227-0689-00L: System Identification HS2023

Michele Zaffalon (Bruker BioSpin AG)

Yannic Hofmann (ETHZ)

January 15, 2024

Unreviewed notes for Prof. Smith's class. Use at own risk!

Contents

Chapter 1

Least-Squares Estimation

A word about notation: in class we used Y and Φ for the measurements and regressor matrix and ϵ for the (possibly correlated) noise. However because the noise is Gaussian, the equations below show that the terms scaled by the Cholesky decomposition C of the noise correlation matrix R are the relevant ones. With the scaling, the scaled noise e becomes uncorrelated and has variance $\sigma^2 = 1$. I personally like this approach.

Consider the model

$$\tilde{Y} = \tilde{\Phi}\theta_0 + \tilde{\epsilon} \tag{1.1}$$

where $\tilde{Y} = \begin{bmatrix} y_0 & \dots & y_{N-1} \end{bmatrix}^\top$ is a vector containing the measurements $\{y_0, \dots, y_{N-1}\}$, $\tilde{\Phi} \in \mathbb{R}^{N \times p}$ is called the *regressor* and $\theta \in \mathbb{R}^p$ is the model parameter to be estimated; p is the model order. The noise vector $\tilde{\epsilon} = \begin{bmatrix} v_0 & \dots & v_{N-1} \end{bmatrix}^\top$ has zero mean $\mathbb{E}\{\tilde{\epsilon}\} = 0$ and covariance[1] $\mathbb{E}\left\{\tilde{\epsilon}\tilde{\epsilon}^\top\right\} = R$, a symmetric positive definite matrix.

The maximum likelihood (ML) probability[2] requires us to minimize

$$(\tilde{Y} - \tilde{\Phi}\theta)^\top R^{-1} (\tilde{Y} - \tilde{\Phi}\theta) = ||C^{-1}(\tilde{Y} - \tilde{\Phi}\theta)||_2^2 = ||Y - \Phi\theta||_2^2 \tag{1.2}$$

[1] Note also that while the matrix $vv^\top$ has rank 1, the noise covariance matrix R for Gaussian distributed white noise is full rank because it is the sum of random vectors that span the whole space.

[2] The expectation value of eq. (1.2) is $N - p$. This can be seen using eq. (1.5)

$$Y - \Phi\hat{\theta}_{\mathrm{LS}} = (\Phi\theta_0 + e) - \Phi\left(\theta_0 + K\Phi^\top e\right) = \left(I - \Phi K\Phi^\top\right) e.$$

The average

$$\mathbb{E}\left\{\left|\left|Y - \Phi\hat{\theta}_{\mathrm{LS}}\right|\right|_2^2\right\} = \mathbb{E}\left\{e^\top \left(I - \Phi K\Phi^\top\right)^\top \left(I - \Phi K\Phi^\top\right) e\right\} = \mathbb{E}\left\{e^\top \left(I - \Phi K\Phi^\top\right) e\right\}.$$

where C is the Cholesky decomposition of $R = CC^\top$ (and $R^{-1} = \left(C^{-1}\right)^\top C^{-1}$), $Y = C^{-1}\tilde{Y}$, $\Phi = C^{-1}\tilde{\Phi}$ and $e = C^{-1}\tilde{\epsilon}$. Note that the scaled noise e is uncorrelated and has unit variance:

$$\mathbb{E}\left\{ee^\top\right\} = \mathbb{E}\left\{C^{-1}\tilde{\epsilon}\tilde{\epsilon}^\top\left(C^{-1}\right)^\top\right\} = C^{-1}R\left(C^{-1}\right)^\top = \mathbb{I}.$$

The *mathematical* solution to eq. (1.2) is found by setting the gradient of the expression with respect to θ to zero, which gives the normal equation

$$\left(\Phi^\top\Phi\right)^\top\hat{\theta}_{\text{LS}} = \Phi^\top Y \;\rightarrow\; \hat{\theta}_{\text{LS}} = \left(\Phi^\top\Phi\right)^{-1}\Phi^\top Y. \tag{1.3}$$

The solution exists provided Φ has full column rank[3]: $\text{rank}\left(\Phi\right) = p$. When this is the case, the system is said to be *persistently excited*, see Sect. 4.0.4, and when given the freedom, one chooses Φ so that it is well conditioned.

Had we not scaled $\tilde{Y}$ and $\tilde{\Phi}$ by the noise covariance, the solution would have been

$$\hat{\theta}_{\text{LS}} = \left(\tilde{\Phi}^\top R^{-1}\tilde{\Phi}\right)^{-1}\tilde{\Phi}^\top R^{-1}\tilde{Y}.$$

Numerically one should not form the normal equation directly because it squares the condition number of Φ and rely either on the QR decomposition or on the SVD to solve eq. (1.2). This is taken care automatically by MATLAB when using the backslash $\backslash$ operator

$$\hat{\theta}_{\text{LS}} = \Phi\backslash Y. \tag{1.4}$$

The term $\mathbb{E}\left\{e^\top e\right\}$ evaluates to N. The other term evaluates to p: using $\Phi = U\Sigma V^\top$,

$$(\Phi^\top\Phi)^{-1} = \left(V\Sigma^\top U^\top U\Sigma V^\top\right)^{-1} = V\left(\Sigma^\top\Sigma\right)^{-1}V^\top$$

and

$$\Phi V\left(\Sigma^\top\Sigma\right)^{-1}V^\top\Phi^\top = U\Sigma\left(\Sigma^\top\Sigma\right)^{-1}\Sigma^\top U$$

where the term between the two U is a tall matrix of dimension $N \times p$ with ones on the top block's main diagonal (of dimension $p \times p$) and zeros in the bottom block.

[3]The usual warning holds for the rank. Instead we want to have the matrix Φ with the smallest condition number for the estimate to be numerically stable, which is a stronger condition than full rank.

1.1 Bias, Covariance and MSE of the Least Squares Estimation

The linear estimator eq. (1.3) is unbiased[4] (but see also Sect. 2.1.2):

$$\mathbb{E}\left\{\hat{\theta}_{\mathrm{LS}}\right\} = \theta_0.$$

The covariance[5]

$$\mathrm{cov}\left(\hat{\theta}_{\mathrm{LS}}\right) = \left(\Phi^\top \Phi\right)^{-1} = \left(\tilde{\Phi}^\top R^{-1} \tilde{\Phi}\right)^{-1}. \tag{1.6}$$

Lastly, we consider the mean squared error

$$\mathrm{MSE}\left(\hat{\theta}_{\mathrm{LS}}\right) = \underbrace{\left|\left|\mathrm{Bias}\left(\hat{\theta}_{\mathrm{LS}}\right)\right|\right|_2^2}_{=0} + \mathrm{tr}\left(\mathrm{cov}\left(\hat{\theta}_{\mathrm{LS}}\right)\right)$$

which reduces to N for the case of uncorrelated noise (I am not sure about this anymore).

1.1.1 Geometric Interpretation of Least-Squares

The least squares problem

$$||b - Ax||_2$$

has the following geometric interpretation: the solution is that for which the residuals $v \doteq b - Ax$ are outside (i.e. orthogonal) of the space spanned by A.

[4] Recalling that $\Phi^\top \Phi$ is a symmetric matrix, letting $K \doteq (\Phi^\top \Phi)^{-1}$ and using eq. (1.1) into eq. (1.3), we obtain

$$\hat{\theta}_{\mathrm{LS}} = K\Phi^\top(\Phi\theta_0 + e) = \theta_0 + K\Phi^\top e \tag{1.5}$$

from which $\mathbb{E}\left\{\hat{\theta}_{\mathrm{LS}}\right\} = \theta_0$ since $\mathbb{E}\{e\} = 0$. Moreover

$$\begin{aligned}
\mathrm{cov}\left(\hat{\theta}_{\mathrm{LS}}\right) &\doteq \mathbb{E}\left\{\left(\hat{\theta}_{\mathrm{LS}} - \mathbb{E}\left\{\hat{\theta}_{\mathrm{LS}}\right\}\right)\left(\hat{\theta}_{\mathrm{LS}} - \mathbb{E}\left\{\hat{\theta}_{\mathrm{LS}}\right\}\right)^\top\right\} \\
&= \mathbb{E}\left\{\left(K\Phi^\top e\right)\left(K\Phi^\top e\right)^\top\right\} \\
&= \mathbb{E}\left\{K\Phi^\top e e^\top \Phi K\right\} = K\Phi^\top \mathbb{E}\left\{ee^\top\right\}\Phi K = K.
\end{aligned}$$

[5] Is there a way to understand the form of covariance matrix without going through the calculation?

In other words, we require[6] the scalar product $\langle Az, v\rangle$ to be zero for all z:

$$\begin{aligned} 0 = \langle Az, v\rangle = (Az)^\top (b - Ax) = z^\top \left(A^\top b - A^\top Ax\right) \qquad \forall z \\ \rightarrow A^\top Ax = A^\top b \end{aligned}$$

which is the normal equation.

1.2 Random Notes: The Covariance Matrix and the Choice of the Measurement Points

These are my considerations that are not part of the lecture.

- The *off-diagonal* elements of the covariance matrix $\operatorname{cov}\left(\hat{\theta}_{\mathrm{LS}}\right)$ eq. (1.6) represent the correlations between the errors of the variables θ. It is therefore not justified to discard them and take θ's standard deviations as the square root of $\operatorname{cov}(\theta)$'s diagonal elements because one discards the correlations: the ball of probability is in general an ellipse with the axes not parallel to the variable directions.

- Given the freedom to choose the measurement points, is there a "best" way of placing them? Is this done in practice?

  In the context of experiment design, this is done.

  There are two factors that determine the covariance matrix: the choice of basis and the choice of points. The choice of basis is determined by the variables that one wants to extract: a linear transformation between one basis and the other will also transform the covariance matrix and the only concern may be the numerical stability (although one should expect that the measurement errors dominate).

  Determining the position of the best measurement points by minimizing one (or more elements of the covariance) is in general a non-convex problem.

[6] I believe the proof given in class is not correct: to span the full column space of A, one has to multiply by a generic vector z; in class $z = x$ was taken.

Chapter 2

Regularized FIR Models

In Sect. 1.1, we have seen that the solution

$$\hat{\theta}_{\mathrm{LS}} = \arg\min_{\theta} ||Y - \Phi\theta||_2^2$$

of the least squares problem is unbiased. We can however choose to have a biased estimate to reduce the mean square error $\mathbb{E}\left\{||\hat{\theta} - \theta_0||_2^2\right\}$. This can be achieved if we modify the minimization problem by adding a regularization term

$$||Y - \Phi\theta||_2^2 + \gamma\theta^\top P^{-1}\theta \tag{2.1}$$

where γP^{-1} is a positive definite matrix. P is called the *kernel* or the regularization matrix[1].

Regularization prevents overfitting, reduces the sensitivity to noise and can improve the estimate by a proper choice of the kernel (*e.g.* for instance if the system is known to be stable, this information can be used to improve the estimation.)

The closed form mathematical solution to eq. (2.1) is given by

$$\hat{\theta}^{\mathrm{R}} = \left(\Phi^\top\Phi + \gamma P^{-1}\right)^{-1}\Phi^\top Y = \left(P\Phi^\top\Phi + \gamma\mathbb{I}\right)^{-1}P\Phi^\top Y. \tag{2.2}$$

[1]When $P = \mathbb{I}_N$ the technique is called ridge regression, otherwise it goes under the name of Tikhonov regularization. Compared to the ridge regression which only tries to decrease $||\theta||^2$, the Tikhonov regularization can use information about the system, see Sect. 2.1.1.

2.0.1 The James-Stein Estimator

For the problem $Y = \theta_o + e$, the James-Stein estimator is a biased estimator that has a smaller MSE[2] than the least squares. This is achieved by "shrinking the estimate towards the origin" [4, page 3]:

$$\hat{\theta}^{\mathrm{JS}} = (1 - r)\hat{\theta}_{\mathrm{LS}}, \qquad r \doteq \frac{(N-2)\sigma^2}{||Y||^2}.$$

where $N = p$. It can be cast into the ridge regression problem

$$||Y - \theta||^2 + \gamma||\theta||^2, \qquad \gamma = \frac{(N-2)\sigma^2}{||Y||^2 - (N-2)\sigma^2}.$$

2.1 Bias, Covariance and MSE of the Regularized Least Squares Estimation

We assume uncorrelated noise with constant variance 1: $\mathbb{E}\left\{ee^\top\right\} = \mathbb{I}_N$.

[2] I am not sure whether the James-Stein estimator is obtained by minimizing the MSE. We have $\hat{\theta}^{\mathrm{JS}} = (1-r)\hat{\theta}_{\mathrm{LS}} = (1-r)(\theta_0 + e)$; the MSE is

$$\mathbb{E}\left\{(\hat{\theta}^{\mathrm{JS}} - \theta_0)(\hat{\theta}^{\mathrm{JS}} - \theta_0)^\top\right\} = (1-r)^2||\theta_0||_2^2 + r^2 \underbrace{\mathbb{E}\left\{||e||_2^2\right\}}_{=N\sigma^2}$$

which has a minimum at

$$r = \frac{N\sigma^2}{||\theta_0||^2 + N\sigma^2}.$$

We also have the identity $\mathbb{E}\left\{||Y||_2^2\right\} = ||\theta_0||_2^2 + N\sigma^2$ but the James-Stein estimator contains $||Y||_2^2$ at the denominator and not $\mathbb{E}\left\{||Y||_2^2\right\}$.

The estimate eq. (2.2) has bias[3], covariance[4] and MSE

$$\begin{aligned}
\text{Bias}\left(\hat{\theta}^{\text{R}}\right) &= -\left(\Phi^\top\Phi + \gamma P^{-1}\right)^{-1}\gamma P^{-1}\theta_0 \\
\text{cov}\left(\hat{\theta}^{\text{R}}\right) &= \left(\Phi^\top\Phi + \gamma P^{-1}\right)^{-1}\Phi^\top\Phi\left(\Phi^\top\Phi + \gamma P^{-1}\right)^{-1} \\
\text{MSE}\left(\hat{\theta}^{\text{R}}\right) &= \left|\left|\text{Bias}\left(\hat{\theta}^{\text{R}}\right)\right|\right|_2^2 + \text{tr}\left(\text{cov}\left(\hat{\theta}^{\text{R}}\right)\right)
\end{aligned}$$

2.1.1 Choice of the Regularization Matrices

The MSE is minimized by this choice of parameters

$$\gamma^\star = 1, \quad P^\star = \theta_0\,\theta_0^\top. \tag{2.3}$$

The optimal regularization matrix $P^\star$ is unknown[5] because it depends on the unknown θ_0 but the approximate knowledge of the solution helps to construct a "good" P.

> Regularization functions are designed to encode prior knowledge on the unknown system; some choices will result in better estimates. The choice can be either subjective or based on cross-validation and different regularization functions/kernels give rise to different model classes. Technically there is no wrong regularization function, but there is an optimal one (in the sense of minimum MSE).

[3] Let $K = \left(\Phi^\top\Phi + \gamma P^{-1}\right)^{-1}$ so that $\hat{\theta}^{\text{R}} = K\Phi^\top Y$. Plugging $Y = \Phi\theta_0 + e$, we have $\mathbb{E}\left\{\hat{\theta}^{\text{R}}\right\} = K\Phi^\top\Phi\theta_0$. The bias is

$$\mathbb{E}\left\{\hat{\theta}^{\text{R}}\right\} - \theta_0 = (K\Phi^\top\Phi - \mathbb{I})\theta_0 = K(\Phi^\top\Phi - K^{-1})\theta_0 = -\gamma K P^{-1}\theta_0.$$

[4] With the same notation as before, we have $\hat{\theta}^{\text{R}} - \mathbb{E}\left\{\hat{\theta}^{\text{R}}\right\} = K\Phi^\top e$. The covariance is

$$\text{cov}\left(\hat{\theta}^{\text{R}}\right) = \mathbb{E}\left\{\left(K\Phi^\top e\right)\left(K\Phi^\top e\right)^\top\right\} = K\Phi^\top\mathbb{E}\left\{ee^\top\right\}\Phi K = K\Phi^\top\Phi K.$$

[5] $P^\star = \theta_0\,\theta_0^\top$ has rank 1 and is therefore not invertible. To get a well-defined problem, the inverse P^{-1} is replaced by the Moore-Penrose pseudoinverse P^+. It turns out that the solution of the problem defined with $P^\dagger$ is also equal to $(P\Phi^\top\Phi + \gamma\mathbb{I})^{-1}P\Phi^\top Y$. [Mohamed Abdalmoaty on Moodle]

> A quadratic regularization function based on the TC kernel is well-suited for exponentially decaying pulse responses of stable linear systems. In that case, a high order FIR model is used to estimate the first significant part of the pulse response of the unknown system, ignoring the tail. An FIR system is always stable, but its coefficients may not be exponentially decaying. An impulse response experiment may be used to reveal some properties of the system if the system and experimental configuration allow it. [Mohamed Abdalmoaty on Moodle]

A commonly used regularization matrix is the diagonally-correlated (DC) kernel

$$[P]_{ij} = c\alpha^{\frac{i+j}{2}}\rho^{|i-j|}$$

with $0 \leq c$, $0 \leq \alpha \leq 1$ and $-1 \leq \rho \leq 1$: ρ describes correlations α and the decays.

A simpler version of it is the tuned-correlated (TC) kernel[6] where one lets $\rho = \sqrt{\alpha}$ in the DC kernel

$$[P]_{ij} = c\alpha^{\max\{i,j\}} = \begin{bmatrix} \alpha & \alpha^2 & \alpha^3 & \dots & \alpha^n \\ \alpha^2 & \alpha^2 & \alpha^3 & \dots & \alpha^n \\ \alpha^3 & \alpha^3 & \alpha^3 & \dots & \alpha^n \\ \dots & & & & \\ \alpha^n & \alpha^n & \alpha^n & \dots & \alpha^n \end{bmatrix}$$

The top left entries are large, the bottom right small. Since the inverse enters the minimization problem (cost function), the top left part is small whereas the bottom down are large and induce a larger cost.

The inverse of the DC kernel is a tridiagonal with the explicit form

$$[P^{-1}]_{ij} = \frac{c_{ij}}{1-\rho^2}(-1)^{i+j}\alpha^{-\frac{i+j}{2}}\rho^{|i-j|}, \qquad c_{ij} = \begin{cases} 1+\rho^2 & \text{if } i=j=2,\dots,p-1 \\ 0 & \text{if } |i-j|>1 \\ 1 & \text{otherwise} \end{cases}$$

[6] Despite having entries that are exponential, the TC kernel is *not* $\theta\theta^\top$

$$\theta\theta^\top = \begin{bmatrix} a & a^2 & a^3 & \dots \\ a^2 & a^3 & \dots & \\ a^3 & \dots & & \\ \dots & & & \end{bmatrix} \quad \text{with } \theta = \begin{bmatrix} a & a^2 & \dots \end{bmatrix}$$

coming from the truncation of the IIR filter $\frac{z^{-1}}{1-az^{-1}} \approx z^{-1}(1 + az^{-1} + a^2z^{-2} + \dots)$.

2.1.2 Estimation Bias

> One cannot say that the linear least-squares method is unbiased in general. There are three elements at play: the data set, the model used, and the estimation method. Least-squares is just the estimation method. To check if it is unbiased, an assumption has to be made on the true system that generated the data, and the model used to get the closed form expression of the least-squares estimator. [Mohamed Abdalmoaty on Moodle]

In the following, the input u is known, the measurement noise e has zero-mean, the data is generated as $Y = \Phi\theta_0 + e$ with $\theta_0 \in R^p$; we estimate $\theta \in R^q$ in the linear regression model $Y = \Phi_q\theta + e$; and we use the least-squares estimation method: then $\hat{\theta} = \left(\Phi_q^\top \Phi_q\right)^{-1} \Phi_q^\top Y$.

Here some examples of source of biased estimate:

- The model does not match the generating data. One underestimates the length of the response by choosing the model order too small, $q < p = \tau_{\max}$: this is the truncation error of Sect.. (see slide 10–10). Indeed we are trying to fit the data with a model of order q but the model has order p. The estimate is biased because

$$\mathbb{E}\left\{\hat{\theta}\right\} = \left(\Phi_q^\top \Phi_q\right)^{-1} \Phi_q^\top \Phi\theta_0 \neq \theta_0.$$

  This means the trade-off bias-variance can also be done with LS by varying the order parameter, but to a lower extent than using regularization where one has extra parameters available.

  Note that here the regressor involves only the known (noise-free) input u;

- The data comes from an ARX system with $p = n + m$ parameters and $\theta_0 \in R^{n+m}$ is the true parameter. As opposed to the case above, here Φ involves also the random outputs: for this reason the resulting LS estimator is *biased* even though we have a correct model order/structure (*i.e.*, the model that matches the data generating system);

- Finally, even with the correct order parameter, one can *on purpose* bias $\hat{\theta}$ by adding the regularization term to $||Y - \Phi\theta||_2$.

2.1.3 Why the LS May Perform Badly

Least-squares performs badly when the regression matrix Φ is ill-conditioned[7], as a result of not being persistently excited and a small perturbation of the measurement data Y can have a large impact on the estimate $\hat{\theta}_{\text{LS}}$. Regularization helps in these cases. (Is there an easy way to show that $\Phi^\top \Phi + \gamma P^{-1}$ has a larger minimum singular value?)

2.2 Cross-Validation Methods

They are methods to select the discrete model orders (in least-squares or ARX) or the regularization parameters (*e.g.* γ and the kernel parameter α in regularized least squares).

A popular method is called hold-out cross-validation and consists of the following steps:

1. split the data into two (equal but not necessarily) non-overlapping parts, one for estimation and the other for validation. The way in which the data is split depends on the type of data: for data generated by dynamical systems, where the data is sequential in time, the order is important. Here the first N_{id} pairs of control inputs and output values $\{u(k), y(k)\}$ are used for identification and the remaining for validation. For static data, the order is not important: one could choose the odd-indexed pairs for identification and the even-indexed pairs for validation;

2. use the estimation data to estimate a model for each mode structure or each regularization parameter. In the case of the continuous parameter γ, one estimates the model on a set of gridded values. If there are two continuous parameters, *e.g.* γ and α, the gridding is done for both parameters;

3. select the model structure / regularization parameters that give a model with least prediction error on validation data;

[7] The condition number of a matrix is defined as the ratio between the largest σ_1 and the smallest σ_n singular value of the matrix

$$\text{cond}(A) = \frac{\sigma_1}{\sigma_n}.$$

Note that from a numerical point of view, the matrix being full rank is not a guarantee that the LS problem can be stably solved.

4. use the selected model structure / regularization parameters and the *complete* data set to estimate a final model.

Numerical example in `10_lect/regularization.m`, for the moment limited to ridge regression.

2.3 Numerical Solution of the Tikhonov Regularization Problem

For the numerical solution I am aware of three methods:

1. Solve directly the normal equation

$$\left(\Phi^\top \Phi + \gamma P^{-1}\right) \hat{\theta} = \Phi^\top y \tag{2.4}$$

   using MATLAB's backslash operator: although this squares Φ's condition number when constructing $\Phi^\top \Phi$, at least it does not require to explicitly construct the inverse;

2. Use the generalized SVD decomposition. We first manipulate eq. (2.1) to obtain

$$||y - \Phi\theta||_2^2 + ||D\theta||_2^2$$

   by letting D be the Cholesky decomposition of γP^{-1}: that is $D^\top D = \gamma P^{-1}$.

   There exists different definitions of generalized SVD: MATLAB implements the call `[U,V,X,C,S] = gsvd(Φ, D)` where U and V are unitary matrices, and C and S are nonnegative diagonal matrices such that

$$\Phi = UCX^\top \qquad D = VSX^\top \qquad C^2 + S^2 = I.$$

   Inserting the decompositions into eq. (2.4) gives

$$X^\top \theta = (UC)^\top y$$

   which can be solved as $\theta = X^\top \backslash (UC)^\top y$;

3. Eq. (2.1) can also be rewritten as

$$\left|\left| \begin{bmatrix} \Phi \\ D \end{bmatrix} \theta - \begin{bmatrix} y \\ 0 \end{bmatrix} \right|\right|_2^2$$

   and solved by backslash.

I tested the three approaches on prob. 8 and found the fastest to be the third which does not construct the normal equation and is therefore better conditioned at the small cost of increasing the problem size by the r rows of P^{-1}. The first method is twice slower and the second is 100 (!) times slower: the problem is most likely the call to `gsvd`. This is strange because this is the method normally suggested, but probably only for larger regularization matrices.

Chapter 3

Frequency Domain Analysis

3.1 Linear Time-Invariant Systems

We consider linear time-invariant systems

$$y(k) \doteq \sum_{l=1}^{\infty} g(l)u(k-l) + v(k) \tag{3.1}$$

with a causal pulse reponse $g(l) = 0$, $\forall l \leq 0$.

Using the *forward and inverse shift operator* [2, p. 24]

$$qu(k) \doteq u(k+1), \qquad q^{-1}u(k) \doteq u(k-1)$$

one rewrites the expression above as

$$\sum_{l=1}^{\infty} g(l)u(k-l) = \sum_{l=1}^{\infty} g(l)q^{-l}u(k) = \underbrace{\left[\sum_{l=1}^{\infty} g(l)q^{-l}\right]}_{\doteq G(q)} u(k).$$

$v(t)$ is a filtered additive disturbance, $v(t) = H(q)e(t)$

$$H(q) \doteq 1 + \sum_{l=1}^{\infty} h(l)q^{-l}$$

and $e(k)$ is scaled so that $h(0) = 1$.

In the rest, we shall be considering the (SISO) linear model

$$y(k) = G(z)u(k) + H(z)e(k) \tag{3.2}$$

where $y, u \in \mathbb{R}$, $e \in \mathcal{N}(0, \sigma^2)$ is the Gaussian uncorrelated independent and identically distributed (i.i.d.) noise with zero mean and variance σ^2 and $v(k) = H(z)e(k)$ the filtered noise.

We wish to estimate $\hat{G}$ from the finite set of measurements

$$\mathcal{Z}_K = \{u(0), y(0), \ldots, u(K-1), y(K-1))\}$$

of length K. The different methods

- transfer function in the frequency domain (this Chapter);
- real-rational transfer function in the time domain[1], Chap. 4;
- state-space approach

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) + Du(k) \end{aligned}$$

  Chap. 7.

are equivalent in the absence of noise. For all methods, the *unmeasured* past induces an error in the estimate because for a causal system, the signal $y(k)$ depends on the control inputs at previous times up to time k which may all not be available. The problem does not arise when the system starts at rest where $u(k) = y(k) = 0, \ \forall k < 0$.

[1]This holds for the frequency and the time approach. Eq. (3.2) is parametrized as

$$y(k) = G(\theta)u(k) + H(\theta)e(k), \quad y, u \in \mathbb{R}$$

where $\theta \in \mathbb{R}^d$ and d is the model parameter. For real-rational transfer functions in the z-domain

$$G(z) = \frac{b_1 z^{-1} + b_2 z^{-2} + \ldots + b_m z^{-m}}{1 + a_1 z^{-1} + \ldots + a_m z^{-n}} \tag{3.3}$$

a common parametrization is

$$\theta = \begin{bmatrix} a_1 & \ldots & a_n & b_1 & \ldots & b_m \end{bmatrix}^\top$$

where $p = m + n$. The transfer function in the frequency domain $G(e^{j\omega})$ is $G(z)$ evaluated at fixed points $z = e^{j\omega}$

$$G(e^{j\omega}) = G(z)\Big|_{z=e^{j\omega}}.$$

3.2 Empirical Transfer-Function Estimate (ETFE)

In the frequency domain,

$$U(e^{j\omega}) \doteq \sum_{k=-\infty}^{\infty} y(k)e^{j\omega k} \qquad Y(e^{j\omega}) \doteq \sum_{k=-\infty}^{\infty} y(k)e^{j\omega k}$$

eq. (3.2) becomes

$$Y(e^{j\omega}) = G(e^{j\omega})U(e^{j\omega}) + V(e^{j\omega}) + R(e^{j\omega}) \tag{3.4}$$

where the system's response $Y(e^{j\omega})$ has been split[2] into a periodic term driven by the periodic input $U(e^{j\omega})$ and the additional term $R(e^{j\omega})$ which accounts for the transient.

To obtain an estimate $\hat{G}(e^{j\omega})$ for $G(e^{j\omega})$, we make the following three considerations:

- the ratio
$$\frac{Y(e^{j\omega})}{U(e^{j\omega})} \tag{3.5}$$
provides an *unbiased* estimate of $G(e^{j\omega_n})$ for a periodic input signal (*i.e.* no transient) and in the absence of noise (see Sect. 3.2.1);

- we only have access to finite length input-output data $\{u(k), y(k)\}$. As the best approximation for the infinite summation $\{U(e^{j\omega}), Y(e^{j\omega})\}$ we use the discrete Fourier transforms (Sect. 3.5)
$$U_N(e^{j\omega_n}) = \sum_{k=0}^{N-1} y(k)e^{j\omega_n k} \qquad Y_N(e^{j\omega_n}) = \sum_{k=0}^{N-1} y(k)e^{j\omega_n k}.$$
$\omega_n = \frac{2\pi}{N}$ are determined by the number of measurement points N;

[2]The initial transient corrupts the measurements: let $W_{[0,N-1]}(k)$

$$W_{[0,N-1]}(k) = \begin{cases} 1 & \text{if } 0 \le k < N \\ 0 & \text{otherwise} \end{cases}$$

the window function; then, for all times up to $N-1$,

$$y(k) = Gu_p(k) - \underbrace{Gu_p(k)W_{(-\infty,-1]}(k)}_{r(r)} + v(k)$$

$$Y_N(e^{j\omega_n}) = G(e^{j\omega_n})U_N(e^{j\omega_n}) + R_N(e^{j\omega_n}) + V(e^{j\omega_n})$$

where $u_p(k)$ is the periodic input signal.

- the Fourier transform of $v(k)$ is assumed to exist. As noise is typically finite power, and not finite energy, this will not be satisfied. However we are more interested in the relationship between $U(e^{j\omega_n})$ and $Y(e^{j\omega_n})$ and so we will overlook this issue for the time being.

3.2.1 Effect of the Input Signal

The statistical properties of the EFTE

$$\hat{G}(e^{j\omega_n}) \doteq \frac{Y_N(e^{j\omega_n})}{U_N(e^{j\omega_n})} = G(e^{j\omega_n}) + \frac{R_N(e^{j\omega_n})}{U_N(e^{j\omega_n})} + \frac{V_N(e^{j\omega_n})}{U_N(e^{j\omega_n})} \tag{3.6}$$

are affected by how the input signal is chosen. The estimate is unbiased, since $\mathbb{E}\left\{V_N(e^{j\omega_n})\right\} = 0$,

$$\mathbb{E}\left\{\hat{G}(e^{j\omega_n})\right\} = G(e^{j\omega_n}) + \frac{R_N(e^{j\omega_n})}{U_N(e^{j\omega_n})}$$

and in the absence of the transient, $R_N(e^{j\omega_n}) = 0$, that is, for a *periodic* signal[3].

The choice of input signal also affects the signal-to-noise ratio (SNR): if the number of measurement points N is taken to be a multiple integer of the period M of the input signal, the magnitude of $U_N(e^{j\omega})$ grows at a rate of N

$$\mathbb{E}\left\{\left|U_N(e^{j\omega_n})\right|^2\right\} = m^2\left|U_M(e^{j\omega_n})\right|^2, \qquad m = \frac{N}{M}. \tag{3.7}$$

Here the only frequencies with non-zero spectral amplitude are those at $\frac{2\pi}{M}n$ with $n \in [0, \ldots, M-1]$, since the signal is periodic with period M, even though there are N measurement points; the remaining $N - M$ components are 0.

Were instead the input signal *random*, we would have

$$\mathbb{E}\left\{|U_N(e^{j\omega_n})|^2\right\} = N\phi_u(e^{j\omega_n}) + 2c \tag{3.8}$$

where here *all* N frequencies $\frac{2\pi}{N}n$ have spectral content. In other words, a periodic signal excites only a limited number of frequencies and concentrates

[3]Periodicity implies that the input signal is known in the interval $[-\infty, +\infty]$ and that the transient has already died out. This is not the case for a random input signal even if the input starts at times earlier than $k = 0$ unless the initial conditions are known: in this case the transient could be taken into account, see Sect 4

all the energy in those frequencies, a random one the whole spectrum. The quantity

$$|c| \leq C = \sum_{\tau=1}^{\infty} |\tau R_v(\tau)|$$

is finite for random noise (50 Hz noise is not finite).

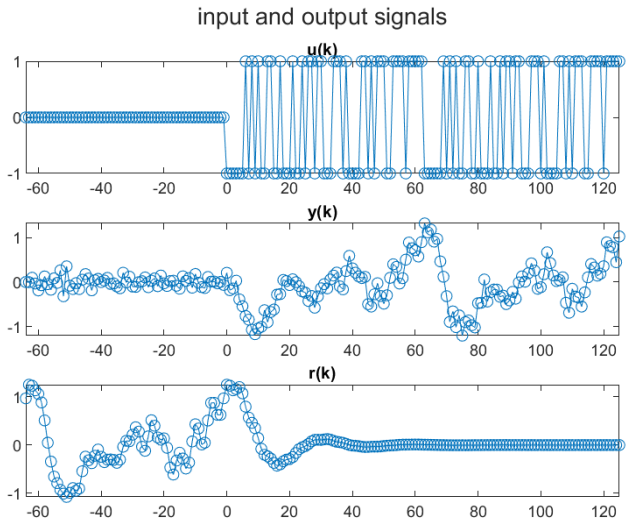

Figure 3.1: Plot of the input $u(k)$ (top), the output $y(k)$ (middle) and transient $r(k)$ (bottom) signals, here shown only for two periods before $t = 0$ and two after $t = 0$.

What is the impact of the noise[4] and the transient? What follows is my hand-waving attempt to understand the motivation to use a periodic signal. The noise signal "behaves" like a random input signal: disregarding the constant term in eq. (3.8), SNR remains constant for a random input

[4]For a periodic input signal, we have

$$\mathbb{E}\left\{\left|\hat{G}_N(e^{j\omega_n}) - G_0(e^{j\omega_n})\right|^2\right\} = \frac{\phi_v(e^{j\omega_n}) + \frac{2c}{N}}{\frac{1}{N}|U_N(e^{j\omega_n})|^2} \tag{3.9}$$

the expression for c is given above, and

$$\mathbb{E}\left\{|\hat{G}(e^{j\omega_n})|^2\right\} = |G(e^{j\omega_n})|^2 + \frac{N\phi_v(e^{j\omega_n})}{|U_N(e^{j\omega_n})|^2}.$$

signal, since

$$\frac{\mathbb{E}\left\{|V_N(e^{j\omega_n})|^2\right\}}{\mathbb{E}\left\{|U_N(e^{j\omega_n})|^2\right\}}$$

does not change for growing N. Instead it decreases as $\frac{1}{N}$ for periodic input signals.

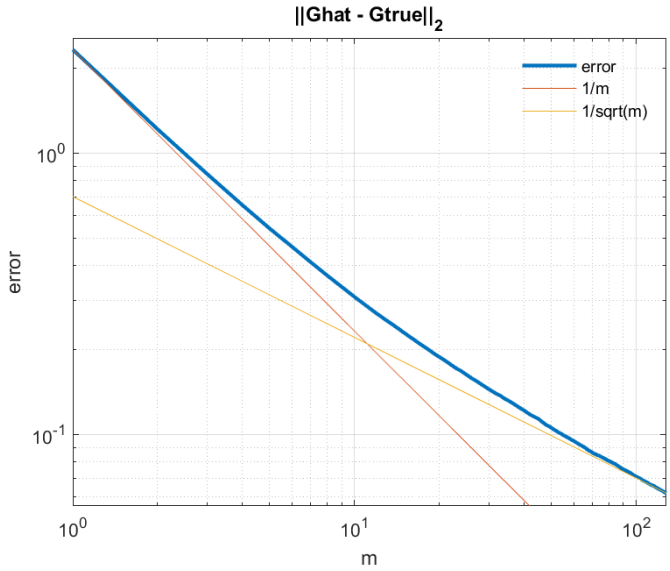

Figure 3.2: Convergence of $\mathbb{E}\left\{||\hat{G} - G||_2\right\}$ as a function of the number of periods m (each of M points). The cross-over between the transient-limited convergence with dependence $\frac{1}{m}$ to noise-limited with convergence $\frac{1}{\sqrt{m}}$ is clearly visible.

The transient instead decays exponentially: its PSD remains constant as a function of N and the convergence of the SNR is $\sim \frac{1}{N}$. Fig. 3.1 shows the output signal $y(k)$ affected by both noise and transient: the cross-over between transient-limited and noise-limited convergence is visible in Fig. 3.2. Implementation in `05_lect/SysID_ETFE.m`.

3.3 Swept-Sine Identification

The system[5] is excited with the periodic input signal $u(k) = \alpha \cos(\omega_u k)$ of *arbitrary* frequency ω_u. The output[6] to this excitation is

$$y(k) = \underbrace{\frac{\alpha}{2}\left[G(e^{j\omega_u})e^{+j\omega_u k} + G(e^{-j\omega_u})e^{-j\omega_u k}\right]}_{\text{steady state response}} + v(k) + \text{transient.}$$

The correlation signal

$$\begin{aligned} I(N) &\doteq \frac{1}{N}\sum_{k=0}^{N-1} y(k)e^{-j\omega_u k} \\ &= \frac{\alpha}{2}G(e^{j\omega_u}) + \frac{\alpha}{2N}\sum_{k=0}^{N-1} G(e^{j\omega_u})e^{-2j\omega_u k} \\ &\to \frac{\alpha}{2}G(e^{j\omega_u}) \quad \text{as } N \to \infty \end{aligned}$$

since the non-contant components (including noise and transient) average out.

Advantages of this method:

- the energy is concentrated at the frequencies of interest;
- arbitrary frequencies can be measured and not only those determined by the Fourier transform;
- the amplitude of $u(k)$ can easily be tuned as a function of frequency;
- it makes it easy to avoid saturation and to tune SNR.

Disadvantages are:

- a large amount of data is required;
- significant amount of time required for experiments because single frequencies are probed sequentially;
- some processes will not allow sinusoidal inputs.

[5] To my knowledge, this method is also called demodulation or lock-in detection.

[6] Plugging the excitation into eq. (3.1), one obtains

$$\begin{aligned} y(k) &= \frac{\alpha}{2}\left[\sum_{l=1}^{\infty} g(l)e^{+j\omega_u(k-l)} + \sum_{l=1}^{\infty} g(l)e^{-j\omega_u(k-l)}\right] + v(k) \\ &= \frac{\alpha}{2}\left[\left(\sum_{l=1}^{\infty} g(l)e^{-j\omega_u l}\right)e^{+j\omega_u k} + \left(\sum_{l=1}^{\infty} g(l)e^{+j\omega_u l}\right)e^{-j\omega_u k}\right] + v(k) \end{aligned}$$

where the terms in round brackets are $G(e^{+j\omega_u})$ and $G(e^{-j\omega_u})$ respectively.

3.4 Spectral Estimation Methods

From eq. (3.6), neglecting the transient and multiplying element-wise[7] by $U^\star(e^{j\omega_n})$, we have

$$\phi_{yu}(e^{j\omega_n}) = G(e^{j\omega_n})\phi_u(e^{j\omega_n}) + \phi_{uv}(e^{j\omega_n})$$

the cross correlation term $\phi_{uv}(e^{j\omega_n})$ being zero if u and v are uncorrelated (for instance in open loop). The estimate $\hat{G}(e^{j\omega_n})$ is

$$\hat{G}(e^{j\omega_n}) = \frac{\hat{\phi}_{yu}(e^{j\omega_n})}{\hat{\phi}_u(e^{j\omega_n})}.$$

It is not clear to me why one would choose spectral methods instead of ETFE since the difference is only the multiplication by the deterministic input signal.

The major point is how to determine the PSD in the case of non-periodic signals.

3.5 Discrete Fourier Transform

The Fourier series is defined (for M even) as

$$X(e^{j\omega_n}) = \sum_{k=0}^{M-1} x(k)e^{-j\omega_n k}, \qquad \omega_n = \frac{2\pi}{M}n \tag{3.10}$$

and the inverse Fourier transform as

$$x(k) = \frac{1}{M}\sum_{n=0}^{M-1} X(e^{j\omega_n})e^{j\omega_n k}. \tag{3.11}$$

3.5.1 Auto-Correlation for Periodic Signals

Given the periodic signal $\{x(k)\}$, the auto-correlation function is the product of x with itself delayed by τ:

$$R_x(\tau) = \frac{1}{N}\sum_{k=0}^{N-1} x(k)x(k-\tau) \tag{3.12}$$

[7] The supplementary notes take a different approach, so I am not sure that my derivation holds.

and the power spectral density (PSD)[8]

$$\phi_x(e^{j\omega_n}) = \sum_{\tau=0}^{N-1} R_x(\tau)e^{-i\omega_n\tau} = \frac{1}{N}|X(e^{j\omega_n})|^2. \tag{3.13}$$

The auto-correlation has the following properties:

- $R_x(-\tau) = R_x^\star(\tau)$;
- $R_x(0) \geq |R_x(\tau)|$, for all $\tau > 0$.

$\phi_x(\omega)$ has the following properties:

- $\phi_x(\omega)$ is non-negative for all ω
- $\phi_x(\omega) = \phi_x(-\omega)$ for all real-valued $x(k)$.

3.5.2 Cross-Correlation for Periodic Signals

Given the periodic signals $y(k)$ and $u(k)$, the cross-correlation function is

$$R_{yu}(\tau) = \frac{1}{N}\sum_{k=0}^{N-1} y(k)u(k-\tau) \tag{3.14}$$

and the cross-spectral density

$$\phi_{yu}(e^{j\omega_n}) = \sum_{\tau=0}^{N-1} R_{yu}(\tau)e^{-i\omega_n\tau} = \frac{1}{N}Y(e^{j\omega_n})U^\star(e^{j\omega_n}). \tag{3.15}$$

3.6 Random Signals

We assume $e(k) \in \mathcal{N}(0, \sigma^2)$; $e(k)$ are independent and identically distributed (i.i.d.).

[8] Using $e^{-j\omega_n\tau} = e^{-j\omega_n k} \cdot e^{-j\omega_n(\tau-k)}$ and eq. (3.12), one has

$$\begin{aligned}\phi_x(e^{j\omega_n}) &= \frac{1}{N}\sum_{k=0}^{N-1} x(k)e^{e^{-j\omega_n k}}\left[\sum_{\tau=0}^{N-1} x(k-\tau)e^{-j\omega_n(k-\tau)}\right]^\star \\ &= \frac{1}{N}X(e^{j\omega_n})X^\star(e^{j\omega_n})\end{aligned}$$

For random signals, the definition is in terms of the expected values: the *autocorrelation* function is defined as

$$R_x(\tau) = \mathbb{E}\left\{x(k)x(k-\tau)\right\}$$

and the *covariance* function as

$$R_x(\tau) = \mathbb{E}\left\{\left(x(k) - \mathbb{E}\left\{x\right\}\right)\left(x(k-\tau) - \mathbb{E}\left\{x\right\}\right)\right\}$$

It will be assumed that random signals are zero mean and the notation $R_x(\tau)$ is used for both the autocorrelation and covariance functions. The stationarity implies that the expectations only depends on the shift τ.

The power spectral density is defined as the *Fourier transform* of the autocovariance, $R_x(\tau)$

$$\phi_x(e^{j\omega}) = \sum_{\tau=-\infty}^{\infty} R_x(\tau)e^{-j\omega\tau} \quad \omega \in [-\pi, \pi)$$

and the inverse transform is given by

$$R_x(\tau) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \phi_x(e^{j\omega})e^{j\omega\tau}\mathrm{d}\omega.$$

Cross-covariance and cross-correlations are similarly defined. For zero mean signals, the cross-correlation is

$$R_{yu}(\tau) = \mathbb{E}\left\{y(k)u(k-\tau)\right\}.$$

The definitions for the auto- and cross-correlation given above are for infinite length signals. Since one can only collect finite-length data, one must make a sensible approximation. The difficulty arises from the fact that in system identification "the measured output is almost always composed of the sum of a stochastic signal (from the noise) and a deterministic, or even periodic, signal from the convolution of the plant with a deterministic input signal, $u(k)$. This means that there is no obvious correct choice in deciding which autocorrelation estimation method to apply. The better method will inevitably be problem dependent." [5].

Chapter 4

Pulse Response Estimation

The frequency-domain approach cannot easily deal with the transient: ignoring it leads to a biased estimate of the transfer function $\hat{G}(e^{j\omega_n})$. In the time domain approach it is easy to include the transient.

Starting[1] from eq. (3.1)

$$y(k) \doteq \sum_{l=0}^{\infty} g(l)u(k-l) + v(k)$$

(note that here the summation[2] starts from $l = 0$) and expanding the relationship above gives

$$\begin{aligned} y(k) &= g(0)u(k) + g(-1)u(k-1) + g(-2)u(k-2) + \ldots + v(k) \\ y(k+1) &= g(0)u(k+1) + g(-1)u(k) + g(-2)u(k-1) + \ldots + v(k+1) \\ y(k+2) &= \ldots \end{aligned}$$

which can be written in Toeplitz matrix form as

$$\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(K-1) \end{bmatrix} = \begin{bmatrix} u(0) & u(-1) & \cdots & u(-K+1) & \cdots \\ u(1) & u(0) & \cdots & u(-K+2) & \cdots \\ \vdots & & \ddots & \vdots & \\ u(K-1) & u(K-2) & & \cdots & \ldots \end{bmatrix} \begin{bmatrix} g(0) \\ g(1) \\ g(2) \\ \vdots \end{bmatrix} + \begin{bmatrix} v(0) \\ v(1) \\ \vdots \\ v(K-1) \end{bmatrix} \tag{4.1}$$

[1] We have seen in the previous chapter that $G(e^{j\omega})$ is related to $g(\tau)$ by a Fourier transformation when the full time reponse $k \in (-\infty, +\infty)$ is available; otherwise it is only approximately true.

[2] Why do we need the pass-through term?

or, in matrix notation, as

$$Y = \Phi_u g + V. \tag{4.2}$$

The existence of the estimate $\hat{g}$ and its properties depends on how Φ_u is constructed. We will consider them here.

In the following we will drop the error term V to keep the notation more readable. Moreover, zero-mean Gaussian distributed error does not induce bias using least-squares estimation.

4.0.1 Estimation of Truncation Error for IIR

A finite impulse response (FIR) is described by a finite $(\tau_{\max} + 1)$ number of coefficients: that is $g(\tau) = 0$ for $\tau \geq \tau_{\max}$. $\hat{g}(\tau)$ in eq. (4.2) can be found by least-squares.

On the contrary, for rational transfer functions (infinite impulse response, IIR) an infinite number of terms is required to describe the impulse response g. If the system is stable, the amplitudes of g decays exponentially.

We cannot solve the least-squares problem with an infinite number of terms but the system can be truncated to a finite number of terms, since g decays exponentially. This is guaranteed by the following theorem: For a strictly stable real-rational system, if all of the poles of g are inside the unit-circle, then, for any $\epsilon > 0$, there exists a $\tau_{\max}$ such that

$$\sum_{i=\tau_{\max}+1}^{\infty} |g(i)| < \epsilon$$

where the truncation of eq. (4.2) gives rise to the error term

$$\begin{bmatrix} y(0) \\ \vdots \\ y(K-1) \end{bmatrix} = \begin{bmatrix} u(0) & \cdots & u(-\tau_{\max}) \\ \vdots & \ddots & \vdots \\ u(K-1) & \cdots & u(K-\tau_{\max}-1) \end{bmatrix} \begin{bmatrix} g(0) \\ \vdots \\ g(\tau_{\max}) \end{bmatrix} + \begin{bmatrix} e(0) \\ \vdots \\ e(K-1) \end{bmatrix}$$

where $Y,\ E \in \mathbb{R}^K$, $\Phi_u \in \mathbb{R}^{K \times (\tau_{\max}+1)}$ and $g \in \mathbb{R}^{\tau_{\max}+1}$. Ignoring the error makes the estimate biased: the estimate is however asymptotically unbiased as it can be made arbitrarily small by selecting a larger $\tau_{\max}$.

4.0.2 Initial Conditions: Negative Times for g

The expression eq. (4.1) requires the knowledge of the control input at negative times. If the initial conditions are not known, the corresponding terms

must be discarded

$$\begin{bmatrix} y(\tau_{\max}) \\ \vdots \\ y(K-1) \end{bmatrix} = \begin{bmatrix} u(\tau_{\max}) & \cdots & u(0) \\ \vdots & \ddots & \vdots \\ u(K-1) & \cdots & u(K-\tau_{\max}-1) \end{bmatrix} \begin{bmatrix} g(0) \\ \vdots \\ g(\tau_{\max}) \end{bmatrix} + \begin{bmatrix} e(\tau_{\max}) \\ \vdots \\ e(K-1) \end{bmatrix}$$

4.0.3 Bad Measurements

A bad measurement can be also easily dealt with in the time domain. Let us assume that $y(j)$ was corrupted. The problem can still be written as before with the faulty line removed:

$$\begin{bmatrix} y(\tau_{\max}) \\ \vdots \\ y(j-1) \\ y(j+1) \\ \vdots \\ y(K-1) \end{bmatrix} = \begin{bmatrix} u(\tau_{\max}) & \cdots & & u(0) \\ \vdots & & & \vdots \\ u(j-1) & u(j-2) & \cdots & u(j-\tau_{\max}-2) \\ u(j+1) & u(j) & \cdots & u(j-\tau_{\max}) \\ \vdots & & & \vdots \\ u(K-1) & \cdots & & u(K-\tau_{\max}-1) \end{bmatrix} \begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ g(\tau_{\max}) \end{bmatrix} + \begin{bmatrix} e(\tau_{\max}) \\ \vdots \\ e(j-1) \\ e(j+1) \\ \vdots \\ e(K-1) \end{bmatrix}$$

and the truncation is still valid.

The problem is different if $u(j)$ were unknown: in this case all elements[3] where $u(j)$ had an effect, must be eliminiated.

4.0.4 Uniqueness of $\hat{g}$ and Persistency of Excitation

For the solution to be unique, Φ_u needs to have full column rank. What input signals satisfy this requirement? For instance, a single sinusoidal input signal would make the matrix have rank 2 and in the frequency domain, this would determine amplitude and phase of a single frequency[4].

- The constant function[5] $u(k) = 1, \forall k$ is persistently exciting of order 1;
- a PRBS signal[6] of period M is persistently exciting of order M;

[3] This is a block of dimentions $\tau_{\max} \times \tau_{\max}$.

[4] In the time domain, trying to solve the rank-deficient problem results in the wrong estimate of the coefficients, see `07_lect/persistency.jl`.

[5] Prof. Smith calls is step function, but a step function makes Φ_u full matrix.

[6] This is a periodic deterministic signal with white noise-like properties. It is generated by the differential equation

$$u(k) = \mod(A(q)u(k), 2), \qquad A(q)u(k) = \sum_{i=1}^{n} a_i u(t-i).$$

- the sum of sinusoidals

$$u(k) = \sum_{s=1}^{S} \alpha_s \cos(\omega_s k + \phi_s)$$

  is persistently exciting of order $2S$ if $\omega_s \in (0, \pi)$. The order decreases by 1 if either one of $\omega_s = 0, \pi$ is included and by 2 if both frequencies are included. For periodic signals, Φ_u is circulant[7].

In class (and in [2, Sect. 13.2]) the uniqueness of $\hat{g}$ was discussed in terms of the rank of the auto-correlation matrix R_u, but discussion in Moodle indicates that checking for the singular values of Φ_u is completely equivalent.

4.0.5 Pulse Response from Cross- and Autocorrelation

A similar relationship holds for the correlations[8]

$$R_{yu}(\tau) = g(\tau) \star R_u(\tau)$$

This is seen by multiplying eq. (4.2) from the left by $\Phi_u^\top$; $R_{yu}(\tau) = \Phi_u^\top y$ and $R_u = \Phi_u^\top \Phi_u^\top$, and taking the expectations, because Φ_u is the matrix containing the shifted entries u.

The equivalent expression in the frequency domain was derived in Sect. 3.4.

The actual period depends on the choice of $A(q)$ and for each n there exists choices of $A(q)$ that give the maximum length. [2, Chap. 13]. The MATLAB command is either `prbs(M,N)` or `idinput("prbs", N)`.

[7] A circulant matrix has the form

$$\begin{bmatrix} u(0) & u(N-1) & \cdots & u(1) \\ u(1) & u(0) & \cdots & u(2) \\ \vdots & & \ddots & \vdots \\ u(N-1) & u(N-2) & \cdots & u(0) \end{bmatrix}.$$

[8] So was derived the expression in class: From eq. (3.1), we have

$$\begin{aligned} \mathbb{E}\{y(k)u(k-\tau)\} &= \mathbb{E}\left\{\sum_{i=0}^{\infty} g(i)u(k-i)u(k-\tau)\right\} + \mathbb{E}\{v(k)u(k-\tau)\} \\ &= \sum_{i=0}^{\infty} g(i)\mathbb{E}\{u(k-i)u(k-\tau)\} \\ &= \sum_{i=0}^{\infty} g(i)R_u(\tau - i) \end{aligned}$$

4.1 Equivalence Between Time and Frequency Domain

Given eq. (4.2), the equivalence with the ETFE is established by multiplying by the Fourier transform matrices F_y and F_g (which may have different dimensions but are nevertheless square)

$$F_y Y = F_y \Phi_u F_g^{-1} \underbrace{F_g g}_{\doteq G} + F_y V \longrightarrow \hat{G} = \left(F_y \Phi_u F_g^{-1}\right) \setminus (F_y Y) \tag{4.3}$$

If the input signal is periodic, Φ_u is a square circulant matrix: F_y and F_g have the same dimensions and $F_y \Phi_u F_g^{-1}$ is a diagonal matrix. The result above simplifies to

$$\hat{G}(e^{j\omega_n}) = \frac{Y_N(e^{j\omega_n})}{U_N(e^{j\omega_n})}.$$

Chapter 5

Error Prediction Methods and Transfer Function Models

We assume that the complete system model is given by

$$y(k) = G(z)u(k) + H(z)e(k). \tag{5.1}$$

Prediction error-based identification methods estimate the transfer functions $G(z)$ and $H(z)$, parametrised as $G(z,\theta)$ and $H(z,\theta)$, by minimizing the objective function $J(\epsilon)$, a function of the prediction error $\epsilon(k,\theta)$

$$\epsilon(k,\theta) \doteq y(k) - \hat{y}(k,\theta).$$

The prediction $\hat{y}(k,\theta)$ for the time k is a function of the previous measurements and inputs at $k-1,\ k-2,\ldots$ only.

5.1 Prediction

We assume that $H(z)$ is monic[1] and stable[2]. In this case, given the filtered noise $v(k) = H(z)e(k)$, $e(k)$ can be reconstructed from $v(k)$ as

$$e(k) = H^{-1}(z)v(k). \tag{5.2}$$

[1] Monic means that $h(0) = 1$:

$$H(z) = 1 + \sum_{k=1}^{\infty} h(k)z^{-k}.$$

[2] $H(z)$ has only poles strictly inside the unit circle.

We now seek to predict $v(k)$ given the past values up to time $k-1$: this is called the *one-step ahead* estimate. Since $H(z)$ is monic, we split the filtered noise contribution into the term $e(k)$ and other terms up to time $k-1$

$$v(k) = H(z)e(k) = e(k) + (H(z) - 1)\, e(k) \tag{5.3}$$

The *predicted* filtered noise $\hat{v}(k|k-1)$ is

$$\hat{v}(k|k-1) = (H(z) - 1)e(k).$$

This can be intuitively understood because the error probability function distribution for $\{e\}$ has zero mean[3]: if we were left to guess $e(k)$, we would guess $e(k) = 0$. Making use of eq. (5.2), the one-step ahead estimate

$$\hat{v}(k|k-1) = \left(1 - H^{-1}(z)\right) v(k) \tag{5.4}$$

is determined only from the knowledge of the past values of v up to $k-1$.

For the model of eq. (5.1), the one-step ahead predictor

$$\hat{y}(k|k-1) = G(z)u(k) + \hat{v}(k|k-1) \tag{5.5}$$

can be rewritten with the help of the expression for $\hat{v}(k|k-1)$ as[4]

$$\hat{y}(k|k-1) = H^{-1}(z)G(z)u(k) + \left(1 - H^{-1}(z)\right) y(k). \tag{5.6}$$

The prediction error[5]

$$\epsilon(k, |k-1) = y(k) - \hat{y}(k|k-1) = e(k) \tag{5.7}$$

is the noise $e(k)$: the *innovation* is the part of the output prediction that cannot be estimated from past measurements.

---

[3] Had the probability distribution $f_e(x)$ not had a zero mean, we would have to modify the predition according to

$$\hat{v}(k|k-1) = \arg\max_x f_e(x - m(k-1)), \qquad m(k-1) = (H(q) - 1)e(k).$$

[4] Using $v(k) = y(k) - G(z)u(k)$

$$\hat{y}(k|k-1) = G(z)u(k) + \left(1 - H^{-1}(z)\right)\left(y(k) - G(z)u(k)\right).$$

Note that $y_0(k) = G(z)u(k)$ is the evolution of the noiseless true system, so that eq. (5.5) could also be written as $\hat{y}(k|k-1) = y_0(k) + \hat{v}(k|k-1)$. However we seek an expression that involves the measured outputs $y(k)$ and expressing the one-step head predictor as a function of the unknown true system is of no use to us.

[5] Although in class the algebra was worked out, that $\epsilon(k|k-1) = e(k)$ is no surprise: $e(k)$ was the term that was discarded from $v(k)$ to compute the predicted filtered noise $\hat{v}(k|k-1)$.

5.1.1 Example: Moving Average

The model

$$v(k) = e(k) + ce(k-1) \rightarrow H(z) = 1 + cz^{-1}$$

is invertible when $|c| < 1$. The one-step ahead predictor eq. (5.4) can be expressed in terms of the error $e(k-1)$ using eq. (5.7)

$$\hat{v}(k|k-1) = \left(1 - H^{-1}(z)\right) H(z)e(k) = cz^{-1}e(k) = ce(k-1)$$

5.2 Family of Transfer-Function Models

The advantage of the transfer-function models is that they can be described by fewer parameters and that the inputs required to identify the system do not have to be persistently exciting as it is the case when one wants to identify frequency or time-response: we have seen in Sect. 4.0.4 that for frequency and time domain methods, the order of excitation must be double the number of complex estimates of the transfer function $G(e^{j\omega_n})$ since gain and phase must be determined for each frequency; for a time response one requires the same persistency order as the number of impulse response terms.

If we control the input, this requirement is easy to satisfy. If on the other hand the data is given, this may not be the case and in these situations, one is better off looking for transfer functions/state space representations because of the reduced numbers of parameters to identify.

Prediction error-based identification methods construct the prediction error[6]

$$\epsilon(k, \theta) = y(k) - \hat{y}(k, \theta) \tag{5.8}$$

from the (one-step ahead) predictor $\hat{y}(k, \theta)$ which is based on the guesses $\hat{G}(z) = G(z, \theta)$ and $\hat{H}^{-1}(z) = H^{-1}(z, \theta)$, the guesses being parametrized by θ. The optimal $\theta^\star$ is the argument that minimizes the cost function $J = J(\epsilon)$

$$\theta^\star = \arg\min_\theta J(\epsilon(k, \theta)).$$

Typical choices for the objective functions $J(\epsilon)$ are the 2-norm $||\epsilon||_2^2$ or the maximum deviation, the ∞-norm $||\epsilon||_\infty$. The kind of minimization depends on how the models $G(q)$ and $H^{-1}(q)$ are parametrised: in general, the

[6]Here we are seeking the unknown parameters θ for a known model for which we can construct the one-step ahead predictor $\hat{y}(k, \theta)$, parametrized by θ. Since θ is unknown, we cannot say just yet that $\epsilon(k, \theta) = e(k)$.

I also do not know if $\min_\theta ||\epsilon(\theta)||_2^2 = \sum_k ||e(k)||_2^2$ but I suspect this is the case when $e(k)$ is Gaussian-distributed.

parametrization will not be linear and the optimization may not be convex. Note moreover that the minimization of $||\epsilon||_2^2$ is not equivalent to the least squres method, unless $H(z) = 1$:

$$y(k) = G(z)u(k) + e(k).$$

Since this approach does not require an a-priori knowlege of the system, it is also called the *black-box* approach.

5.2.1 Equation Error Model Structure (ARX)

The ARX model

$$y(k) = B(z)u(k) + (1 - A(z))\, y(k) + e(k) \tag{5.9}$$

is a simple input-output relationship where the error enters as a direct term: this model covers a limited set of real-world problems, for instance those where the perturbation act as a force. We take

$$A(z) = 1 + a_1 z^{-1} + \ldots + a_n z^{-n}, \qquad B(z) = b_1 z^{-1} + \ldots + b_m z^{-m}$$

where $A(z)$ is monic and $B(z)$ does not contain a constant term, *i.e.* the model has no feed-through. It corresponds to the model of eq. (5.1) with

$$G(z) = \frac{B(z)}{A(z)}, \qquad H(z) = \frac{1}{A(z)}.$$

and generates the one-step ahead predictor

$$\hat{y}(k|k-1) = (1 - A(z))\, y(k) + B(z)u(k) \tag{5.10}$$

either by plugging $G(z)$ and $H^{-1}(z)$ in eq. (5.6) or by using the result of eq. (5.7) in eq. (5.9).

We now seek the estimates $\hat{B}(z, \theta)$ and $\hat{A}(z, \theta)$ parametrized by the *parameter vector* θ which contains the unknown coefficients of the polynomials $A(z)$ and $B(z)$

$$\theta = \begin{bmatrix} a_1 & \ldots & a_n & b_1 & \ldots & b_m \end{bmatrix}^\top$$

This generates the one-step ahead predictor

$$\hat{y}(k, \theta) = \left(1 - \hat{A}(z, \theta)\right) y(k) + \hat{B}(z)u(k).$$

that can be written in linear form

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

using the *regressor vector* $\varphi(k)$ that contains the outputs and inputs.

θ is found by minimization of the prediction error $\epsilon(k,\theta) = y(k) - \hat{y}(k,\theta)$: when θ is the true parameter vector θ_0, then $\epsilon(k,\theta_0) = e(k)$. Generally, one minimizes the 2–norm $||\epsilon(k,\theta)||_2^2$: this is solved in MATLAB by

$$\hat{\theta} = \Phi \backslash Y, \qquad \Phi \doteq \begin{bmatrix} \varphi^\top(1) \\ \varphi^\top(2) \\ \dots \\ \varphi^\top(N) \end{bmatrix}, \quad Y \doteq \begin{bmatrix} y(1) \\ y(2) \\ \dots \\ y(N) \end{bmatrix} \tag{5.11}$$

The first element of the regressor is $\varphi(1)$ that depends on the past indeces up to $k = 0$: these are the initial conditions. They are usually specified by the system being at rest; otherwise, one must drop a certain amount of entries (how many exactly?)

5.2.2 ARX: Estimate Bias

The numerical example in class (slides 9-30 to 9–34) confused me. In the first situation we have

$$\frac{B(z)}{A(z)} = \frac{bz^{-1}}{1 + az^{-1}}, \qquad H(z) = 1.$$

The one-step ahead predictor is

$$\hat{y}(k|k-1) = y(k) - e(k) = \frac{B(z)}{A(z)} u(k) = \varphi_0^\top(k)\theta_0$$

is nothing else than the evolution of the true system. The regressor Φ_0 must be constructed from the true $y_0(k) = \frac{B(z)}{A(z)}u(k)$: using instead the measured $y(k)$ gives the bias of slide 9–30. (See also discussion in Sect. 2.1.2.)

In the second situation, we take the noise to be $H(z) = \frac{1}{A(z)}$. In Fig. 5.1 I compare the solutions of the least square problems

$$\hat{\theta}_{tn} \doteq \Phi_0 \backslash Y, \qquad \hat{\theta}_{nn} \doteq \Phi \backslash Y$$

as a function of the number of elements in the regressor, just as it was done in class for the first problem. Only by looking at the graph, it is not possible to tell whether an estimate is biased, but repeating the experiment reveals that neither $\hat{\theta}_{nn}$ (as expected) nor $\hat{\theta}_{tn}$ (perhaps unexpectedly) are biased.

The explanation is the following: the (noisy) system's evolution is

$$y(k) = (1 - A(z))y(k) + B(z)u(k) + e(k) = \varphi^\top(k)\theta_0 + e(k)$$

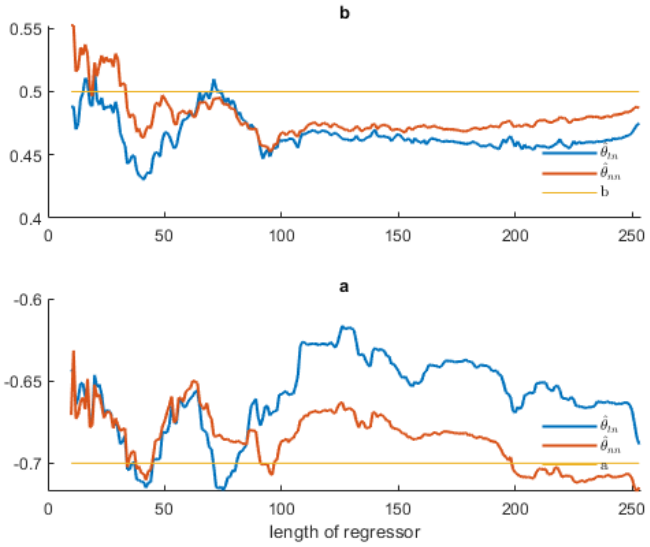

Figure 5.1: Comparison of estimates with true and noisy regressors and data for a PRBS input signal with $O = 7$ and $N = 127$. The estimates start from a minimum regressor length of 10. The code is in `09_lect/SysID_ARX.m`.

and I *pick*[7] the one-step ahead predictor to be $\hat{y}(k,\theta) = \varphi^\top(k)\theta$. Their difference

$$y(k) - \hat{y}(k,\theta) = \varphi^\top(k)(\theta_0 - \theta) + e(k)$$

is minimized when $\theta = \theta_0$ since $e(k)$ is Gaussian distributed: this is the case $\Phi\backslash Y$. On the other hand, we can also write

$$y(k) = y_0(k) + \frac{1}{A(z)}e(k) = \varphi_0^\top(k)\theta_0 + \frac{1}{A(z)}e(k)$$

and I *choose*[8] the predictor to be $\hat{y}(k,\theta) = \varphi_0^\top(k)\theta$. Their difference

$$y(k) - \varphi_0^\top(k)\theta = \varphi_0^\top(k)(\theta_0 - \theta) + \frac{1}{A(z)}e(k)$$

[7] The word "pick" is emphasized: it is the true predictor but, at least for ARX, it seems to me that the most convenient predictor for the solution of the least-squares is whatever regressor the true system uses.

[8] As mentioned already in footnote 4, the noiseless regressor Φ_0 cannot be constructed from the noisy data unless (I think) $H(z) = 1$ when $\hat{y}(k|k-1) = y_0(k)$.

is also minimized when $\theta = \theta_0$ since $\frac{1}{A(z)}e(k)$ has Gaussian distribution (although it is now correlated). This means

$$\mathbb{E}\{\Phi \backslash Y\} = \mathbb{E}\{\Phi_0 \backslash Y\} = \theta_0$$

are both unbiased.

5.2.3 ARMAX Model Structure

The ARX model structure is not very flexible with regards to the noise model requiring it to have the particular structure $\frac{1}{A(z)}$. The ARMAX transfer function model partially relaxes this: it has the form

$$A(z)y(k) = B(z)u(k) + C(z)e(k) \tag{5.12}$$

where $A(z)$ and $B(z)$ are as in ARX and $C(z)$ is monic. It corresponds to the model of eq. (5.1) with

$$G(z) = \frac{B(z)}{A(z)}, \qquad H(z) = \frac{C(z)}{A(z)}$$

and with the one-step ahead predictor[9]

$$\hat{y}(k|k-1) = (1 - A(z))y(k) + B(z)u(k) + (C(z) - 1)\left(y(k) - \hat{y}(k|k-1)\right) \tag{5.13}$$

Introducing the regression vector

$$\varphi^\top(k,\theta) = \begin{bmatrix} -y(k-1) & \dots & u(k-1) & \dots & \epsilon(k-1,\theta) & \dots \end{bmatrix}$$

eq. (5.13) induces the pseudolinear regression

$$\hat{y}(k,\theta) = \varphi^\top(k,\theta)\theta \tag{5.14}$$

the equation is non-linear because the unknown appears also in the regressor $\varphi^\top(k,\theta)$.

Since the regressor is pseudolinear, using non-linear least-squares to estimate θ will result in a biased estimate (why exactly?).

[9] In alternative to plugging the expressions for $G(z)$ and $H^{-1}(z)$ into eq. (5.6), the one-step ahead predictor can be equally easily obtained from eq. (5.12) and eq. (5.7):

$$\begin{aligned}\hat{y}(k|k-1) &= y(k) - e(k)\\ &= (1 - A(z))\, y(k) + B(z)u(k) + (C(z) - 1)\, e(k)\\ &= (1 - A(z))\, y(k) + B(z)u(k) + (C(z) - 1)\left(y(k) - \hat{y}(k|k-1)\right).\end{aligned}$$

This operation only makes sense when $H(z)$ is stably invertible though: at least $C(z)$ cannot have unstable zeros.

Consistency check: this expression reduces to ARX one-step ahead predictor eq. (5.10) when $C(z) = 1$.

5.2.4 Constrained Minimization

Using eq. (5.14) and eq. (5.8), we have that $y(k) = \varphi^\top(k,\theta)\theta + \epsilon(k,\theta)$: this is used as the constraint in the optimization-based algorithm for the solution to the ARMAX problem:

$$\begin{aligned} &\min_{\theta,\epsilon} \; ||\epsilon||_2^2 \\ &\text{subject to } Y = \Phi(\epsilon)\theta + \epsilon \end{aligned}.$$

A reference implementation is in `11_lect/SysID_ARMAX.m`.

5.2.5 General Family of Model Structures

The most general family of model structure is [2, Sect. 4]

$$A(z)y(k) = \frac{B(z)}{F(z)}u(k) + \frac{C(z)}{D(z)}e(k) \tag{5.15}$$

Some of the common cases that we have seen so far are summarized in Table 5.1.

Polynomials	Name of Model Structure
B	FIR
A B	ARX
A B C	ARMAX
A B C D	ARARMAX
B F C D	Box-Jenkins

Table 5.1: Some common black-box SISO models using the polynomials of eq. (5.15).

5.2.6 Known Noise Model (with AR Noise Dynamics)

This was partially addressed in the exercises. If the noise model $L(z)$ is known

$$y(k) = \frac{B(z)}{A(z)}u(k) + \frac{L(z)}{A(z)}e(k)$$

and provided $L(z)$ is stably invertible, by letting $y_L \doteq L^{-1}y$ and $u_L \doteq L^{-1}u$, one obtains the standard ARX form

$$\begin{aligned} y_L(k) &= \frac{B(z)}{A(z)}u_L(k) + \frac{1}{A(z)}e(k) \\ \hat{y}_L(k|k-1) &= \underbrace{(1-A)\,y_L + Bu_L}_{\Phi_L\theta_0} \end{aligned}$$

for which, as usual, $y_L(k) - \hat{y}_L(k|k-1) = e(k)$. The estimate for $A(k,\theta)$, $B(k,\theta)$ is obtained by least-squares minimization of the prediction error $y_L(k) - \hat{y}_L(k,\theta)$. In terms of $y(k)$, the prediction error

$$y(k) - L(z)\hat{y}_L(k|k-1) = L(z)e(k)$$

is the filtered error $L(z)e(k)$ but since filtered Gaussian noise is still Gaussian, the estimate remains unbiased. Note that

$$L(z)\hat{y}_L(k|k-1) \neq \hat{y}(k|k-1)$$

is *not* the one-step ahead prediction for $y(k)$, because we expect the difference to be $e(k)$.

If one were to use eq. (5.6) to compute the one-step ahead predictor for which the prediction error is $e(k)$, one would obtain the linear expression $\Phi_L\theta_0$ plus an offset term

$$\begin{aligned} \hat{y}(k|k-1) &= \frac{B(z)}{L(z)}u(k) + \frac{L(z)-A(z)}{L(z)}y(k) \\ &= Bu_L + (L-A)\,y_L \\ &= \underbrace{(1-A)\,y_L + Bu_L}_{\Phi_L\theta_0} + (L-1)\,y_L \end{aligned}$$

which must be subtracted from $y(k)$.

Chapter 6

Closed-Loop Identification

There are reasons to use system identification in closed loop:

- an unstable system must be operated in closed-loop. A simple controller stabilizes the system, but one may want to have a better model to improve the performance or because the system may change with time;
- operational constraints may require closed-loop: *e.g.* an operational industrial process cannot run in open loop for the purpose of identification because the specs on the final product must still be met;
- closed-loop controller maintains the system close to the operating point of interest (the system may be non-linear and closed loop linearizes it);
- this emphasizes plant dynamics close to the cross-over frequency range by removing a possibly large-scale zero-frequency response which is easy to control with a slow controller.

In open-loop, system identification can be performed in the frequency and time domain. In frequency domain for instance the estimate

$$\hat{G}(e^{j\omega_n}) = \frac{\hat{Y}}{\hat{U}}$$

with

$$\text{bias:} \quad \mathbb{E}\left\{\hat{G}(e^{j\omega_n}) - G(e^{j\omega_n})\right\} \longrightarrow 0as$$

$$\text{variance:} \quad \mathbb{E}\left\{|\hat{G}(e^{j\omega_n}) - G(e^{j\omega_n})|^2\right\} \longrightarrow \frac{\phi_v(e^{j\omega_n})}{\phi_u(e^{j\omega_n})}$$

as $N \to 0$. In closed-loop, the identification results may not be as good.

6.1 Methods for Closed-Loop Identification

The fundamental assumption to derive the results until now was that control input u and noise e were uncorrelated: $\phi_{ue}(e^{j\omega}) = 0$. In closed-loop this is no longer the case: the noise on the output is seen at the input through the feedback loop.

For identification, we assume

- a generalized reference $r(k) = r_2(k) + C(z)r_1(k)$;
- $y(k)$ and $u(k)$ are available;
- $C(z)$ stabilize the system and makes it internally stable: that is, all transfer functions (the Gang of Four)

$$\frac{GC}{1+GC}, \frac{G}{1+GC}, \frac{C}{1+GC}, \frac{1}{1+GC}$$

are stable.

There are four main methods for closed-loop identification:

- direct methods;
- indirect methods;
- joint input-output methods;
- dual-Youla parametrization.

6.1.1 Direct Methods

It applies the basic prediction error method Sect. in a straightforward manner: use the output y of the process and the input u in the same way as for open loop operation, ignoring any possible feedback, and not using the reference signal r. The method works regardless of the complexity of the regulator and requires no knowledge about the character of the feedback [2].

The closed loop transfer functions are

$$y = SGr + Sv$$
$$u = Sr - SCv$$

where $S(e^{j\omega})$ is the stable sensitivity function for the closed loop system

$$S(e^{j\omega}) = \frac{1}{1 + C(e^{j\omega})G_0(e^{j\omega})}.$$

Using spectral analysis[1], and assuming $\phi_{rv} = 0$, we have that

$$\hat{\phi}_{yu}(e^{j\omega}) = |S|^2 G\phi_r - |S|^2 \bar{C}\phi_v$$
$$\hat{\phi}_u(e^{j\omega}) = |S|^2 \phi_r - |S|^2 |C|^2 \phi_v.$$

The direct method estimates G directly from $\hat{\phi}_{yu}$ and $\hat{\phi}_u$:

$$\hat{G}(e^{j\omega}) = \frac{\hat{\phi}_{yu}(e^{j\omega})}{\hat{\phi}_u(e^{j\omega})} = \frac{G\phi_r - \bar{C}\phi_v}{\phi_r - |C|^2\phi_v}$$

which converges to G when the contribution from the reference signal dominates the noise.

The simplification of $|S|^2$ hides the fact that for frequencies for which $|S|^2 \sim 0$, *e.g.* when the loop transfer function $C(z)G(z)$ contains an integrator $\sim s^{-1}$, the measured signals $\hat{\phi}_{yu}$ and $\hat{\phi}_u$ are zero. On the other hand, in every practical control system with tracking, S has a bump at around the closed loop BW (is this true?): those are the frequencies that get emphasized and are relevant for the stability.

6.1.2 Indirect Methods

It identifies the closed loop transfer function[2] $T_{yr}(z)$ from reference input $r(k)$ to output $y(k)$, and retrieves the open loop system $G(z)$, making use of the knowledge of the regulator $C(z)$ [2].

Given the closed loop system

$$y(k) = T_{yr}(z)r(k) + v_{cl}(k) = \frac{G(z)}{1 + G(z)C(z)}r(k) + \frac{1}{1 + G(z)C(z)}v(k)$$

the open loop transfer function estimate $\hat{G}(z)$ is retrieved from

$$T_{yr}(z) = \frac{\hat{G}(z)}{1 + \hat{G}(z)C(z)}.$$

Only the estimate $T_{yr}(z)$ is asymptotically unbiased because the reference r is known; $\hat{G}$ (probably) does not converge to G because the transformation is non-linear which does not preserve the mean.

The advantage with the indirect method is that any idenfication method can be applied to estimate $T_{yr}(z)$. On the other hand, any error in the knowledge of $C(z)$ will be reflected in $\hat{G}(z)$.

[1] One more time, I have the impression that the result could have equally well been expressed in terms of ETFE without the need of using the correlations.

[2] In class the method was described in the frequency domain; Ljung does it in the time-domain.

6.1.3 Joint Input-Output Methods

It relies on independent measurements of y and u

$$y = SGr + Sv = T_{yr}r + Sv$$
$$u = Sr - SCv = T_{ur}r - SCv$$

so that their noise is uncorrelated, to estimate (asymptotically unbiased) $\hat{T}_{yr}(z)$ and $\hat{T}_{ur}(z)$; their noise is also uncorrelated. Since

$$\frac{T_{yr}}{T_{ur}} = \frac{SG}{S} = G$$

the estimate for $\hat{G}$ follows as the ratio

$$\hat{G}(z) = \frac{\hat{T}_{yr}(z)}{\hat{T}_{ur}(z)}.$$

As before, since the estimated spectra are weighted by S or $S(z)C(z)$ and S may become small, some frequencies may not be reliably resolved when taking the ratio. Key points:

- $\hat{G}$ may not be unbiased (unless the input signal is periodic?);
- the noise enters in a complicated manner;
- one key advantage: C does not need to be known.

This method can be seen as the specific case of a more general framework [2, Sect. 13.5] that works also for large interconnected systems, where there is no measurable reference r (*e.g.* large interconnected systems where it is also not possible to model the controller). The model is

$$y = GS(r + w) + Sv = G_{cl}r + v_1$$
$$u = S(r + w) - CSv = T_{ru}r + v_2$$

When including the correlations between $v_1 = Sv + GSw$ and $v_2 = -CSv + Sw$ gives

$$\begin{bmatrix} y \\ u \end{bmatrix} = \mathcal{G}r + \mathcal{H}v.$$

When instead the correlations are ignored gives the method described at the beginning of this section.

6.1.4 Dual-Youla Methods

It relies on coprime factorization of transfer functions

$$G(s) = \frac{N_0(s)}{D_0(s)}$$

where $N_0(s)$ and $D_0(s)$ are stable and have no common zeros.

The Bezout identity states that $N_0(s)$ and $D_0(s)$ are coprime iff there exists U and V such that

$$UN_0 + VD_0 = I.$$

A coprime factorization is "normalised" if

$$D_0^* D_0 + N_0^* N_0 = I.$$

The MATLAB command is `sncfbal`.

The Youla parametrisation is a way of parametrize all stable controllers: given a controller $C_0 = \frac{X_0}{Y_0}$ with X_0, Y_0 a coprime factorization and stable (an integral controller would not work) which stabilizes G_0, all controllers C stabilizing $G_0 = N_0/D_0$ have the form

$$C_Q = \frac{X_0 + QD_0}{Y_0 - QN_0}$$

with Q stable.

The dual Youla parametrization method takes the opposite route: given the known controller C that stabilizes the system, the plant G must be one of those that can be stabilized by C. The problem can be therefore formulated as a search on stable[3] Q: find the estimate $\hat{G}$ from the set of all plants stabilized by $C(s)$.

We model the open-loop system as

$$y(k) = \frac{N}{D}u(k) + \frac{F}{D}e(k) \rightarrow Dy = Nu + Fe \tag{6.1}$$

with F stable and stably invertible[4]. Let $C_0 = \frac{X_0}{Y_0}$ any stabiling controller: the choice of X_0, Y_0 makes a difference only from a numerical point of view. The parametrization gives

$$G_Q = \frac{N}{D} = \frac{N_0 + QY_0}{D_0 - QX_0}, \qquad H_{Q,F} = \frac{F}{D} = \frac{F}{D_0 - QX_0}$$

[3] In class we used R as the stable search transfer function, but R was used for the transient in frequency-domain and r for the closed-loop reference in time-domain. To avoid confusion, I use Q.

[4] I guess this means all zeros and poles strictly inside the unit circle.

The equivalent open loop identification experiment is obtained by rewriting eq. 6.1 as

$$(D_0 - QX_0)y = (N_0 + QY_0)u + Fe$$

or equivalently, after rearranging the terms, as

$$\begin{aligned}
\beta &\doteq D_0 y - N_0 u \\
\alpha &\doteq X_0 y + Y_0 u = X_0 \left(y + \frac{Y_0}{X_0} u \right) = X_0 r \\
\beta &= Q\alpha + Fe.
\end{aligned}$$

where the quantity $r = y + \frac{Y_0}{X_0} u$ is the reference signal r.

As it is written, this is an open-loop since there is no feedback between β and α. The procedure for the dual-Youla method is the following: given a stabilizing controller C_0

- factorise $C_0 = X_0/Y_0$;
- choose the excitation r;
- run closed-loop experiments with C_0, measuring y and u.
- choose an initial model, $G_0 = \frac{N_0}{D_0}$ (must be stabilised by C_0);
- filter the measurements, $\beta = D_0 y - N_0 u$ (time or frequency domain);
- filter the excitation $\alpha = Y_0 r$;
- estimate $\hat{Q}$ and $\hat{F}$ from $\beta = Q\alpha + Fe$;
- calculate the plant estimate $\hat{G} = (N_0 + \hat{Q}Y_0)(D_0 - \hat{Q}X_0)$.

Chapter 7

Subspace Identification

Subspace identification aims at constructing a state space representation

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \qquad & A \in \mathbb{R}^{n_x \times n_x} \\ y(k) &= Cx(k) + Du(k) \qquad & D \in \mathbb{R}^{n_y \times n_u} \end{aligned} \tag{7.1}$$

from the (time- or frequency-domain) data. n_x, n_u and n_y are the number of states, control inputs and outputs. This approach handles also MIMO systems and use only linear algebra.

The subspace identification method presented in this chapter makes use of the identity in the frequency-domain

$$G(e^{j\omega}) = C\left(e^{j\omega}\mathbb{I} - A\right)^{-1} B + D \qquad G(e^{j\omega}) \in \mathbb{R}^{n_y \times n_u}$$

where for the estimates, we require *consistency*

$$\lim_{N\to\infty} \text{Prob}\left\{ \left|\left| \hat{G}(e^{j\omega_n}) - G(e^{j\omega_n}) \right|\right|_\infty > \delta \right\} = 0$$
$$\forall \delta > 0 \text{ and } \forall n = 0, \cdots, N-1$$

on a uniform grid of frequencies ω_n. We also require *correctness*, that the algorithms produce the true model if the noise is zero given a finite amount of data M [3].

In constructing the matrix $\hat{A}$, the number of states n_x is, in general, not known in advance since they may all not be measurable and may be unphysical; on the other hand, prior knowledge of the system, when available, must be used. A second issue for state-stace representation arised from the fact that the matrices are not unique, and only related by a non-singular transformation[1].

[1] Given the transformation T, the two systems described by

$$A,\ B,\ C,\ D$$

In the time-domain, the pulse response coefficients for a causal system[2] are given by

$$g(k) = \begin{cases} 0 & k < 0 \\ D & k = 0 \\ CA^{k-1}B & k > 0 \end{cases} \tag{7.2}$$

The observability and controllability matrices play an important role in subspace identification. The (extended) observability matrix is

$$\mathcal{O}_q \doteq \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix} \in \mathbb{R}^{n_y q \times n_x}$$

is a tall matrix since $n_y q \geq n_x$. A system is *observable* if for every $T > 0$ it is possible to determine the state of the system $x(T)$ through measurements of $y(t)$ and $u(t)$ on the interval $[0, T]$ [1, Sect. 8]. If $\operatorname{rank}(\mathcal{O}_q) = n_x$, the matrix has full column rank and the system is observable.

The extended controllability/reachability matrix

$$\mathcal{C}_r \doteq \begin{bmatrix} B & AB & \cdots & A^{r-1}B \end{bmatrix} \in \mathbb{R}^{n_x \times n_u r}$$

where $n_u r \times n_x$ is a fat matrix. A system is *reachable* if there exists a T and a sequence of inputs that brings the initial state x_0 into the final state x_f. If $\operatorname{rank}(\mathcal{C}_r) = n_x$, the matrix has full row rank and the system is controllable.

---

and

$$\tilde{A} \doteq T^{-1}AT,\ \tilde{B} \doteq TB,\ \tilde{C} \doteq CT,\ \tilde{D} \doteq D$$

have the same input-output behavior

$$C(z\mathbb{I} - A)B + D.$$

[2]The time-domain $g(k)$ and the frequency-domain transfer function $G(e^{j\omega})$ are related as usual by the Fourier transform

$$G(e^{j\omega}) = \sum_{k=0}^{\infty} g(k)e^{-j\omega k} \qquad 0 \leq \omega \leq \pi.$$

7.1 Subspace Identification using the Hankel Matrix

The algorithm described in [3] assumes an observable system (*i.e.* $\mathcal{O}_q$ has full column rank) and consists of the following steps:

1. compute the time-alised estimate of the pulse response $\hat{h}_k$ from the estimated frequency response $\hat{G}(e^{j\omega_n})$;
2. constructs the Hankel matrix $\hat{\mathcal{H}}$;
3. compute a singular value decomposition of $\hat{\mathcal{H}}$ and select the state space order $\hat{n}_x$;
4. estimate the state space matrices $\hat{A}$, $\hat{B}$, $\hat{C}$ and $\hat{D}$.

As for the notations: $\mathcal{H}$ is the Hankel matrix for noise-free data and is exactly constructed from the real A, B, C and D; $\hat{\mathcal{H}}$ is the Hankel matrix constructed from noisy data from which the estimates $\hat{A}$, $\hat{B}$, $\hat{C}$ and $\hat{D}$ are computed.

Step 1: Construct $\hat{h}_k$

Starting from uniformly spaced data for $\hat{G}(e^{j\omega_n})$, compute the time-alised estimate of the pulse response using the inverse DFT

$$\hat{h}_k = \frac{1}{N}\sum_{n=0}^{N-1} \hat{G}(e^{j\omega_n}) e^{j2\pi kn/N}, \quad k = 0, \ldots, N-1.$$

The time-aliased response is also given by[3]

$$h_k \doteq \sum_{l=0}^{\infty} g(k+Nl) = CA^{k-1}\left(\mathbb{I} - A^N\right)^{-1} B$$

provided all eigenvalues of A are strictly inside the unit circle, $\rho(A) < 1$. For a MIMO system, $\hat{h}_k \in \mathbb{R}^{n_y \times n_u}$. If $N \geq \tau_{\max}$, the aliasing effect is small and the term $A^N = 0$.

[3] Using the alised pulse response and eq. (7.2)

$$h_k \doteq \sum_{l=0}^{\infty} g(k+Nl) = CA^{k-1}\left(\sum_{l=0}^{\infty} A^{Nl}\right) B = CA^{k-1}\left(I - A^N\right)^{-1} B$$

Step 2: Construct $\hat{\mathcal{H}}$

Assemble the block[4] Hankel matrix with equal elements on the skew diagonal

$$\hat{\mathcal{H}} = \begin{bmatrix} \hat{h}_1 & \hat{h}_2 & \hat{h}_3 & \cdots & \hat{h}_r \\ \hat{h}_2 & \hat{h}_3 & \ddots & & \hat{h}_{r+1} \\ \hat{h}_3 & \ddots & & & \\ \cdots & & & & \vdots \\ \hat{h}_q & \hat{h}_{q+1} & & \cdots & \hat{h}_{q+r-1} \end{bmatrix} \in \mathbb{R}^{n_y q \times n_u r}$$

where $q > n_x$, $r > n_x$ and $q + r - 1 \leq N - 1$. The matrix does not need to be square, although a square matrix has better numerical properties. The matrix can be factored out in terms of the observability and controllability matrices[5]

$$\mathcal{H} = \mathcal{O}_q \left(I - A^N\right)^{-1} \mathcal{C}_r.$$

Step 3: Compute the SVD of $\hat{\mathcal{H}}$

The singular value decomposition of $\hat{\mathcal{H}}$ gives

$$\hat{\mathcal{H}} = \begin{bmatrix} \hat{U}_1 & \hat{U}_2 \end{bmatrix} \begin{bmatrix} \hat{\Sigma}_1 & 0 \\ 0 & \hat{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \hat{V}_1^\top \\ \hat{V}_2^\top \end{bmatrix}$$

In the noise-free case, the rank of $\mathcal{H}$ is n_x since $\mathcal{H}$ is the product of matrices of rank n_x: $\hat{\Sigma}_1 \in \mathbb{R}^{n_x \times n_x}$ and $\hat{\Sigma}_2 = 0$.

In the presence of noise, $\hat{\mathcal{H}}$ becomes full rank and one must select the state space size $\hat{n}_x$ such that $\Sigma_1 \in \mathbb{R}^{\hat{n}_x \times \hat{n}_x}$. One way of estimating $\hat{n_x}$ is corresponding to the singular value with the largest drop although this is not always clear. Cross-validation is also a possible approach if the tional data is available.

Step 4: Estimate the state-space matrices

Approximating $\hat{\Sigma}_2 = 0$, from

$$\hat{\mathcal{H}} \approx \hat{U}_1 \hat{\Sigma}_1 \hat{V}_1 \top$$

[4]Block because the underlaying system may be a MIMO.

[5]The equivalency can be seen using the property that A commutes with $I - A^N$: for instance

$$h_3 = [\mathcal{H}]_{22} = CA \left(I - A^N\right)^{-1} AB = CA^2 \left(I - A^N\right)^{-1} B.$$

and remembering that $\mathcal{O}_q$ is full column rank (although not orthonormal), we conclude that $\hat{U}_1$ and the estimate $\hat{\mathcal{O}}_q$ are equivalent within a similarity transformation:

$$\hat{U}_1 = \hat{\mathcal{O}}_q T = \begin{bmatrix} \hat{C}T \\ \hat{C}T\left(T^{-1}\hat{A}T\right) \\ \vdots \\ \hat{C}T\left(T^{-1}\hat{A}T\right)^{q-1} \end{bmatrix}$$

Since we are just interested in any estimate, for readability I will make the substitutions $CT \to C$ and $T^{-1}AT \to A$ in the equation above.

The estimate for $\hat{C}$ can be obtained from the first $n_y \times n_x$ block of $\hat{U}_1$: define

$$J_3 \doteq \begin{bmatrix} \mathbb{I}_{n_y} & 0_{n_y \times n_y(q-1)} \end{bmatrix}$$

and

$$\hat{C} = J_3 \hat{U}_1.$$

The estimate for $\hat{A}$ can be found by solving a least-squares problem on $\hat{U}_1$: define

$$J_1 \doteq \begin{bmatrix} \mathbb{I}_{n_y(q-1)} & 0_{n_y(q-1) \times n_y} \end{bmatrix}$$
$$J_2 \doteq \begin{bmatrix} 0_{n_y(q-1) \times n_y} & \mathbb{I}_{n_y(q-1)} \end{bmatrix}$$

and solve for $\hat{A}$

$$J_1 \hat{U}_1 \hat{A} = J_2 \hat{U}_1.$$

For the estimates $\hat{B}$ and $\hat{D}$, solve the linear least-squares problem[6]

$$\hat{B}, \hat{D} = \underset{B,D}{\operatorname{argmin}} \sum_{n=0}^{N-1} \left|\left| (e^{j\omega_n}) - D - \hat{C}\left(e^{j\omega_n}\mathbb{I} - \hat{A}\right)^{-1} B \right|\right|_F^2$$

in the Frobinius norm

$$||M||_F^2 = \sum_i \sum_j [M]_{ij}^2.$$

[6]In MATLAB, to obtained real-valued $\hat{B}$ and $\hat{D}$, solve the problems

$$\begin{bmatrix} \text{real}\left(\hat{C}\left(e^{j\omega_n}\mathbb{I} - \hat{A}\right)^{-1}\right) & \mathbb{I} \\ \text{imag}\left(\hat{C}\left(e^{j\omega_n}\mathbb{I} - \hat{A}\right)^{-1}\right) & 0 \end{bmatrix} \begin{bmatrix} B \\ D \end{bmatrix} = \begin{bmatrix} \text{real}\left(\hat{G}(e^{j\omega_n})\right) \\ \text{imag}\left(\hat{G}(e^{j\omega_n})\right) \end{bmatrix}$$

where the matrices for each frequency ω_n are stacked one on top of the other.

7.2 Nonuniformly-Spaced Frequency Case

In the frequency domain, eq. (7.1) becomes

$$\begin{aligned} e^{j\omega}X(e^{j\omega}) &= AX(e^{j\omega}) + BU(e^{j\omega}) \\ Y(e^{j\omega}) &= CX(e^{j\omega}) + DU(e^{j\omega}) \end{aligned}$$

For each input

$$U_i(e^{j\omega}) = e_i, \quad i = 1, \cdots, n_u$$

where e_i is the vector with entry 1 at the i-th location and zero elsewhere, we define the equations

$$\begin{aligned} e^{j\omega}X_i(e^{j\omega}) &= AX_i(e^{j\omega}) + BU_i(e^{j\omega}) \\ Y_i(e^{j\omega}) &= CX_i(e^{j\omega}) + DU_i(e^{j\omega}) \end{aligned}$$

where $X_i(e^{j\omega})$ and $Y_i(e^{j\omega})$ are the states and output when $U_i(e^{j\omega})$ is active. Defining

$$X_c(e^{j\omega}) \doteq \begin{bmatrix} X_1(e^{j\omega}) & \cdots & X_{n_u}(e^{j\omega}) \end{bmatrix}$$

and stacking the equations column-wise

$$\begin{aligned} e^{j\omega}X_C(e^{j\omega}) &= AX_C(e^{j\omega}) + B \\ G(e^{j\omega}) &= CX_C(e^{j\omega}) + D \end{aligned}$$

By multiplying the second equation by $e^{j\omega}$ and using the first one, one obtains

$$e^{j\omega}G(e^{j\omega}) = CAX_C(e^{j\omega}) + CB + De^{j\omega}.$$

Repeating this operation, one obtains

$$\begin{bmatrix} G(e^{j\omega}) \\ e^{j\omega}G(e^{j\omega}) \\ \vdots \\ e^{j(q-1)\omega}G(e^{j\omega}) \end{bmatrix} = \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix}}_{\doteq \mathcal{O}} X_C(e^{j\omega}) + \Gamma \begin{bmatrix} \mathbb{I}_{n_u} \\ e^{j\omega}\mathbb{I}_{n_u} \\ \vdots \\ e^{j(q-1)\omega}\mathbb{I}_{n_u} \end{bmatrix}$$

with

$$\Gamma \doteq \begin{bmatrix} D & 0 & \cdots & & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & \ddots & & \\ \vdots & \ddots & & & \\ CA^{q-2}B & & \cdots & CB & D \end{bmatrix}$$

unknown, is the matrix of impulse responses, this time starting with $g_0 = D$. For each of the N available frequencies ω_n (non necessarily uniformly spaced), stacking the results above column-wise, one obtains

$$\mathcal{G} = \mathcal{O}\mathcal{X}_C + \Gamma\mathcal{W} \tag{7.3}$$

where

$$\mathcal{G} \doteq \frac{1}{\sqrt{N}} \begin{bmatrix} G(e^{j\omega_1}) & \cdots & G(e^{j\omega_N}) \\ e^{j\omega_1}G(e^{j\omega_1}) & \cdots & e^{j\omega_N}G(e^{j\omega_N}) \\ \vdots & & \vdots \\ e^{j(q-1)\omega_1}G(ej\omega_1) & \cdots & e^{j(q-1)\omega_N}G(ej\omega_N) \end{bmatrix}$$

$$\mathcal{W} \doteq \frac{1}{\sqrt{N}} \begin{bmatrix} \mathbb{I} & \cdots & \mathbb{I} \\ e^{j\omega_1}\mathbb{I} & \cdots & e^{j\omega_N}\mathbb{I} \\ \vdots & & \vdots \\ e^{j(q-1)\omega_1}\mathbb{I} & \cdots & e^{j(q-1)\omega_N}\mathbb{I} \end{bmatrix}$$

$$\mathcal{X}_C \doteq \frac{1}{\sqrt{N}} \begin{bmatrix} X_C(e^{j\omega_1}) & \cdots & X_C(e^{j\omega_N}) \end{bmatrix}$$

The equation can be split into real and imaginary parts:

$$\underbrace{\begin{bmatrix} \text{real}\,(G) & \text{imag}\,(G) \end{bmatrix}}_{\mathcal{G}_r} = \mathcal{O}\,\underbrace{\begin{bmatrix} \text{real}\,(\mathcal{X}_C) & \text{imag}\,(\mathcal{X}_C) \end{bmatrix}}_{\mathcal{X}_C r} + \Gamma\,\underbrace{\begin{bmatrix} \text{real}\,(\mathcal{W}) & \text{imag}\,(\mathcal{W}) \end{bmatrix}}_{\mathcal{W}_r}$$

Since $\mathcal{W}_r$ is a fat matrix, if $n_y q < n_u r$, there exists a nullspace

$$\mathcal{W}_r^{\perp} = \mathbb{I} - \mathcal{W}_r^{\top}\left(\mathcal{W}_r\mathcal{W}_r^{\top}\right)^{-1}\mathcal{W}_r$$

such that $\mathcal{W}_r\mathcal{W}_r^{\perp} = 0$. This can be used to multiply from the right to get rid of the unknown Γ

$$\mathcal{G}_r\mathcal{W}_r^{\perp} = \mathcal{O}\mathcal{X}_{Cr}\mathcal{W}_r^{\perp}.$$

$\mathcal{G}_r\mathcal{W}_r^{\perp}$ is the equivalent of the Hankel matrix for the uniformly spaced frequency response, on which to perform the SVD and extract the estimates $\hat{A}$, $\hat{B}$, $\hat{C}$ and $\hat{D}$.

7.3 Summary

Subspace identification is a popular model with the following advantages:

- time- and frequency-domain versions available: N4SID, *etc*;
- many variants which depend on weighting for noise;
- gives a state-space model directly;
- can be effective in determining system order;
- works equally well for MIMO systems;

and disadvantages:

- unusual noise weighting in frequency-domain case;
- truncated SVD reconstructions are not Hankel;
- $\hat{U}_1$ does not have the "shift" structure;
- least-squares noise assumptions are not correct.;
- can give unstable models for stable systems.

Bibliography

[1] Karl Johan Åström and Richard M. Murray. *Feedback Systems, An Introduction for Scientists and Engineers*. second edition. Available at `http://fbsbook.org`, version v3.1.5. July 24, 2020.

[2] Lennart Ljung. *System Indentification: Theory for the User*. Second edition. Prentice Hall PTR, 2009.

[3] Tomas McKelvey, Hüseyin Akçay, and Lennart Ljung. "Subspace-based multivariable system identification from frequency response data". In: *IEEE Trans. Auto. Ctrl* 41.7 (1996), pp. 960–979.

[4] *Regularized System Identification: Learning Dynamic Models from Data*. Springer, 2021.

[5] Roy Smith. "Supplementary notes: lecture 4". `https://people.ee.ethz.ch/~rsmith/idfiles/SysID_lecture04_supplement.pdf`.