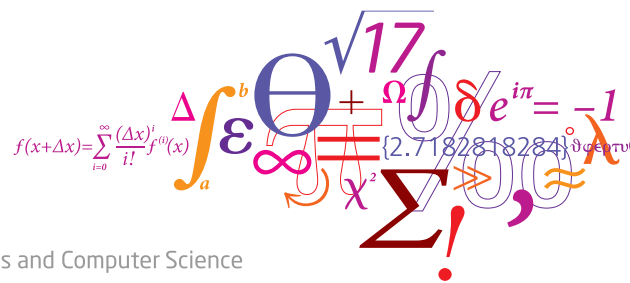
Stochastic Adaptive Control (02421)

Lecture 11

Tobias K. S. Ritschel

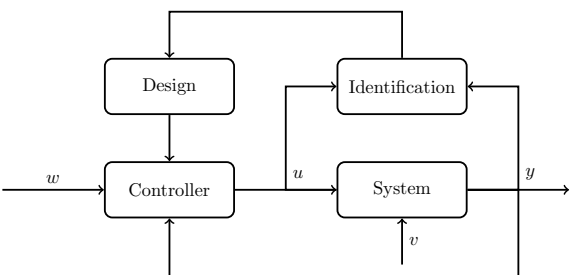
Section for Dynamical Systems, DTU Compute



DTU Compute
Department of Applied Mathematics and Computer Science

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
- MV: Model validation
  - Model reduction
  - Zeros and poles
  - Residual Analysis
  - Model Comparison
- Closed-loop identification
  - CL-Id - properties
  - Direct Approach
  - Indirect Approach
  - Joint Input-Output Approach

Follow-up from last time, Question 1.1 and 1.2

We want to estimate a MARX model of the two tank system.
If we remember from last time, we write the MIMO system as:

$$y_t^T = \phi^T \theta + e_t \tag{1}$$

$$\theta^T = \begin{bmatrix} A_1 & A_2 & B_1 \end{bmatrix}^T \tag{2}$$

$$\phi^T = \begin{bmatrix} -y_{t-1}^T & -y_{t-2}^T & u_{t-1}^T \end{bmatrix} \tag{3}$$

In this example the matrices takes the form:

$$\theta = \left[\begin{bmatrix} a_{11,1} & a_{12,1} \\ a_{21,1} & a_{22,1} \end{bmatrix} \quad \begin{bmatrix} a_{11,2} & a_{12,2} \\ a_{21,2} & a_{22,2} \end{bmatrix} \quad \begin{bmatrix} b_{1,1} \\ b_{2,1} \end{bmatrix} \right]^T \tag{4}$$

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

Matlab: $arx(DATA, \left[\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right]) = arx(data, [n_a, n_b, k])$

Follow-up from last time, Question 1.1 and 1.2

If we compare the estimates of the deterministic and stochastic simulations

	1	z^{-1}	z^{-2}
A_{11}	1.0000	-0.9693	-0.0005
A_{12}	0.0000	0.0000	-0.0284
A_{21}	0.0000	-0.0452	-0.0142
A_{22}	1.0000	0.0000	-0.8644
B_1	0.0000	0.0320	
B_2	0.0000	0.0005	

	1	z^{-1}	z^{-2}	dCl0	dCl1	dCl2
A_{11}	1.0000	-1.1383	-0.0289	0.0000	0.1200	0.1262
A_{12}	0.0000	0.7321	-0.2149	0.0000	0.1552	0.1625
A_{21}	0.0000	-0.1988	-0.0252	0.0000	0.1187	0.1248
A_{22}	1.0000	-0.1943	-0.2043	0.0000	0.1535	0.1607
B_1	0.0000	0.0292		0.0000	0.0038	
B_2	0.0000	-0.0023		0.0000	0.0037	

Follow-up from last time

Questions?

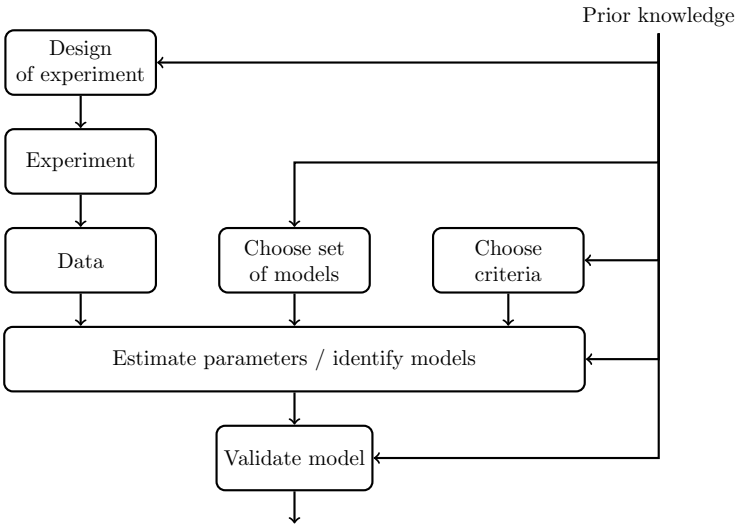
Model Validations

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

In some sense, we are asking the following two questions.

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

Model Validations



Model Validations

In order to validate the model we have estimated, we can utilize the three components of an estimated model:

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

where the last is the source of measurement deviations:

$$\text{measurement}(y) = \text{model}(\theta, u) + \text{residual}(\epsilon) \tag{6}$$

Model Validations - too complex?

Let us consider the case were we want to check, if our model has too many parameters. Consider the unbiased estimate,

$$\hat{\theta} \in \mathcal{F}(\theta, P) \tag{7}$$

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

Conversely, if a parameter might be zero, it might be insignificant, i.e., it might not be needed in the model. If we decide that a parameter is not needed, we can reduce our model.

Model validations - parameter insignificant?

In order to check if a parameter might be insignificant, we can do a marginal parameter test to check our confidence in the parameter.

If we consider the "Central Limit Theorem" from statistics, and we have a high number of measurements, then the parameter distribution will approach the normal distribution:

$$\hat{\theta} \in N(\theta, P) \tag{8}$$

We can then say, with $1 - \alpha$ percent confidence, that θ_i is not insignificant if the following holds

$$|\hat{\theta}_i| > f_{1-\frac{\alpha}{2}} \sqrt{P_{i,i}} \tag{9}$$

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

Model validations - parameter insignificant?

If the variance P was estimated during the estimation, it is more correct to consider the t-distribution:

$$z_i = \frac{\hat{\theta}_i}{\sqrt{P_{i,i}}} \in t(M - d_p) \tag{10}$$

where d_p is the number of parameters and M is the number of observations in the estimate.

We can then say with $1 - \alpha$ percent confidence that a given parameter is not insignificant if the following holds

$$|\hat{\theta}_i| > f^t_{1-\frac{\alpha}{2}}(M - d_p)\sqrt{P_{i,i}} \tag{11}$$

where 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?

If more than one parameter might be insignificant, then, due to parameter correlation, we cannot say whether all or some are insignificant based on the marginal tests.

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

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_a \\ \hat{\theta}_b \end{bmatrix} \in N\left(\begin{bmatrix} \theta_a \\ \theta_b \end{bmatrix}, \begin{bmatrix} P_a & P_{ab} \\ P_{ab}^T & P_b \end{bmatrix} \right) \tag{12}$$

For a subset of the parameters, θ_b, we test the hypothesis of insignificant parameters ($\theta_b = 0$), by considering the squared value:

$$z_b = \hat{\theta}_b^T P_b^{-1} \hat{\theta}_b \in F(d_b, M - d_p) \tag{13}$$

We can then say with $1 - \alpha$ percent confidence that all the parameters in θ_b are significant if

$$z_b > f_{1-\alpha}^F(d_b, M - d_p) \tag{14}$$

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

Model Reduction

If we find that we have subset $\hat{\theta}_b$ where none of its elements can be added to the significant subset $\hat{\theta}_a$, then, by considering the correlations between the subsets,

$$\begin{bmatrix}\theta_a \\ \theta_b\end{bmatrix} \in N\left(\begin{bmatrix}\hat{\theta}_a \\ \hat{\theta}_b\end{bmatrix}, \begin{bmatrix}P_a & P_{ab} \\ P_{ab}^T & P_b\end{bmatrix}\right) \tag{15}$$

we can perform a model reduction by applying the projection theorem:

$$\theta_a|\theta_b \in N(\hat{\bar{\theta}}_a, \bar{P}_a) \tag{16}$$

$$\hat{\bar{\theta}}_a = \hat{\theta}_a - P_{ab}^T P_b^{-1} \hat{\theta}_b \tag{17}$$

$$\bar{P}_a = P_a - P_{ab} P_b^{-1} P_{ab}^T \tag{18}$$

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 are related to solving linear equations in the form

$$H\hat{\theta} = g \tag{19}$$

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

$$P = \left(\sum_{i=0}^{N} \psi_i \psi_i^T \right)^{-1} \sigma^2 \tag{20}$$

If a model is overparameterized, then in the ideal case H will be a singular matrix (values close to zero), or in less ideal case is invertible, but has some eigenvalues that are significantly smaller than the rest (larger eigenvalues in terms of P).

$$eig(P)_i \gg eig(P)_j \tag{21}$$

This requires that a system is sufficiently excited; insufficiently excited systems will produce similar issues. Similarly one should be aware of the effect of the size of the noise.

Condition number

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

$$Cond(P) = \frac{|\lambda_{\max}|}{|\lambda_{\min}|}, \quad \Lambda = eig(P) \tag{22}$$

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

Then, if $Cond(P)$ is large, it is an indication of overparameterization. Example:

$$\text{model 1: } Cond(P_1) = 1000 \tag{23}$$

$$\text{model 2: } Cond(P_2) = 40 \tag{24}$$

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

Zeros and poles: cancellation?

If we have an overparameterized model, our external models might have zeros and poles that are close to each other.

$$y_t = H_{y,u}(q)u_t + H_{y,e}(q)e_t \tag{25}$$

Using the model structure used for the estimation, we can compute the poles and zeros of the estimated model. Using a linearization, we can derive the certainties of the zeros and poles:

$$\hat{p}_i = f_i(\hat{\theta}) \simeq f_i(\theta) + \frac{\partial f_i}{\partial \theta}\tilde{\theta}, \qquad \tilde{\theta} \in N(0, P) \tag{26}$$

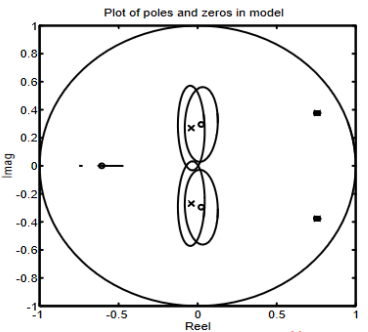
$$\hat{p}_i \in N\left(p_i, \frac{\partial f_i}{\partial \theta} P \left(\frac{\partial f_i}{\partial \theta}\right)^T\right) \tag{27}$$

Therefore, we can compute the confidence intervals in the complex plane. If the confidence intervals of a pole and a zero overlap, it is a strong indication that they cancel out.

Zeros and poles: an example of cancellation

$$(1 - 1.5q^{-1} + 0.7q^{-2})y_t = (1 - 0.5q^{-1})u_t + e_t \tag{28}$$

$$(1 - a_1q^{-1} + ... + a_4q^{-4})y_t = (b_0 + ... + b_3q^{-3})u_t + e_t \tag{29}$$



Residual Analysis

Let us now consider how to check whether a model is too simple or describes the system sufficiently well. For this, we consider the residuals of the model:

$$\text{measurement}(y) = \text{model}(\theta, u) + \text{residual}(\epsilon) \tag{30}$$

In a perfect model, the residual would have the following properties.

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

or stated with co-variance functions:

$$r_\epsilon(k) = E\{\epsilon_{t+k}\epsilon_t\} = \begin{cases} \sigma^2 & k = 0 \\ 0 & otherwise \end{cases} \qquad r_{\epsilon_t, u_t}(k) = E\{\epsilon_{t+k}u_t\} = 0 \tag{31}$$

An important part of validation is to use **one** data set for the **estimation**, and **one** for the **validation** (cross-validation).

Residual Analysis - mean and variance test

A simple approach is to test whether the distribution of the residual has the right mean and variance.

If the below holds, we can then say that our residual does not have zero mean.

$$|\bar{\epsilon}| > f^t_{1-\frac{\alpha}{2}}(M-1)\sqrt{\frac{S^2}{M}} \tag{32}$$

$$\bar{\epsilon} = \frac{1}{M}\sum_{i=1}^{M}\epsilon_i, \quad S^2 = \frac{1}{M-1}\sum_{i=1}^{M}(\epsilon_i - \bar{\epsilon})^2 \tag{33}$$

For the variance, we can test whether the variance is constant, by using two non-overlapping data sequences. The variance is time-varying if either of these two conditions hold,

$$\frac{S_1^2}{S_2^2} < f^F_{\alpha/2}(M_1, M_2) \text{ or } \frac{S_1^2}{S_2^2} > f^F_{1-\alpha/2}(M_1, M_2) \tag{34}$$

$$S_i^2 = \frac{1}{M_i}\sum_{j=1}^{M_i}\epsilon^2_{j+t_i} \tag{35}$$

Residual Analysis - sign test

A simple way to test for whiteness is by looking at how many times the residual changes sign. For a white data sequence of length M, the number of sign changes z should follow

$$z \in N\left(\frac{M-1}{2}, \frac{M-1}{4}\right) \tag{36}$$

We can reject this hypothesis if either of the below holds

$$z < \frac{M-1}{2} - \sqrt{\frac{M-1}{4}} f^N_{1-\frac{\alpha}{2}} \text{ or } z > \frac{M-1}{2} + \sqrt{\frac{M-1}{4}} f^N_{1-\frac{\alpha}{2}} \tag{37}$$

in other words, it is rejected if it is outside the confidence interval.

Residual Analysis - test of co-variance function

Another way to test for whiteness is to consider the autocovariance function, which must have the form:

$$r_\epsilon(k) = E\{\epsilon_{t+k}\epsilon_t\} = \begin{cases} \sigma^2 & k = 0 \\ 0 & \text{otherwise} \end{cases} \tag{38}$$

Sample-wise we can compute it and its normalized version as

$$\hat{r}_\epsilon(k) = \frac{1}{M}\sum_{t=1}^{M-k} \epsilon_{t+k}\epsilon_t, \qquad \hat{\rho}_\epsilon(k) = \frac{\hat{r}_\epsilon(k)}{\hat{r}_\epsilon(0)} \tag{39}$$

We can test the covariance at each time step, by evaluating the hypothesis

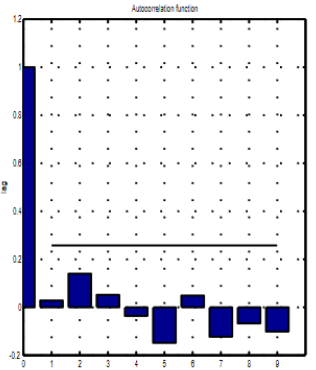
$$H_0: \quad \sqrt{M}\hat{\rho}_\epsilon(k) \in N(0,1), \text{ reject if: } |\hat{\rho}_\epsilon(k)| > \frac{f^N_{1-\frac{\alpha}{2}}}{\sqrt{M}} \tag{40}$$

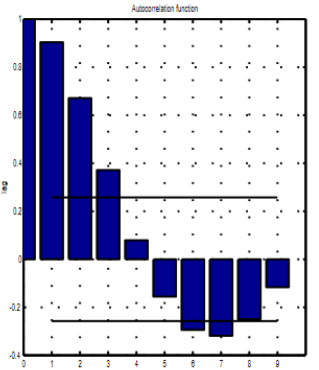
We can test if the covariance is zero for $k \neq 0$, by evaluating the hypothesis

$$H_0: \quad z = M\sum_{i=1}^{m} \hat{\rho}^2_\epsilon(i) \in \chi^2(m) \text{ reject if: } z > f^{\chi^2}_{1-\alpha}(m) \tag{41}$$

also called a Portmanteau test.

Residual Analysis - test of autocorrelation





Residual Analysis - cross-covariance function test

Similarly, we can test the assumption on the cross-covariance:

$$r_{\epsilon,u}(k) = E\{\epsilon_{t+k} u_t\} = 0 \tag{42}$$

$$\hat{r}_{\epsilon,u}(k) = \frac{1}{M} \sum_{i=1}^{M-k} \epsilon_{t+k} u_t, \qquad \hat{\rho}_{\epsilon,u}(k) = \frac{\hat{r}_{\epsilon,u}(k)}{\sqrt{\hat{r}_\epsilon(0)\hat{r}_u(0)}} \tag{43}$$

The marginal test of the cross-covariance is done by:

$$H_0: \quad \sqrt{M}\hat{\rho}_{\epsilon,u}(k) \in N(0,1) \text{ reject if: } |\hat{\rho}_{\epsilon,u}(k)| > \frac{f^N_{1-\frac{\alpha}{2}}}{\sqrt{M}} \tag{44}$$

Conversely, we can check if the covariance is zero for $k \neq 0$, by evaluating the hypothesis

$$H_0: \quad z = M \sum_{i=1}^{m} \hat{\rho}^2_{\epsilon,u}(i) \in \chi^2(m) \text{ reject if: } z > f^{\chi^2}_{1-\alpha}(m) \tag{45}$$

Residual Analysis - spectral density test

It is also possible to check the whiteness assumption by looking to the frequency domain. Consider the Fourier transformed residuals:

$$X(w_k) = \frac{1}{M} \sum_{t=1}^{M} \epsilon_t e^{jw_k t} \tag{46}$$

Hint: Matlab's `fft` can be used to compute $X(w_k)$. The estimated spectral density (periodogram) is then given by

$$\hat{\phi}(w_k) = |X(w_k)|^2 \tag{47}$$

Hint: Matlab's `etfe` can be used to compute $\hat{\phi}(w_k)$. If $\{x_t\}$ is white noise,

$$E\{\hat{\phi}(w_k)\} = 2\sigma^2 \tag{48}$$

Residual Analysis - model comparison tests

While it is best to use two data sets to estimate and validate a model, there are ways to evaluate a model using a single data set.
One way is to evaluate a model's coefficient of determination, R^2:

$$R^2 = \frac{J_0 - J(\hat{\theta})}{J_0} \tag{49}$$

$$J_0 = \frac{1}{2}\sum_{i=1}^{M}(y_i - \bar{y})^2, \quad J(\hat{\theta}) = \frac{1}{2}\sum_{i=1}^{M}\epsilon_i^2 \tag{50}$$

where $J(\hat{\theta})$ is known as the models loss-function. A perfect model thus gives an $R^2 = 1$, and the lower R^2 is the worse the model.
The loss function are occasionally formulated as:

$$W(\hat{\theta}) = \sum_{i=1}^{M}\epsilon_i^2, \qquad W_M(\hat{\theta}) = \frac{1}{M}\sum_{i=1}^{M}\epsilon_i^2 \tag{51}$$

The loss-functions are monotonically decreasing with the rise in model complexity, when using a single data-set.

Residual Analysis - model comparison: F-test

One type of single data test is the F-test; where the loss-function of two model classes $\mathcal{M}_1$ and $\mathcal{M}_2$ are compared statistically.

We are interested in the hypothesis that $\mathcal{M}_1 \in \mathcal{M}_2$ (assuming that $\mathcal{M}_{\text{true}} \in \mathcal{M}_1$), where $d_2 \geq d_1$ are the number of model parameters. Specifically, it means that if $\mathcal{M}_1 \in \mathcal{M}_2$, the loss-function $J_i = J_i(\hat{\theta})$ does not improve significantly by increasing the model size.

We can test this by considering the hypothesis:

$$H_0 : z = \frac{J_1 - J_2}{J_2} \times \frac{M - d_2}{d_2 - d_1} \in F(d_2 - d_1, M - d_2) \tag{52}$$

with a rejection being defined as

$$z > f^F_{1-\alpha}(d_2 - d_1, M - d_2) \tag{53}$$

So we use the test to check if we should increase the model order.

Residual Analysis - model comparison: Information Criteria

The loss functions can also be used to evaluate the model structure, both over and under parametrization. This is done considering an information criterion, and finding which model it reaches its minimum.

There exist several information criteria:

1. Akaike's Information Criterion (AIC); tends towards higher complexity
2. Bayesian Information Criterion (BIC);
3. Akaike's Final Prediction Error (FPE) Criterion; expresses the variance of the prediction error, also $FPE \rightarrow AIC, M \gg d$,

The criteria is computed as follows.

$$AIC = \left(1 + \frac{2d}{M}\right) W_M \tag{54}$$

$$BIC = \left(1 + \frac{\log(M)d}{M}\right) W_M \tag{55}$$

$$FPE = \frac{M+d}{M-d} W_M = \left(1 + \frac{2d}{M-d}\right) W_M \tag{56}$$

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

Residual Analysis - model comparison: Information Criteria

The criteria is computed as follows.

$$AIC = \left(1 + \frac{2d}{M}\right) W_M \tag{57}$$

$$BIC = \left(1 + \frac{\log(M)d}{M}\right) W_M \tag{58}$$

$$FPE = \frac{M+d}{M-d} W_M = \left(1 + \frac{2d}{M-d}\right) W_M \tag{59}$$

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

What happens to BIC as $M \to \infty$?

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

Validation examples

Let us jump to some Matlab examples:

- Model reduction: Zero-pole cancellation.
- Singular analysis.
- Insignificant parameter.
- Model Comparison.

Closed-Loop Identification - why?

Given the methods for open-loop identification, why would we want to consider closed-loop identification?

1. The open-loop system is unstable.
2. The process is running (production / economics).
3. Safety reasons.
4. Adaptive control.

Closed-Loop Identification - Pitfalls

Let us consider some of the potential issues of closed-loop identification. Consider the following system and control law

$$y_t + ay_{t-1} = bu_{t-1} + e_t, \quad e_t \in N_{iid}(0, \sigma^2) \tag{60}$$

$$u_t = -fy_t \tag{61}$$

The closed-loop system is then given in the time-domain by

$$y_t + (a + bf)y_{t-1} = e_t \tag{62}$$

or, in the frequency domain using spectral analysis, by

$$\phi_{yu}(w) = G(e^{jw})\phi_u(w), \qquad \hat{G}(e^{jw}) = \frac{\hat{\phi}_{yu}(w)}{\hat{\phi}_u(w)} \tag{63}$$

where the closed-loop transfer functions from the noise to the output and the control are given by

$$H_y(z) = \frac{1}{1 + (a + bf)z^{-1}}, \quad H_u(z) = \frac{-f}{1 + (a + bf)z^{-1}} \tag{64}$$

Closed-Loop Identification - Pitfalls

If we start with the spectral analysis:

$$\phi_u(w) = \Phi_u(e^{jw}), \qquad \Phi_u(z) = H_u(z)H_u(z^{-1})\sigma^2 \tag{65}$$

$$\phi_{yu}(w) = \Phi_{yu}(e^{jw}), \qquad \Phi_{yu}(z) = H_y(z)H_u(z^{-1})\sigma^2 \tag{66}$$

the estimated closed-loop system is then given by

$$\hat{G}(w) = \frac{\hat{\phi}_{yu}(w)}{\hat{\phi}_u(w)} = \frac{H_y(w)}{H_u(w)} = -\frac{1}{f} \tag{67}$$

only informing us of the controller. If we consider estimation of the CL:

$$y_t + (a + bf)y_{t-1} = e_t \tag{68}$$

we find that the fitting parameters are not unique:

$$a = a_0 + \gamma f \tag{69}$$

$$b = b_0 - \gamma \tag{70}$$

where γ is arbitrary, and the controller therefore is too simple.

Closed-Loop Identification - Pitfalls

If we consider estimation of the CL:

$$y_t + (a + bf)y_{t-1} = e_t \tag{71}$$

we find that the fitting parameters are not unique:

$$a = a_0 + \gamma f \tag{72}$$
$$b = b_0 - \gamma \tag{73}$$

where γ is arbitrary, and the controller therefore is too simple.

What would you do to prevent this issue of non-uniqueness?

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

Closed-Loop Identification

The question is then, how complex should a controller be? Let us consider the general system

$$A(q)y_t = q^{-k}B(q)u_t + C(q)e_t, \quad e_t \in N_{iid}(0, \sigma^2) \tag{74}$$

with the control given by

$$u_t = -\frac{S(q)}{R(q)}y_t \tag{75}$$

The closed-loop system is then given by

$$\Big(A(q)R(q) + q^{-k}B(q)S(q)\Big)y_t = R(q)C(q)e_t \tag{76}$$

A controller is adequately complex if the order of the closed-loop system is higher than the number of parameters to be estimated.

$$\max(n_r - n_b, n_s + k - n_a) \geq 1 + n_p \tag{77}$$

where n_p is the number of common factors in RC and $AR + q^{-k}BS$

Closed-Loop Identification - examples:

Let us consider the system from before:

$$y_t + ay_{t-1} = bu_{t-1} + e_t, \quad e_t \in N_{iid}(0, \sigma^2) \tag{78}$$

$$u_t = -fy_t \tag{79}$$

We can then see the orders of each polynomials is

$$n_a = 1, \quad n_b = 0, \quad k = 1, \quad n_r = 0, \quad n_s = 0, \quad n_p = 0 \tag{80}$$

We can then evaluate the controller:

$$\max(n_r - n_b, n_s + k - n_a) \geq 1 + n_p \tag{81}$$

$$\max(0 - 0, 0 + 1 - 1) \geq 1 + 0 \tag{82}$$

$$0 \geq 1 \tag{83}$$

The order is too low to identify the system.

What order should the controller have?

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

Closed-Loop Identification - examples: Minimum Variance

Let us consider the general ARMAX system:

$$A(q)y_t = q^{-k}B(q)u_t + C(q)e_t, \quad e_t \in N_{iid}(0, \sigma^2) \tag{84}$$

and let us utilize a minimum variance controller:

$$J = E\{y_{t+k}^2\}, \qquad y_{t+k} = \frac{1}{C}(BGu_t + Sy_t) + Ge_{t+k} \tag{85}$$

The controller and closed-loop system are then given by

$$u_t = -\frac{S}{BG}y_t \qquad (AG + q^{-k}S)y_t = CGe_t \tag{86}$$

with the orders of R,S and n_p being

$$n_r = n_b + k - 1, \quad n_s = n_a - 1, \quad n_p = n_c \tag{87}$$

We can now evaluate the controller:

$$\max(n_r - n_b, n_s + k - n_a) = \max(k-1, k-1) \geq 1 + n_c \tag{88}$$

$$k \geq n_c + 2 \tag{89}$$

Closed-Loop Identification - informative Data

We say that the examples were not informative enough (for $k < n_c + 2$).
As in the open-loop case, we need the system to be sufficiently informative in order to estimate the system.
As earlier, a data set z_t is sufficiently informative if

$$\overline{E}\{||(\mathcal{M}_1 - \mathcal{M}_2)z_t||^2\} = 0 \quad \Rightarrow \quad \mathcal{M}_1(w) \equiv \mathcal{M}_2(w) \tag{90}$$

such that we can distinguish two models $\mathcal{M}_i$ of the model set $\mathcal{M}$ from each other.

Generally the closed loop experiment is informative if the reference w_t (or another probe signal) is persistently exciting.

Similarly, time-invariant, nonlinear or higher order feedback controllers should also provide informative experiments.

Closed-Loop Identification - issues

As opposed to the open-loop identification, closed-loop identification has some potential pitfalls to be aware of:

1. The closed-loop experiment may be non-informative even if the input in itself is persistently exciting. Reason: the controller might be too simple.
2. Spectral analysis applied in a straightforward manner will give erroneous results. The estimate of G will converge to

$$G^* = \frac{G_0\phi_w - F\phi_v}{\phi_w + |F|^2\phi_v}, \quad y = Gu + v \tag{91}$$

3. Correlation analysis will give a biased estimate of the impulse response because the assumption that $\bar{E}\{u_t v_{t-\tau}\} = 0$ is violated.
4. OE methods give unbiased estimates of G in open-loop experiments, even if the additive noise (v) is not white. This is not true in closed-loop.

Closed-Loop Identification - A strength

If we consider the PE methods, they will give consistent estimates of the closed-loop system if

1. The data is informative.
2. The model set contains the true system ($\mathcal{S} \in \mathcal{M}$).

meaning we only need to consider the data set $[y, u]$, and proceed as with the open-loop case.

Closed-Loop Identification - Time-varying

If we consider a time-varying control law without a reference, we will need m different LTI Controllers:

$$u_t = F_i y_t, \quad i = 1, ..., m \tag{92}$$

In the SISO case, it is sufficient to use 2 controllers for which

$$F_1(e^{jw}) - F_2(e^{jw}) \neq 0 \quad \forall w \tag{93}$$

Closed-Loop Identification - Direct approach

The system is identified in exactly the same way as in open-loop identification, using the data set $[y, u]$ and ignoring any information about the feedback structures.

Strengths:

1. It works regardless of the complexity of the controller and requires no knowledge about the character of the feedback.
2. No special algorithms or software are required.
3. Consistency and optimal accuracy are obtained if model structure contains the true system.
4. Unstable systems can be handled without problems (as long as the closed loop and the predictor are stable).

Drawbacks:

1. We need good noise models. (Not a problem if true system (G,H) is contained in model structure).
2. If noise model is incorrect (fixed incorrectly or not containing the true noise model) bias in G will be introduced.

Closed-Loop Identification - Indirect Approach

In the indirect approach, we estimate the closed-loop transfer functions:

$$y_t = Gu_t + He_t, \quad u_t = w_t - F_y y_t \tag{94}$$

$$y_t = G_{cl} w_t + H_{cl} e_t, \quad G_{cl} = GS, \quad H_{cl} = HS, \quad S = \frac{1}{1 + F_y G} \tag{95}$$

From these and the knowledge of the control, we can derive the identified system:

$$\hat{G} = \frac{\hat{G}_{cl}}{1 - F_y \hat{G}_{cl}}, \quad \hat{H} = \hat{H}_{cl}(1 + F_y \hat{G}) \tag{96}$$

Properties of the indirect approach:

1. $+$ Any (open-loop) method such as spectral analysis, instrumental variable, subspace and prediction error methods can be applied.
2. $-$ Any error in F_y will transported directly to the estimate of the model (e.g., saturation, manual operation).

Closed-Loop Identification - joint input-output

The last approach considers the full closed-loop description:

$$y_t = GSw_t + HSe_t + GSz_t = GSw_t + v_{1,t} \tag{97}$$

$$u_t = Sw_t - FSHe_t + Sz_t = Sw_t + v_{2,t} \tag{98}$$

where z_t is a partial unknown signal part of u_t.

In this approach, we utilize the structure of both the input and output to estimate the closed-loop plus the sensitivity function S.

Closed-Loop Identification - joint input-output V1

The first version, takes the correlation between v_1 and v_2 into account:

$$\begin{bmatrix} y_t \\ u_t \end{bmatrix} = S \begin{bmatrix} G \\ 1 \end{bmatrix} w_t + S \begin{bmatrix} H & G \\ -FH & 1 \end{bmatrix} \begin{bmatrix} e_t \\ z_t \end{bmatrix} = \mathcal{G} w_t + \mathcal{H} \begin{bmatrix} e_t \\ z_t \end{bmatrix} \tag{99}$$

Using the variance of $[e_t, z_t]^T$, we can estimate the system parameters using, e.g., a ML or PEM method:

$$J = \sum_{i=1}^{t} \epsilon_i^T R^{-1} \epsilon_i, \quad R = Var\left\{ \begin{bmatrix} e_i \\ z_i \end{bmatrix} \right\}, \quad \epsilon_i = \mathcal{H}^{-1} \left[\begin{bmatrix} y_i \\ u_i \end{bmatrix} - \mathcal{G} w_i \right] \tag{100}$$

This is essentially the direct approach extended to controller estimation.

Closed-Loop Identification - joint input-output V2

In the second version, we disregard the correlation:

$$\begin{bmatrix} y_t \\ u_t \end{bmatrix} = \begin{bmatrix} G_{cl} \\ G_{uw} \end{bmatrix} w_t + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} \tag{101}$$

For our estimation we minimize the cost:

$$J = \frac{1}{\sigma_1^2} \sum_{i=1}^{t} (y_i - G_{cl} w_i)^2 + \frac{1}{\sigma_2^2} \sum_{i=1}^{t} (u_i - G_{uw} w_i)^2, \tag{102}$$

The system can then be determined by

$$\hat{G} = \frac{\hat{G}_{cl}}{\hat{G}_{wu}} \approx \frac{\hat{G}\hat{S}}{\hat{S}} \tag{103}$$

Given these polynomials are estimated the cancellation is not perfect, making $\hat{G}$ higher order than it should be.
A trick against this, is to use independent parametrization of G and S: $G(\theta)$ and $S(\eta)$

Closed-Loop Identification - joint input-output V2 - 2 step

For the independently parametrized estimation, we consider the cost:

$$J = \beta \frac{1}{\sigma_1^2} \sum_{i=1}^{t} (y_i - G(\theta)S(\eta)w_i)^2 + \sum_{i=1}^{t} (u_i - S(\eta)w_i)^2, \tag{104}$$

We then do the estimation in two steps

1. first for $\beta = 0$, to estimate the parameters of $S(\eta)$.
2. then use $\hat{u}_t = S(\hat{\eta})w_t$ to estimate $G(\theta)$ from $y_t = G(\theta)\hat{u}_t + v_{1,t}$

One previously suggested parametrization of S is the non-causal filter:

$$S(\eta) = \sum_{i=-m}^{m} s_k q^{-k} \tag{105}$$