Stochastic Adaptive Control (02421)

Lecture 11

Tobias K. S. Ritschel

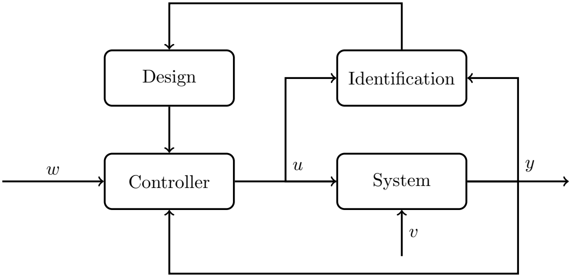
Section for Dynamical Systems

Department of Applied Mathematics and Computer Science

Technical University of Denmark

Lecture Plan

1. System theory
2. Stochastics
3. State estimation 1
4. State estimation 2
5. Optimal control 1
6. System identification 1 + adaptive control 1
7. External models + prediction
8. Optimal control 2
9. Optimal control 3
10. System identification 2
11. **System identification 3 + model validation**
12. System identification 4 + adaptive control 2
13. Adaptive control 3



Today's Agenda

- Time-varying estimation
- Design of experiments
- Model validation

Follow-up from last time

Questions?

Heuristics for time-varying systems

Time-varying estimation - first example

ARX model

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

$$b_1(t) = b_{1,0} + b_{1,1}t \tag{2}$$

Treat time-varying coefficient as two coefficients with their own inputs

$$y_t = \phi^T \theta + e_t \tag{3}$$

$$\theta^T = \begin{bmatrix} a_1 & a_2 & \cdots & a_{n_a} & b_{1,0} & b_{1,1} & b_2 & \cdots & b_{n_b} \end{bmatrix} \tag{4}$$

$$\phi^T = \begin{bmatrix} -y_{t-1} & -y_{t-2} & \cdots & -y_{t-n_a} & u_{t-1} & tu_{t-1} & u_{t-2} & \cdots & u_{t-n_b} \end{bmatrix} \tag{5}$$

Time-varying estimation

For deterministic time varying systems, rearrange the parameters

$$y_t = \phi_t^T \theta_t + e_t \tag{6}$$

$$\theta_t = \alpha + f(t)\beta = \begin{bmatrix} I & f(t) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{7}$$

$$y_t = \begin{bmatrix} \phi_t^T & \phi_t^T f(t) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + e_t \tag{8}$$

For piece-wise linear parameters, rearrange the parameters

$$y_t = \phi_t^T \theta_t + e_t \tag{9}$$

$$\theta_t = \alpha_i + (t - T_i)\beta_i, \quad T_i \leq t \leq T_{i+1} \tag{10}$$

$$y_t = \begin{bmatrix} \phi_t^T & \phi_t^T (t - T_i) \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} + e_t \tag{11}$$

Time-varying systems

System with general time-varying parameters

$$\theta_{t+1} = f(t, \theta_t, v_t) \tag{12}$$

- 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 diminishes over time

$$P_t \to 0 \tag{13}$$

Time-varying systems - Forgetting methods: Resetting

Reset the covariance after some time, t_i

$$P_{t_i} = P_i > P_{t_i - 1}, \quad \hat{\theta}_{t_i} = \hat{\theta}_{t_i - 1} \tag{14}$$

The appropriate restarting time depends on the application

For instance, restart at fixed intervals

$$t_i = Ni \tag{15}$$

This can be useful for periodic systems

Time-varying systems - Forgetting methods: Constant Gain

Another method: Keep the correction term large

For instance, keep the correction term κ constant

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \kappa \epsilon_t \tag{16}$$

$$\tilde{\theta}_t = (I - \kappa \phi_t^T)\tilde{\theta}_t - \kappa e_t \tag{17}$$

Alternatively, keep the variance constant

$$P_t = P \tag{18}$$

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \kappa \epsilon_t \tag{19}$$

$$\kappa_t = \frac{P\phi_t}{1 + \phi_t^T P \phi_t} \tag{20}$$

Time-varying systems - Forgetting methods: Exponential Forgetfulness

Another approach: Forget a little bit all the time (exponential forgetfulness)

$$J_t = \frac{1}{2}\sum_{i=1}^{t} \lambda^{t-i}\epsilon_i^2 = \lambda J_{t-1} + \frac{1}{2}\epsilon_t^2 \tag{21}$$

The recursion is similar to the previous methods

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t\phi_t\epsilon_t \tag{22}$$

$$\epsilon_t = y_t - \phi_t^T\hat{\theta}_{t-1} \tag{23}$$

$$P_t^{-1} = \lambda P_{t-1}^{-1} + \phi_t\phi_t^T \tag{24}$$

The forgetting factor λ can be expressed in terms of a horizon, N_∞

$$\lambda = 1 - \frac{1}{N_\infty} \tag{25}$$

Exponential forgetting - Example

Model

$$y_t = \frac{-1}{4} y_{t-1} + \frac{1}{2} y_{t-2} + u_{t-1} + e_t \tag{26}$$

New measurement ($t = 10$)

$$y_t = 1.47, \qquad u_{t-1} = 2 \tag{27}$$

Covariates, parameter estimate and covariance

$$\phi_t = \begin{bmatrix} 2.42 \\ 2.57 \\ 2 \end{bmatrix}, \quad \hat{\theta}_{t-1} = \begin{bmatrix} -0.2505 \\ 0.4960 \\ 0.9991 \end{bmatrix}, \quad P_{t-1} = \begin{bmatrix} 0.1355 & -0.0431 & -0.1057 \\ -0.0431 & 0.0620 & -0.0088 \\ -0.1057 & -0.0088 & 0.1242 \end{bmatrix}$$

Forgetting factor

$$\lambda = 0.95 \tag{28}$$

Exponential forgetting - Example

Residual

$$\epsilon_t = y_t - \phi_t^T \hat{\theta}_{t-1} = 1.47 - \begin{bmatrix} 2.42 & 2.57 & 2 \end{bmatrix} \begin{bmatrix} -0.2505 \\ 0.4960 \\ 0.9991 \end{bmatrix} = -1.1967 \quad (29)$$

Covariance

$$P_t^{-1} = \lambda P_{t-1}^{-1} + \phi_t \phi_t^T = 0.95 \begin{bmatrix} 259.9763 & 214.2840 & 236.4348 \\ 214.2840 & 192.9152 & 196.0344 \\ 236.4348 & 196.0344 & 223.1583 \end{bmatrix} \quad (30)$$

$$+ \begin{bmatrix} 2.42 \\ 2.57 \\ 2 \end{bmatrix} \begin{bmatrix} 2.42 & 2.57 & 2 \end{bmatrix} = \begin{bmatrix} 252.8339 & 209.7892 & 229.4530 \\ 209.7892 & 189.8743 & 191.3727 \\ 229.4530 & 191.3727 & 216.0004 \end{bmatrix} \quad (31)$$

Exponential forgetting - Example

Parameter estimate

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t \phi_t \epsilon_t = \begin{bmatrix} -0.2505 \\ 0.4960 \\ 0.9991 \end{bmatrix} \tag{32}$$

$$- 1.1967 \begin{bmatrix} 0.1426 & -0.0456 & -0.1111 \\ -0.0456 & 0.0638 & -0.0081 \\ -0.1111 & -0.0081 & 0.1298 \end{bmatrix} \begin{bmatrix} 2.42 \\ 2.57 \\ 2 \end{bmatrix} = \begin{bmatrix} -0.2574 \\ 0.4512 \\ 1.0350 \end{bmatrix} \tag{33}$$

Exponential forgetting - Exercise

Model

$$y_t = \frac{-1}{4} y_{t-1} + \frac{1}{2} y_{t-2} + u_{t-1} + e_t \tag{34}$$

New measurement ($t = 10$)

$$y_t = 1.47, \qquad u_{t-1} = 2 \tag{35}$$

Covariates, parameter estimate and covariance

$$\phi_t = \begin{bmatrix} 2.42 \\ 2.57 \\ 2 \end{bmatrix}, \quad \hat{\theta}_{t-1} = \begin{bmatrix} -0.2505 \\ 0.4960 \\ 0.9991 \end{bmatrix}, \quad P_{t-1} = \begin{bmatrix} 0.1355 & -0.0431 & -0.1057 \\ -0.0431 & 0.0620 & -0.0088 \\ -0.1057 & -0.0088 & 0.1242 \end{bmatrix}$$

Forgetting factor

$$\lambda = 0.90 \tag{36}$$

Solve the exercise in 15 min.

Exponential forgetting - Exercise

Residual

$$\epsilon_t = y_t - \phi_t^T \hat{\theta}_{t-1} = 1.47 - \begin{bmatrix} 2.42 & 2.57 & 2 \end{bmatrix} \begin{bmatrix} -0.2505 \\ 0.4960 \\ 0.9991 \end{bmatrix} = -1.1967 \quad (37)$$

Covariance

$$P_t^{-1} = \lambda P_{t-1}^{-1} + \phi_t \phi_t^T = 0.90 \begin{bmatrix} 259.9763 & 214.2840 & 236.4348 \\ 214.2840 & 192.9152 & 196.0344 \\ 236.4348 & 196.0344 & 223.1583 \end{bmatrix} \quad (38)$$

$$+ \begin{bmatrix} 2.42 \\ 2.57 \\ 2 \end{bmatrix} \begin{bmatrix} 2.42 & 2.57 & 2 \end{bmatrix} = \begin{bmatrix} 239.8351 & 199.0750 & 217.6313 \\ 199.0750 & 180.2286 & 181.5709 \\ 217.6313 & 181.5709 & 204.8424 \end{bmatrix} \quad (39)$$

Exponential forgetting - Exercise

Parameter estimate

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t \phi_t \epsilon_t = \begin{bmatrix} -0.2505 \\ 0.4960 \\ 0.9991 \end{bmatrix} \tag{40}$$

$$- 1.1967 \begin{bmatrix} 0.1505 & -0.0481 & -0.1172 \\ -0.0481 & 0.0672 & -0.0085 \\ -0.1172 & -0.0085 & 0.1369 \end{bmatrix} \begin{bmatrix} 2.42 \\ 2.57 \\ 2 \end{bmatrix} = \begin{bmatrix} -0.2577 \\ 0.4488 \\ 1.0369 \end{bmatrix} \tag{41}$$

Time-varying systems - Fortescue's Method

Improve with a time-varying forgetting factor depending on the prediction error, ϵ_t

$$\lambda_t = 1 - \frac{1}{N_0} \frac{\epsilon_t^2}{\sigma^2 s_t} \tag{42}$$

N_0 is the approx. horizon over which the parameter is roughly constant

Recursion

$$\epsilon_t = y_t - \phi_t^T \hat{\theta}_{t-1} \tag{43}$$

$$s_t = 1 + \phi_t^T P_{t-1} \phi_t \tag{44}$$

$$K_t = \frac{P_{t-1}\phi_t}{\lambda_t + s_t} \tag{45}$$

$$\hat{\theta}_t = \hat{\theta}_{t-1} + K_t \epsilon_t \tag{46}$$

$$P_t = (I - K_t \phi_t^T) P_{t-1} \frac{1}{\lambda_t} \tag{47}$$

Time-varying systems - Fortescue's Method

If the variance is unknown, we can introduce an estimate

$$\lambda_t = 1 - \frac{1}{N_0}\frac{\epsilon_t^2}{r_t s_t} \tag{48}$$

$$r_t = r_{t-1} + \frac{1}{t}\left(\frac{\epsilon_t^2}{s_t} - r_{t-1}\right), \qquad r_0 = \epsilon_0^2 \tag{49}$$

Time-varying systems - Model Estimators

Introduce model of parameters

$$\theta_{t+1} = \theta_t + v_t, \qquad v_t \sim N(0, R_1\sigma^2) \tag{50}$$

$$y_t = \phi_t^T\theta_t + e_t, \qquad e_t \sim N(0, \sigma^2) \tag{51}$$

Estimate parameters using the Kalman filter

Data update

$$\hat{\theta}_{t|t} = \hat{\theta}_{t|t-1} + P_{t|t-1}\phi_t(y_t - \phi_t^T\hat{\theta}_{t|t-1}) \tag{52}$$

$$P_{t|t}^{-1} = P_{t|t-1}^{-1} + \phi_t\phi_t^T \tag{53}$$

Time update

$$\hat{\theta}_{t+1|t} = \hat{\theta}_{t|t} \tag{54}$$

$$P_{t+1|t} = P_{t|t} + R_1 \tag{55}$$

Experiment design

Design Configurations

When attempting to identify a system, consider the following:

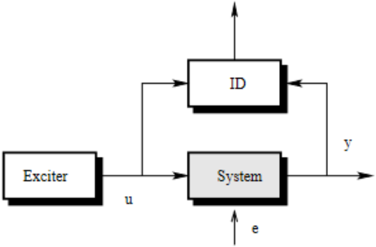
1. What are the outputs?
2. What are the inputs?
3. What are the disturbances?

Also consider some practical aspects of the system

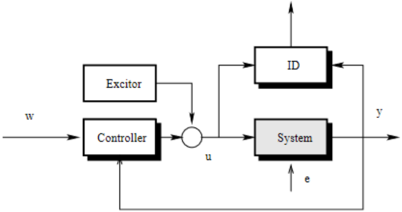
1. What are we allowed to do?
2. What type of model are we interested in?

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

$$\mathcal{S}: \quad y = G_0(q)u + H_0(q)e \tag{56}$$

$$\mathcal{M} = \{G(q,\theta), H(q,\theta) | \theta \in \mathcal{D}\} \tag{57}$$

Ideally, the system should be included in the set of possible models

$$\mathcal{S} \in \mathcal{M} \tag{58}$$

Given two models in $\mathcal{M}$

$$\mathcal{M}_1 : y = G_1(q)u + H_1(q)e_1 \tag{59}$$

$$\mathcal{M}_2 : y = G_2(q)u + H_2(q)e_2 \tag{60}$$

we want to be able to determine which that describes 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 $\mathcal{M}$

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

$$\overline{\mathbb{E}}[\Delta\epsilon^2] = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\mathbb{E}[\Delta\epsilon_t^2] = \int_{-\pi}^{\pi}\phi_1(w)+\phi_2(w)dw = 0 \tag{61}$$

only holds if

$$\phi_2(w) = \left|\frac{H_0\Delta H}{H_1 H_2}\right|^2\sigma^2 = 0 \quad \Rightarrow \quad \Delta H(e^{jw}) \equiv 0 \tag{62}$$

$$\phi_1(w) = \left|\frac{1}{H_1}\right|^2\left|\Delta G + \frac{G_0 - G_2}{H_2}\Delta H\right|^2\Phi_u(w) = 0 \tag{63}$$

$$\Rightarrow |\Delta G(e^{jw})|^2\Phi_u(w) \equiv 0 \Rightarrow \Delta G(e^{jw}) \equiv 0 \tag{64}$$

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

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

$$M(q^{-1}) = m_0 + m_1 q^{-1} + ... + m_{n-1} q^{-(n-1)} \tag{65}$$

the relation

$$\Phi_z(w) = |M(e^{jw})|^2 \Phi_u(w) = 0, \quad z_t = M(q^{-1}) u_t \tag{66}$$

implies that for all w

$$M(e^{jw}) = 0 \tag{67}$$

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

Informative Experiments - Persistently excited signal

Equivalently, 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 is that a signal which is pe(n) cannot be filtered to zero by an MA filter of order $n-1$, but n or higher might do it

$$u_t = \text{const} \neq 0, \text{ signal is pe(1)} \tag{68}$$

$$M_1(q^{-1}) = 1 - q^{-1} : \quad M_1(q^{-1})u_t = u_t - u_{t-1} = 0 \tag{69}$$

$$M_0(q^{-1}) = 1 : \qquad M_0(q^{-1})u_t = u_t \neq 0 \tag{70}$$

or looking at the spectrum: it is always zero

$$\Phi_u = \tilde{d}\delta(w) \tag{71}$$

$$\Phi_{M_1 u} = 2(1 - \cos(w))\tilde{d}\delta(w) = 0 \tag{72}$$

Informative Experiments - Persistently excited signal

Transfer function

$$G = q^{-k}\frac{B(q)}{A(q)} = q^{-k}\frac{b_0 + b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}}{1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}} \tag{73}$$

The signal u_t has to be $pe(n_b + n_a + 1)$

$$\Delta G = \frac{B_1}{A_1} - \frac{B_2}{A_2} = \frac{B_1 A_2 - B_2 A_1}{A_1 A_2} = 0 \quad \Rightarrow \quad |B_1 A_2 - B_2 A_1|^2 \Phi_u(w) = 0 \tag{74}$$

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

Informative Experiments - common signals

Crest factor (for zero-mean signals)

$$C_r^2 = \frac{\max_t u_t^2}{\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} u_t^2} \tag{75}$$

The crest factor should be as low as possible (the minimum is 1)

For binary signals, $u_t = \pm\bar{u}$, the crest factor is minimum, $C_r^2 = 1$

Consequently, binary signals are useful for linear systems, but cannot, in general, handle nonlinear systems

$$y_t = \frac{B(q)}{A(q)} f(u_t) \tag{76}$$

$$f(u_t) = \alpha\cos(\pm\bar{u}) = \alpha\cos(\bar{u}) \tag{77}$$

Informative Experiments - common signals

Single harmonic signal

$$u_t = A \sin(wt), \tag{78}$$

- Two non-zero frequency components in its spectrum (at $\pm w$)
- It is pe(2)
- Its crest factor is $C_r^2 = 2$

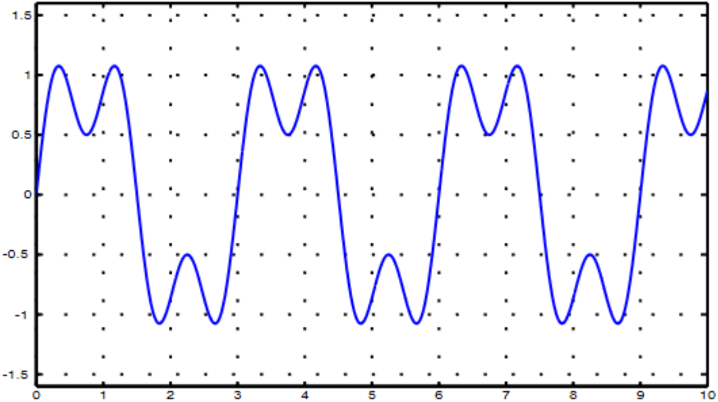
Sum of sines

$$u_t = \sum_{k=1}^{n} A_k \sin(w_k t + \phi_t) \tag{79}$$

- Two 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

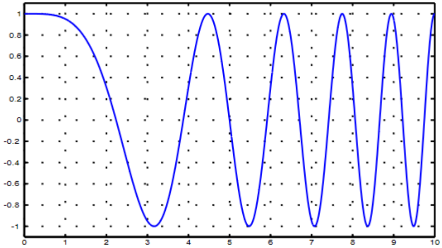
Sum of 2 harmonics, with maximum phase difference $(180°)$



Informative Experiments - common signals

Single sine function: The chirp signal

$$u_t = A\sin((w_0 + \alpha t)t), \quad C_r^2 = \sqrt{2} \tag{80}$$

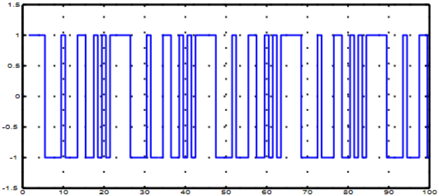


Informative Experiments - common signals

PRBS signal

$$z_t = \mathrm{mod}(B(q^{-1})z_{t-1}, 2) \tag{81}$$

B is order m and the signal has the maximum length $M = 2^m - 1$



- PRBS signals are deterministic, but have properties similar to those of white noise
- A PRBS signal is pe($M-1$) and $C_r^2 = 1$

Informative Experiments - common signals

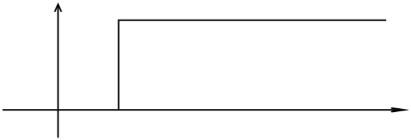
Alternative: Apply random Gaussian signals that are filtered/colored white noise signals

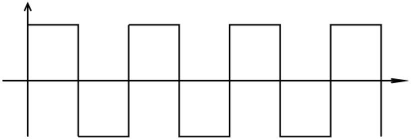
$$u_t = H_u(q)\check{e}_t, \quad \check{e}_t \sim \mathcal{F}_{iid}(0, \sigma_u^2)(white) \tag{82}$$

- In practice, we would have to use a truncated Gaussian to keep the control bounded, e.g., within $\pm 3\sigma$ ($\approx 99\%$ coverage), resulting in $C_r^2 = 3$
- Random binary signals can be generated by taking the sign of a suitable Random Gaussian signal

Informative Experiments - common signals

Step and square wave signals are also commonly used





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

Informative experiments - Example

ARMAX model

$$A(q^{-1})y_t = B(q^{-1})u_t + C(q^{-1})e_t, \qquad e_t \sim N(0, 0.05) \tag{83}$$

Polynomials

$$A(q^{-1}) = 1 - 1.2q^{-1} + 0.8q^{-2}, \tag{84}$$

$$B(q^{-1}) = \quad - 0.5q^{-1} + 0.2q^{-2}, \tag{85}$$

$$C(q^{-1}) = 1 + 0.3q^{-1} \tag{86}$$

Informative experiments - Example

Define system

```
>> A  = [1, −1.2, 0.8];      D = 1;
>> B  = [0, −0.5, 0.2];      F = 1;
>> C  = [1, 0.3];            R = 0.05;
>> Ts = 1;
>> M  = idpoly(A, B, C, D, F, R, Ts);
```

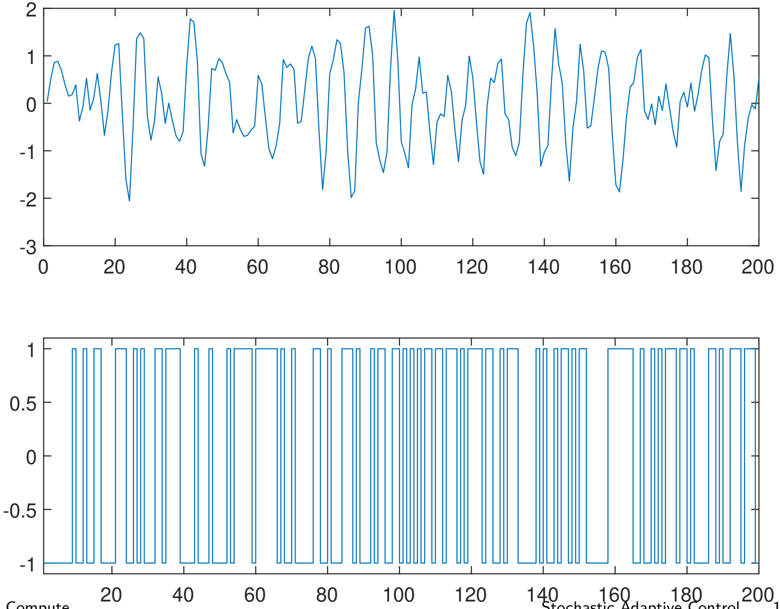
Simulate system

```
>> N = 200;
>> u = iddata([], idinput(N, 'prbs'));
>> e = iddata([], sqrt(R)*randn(N, 1));
>> y = sim(M, [u, e]);
>> simdata = [y.y, u.u];
```

Visualize simulation

```
>> figure(1);
>> subplot(211);
>> plot(y.y);
>> subplot(212);
>> stairs(u.u([1:end, end]));
>> xlim([1, numel(u.u)]);
>> ylim([−1.1, 1.1]);
```

Informative experiments - Example



Informative experiments - Exercise

ARMAX model

$$A(q^{-1})y_t = B(q^{-1})u_t + C(q^{-1})e_t, \qquad e_t \sim N(0, 0.05) \tag{87}$$

Polynomials

$$A(q^{-1}) = 1 + 1.3q^{-1} - 0.7q^{-2}, \tag{88}$$
$$B(q^{-1}) = \quad - 0.3q^{-1} - 0.1q^{-2}, \tag{89}$$
$$C(q^{-1}) = 1 + 0.5q^{-1} \tag{90}$$

Informative experiments - Exercise

Define system

```
>> A  = [1,  0.8,  -0.1];     D = 1;
>> B  = [0,  -0.3,  -0.2];    F = 1;
>> C  = [1,  0.5];            R = 0.05;
>> Ts = 1;
>> M  = idpoly(A, B, C, D, F, R, Ts);
```

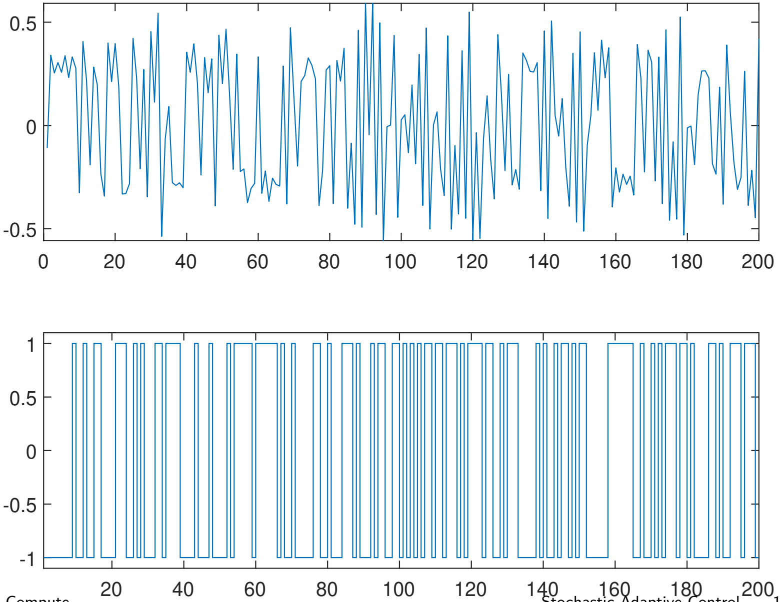
Simulate system

```
>> N = 200;
>> u = iddata([], idinput(N, 'prbs'));
>> e = iddata([], sqrt(R)*randn(N, 1));
>> y = sim(M, [u, e]);
>> simdata = [y.y, u.u];
```

Visualize simulation

```
>> figure(1);
>> subplot(211);
>> plot(y.y);
>> subplot(212);
>> stairs(u.u([1:end, end]));
>> xlim([1, numel(u.u)]);
>> ylim([-1.1, 1.1]);
```

Informative experiments - Exercise



Model validation

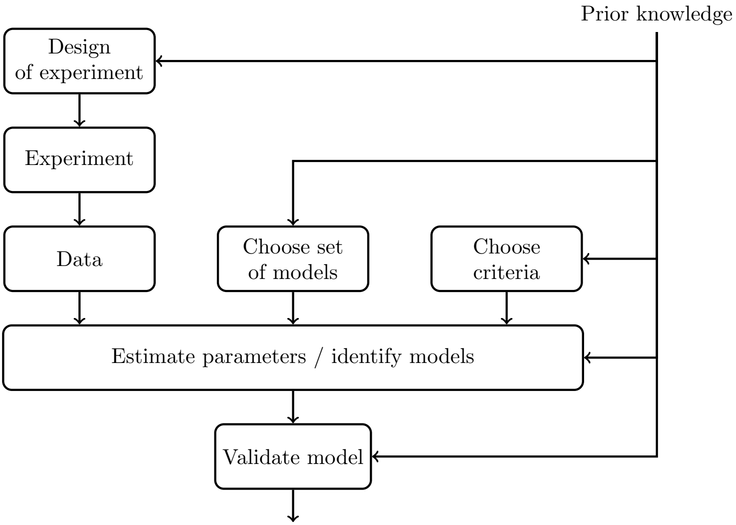
Model Validation

We now know how to estimate a model, but how do we check if it is correctly estimated when we don't know the true parameters?

Essentially, we are asking the following two questions

1. Is our model too simple?
2. Is our model too complex?

Model Validations



Model Validation

Three available quantities for validation

1. the estimated parameters
2. the uncertainty (the variance)
3. the undescribed model parts (the residuals)

The last is the source of measurement deviations

$$\text{measurement}(y) = \text{model}(\theta, u) + \text{residual}(\epsilon) \tag{91}$$

Model Validation - too complex?

Question: Does our model have too many parameters?

Unbiased estimate

$$\hat{\theta} \sim \mathcal{F}(\theta, P) \tag{92}$$

θ_i is *significant* if it, with reasonable certainty, is different from zero

Model validation - parameter insignificant?

Use a *marginal parameter test* to validate that a parameter is significant

For sufficiently many measurements, the distribution approaches a normal distribution

$$\hat{\theta} \sim N(\theta, P) \tag{93}$$

If the following holds, θ_i is, with $(1-\alpha)$% confidence not insignificant

$$|\hat{\theta}_i| > f_{1-\frac{\alpha}{2}} \sqrt{P_{i,i}} \tag{94}$$

f_x is the xth quantile of the standard normal distribution. This approach requires that the variance, P, is known

Model validation - parameter insignificant?

If the variance, P, was estimated, use the t-distribution

$$z_i = \frac{\hat{\theta}_i}{\sqrt{P_{i,i}}} \sim t(M - d_p) \tag{95}$$

d_p is the number of parameters and M is the number of measurements

If the following holds, θ_i is, with $(1-\alpha)$% confidence not insignificant

$$|\hat{\theta}_i| > f^t_{1-\frac{\alpha}{2}}(M - d_p)\sqrt{P_{i,i}} \tag{96}$$

f^t_x is the xth quantile of the t-distribution. Again, if $M \gg d_p$, this will approach the normal distribution

Model Validation - Multiple Insignificant Parameters?

More than one parameter might be insignificant, but we cannot tell whether its some or all

But we can test whether all parameters in a subset θ_b are significant

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_a \\ \hat{\theta}_b \end{bmatrix} \sim N\left(\begin{bmatrix} \theta_a \\ \theta_b \end{bmatrix}, \begin{bmatrix} P_a & P_{ab} \\ P_{ab}^T & P_b \end{bmatrix} \right) \tag{97}$$

Test statistic for the hypothesis of insignificant parameters $(\theta_b = 0)$

$$z_b = \hat{\theta}_b^T P_b^{-1} \hat{\theta}_b \sim F(d_b, M - d_p) \tag{98}$$

If the following holds, all parameters in θ_b are, with $(1-\alpha)$% confidence significant

$$z_b > f_{1-\alpha}^F(d_b, M - d_p) \tag{99}$$

d_b is the size of the subset and f_x^F is the xth quantile of the F-distribution. For large M, we can apply a $\chi^2(d_b)$ instead of the F-distribution

Model Reduction

Distribution of parameter estimates

$$\begin{bmatrix} \theta_a \\ \theta_b \end{bmatrix} \sim N\left(\begin{bmatrix} \hat{\theta}_a \\ \hat{\theta}_b \end{bmatrix}, \begin{bmatrix} P_a & P_{ab} \\ P_{ab}^T & P_b \end{bmatrix} \right) \tag{100}$$

If a subset of the parameters, $\hat{\theta}_b$, is insignificant, we can reduce the model using the projection theorem

$$\theta_a | \theta_b \sim N(\hat{\hat{\theta}}_a, \bar{P}_a) \tag{101}$$

$$\hat{\hat{\theta}}_a = \hat{\theta}_a - P_{ab}^T P_b^{-1} \hat{\theta}_b \tag{102}$$

$$\bar{P}_a = P_a - P_{ab} P_b^{-1} P_{ab}^T \tag{103}$$

When have we used the projection theorem before and what for?

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

Insignificant: singular analysis of the variance matrix P

Most estimations methods involve solving linear equations in the form

$$H\hat{\theta} = g \tag{104}$$

H is a measure of the data set related to the variance, $H^{-1} = P$

$$P = \left(\sum_{i=0}^{N} \psi_i \psi_i^T \right)^{-1} \sigma^2 \tag{105}$$

- If a model is overparameterized, then (in the ideal case) H will be singular
- In the less ideal case, H is invertible, but has eigenvalues that are significantly smaller than the rest

$$\mathrm{eig}(H)_i \ll \mathrm{eig}(H)_j \qquad \Leftrightarrow \qquad \mathrm{eig}(P)_i \gg \mathrm{eig}(P)_j \tag{106}$$

- This requires that the system is sufficiently excited – insufficiently excited systems will result in similar issues

Condition number

Another way to evaluate if a model is overparameterized is to consider the condition number of its variance.

$$\mathrm{cond}(P) = \frac{|\lambda_{\max}|}{|\lambda_{\min}|}, \quad \lambda = \mathrm{eig}(P) \tag{107}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of P

If $\mathrm{cond}(P)$ is large, it indicates overparameterization

Example:

$$\text{Model 1: } \mathrm{cond}(P_1) = 1000 \tag{108}$$

$$\text{Model 2: } \mathrm{cond}(P_2) = 40 \tag{109}$$

Model 1 appears to be too complex, while model 2 is more balanced

Zeros and poles: Cancellation?

If the model is overparameterized, some zeros and poles might be close to each other

$$y_t = H_{yu}(q)u_t + H_{ye}(q)e_t \tag{110}$$

Use linearization to approximate uncertainty in zero and poles

$$\hat{p}_i = f_i(\hat{\theta}) \simeq f_i(\theta) + \frac{\partial f_i}{\partial \theta}\tilde{\theta}, \qquad \tilde{\theta} \sim N(0, P) \tag{111}$$

$$\hat{p}_i \sim N\left(p_i, \frac{\partial f_i}{\partial \theta} P \left(\frac{\partial f_i}{\partial \theta}\right)^T\right) \tag{112}$$

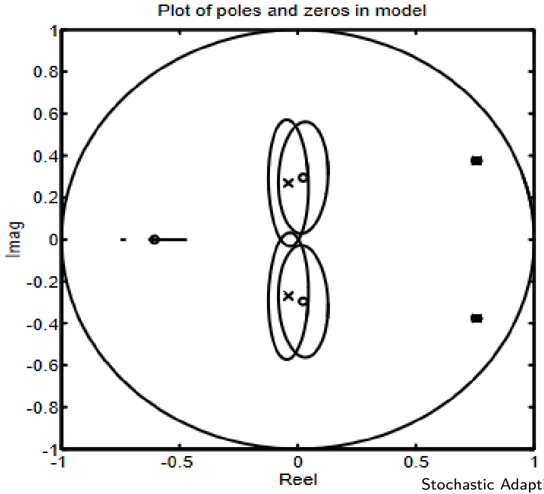
If the confidence intervals of a pole and a zero overlap, it is a strong indication that they cancel each other out

Hint: Use Matlab's `zpplot`

Zeros and poles: Example of cancellation

$$(1 - 1.5q^{-1} + 0.7q^{-2})y_t = (1 - 0.5q^{-1})u_t + e_t \tag{113}$$

$$(1 - a_1 q^{-1} + ... + a_4 q^{-4})y_t = (b_0 + ... + b_3 q^{-3})u_t + e_t \tag{114}$$



Residual Analysis

Question: Is the model too simple?

Residuals

$$\text{measurement}(y) = \text{model}(\theta, u) + \text{residual}(\epsilon) \quad (115)$$

For a perfect model, the residuals would have the following properties

1. $\epsilon_t \sim \mathcal{F}(0, \sigma^2)$.
2. ϵ_t has a symmetric distribution
3. ϵ_t is white
4. ϵ_t is uncorrelated with current and prior inputs

Equivalently (in terms of co-variance functions)

$$r_\epsilon(k) = \mathbb{E}[\epsilon_{t+k}\epsilon_t] = \begin{cases} \sigma^2 & k = 0, \\ 0 & \text{otherwise}, \end{cases} \qquad r_{\epsilon_t,u_t}(k) = \mathbb{E}[\epsilon_{t+k}u_t] = 0 \quad (116)$$

Important: use one data set for *estimation* and another for the *validation* (cross-validation)

Residual Analysis - mean and variance test

Simple approach: Test whether the distribution of the residuals has the right mean and variance

If the below holds, the residuals are not zero mean

$$|\bar{\epsilon}| > f^{t}_{1-\frac{\alpha}{2}}(M-1)\sqrt{\frac{S^2}{M}} \tag{117}$$

$$\bar{\epsilon} = \frac{1}{M}\sum_{i=1}^{M}\epsilon_i, \quad S^2 = \frac{1}{M-1}\sum_{i=1}^{M}(\epsilon_i - \bar{\epsilon})^2 \tag{118}$$

If either of the below hold, the variance is time-varying

$$\frac{S_1^2}{S_2^2} < f^{F}_{\alpha/2}(M_1, M_2) \text{ or } \frac{S_1^2}{S_2^2} > f^{F}_{1-\alpha/2}(M_1, M_2) \tag{119}$$

$$S_i^2 = \frac{1}{M_i}\sum_{j=1}^{M_i}\epsilon_{i+j}^2 \tag{120}$$

Note: The intervals must be non-overlapping

Residual Analysis - sign test

Test for whiteness: The number of sign changes, z, should follow (M is the number of data points)

$$z \sim N\left(\frac{M-1}{2}, \frac{M-1}{4}\right) \tag{121}$$

We reject the hypothesis if either of the below holds

$$z < \frac{M-1}{2} - \sqrt{\frac{M-1}{4}} f^N_{1-\frac{\alpha}{2}} \text{ or } z > \frac{M-1}{2} + \sqrt{\frac{M-1}{4}} f^N_{1-\frac{\alpha}{2}} \tag{122}$$

That is, the hypothesis is rejected if the test statistic is outside the confidence interval

Residual Analysis - test of co-variance function

Alternative test for whiteness: The auto-covariance must be in the form

$$r_\epsilon(k) = \mathbb{E}[\epsilon_{t+k}\epsilon_t] = \begin{cases} \sigma^2 & k = 0 \\ 0 & \text{otherwise} \end{cases} \tag{123}$$

Estimates of auto-covariance and auto-correlation

$$\hat{r}_\epsilon(k) = \frac{1}{M}\sum_{t=1}^{M-k} \epsilon_{t+k}\epsilon_t, \qquad \hat{\rho}_\epsilon(k) = \frac{\hat{r}_\epsilon(k)}{\hat{r}_\epsilon(0)} \tag{124}$$

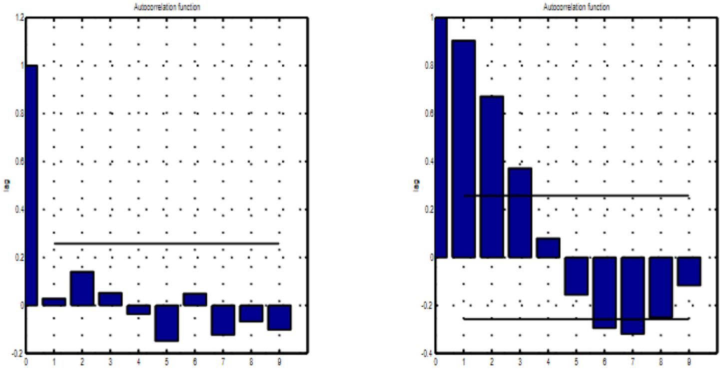
Test the covariance at each time step

$$H_0: \quad \sqrt{M}\hat{\rho}_\epsilon(k) \sim N(0,1), \text{ reject if } |\hat{\rho}_\epsilon(k)| > \frac{f^N_{1-\frac{\alpha}{2}}}{\sqrt{M}} \tag{125}$$

Test if the covariance is zero for $k \neq 0$

$$H_0: \quad z = M\sum_{i=1}^{m} \hat{\rho}^2_\epsilon(i) \sim \chi^2(m), \text{ reject if } z > f^{\chi^2}_{1-\alpha}(m) \tag{126}$$

Residual Analysis - test of autocorrelation



Residual Analysis - cross-covariance function test

Test the cross-covariance

$$r_{\epsilon,u}(k) = \mathbb{E}[\epsilon_{t+k} u_t] = 0 \qquad (127)$$

Cross-covariance and cross-correlation

$$\hat{r}_{\epsilon,u}(k) = \frac{1}{M} \sum_{i=1}^{M-k} \epsilon_{t+k} u_t, \qquad \hat{\rho}_{\epsilon,u}(k) = \frac{\hat{r}_{\epsilon,u}(k)}{\sqrt{\hat{r}_{\epsilon}(0)\hat{r}_u(0)}} \qquad (128)$$

Marginal test of the cross-covariance

$$H_0 : \quad \sqrt{M}\hat{\rho}_{\epsilon,u}(k) \sim N(0,1), \text{ reject if } |\hat{\rho}_{\epsilon,u}(k)| > \frac{f^N_{1-\frac{\alpha}{2}}}{\sqrt{M}} \qquad (129)$$

Check if the covariance is zero for $k \neq 0$

$$H_0 : \quad z = M \sum_{i=1}^{m} \hat{\rho}^2_{\epsilon,u}(i) \sim \chi^2(m), \text{ reject if } z > f^{\chi^2}_{1-\alpha}(m) \qquad (130)$$

Residual Analysis - spectral density test

Alternative test for whiteness: Consider the Fourier transformed residuals

$$X(w_k) = \frac{1}{M} \sum_{t=1}^{M} \epsilon_t e^{j w_k t} \tag{131}$$

Hint: Matlab's `fft` can be used to compute $X(w_k)$

Estimated spectral density (periodogram)

$$\hat{\phi}(w_k) = |X(w_k)|^2 \tag{132}$$

Hint: Matlab's `etfe` can be used to compute $\hat{\phi}(w_k)$

If x_t is white noise

$$\mathbb{E}[\hat{\phi}(w_k)] = 2\sigma^2 \tag{133}$$

Residual Analysis - model comparison tests

Question: Can we validate a model using a single data set?

Coefficient of determination

$$R^2 = \frac{J_0 - J(\hat{\theta})}{J_0} \tag{134}$$

$$J_0 = \frac{1}{2}\sum_{i=1}^{M}(y_i - \bar{y})^2, \quad J(\hat{\theta}) = \frac{1}{2}\sum_{i=1}^{M}\epsilon_i^2 \tag{135}$$

$J(\hat{\theta})$ is the loss-function and a perfect model results in $R^2 = 1$. Lower values of R^2 indicate worse models

Alternative loss functions

$$W(\hat{\theta}) = \sum_{i=1}^{M}\epsilon_i^2, \qquad W_M(\hat{\theta}) = \frac{1}{M}\sum_{i=1}^{M}\epsilon_i^2 \tag{136}$$

The loss functions are monotonically decreasing with model complexity

Residual Analysis - model comparison: F-test

Objective: Compare two model classes, $\mathcal{M}_1$ and $\mathcal{M}_2$ using the F-test

Hypothesis: $\mathcal{M}_{\text{true}} \subset \mathcal{M}_1 \subset \mathcal{M}_2$ where $d_2 \geq d_1$ are the number of model parameters. Consequently, the loss-function $J_i = J_i(\hat{\theta})$ does not decrease significantly by increasing the model size if $\mathcal{M}_1 \subset \mathcal{M}_2$

Test statistic

$$H_0 : z = \frac{J_1 - J_2}{J_2} \frac{M - d_2}{d_2 - d_1} \sim F(d_2 - d_1, M - d_2) \tag{137}$$

Reject hypothesis if

$$z > f^F_{1-\alpha}(d_2 - d_1, M - d_2) \tag{138}$$

Residual Analysis - model comparison: Information Criteria

Information criteria

1. Akaike's Information Criterion (AIC); tends towards higher complexity

$$AIC = \left(1 + \frac{2d}{M}\right) W_M \tag{139}$$

2. Bayesian Information Criterion (BIC);

$$BIC = \left(1 + \frac{\log(M)d}{M}\right) W_M \tag{140}$$

3. Akaike's Final Prediction Error (FPE) Criterion; expresses the variance of the prediction error, also $FPE \rightarrow AIC, M \gg d$

$$FPE = \frac{M+d}{M-d} W_M = \left(1 + \frac{2d}{M-d}\right) W_M \tag{141}$$

If two models have the same d, choose the one with the lowest loss function

Residual Analysis - model comparison: Information Criteria

Information criteria

$$AIC = \left(1 + \frac{2d}{M}\right) W_M \tag{142}$$

$$BIC = \left(1 + \frac{\log(M)d}{M}\right) W_M \tag{143}$$

$$FPE = \frac{M+d}{M-d} W_M = \left(1 + \frac{2d}{M-d}\right) W_M \tag{144}$$

What happens to the criteria as $M \to \infty$?

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

Model validation - Example

ARMAX model

$$A(q^{-1})y_t = B(q^{-1})u_t + C(q^{-1})e_t, \qquad e_t \sim N(0, 0.05) \tag{145}$$

Polynomials

$$A(q^{-1}) = 1 - 1.2q^{-1} + 0.8q^{-2}, \tag{146}$$
$$B(q^{-1}) = \quad - 0.5q^{-1} + 0.2q^{-2}, \tag{147}$$
$$C(q^{-1}) = 1 + 0.3q^{-1} \tag{148}$$

Define system

```
>> A  = [1, -1.2, 0.8];     D = 1;
>> B  = [0, -0.5, 0.2];     F = 1;
>> C  = [1, 0.3];           R = 0.05;
>> Ts = 1;
>> M  = idpoly(A, B, C, D, F, R, Ts);
```

Simulate system

```
>> N = 200;
>> u = iddata([], idinput(N, 'prbs'));
>> e = iddata([], sqrt(R)*randn(N, 1));
>> y = sim(M, [u, e]);
>> simdata = [y.y, u.u];
```

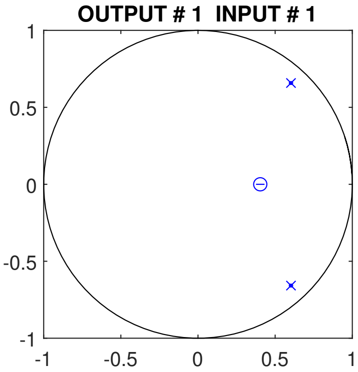
Model validation - Example

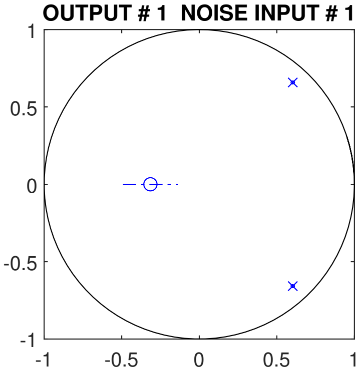
Estimate parameters

```
>> th = armax(simdata, [2, 2, 1, 1])
```

Zero-pole cancellation

```
>> figure(1)
>> subplot(121)
>> zpplot(th2zp(th, 1), norminv(0.995)) % 99% Gaussian CI
>> subplot(122)
>> zpplot(th2zp(th, 0), norminv(0.995)) % 99% Gaussian CI
```





Model validation - Example

Eigenvalue analysis

```
>> [TH,P] = th2par(th);
>> log10(eig(P))

ans =

   -5.1145
   -4.8291
   -4.7172
   -3.8705
   -2.3304
```

Insignificant parameters

```
>> TH + sqrt(diag(P))*[-1,1]*norminv(0.995)

ans =

   -1.2319   -1.1860
    0.7790    0.8178
   -0.5121   -0.4936
    0.1862    0.2139
    0.1555    0.5077
```

Model validation - Exercise

ARMAX model

$$A(q^{-1})y_t = B(q^{-1})u_t + C(q^{-1})e_t, \qquad e_t \sim N(0, 0.05) \tag{149}$$

Polynomials

$$A(q^{-1}) = 1 - 1.2q^{-1} + 0.8q^{-2}, \tag{150}$$

$$B(q^{-1}) = \quad - 0.5q^{-1} + 0.2q^{-2}, \tag{151}$$

$$C(q^{-1}) = 1 + 0.3q^{-1} \tag{152}$$

Estimate the parameters in a model with

$$n_a = 3, \tag{153}$$

$$n_b = 3, \tag{154}$$

$$n_c = 2, \tag{155}$$

$$k = 1 \tag{156}$$

Solve the exercise in 15 min.

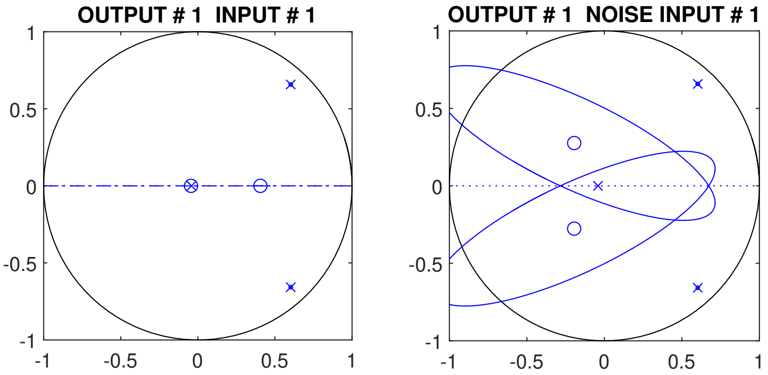
Model validation - Exercise

Estimate parameters

```
>> th = armax(simdata, [3, 3, 2, 1])
```

Zero-pole cancellation

```
>> figure(1)
>> subplot(121)
>> zpplot(th2zp(th, 1), norminv(0.995)) % 99% Gaussian CI
>> subplot(122)
>> zpplot(th2zp(th, 0), norminv(0.995)) % 99% Gaussian CI
```



Model validation - Exercise

Eigenvalue analysis

```
>> [TH,P] = th2par(th);
>> log10(eig(P))

ans =

   -5.2588
   -5.0938
   -4.8826
   -4.7356
   -4.2485
   -2.5110
   -2.2328
    0.3504
```

Insignificant parameters

```
>> TH + sqrt(diag(P))*[-1,1]*norminv(0.995)

ans =

   -2.9920    0.6649
   -1.4583    2.9499
   -1.4232    1.4888
   -0.5054   -0.4884
   -0.7291    1.0882
   -0.3571    0.3749
   -1.4256    2.2050
   -0.4790    0.7077
```

Questions

Questions?