Subspace identification algorithms:

an introduction

Hüseyin Akçay

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

October 28, 2008

Outline

Subspace identification algorithms are often based on geometric concepts. Some system characteristics can be revealed by geometric manipulation of the row spaces of certain matrices.

Let us assume that the three matrices

$$A \in \mathbf{R}^{p\times j}, \qquad B \in \mathbf{R}^{q\times j}, \qquad C \in \mathbf{R}^{r\times j}$$

are given.

Orthogonal projections

The operator that projects the row space of a matrix onto the row space of $B \in \mathbf{R}^{q\times j}$ is defined by

$$\Pi_B \triangleq B^T(BB^T)^\dagger B$$

where † denotes the the Moore-Penrose pseudo inverse.

Thus,

$$A/B \triangleq A\Pi_B = AB^T(BB^T)^\dagger B$$

is shorthand for the projection of the row space of $A \in \mathbf{R}^{p\times j}$ on the row space of B.

$\Pi_{B^\perp}$ is the geometric operator that projects the row space of a matrix onto the orthogonal complement of the row space of B:

$$A/B^\perp \triangleq A\Pi_{B^\perp},$$

where

$$\Pi_{B^\perp} = I_j - \Pi_B.$$

$$A = A\Pi_B + A\Pi_{B^\perp}.$$

Alternatively, the projections decompose A as linear combination of the rows of B and of the rows of the orthogonal complement of B. With

$$\begin{aligned} L_B B &\triangleq A/B, \\ L_{B^\perp} B^\perp &\triangleq A/B^\perp \end{aligned}$$

where $B^\perp$ is a basis for the orthogonal complement of the row space of B, we find

$$A = L_B B + L_{B^\perp} B^\perp.$$

- A decomposition of A into a sum of linear combinations of the rows of B and $B^\perp$.

Instead of decomposing A as linear combinatons of two orthogonal matrices B and $B^{\perp}$, it can also be decomposed as linear combination of two non-orthogonal matrices B and C and of the orthogonal complement f B and C. Thus,

$$A = L_B B + \underbrace{L_C C}_{A/_B C} + L_{B^{\perp},C^{\perp}} \begin{pmatrix} B \\ C \end{pmatrix}^{\perp}.$$

Definition *Oblique projections*

The oblique projection of the row space $A \in \mathbf{R}^{p \times j}$ along the row spaces of $C \in \mathbf{R}^{r \times j}$ is defined as:

$$A/_B C \triangleq A \begin{pmatrix} C^T & B^T \end{pmatrix} \left[\begin{pmatrix} CC^T & CB^T \\ BC^T & BB^T \end{pmatrix}^{\dagger} \right]_{\text{first r columns}} C.$$

Some properties of the oblique projections are:

- $B/_B C = 0$,
- $C/_B C = C$,
- $A/_B C = [A/B^{\perp}][C/B^{\perp}]^{\dagger} C$.

Orthogonal projections can be easily expressed in functions of the RQ decomposition. Let us first treat the case A/B where A and B can be expressed as linear combinations of the orthonormal matrix Q^T as:

$$\begin{aligned} A &= R_A Q^T, \\ B &= R_B Q^T. \end{aligned}$$

Here, A and B are lower triangular matrices.

Then,

$$\begin{aligned} A/B &= AB^T(BB^T)^{\dagger}B \\ &= [R_A Q^T Q R_B^T] R_B Q^T Q R_B^T]^{\dagger} R_B Q^T \\ &= R_A R_B^T [R_B R_B^T]^{\dagger} R_B Q^T. \end{aligned}$$

- The oblique projections can also be written as functions of the RQ decompositions by noting that

$$\begin{aligned} A/B^{\perp} = A - A/B &= R_A Q^T - R_A R_B^T [R_B R_B^T]^{\dagger} R_B Q^T \\ &= R_A [I - R_B^T [R_B R_B^T]^{\dagger} R_B] Q^T. \end{aligned}$$

Outline

Given:

s measurements of the input $u_k \in \mathbf{R}^m$ and the output $y_k \in \mathbf{R}^l$ generated by the unknown system of order n:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k, \\ y_k &= Cx_k + Du_k. \end{aligned}$$

Determine:

- The order of the unknown system
- The system matrices $A \in \mathbf{R}^{n \times n}$, $B \in \mathbf{R}^{n \times m}$, $C \in \mathbf{R}^{l \times n}$ (up to within a similarity transformation).

Outline

$$
U_{0|2i-1} \triangleq \left(\begin{array}{cccc}
u_0 & u_1 & \cdots & u_{j-1} \\
u_1 & u_2 & \cdots & u_j \\
\vdots & \vdots & \ddots & \vdots \\
u_{i-1} & u_i & \cdots & u_{i+j-2} \\
\hline
u_i & u_{i+1} & \cdots & u_{i+j-1} \\
u_{i+1} & u_{i+2} & \cdots & u_{i+j} \\
\vdots & \vdots & \ddots & \vdots \\
u_{2i-1} & u_{2i} & \cdots & u_{2i+j-2}
\end{array}\right)
\triangleq \left(\frac{U_{0|i-1}}{U_{i|2i-1}}\right) \triangleq \left(\frac{U_p}{U_f}\right)
$$

$$
\triangleq \left(\begin{array}{cccc}
u_0 & u_1 & \cdots & u_{j-1} \\
u_1 & u_2 & \cdots & u_j \\
\vdots & \vdots & \ddots & \vdots \\
u_{i-1} & u_i & \cdots & u_{i+j-2} \\
u_i & u_{i+1} & \cdots & u_{i+j-1} \\
\hline
u_{i+1} & u_{i+2} & \cdots & u_{i+j} \\
\vdots & \vdots & \ddots & \vdots \\
u_{2i-1} & u_{2i} & \cdots & u_{2i+j-2}
\end{array}\right)
$$

$$
\triangleq \left(\frac{U_{0|i}}{U_{i+1|2i-1}}\right) \triangleq \left(\frac{U_p^+}{U_f^-}\right).
$$

- The output block Hankel matrices $Y_{0|2i-1}$, Y_p, Y_f, Y_p^+, Y_f^- are defined in a similar way.

$$W_{0|i} \triangleq \begin{pmatrix} U_{0|i-1} \\ Y_{0|i-1} \end{pmatrix} = \begin{pmatrix} U_p \\ Y_p \end{pmatrix} = W_p.$$

Similarly as before, W_p^+ is defined as

$$W_p^+ \triangleq \begin{pmatrix} U_p^+ \\ Y_p^+ \end{pmatrix}.$$

The state sequence X_i is defined as

$$X_i \triangleq \begin{pmatrix} x_i & \cdots & x_{i+j-1} \end{pmatrix} \in \mathbf{R}^{n \times j}.$$

Analogous to the past inputs and outputs,

$$X_p = X_0, \qquad X_f = X_i.$$

System related matrices

The extended ($i > n$) observability matrix

$$\Gamma_i \triangleq \begin{pmatrix} C \\ \vdots \\ CA^{i-1} \end{pmatrix} \in \mathbf{R}^{li \times n}.$$

We assume that the pair $\{A, C\}$ to be observable, which implies that the rank of Γ_i is equal to n.

The reversed extended controllability matrix

$$\Delta_i \triangleq \begin{pmatrix} A^{i-1}B & \cdots & B \end{pmatrix} \in \mathbf{R}^{n \times mi}.$$

We assume that the pair $\{A, B\}$ to be controllable, which implies that the rank of Δ_i is equal to n.

The lower block triangular Toeplitz matrix

$$H_i \triangleq \begin{pmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{i-2}B & CA^{i-3}B & \cdots & D \end{pmatrix} \in \mathbf{R}^{li \times mi}.$$

Definition The input sequence $u_k \in \mathbf{R}^m$ is *persistently exciting of order* $2i$ if the input covariance matrix

$$R^{uu} \triangleq \lim_{j \to \infty} \frac{1}{j} U_{0|2i-1} U_{0|2i-1}^T$$

has full rank, which is $2mi$.

Outline

Theorem Matrix input-output equations.

$$
\begin{aligned}
Y_p &= \Gamma_i X_p + H_i U_p, \\
Y_f &= \Gamma_i X_f + H_i U_f, \\
X_f &= A^i X_p + \Delta_i U_p.
\end{aligned}
$$

Theorem *Deterministic time-domain identification*.

Under the assumptions that:

- 1. The input u_k is persistently exciting of order $2i$.
- 2. The intersection of the row space of U_f (the future inputs) and the row space of X_p (the past states) is empty.

- 3. The user defined weighting matrices $W_1 \in \mathbf{R}^{li \times li}$ and $W_2 \in \mathbf{R}^{j \times j}$ are such that W_1 is of full rank and W_2 obeys: $\mathrm{rank}(W_p) = \mathrm{rank}(W_p W_2)$, where W_p is the block Hankel matrix containing the past inputs and outputs.

And with $\mathcal{O}_i$ defined as the oblique projection:

$$\mathcal{O}_i \triangleq Y_f /_{U_f} W_p,$$

and the singular value decomposition:

$$\begin{aligned} W_1 \mathcal{O}_i W_2 &= \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} \\ &= U_1 S_1 V_1^T, \end{aligned}$$

we have:

1. $\mathcal{O}_i$ is equal to the product of the extended observability matrix and the states:

$$\mathcal{O}_i = \Gamma_i X_f.$$

2. n is equal to the number of non-zero singular values.
3. The extended observability matrix Γ_i is equal to:

$$\Gamma_i = W_1^{-1} U_1 S_1^{1/2} T.$$

4. The part of the state sequence X_f that lies in the column space of W_2 can be recovered from:

$$X_f W_2 = T^{-1} S_1^{1/2} V_1^T.$$

5. The state sequence X_f is equal to

$$X_f = \Gamma_i^{\dagger} \mathcal{O}_i.$$

Proof

From matrix input-output equations,

$$\begin{aligned} X_f &= A^i X_p + \Delta_i U_p \\ &= A^i[\Gamma_i^{\dagger} Y_p - \Gamma_i^{\dagger} H_i U_p] + \Delta_i U_p \\ &= [\Delta_i - A^i \Gamma_i^{\dagger} H_i] U_p + [A^i \Gamma_i^{\dagger}] Y_p \\ &= L_p W_p \end{aligned}$$

with

$$L_p = (\Delta_i - A^i \Gamma_i^{\dagger} H_i \;\; A^i \Gamma_i^{\dagger}).$$

Thus,

$$\begin{aligned} Y_f &= \Gamma_i L_p W_p + H_i U_f, \\ Y_f \Pi_{U_f^{\perp}} &= \Gamma_i L_p W_p \Pi_{U_f^{\perp}}, \end{aligned}$$

$$Y_f/U_f^{\perp} = \Gamma_i L_p W_p/U_f^{\perp},$$

$$[Y_f/U_f^{\perp}][W_p/U_f^{\perp}]^{\dagger} W_p = \Gamma_i L_p W_p,$$

$$\mathcal{O}_i = \Gamma_i X_f$$

where we have used the fact that $[W_p/U_f^{\perp}][W_p/U_f^{\perp}]^{\dagger} W_p = W_p$ to be shown shortly.

The second claim follows from the fact that $W_1\mathcal{O}_i W_2$ is equal to the product of two matrices $W_1\Gamma_i$ (n columns) and $X_f W_2$ (n rows). Since W_1 is of full rank due to assumption 3 of the theorem, the product $W_1\Gamma_i$ is also of rank n. Multiplying both sides of $X_f = L_p W_p$ with W_2, we get $X_f W_2 = L_p W_p W_2$. Then, from assumption 3, the rank of $X_f W_2$ is equal to the rank of X_f. Hence, $W_1\Gamma_i$, $X_f W_2$, and their products are all of rank n.

Thus, the second formula in the SVD can be split into two parts for some non-singular $T \in \mathbf{R}^{n \times n}$ as follows:

$$\begin{aligned} W_1 \Gamma_i &= U_1 S_1^{1/2} T, \\ X_f W_2 &= T^{-1} S_1^{1/2} V_1^T \end{aligned}$$

which leads to claim 3 and 4 of the theorem. Claim 5 easily follows from the first claim.

The proof of $[W_p/U_f^{\perp}][W_p/U_f^{\perp}]^{\dagger} W_p = W_p$*:*

Let us first show that

$$\mathrm{rank}\, W_p = rank W_p/U_f^{\perp}.$$

W_p can be written as

$$W_p = \begin{pmatrix} I_{mi} & 0 \\ H_i & \Gamma_i \end{pmatrix} \begin{pmatrix} U_p \\ X_p \end{pmatrix},$$

which implies that $W_p/U_f^{\perp}$ can be written as:

$$W_p/U_f^{\perp} = \begin{pmatrix} I_{mi} & 0 \\ H_i & \Gamma_i \end{pmatrix} \begin{pmatrix} U_p/U_f^{\perp} \\ X_p/U_f^{\perp} \end{pmatrix},$$

Due to the persistency of excitation $U_p/U_f^{\perp}$ does not lose rank and due to the second assumption $x_p/U_f^{\perp}$ either. Hence,

$$\mathrm{rank} \begin{pmatrix} U_p \\ X_p \end{pmatrix} = \mathrm{rank} \begin{pmatrix} U_p/U_f^{\perp} \\ X_p/U_f^{\perp} \end{pmatrix}$$

proving the first claim. Now, denote the SVD of $W_p/U_f^{\perp}$ as:

$$W_p/U_f^{\perp} = (U_1 \quad U_2) \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = U_1 S_1 V_1^T.$$

Since $W_p/U_f^{\perp}$ is a linear combination of the columns of W_p and since the rank of W_p and $W_p/U_f^{\perp}$ are equal, we find that the column spaces of W_p and $W_p/U_f^{\perp}$ are equal. This implies that W_p can be written as:

$$W_p = U_1 R.$$

Finally,

$$\begin{aligned}[W_p/U_f^{\perp}][W_p/U_f^{\perp}]^{\dagger} W_p &= [U_1 S_1 V_1^T][V_1 S_1^{-1} U_1^T] U_1 R \\ &= U_1 R = W_p.\end{aligned}$$

Remarks

- $\mathrm{rank}(Y_f/_{U_f}W_p) = n$
- $\text{row space}(Y_f/_{U_f}W_p) = \text{row space}(X_f)$
- $\text{column space}(Y_f/_{U_f}W_p) = \text{column space}(\Gamma_i)$

Summarize why these algorithms are called *subspace algorithms*!

Computing the system matrices

Similarly, we can show that

$$\mathcal{O}_{i-1} \triangleq Y_f^- / _{U_f^-} W_p^+ = \Gamma_{i-1} X_{i+1}.$$

Let $\underline{\Gamma_i}$ denote the matrix Γ_i without the last l rows. Then,

$$\Gamma_{i-1} = \underline{\Gamma_i}$$

and X_{i+1} can be calculated as

$$X_{i+1} = \Gamma_{i-1}^{\dagger} \mathcal{O}_{i-1}.$$

We have calculated X_i and X_{i+1} using only input-output data. The matrices A, B, C, D can be solved in an LS sense from:

$$\begin{pmatrix} X_{i+1} \\ Y_{i|i} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} X_i \\ U_{i|i} \end{pmatrix}.$$

Deterministic algorithm

- 1. Calculate the oblique projections:

$$\mathcal{O}_i = Y_f /_{U_f} W_p, \qquad \mathcal{O}_{i-1} = Y_f^- /_{U_f^-} W_p^+.$$

- 2. Calculate the SVD of the weighted oblique projection:

$$W_1 \mathcal{O}_i W_2 = USV^T.$$

- 3. Determine the order by inspecting the singular values in S and partition the SVD accordingly to obtain U_1 and S_1.
- 4. Determine Γ_i and Γ_{i-1} as:

$$\Gamma_i = W_1^{-1} U_1 S_1^{1/2}, \qquad \Gamma_{i-1} = \underline{\Gamma_i}.$$

- 5. Determine X_i and X_{i+1} as:

$$X_i = \Gamma_i^{\dagger} \mathcal{O}_i, \qquad X_{i+1} = \Gamma_{i-1}^{\dagger} \mathcal{O}_{i-1}.$$

- 6. Solve the set of linear equations for A, B, C and D:

$$\begin{pmatrix} X_{i+1} \\ Y_{i|i} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} X_i \\ U_{i|i} \end{pmatrix}.$$

Outline

Consider the continuous-time system with m inputs, l outputs and n states:

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

With the assumption $x(0) = 0$, the system equations can be transformed to the Laplace domain:

$$\begin{aligned} sX(s) &= AX(s) + BU(s), \\ Y(s) &= CX(s) + DU(s). \end{aligned}$$

The frequency domain response is given by

$$H(s) = D + C(sI - A)^{-1}B.$$

With an input $U(s) = I_m$, Laplace domain equations are rewritten as

$$
\begin{aligned}
sX^H(s) &= AX^H(s) + BI_m, \\
Y(s) &= CX^H(s) + DI_m.
\end{aligned}
$$

- $X^H(s)$ is $n \times m$ matrix where the kth column of $x^H(s)$ contains the transformed state trajectory induced by an impulse applied to the kth input.

Problem Given N frequency response samples $H(j\omega_k)$, $k = 1, \cdots, N$, find the system matrices A, B, C, D.

Outline

The extended observability matrix Γ_i and the block Toeplitz matrix Θ_i are given by

$$\Gamma_i \triangleq \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{i-1} \end{bmatrix} \in \mathbf{R}^{li\times n},$$

$$\Theta_i \triangleq \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & vdots \\ CA^{i-2}B & CA^{i-3}B & \cdots & D \end{bmatrix} \in \mathbf{R}^{li\times mi}$$

with $i > n$, a user defined index.

Let

$$\mathcal{H}^c \triangleq \begin{bmatrix} H(j\omega_1) & H(j\omega_2) & \cdots & H(j\omega_N) \\ (j\omega_1)H(j\omega_1) & (j\omega_2)H(j\omega_2) & \cdots & (j\omega_N)H(j\omega_N) \\ \vdots & \vdots & \ddots & \vdots \\ (j\omega_1)^{i-1}H(j\omega_1) & (j\omega_2)^{i-1}H(j\omega_2) & \cdots & (j\omega_N)^{i-1}H(j\omega_N) \end{bmatrix},$$

$$\mathcal{I}^c \triangleq \begin{bmatrix} I_m & I_m & \cdots & I_m \\ (j\omega_1)I_m & (j\omega_2)I_m & \cdots & (j\omega_N)I_m \\ \vdots & \vdots & \ddots & \vdots \\ (j\omega_1)^{i-1}I_m & (j\omega_2)^{i-1}I_m & \cdots & (j\omega_N)^{i-1}I_m \end{bmatrix},$$

$$\mathcal{X}^c \triangleq [X^H(j\omega_1) \;\; X^H(j\omega_2) \;\; \cdots X^H(j\omega_N)],$$

with $\mathcal{H} \in \mathbf{C}^{li\times mN}, \mathcal{I} \in \mathbf{C}^{mi\times mN}$ and $\mathcal{X} \in \mathbf{C}^{n\times mN}$.

$$\begin{aligned}
\mathcal{H} &\triangleq [\mathrm{Re}(\mathcal{H}^c) \;\; \mathrm{Im}(\mathcal{H}^c)] \in \mathbf{R}^{li\times 2mN}, \\
\mathcal{I} &\triangleq [\mathrm{Re}(\mathcal{I}^c) \;\; \mathrm{Im}(\mathcal{I}^c)] \in \mathbf{R}^{mi\times 2mN}, \\
\mathcal{X} &\triangleq [\mathrm{Re}(\mathcal{X}^c) \;\; \mathrm{Im}(\mathcal{X}^c)] \in \mathbf{R}^{n\times 2mN}.
\end{aligned}$$

Lemma (Input-output equation)

$$\begin{aligned}
\mathcal{H}^c &= \Gamma_i \mathcal{X}^c + \Theta_i \mathcal{I}^c, \\
\mathcal{H} &= \Gamma_i \mathcal{X} + \Theta_i \mathcal{I}.
\end{aligned}$$

- This lemma is obtained by recursive use and evaluation of the Laplace domain equations.

By projecting the second equation in the lemma onto the orthogonal complement of $\mathcal{I}$, we obtain:

Theorem (Orthogonal projection) If $\operatorname{rank}[\mathcal{X}/\mathcal{I}^{\perp}] = n$, then,

$$
\begin{aligned}
\operatorname{rank}[\mathcal{H}/\mathcal{I}^{\perp}] &= n, \\
\operatorname{range}[\mathcal{H}/\mathcal{I}^{\perp}] &= \operatorname{range}[\Gamma_i].
\end{aligned}
$$

Simple frequency domain algorithm

- Construct $\mathcal{I}$ and $\mathcal{H}$ from the given frequencies ω_k and the frequency response points $H(j\omega_k)$.
- Compute $\mathcal{H}/\mathcal{I}^{\perp}$.

- Compute the SVD:

$$\mathcal{H}/\mathcal{I}^{\perp} = (U_1 \;\; U_2) \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix}.$$

- Determine the order from the number of singular values S_1 different from zero.
- Determine $= U_1 S_1^{1/2}$, which is one possible estimate for the extended observability matrix Γ_i.
- Determine A and C as:

$$C = \mathcal{G}_{\text{first } l \text{ rows}}, \qquad A = [\underline{\mathcal{G}}]^{\dagger} \bar{\mathcal{G}}$$

where $\underline{\mathcal{G}}$ and $\bar{\mathcal{G}}$ denote $\mathcal{G}$ without the last and first l rows.

- Determine B and D through the (least squares) solution of the linear set:

$$\begin{pmatrix} \mathcal{L}_R \\ \mathcal{L}_I \end{pmatrix} = \begin{pmatrix} \mathcal{L}_R \\ \mathcal{L}_I \end{pmatrix} \begin{pmatrix} B \\ D \end{pmatrix},$$

where $\mathcal{L} \in \mathbf{C}^{lN \times m}$ and $\mathcal{M} \in \mathbf{C}^{lN \times (n+l)}$ are defined as:

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_R + j\mathcal{L}_I = \begin{pmatrix} H(j\omega_1) \\ \vdots \\ H(j\omega_N) \end{pmatrix}, \\ \mathcal{M} &= \mathcal{M}_R + j\mathcal{M}_I = \begin{pmatrix} C(j\omega_1 I_n - A)^{-1} & I_l \\ \vdots & \vdots \\ C(j\omega_N I_n - A)^{-1} & I_l \end{pmatrix}. \end{aligned}$$

- This algorithm is academic since it is limited to small values of n and the frequency range:
  - Due to block-Vandermonde structure the condition numbers of $\mathcal{H}$ and $\mathcal{I}$ become extremely large when n gets larger.
  - The larger the frequency range, the poorer numerical conditionings of $\mathcal{H}$ and $\mathcal{I}$.
- It is possible to improve the numerical conditioning by implicitly constructing a well-conditioned basis for the row spaces of $\mathcal{H}$ and $\mathcal{I}$ through Forsythe recursions (Van Overschee and De Moor:1996).