ETH zürich !!! Apéro and Pizza available !!! More details at control.ee.ethz.ch

AUTOMATIC CONTROL LABORATORY

OPEN HOUSE

Automatic Control Laboratory

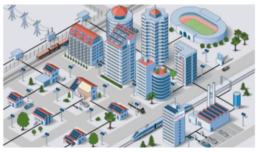
Prof. F. Dörfler

Prof. J. Lygeros

Prof. R. Smith

Tuesday, 28. Nov. 2023

17:15, ETF E1

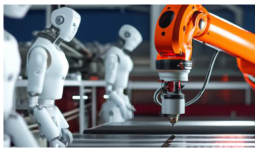


Control of energy systems and smart grids

Control: a world of data

Reinforcement learning and optimal control

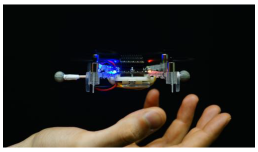
Game-theoretical control of socio-technical systems



Industrial control and robotics

Optimization, games, and learning in closed loop

RoboCup – NomadZ team



Experimental validation of control methods

System Identification

Lecture 9: Prediction error methods & ARX models

Roy Smith

Identification framework

Measurement data

Define the measurement data set:

$$Z_K = \{u(0), y(0), \ldots, u(K-1), y(K-1)\}.$$

Objective

Define the cost function (includes the model structure)

$$J(\theta, Z_K)$$

General optimisation formulation

$$\underset{\theta}{\text{minimise}}\, J(\theta, Z_K) \quad \text{or} \quad \hat{\theta} = \text{argmin}\, J(\theta, Z_K).$$

Optimisation objectives

Residual error objectives

error: $e(k, \theta) = y(k) - G(\theta)u(k).$

$J(\theta) = \|e(\theta)\|_2^2$ or $\|e(\theta)\|_\infty$, or $\|e(\theta)\|_1$?

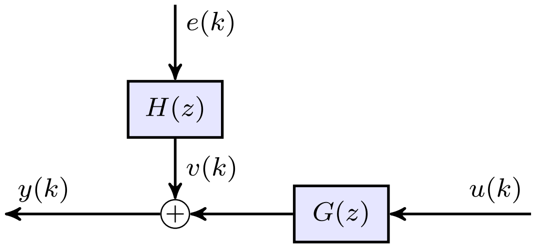
Parametric error objectives

$J(\theta) = \|\theta - \theta_0\|_2$ or $E\{\theta - \theta_0\}$

Prediction error objective

$$J(\theta) = E\{y(k+1) - \hat{y}(k+1, \theta|k)\}$$

Prediction



Typical assumptions

$G(z)$ and $H(z)$ are stable,

$H(z)$ is stably invertible (no zeros outside the unit disk)

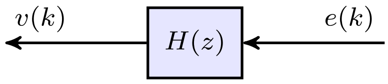
$e(k)$ has known statistics: known pdf or known moments.

One-step ahead prediction

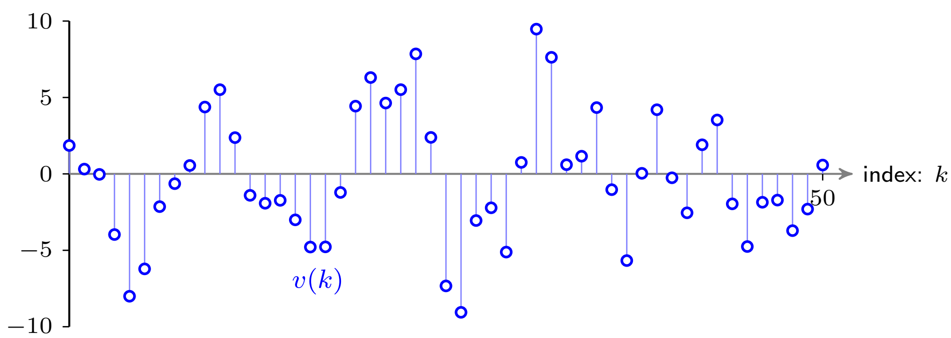
Given $Z_K = \{u(0), y(0), \ldots, u(K-1), y(K-1)\}$,

what is the best estimate of $y(K)$?

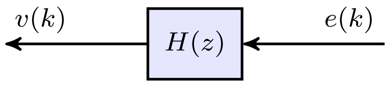
Prediction



What is the next output, $v(k)$?



Prediction



Noise model invertibility

Given, $v(k)$, $k = 0, \ldots, K-1$, can we determine $e(k)$, $k = 0, \ldots, K-1$?

Inverse filter: $H_{\mathsf{inv}}(z)$: $$e(k) = \sum_{i=0}^{\infty} h_{\mathsf{inv}}(i) v(k-i)$$

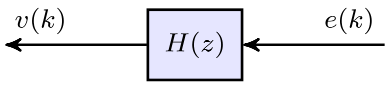
We also want the inverse filter to be causal and stable:

$$h_{\mathsf{inv}}(k) = 0, k < 0, \quad \text{and} \quad \sum_{k=0}^{\infty} |h_{\mathsf{inv}}(k)| < \infty.$$

If $H(z)$ has no zeros for $|z| \geqslant 1$, then,

$$H_{\mathsf{inv}}(z) = \frac{1}{H(z)}.$$

Prediction



One step ahead prediction

Given measurements of $v(k)$, $k = 0, \ldots, K-1$, can we predict $v(K)$?

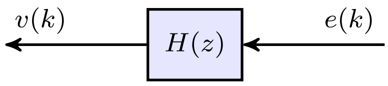
Assume that we know $H(z)$, how much can we say about $v(K)$?

Assume also that $H(z)$ is monic ($h(0) = 1$).

$$v(k) = \sum_{i=0}^{\infty} h(i) e(k-i)$$

$$= e(k) + \underbrace{\sum_{i=1}^{\infty} h(i) e(k-i)}_{\substack{= m(k-1) \\ \text{"observed"}}}$$

Prediction



One-step ahead prediction

The prediction of $v(k)$, based on measurements up to time $k-1$ is,

$$\hat{v}(k|k-1).$$

We will argue that a good choice in this case is,

$$\hat{v}(k|k-1) = m(k-1) = \sum_{i=1}^{\infty} h(i)e(k-i).$$

The error in our prediction is $e(k)$ — which we clearly can't reduce.

One-step ahead prediction statistics

General case

Say $e(k)$ is i.i.d. with probability density function: $f_e(x)$,

A posteriori distribution

What are the statistics of $v(k)$ given $v_{-\infty}^{k-1} = \{v(-\infty), \ldots, v(k-1)\}$?

$$v(k|k-1) \;=\; e(k) \;+\; m(k-1),$$

so $f_{v(k|k-1)}(x) \;=\; f_e(v(k) - m(k-1)).$ $\longleftarrow$ $e(k)$ density shifted by $m(k-1)$.

Choosing an estimate: the expected value

$$\hat{v}(k|k-1) \;=\; E\{v(k|k-1)\} \;=\; m(k-1) + E\{e(k)\} \;=\; m(k-1).$$

One-step ahead prediction

Calculation

$$\hat{v}(k|k-1) = m(k-1) = \sum_{i=1}^{\infty} h(i)e(k-i)$$
$$= (H(z)-1)\,e(k) \qquad \text{(assuming } H(z) \text{ is monic)}$$
$$= \frac{H(z)-1}{H(z)} v(k)$$
$$= (1 - H_{\text{inv}}(z))\,v(k)$$
$$= -\sum_{i=1}^{\infty} h_{\text{inv}}(i)v(k-i)$$

Note that $\hat{v}(k|k-1)$ depends only on values up to time $k-1$.

The best we can do is:

$$\hat{v}(k|k-1) = -\sum_{i=1}^{k} h_{\text{inv}}(i)v(k-i) \approx -\sum_{i=1}^{\infty} h_{\text{inv}}(i)v(k-i).$$

Example

Moving average model

$$v(k) = e(k) + ce(k-1), \quad \Longrightarrow \quad H(z) = 1 + cz^{-1}.$$

For $H(z)$ to be stably invertible we require $|c| < 1$.

$$H_{\text{inv}}(z) = \frac{1}{1+cz^{-1}} = \sum_{i=0}^{\infty}(-c)^i z^{-i}.$$

One-step ahead predictor

$$\hat{v}(k|k-1) = (1 - H_{\text{inv}}(z))v(k) = -\sum_{i=1}^{\infty}(-c)^i v(k-i) \approx -\sum_{i=1}^{k}(-c)^i v(k-i)$$
$$= cv(k-1) - c^2 v(k-2) + c^3 v(k-3) + \cdots - (-c)^k v(0).$$

Exercise: show $\hat{v}(k|k-1) = c \underbrace{(v(k-1) - \hat{v}(k-1|k-2))}_{\epsilon(k-1)}$ (prediction error at $k-1$)

$$= c\epsilon(k-1)$$

Another example

Autoregressive noise model

Our noise model is:

$$v(k) = \sum_{i=0}^{\infty} a^i e(k-i) \qquad |a| < 1 \text{ for stability.}$$

So, $H(z) = \sum_{i=0}^{\infty} a^i z^{-i} = \frac{1}{1 - az^{-1}}$,

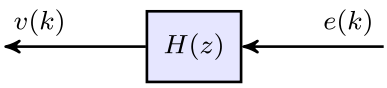
and $H_{\mathsf{inv}}(z) = 1 - az^{-1}$ (a moving average process)

Our one-step ahead predictor is,

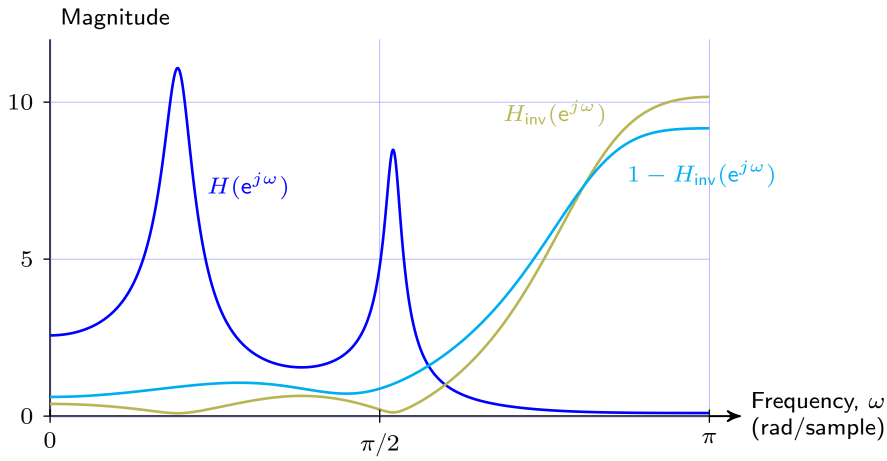
$$\hat{v}(k|k-1) = (1 - H_{\mathsf{inv}}(z))v(k) = av(k-1).$$

A more complicated example

Noise model

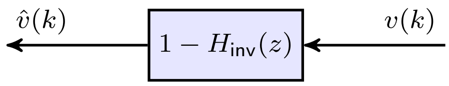


$$H(z) = \frac{z^4 + 0.5z^3 + 0.1z^2 + 0.1z + 0.1}{z^4 - 1.4z^3 + 1.6z^2 - 1.3z + 0.8}$$

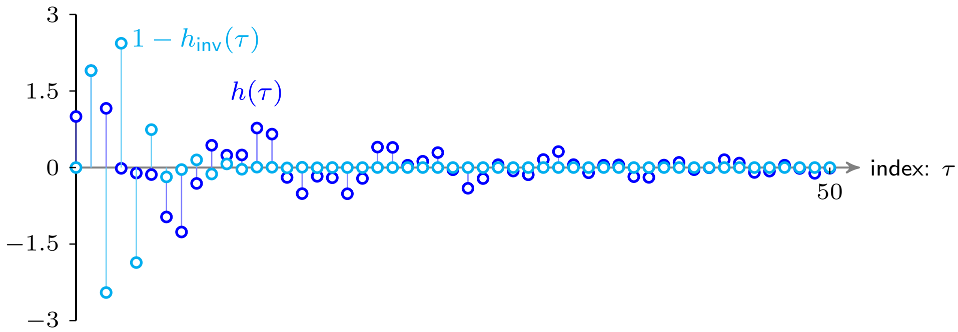


A more complicated example

Predictor

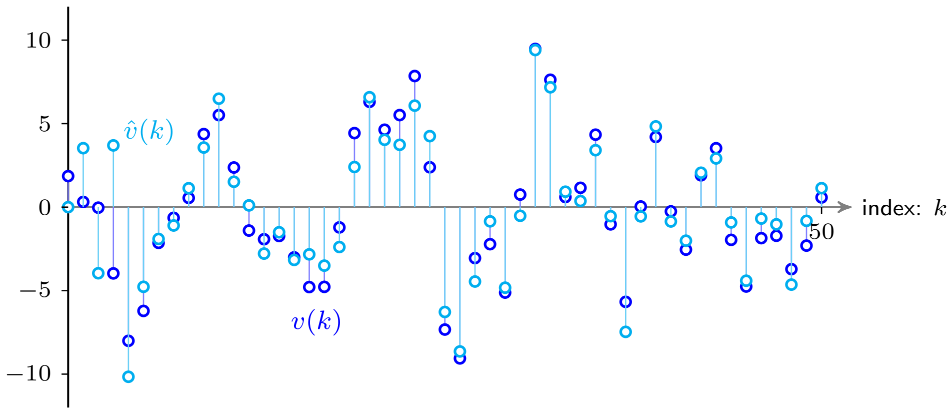


Impulse responses: $H(z)$ and $1 - H_{\mathsf{inv}}(z)$



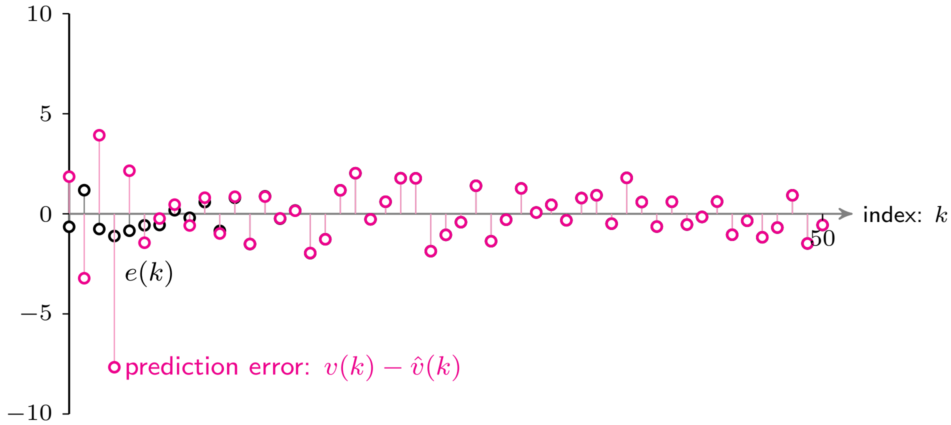
A more complicated example

One-step ahead predictions



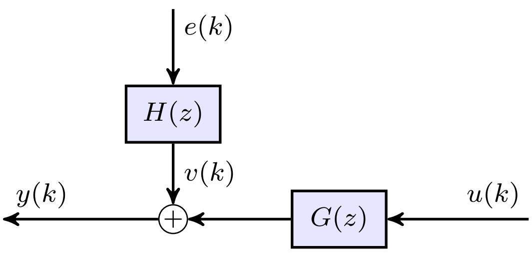
A more complicated example

One-step ahead prediction errors



Output prediction

$$y(k) = G(z)u(k) + v(k)$$



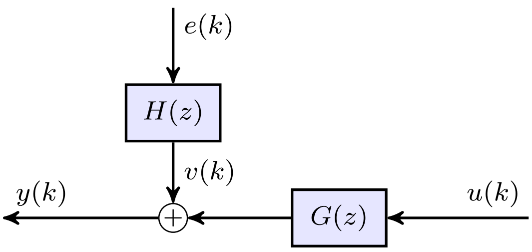
One-step ahead prediction

Choose the expected value of the conditional distribution,

$$\begin{aligned}\hat{y}(k|k-1) &= E\{y(k|k-1)\} = G(z)u(k) + \hat{v}(k|k-1) \\ &= G(z)u(k) + (1 - H_{\text{inv}}(z))v(k) \\ &= H_{\text{inv}}(z)G(z)u(k) + (1 - H_{\text{inv}}(z))y(k)\end{aligned}$$

Output prediction

$$y(k) = G(z)u(k) + v(k)$$



Prediction error

$$\begin{aligned} y(k) - \hat{y}(k|k-1) &= -H_{\mathsf{inv}}(z)G(z)u(k) + H_{\mathsf{inv}}(z)y(k) \\ &= H_{\mathsf{inv}}(z)(y(k) - G(z)u(k)) = H_{\mathsf{inv}}(z)v(k) \\ &= e(k) \end{aligned}$$

The innovation is the part of the output prediction that cannot be estimated from past measurements.

Prediction error based identification

Prediction error

The one-step ahead predictor is parametrised by θ,

$$\hat{y}(k|\theta, Z_K) = H_{\mathsf{inv}}(\theta, z)G(\theta, z)u(k) + (1 - H_{\mathsf{inv}}(\theta, z))y(k)$$

Define a parametrised prediction error,

$$\epsilon(k, \theta) \;=\; y(k) - \hat{y}(k, \theta) \;=\; -H_{\mathsf{inv}}(\theta, z)G(\theta, z)u(k) \;+\; H_{\mathsf{inv}}(\theta, z)y(k).$$

Prediction error-based identification

Typical cost function: $\quad J(\theta, Z_K) \;=\; \|\epsilon(k, \theta)\|_2.$

Pick θ to minimise the error.

Optimisation formulation: $\quad \hat{\theta} \;=\; \underset{\theta}{\operatorname{argmin}} \; J(\theta, Z_K).$

ARX and ARMAX models

Autoregressive moving average models

Model form (without noise):

$$G(z) = \frac{b_1 z^{-1} + \cdots + b_m z^{-m}}{1 + a_1 z^{-1} + \cdots + a_n z^{-n}}.$$

Input output reationship:

$$\begin{aligned} y(k) &= G(z)u(k) \\ &= -a_1 y(k-1) - \cdots - a_n y(k-n) + b_1 u(k-1) + \cdots + b_m u(k-m) \\ &= \varphi^T(k)\theta \end{aligned}$$

where

$$\varphi(k) = \begin{bmatrix} -y(k-1) & \cdots & -y(k-n) & u(k-1) & \cdots & u(k-m) \end{bmatrix}^T$$

$$\theta = \begin{bmatrix} a_1 & \cdots & a_n & b_1 & \cdots & b_m \end{bmatrix}^T.$$

ARX and ARMAX models

Autoregressive moving average models

$$y(k) = G(\theta, z)u(k) = \varphi^T(k)\theta.$$

where $\varphi(k)$ is called the regressor vector
θ is the parameter vector.

Because we can calculate $y(k)$ from the model, with knowledge of past $y(k)$, $u(k)$ and model parameters, θ, the output $y(k)$ is sometimes written as,

$$y(k) = \varphi^T(k)\theta = y(k|\theta),$$

to emphasize the parameter dependence.

ARX and ARMAX models

Regressor formulation

Least squares estimation

Model framework (ARX or ARMAX)

Consider $y(k) = \varphi^T(k)\theta$ for $k = 0, \ldots, N-1$.

$$\underbrace{\begin{bmatrix} y(0) \\ \vdots \\ y(N-1) \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} \varphi^T(0) \\ \vdots \\ \varphi^T(N-1) \end{bmatrix}}_{\Phi} \theta.$$

or, in matrix form, $Y = \Phi\theta$

If $N > (n+m)$ this system is overdetermined.

Least squares solution

$$\hat{\theta} = \left(\Phi^T\Phi\right)^{-1}\Phi^T Y$$ (MATLAB: `Theta = Phi\Y;`)

Least squares estimation

Optimisation framework

The least squares solution, $\hat{\theta} = \left(\Phi^T \Phi\right)^{-1} \Phi^T Y$, solves the problem,

$$\begin{aligned} \underset{\theta}{\text{minimise}} \quad & \|\epsilon\|_2 \\ \text{subject to} \quad & Y = \Phi\theta + \epsilon \end{aligned}$$

Define the cost function: $J(\theta, Z_N) = \|Y - \Phi\theta\|_2 = \|\epsilon\|_2$

General formulation:

$$\begin{aligned} \hat{\theta} &= \underset{\theta}{\operatorname{argmin}}\, J(\theta, Z_N) \\ &= \underset{\theta}{\operatorname{argmin}}\, \|\epsilon\|_2 \\ & \quad \text{subject to}: \quad \epsilon = Y - \Phi\theta. \end{aligned}$$

Example

Single state system: 2 parameters

System: $G(z, \theta) = \dfrac{bz^{-1}}{(1 + az^{-1})} = \dfrac{b}{(z + a)}$ where $\theta = \begin{bmatrix} a \\ b \end{bmatrix}$.

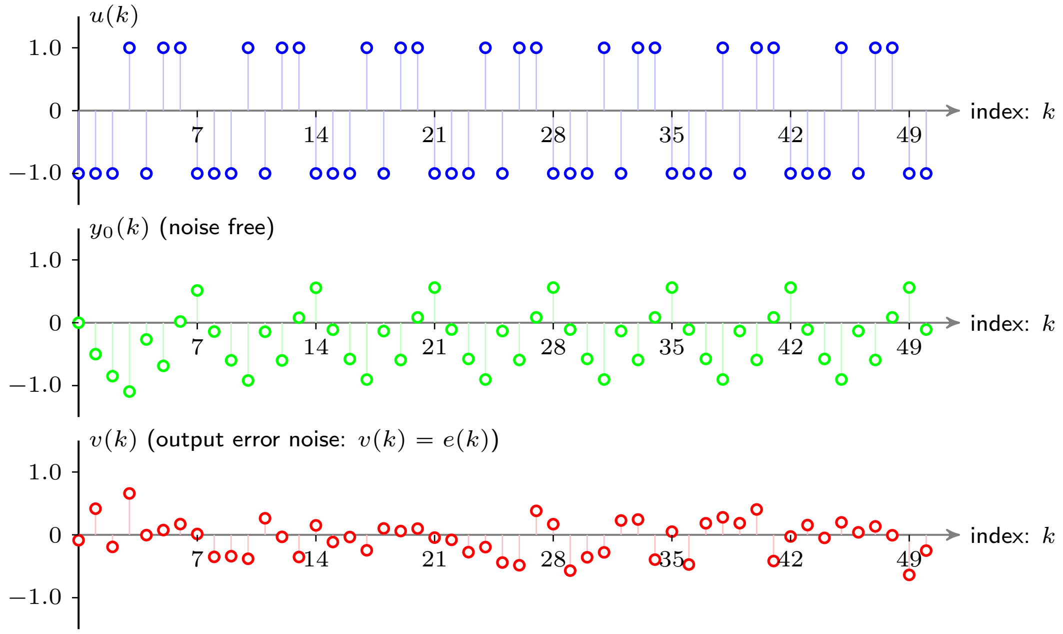
Difference equation: $y(k) = -ay(k-1) + bu(k-1)$.

Least squares solution:

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \left(\Phi^T \Phi\right)^{-1} \Phi^T Y$$

Example

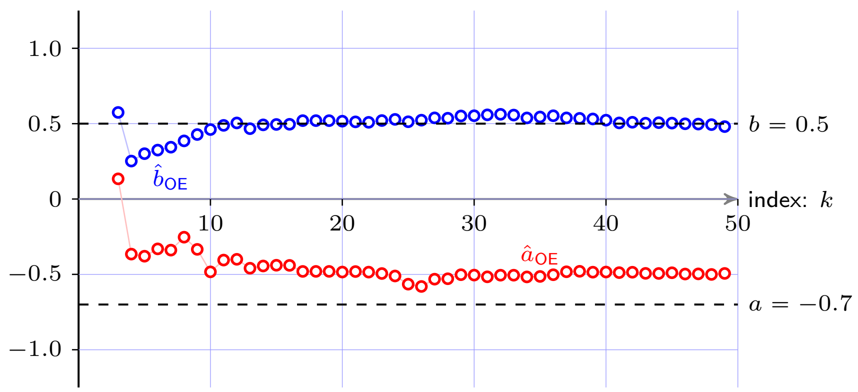
Output error experiment



$\mathrm{var}(e(k)) = 0.25^2$

Example

Least-squares estimates: output error structure



True system

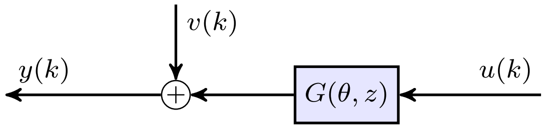
$$G(z) = \frac{bz^{-1}}{1 + az^{-1}}, \quad H(z) = 1.$$

Estimation:

$$y(k) = -\hat{a}_{\mathsf{OE}} y(k-1) + \hat{b}_{\mathsf{OE}} u(k-1)$$

Error statistics and noise statistics

Experimental configuration: output noise



Estimation framework

$$\begin{bmatrix} y(0) \\ \vdots \\ y(N-1) \end{bmatrix} = \begin{bmatrix} \varphi^T(0) \\ \vdots \\ \varphi^T(N-1) \end{bmatrix} \theta + \begin{bmatrix} v(0) \\ \vdots \\ v(N-1) \end{bmatrix},$$

But this was worked out for the noise-free case. So we really have,

$$\varphi_0^T(k) = \begin{bmatrix} -y_0(k-1) & \cdots & -y_0(k-n) & u(k-1) & \cdots & u(k-m) \end{bmatrix}$$

And by stacking,

$Y = \Phi_0 \, \theta + V.$ What happens if we use $Y = \Phi \, \theta + V$?

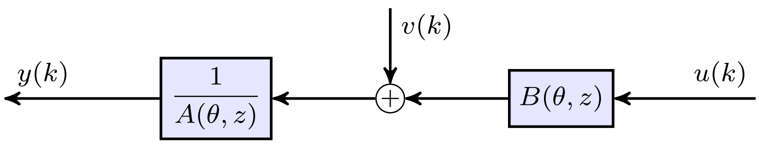
Equation error models

Model form

$$y(k) + a_1 y(k-1) + \cdots + a_n y(k-n) = b_1 u(k-1) + \cdots + b_m u(k-m) + v(k)$$

$$A(z)Y(z) = B(z)U(z) + V(z)$$

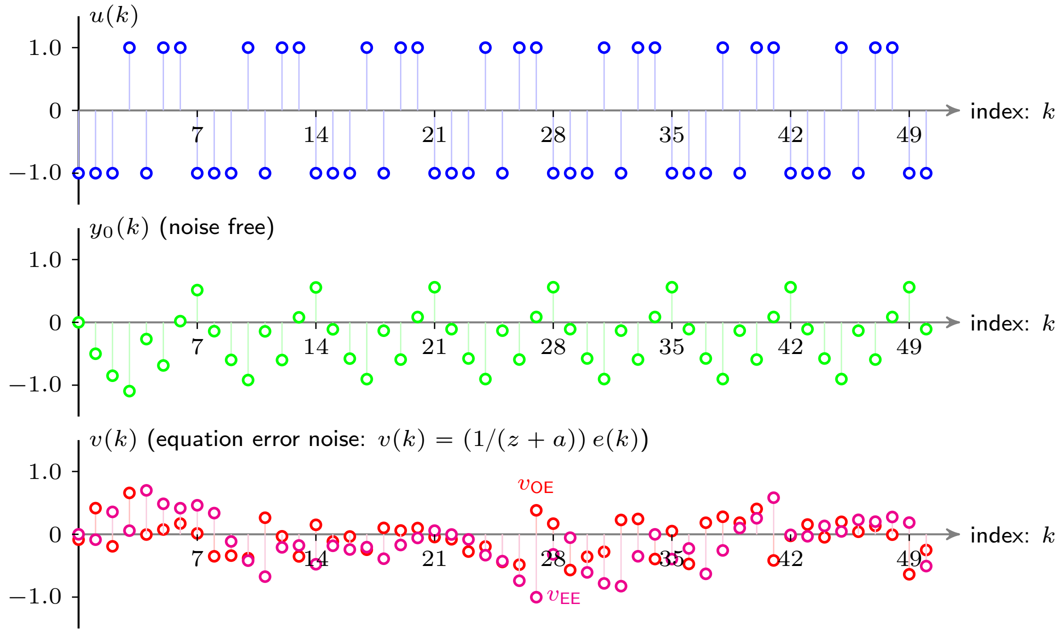
Now we have: $Y = \Phi\theta + V.$



The noise also has the autoregressive dynamics (in $A(z)$).

Example

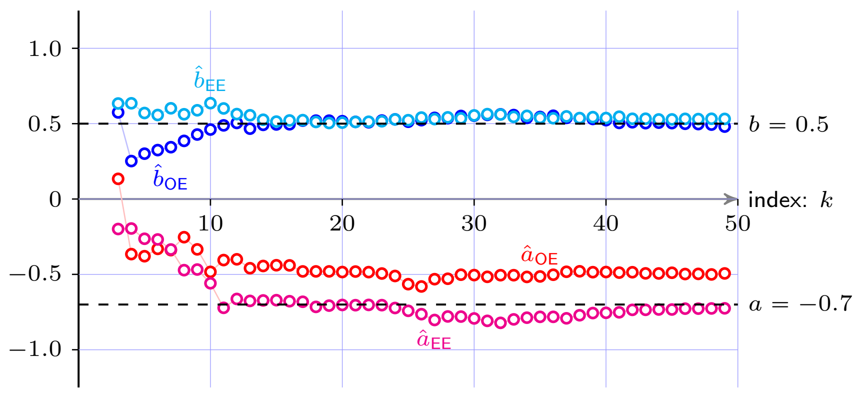
Equation error experiment



$\mathrm{var}(e(k)) = 0.25^2$

Example

Least-squares estimates: equation error structure



True system

$$G(z) = \frac{bz^{-1}}{1 + az^{-1}}, \quad H(z) = \frac{1}{1 + az^{-1}}.$$

Estimation:

$$y(k) \;=\; -\hat{a}_{\mathsf{EE}} y(k-1) \;+\; \hat{b}_{\mathsf{EE}} u(k-1)$$

Output error and Equation error models

Output error framework

Equation error framework

Bibliography

Prediction
Lennart Ljung, *System Identification;Theory for the User*, 2nd Ed., Prentice-Hall, 1999, [section 3.2].

ARX models
Lennart Ljung, *System Identification;Theory for the User*, 2nd Ed., Prentice-Hall, 1999, [sections 1.3 and 4.2].