System Identification

Lecture 13: Subspace Identification

Roy Smith

Subspace identification

State-space:

$$x(k+1) = Ax(k) + Bu(k), \qquad A \in \mathcal{R}^{n_x \times n_x}$$
$$y(k) = Cx(k) + Du(k), \qquad D \in \mathcal{R}^{n_y \times n_u}$$

Pulse response coefficients:

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

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

Subspace identification: Frequency domain

Frequency Domain Problem:

Given: $G(n) = G(\mathsf{e}^{j\omega_n}) + V(\mathsf{e}^{j\omega_n}), \quad n = 0, \ldots, N/2,$ (positive frequencies)

Find: $\hat{G}(\mathsf{e}^{j\omega}) = \hat{C}\left(\mathsf{e}^{j\omega} I - \hat{A}\right)^{-1} \hat{B} + \hat{D},$

such that:

$$\lim_{N \longrightarrow \infty} \left\| \hat{G}(\mathsf{e}^{j\omega_n}) - G(\mathsf{e}^{j\omega_n}) \right\|_{\infty} = 0.$$

Key system concept: extended observability subspace

Observability:

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

$\operatorname{rank}(\mathcal{O}) = n_x$ for all $n_y q \geqslant n_x$.

Controllability:

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

$\operatorname{rank}(\mathcal{C}) = n_x$ for all $n_u r \geqslant n_x$.

Time-aliased pulse response

Frequency domain data

ETFE:

$$G(n) \;\approx\; G(\mathrm{e}^{j\omega_n}), \quad \omega_n = \frac{2\pi n}{N}, \quad n = 0, \ldots, N-1.$$

Noise free case ($V(\mathrm{e}^{j\omega_n}) = 0$)

$$G(n) = G(\mathrm{e}^{j\omega_n}), \qquad n = 0, \ldots, N-1.$$

Inverse Fourier transform:

$$h_k = \frac{1}{N}\sum_{n=0}^{N-1} G(\mathrm{e}^{j\omega_n})\mathrm{e}^{j2\pi kn/N} = CA^{k-1}\left(\sum_{l=0}^{\infty} A^{Nl}\right)B$$

$$= CA^{k-1}\left(I - A^N\right)^{-1}B \quad (\text{if } \rho(A) < 1).$$

Hankel matrix (noise-free case)

$$h_k = \frac{1}{N}\sum_{n=0}^{N-1} G(\mathrm{e}^{j\omega_n})\mathrm{e}^{j2\pi kn/N} = CA^{k-1}\left(\sum_{l=0}^{\infty} A^{Nl}\right)B$$

$$H = \begin{bmatrix} h_1 & h_2 & h_3 & \cdots & h_r \\ h_2 & h_3 & \iddots & \iddots & h_{r+1} \\ h_3 & \iddots & \iddots & \iddots & \\ & \iddots & \iddots & & \vdots \\ h_q & h_{q+1} & & \cdots & h_{q+r-1} \end{bmatrix}$$

(choose $q > n_x, r > n_x$
and $q + r - 1 \leqslant N - 1$)

$$= \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix} \left(I - A^N\right)^{-1} \begin{bmatrix} B & AB & \cdots & A^{r-1}B \end{bmatrix}$$

SVD and extended observability space

$$H = \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix}}_{\text{rank} = n_x} \underbrace{\left(I - A^N\right)^{-1}}_{\text{rank} = n_x} \underbrace{\begin{bmatrix} B & AB & \cdots & A^{r-1}B \end{bmatrix}}_{\text{rank} = n_x}$$

$$H = U\Sigma V^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}, \quad \Sigma_1 \in \mathcal{R}^{n_x \times n_x}.$$

$$\text{range}(H) = \text{range}(U_1) = \mathcal{O}$$

Estimating $\hat{C}$ and $\hat{A}$

Estimating $\hat{A}$:

Define J_1 and J_2 so that,

$$J_1 \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-2} \end{bmatrix} \quad \text{and} \quad J_2 \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix} = \begin{bmatrix} CA \\ \vdots \\ CA^{q-1} \end{bmatrix}$$

Then, $J_1 \mathcal{O} A = J_2 \mathcal{O}$,

So, $\hat{A}$ is the least-squares solution to $J_1 \hat{U}_1 \hat{A} = J_2 \hat{U}_1$.

Estimating $\hat{C}$:

$$J_3 \mathcal{O} = C \quad \text{so} \quad J_3 \hat{U}_1 = \hat{C}.$$

B and D matrix estimation

Linear least-squares problem

With $\hat{C}$ and $\hat{A}$ already estimated...

The value of $\hat{C}(\mathrm{e}^{j w_n} I - \hat{A})^{-1}$ is calculated for all ω_n, $n = 0, \ldots, N-1$.

Find $\hat{B}$ and $\hat{D}$ via least-squares:

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

Subspace identification algorithm

Uniform data spacing case:

$$G(n), \quad \omega_n = \frac{\pi n}{N}, \quad n = 0, \ldots, N/2. \qquad \text{(positive frequencies)}.$$

Algorithm [McKelvey *et al.*, 1996]:

1. Extend data to negative frequencies:

$$G(n) = \bar{G}(N-n), \quad n = N/2+1, \ldots, N-1.$$

2. Calculate inverse DFT:

$$\hat{h}_k = \frac{1}{N} \sum_{n=0}^{N-1} G(n) \mathrm{e}^{j 2\pi k n / N}, \quad k = 0, \ldots, N-1.$$

Subspace identification algorithm

3. Form a block-Hankel matrix:

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

4. Calculate a singular value decomposition:

$$\hat{H} = \hat{U}\hat{\Sigma}\hat{V}^T.$$

Subspace identification algorithm

5. Select a model order, $\hat{n}_x$ and partition the SVD:

$$\hat{U}\hat{\Sigma}\hat{V}^T = \begin{bmatrix} \hat{U}_1 & \hat{U}_2 \end{bmatrix} \begin{bmatrix} \hat{\Sigma}_1 & 0 \\ 0 & \hat{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \hat{V}_1^T \\ \hat{V}_2^T \end{bmatrix}, \quad \hat{\Sigma}_1 \in \mathcal{R}^{\hat{n}_x \times \hat{n}_x}.$$

6. Estimate $\hat{A}$ via:

Define $J_1 = \begin{bmatrix} I_{n_y(q-1)} & 0_{n_y(q-1) \times n_y} \end{bmatrix}$

$J_2 = \begin{bmatrix} 0_{n_y(q-1) \times n_y} & I_{n_y(q-1)} \end{bmatrix}$

Solve for $\hat{A}$ via LS: $J_1 \hat{U}_1 \hat{A} = J_2 \hat{U}_1.$

Subspace identification algorithm

7. Estimate $\hat{C}$ via:

Define: $J_3 = \begin{bmatrix} I_{n_y} & 0_{n_y \times n_y(q-1)} \end{bmatrix}$,

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

8. Find $\hat{B}$ and $\hat{D}$ via least-squares:

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

9. Form the estimate:

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

Properties

Properties:

Asymptotic convergence:

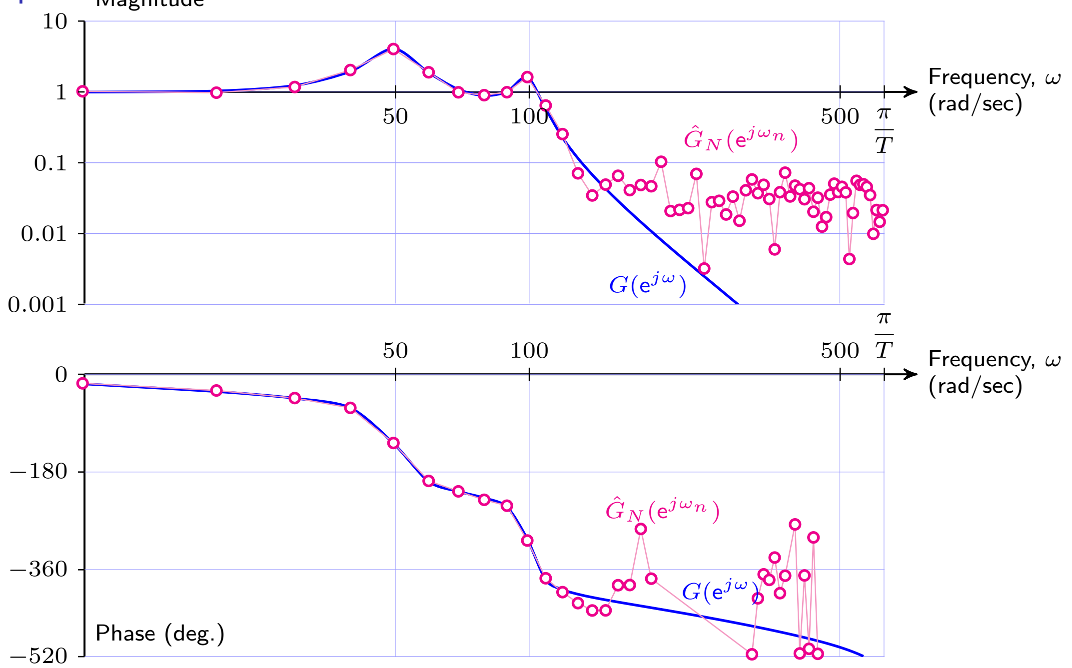
$$\lim_{N \longrightarrow \infty} \left\| \hat{G}(\mathrm{e}^{j\omega_n}) - G(\mathrm{e}^{j\omega_n}) \right\|_\infty = 0, \quad n = 0, \ldots, N-1, \quad \text{w.p. } 1$$

Algorithm is "correct":

If $V(\mathrm{e}^{j\omega_n}) = 0$ then there exists a data length $N_0 < \infty$, such that

$$\left\| \hat{G}_N(\mathrm{e}^{j\omega}) - G(\mathrm{e}^{j\omega}) \right\|_\infty = 0 \quad \text{for all } N > N_0.$$

Example

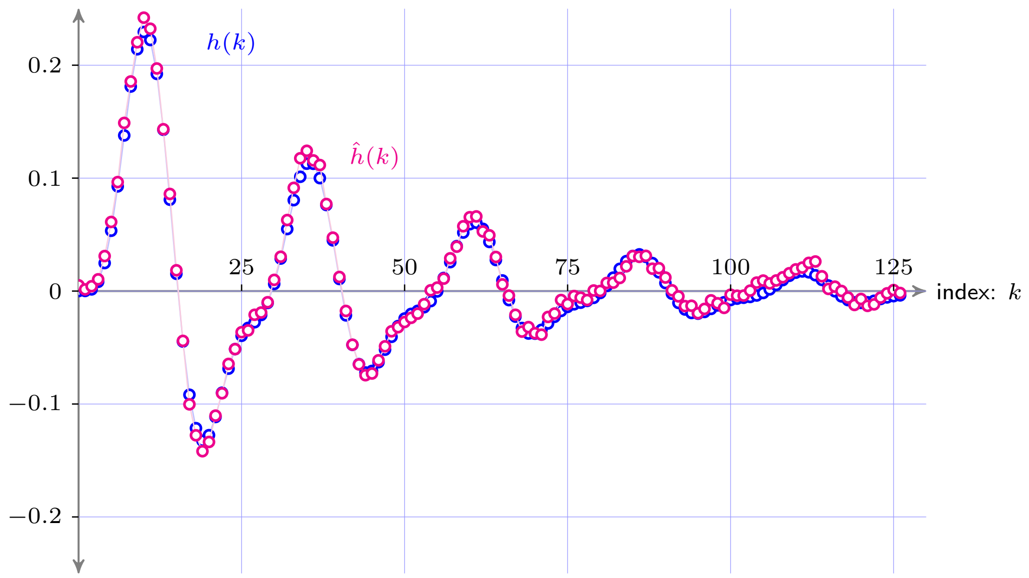


Sampling period, $T = 1/200$, Data & DFT length, $N = 127$

$\Longrightarrow$ Nyquist frequency, $\pi/T = 628.3$ rad/sec

Example

Aliased impulse response comparison



Example

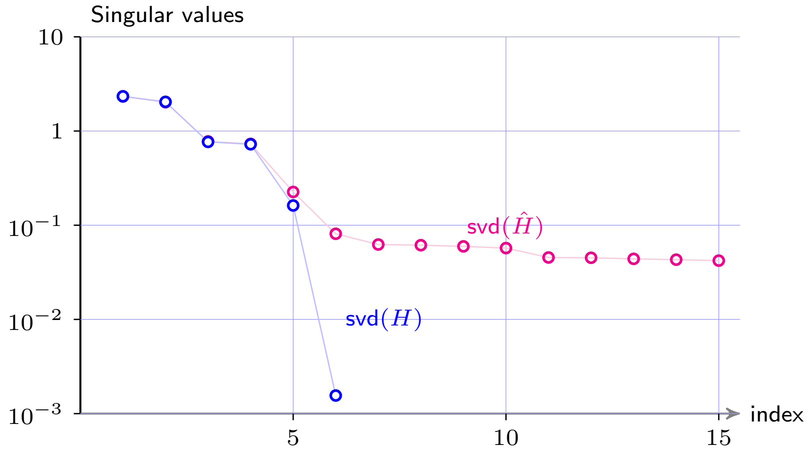
Hankel matrix

$$\hat{H} = \begin{bmatrix} \hat{h}_1 & \hat{h}_2 & \hat{h}_3 & \cdots & \hat{h}_r \\ \hat{h}_2 & \hat{h}_3 & \iddots & \iddots & \hat{h}_{r+1} \\ \hat{h}_3 & \iddots & \iddots & \iddots & \\ & \iddots & \iddots & & \vdots \\ \hat{h}_q & \hat{h}_{q+1} & & \cdots & \hat{h}_{q+r-1} \end{bmatrix} \in \mathcal{R}^{63\times 63}$$

$$= \begin{bmatrix} 0.0015 & 0.0043 & 0.0104 & 0.0310 & \cdots \\ 0.0043 & 0.0104 & 0.0310 & 0.0612 & \\ 0.0104 & 0.0310 & 0.0612 & 0.0965 & \\ 0.0310 & 0.0612 & 0.0965 & 0.1489 & \\ \vdots & & & & \ddots \end{bmatrix}.$$

Example

Singular value decomposition of H and $\hat{H}$



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

Choose order, $n_x = 5$.

Example

Estimate $\hat{A}$ and $\hat{C}$

$\hat{A}$ is the least squares solution to $J_1 \hat{U}_1 \hat{A} = J_2 \hat{U}_1$.

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

Estimate $\hat{B}$ and $\hat{D}$

Least squares solution to:

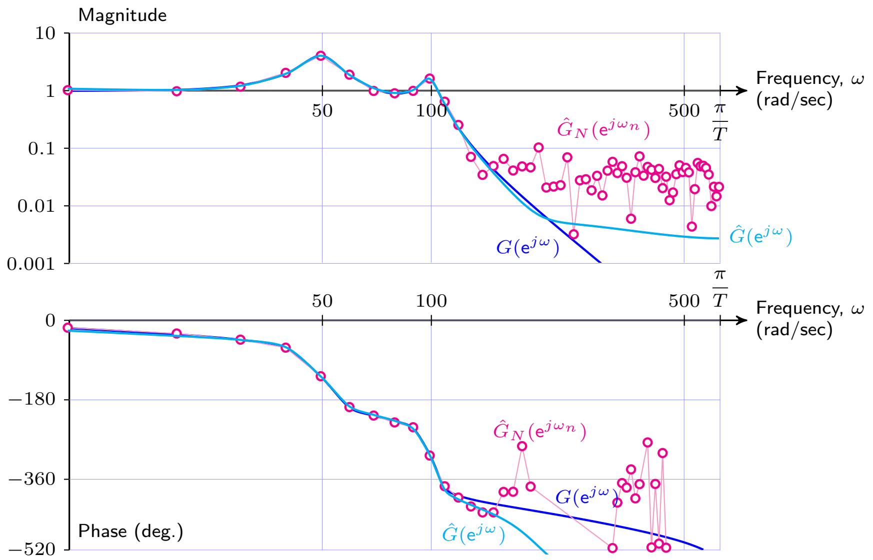
$$\begin{bmatrix} \hat{C}(\mathrm{e}^{j\omega_n} I - \hat{A})^{-1} & I \end{bmatrix} \begin{bmatrix} B \\ D \end{bmatrix} = G(n).$$

or, to get real-valued B and D,

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

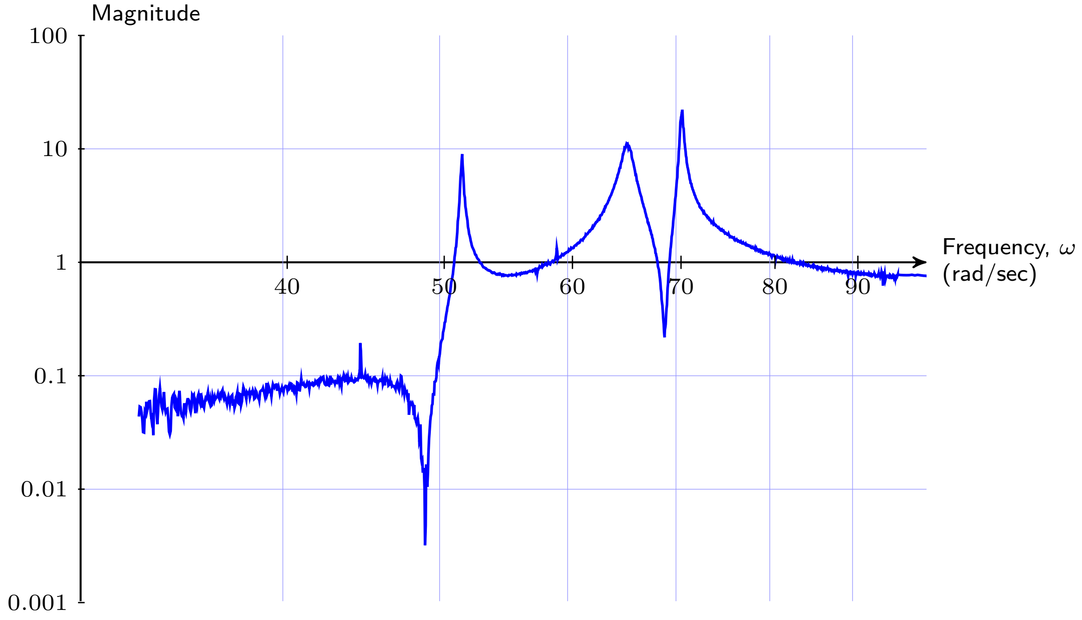
Example

Estimated result ($n_x = 5$)



Nonuniformly spaced frequency case

Swept-sine experiments



The frequency grid is linear and over a finite range.

Nonuniformly spaced frequency case

Time-domain:

$$x(k+1) = Ax(k) + Bu(k)$$
$$y(k) = Cx(k) + Du(k)$$

Frequency-domain:

$$\mathrm{e}^{j\omega} X(\omega) = AX(\omega) + BU(\omega)$$
$$Y(\omega) = CX(\omega) + BU(\omega)$$

For a specific frequency, ω, on each channel;

$$U_i(\omega) = e_i, \quad i = 1, \ldots, n_u.$$

Resulting system equations:

$$\mathrm{e}^{j\omega} X_i(\omega) = AX_i(\omega) + BU_i(\omega)$$
$$Y_i(\omega) = CX_i(\omega) + DU_i(\omega)$$

Nonuniformly spaced frequency case

Defining,

$$X_c(\omega) = \begin{bmatrix} X_1(\omega) & \cdots & X_{n_u}(\omega) \end{bmatrix},$$

and stacking the equations column-wise gives,

$$\begin{aligned} \mathrm{e}^{j\omega} X_c(\omega) &= A X_c(\omega) + B \\ G(\omega) &= C X_c(\omega) + D \end{aligned}$$

Multiplying by $\mathrm{e}^{j\omega}$ and substituting,

$$\mathrm{e}^{j\omega} G(\omega) = CAX_c(\omega) + CB + D\mathrm{e}^{j\omega}$$

Nonuniformly spaced frequency case

Repeating and stacking row-wise,

$$\begin{bmatrix} G(\mathrm{e}^{j\omega}) \\ \mathrm{e}^{j\omega} G(\mathrm{e}^{j\omega}) \\ \vdots \\ \mathrm{e}^{j(q-1)\omega} G(\mathrm{e}^{j\omega}) \end{bmatrix} = \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix}}_{\mathcal{O}} X_c(\omega) + \Gamma \begin{bmatrix} I_{n_u} \\ \mathrm{e}^{j\omega} I_{n_u} \\ \vdots \\ \mathrm{e}^{j(q-1)\omega} I_{n_u} \end{bmatrix},$$

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

Nonuniformly spaced frequency case

Repeat for multiple frequencies and stack column-wise,

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

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

$$\mathcal{X}_c = \frac{1}{\sqrt{N}} \begin{bmatrix} X_c(\omega_1) & \cdots & X_c(\omega_N) \end{bmatrix}$$

$$\mathcal{G} = \mathcal{O}\mathcal{X}_c + \Gamma\mathcal{W}$$

Nonuniformly spaced frequency case

$$\mathcal{G} = \mathcal{O}\mathcal{X}_c + \Gamma\mathcal{W}$$

As $\mathcal{O}$ and Γ are real-valued,

$$\underbrace{\begin{bmatrix} \mathrm{real}(\mathcal{G}) & \mathrm{imag}(\mathcal{G}) \end{bmatrix}}_{\mathcal{G}_r} = \mathcal{O} \underbrace{\begin{bmatrix} \mathrm{real}(\mathcal{X}_c) & \mathrm{imag}(\mathcal{X}_c) \end{bmatrix}}_{\mathcal{X}_{cr}}$$
$$+ \Gamma \underbrace{\begin{bmatrix} \mathrm{real}(\mathcal{W}) & \mathrm{imag}(\mathcal{W}) \end{bmatrix}}_{\mathcal{W}_r}$$

If $n_y q < n_u r$ then $\exists \mathcal{W}_r^\perp$ such that $\mathcal{W}_r \mathcal{W}_r^\perp = 0$.

$$\mathcal{G}_r \mathcal{W}_r^\perp = (\mathcal{O}\mathcal{X}_{cr} + \Gamma\mathcal{W}_r)\,\mathcal{W}_r^\perp = \mathcal{O}\mathcal{X}_{cr}\mathcal{W}_r^\perp$$

$$\mathrm{range}\left(\mathcal{G}_r \mathcal{W}_r^\perp\right) = \mathrm{range}(\mathcal{O})$$

Subspace identification

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.

Potential pitfalls

- 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

Subspace ID:

Tomas McKelvey, Hüseyin Akçay & Lennart Ljung, "Subspace-based multivariable system identification from frequency response data," *IEEE Trans. Auto. Ctrl,* Vol. 41, No. 7, pp. 960–979, 1996.

Mats Viberg, Björn Ottersten, Bo Wahlberg & Lennart Ljung, "A statistical perspective on state-space modeling using subspace methods," *Proc. Control & Decision Conf.,* pp. 1337–1342, 1991.

Peter Van Overschee & Bart De Moor, "N4SID: subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica,* Vol. 30, No. 1, pp. 75–93, 1994.

Lennart Ljung, *System Identification;Theory for the User,* 2nd Ed., Prentice-Hall, 1999, [section 10.6].