Subspace-based Multivariable System Identification

from Frequency Response Data

Hüseyin Akçay

Department of Electrical and Electronics Engineering
Anadolu University, Eskişehir, Turkey

October 10, 2008

Outline

The properties of a closed-loop system can, for single-input, single-output (SISO) systems, very accurately and intuitively be determined by studying the frequency response function.

- The classical lead-lag compensator design is done entirely by shaping the Nyquist plot or the Bode plot of the open-loop system.
- From this perspective, it is quite natural to also consider performing system identification in the frequency domain, i.e., determining low order, linear models given samples of the frequency response.

- Most modern multivariable control design techniques are based on state-space models of the systems.
- Algorithms using time-domain measurements:
  - Iterative or prediction error methods (Ljung:1999).
  - Non-iterative, *i.e.,* subspace based algorithms (De Moor and Vandewalle:1987,Verhaegen and Dewilde:1992, Van Overschee and De Moor: 1994).
- Non-iterative methods do not involve nonlinear parametric optimization.
- Subspace-based algorithms deliver state-space models without the need for an explicit parameterization of the model set.

- There is no difference between MIMO system identification and SISO system identification for a subspace-based algorithm.
- Estimated models delivered in a state-space basis, wherein the transfer function is insensitive to small perturbations in the matrix elements, leading to the ability to identify high-order systems.
- Subspace-based algorithms analyzed with respect to consistency, and asymptotic expressions for the quality of the estimates derived (Viberg *etal.*: 1991, De Moor: 1993).

- If the excitation of the system is well-designed, each transfer function measurement is of high quality. Data originating from different experiments can easily be combined in the frequency domain.
- The problem of fitting a real-rational model to a given frequency response has been addressed by many authors (see, Pintelon *etal.*: 1994).
- A system is modeled as a fraction of two polynomials and a NLS fit to the frequency response data is sought. The solution to this nonlinear parametric optimization problem is obtained by iterative, numerical search, *i.e.,* SK-iterations (Sanathanan and Koerner: 1963).

- A non-iterative algorithm based on Markov parameter estimates proposed in Juang and Suzuki: 1988 yields very poor estimates for lightly damped systems.
- A frequency domain approach proposed by Liu and coworkers (1994) is a frequency domain counterpart of the time domain subspace methods by De Moor and Vandewalle: 1987 and Liu and Skelton: 1993.

Objectives

To introduce two new algorithms and provide stochastic analysis regarding their consistency properties

The features:

- Given samples of the frequency function, minimal MIMO state-space models are delivered by the algorithms.
- A key step is the extraction of a low-dimensional subspace by the use of a truncated SVD of a noisy data matrix.
- Non-iterative.
- Strongly consistent.
- Correct.

- G stable, MIMO, linear-time invariant, discrete-time system with input-output properties characterized by the impulse response coefficients g_k through the equation:

$$y(t) = \sum_{k=0}^{\infty} g_k u(t-k)$$

where $y(t) \in \mathbf{R}^p$, $u(t) \in \mathbf{R}^m$, and $g_k \in \mathbf{R}^{p\times m}$.

- The system is of finite order n.

State-space model:

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

The frequency response is calculated as

$$G(e^{j\omega}) = \sum_{k=1}^{\infty} g_k e^{-j\omega k}, \qquad 0 \leq \omega \leq \pi$$

or for the state-space model it can be written as

$$G(e^{j\omega}) = D + C(e^{j\omega} I_n - A)^{-1} B.$$

- If (A, B, C, D) has a minimal McMillan degree, the extended observability matrix $\mathcal{O}$ and $\mathcal{C}$ defined by,

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix} \in \mathbf{R}^{qp \times n},$$

$$\mathcal{C} = \begin{bmatrix} B & AB & \cdots & A^{r-1} \end{bmatrix} \in \mathbf{R}^{n \times rm}$$

both have full rank n for all values $q, r \geq n$.

Given: Noise-corrupted M samples of the frequency response

$$G_k = G(e^{j\omega_k}) + n_k, \qquad k = 1, \cdots, M.$$

Find: An identification algorithm which maps data G_k to a finite-dimensional transfer matrix $\hat{G}^M(e^{j\omega})$ such that

1. with probability one (w.p.1),

$$\lim_{M\to\infty} \|\hat{G}^M - G\|_\infty = 0 \tag{1}$$

   where $\|X\|_\infty = \sup_\omega \sigma_1(X(e^{j\omega}))$ and σ_1 denotes the largest singular value.

2. The algorithm produces the true model if the noise is zero ($n_k = 0$) given a finite amount of data M, *i.e.,* there exists some $M_0 < \infty$ such that

$$\|\hat{G}^M - G\|_\infty = 0, \qquad \text{for all } M > M_0. \tag{2}$$

- Algorithms satisfying (1) are called *strongly consistent*.
–As the amount of data increases, the estimate should improve and asymptotically the correct model should be obtained.
- Algorithms with the property (2) are called *correct*.
–In practice, only a finite amount of data is available. Particularly important for lightly damped systems.
- By use of the bilinear transformation, continuous-time identification problem can be converted to a discrete-time identification problem.

Assume that $M+1$ frequency response data G_k on a set of uniformly spaced frequencies

$$\omega_k = \frac{\pi k}{M}, \qquad k = 0, \cdots, M$$

are given.

- If g_k are given, well-known realization algorithm can be used to obtain a state-space realization (Ho and Kalman: 1966; Kung: 1978; Juang and Pappa: 1985; Juang and Suzuki: 1988).

  – The algorithm to be proposed is closely related to these algorithms, but uses the coefficients of the inverse DFT from samples of the frequency response function.

Outline

Algorithm 1

- 1. Extend the transfer function samples to the full circle

$$G_{M+k} = G^*_{M-k}, \qquad k = 1, \cdots, M-1$$

where $()^*$ denotes complex conjugate.

- 2. Let $\hat{h}_i$ defined by the 2M-point IDFT

$$\hat{h}_i = \frac{1}{2M} \sum_{k=0}^{2M-1} G_k e^{j2\pi ik/2M}, \qquad i = 0, \cdots, 2M-1.$$

- 3. Let the block Hankel matrix $\hat{H}$ be defined as

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

with number of block rows $q > n$ and block columns $r \geq n$.
–The dimension of $\hat{H}$ is bounded by $q + r \leq 2M$.

- 4. Calculate the singular value decomposition of the Hankel matrix
$$\hat{H} = \hat{U}\hat{\Sigma}\hat{V}^T.$$
- 5. Determine the system order n by inspecting the singular values and partition the SVD such that $\hat{\Sigma}_s$ contains the n largest singular values
$$\hat{H} = \begin{bmatrix} \hat{U}_s & \hat{U}_o \end{bmatrix} \begin{bmatrix} \hat{\Sigma}_s & 0 \\ 0 & \hat{\Sigma}_o \end{bmatrix} \begin{bmatrix} \hat{V}_s^T \\ \hat{V}_o^T \end{bmatrix}.$$
- 6. Determine the system matrices $\hat{A}$ and $\hat{C}$ as
$$\hat{A} = (J_1\hat{U}_s)^\dagger J_2\hat{U}_s, \qquad \hat{C} = J_3\hat{U}_s$$

where

$$
\begin{aligned}
J_1 &= \begin{bmatrix} I_{(q-1)p} & 0_{(q-1)p\times p} \end{bmatrix}, \\
J_2 &= \begin{bmatrix} 0_{(q-1)p\times p} & I_{(q-1)p} \end{bmatrix}, \\
J_3 &= \begin{bmatrix} I_p & 0_{p\times (q-1)p} \end{bmatrix},
\end{aligned}
$$

and I_i denotes the $i \times i$ identity matrix, $O_{i\times j}$ the $i \times j$ zero matrix, and $X^{\dagger} = (X^T X)^{-1} X^T$ the Moore-penrose pseudo-inverse of the full column rank matrix X.

- 7. Solve a least-squares problem to determine $\hat{B}$ and $\hat{D}$

$$
\hat{B}, \hat{D} = \arg \min_{\substack{B\in\mathbf{R}^{n\times m} \\ D\in\mathbf{R}^{p\times m}}} \sum_{k=0}^{M} \left\| G_k - D - \hat{C}(e^{j\omega_k} I - \hat{A})^{-1} B \right\|_F^2
$$

where $\|X\|_F = (\sum_k \sum_s |x_{ks}|^2)^{1/2}$ denotes the Frobenius norm.

- 8. The estimated transfer function is defined as

$$\hat{G}^M(z) = \hat{D} + \hat{C}(zI - \hat{A})^{-1}\hat{B}.$$

• B and D appear linearly in $G(z)$ for fixed A and C. Hence, Step 7 has an analytical solution.

Uniqueness of the least-squares solution?

Lemma Let $A \in \mathbf{R}^{n \times n}$, $C \in \mathbf{R}^{p \times n}$, $M \geq n$ and define

$$\mathcal{X} = \begin{bmatrix} C(z_0 I - A)^{-1} & I_p \\ C(z_1 I - A)^{-1} & I_p \\ \vdots & \vdots \\ C(z_M I - A)^{-1} & I_p \end{bmatrix} \in \mathbf{R}^{(M+1)p \times (n+p)}$$

with distinct z_i ($z_i \neq z_j$, $i \neq j$) and z_i's do not coincide with any of the eigevenvalues of A. Then

$$\mathrm{rank} \mathcal{X} = n + p \Leftrightarrow (A, C) \text{ is observable.}$$

- When data are uniformly-spaced and we use (A, C) which are observable and A is stable, the conditions in the lemma are naturally met.

A dual result

Lemma Let $A \in \mathbf{R}^{n\times n}$, $B \in \mathbf{R}^{n\times m}$, and define

$$\mathcal{X} = \left[(z_1 I - A)^{-1}B \;\; z_2 I - A)^{-1}B \;\; \cdots \;\; (z_n I - A)^{-1}B \right]$$

with distinct z_i ($z_i \neq z_j$, $i \neq j$) and z_i's do not coincide with any of the eigevenvalues of A. Then

$$\mathrm{rank}\mathcal{X} = n \Leftrightarrow (A, B) \text{ is controllable.}$$

- A, B, C, D are all real matrices since g_k is real-valued.

The full rank properties of a complex matrix are transferred to the compound matrix constructed from the real and the imaginary parts:

Lemma Let $Z \in \mathbf{C}^{n\times m}$, $n > m$. Then,

$$Z \text{ has full rank } \Leftrightarrow \begin{bmatrix} \mathrm{Re}Z \\ \mathrm{Im}Z \end{bmatrix} \text{ has full rank.}$$

- Algorithm 1 is correct.

Theorem 1 Let G be a stable discrete-time system of order n and let $G_k = G(e^{j\pi k/M}),\ k = 0, \cdots, M$. Let $\hat{G}^M(e^{j\omega})$ be given by Algorithm 1 with $q > n$ and $r \geq n$. Then for all $M > n$

$$\|\hat{G}^M - G\|_\infty = 0.$$

- This result is by no means unique, *i.e.,* the same result is achieved by many identification algorithms such as Levy's method (1959).
- Algorithm 1 exactly recovers finite-dimensional rational transfer functions from a finite number of data in contrast to algorithms wherein the estimated model parameters depend linearly on the measured data. The difference between algorithms becomes more pronounced as the poles of the system move toward the unit circle.
- In the limit, Kung's realization algorithm, *i.e.,*

$$\lim_{M\to\infty} \hat{h}_k = g_k .$$

- $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ is balanced:

$$\hat{\mathcal{O}}^T \hat{\mathcal{O}} = I, \qquad \hat{\mathcal{C}} \hat{\mathcal{C}}^T = (I - \hat{A}^{2M}) \hat{\Sigma}_s^2 (I - \hat{A}^{2M})^T.$$

- As $M, q, r \to \infty$, $\hat{\mathcal{O}}$ and $\hat{\mathcal{C}}$ converge to the observability and the controllability Gramians and the diagonal elements of $\hat{\Sigma}_s$ converge to the Hankel singular values of $G(z)$.

Outline

Noise assumptions.

Brillinger: 1981; Schoukens and Pintelon: 1991.

1. n_k is a zero-mean complex random variable with covariance

$$\mathrm{E}\begin{bmatrix} \mathrm{Re} n_k \\ \mathrm{Im} n_k \end{bmatrix} [\mathrm{Re} n_s^T \ \ \mathrm{Im} n_s^T] = \begin{bmatrix} \frac{1}{2} R_k & 0 \\ 0 & \frac{1}{2} R_k \end{bmatrix} \delta_{ks}.$$

2. The covariance function is uniformly bounded

$$R_k \leq R.$$

– The noise terms for different frequencies are independent.

– The real and imaginary parts of n_k are independent.

Theorem 2 Let G be a stable linear system of order n and let G_k be its samples corrupted by n_k. Let $\hat{G}^M(z)$ denote the transfer function obtained by Algorithm 1 with $q > n$ and $r \geq n$ using $M + 1$ data points. Then, w.p.1

$$\lim_{M \to \infty} \|\hat{G}^M - G\|_\infty = 0.$$

Alternative ways of calculating A

- Step 6 of Algorithm 1 is based on the relation

$$J_1 \hat{U}_s \hat{A} = J_2 \hat{U}_s$$

which exactly holds when $n_k = 0$, and in the noisy case

$$J_1 \hat{U}_s \hat{A} = J_2 \hat{U}_s + N,$$

$\hat{A}$ minimizes the Frobenius norm of N:

$$\hat{A} = \arg \min_{A \in \mathbf{R}^{n \times n}} \| J_1 \hat{U}_s A = J_2 \hat{U}_s \|_F .$$

Inconsistent with the original equation since both $J_1 \hat{U}_s$ and $J_2 \hat{U}_s$ contain errors!

- With this more correct view we obtain the error model

$$(J_1 \hat{U}_s + N_1)\hat{A} = J_2 \hat{U}_s + N_2$$

and the total least-squares (TLS) method can be applied:

$$\hat{A} = \arg \min_{\substack{A \in \mathbf{R}^{n \times n} \\ (J_1 \hat{U}_s + N_1)\hat{A} = J_2 \hat{U}_s + N_2}} \|[N_1 \quad N_2]\|_F.$$

– The TLS technique for calculating A is also found in the signal processing algorith ESPRIT (Roy and Kailath: 1989).

– Similar performance observed in practice for both the LS and the TLS methods when applied to noisy data.

– The poles of the system, or the eigenvalues of A, can be optimally calculated given the statistical properties of $\hat{O}$ by applying the array signal processing technique of weighted subspace fitting (Swindlehurst *etal.*: 1995). A disadvantage is the introduction of a nonlinear optimization step.

We will develop an algorithm which is applicable for the case when samples of the frequency response

$$G_k = G(e^{j\omega_k}) + n_k, \qquad k = 1, \cdots, M$$

are given at arbitrary, distinct frequencies.

- The algorithm is a direct frequency domain formulation of the time-domain subspace algorithm in Verhaegen and Dewilde: 1992.
- It has some connections with a frequency domain algorithm presented in Liu *etal.*: 1994.
- The consistency of the algorithm will be established for a much larger class of noise sources in comparison with the algorithm in Liu *etal.*: 1994.

Outline

Take the discrete Fourier transform of the state-space equations:

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

where $X(w)$, $U(\omega)$, and $Y(\omega)$ denote the transformed time-domain signals. Let $X^i(\omega)$ denote the resulting state-transform when $U(\omega) = e_i$.

Define the compound state-transform matrix:

$$
X^C(\omega) = \begin{bmatrix} X^1(\omega) & X^2(\omega) & \cdots & X^m(\omega) \end{bmatrix} \in \mathbf{C}^{n \times m}.
$$

The transfer function $G(z)$ can be described in the state-space form

$$\begin{aligned} e^{j\omega} X^C(\omega) &= AX^C(\omega) + B, \\ G(e^{j\omega}) &= CX^C(\omega) + D. \end{aligned}$$

By recursive use, we obtain the relation

$$\begin{bmatrix} G(e^{j\omega}) \\ e^{j\omega} G(e^{j\omega}) \\ \vdots \\ e^{j(q-1)\omega} G(e^{j\omega}) \end{bmatrix} = \mathcal{O} X^C(\omega) + \Gamma \begin{bmatrix} I_m e^{j\omega} \\ e^{j\omega} I_m \\ \vdots \\ e^{j(q-1)\omega} I_m \end{bmatrix}$$

where Γ is the lower triangular block Toeplitz matrix defined by

$$\Gamma = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{q-2}B & CA^{q-3} & \cdots & D \end{bmatrix} \in \mathbf{R}^{pq\times mq}.$$

Form a matrix from the frequency response samples

$$\mathbf{G} = \frac{1}{\sqrt{M}} \begin{bmatrix} G_1 & G_2 & \cdots & G_M \\ e^{j\omega_1})G_1 & e^{j\omega_2}G_2 & \cdots & e^{j\omega_M}G_M \\ \vdots & \vdots & \ddots & \vdots \\ e^{j(q-1)\omega_1}G_1 & e^{j(q-1)\omega_2}G_2 & \cdots & e^{j(q-1)\omega_M}G_M \end{bmatrix} \in \mathbf{C}^{qp\times mM}$$

and let **N** denote a matrix with the same structure as **G** with n_i inserted instead of G_i.

Define the block Vandermonde matrix

$$\mathbf{W}_m = \frac{1}{\sqrt{M}} \begin{bmatrix} I_m & I_m & \cdots & I_m \\ e^{j\omega_1} I_m & e^{j\omega_2} I_m & \cdots & e^{j\omega_M} I_m \\ \vdots & \vdots & \ddots & \vdots \\ e^{j(q-1)\omega_1} I_m & e^{j(q-1)\omega_2} I_m & \cdots & e^{j(q-1)\omega_M} I_m \end{bmatrix} \in \mathbf{C}^{qm \times mM}.$$

Then, we have the matrix equation

$$\mathbf{G} = \mathcal{O}X^C + \Gamma \mathbf{W}_m + \mathbf{N}.$$

- Since the system is minimal, (A, B) is a controllable pair and from the dual lemma $\mathrm{rank}(X^C) = n$. Hence, $\mathrm{rank}(\mathcal{O}X^C) = n$ and $\mathrm{range}(\mathcal{O}X^C) = \mathrm{range}(\mathcal{O})$.

A relation involving only real valued matrices

$$\underbrace{[\mathrm{Re}\mathbf{G}\ \mathrm{Im}\mathbf{G}]}_{\mathcal{G}} = \mathcal{O}\underbrace{[\mathrm{Re}X^C\ \mathrm{Im}X^C]}_{\mathcal{X}} + \Gamma\underbrace{[\mathrm{Re}\mathbf{W}_m\ \mathrm{Im}\mathbf{W}_m]}_{\mathcal{W}} + \underbrace{[\mathrm{Re}\mathbf{N}\ \mathrm{Im}\mathbf{N}]}_{\mathcal{N}}.$$

- When $\mathbf{W}_m$ and X^C are matrices of full rank, $\mathcal{W}$ and $\mathcal{X}$ also have full rank.

Projection matrix onto the nullspace of $\mathcal{W}$:

$$\mathcal{W}^\perp = I - \mathcal{W}^T(\mathcal{W}\mathcal{W}^T)^{-1}\mathcal{W}.$$

After the projection, we obtain the relation

$$\mathcal{G}\mathcal{W}^\perp = \mathcal{O}\mathcal{X}\mathcal{W}^\perp + \mathcal{N}\mathcal{W}^\perp.$$

When $\mathcal{N} = 0$, $\mathrm{range}(\mathcal{G}\mathcal{W}^{\perp}) = \mathrm{range}(\mathcal{O})$ unless rank cancellations occur. A sufficient condition is that the intersection between the row spaces of $\mathcal{W}$ and $\mathcal{X}$ is empty.

Lemma Let $M \geq q + n$. Assume that the frequencies ω_i are distinct and not in the spectrum of A. Then,

$$\mathrm{rank} \begin{bmatrix} \mathbf{W}_m \\ X^C \end{bmatrix} = qm + n \Leftrightarrow (A, B) \text{ controllable.}$$

- When $\mathcal{N} = 0$, $\mathrm{range}(\mathcal{G}\mathcal{W}^{\perp}) = \mathrm{range}(\mathcal{O})$ and a state-space model similar to (A, B, C, D) can be obtained proceeding according to Steps 5-7 in Algorithm 1.

Outline

$\mathcal{G}\mathcal{W}^\perp$ is generically of full rank $qp > n$. Some type of approximation is necessary to obtain a good estimate of the observability range space.

- The n left singular vectors corresponding to the n largest singular values of $\mathcal{G}\mathcal{W}^\perp$ form a strongly consistent estimate of $\mathrm{range}(\mathcal{O})$ (De Moor: 1993) if w.p.1,

$$\text{(i)} \qquad \lim_{M\to\infty} \mathcal{O}\mathcal{X}\mathcal{W}^\perp(\mathcal{N}\mathcal{W}^\perp)^T = 0$$

$$\text{(ii)} \qquad \lim_{M\to\infty} \mathcal{N}\mathcal{W}^\perp(\mathcal{N}\mathcal{W}^\perp)^T = \alpha I$$

for some scalar α.

Condition (ii) is satisfied if $R_k = \alpha I$ for all k and the frequencies are equally spaced.

If the covariance function is known, the row space of $\mathcal{G}\mathcal{W}^{\perp}$ can be weighted with a matrix $\mathbf{K} \in \mathbf{R}^{qp \times qp}$ satisfying

$$\mathbf{K}\mathbf{K}^T = \alpha \mathrm{Re}(\mathbf{W_p} \mathrm{diag}(R_1, \cdots, R_M) \mathbf{W}_p^H) \tag{3}$$

for some $\alpha > 0$.

- The matrix $\mathbf{K}$ can be found by the Cholesky factorization given the covariance data R_k.

The weighted version satisfying the requirement (ii) is

$$\mathbf{K}^{-1}\mathcal{G}\mathcal{W}^{\perp} = \mathbf{K}^{-1}\mathcal{O}\mathcal{X}\mathcal{W}^{\perp} + \mathbf{K}^{-1}\mathcal{N}\mathcal{W}^{\perp}.$$

- The n left singular vectors $\hat{U}_s$ corresponding to the n largest singular values of $\mathbf{K}^{-1}\mathcal{G}\mathcal{W}^{\perp}$ form a strongly consistent estimate of $\mathbf{K}^{-1}\mathcal{O}$. The observability range space is then recovered by $\mathbf{K}\hat{U}_s$.

A numerically efficient way of forming $\mathcal{G}\mathcal{W}^{\perp}$

- Perform a QR-factorization:

$$\begin{bmatrix} \mathcal{W} \\ \mathcal{G} \end{bmatrix} = \begin{bmatrix} R_{11} & 0 \\ R_{21} & R_{22} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix}.$$

- Then, $\mathcal{G}\mathcal{W}^{\perp} = R_{22} Q_2^T$.
- Since Q_2^T is a full rank matrix, $\mathcal{G}\mathcal{W}^{\perp} \sim R_{22}$.

Algorithm 2

- 1. Given the data G_k, ω_k, and R_k, form the matrices $\mathbf{G}$, $\mathbf{W}_m$, and $\mathbf{K}$.
- 2. Calculate the QR-factorization

$$\begin{bmatrix} \mathrm{Re}\mathbf{W}_m & \mathrm{Im}\mathbf{W}_m \\ \mathrm{Re}\mathbf{G} & \mathrm{Im}\mathbf{G} \end{bmatrix} = \begin{bmatrix} R_{11} & 0 \\ R_{21} & R_{22} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix}.$$

- 3. Calculate the SVD

$$\mathbf{K}^{-1} R_{22} = \hat{U} \hat{\Sigma} \hat{V}^T.$$

- 4. Determine the system order by inspecting the singular values, and partition the SVD such that $\widehat{\Sigma}_s$ contains the n largest singular values.

$$\mathbf{K}^{-1} R_{22} = \begin{bmatrix} \hat{U}_s & \hat{U}_o \end{bmatrix} \begin{bmatrix} \hat{\Sigma}_s & 0 \\ 0 & \hat{\Sigma}_o \end{bmatrix} \begin{bmatrix} \hat{V}_s^T \\ \hat{V}_o^T \end{bmatrix}.$$

- 5. Determine the system matrices $\hat{A}$ and $\hat{C}$ as

$$\hat{A} = (J_1 \mathbf{K} \hat{U}_s)^{\dagger} J_2 \mathbf{K} \hat{U}_s, \qquad \hat{C} = J_3 \mathbf{K} \hat{U}_s.$$

- 6. Solve the least-squares problem to determine $\widehat{B}$ and $\widehat{D}$.

$$\hat{B}, \hat{D} = \arg \min_{\substack{B \in \mathbf{R}^{n \times m} \\ D \in \mathbf{R}^{p \times m}}} \sum_{k=1}^{M} \left\| R_k^{-1/2} (G_k - D - \hat{C}(e^{j\omega_k} I - \hat{A})^{-1} B) \right\|_F^2 .$$

- 7. The estimated transfer function is defined as

$$\hat{G}^M(z) = \hat{D} + \hat{C}(zI - \hat{A})^{-1} \hat{B}.$$

Theorem 3 Let G be a stable system of order n and G_k, $k = 1, \cdots, M$ be noise-free samples of the transfer function $G(e^{j\omega})$ at M distinct frequencies ω_k. Furthermore, let $q > n$, $M_0 \geq n + q$, and $K \in \mathbf{R}^{qp \times qp}$ be any nonsingular matrix. Finally, let $\hat{G}^M(z)$ be given by Algorithm 2. Then,

$$\|\hat{G}^M - G\|_\infty = 0, \qquad \text{for all } M \geq M_0.$$

Theorem 4 Let G be a stable system of order n. Let $G_k = G(e^{j\omega_k}) + n_k$, $k = 1, \cdots, M$ be noisy samples of G at M distinct frequencies ω_k. Let n_k satisfy the noise assumptions and have bounded fourth-order moments. Furthermore let $q > n$, **K** be given by (3), and $\hat{G}^M(z)$ be given by Algorithm 2. Then, w.p.1

$$\lim_{M \to \infty} \|\hat{G}^M - G\|_\infty = 0, \qquad \text{for all } M \geq M_0.$$

Relations between the algorithms

- Suppose ω_k are uniformly spaced and cover the full unit circle. Assume that $n_k = 0$ and $\mathbf{K} = I$.

  In Algorithm 2, the observability range space is determined from $\mathbf{GW}^{\perp} \in \mathbf{C}^{qp \times mM}$ where a maximal rank annihilator of $\mathbf{W}$ is used. With equidistant frequencies, the following

$$\tilde{\mathbf{W}}^{\perp} = \frac{1}{\sqrt{M}} \begin{bmatrix} I_m & I_m & \cdots & I_m \\ e^{j2\pi/M} I_m & e^{j4\pi/M} I_m & \cdots & e^{j2\pi r/M} I_m \\ \vdots & \vdots & \ddots & \vdots \\ e^{j2\pi(M-1)/M} I_m & e^{j4\pi(M-1)/M} I_m & \cdots & e^{j2\pi r(M-1)/M} I_m \end{bmatrix}$$

  also annihilates $\mathbf{W}$; yet, it is the smallest rank annihilator among all annihilators of $\mathbf{W}$.

Moreover,

$$\hat{H} = \mathbf{G}\tilde{W}^{\perp}.$$

- Algorithm 2 with this particular choice of the annihilator matrix coincides with Algorithm 1.
- Algorithm in Liu *etal.*: 1994 coincides with Algorithm 2 in the special case that the the Fourier transform of the input signal equals identity at all frequencies. It is only consistent when the frequencie are equidistantly spaced and the noise covariance function is constant and proportial to the identity matrix.

Outline

A stable A can be guaranteed by the following procedure (Maciejowski:1995):

$$\hat{A} = \hat{U}_s^{\dagger} \begin{bmatrix} J_2 \hat{U}_s \\ 0_{p \times n} \end{bmatrix}.$$

The price paid is that the method will not yield the true A matrix even for he noise-free case unless the true system has a finite impulse response or if $q \to \infty$.

We would like to suggest a different approach to guarantee stability by adding an extra projection Step 5b. In this step all unstable eigenvalues are projected into the unit circle. The idea can be implemented in the following way.

- Transform A to the complex Schur form with the eigenvalues λ_i on the diagonal.
- Project any diagonal elements (eigenvalues) satisfying $1 < |\lambda_i| \leq 2$ into the unit disc by $\lambda_i = \lambda_i \left(\frac{2}{|\lambda_i|} - 1 \right)$. Eigenvalues with magnitude $\lambda_i| > 2$ are set to zero. Eigenvalues on the unit circle can be moved into the unit disc by changing the magnitude of the eigenvalue to $1 - \epsilon$ for some small positive ϵ, *i.e.*, $\lambda_i = \lambda_i(1 - \epsilon)$.
- Finally transform $\hat{A}$ back to its original form before proceeding further to determine $\hat{B}$ and $\hat{D}$.

Outline

1. ZOH: A discrete-time model can be estimated using the sampled data, and the continuous-time model is obtained by inverse ZOH-sampling of the discrete-time system.
2. The bilinear transformation

$$s = \frac{2(z-1)}{T(z+1)}.$$

   1. The frequency response is invariant if the frequency scale is prewarped.
   2. The observability and the controllability Gramians, hence the Hankel singular values, are invariant under the bilinear map. Balanced realizations remain balanced, hence well-conditioned.

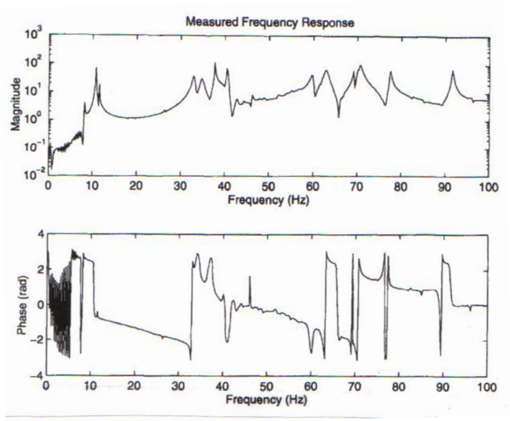
Outline

The data

This application considers the identification of the transfer function between a force-actuator and an accelerometer located on a flexible mechanical structure. The structure is the advanced reconfigurable control (ARC) testbed at the JPL. The ARC testbed is a mechanical truss structure with several active struts and accelerometers at different locations.

The frequency data obtained with a sampling frequency of 200 Hz using a multisine input with 512 equidistant spectral lines.

Measured frequency response of the ARC testbed at JPL.



Quality measures

To assess the quality of estimated models, we will use two measures based on the fit between the data and the model:

1. The maximum error

$$\|\hat{G}^M - G\|_{m,\infty} = \max_{\omega_k} |\hat{G}^M(e^{j\omega_k}) - G_k|$$

2. and the root-mean-square error (rms)

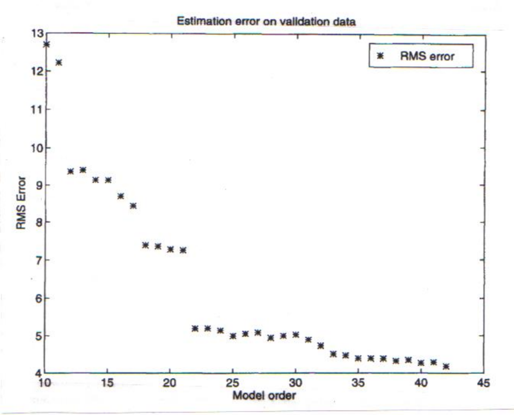
$$\|\hat{G}^M - G\|_{m,2} = \sqrt{\frac{1}{N}\sum_{k=1}^{N} |\hat{G}^M(e^{j\omega_k}) - G_k|^2}.$$

Model Order Determination

The data set is divided into two disjoint sets: the estimation data and the validation data. The division is made such that every second frequency response sample is put in the estimation set and every other in the validation set.

By only using the estimation data, models of different orders are estimated. From the validation data the model error is determined at the frequency points of the validation data. The underlying assumption is that if the model order is low, the error on validation data will decrease as the model order increases until an appropriate model order is found.

Model errors calculated on independent validation data plotted versus order of the estimated models using Algorithm 1.



A comparison study

Let us study the performance of the two new algorithms in comparison with the LS estimate introduced by Levy and the NLS estimate.

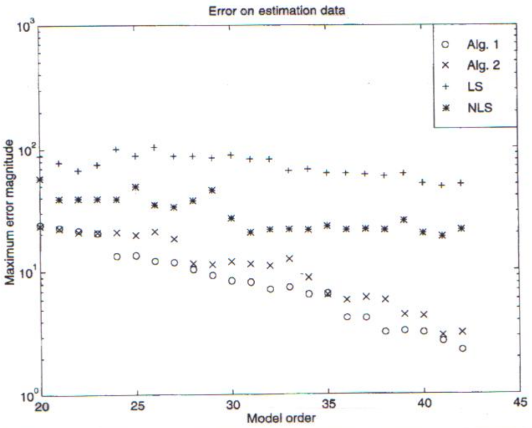
The LS estimate is calculated by minimizing

$$V_{LS}(\theta) = \sum_{k=1}^{M} |G_k a(e^{j\omega_k}, \theta) - b(e^{j\omega_k}, \theta)|^2.$$

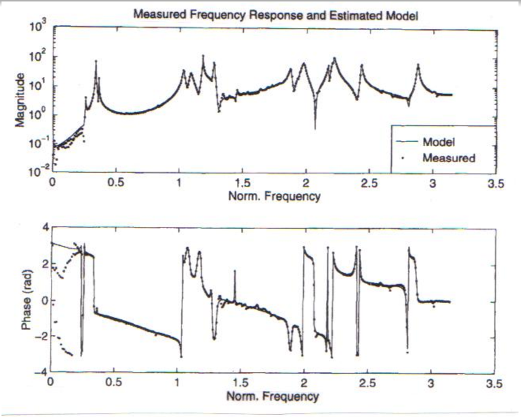
The NLS estimate uses this estimate before proceeding with Gauss- Newton iterations to minimize

$$V_{NLS}(\theta) = \sum_{k=1}^{M} |G_k - \frac{b(e^{j\omega_k}, \theta)}{a(e^{j\omega_k}, \theta)}|^2.$$

Estimation errors for four different methods.



Measured frequency response and estimated model using Algorithm 1. The model is stable and of order 42.



- Two simple state-space identification algorithms to identify linear MIMO systems from samples of the frequency response function were developed.
- Both algorithms were shown to be correct.
- The first algorithm uses uniformly-spaced data and is strongly consistent if the noise is zero mean and has a uniformly bounded covariance function.
- The consistency of the second algorithm is dependent on the a priori knowledge of the noise covariance function. However, it can use an arbitrary frequency spacing.
- The algorithms were used to identify a high-order flexible structure and a comparison with a nonlinear least-squares iterative algorithm was made.