Stochastic Adaptive Control (02421)

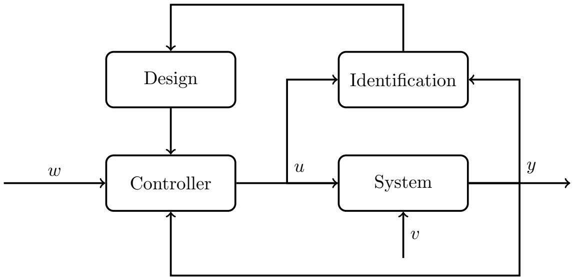
Lecture 10

Tobias K. S. Ritschel

Section for Dynamical Systems, DTU Compute

Lecture Plan

1. Systems theory
2. Stochastics
3. State estimation - Kalman filter 1
4. State estimation - Kalman filter 2
5. Optimal control 1 - internal models
6. External models
7. Prediction + optimal control 1 - external models
8. Optimal control 2 - external models
9. System identification 1
10. **System identification 2**
11. System identification 3 + model validation
12. Adaptive control 1
13. Adaptive control 2



Today's Agenda

- Follow-up from last lecture
- Estimation - State space models: EKF estimation
- Estimation - MIMO estimation
- Estimation - Recursive estimation
- Estimation - Time-varying estimation

Follow-up from last time, Question 3.1-3.4

We are considering the system

$$y_t = \frac{B}{A} u_{t-1} + e_t, \quad e \in N_{iid}(0, 0.1888) \tag{1}$$

$$A(q^{-1}) = 1 - 1.5q^{-1} + 0.7q^{-2}, \quad B(q^{-1}) = 1 + 0.5q^{-1} \tag{2}$$

and we are asked to estimate it using and ARX, OE and IV model estimation, and a PRBS signal.

First we need to choose the noise variance (hint: use trfvar or trfvar2):

$$\sigma_e^2 = \sigma_{yu}^2/100, \quad \sigma_{yu}^2 = Var\left(\frac{q^{-1}B}{A}\right) Var(Prbs) \simeq Var\left(\frac{q^{-1}B}{A}\right) \tag{3}$$

Using a least-squares approach for an ARX-model, arx(data,[2,2,1]):

$$a_1 = -1.3171, \qquad CI_{a_1} = [-1.4430, \quad -1.1911] \tag{4}$$

$$a_2 = 0.5231, \qquad CI_{a_2} = [0.3991, \quad 0.6472] \tag{5}$$

$$b_0 = 0.9360, \qquad CI_{b_0} = [0.7237, \quad 1.1484] \tag{6}$$

$$b_1 = 0.7005, \qquad CI_{b_1} = [0.4593, \quad 0.9417] \tag{7}$$

For a_1 and a_2, the confidence intervals do not include the true values.

Follow-up from last time, Question 3.1-3.4

Using the IV approach, iv4(data,[2,2,1]), we get

$$a_1 = -1.4983, \qquad CI_{a_1} = [-1.5247, \quad -1.4718] \tag{8}$$
$$a_2 = 0.6999, \qquad CI_{a_2} = [0.6767, \quad 0.7231] \tag{9}$$
$$b_0 = 0.9905, \qquad CI_{b_0} = [0.8692, \quad 1.1118] \tag{10}$$
$$b_1 = 0.5297, \qquad CI_{b_1} = [0.4383, \quad 0.6211] \tag{11}$$

Our confidence intervals now cover the actual parameters.
And using the OE approach, oe(data, [2, 2, 1])

$$a_1 = -1.4977, \qquad CI_{a_1} = [-1.5181, \quad -1.4774] \tag{12}$$
$$a_2 = 0.6988, \qquad CI_{a_2} = [0.6821, \quad 0.7156] \tag{13}$$
$$b_0 = 0.9997, \qquad CI_{b_0} = [0.9070, \quad 1.0923] \tag{14}$$
$$b_1 = 0.5174, \qquad CI_{b_1} = [0.3897, \quad 0.6451] \tag{15}$$

Again, our confidence intervals cover the actual parameters.

Follow-up from last time

Questions?

State-Space estimation

Instead of identifying an external model, we might want to identify an internal model. For example:

$$x_{t+1} = A(\theta)x_t + B(\theta)u_t \tag{16}$$

$$y_t = C(\theta)x_k + D(\theta)u_t + e_t \tag{17}$$

$$x_0 = m_0(\theta) \tag{18}$$

Depending on the estimation approach, there are different types of parameters that can be handled.

State-Space estimation

The fully parametrized linear state space model is

$$x_{t+1} = A(\theta)x_t + B(\theta)u_t + v_t, \quad v_t \in \mathbb{F}(0, R_v(\theta)) \tag{19}$$

$$y_t = C(\theta)x_t + D(\theta)u_t + e_t, \quad e_t \in \mathbb{F}(0, R_e(\theta)) \tag{20}$$

$$x_0 \in \mathbb{F}(m_0(\theta), P_0(\theta)) \tag{21}$$

If we apply LS or ML estimation, we still use the same formulation based on the measurements:

$$\text{LS: } J = \sum_{t=1}^{N} \epsilon_t^2 = \sum_{t=1}^{N} (y_t - \hat{y}_t)^2 \tag{22}$$

$$\text{ML: } J = -\sum_{t=1}^{N} \log(f(y_t | Y_{t-1,\theta})) \tag{23}$$

$$y_t | Y_{t-1,\theta} \in N(\hat{y}_{t|t-1}, Q_{t|t-1}) \tag{24}$$

though the estimation might not be linear in the parameters.

Extended Kalman Filter

We can also use a Kalman filter for the estimation. Consider the linear system

$$x_{t+1} = A(\theta_t)x_t + B(\theta_t)u_t + v_t, \quad v_t \in \mathbb{F}(0, \Sigma_v) \tag{25}$$

$$\theta_{t+1} = \theta_t + \eta_t, \quad \eta_t \in \mathbb{F}(0, \Sigma_\eta) \tag{26}$$

$$y_t = C(\theta_t)x_t + D(\theta_t)u_t + e_t, \quad e_t \in \mathbb{F}(0, \Sigma_e), \quad Cov(v_t, e_t) = \Sigma_{ve} \tag{27}$$

$$v_t, \eta_t, e_t \text{ white and } \eta_t \perp v_t, e_t \tag{28}$$

We can write it as a nonlinear model with an augmented state vector:

$$\begin{bmatrix} x_{t+1} \\ \theta_{t+1} \end{bmatrix} = \begin{bmatrix} f(x_t, \theta_t, u_t, v_t) \\ \theta_t + \eta_t \end{bmatrix} \tag{29}$$

$$y_t = g(x_t, \theta_t, u_t, e_t) \tag{30}$$

Extended Kalman filter

Next, we linearize to obtain a model in the standard form:

$$\begin{bmatrix} x \\ \theta \end{bmatrix}_{t+1} = A_l \begin{bmatrix} x \\ \theta \end{bmatrix}_t + B_l u_t + w_t, \quad w_t \in \mathbb{F}(0, R_1) \tag{31}$$

$$y_t = C_l \begin{bmatrix} x \\ \theta \end{bmatrix}_t + D_l u_t + \epsilon_t, \quad \epsilon_t \in \mathbb{F}(0, R_2), \quad Cov(w_t, \epsilon_t) = R_{12} \tag{32}$$

where the matrices are given by

$$A_l = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial \theta} \\ 0 & I \end{bmatrix}, \quad B_l = \begin{bmatrix} \frac{\partial f}{\partial u} \\ 0 \end{bmatrix}, \quad G_l = \begin{bmatrix} \frac{\partial f}{\partial v} & 0 \\ 0 & I \end{bmatrix} \tag{33}$$

$$C_l = \begin{bmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial \theta} \end{bmatrix}, \quad D_l = \frac{\partial g}{\partial u}, \quad H_l = \frac{\partial g}{\partial e} \tag{34}$$

$$R_1 = G_l \begin{bmatrix} \Sigma_v & 0 \\ 0 & \Sigma_\eta \end{bmatrix} G_l^T, \quad R_2 = H_l \Sigma_e H_l^T, \quad R_{12} = \begin{bmatrix} \frac{\partial f}{\partial v} \Sigma_{ve} \left(\frac{\partial g}{\partial e} \right)^T \\ 0 \end{bmatrix} \tag{35}$$

Extended Kalman filter

The extended Kalman filter is used to estimate both states *and* parameters:

$$\begin{bmatrix} \hat{x}_{t+1} \\ \hat{\theta}_{t+1} \end{bmatrix} = \begin{bmatrix} f(\hat{x}_t, \hat{\theta}_t, u_t) \\ \hat{\theta}_t \end{bmatrix} + K_t(y_t - g(\hat{x}_t, \hat{\theta}, u_t)) \tag{36}$$

$$K_t = (A_l P_t C_l^T + R_{12})(C_l P_t C_l^T + R_2)^{-1} \tag{37}$$

$$P_{t+1} = (A_l - K_t C_l) P_t A_l^T + R_1 - K_t R_{12}^T \tag{38}$$

The Jacobians used to form A_l, B_l, G_l, C_l, D_l, and H_l are evaluated at $\hat{x}_t$ and $\hat{\theta}_t$, i.e., the system matrices vary over time.

Can you think of any particular limitations of this approach?

Think about it for yourself for <u>one minute</u> and
then discuss with the person next to you for <u>one minute</u>.

Multiple inputs, multiple outputs (MIMO) estimation

Let us consider a system with multiple inputs and outputs

$$\mathbf{A}(q^{-1})y_t = \mathbf{B}(q^{-1})u_t + e_t \tag{39}$$

$$\mathbf{A}(q^{-1}) = I + A_1 q^{-1} + \cdots + A_{n_a} q^{-n_a} \tag{40}$$

$$\mathbf{B}(q^{-1}) = B_0 + B_1 q^{-1} + \cdots + B_{n_b} q^{-n_b} \tag{41}$$

$$A_i \in \mathbb{R}^{n_y \times n_y}, \quad B_i \in \mathbb{R}^{n_y \times n_u} \tag{42}$$

We can then write it on the matrix form as

$$y_t^T = \phi_t^T \theta + e_t^T \tag{43}$$

$$\theta = \begin{bmatrix} A_1 & A_2 & \cdots & A_{n_a} & B_0 & B_1 & \cdots & B_{n_b} \end{bmatrix}^T \tag{44}$$

$$\phi_t^T = \begin{bmatrix} -y_{t-1}^T & -y_{t-2}^T & \cdots & -y_{t-n_a}^T & u_t^T & u_{t-1}^T & \cdots & u_{t-n_b}^T \end{bmatrix} \tag{45}$$

Multi-input-Multi-Output estimation

The estimate is obtained by

$$\hat{\theta} = (\Phi^T \Phi)^{-1} \Phi^T Y \tag{46}$$

$$Y = \Phi\theta + E \tag{47}$$

$$Y = \begin{bmatrix} y_1^T \\ y_2^T \\ \vdots \\ y_N^T \end{bmatrix} \quad \Phi = \begin{bmatrix} \phi_1^T \\ \phi_2^T \\ \vdots \\ \phi_N^T \end{bmatrix} \quad E = \begin{bmatrix} e_1^T \\ e_2^T \\ \vdots \\ e_N^T \end{bmatrix} \tag{48}$$

Recursive Estimation

The previous methods are in the form

$$\hat{\theta}_t = func(Y_t) \tag{49}$$

That is, we use all measurements up to and including time t. Over time, that becomes computationally intensive.

In contrast, a recursive method only relies on the current measurement and the past estimate:

$$\hat{\theta}_t = func(y_t, \hat{\theta}_{t-1}) \tag{50}$$

This approach assumes that $\hat{\theta}_{t-1}$ is a sufficient statistic of Y_{t-1}. One advantage of this approach is that it can easily be adapted to account for time-varying parameters.

RLS/RARX - Recursive Least Squares

If our system is an ARX model:

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + e_t, \quad e_t \in \mathcal{F}(0, \sigma^2) \tag{51}$$

$$y_t = \phi_t^T \theta + e_t, \quad e_t \perp e_s \quad s > t \tag{52}$$

$$\phi_t = [-y_{t-1}, \dots, -y_{t-n_a}, u_{t-k}, \dots, u_{t-n_b-k}]^T \tag{53}$$

$$\theta = [a_1, \dots, a_{n_a}, b_0, \dots, b_{n_b}]^T \tag{54}$$

For a least squares approach based on t measurements, the estimator is

$$\hat{\theta}_t = \left(\sum_{i=1}^{t} \phi_i \phi_i^T \right)^{-1} \sum_{i=1}^{t} \phi_i y_i \tag{55}$$

$$P_t^{-1} = \sum_{i=1}^{t} \phi_i \phi_i^T, \quad \sum_{i=1}^{t} \phi_i \epsilon_i = 0 \tag{56}$$

RLS/RARX - Recursive Least Squares

The recursive formulation is

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t \times \sum_{i=1}^{t} \phi_i \epsilon_i \tag{57}$$

This allows us write the recursion in a computationally suitable form:

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t \phi_t \epsilon_t \tag{58}$$

$$\epsilon_t = y_t - \phi_t^T \hat{\theta}_{t-1} \tag{59}$$

$$P_t^{-1} = P_{t-1}^{-1} + \phi_t \phi_t^T \tag{60}$$

$$Var(\hat{\theta}_t | Y_t) = P_t \sigma^2 \approx Var(\hat{\theta}_t) \tag{61}$$

If no a priori knowledge about the parameter values is available, this initial estimate is suitable

$$\hat{\theta}_0 = 0, \quad P_0 = \beta I, \quad \beta \gg 0 \tag{62}$$

RLS/RARX - Recursive Least Squares

The recursion can also be computed using alternative formulations. Inspired by the Hemes' inversion lemma and square-root/factorization algorithms, we can write it as

$$\epsilon_t = y_t - \phi_t^T \hat{\theta}_{t-1} \tag{63}$$

$$s_t = 1 + \phi_t^T P_{t-1} \phi_t \tag{64}$$

$$K_t = \frac{P_{t-1}\phi_t}{s_t} \tag{65}$$

$$\hat{\theta}_t = \hat{\theta}_{t-1} + K_t \epsilon_t \tag{66}$$

$$P_t = P_{t-1} - K_t s_t K_t^T \tag{67}$$

RELS - Recursive Extended Least Squares

If we consider the ARMAX structure

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + C(q^{-1})e_t \tag{68}$$

$$y_t = \phi_t^T\theta + e_t \tag{69}$$

$$\phi_t = [-y_{t-1}, \dots, -y_{t-n_a}, u_{t-k}, \dots, u_{t-n_b-k}, e_{t-1}, \dots, e_{t-n_c}]^T \tag{70}$$

$$\theta = [a_1, \dots, a_{n_a}, b_0, \dots, b_{n_b}, c_1, \dots, c_{n_c}]^T \tag{71}$$

As with the LS method, we formulate a recursive version of the extended LS method by estimating e_i as ϵ_i in ϕ:

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t\phi_t\epsilon_t \tag{72}$$

$$\epsilon_i = y_i - \phi_i^T\hat{\theta}_{i-1} \tag{73}$$

$$P_t^{-1} = P_{t-1}^{-1} + \phi_t\phi_t^T \tag{74}$$

RML - Recursive Maximum Likelihood

If we consider the ARMAX structures

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + C(q^{-1})e_t \tag{75}$$

$$y_t = \phi_t^T\theta + e_t \tag{76}$$

$$\phi_t = [-y_{t-1}, \dots, -y_{t-n_a}, u_{t-k}, \dots, u_{t-n_b-k}, e_{t-1}, \dots, e_{t-n_c}]^T \tag{77}$$

$$\theta = [a_1, \dots, a_{n_a}, b_0, \dots, b_{n_b}, c_1, \dots, c_{n_c}]^T \tag{78}$$

Using the same trick of replacement: estimating e_i as ϵ_i in ϕ; we can formulate the recursive maximum likelihood method:

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t\psi_t\epsilon_t, \quad \psi_t = \frac{1}{\hat{C}(q^{-1})}\phi_t \tag{79}$$

$$\epsilon_i = y_i - \phi_i^T\hat{\theta}_{i-1} \tag{80}$$

$$P_t^{-1} = P_{t-1}^{-1} + \psi_t\psi_t^T \tag{81}$$

Recursive methods for L-structures

Consider the L-Structure,

$$A(q^{-1})y_t = \frac{B(q^{-1})}{F(q^{-1})}u_t + \frac{C(q^{-1})}{D(q^{-1})}e_t + d \tag{82}$$

$$y_t = \phi_t^T\theta + e_t \tag{83}$$

$$\phi_t = [-y_{t-1}, \dots, u_t, \dots, -y^u_{t-1}, \dots, e_{t-1}, \dots, -y^e_{t-1}, \dots, 1]^T \tag{84}$$

$$\theta = [a_1, \dots, b_0, \dots, f_1, \dots, c_1, \dots, d_1, \dots, d]^T \tag{85}$$

We estimate the unknown regressors using our prior parameter estimate

$$\hat{y}^u_t = q^{-k}\frac{\hat{B}}{\hat{F}}u_t, \quad \hat{y}^e_t = \hat{A}y_t - \hat{y}^u_t - \hat{d}, \quad \epsilon_t = \hat{e}_t = \frac{\hat{D}}{\hat{C}}y^e_t \tag{86}$$

RPLR - Recursive Pseudo Linear Regression

For a PLR method (like ELS),

$$\phi_t = [-y_{t-1}, \ldots, u_t, \ldots, -\hat{y}^u_{t-1}, \ldots, \epsilon_{t-1}, \ldots, -\hat{y}^e_{t-1}, \ldots, 1]^T \quad (87)$$

$$\theta = [a_1, \ldots, b_0, \ldots, f_1, \ldots, c_1, \ldots, d_1, \ldots, d]^T \quad (88)$$

Then, the recursive algorithm is

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t \phi_t \epsilon_t \quad (89)$$

$$\epsilon_t = y_t - \phi_t^T \hat{\theta}_{t-1} \quad (90)$$

$$P_t^{-1} = P_{t-1}^{-1} + \phi_t \phi_t^T \quad (91)$$

RPEM - Recursive Prediction Error Method

if we instead consider a PEM algorithm (ML), we have that

$$\psi_t = [-\check{y}_{t-1}, \dots, \check{u}_t, \dots, -\check{y}^u_{t-1}, \dots, \check{e}_{t-1}, \dots, -\check{y}^e_{t-1}, \dots, \delta]^T \tag{92}$$

$$\theta = [a_1, \dots, b_0, \dots, f_1, \dots, c_1, \dots, d_1, \dots, d]^T \tag{93}$$

Where the estimated variables are given as

$$\check{y}_t = \frac{\hat{D}}{\hat{C}} y_t, \quad \check{u}_t = \frac{\hat{D}}{\hat{C}\hat{F}} u_t, \quad \check{y}^u_t = -\frac{\hat{D}}{\hat{C}\hat{F}} y^u_t \tag{94}$$

$$\check{e}_t = \frac{1}{\hat{C}} \epsilon_t, \quad \check{y}^e_t = -\frac{1}{\hat{C}} y^e_t, \quad \delta = \frac{\hat{D}}{\hat{C}} 1 \tag{95}$$

Then, the recursion is given by

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t \psi_t \epsilon_t \tag{96}$$

$$\epsilon_t = y_t - \psi_t^T \hat{\theta}_{t-1} \tag{97}$$

$$P_t^{-1} = P_{t-1}^{-1} + \psi_t \phi_t^T \tag{98}$$

STA Algorithm - Stochastic Approximation Algorithm

The above recursive algorithms are based on the Newton-Raphson method.

An alternative recursive algorithm is the STA or gradient algorithm:

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{r_t}\phi_t\epsilon_t \tag{99}$$

$$\epsilon_t = y_t - \phi_t^T\hat{\theta}_{t-1} \tag{100}$$

$$r_t = r_{t-1} + 1 \text{ or } r_t = r_{t-1} + \phi_t^T\phi_t \tag{101}$$

Time-varying estimation - first example

Let us consider the case where we have a time varying ARX model

$$A(q^{-1})y_t = B(t, q^{-1})u_t + e_t \tag{102}$$

$$b_1(t) = b_{1,0} + b_{1,1}t \tag{103}$$

We then treat the time-varying coefficient as two coefficients with their own inputs:

$$y_t = \phi^T \theta + e_t \tag{104}$$

$$\theta^T = \begin{bmatrix} a_1 & a_2 & \dots & a_{n_a} & b_{1,0} & b_{1,1} & b_2 & \dots & b_{n_b} \end{bmatrix} \tag{105}$$

$$\phi^T = \begin{bmatrix} -y_{t-1} & -y_{t-2} & \dots & -y_{t-n_a} & u_{t-1} & t \times u_{t-1} & u_{t-2} & \dots & u_{t-n_b} \end{bmatrix} \tag{106}$$

Similar approaches can be used for nonlinear time-varying coefficients.

Time-varying estimation

For deterministic time varying systems, we can rearrange the parameters as:

$$y_t = \phi_t^T \theta_t + e_t \tag{107}$$

$$\theta_t = \alpha + f(t)\beta = \begin{bmatrix} I & f(t) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{108}$$

$$y_t = \begin{bmatrix} \phi_t^T & \phi_t^T f(t) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + e_t \tag{109}$$

while for piece-wise linear parameters, we have

$$y_t = \phi_t^T \theta_t + e_t \tag{110}$$

$$\theta_t = \theta_{T_i} + (t - T_i)\alpha, \quad T_i \leq t \leq T_{i+1} \tag{111}$$

$$y_t = \begin{bmatrix} \phi_t^T & \phi_t^T (t - T_i) \end{bmatrix} \begin{bmatrix} \theta_{T_i} \\ \alpha \end{bmatrix} + e_t \tag{112}$$

But what do we do in the general case?

Time-varying systems

Let us now consider the case of systems with general time-varying parameters:

$$\theta_{t+1} = f(t, \theta_t, v_t) \tag{113}$$

The methods discussed so far cannot estimate the time-varying dynamics and were not designed to do it.

In practice, the problem is that the correction factor is diminishing as time goes on.

$$P_t \to 0 \tag{114}$$

Time-varying systems - Forgetting methods: Resetting

One approach is to restart the estimation after some time t_i:

$$P_{t_i} = P_i > P_{t_i-1}, \quad \hat{\theta}_{t_i} = \hat{\theta}_{t_i-1} \tag{115}$$

when to restart depends on the application.

An example is to restart at fixed intervals:

$$t_i = N * i \tag{116}$$

This can be useful for periodic systems.

Time-varying systems - Forgetting methods: Constant Gain

Another method is to simply keep the correction term large. One variant is to keep the correction term κ constant:

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \kappa \epsilon_t \tag{117}$$

$$\tilde{\theta}_t = (I - \kappa \phi_t^T)\tilde{\theta}_t - \kappa e_t \tag{118}$$

Alternatively, we can keep the variance constant:

$$P_t = P \tag{119}$$

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \kappa \epsilon_t \tag{120}$$

$$\kappa_t = \frac{P\phi_t}{1 + \phi_t^T P \phi_t} \tag{121}$$

Time-varying systems - Forgetting methods: Exponential Forgetfulness

A third method is to forget a little bit all the time. This is also known as exponential forgetfulness:

$$J_t = \frac{1}{2}\sum_{i=1}^{t} \lambda^{t-i}\epsilon_i^2 = \lambda J_{t-1} + \frac{1}{2}\epsilon_t^2 \tag{122}$$

The recursion is then similar to the previous methods:

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t\phi_t\epsilon_t \tag{123}$$

$$\epsilon_t = y_t - \phi_t^T\hat{\theta}_{t-1} \tag{124}$$

$$P_t^{-1} = \lambda P_{t-1}^{-1} + \phi_t\phi_t^T \tag{125}$$

The forgetting factor λ can be expressed in terms of a horizon, N_∞, which is roughly the period affecting the estimate

$$\lambda = 1 - \frac{1}{N_\infty} \tag{126}$$

This method relies on the system being sufficiently excited.

Time-varying systems - Fortescue's Method

We can improve the method by using a time-varying forgetting factor depending on the prediction error ϵ_t:

$$\lambda_t = 1 - \frac{1}{N_0} \times \frac{\epsilon_t^2}{\sigma^2 s_t} \tag{127}$$

where N_0 is the approx. horizon for which the parameter can be assumed to be constant.

The full recursion is then given as

$$\epsilon_t = y_t - \phi_t^T \hat{\theta}_{t-1} \tag{128}$$

$$s_t = 1 + \phi_t^T P_{t-1} \phi_t \tag{129}$$

$$K_t = \frac{P_{t-1}\phi_t}{\lambda_t + s_t} \tag{130}$$

$$\hat{\theta}_t = \hat{\theta}_{t-1} + K_t \epsilon_t \tag{131}$$

$$P_t = (I - K_t \phi_t^T) P_{t-1} \frac{1}{\lambda_t} \tag{132}$$

Time-varying systems - Fortescue's Method

If the variance is unknown, we can introduce an estimate r_t given by

$$\lambda_t = 1 - \frac{1}{N_0} \times \frac{\epsilon_t^2}{r_t s_t} \tag{133}$$

$$r_t = r_{t-1} + \frac{1}{t}\left(\frac{\epsilon_t^2}{s_t} - r_{t-1}\right), \quad r_0 = \epsilon_0^2 \tag{134}$$

Time-varying systems - Model Estimators

In these types of methods, a model of the parameters are introduced:

$$\theta_{t+1} = \theta_t + v_t, \qquad v_t \in N(0, R_1\sigma^2) \tag{135}$$

$$y_t = \phi_t^T \theta_t + e_t, \qquad e_t \in N(0, \sigma^2) \tag{136}$$

and we can therefore utilize a Kalman filter to do the estimation:

$$\text{Data Update:} \tag{137}$$

$$\hat{\theta}_{t|t} = \hat{\theta}_{t|t-1} + P_{t|t-1}\phi_t(y_t - \phi_t^T \hat{\theta}_{t|t-1}) \tag{138}$$

$$P_{t|t}^{-1} = P_{t|t-1}^{-1} + \phi_t \phi_t^T \tag{139}$$

$$\text{Time Update:} \tag{140}$$

$$\hat{\theta}_{t+1|t} = \hat{\theta}_{t|t} \tag{141}$$

$$P_{t+1|t} = P_{t|t} + R_1 \tag{142}$$

Estimation examples

Today's Matlab example topics:

- Recursive least-squares Method
- Linear time-varying estimation
- Nonlinear time-varying estimation

Design Configurations

When attempting to identify a system, we should consider the following:

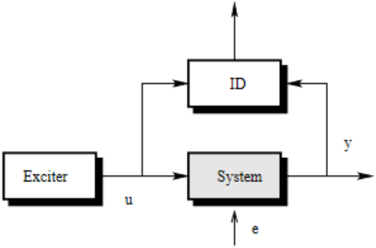
1. What are the outputs?
2. What are the inputs?
3. What are the disturbances?

We should further consider some practical aspects of the system:

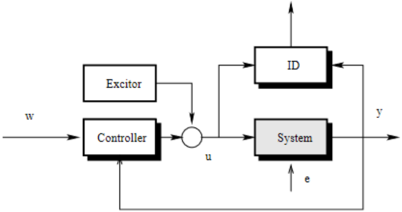
1. What are we allowed to do?
2. What type of model are we interested in?

Design Configurations

Open Loop



Closed Loop



Informative Experiments - Objectives

For any system $\mathcal{S}$, we can construct a set of models $\mathcal{M}$ to describe it:

$$\mathcal{S}: \quad y = G_0(q)u + H_0(q)e \tag{143}$$

$$\mathcal{M} = \{G(q,\theta), H(q,\theta) | \theta \in \mathcal{D}\} \tag{144}$$

Ideally we would have the system included within the possible models:

$$\mathcal{S} \in \mathcal{M} \tag{145}$$

If we have two models within $\mathcal{M}$,

$$\mathcal{M}_1 : y = G_1(q)u + H_1(q)e_1 \tag{146}$$

$$\mathcal{M}_2 : y = G_2(q)u + H_2(q)e_2 \tag{147}$$

we want to be able to determine which that approximates the system better.

Therefore, we need to perform an *informative* open-loop experiment.

Informative Experiments

We want to determine an input signal resulting in data that is *sufficiently informative* to dinstinguish between models in $\mathcal{M}$.

Consequently, for two models identified using data that is sufficiently informative, the expectation

$$\overline{E}\{\Delta\epsilon^2\} = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N} E\{\Delta\epsilon_t^2\} = \int_{-\pi}^{\pi}\phi_1(w)+\phi_2(w)dw = 0 \tag{148}$$

only holds if

$$\phi_2(w) = \left|\frac{H_0\Delta H}{H_1 H_2}\right|^2\sigma^2 = 0 \quad \Rightarrow \quad \Delta H(e^{jw}) \equiv 0 \tag{149}$$

$$\phi_1(w) = \left|\frac{1}{H_1}\right|^2\left|\Delta G + \frac{G_0 - G_2}{H_2}\Delta H\right|^2\Phi_u(w) = 0 \tag{150}$$

$$\Rightarrow |\Delta G(e^{jw})|^2\Phi_u(w) \equiv 0 \Rightarrow \Delta G(e^{jw}) \equiv 0 \tag{151}$$

Consequently, the input should have a spectrum $\Phi_u(w)$ for which the above expectation only becomes zero for identical models in $\mathcal{M}$.

Informative Experiments - Persistently excited signal

We say that such an input is persistently excited, with the following definition.

A quasi-stationary signal with spectrum $\Phi_u(w)$ is said to be persistently excited of order n (pe(n)) if, for all filters in the form

$$M(q) = m_0 + m_1 q^{-1} + ... + m_{n-1} q^{-(n-1)} \tag{152}$$

the relation

$$\Phi_z(w) = |M(e^{jw})|^2 \Phi_u(w) = 0, \quad z_t = M(q)u_t \tag{153}$$

implies that for all w

$$M(e^{jw}) \equiv 0 \tag{154}$$

$M(q)$ has n parameters and $n-1$ zeros; implying $M(q)M(q^{-1})$ has at most $n-1$ different zeros.

Informative Experiments - Persistently excited signal

In order to uniquely to determine the n coefficients in M, the spectrum, $\Phi_u(w)$, has to be non-zero at at least n different points in the interval $w \in [-\pi, \pi]$.

The reason for this is that a signal which is pe(n) can not be filtered to zero by an MA filter of order $n-1$, but n or higher might do it

$$u_t = const \neq 0, \text{ signal is pe(1)} \tag{155}$$

$$M_1(q) = 1 - q^{-1}: \quad M_1(q)u_t = u_t - u_{t-1} = 0 \tag{156}$$

$$M_0(q) = 1: \quad M_0(q)u_t = u_t \neq 0 \tag{157}$$

or looking at the spectrum: it is always zero

$$\Phi_u = \tilde{d}\delta(w) \tag{158}$$

$$\Phi_{M_1 u} = 2(1 - \cos(w))\tilde{d}\delta(w) = 0 \tag{159}$$

Alternatively, it can be stated that for a signal that is pe(n), $M(q)$ has at most n parameters. Which means there are maximum n estimated parameters in the model.

Informative Experiments - Persistently excited signal

Let us consider the transfer function:

$$G = q^{-k}\frac{B(q)}{F(q)} = q^{-k}\frac{b_0 + b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}}{1 + f_1 q^{-1} + \cdots + f_{n_f} q^{-n_f}} \tag{160}$$

noticing the order of the polynomials, we can say the signal u_t has to be $pe(n_b + n_f + 1)$

$$\Delta G = \frac{B_1}{F_1} - \frac{B_2}{F_2} = \frac{B_1 F_2 - B_2 F_1}{F_1 F_2} = 0 \quad \Rightarrow \quad |B_1 F_2 - B_2 F_1|^2 \Phi_u(w) = 0 \tag{161}$$

where it can be seen that the effective part of ΔG has the order $n_b + n_f$.

Informative Experiments - common signals

Let us now consider some signals that are persistently exciting.

A measure of the input power of a signal is beneficial, given that the variance of the estimation is inversely proportional to the input power. Given a practical signal is finitely bound, the measure can be expressed in terms of the crest factor, for zero-mean signals:

$$C_r^2 = \frac{\max_t u_t^2}{\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} u_t^2} \tag{162}$$

which for a good signal is as low as possible (minimum is 1).

For binary signals, $u_t = \pm\bar{u}$, the crest factor is minimum, $C_r^2 = 1$. This makes binary signals very useful for linear systems, but cannot in general handle nonlinear functions:

$$y_t = \frac{B(q)}{A(q)} f(u_t) \tag{163}$$

$$f(u_t) = A_f \cos(\pm\bar{u}) = A_f \cos(\bar{u}) \tag{164}$$

Informative Experiments - common signals

Let us now consider harmonic signals:

A single harmonic signal,

$$u_t = A\sin(wt), \tag{165}$$

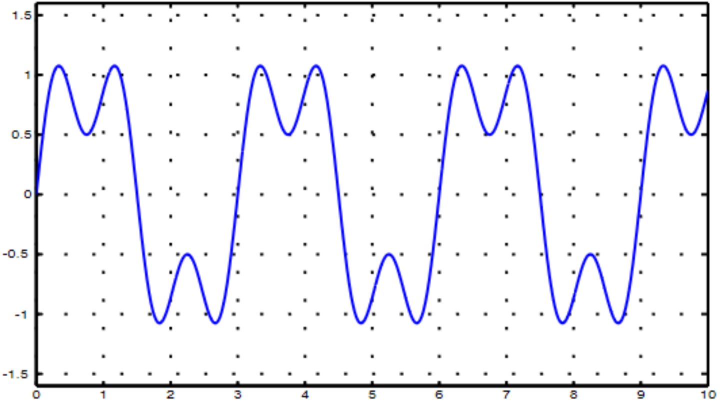
has two non-zero frequency components in its spectrum at $\pm w$, and is pe(2). However, the crest factor is $C_r^2 = 2$. If we instead consider sums of sinusoids:

$$u_t = \sum_{k=1}^{n} A_k \sin(w_k t + \phi_t) \tag{166}$$

Then we have 2 components for each w_k, so the signal is $pe(2n)$.
If $w_k = 0$ or $w_k = \frac{\pi}{T_s}$, the order goes down by 1 to $pe(2n-1)$ (by 2 if both)
The crest factor is, in the worst case, $C_r^2 = 2n$, and lowest if the sinusoids are maximally out of phase.

Informative Experiments - common signals

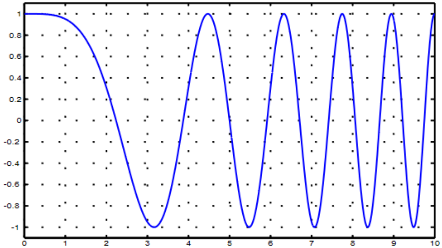
Sum of 2 harmonics, with maximum phase difference (180°)



Informative Experiments - common signals

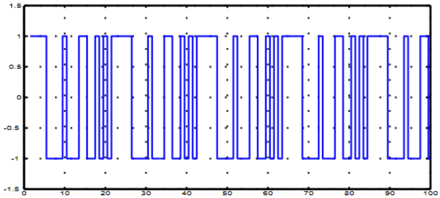
Among the single sine functions we have the chirp signal:

$$u_t = A\sin((w_0 + \alpha t)t), \quad C_2^2 = \sqrt{2} \tag{167}$$



Informative Experiments - common signals

If we now consider the binary signals, we have the option of using a PRBS signal



PRBS signals are deterministic, but has white noise-ish properties.

$$z_t = \mathrm{mod}(B(q)z_{t-1}, 2) \tag{168}$$

where B has the order m, and PRBS has the maximum length $M = 2^m - 1$. A PRBS signal is $pe(M-1)$, with $C_r^2 = 1$.

Informative Experiments - common signals

We can also apply Random Gaussian signals, which are filtered/colored white noise signals:

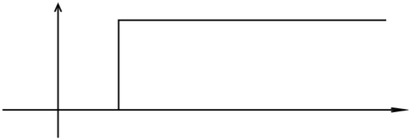
$$u_t = H_u(q)\check{e}_t, \quad \check{e}_t \in \mathcal{F}_{iid}(0, \sigma_u^2)(white) \tag{169}$$

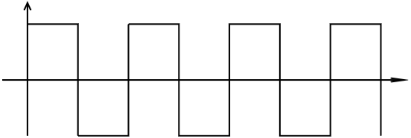
In practice, we would have to use a truncated Gaussian to keep the control bounded, e.g., within $\pm 3\sigma$ ($\approx 99\%$ coverage), giving $C_r^2 = 3$.

Random binary signals can be generated by taking the sign of a suitable Random Gaussian signal.

Informative Experiments - common signals

Finally the step and square wave signals are also quite common:





where, for a step at time M and a square (both between d_0 and d_1),

$$C_r^2 = \frac{d_1^2}{\lim\limits_{N\to\infty} \frac{Md_0^2+(N-M)d_1^2}{N}} = \frac{d_1^2}{d_1^2 + \lim\limits_{N\to\infty} \frac{M}{N} d_0^2} = 1, \qquad C_r^2 = \frac{d_1^2}{\frac{1}{2}d_1^2 + \frac{1}{2}d_0^2}$$

The pulse can also be represented as an infinite harmonic sum.

Questions

Questions?