k) - 2K\hat{x}(k/k-1) - K\hat{x}_i(k/k-1) + 0.2u(k)$$

$$\hat{x}_i(k+1/k) = \hat{x}_i(k/k-1) + K_i y(k) - 2K_i\hat{x}(k/k-1) - K_i\hat{x}_i(k/k-1)$$

$$\begin{bmatrix} \hat{x}(k+1/k) \\ \hat{x}_i(k+1/k) \end{bmatrix} = \begin{bmatrix} 0.8-2K & -K \\ -2K_i & 1-K_i \end{bmatrix} \begin{bmatrix} \hat{x}(k/k-1) \\ \hat{x}_i(k/k-1) \end{bmatrix} + \begin{bmatrix} 0.2 \\ 0 \end{bmatrix} u(k) + \begin{bmatrix} K \\ K_i \end{bmatrix} y(k)$$

Identical to the general form:

$$\hat{x}(k+1/k) = (A-KC)\hat{x}(k/k-1) + Bu(k) + Ky(k)$$

Observers – Luenberger with integration

$$\det\begin{bmatrix} z-0.8-2K & K \\ 2K_i & z-1+K_i \end{bmatrix} = z^2+(K_i+2K-1.8)z+(0.8-0.8K_i-2K)$$

If we want the poles $z = 0$ and $z = 0.7$:

$$\begin{aligned} z^2+(K_i+2K-1.8)z+(0.8-0.8K_i-2K) &= z(z-0.7) \\ &= z^2-0.7z+0 \end{aligned}$$

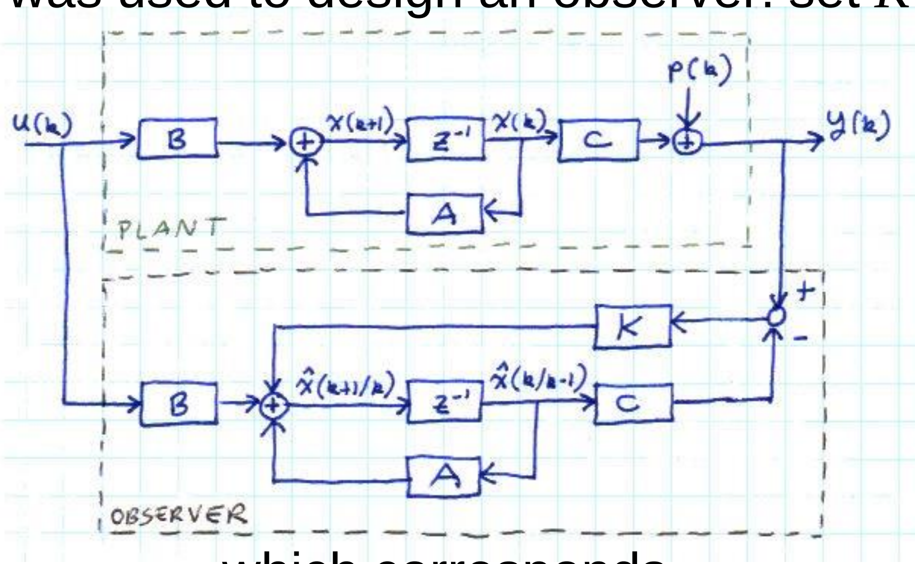
leading to $K = -0.2$ and $K_i = 1.5$

Observers – Luenberger with integration

Direct design with an augmented model

How can we obtain an observer with integration for a plant model by designing an observer without integration for an *augmented model* of the plant? In that case, what is the *augmented model*?

To retrieve the model that was used to design an observer: set $K = 0$ (open-loop observer = model)



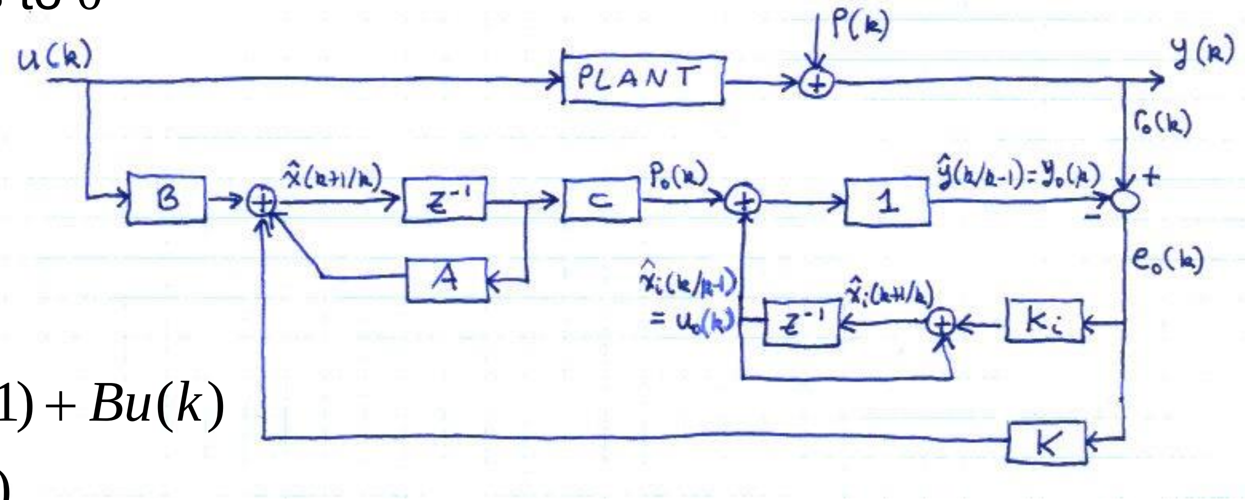
$$\hat{y}(k/k-1) = C\hat{x}(k/k-1)$$
$$\hat{x}(k+1/k) = A\hat{x}(k/k-1) + Bu(k)$$

which corresponds to the model used for its design

$$y(k) = Cx(k)$$
$$x(k+1) = Ax(k) + Bu(k)$$

Observers – Luenberger with integration

Retrieve the model that was used to design an observer with integration by setting the observer gains to 0



$$\hat{x}(k+1/k) = A\hat{x}(k/k-1) + Bu(k)$$

$$\hat{x}_i(k+1/k) = \hat{x}_i(k/k-1)$$

$$\hat{y}(k/k-1) = C\hat{x}(k/k-1) + \hat{x}_i(k/k-1)$$

$$\left.\begin{aligned}
\begin{bmatrix} \hat{x}(k+1/k) \\ \hat{x}_i(k+1/k) \end{bmatrix} &= \begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}(k/k-1) \\ \hat{x}_i(k/k-1) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u(k) \\
\hat{y}(k/k-1) &= \begin{bmatrix} C & 1 \end{bmatrix} \begin{bmatrix} \hat{x}(k/k-1) \\ \hat{x}_i(k/k-1) \end{bmatrix}
\end{aligned}\right\}
\quad
\begin{aligned}
x_a(k+1) &= A_a x_a(k) + B_a u(k) \\
y(k) &= C_a x_a(k)
\end{aligned}$$

Observers – Luenberger with integration

Indeed, if we design an observer without integration with the augmented model

$$\begin{bmatrix} x(k+1) \\ x_i(k+1) \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x(k) \\ x_i(k) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} C & 1 \end{bmatrix} \begin{bmatrix} x(k) \\ x_i(k) \end{bmatrix}$$

we obtain

$$\begin{bmatrix} \hat{x}(k+1/k) \\ \hat{x}_i(k+1/k) \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}(k/k-1) \\ \hat{x}_i(k/k-1) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u(k) + \begin{bmatrix} K \\ K_i \end{bmatrix} \left(y(k) - \begin{bmatrix} C & 1 \end{bmatrix} \begin{bmatrix} \hat{x}(k/k-1) \\ \hat{x}_i(k/k-1) \end{bmatrix} \right)$$

$$\hat{x}_a(k+1/k) = A_a \hat{x}_a(k/k-1) + B_a u(k) + K_a \left(y(k) - C_a \hat{x}_a(k/k-1) \right)$$

which corresponds to the figure of the previous page

Observers – Luenberger with integration

For a system with n states, n_u inputs, n_y outputs, with one integrator for each output, the augmented model is:

$$A_a = \begin{bmatrix} A & 0_{n\times n_y} \\ 0_{n_y\times n} & I_{n_y} \end{bmatrix}$$

$$B_a = \begin{bmatrix} B \\ 0_{n_y\times n_u} \end{bmatrix}$$

$$C_a = \begin{bmatrix} C & I_{n_y} \end{bmatrix}$$

where I_{n_y} is a $n_y \times n_y$ identity matrix and $0_{i\times j}$ is a null matrix of dimensions $i\times j$

Observers\Luenb_with_integr.mlx

Observers – Luenberger with integration

Other interpretation: estimation of the output disturbance

Design of an observer based on the augmented model:

$$\begin{bmatrix} \hat{x}(k+1/k) \\ \hat{x}_i(k+1/k) \end{bmatrix} = \begin{bmatrix} A & 0_{n\times n_y} \\ 0_{n_y\times n} & I_{n_y} \end{bmatrix} \begin{bmatrix} \hat{x}(k/k-1) \\ \hat{x}_i(k/k-1) \end{bmatrix} + \begin{bmatrix} B \\ 0_{n_y\times n_u} \end{bmatrix} u(k) + \begin{bmatrix} K \\ K_i \end{bmatrix} \left(y(k) - \begin{bmatrix} C & I_{n_y} \end{bmatrix} \begin{bmatrix} \hat{x}(k/k-1) \\ \hat{x}_i(k/k-1) \end{bmatrix} \right)$$

Suppose that it converges, in steady-state we have:

$$\hat{x}(k+1/k) = \hat{x}(k/k-1) = \hat{x} \qquad y(k) = y$$

$$\hat{x}_i(k+1/k) = \hat{x}_i(k/k-1) = \hat{x}_i \qquad u(k) = u$$

Under theses conditions, solving the observer equations leads to:

$$\hat{x} = (I-A)^{-1}Bu$$

$$\hat{x}_i = y - \underbrace{C(I-A)^{-1}B}_{\text{Model transfer function}}u$$

$$\Rightarrow$$

$$\hat{y} = C\hat{x} + \hat{x}_i$$

$$= C(I-A)^{-1}Bu + y - C(I-A)^{-1}Bu$$

$= y$ as expected!

Similar to an IMC structure but with feedback (K, K_i): OK with unstable systems

Observers – Luenberger with integration

In presence of a constant output disturbance the plant is:

$$x(k+1) = Ax(k) + Bu(k)$$
$$y(k) = Cx(k) + p$$

In steady-state:

$$x = (I - A)^{-1} Bu$$
$$y = Cx + p = C(I - A)^{-1} Bu + p$$

And thus the estimation of the integration state is:

$$\begin{aligned}\hat{x}_i &= y - C(I - A)^{-1} Bu \\ &= C(I - A)^{-1} Bu + p - C(I - A)^{-1} Bu \\ &= p\end{aligned}$$

The estimate of the integration state is p, thus compensating to ensure that $\hat{y}(k / k-1) = y(k)$ in steady state!

Observers\Luenb_with_without_integr.mlx

Observers – Kalman filter

Observer based on the stochastic plant model:

$$x(k+1) = Ax(k) + Bu(k) + w(k)$$
$$y(k) = Cx(k) + v(k)$$

$$x = [x_1 \quad \dots \quad x_n]^T : n \times 1$$
$$w : n \times 1$$
$$v : n_y \times 1$$

where $w(k)$ is the process noise and $v(k)$ the measurement noise. They are both white noises and are uncorrelated:

$$E\{w(k)\} = E\{v(k)\} = 0 \quad \text{(zero mean)}$$
$$E\{w(k)v^T(i)\} = 0 \quad \forall k, i \qquad (w \text{ and } v \text{ uncorrelated})$$
$$E\{w(k)w^T(i)\} = 0 \quad k \neq i \qquad \text{(random)}$$
$$E\{w(k)w^T(k)\} = Q \qquad \text{(diagonal cov. matrix)}$$
$$E\{v(k)v^T(i)\} = 0 \quad k \neq i \qquad \text{(random)}$$
$$E\{v(k)v^T(k)\} = R \qquad \text{(diagonal cov. matrix)}$$

Observers – Kalman filter

Kalman filter

$$\hat{x}(k+1/k) = A\hat{x}(k/k-1) + Bu(k) + K(k)\left[y(k) - C\hat{x}(k/k-1)\right]$$

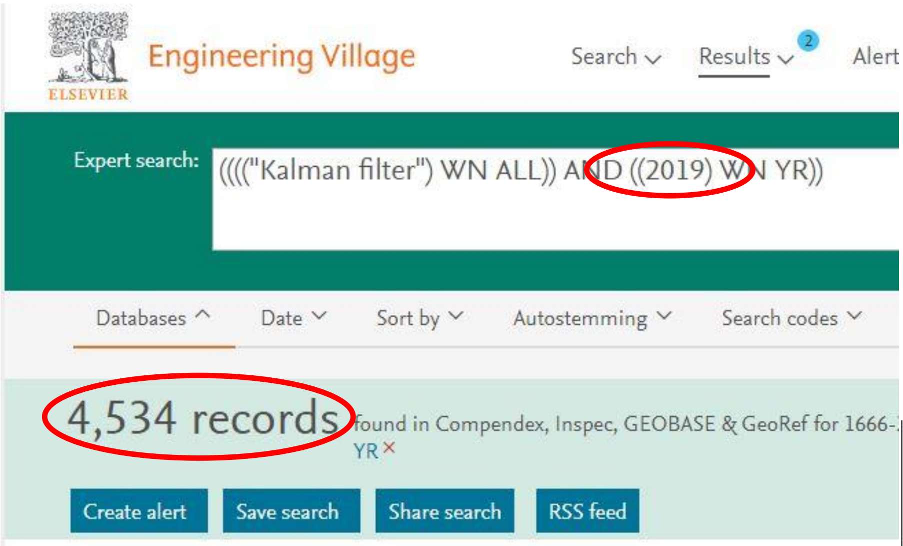
where the gain $K(k)$ is calculated at each sampling period to minimize the variance of the estimation error.

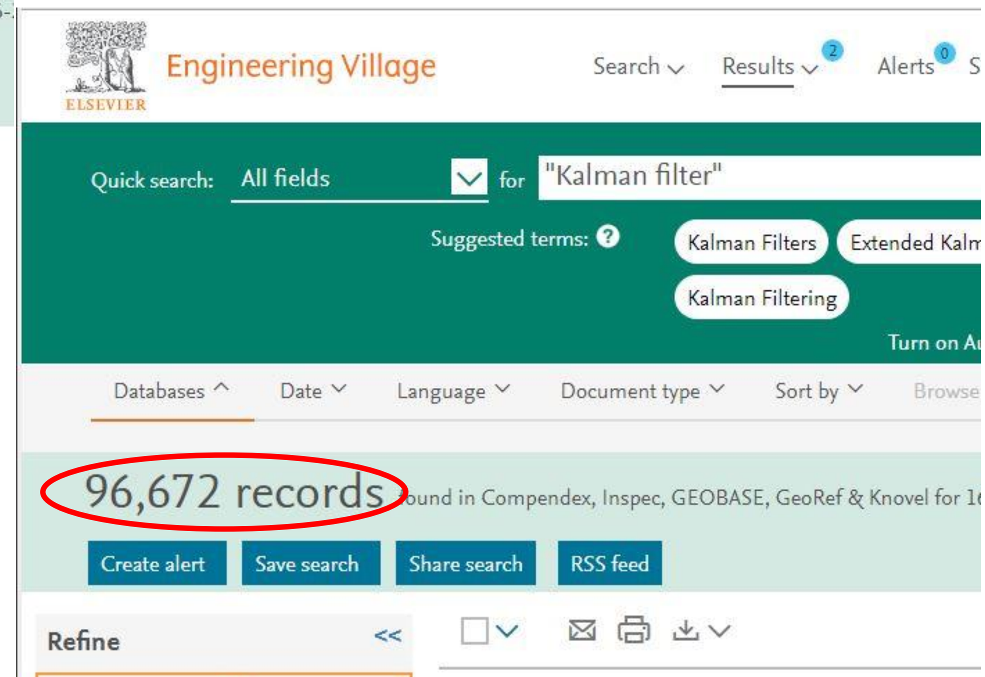
- With Gaussian noises: Kalman filter is optimal (i.e. minimum variance estimate of the states)
- Without the Gaussian assumptions: linear observer with minimum variance estimate of the states
- Date back to 1960:

  R.E. Kalman, A New Approach to Linear Filtering and Prediction Problems, Trans. of the ASME – Journal of Basic Engineering, 82 (Series D), pp. 35-45, 1960. [8]

- www.cs.unc.edu\~welch\kalman

Observers – Kalman filter

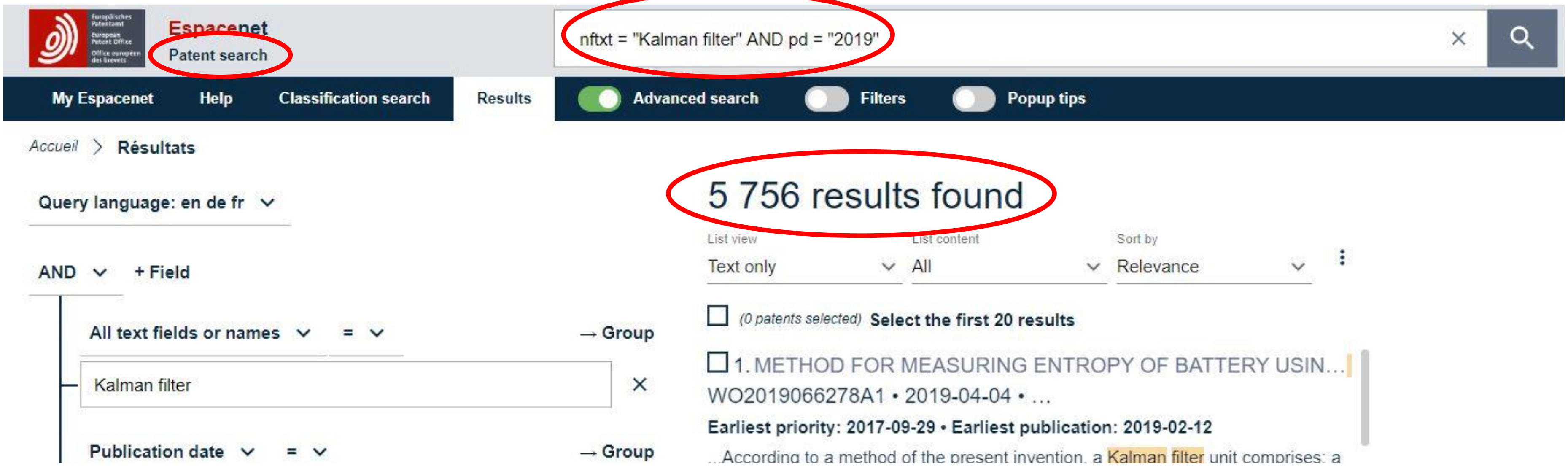


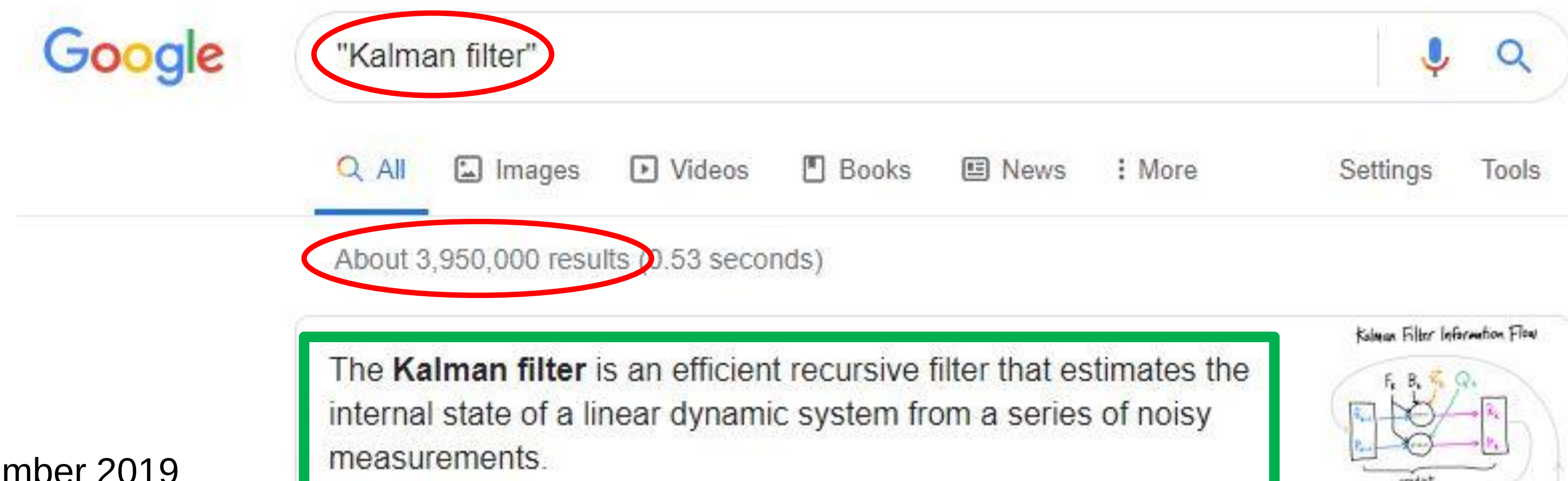


December 2019

Observers – Kalman filter

pd: publication date





December 2019

Observers – Kalman filter

Kalman filter:

$$\hat{x}(k+1/k) = A\hat{x}(k/k-1) + Bu(k) + K(k)\left[y(k) - C\hat{x}(k/k-1)\right]$$

where the gain $K(k)$ is calculated at each sampling period to minimize

$$\text{trace}\left[E\left\{\left[x(k+1) - \hat{x}(k+1/k)\right]\left[x(k+1) - \hat{x}(k+1/k)\right]^T\right\}\right]$$

i.e. the sum of the variance of all n estimation errors

$$\sum_{i=1}^{n} E\left\{\left[x_i(k+1) - \hat{x}_i(k+1/k)\right]^2\right\}$$

The covariance matrix of the estimation error is denoted P:

$$P(k+1/k) = E\left\{\left[x(k+1) - \hat{x}(k+1/k)\right]\left[x(k+1) - \hat{x}(k+1/k)\right]^T\right\}$$

Observers – Kalman filter

The result is (prediction form – proof in Annex A):

1. Initialization of $P(0/-1)$ and $\hat{x}(0/-1)$
   If the confidence in $\hat{x}(0/-1)$ is low than $P(0/-1)$ is large, e.g. $1000I$

2. Correction gain: $K(k) = AP(k/k-1)C^T\left(CP(k/k-1)C^T + R\right)^{-1}$

3. Optimal states:

$$\hat{x}(k+1/k) = A\hat{x}(k/k-1) + Bu(k) + K(k)\overbrace{\left[y(k) - C\hat{x}(k/k-1)\right]}^{\text{Innovation sequence}}$$

4. Covariance matrix update (dynamic Riccati equation):

$$P(k+1/k) = A\left[P(k/k-1) - P(k/k-1)C^T\left(CP(k/k-1)C^T + R\right)^{-1}CP(k/k-1)\right]A^T + Q$$

5. Wait for the next sampling time and go to step 2

Observers – Kalman filter

- A, B and C can be time-varying as long as it is known
- Q and R can be time-varying as long as it is known
- If A, B, C, Q and R are constant than after quite a short transient, K and P will become constant, with values that are independent of u and y:

$$P(\infty) = A\left[P(\infty) - P(\infty)C^T\left(CP(\infty)C^T + R\right)^{-1}CP(\infty)\right]A^T + Q$$

$$K(\infty) = AP(\infty)C^T\left(CP(\infty)C^T + R\right)^{-1}$$

The first equation (algebraic Riccati equation) can be solved (Matlab function idare) to get $P(\infty)$

- If from the beginning we use $K(\infty)$: steady-state Kalman filter (which is the case in the Matlab MPC toolbox) – not much difference because short transient

Observers – Kalman filter

$$x(k+1) = Ax(k) + Bu(k) + w(k) \rightarrow Q$$
$$y(k) = Cx(k) + v(k) \rightarrow R$$

- Selection of R: diagonal matrix with the variance of the measurements on the diagonal ⇨ from the specifications of the sensors or by an estimation from the measurements in steady-state

- Selection of Q: diagonal matrix of the process noises, include the dynamic model mismatches… not obvious! Often in practice: tuning knobs

- Matlab: $$x(k+1) = Ax(k) + Bu(k) + \overset{n\times n}{G}w(k)$$
$$y(k) = Cx(k) + Du(k) + \underset{n_y\times n}{H}w(k) + v(k)$$

$$E\{w(k)\} = E\{v(k)\} = 0$$
$$E\{w(k)w^T(i)\} = 0 \quad k \neq i$$
$$E\{w(k)w^T(k)\} = Q$$
$$E\{v(k)v^T(i)\} = 0 \quad k \neq i$$
$$E\{v(k)v^T(k)\} = R$$
$$E\{w(k)v^T(i)\} = N$$

```
Gmodel=ss(A,[B G],C,[D H],Ts);
[KF,K,P]=kalman(Gmodel,Q,R,N);
```

$$\begin{array}{l} u \rightarrow \\ y \rightarrow \end{array} \boxed{\text{KF}} \begin{array}{l} \rightarrow \hat{y} \\ \rightarrow \hat{x} \end{array}$$

Observers\Kalman_Filter.mlx

Observers – Kalman filter

Filtering form (proof in Annex A):

1. Initialization of $P(0/-1)$ and $\hat{x}(0/-1)$

2. Correction gain: $K_f(k) = P(k/k-1)C^T\left(CP(k/k-1)C^T + R\right)^{-1}$

3. Optimal states: $\hat{x}(k/k) = \hat{x}(k/k-1) + K_f(k)\left[y(k) - C\hat{x}(k/k-1)\right]$

4. Prediction of the states: $\hat{x}(k+1/k) = A\hat{x}(k/k) + Bu(k)$

5. Covariance matrix update:

$$P(k+1/k) = A\left[P(k/k-1) - P(k/k-1)C^T\left(CP(k/k-1)C^T + R\right)^{-1}CP(k/k-1)\right]A^T + Q$$

6. Wait for the next sampling time and go to step 2

Observers – Kalman filter

- Interpretation by an optimal batch state estimation [9]:

$$\min_{\{\hat{x}(i/k)\}_{i=0}^{k}} \left[\begin{array}{c} \left[\hat{x}(0/k)-\hat{x}(0/-1)\right]^T P(0/-1)\left[\hat{x}(0/k)-\hat{x}(0/-1)\right] \\ +\sum_{j=0}^{k-1}\hat{w}^T(j)Q^{-1}\hat{w}(j)+\sum_{j=0}^{k}\hat{v}^T(j)R^{-1}\hat{v}(j) \end{array} \right]$$

where $\hat{w}(j)=\hat{x}(j+1/k)-A\hat{x}(j/k)+Bu(j)$

$$\hat{v}(j)=y(j)-C\hat{x}(j/k)$$

- The first term in the cost function: quickly with little influence since only one point
- Dimensions of the problem are always growing: at time k we need to estimate $n\times(k+1)$ variables
- When the number of data becomes large, adding new data does not influence much the results (this is why $K(k)$ quickly converges)

Observers – Kalman filter

$$\min_{\{\hat{x}(i/k)\}_{i=0}^{k}} \left[\begin{array}{c} \left[\hat{x}(0/k)-\hat{x}(0/-1)\right]^T P(0/-1)\left[\hat{x}(0/k)-\hat{x}(0/-1)\right] \\ +\sum_{j=0}^{k-1} \hat{w}^T(j)Q^{-1}\hat{w}(j)+\sum_{j=0}^{k} \hat{v}^T(j)R^{-1}\hat{v}(j) \end{array} \right]$$

where $\hat{w}(j)=\hat{x}(j+1/k)-A\hat{x}(j/k)+Bu(j)$

$$\hat{v}(j)=y(j)-C\hat{x}(j/k)$$

- Clearly shows the influence of Q and R:
  - If Q is large and R small: the estimation relies on the measurement equation: $$y(k)=Cx(k)$$
  - If R is large and Q small: the estimation relies on the state equation: $$x(k+1)=Ax(k)+Bu(k)$$
- Leads to the moving horizon estimator for nonlinear systems

Observers – Kalman filter with integration

Kalman filter with integration

- A stochastic output disturbance $p(k)$ can be added

$$x(k+1) = Ax(k) + Bu(k) + w(k)$$

$$y(k) = Cx(k) + v(k) + p(k)$$

where $x_i(k+1) = A_i x_i(k) + B_i e(k)$ with $e(k)$ being a white noise

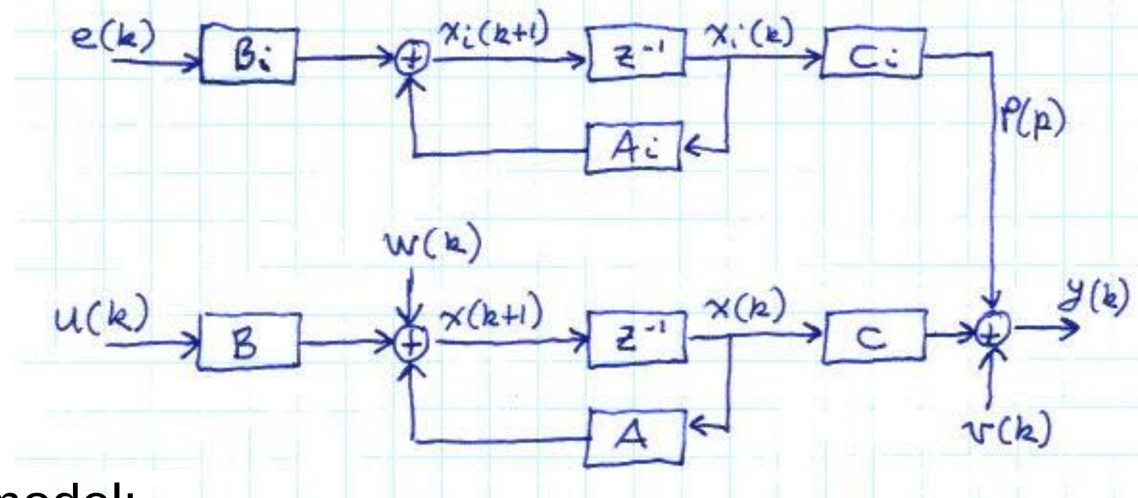
$$p(k) = C_i x_i(k)$$

- The output disturbance model is defined by A_i, B_i and C_i
- If it is an integrator then the disturbance $p(k)$ is then not stationary and its mean is $p(0)$
  - ➢ This is the stochastic equivalent to a constant disturbance with the Luenberger observer
- As with the Luenberger observer with integration, we may want to use two or more integrators in series

Observers – Kalman filter with integration

$$x(k+1) = Ax(k) + Bu(k) + w(k) \qquad x_i(k+1) = A_i x_i(k) + B_i e(k)$$

$$y(k) = Cx(k) + v(k) + p(k) \qquad p(k) = C_i x_i(k)$$



Augmented model:

$$\begin{bmatrix} x(k+1) \\ x_i(k+1) \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & A_i \end{bmatrix} \begin{bmatrix} x(k) \\ x_i(k) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u(k) + \begin{bmatrix} w(k) \\ B_i e(k) \end{bmatrix}$$

$$y(k) = \begin{bmatrix} C & C_i \end{bmatrix} \begin{bmatrix} x(k) \\ x_i(k) \end{bmatrix} + v(k)$$

$$x_a(k+1) = A_a x_a(k) + B_a u(k) + w_a(k)$$

$$y_a(k) = C_a x_a(k) + v(k)$$

Observers – Kalman filter with integration

- Augmented model:

$$\begin{bmatrix} x(k+1) \\ x_i(k+1) \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & A_i \end{bmatrix} \begin{bmatrix} x(k) \\ x_i(k) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u(k) + \begin{bmatrix} w(k) \\ B_i e(k) \end{bmatrix}$$

$$y(k) = \begin{bmatrix} C & C_i \end{bmatrix} \begin{bmatrix} x(k) \\ x_i(k) \end{bmatrix} + v(k)$$

$$x_a(k+1) = A_a x_a(k) + B_a u(k) + w_a(k)$$

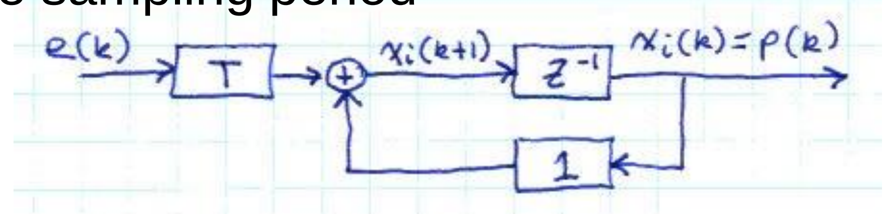
$$y_a(k) = C_a x_a(k) + v(k)$$

- Corresponding Kalman filter:

$$\hat{x}_a(k+1/k) = A_a \hat{x}_a(k/k-1) + B_a u(k) + K(k)\left[y(k) - C_a \hat{x}_a(k/k-1)\right]$$

Example

The output disturbance model is an integrator: $p(k) = p(k - 1) + Te(k)$ where T is the sampling period



$$x_i(k+1) = 1x_i(k) + Te(k)$$

$$p(k) = 1x_i(k)$$

Observers – Kalman filter with integration

- The augmented model is therefore: $A_a = \begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix}, B_a = \begin{bmatrix} B \\ 0 \end{bmatrix}, C_a = \begin{bmatrix} C & 1 \end{bmatrix}$

  and therefore the corresponding observer

$$\hat{x}_a(k+1/k) = A_a \hat{x}_a(k/k-1) + B_a u(k) + K(k)\left[y(k) - C_a \hat{x}_a(k/k-1)\right]$$

  is identical to the [Luenberger with integration](), except for the gain $K(k)$ which is calculated differently

- After convergence it will therefore ensure that, despite the non stationary disturbance, the mean of $\hat{y}$ will be equal to the mean of y
  - This is the stochastic equivalent for a deterministic observer of having $\hat{y} = y$ in steady–state for a constant disturbance
- In the MIMO case, the model can be [augmented as with the Luenberger observer]()

ANNEX A – KALMAN FILTER

Annex A – Kalman filter

Prediction form – proof

Estimation error:

$$
\begin{aligned}
\varepsilon(k+1) &= x(k+1) - \hat{x}(k+1) \\
&= Ax(k) + Bu(k) + w(k) - A\hat{x}(k) - Bu(k) - K(k)\left[y(k) - C\hat{x}(k)\right] \\
&= A\left[x(k) - \hat{x}(k)\right] + w(k) - K(k)\left[Cx(k) + v(k) - C\hat{x}(k)\right] \\
&= A\varepsilon(k) + w(k) - K(k)\left[C\varepsilon(k) + v(k)\right]
\end{aligned}
$$

Cost function to minimize:

$$
\begin{aligned}
J &= \mathrm{tr}\left[E\left\{ \varepsilon(k+1)\varepsilon^T(k+1) \right\} \right] \\
&= \mathrm{tr}\left[E\left\{ \left(A\varepsilon(k) + w(k) - K(k)\left[C\varepsilon(k) + v(k)\right] \right) \left(A\varepsilon(k) + w(k) - K(k)\left[C\varepsilon(k) + v(k)\right] \right)^T \right\} \right] \\
&= \mathrm{tr}\left[E\left\{ \begin{array}{l} A\varepsilon(k)\varepsilon^T(k)A^T + w(k)w^T(k) + K(k)\left[C\varepsilon(k) + v(k)\right]\left[C\varepsilon(k) + v(k)\right]^T K^T(k) \\ -A\varepsilon(k)\left[C\varepsilon(k) + v(k)\right]^T K^T(k) - K(k)\left[C\varepsilon(k) + v(k)\right]\varepsilon^T(k)A^T \end{array} \right\} \right]
\end{aligned}
$$

Annex A – Kalman filter

$$J = \mathrm{tr}\left[E\left\{ \begin{array}{l} A\varepsilon(k)\varepsilon^T(k)A^T + w(k)w^T(k) + K(k)\left[C\varepsilon(k)+v(k)\right]\left[C\varepsilon(k)+v(k)\right]^T K^T(k) \\ -A\varepsilon(k)\left[C\varepsilon(k)+v(k)\right]^T K^T(k) - K(k)\left[C\varepsilon(k)+v(k)\right]\varepsilon^T(k)A^T \end{array} \right\} \right]$$

$$= \mathrm{tr}\left[E\left\{ \begin{array}{l} A\varepsilon(k)\varepsilon^T(k)A^T + w(k)w^T(k) + K(k)\left[C\varepsilon(k)\varepsilon^T(k)C^T + v(k)v^T(k)\right]K^T(k) \\ -A\varepsilon(k)\varepsilon^T(k)C^T K^T(k) - K(k)C\varepsilon(k)\varepsilon^T(k)A^T \end{array} \right\} \right]$$

Since $P(k+1/k) = E\{\varepsilon(k+1)\varepsilon^T(k+1)\}$, $Q = E\{w(k)w^T(k)\}$ and $R = E\{v(k)v^T(k)\}$:

$$J = \mathrm{tr}\left[P(k+1/k)\right]$$

$$= \mathrm{tr}\left[\begin{array}{l} AP(k/k-1)A^T + Q + K(k)\left[CP(k/k-1)C^T + R\right]K^T(k) \\ -AP(k/k-1)C^T K^T(k) - K(k)CP(k/k-1)A^T \end{array} \right] \qquad (A.1)$$

Annex A – Kalman filter

Derivative of the trace of some matrices:
$$\frac{d}{dK(k)}\operatorname{tr}\left[K(k)A\right]=A^H$$
$$\frac{d}{dK(k)}\operatorname{tr}\left[AK^H(k)\right]=A$$
$$\frac{d}{dK(k)}\operatorname{tr}\left[K(k)AK^H(k)\right]=2K(k)A$$

where ^{H} is the conjugate transpose for complex matrices. In our case, they are real, thus ^{H} becomes ^{T}:, The minimum of the cost function is found with:

$$\frac{d}{K(k)}\operatorname{tr}\begin{bmatrix}AP(k/k-1)A^T+Q+K(k)\left[CP(k/k-1)C^T+R\right]K^T(k)\\-AP(k/k-1)C^TK^T(k)-K(k)CP(k/k-1)A^T\end{bmatrix}=$$
$$=2K(k)\left[CP(k/k-1)C^T+R\right]-AP(k/k-1)C^T-AP(k/k-1)C^T=0$$

which leads to: $\boxed{K(k)=AP(k/k-1)C^T\left[CP(k/k-1)C^T+R\right]^{-1}}$

Annex A – Kalman filter

From A.1: $P(k+1/k) = AP(k/k-1)A^T + Q + K(k)\left[CP(k/k-1)C^T + R\right]K^T(k)$

$$- AP(k/k-1)C^T K^T(k) - K(k)CP(k/k-1)A^T$$

Inserting $K(k) = AP(k/k-1)C^T\left[CP(k/k-1)C^T + R\right]^{-1}$ into the previous result gives (note that $P^T = P$ and $(CPC^T + R)^T = CPC^T + R$):

$$\begin{aligned}
P(k+1/k) &= AP(k/k-1)A^T + Q \\
&\quad + \left[\begin{array}{l} AP(k/k-1)C^T\left[CP(k/k-1)C^T + R\right]^{-1}\left[CP(k/k-1)C^T + R\right] \\ \times\left[CP(k/k-1)C^T + R\right]^{-1} CP(k/k-1)A^T \end{array}\right] \\
&\quad - AP(k/k-1)C^T\left[CP(k/k-1)C^T + R\right]^{-1} CP(k/k-1)A^T \\
&\quad - AP(k/k-1)C^T\left[CP(k/k-1)C^T + R\right]CP(k/k-1)A^T \\
&= AP(k/k-1)A^T + Q - AP(k/k-1)C^T\left[CP(k/k-1)C^T + R\right]^{-1} CP(k/k-1)A^T
\end{aligned}$$

Finally:

$$\boxed{P(k+1/k) = A\left[P(k/k-1) - P(k/k-1)C^T\left(CP(k/k-1)C^T + R\right)^{-1} CP(k/k-1)\right]A^T + Q}$$

Annex A – Kalman filter

Filtering form – proof

Let suppose that at time k the best possible estimate $\hat{x}(k/k)$ is available, then, since $w(k)$ is a random noise with a zero mean, the best prediction is:

$$\hat{x}(k+1/k) = A\hat{x}(k/k) + Bu(k)$$

and it must be equal to:

$$\hat{x}(k+1/k) = A\hat{x}(k/k-1) + Bu(k) + K(k)\left[y(k) - C\hat{x}(k/k-1)\right]$$

Therefore:

$$\hat{x}(k/k) = \hat{x}(k/k-1) + A^{-1}K(k)\left[y(k) - C\hat{x}(k/k-1)\right]$$

If we define $K_f(k) = A^{-1}K(k)$ then:

$$\boxed{\begin{aligned} K_f(k) &= P(k/k-1)C^T\left[CP(k/k-1)C^T + R\right]^{-1} \\ \hat{x}(k/k) &= \hat{x}(k/k-1) + K_f(k)\left[y(k) - C\hat{x}(k/k-1)\right] \end{aligned}}$$