Subspace-based Identification

of Infinite-dimensional Multivariable Systems from Frequency-response Data

Hüseyin Akçay

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

October 12, 2008

Outline

1. Introduction
2. Problem Formulation
3. Subspace-based algorithm
4. Convergence analysis
   - Noise-free data case
   - Consistency analysis
   - Two related consistent algorithms
5. identification Examples
6. Conclusions

A subspace-based identification algorithm, which takes samples of an infinite-dimensional transfer function, is shown to produce estimates which converge to a balanced truncation of the system.

- Identification of infinite-dimensional systems studied in time-domain in Ljung and Yuan: 1985; Huang and Guo: 1990, Hjalmarsson: 1993.
- In frequency-domain studied in Helmicki *etal.*: 1991; Mäkilä and Partington: 1991; Gu and Khargonekar: 1992.
- Despite that low-order nominal models are preferred in most practical applications as in the design of model-based controllers, the true systems are usually of high or infinite order with unmodeled dynamics and random/deterministic noise.

- Thus, the basic task of system identification is to construct a simple nominal model based on the measured data generated from a complex system.

Based on how the disturbances are characterized, problem formulations can be divided into two categories:

1. Stochastic formulation leading to instrumental variable and prediction error methods (Ljung: 1997, Södestrom and Stoica: 1989). "The least-squares method".
2. Deterministic formulation leading to "robustly convergent" non-linear algorithms (Helmicki *etal.*: 1991, Gu and Khargonekar: 1992).

- In both approaches, a prejudice-free model set of high complexity is the underlying model structure.
- In most practical applications, the model is required to be of restricted complexity despite the fact that the true system might have infinite order. Thus, model reduction is a complementary step to the black-box identification.
  - Besides the computational complexity, this step induces large approximation errors.

- An alternative method is to directly realize low complexity models from the experimental data.
  - Nonlinear parametric optimization (Ljung: 1993, Pintelon *et al.*: 1994b) where the solution is obtained by iterations.
  - Non-iterative subspace-based algorithms delivering state-space models without any parametric optimization (Verhaegen and Dewilde: 1992, Van Overschee and De Moor: 1994).
- Models in canonical minimal parametrizations are numerically sensitive for high-order models, in comparison with state-space models in a balanced realization.
- Subspace-based algorithms are more robust to numerical inaccuracies than the canonically parametrized models since the model obtained is normally close to being balanced.

- Frequency-domain subspace algorithms (Juang and Suziki: 1988, Liu *etal.*: 1994, McKelvey *etal.,*: 1996 based on the realization algorithms by Ho and Kalman (1966) and Kung (1978).
  - Ho and Kalman: 1966 and Kung: 1978 find a minimal state-space realization given a finite sequence of the Markov parameters estimated from the inverse discrete Fourier transform (DFT) of the frequency-response data.
  - Juang and Suziki: 1988 is exact only if the system has a finite impulse response, therefore for lightly damped systems yields very poor estimates.
  - In McKelvey *etal.*: 1994, the inverse DFT technique is combined with a subspace identification step yielding the true finite-dimensional system in spite of this aliasing effect.
  - Current work reporting extensions of McKelvey *etal.*: 1996 to infinite-dimensional systems.

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

$$y(t) = \sum_{k=0}^{\infty} g_k u(t-k) \tag{1}$$

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

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

$$G(e^{-j\omega}) = G^*(e^{j\omega}), \qquad 0 \leq \omega \leq \pi.$$

- For engineering purposes, a more practical model is a state-space model:

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

  where $x(t) \in \mathbf{R}^n$.

- The state-space model is a special case of (1) with

$$g_k = \begin{cases} D, & k = 0, \\ CA^{k-1}B, & k > 0. \end{cases}$$

- Identify a finite-dimensional model which is a good approximation of the infinite-dimensional system (1).

System assumptions

Some further assumptions must be imposed on the system to obtain good approximations.

The Hankel operator of $G(z)$ with symbol Γ is defined on ℓ_2^m by

$$(\Gamma u)(t) = \sum_{i=0}^{\infty} g_{t+i+1} u(i), \qquad t \geq 0$$

is a mapping into ℓ_2^p. Let Γ^* be the adjoint of Γ.

The Hankel singular values $\Gamma_i(G)$ are defined to be the square roots of the eigenvalues of $\Gamma\Gamma^*$.

Let u_i and v_i be the corresponding normalized eigenvectors of $\Gamma\Gamma^*$ and $\Gamma^*\Gamma$, respectively.

The pair (v_i, u_i) is called the Schmidt pair and satisfies

$$\begin{aligned} \Gamma v_i &= \Gamma_i(G) u_i, \\ \Gamma^* u_i &= \Gamma_i(G) v_i. \end{aligned}$$

A system G is said to be *Hilbert-Schmidt* if its Hankel singular values satisfy

$$\sum_{k=1}^{\infty} \Gamma_k^2(G) < \infty$$

and *nuclear* if

$$\sum_{k=1}^{\infty} \Gamma_k(G) < \infty.$$

- All finite-dimensional linear systems form a subset of nuclear systems and nuclear systems themselves are contained in the set of Hilbert-Schmidt systems.

These classes can be identified with impulse-response decay rates.

- G has Hilbert-Schmidt Hankel operator if

$$\|g_k\| = O(k^{-\alpha}), \qquad \alpha > 1$$

  or

$$\|g_k\| = O(1/(k \log k))$$

  which follows from the identity

$$\sum_{k=1}^{\infty} \Gamma_k^2(G) = \sum_{k=1}^{\infty} k\|g_k\|^2.$$

- A sufficient condition for the nuclearity (Bonnet: 1993)

$$\|g_k\| = O(k^{-\alpha}), \qquad \alpha > 3/2.$$

- Conversely, sufficient conditions for a system to have a Hilbert-Schmidt or nuclear Hankel operator can be stated in terms of boundary behavior of the system transfer function and its derivatives (Curtain:1985)

Assumption 1 The system $G \in \mathcal{H}_\infty$ has a continuous transfer function and a Hilbert-Schmidt Hankel operator Γ. For a fixed n, the Hankel singular values satisfy

$$\Gamma_n(G) > \Gamma_{n+1}(G).$$

- Let f be a complex-valued periodic function on the unit circle. Its *modulus of continuity* is defined by

$$\omega_f(t) = \sup_{|x-y|\le t} \|f(e^{ix}) - f(e^{iy})\|.$$

- f is of class Λ_α $(0 < \alpha \leq 1)$ if $\omega_f(t) = O(t^\alpha)$ as $t \to 0$.
- Optimal Hankel norm and balanced truncations are two popular model reduction techniques for nuclear systems:

$$\|G_n - G\|_\infty \leq 2 \sum_{k=n+1}^{\infty} \Gamma_k(G) \tag{2}$$

where repeated singular values are omitted in the sum and G_n is nth-order balanced truncation of G (Hinrichsen and Pritchard: 1990).

Noise assumptions

Data: $G_k = G(e^{j\pi k/M}) + e_k, \qquad k = 0, \cdots, M.$

Assumption 2 The noise e_k, $k = 0, \cdots, M$ are independent zero-mean complex random variables with uniformly-bounded second moments

$$R_k = \mathrm{E}\{e_k e_k^H\} \leq \bar{R}, \qquad \forall k.$$

1. **Objective:** to achieve (2) with probability one:

$$\lim_{M\to\infty} \|\hat{G}_{n,M} - G\|_\infty \leq 2 \sum_{k=n+1}^{\infty} \Gamma_k(G) \qquad \text{w.p.1}, \qquad (3)$$

where $\hat{G}_{n,M}$ is the nth-order identified model.

Algorithm 1

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

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

and perform the 2M-point inverse DFT

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

to obtain the estimates of g_i.
- 2. Construct the $q \times r$-block Hankel matrix

$$\hat{H}_{qr} = \begin{bmatrix} \hat{g}_1 & \hat{g}_2 & \cdots & \hat{g}_r \\ \hat{g}_2 & \hat{g}_3 & \cdots & \hat{g}_{r+1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{g}_q & \hat{g}_{q+1} & \cdots & \hat{g}_{q+r-1} \end{bmatrix}$$

and perform an SVD for $\hat{H}_q r$ as follows

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

where $\hat{\Sigma}_1$ contains the n dominant singular values.

- 3. Determine the system matrices as

$$\hat{A} = (J_1^q \hat{U}_1)^\dagger J_2^q \hat{U}_1 ,$$

$$
\begin{aligned}
\hat{C} &= J_3^q \hat{U}_1, \\
\hat{B} &= (I - \hat{A}^{2M}) \hat{\Sigma}_1 \hat{V}_1^T J_4^r, \\
\hat{D} &= \hat{g}_0 - \hat{C} \hat{A}^{2M-1} (I - \hat{A}^{2M})^{-1} \hat{B},
\end{aligned}
$$

where

$$
\begin{aligned}
J_1^q &= \begin{bmatrix} I_{(q-1)p} & 0_{(q-1)p \times p} \end{bmatrix}, \\
J_2^q &= \begin{bmatrix} 0_{(q-1)p \times p} & I_{(q-1)p} \end{bmatrix}, \\
J_3^q &= \begin{bmatrix} I_p & 0_{p \times (q-1)p} \end{bmatrix}, \\
J_4^r &= \begin{bmatrix} I_m \\ 0_{(r-1)m \times m} \end{bmatrix}.
\end{aligned}
$$

- 4. The estimated transfer function is

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

Theorem 1 Let G be a stable system of order n. Assume $q > n$, $r \geq n$ and $2M \geq q + r$. Suppose that $M + 1$ equidistant noise-free frequency-response measurements of G on $[O, \pi]$ are available and let $\hat{G}_{q,r,n,M}$ be given by Algorithm 1. Then

$$\|\hat{G}_{q,r,n,M} - G\|_\infty = 0.$$

- Let $q = n + 1$ and $r = n$ to meet the requirements on r and q which imply that $M = n + 1$, and consequently $n + 2$ equidistant samples of the frequency-response function on $[0, \pi]$ are required.
- Algorithm 1 is in the class of correct algorithms when applied to data from systems of finite dimension and uses a minimum amount of data among all such algorithms.
- Remarkable advantage with respect to black-box identification algorithms which use linearly parametrized model structures and satify (3).

Outline

Theorem 2 Let G be a linear system satisfying Assumption 1. Let ω_G be the modulus of continuity of G and assume that q and r satisfy the conditions

$$\text{(i)} \qquad \lim_{q,r,M\to\infty} \sqrt{qr}\frac{q+r}{M} = 0,$$

$$\text{(ii)} \qquad \lim_{q,r,M\to\infty} \sqrt{qr}\,\omega_G\left(\frac{\pi}{M}\right) = 0.$$

Let G_n be the balanced truncation of G be the balanced truncation of order n. Let $\hat{G}_{q,r,n,M}$ be given by Algorithm 1 using $M+1$ noise-free frequency-response measurements of G equidistantly spaced on $[O,\pi]$. Then

$$\lim_{q,r,M\to\infty} \|\hat{G}_{q,r,n,M} - G_n\|_\infty = 0.$$

- The Hilbert-Schmidt assumption on G merely implies that $\|g_k\| = o(1/\sqrt{k})$. The set of Hilbert-Schmidt systems is not contained in $\mathcal{H}_\infty$ (Duren: 1970, Exercise 6-7 in Chapter 6).

- $\ell_1^{p\times m}$ is not contained in the set of Hilbert-Schmidt systems either. Example:

$$g_k = \begin{cases} \frac{1}{\sqrt{k}}, & \text{for } k = 1, 2^4, 3^4, \cdots, \\ 0, & \text{otherwise.} \end{cases}$$

$\|g\|_1 = \sum_{k=1}^{\infty} k^{-2} < \infty$ while $\sum_{k=1}^{\infty} k|g_k|^2 = \infty$.

- Assumption 1 imposed on the system is rather weak!

Convergence condition $\sqrt{qr}\omega_G(\pi/M)$

Suppose $\|g_k\| = O(k^{-\alpha})$. If $\alpha > 1$, such systems are Hilbert-Schmidt and in ℓ_1. Moreover,

Lemma Assume that $\|g_k\| = O(k^{-\alpha})$ for some $\alpha > 1$. Then, $G(e^{j\theta}) \in \Lambda_{\min\{2,\alpha\}-1}$.

- Hence, for this class, we have a convergence requirement

$$qr = o(M^{2\alpha-2}), \qquad \text{for } 1 < \alpha \leq 2$$

  which drops out for $\alpha > 2$ since have $M > q, r$.

- Lemma 5 is sharp. Thus, as α gets closer to one, more and more data are required for the convergence to take place.

- Let $r = O(q)$. Then, condition (i) in Theorem 2 reads off $q = o(\sqrt{M})$ and for the class in the lemma, condition (ii) in Theorem 2 becomes $q = O(M^{\alpha-1})$.

$$q = \begin{cases} o(\sqrt{M}), & \alpha \geq 3/2, \\ o(M^{\alpha-1}), & 1 < \alpha < 3/2. \end{cases}$$

- For nuclear systems characterized by $\alpha > 3/2$, the only convergence requirement is $q = o(\sqrt{M})$ if $q = O(r)$.

Outline

Theorem 3 Let G be a system satisfying Assumption 1. Let ω_G be the modulus of continuity of G. Assume q, r satisfy condition (ii) in Theorem 2 and be at most $O(\sqrt{M}(\log M)^{-\beta})$ for some $\beta > 1/2$. Let G_n be the balanced truncation of G of order n. Let $\hat{G}_{q,r,n,M}$ be given by Algorithm 1 using equidistantly spaced $M+1$ frequency-response measurements of G on $[0, \pi]$. Let e_k satisfy Assumption 2. Then

$$\lim_{q,r,M\to\infty} \|\hat{G}_{q,r,n,M} - G_n\|_\infty = 0, \qquad \text{w.p.1.}$$

Furthermore if the Hankel operator of G is nuclear, then

$$\lim_{q,r,M\to\infty} \|\hat{G}_{q,r,n,M} - G_n\|_\infty \leq 2 \sum_{k=n+1}^{\infty} \Gamma_k(G), \qquad \text{w.p.1.}$$

where repeated singular values are omitted in the sum.

Corollary Let G be a system satisfying Assumption 1. Assume that $\|g_k\| = O(k^{-\alpha})$ for some $\alpha > 1$. Let q and r be at most $O(\sqrt{M}(\log M)^{-\beta})$ for some $\beta > 1/2$ and satisfy $qr = o(M^{2\min\{\alpha,2\}-2})$. Let $\hat{G}_{g,r,n,M}$ be given by Algorithm 1 using equidistantly spaced $M+1$ frequency-response measurements of G on $[0,\pi]$. Let e_k satisfy Assumption 2. Let G_n be the balanced truncation of G of order n. Then

$$\lim_{q,r,M\to\infty} \|\hat{G}_{q,r,n,M} - G_n\|_\infty = 0, \qquad \text{w.p.1.}$$

Furthermore, if $\alpha > 3/2$ then

$$\lim_{q,r,M\to\infty} \|\hat{G}_{q,r,n,M} - G_n\|_\infty \leq 2 \sum_{k=n+1}^{\infty} \Gamma_k(G), \qquad \text{w.p.1.}$$

where repeated singular values are omitted in the sum.

- Assume $q = O(r)$, then the consistency condition in Corollary 1 is

$$q = \begin{cases} o(M^{\alpha-1}), & \text{if } \alpha < 3/2, \\ O(\sqrt{M}(\log M)^{-\beta}); \beta > 1/2, & \text{if } \alpha \geq 3/2. \end{cases}$$

  - If $\alpha < 3/2$, rates for q and r depend on the smoothness of the system impulse response.
  - For the nuclear systems characterized by $\alpha > 3/2$, rates are determined by approximation errors caused by the measurement noise.

Outline

- Let $\hat{A}$, $\hat{C}$ be calculated as in Algorithm 1, $\hat{B}$ as

$$\hat{B} = \hat{\Sigma}_1 \hat{V}_1^T \begin{bmatrix} I_m \\ 0_{(r-1)m \times m} \end{bmatrix}$$

  and $\hat{D} = \hat{g}_0$. This algorithm, which we call Algorithm 2, studied in a modal analysis context by Juang and Suziki (1988) is a biased algorithm. Indeed, Example 1 illustrates poor performance of Algorithm 2 on real data of finite length when it is applied to lightly damped systems.

  - The bias term vanishes asymptotically and the algorithm yields truncated balanced realizations of the identified system under same assumptions in Theorem 3.

- In the third algorithm, which we call Algorithm 3, g_i are estimated as in Algorithm 1 and a pre-identified model is calculated by

$$\hat{G}_{\mathrm{pi}}(z) = \sum_{i=0}^{k} \hat{g}_i z^{-i}.$$

  The nth-order identified model is obtained from $\hat{G}_{\mathrm{pi}}$ by a recursively implemented balanced truncation technique. Thanks to the FIR structure!

  - Algorithm 1 contains Algorithm 3 as a special case. Thus, Algorithm 3 is also consistent under the assumptions of Theorem 3 though it is biased for finite data sets. (Algorithm 1 yields the nth-order balanced truncation of $\hat{G}_{\mathrm{pi}}$).

- The bias error of Algorithm 3 has two components:
  1. the first-stage error $\|G - \hat{G}_{\text{pi}}\|_\infty$
  2. approximation error: $\|\hat{G}_{\text{pi}} - \hat{G}\|_\infty$.

  The total error is bounded above by the sum of $\|G - \hat{G}_{\text{pi}}\|_\infty$ and $\|\hat{G}_{\text{pi}} - \hat{G}\|_\infty$. In the same example, Algorithm 3 performs poorly on the same data due to large approximation errors.
- In the choice of a potential identification algorithm, the posterior error caused by model reduction and correctness in addition to asymptotic properties must be taken into account.

- Algorithm 1 differs from McKelvey *etal.*: 1996, which we call Algorithm 4, only in the calculation of $\hat{B}$ and $\hat{D}$:

$$\hat{B}, \hat{D} = \arg \min_{\substack{B \in \mathbf{R}^{n \times m} \\ D \in \mathbf{R}^{p \times m}}} \sum_{i=0}^{2M-1} \| G_i - \tilde{G}(e^{j\pi i/M}) \|_F^2$$

where

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

- Algorithm 4 is also correct and uses minimal data when restricted to finite-dimensional systems, In Algorithm 1, $\hat{B}$ and $\hat{D}$ were modified to obtain truncation error formula while maintaining correctness of Algorithm 2.

- The least-squares procedure to estimate $\hat{B}$ and $\hat{D}$ is a particular case of the NLS identification algorithm where $\hat{A}$ and $\hat{C}$ as well are estimated. The NLS is not suitable for narrow band data if model fit is measured in the $\mathcal{H}_\infty$-norm.
- To reduce model mismatch, model orders should be increased. As this happens, pole-zero sensitivity of the model increases. Example 1 of this section illustrates a model error fluctuation at high orders for the NLS.
- Since Algorithm 1 and Algorithm 4 yield identical asymptotic poles, the asymptotic performance of Algorithm 4 should be expected between the NLS and Algorithm 1.

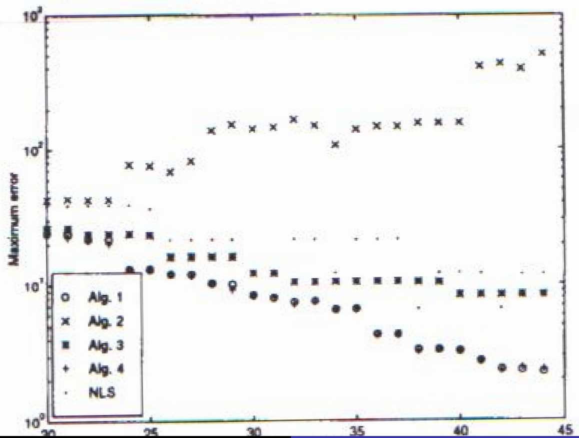
Example 1 Real data set obtained at the Jet Propulsion Laboratory, Pasadena, California originating from a frequency-response experiment on a flexible structure.

• The JPL-data consist of a total of $M = 512$ complex frequency samples in the frequency range [1.23,628] and have several lightly damped modes.

• The discrete-time models matching the given frequency response were constructed applying zero-order hold sampling equivalence and five algorithms.

• $q = r = 512$.

Plot of $\|\hat{G}^M - G\|_{m,\infty} = \max_{\omega_k} |\hat{G}^M(e^{j\omega_k}) - G_k|$ for different model orders and algorithms in Example 1.



- Algorithm 3 was tested by Friedman and Khargonekar: 1995 on the JPL-data.
- The pre-identified model had a finite-impulse response represented by 1024 coefficients and was reduced by a recursively implemented model reduction procedure. (With this choice of model order, the data are entirely explained by the model!)
- The use of an FIR model as an intermediate step in the identification leads to less accurate models as compared with a direct approximation of a rational model to the given data using a correct algorithm.

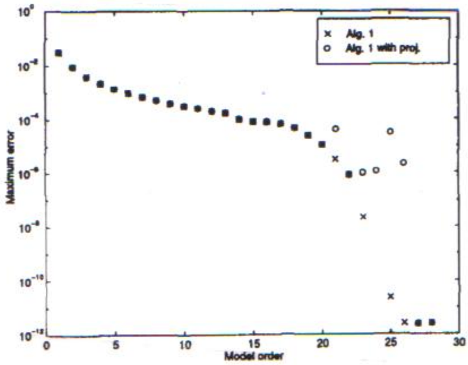
Example 2 Consider the problem of approximating

$$G(s) = \frac{1}{s + 1 - e^{-2-s}} \tag{4}$$

with a finite-dimensional linear model (Gu *etal.*: 1989).

- 512 uniformly spaced noise-free frequency-response data on $[0, \pi]$ derived from (4) by use of the bilinear map.
- $q = r = 512$ which gives the maximal size Hankel matrix.
- Aprroximation errors: 1st order: 3.1×10^{-2} of Algorithm 1 versus 3.2×10^{-2} of Gu *etal.*: 1989; 24th-order: 1.4×10^{-6} of Algorithm 1 versus 7.9×10^{-3} of Gu *etal.*: 1989; 27th-order: 2.4×10^{-12} of Algorithm 1.

Plot of $\|\hat{G}^M - G\|_{m,\infty}$ for different model orders in Example 2 using Algorithm 1 with "x" and without "o" projecting unstable eigenvalues of $\hat{A}$ into the unit disk.



- We presented a correct, frequency domain subspace-based algorithm yielding w.p.1 balanced truncations of the identified system.
- Two examples were used to illustrate the properties of different algorithms and to show the practical applicability of the algorithms.