Frequency Domain Subspace-Based Identification

of Discrete-Time Power Spectra from Nonuniformly Spaced Measurements

Hüseyin Akçay

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

October 17, 2008

Outline

Identification of multi-input/multi-output systems from a measured power spectrum arises in certain applications; for example, the design of shaping filters for noise processes.

- A practical application is the modeling of stochastic road disturbances experienced by a vehicle moving forward.
  - Model road spectrum by a rational transfer function of reasonably low order and to use this approximation for a design of a *linear shaping filter* with a white noise input.
  - Formulate the vehicle control problem in standard form.
  - The algorithm of this paper determines a state-space realization of road spectrum.
- Other applications : modeling of acoustic power spectra and modeling of passenger sensitivity for car accelerations (Van Overschee *etal.*:1997).

- A *parametric* approach uses a non-linear least-squares criterion optimized by an iterative search in the parameter space (Kay:1988,Priestley:1989,Stoica and Moses:1997):
  - Convergence problems and difficulty of parameterizing MIMO systems to determine the canonical models.
- The subspace approach does not suffer from any of these inconveniences:
  - No explicit need for parameterization since full state-space models are used and the only parameter is the order of the system.
  - Non-iterative.

- A subspace algorithm using spectrum samples at uniformly spaced frequencies was presented in Van Overschee *etal.*:1997.
  - It is based on McKelvey *etal.*:1996 and uses biased impulse response coefficients; however, it generates *strongly consistent* power spectrum estimates.
  - A related work is Verhaegen:1996, where a subspace algorithm for the time domain identification of mixed causal and anti-causal systems was proposed. Its frequency domain extension was given in Fraanje *etal.*:2003.

Current work The restriction on the frequencies are removed.

- A subspace-based algorithm to identify MIMO systems from power spectrum samples measured at nonuniformly spaced frequencies is developed.
- This algorithm is shown not only strongly consistent but also recovers finite-dimensional rational spectra given a finite number of noise-free data (depending on the model order).

Consider a multi-input/multi-output square linear-time invariant discrete-time system represented by the state-space equations:

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

where $x(t) \in \mathbf{R}^n$ is the state, $u(t) \in \mathbf{R}^m$ and $y(t) \in \mathbf{R}^m$ are, respectively, the input and the output of the system.

- The transfer function of the system:

$$G(z) = D + C(zI_n - A)^{-1}B.$$

Assumption 1 The system (1) is stable and strictly minimum phase: all eigenvalues of A and $A - BD^{-1}C$ lie strictly inside the unit circle. The pairs $\{A, B\}$ and $\{A, C\}$ are controllable and observable, respectively. All eigenvalues of A are nonzero and distinct.

- The system (1) is a minimal stochastic system.
- Since the *Jordan canonical form* is not numerically stable, a slight perturbation of A will lead to distinct eigenvalues if A has repeated eigenvalues.

Assume $u(t)$ is zero-mean, unity-variance white noise process. Then, the power spectrum associated with (1) is defined as

$$S(z) = G(z)G^T(z^{-1}). \tag{2}$$

- The system (1) is called the *innovation form*, unity variance, minimum phase spectral factor associated with the power spectrum $S(z)$.
- From (2) and Assumption 1,

$$S(e^{j\theta}) > 0, \qquad \text{for all } \theta. \tag{3}$$

  - The positive realness condition imposing a constraint on the given spectrum samples S_k, *i.e.*, $S_k > 0$ for each k, as well as on the identified power spectrum $\widehat{S}_N(z)$.

Noise assumptions

The noise η corrupting the spectrum samples is a zero-mean, complex, white-noise process with a covariance function satisfying

$$\mathbf{E}\begin{bmatrix} \operatorname{Re}\eta_k \\ \operatorname{Im}\eta_k \end{bmatrix}\begin{bmatrix} \operatorname{Re}\eta_s^T & \operatorname{Im}\eta_s^T \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\mathcal{R}_k & 0 \\ 0 & \frac{1}{2}\mathcal{R}_k \end{bmatrix}\delta_{ks}. \tag{4}$$

Furthermore, we assume that the fourth order moments are bounded above by some $M_\eta < \infty$ as

$$\mathbf{E}\,\|\eta_k\|_F^4 \leq M_\eta, \qquad \text{for all } k. \tag{5}$$

Assumption on the frequencies

$\mathcal{C}(X)$: the number of elements in a given set X.

" $\mathcal{C}(\{\theta_k\}_{k=1}^N \cap [a,b])$ is the number of frequencies contained in $[a,b] \subseteq [0,2\pi]$ ".

Assumption 2

$$\lim_{N\to\infty} \inf \frac{1}{N} \mathcal{C}(\{\theta_k\}_{k=1}^N \cap [a,b]) \geq \delta(b-a) \tag{6}$$

for every $[a,b] \subseteq [0,2\pi]$ and some fixed $\delta > 0$.

- Every point on the unit circle has a nonzero asymptotic density of frequencies relative to N.

Given: N noisy samples $S_k \in \mathbf{C}^{m\times m}$ of the power spectrum $S(z)$ evaluated at N points on the unit circle:

$$S_k = S(e^{j\theta_k}) + \eta_k, \qquad k = 1, 2, \cdots, N, \tag{7}$$

Find: $(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{D})$ such that the estimated power spectrum

$$\widehat{S}_N(z) = \widehat{G}(z)\widehat{G}^T(z^{-1})$$

is strongly consistent, *i.e.,*

$$\lim_{N\to\infty} \|\widehat{S}_N - S\|_\infty = 0, \qquad \text{w.p.1.}$$

where

$$\begin{aligned} \widehat{G}(z) &= \widehat{C}(zI_n - \widehat{A})^{-1}\widehat{B} + \widehat{D}, \\ \|G\|_\infty &= \sup_{0\le\theta<2\pi} \sigma_{\max}(G(e^{j\theta})). \end{aligned}$$

We also require the algorithm to produce the true model if the noise is zero given a finite amount of data N, *i.e.,* there exists an $N_0 < \infty$ such that

$$\|\widehat{S}_N - S\|_\infty = 0, \qquad \text{for all } N \geq N_0. \tag{8}$$

- An identification algorithm which satisfies (8) is called *correct algorithm*. Strong consistency is a most natural requirement for any useful algorithm. The correctness is particularly important for spectra with sharp peaks.
- This identification problem can be thought as the design of a linear shaping filter (A, B, C, D) from (corrupted) power spectrum measurements where the zeros of $G(z)$ can be restricted, without loss generality, to be minimum phase.

Outline

Theorem 1 [Caines:1988] Consider the power spectrum $S(z)$ in (2). Suppose that Assumption 1 holds. Let P be the solution of the discrete-time Lyapunov equation:

$$P = APA^T + BB^T.$$

Let

$$E = CPC^T + DD^T,$$
$$F = APC^T + BD^T,$$

Then $S(z)$ can be split into the sum of two system transfer matrices as follows

$$S(z) = H(z) + H^T(z^{-1})$$

with

$$H(z) = \frac{1}{2}E + C(zI_n - A)^{-1}F.$$

- This splitting of $S(z)$ into the sum of a causal transfer function $H(z)$ and an anti-causal transfer function $H^T(z^{-1})$ is the first step of our algorithm. (It is also the starting point of the algorithm of Van Overschee *etal.*:1997).
- From the spectrum samples, $(A, F, C, \frac{1}{2}E)$ describing the spectral summand $H(z)$ is estimated.
- The algorithm proposed in Van Overschee *etal.*:1997 uses biased Markov parameters of $S(z)$ as in McKelvey *etal.*:1996; and requires the discrete frequencies θ_k, $k = 1, 2, \cdots, N$ be uniformly spaced.
- The restriction on the frequencies is removed in this work.

From Theorem 1, we write a state-representation of $S(z)$ as follows:

$$x^{c}(t+1) = Ax^{c}(t) + Fu(t) \quad (9)$$
$$x^{ac}(t-1) = A^{T}x^{ac}(t) + C^{T}u(t) \quad (10)$$
$$y^{s}(t) = Cx^{c}(t) + F^{T}x^{ac}(t) + Eu(t). \quad (11)$$

- These equations are the special cases of the equations considered in Verhaegen:96 for the time-domain subspace identification of mixed causal and anti-causal linear-time invariant systems.

We take the discrete Fourier transforms of equations (9)–(11) where we shift equation (10) by $p-1$ samples forward in time:

$$
\begin{aligned}
e^{j\theta} X^{\mathrm{c}}(\theta) &= AX^{\mathrm{c}}(\theta) + FU(\theta) \\
e^{-j\theta} X^{\mathrm{ac,p}}(\theta) &= A^T X^{\mathrm{ac,p}}(\theta) + C^T e^{j(p-1)\theta} U(\theta) \\
Y^{\mathrm{s}}(\theta) &= CX^{\mathrm{c}}(\theta) + F^T e^{-j(p-1)\theta} X^{\mathrm{ac,p}}(\theta) + EU(\theta)
\end{aligned}
$$

where $X^{\mathrm{c}}(\theta)$, $X^{\mathrm{ac,p}}(\theta)$, $U(\theta)$, and $Y^{\mathrm{s}}(\theta)$ denote the discrete Fourier transforms of $x^{\mathrm{c}}(t)$, $x^{\mathrm{ac}}(t+p-1)$, $u(t)$, and $y(t)$, respectively, and $p > 2n$.

Let $X_i^{\mathrm{c}}(\theta)$ be the resulting state transform when $U(\theta) = e_i$, the unit vector with 1 on the i^{th} position; and $X_i^{\mathrm{ac,p}}(\theta)$ is defined similarly.

By defining the compound state matrices:

$$\begin{aligned} X_{\mathrm{C}}^{\mathrm{c}}(\theta) &\triangleq \left[X_1^{\mathrm{c}}(\theta) \;\; X_2^{\mathrm{c}}(\theta) \;\; \cdots \;\; X_m^{\mathrm{c}}(\theta)\right], \\ X_{\mathrm{C}}^{\mathrm{ac,p}}(\theta) &\triangleq \left[X_1^{\mathrm{ac,p}}(\theta) \;\; X_2^{\mathrm{ac,p}}(\theta) \;\; \cdots \;\; X_m^{\mathrm{ac,p}}(\theta)\right], \end{aligned}$$

$S(e^{j\theta})$ can be implicitly described as

$$S(e^{j\theta}) = CX_{\mathrm{C}}^{\mathrm{c}}(\theta) + F^T e^{-j(p-1)\theta} X_{\mathrm{C}}^{\mathrm{ac,p}}(\theta) + E$$

with

$$\begin{aligned} e^{j\theta} X_{\mathrm{C}}^{\mathrm{c}}(\theta) &= AX_{\mathrm{C}}^{\mathrm{c}}(\theta) + F, \\ e^{-j\theta} X_{\mathrm{C}}^{\mathrm{ac,p}}(\theta) &= A^T X_{\mathrm{C}}^{\mathrm{ac,p}}(\theta) + C^T e^{j(p-1)\theta}. \end{aligned}$$

By iteratively substituting the state-equations, we obtain the relation

$$\begin{bmatrix} S(e^{j\theta}) \\ e^{j\theta}S(e^{j\theta}) \\ \vdots \\ e^{j(p-2)\theta}S(e^{j\theta}) \\ e^{j(p-1)\theta}S(e^{j\theta}) \end{bmatrix} = \Gamma_p \begin{bmatrix} I_m \\ e^{j\theta}I_m \\ \vdots \\ e^{j(p-2)\theta}I_m \\ e^{j(p-1)\theta}I_m \end{bmatrix} + \mathcal{O}_p \begin{bmatrix} X_{\mathrm{C}}^{\mathrm{c}}(\theta) \\ X_{\mathrm{C}}^{\mathrm{ac,p}}(\theta) \end{bmatrix}$$

where

$$\mathcal{O}_p = \begin{bmatrix} C & F^T(A^T)^{p-1} \\ CA & F^T(A^T)^{p-2} \\ \vdots & \vdots \\ CA^{p-2} & F^TA^T \\ CA^{p-1} & F^T \end{bmatrix}$$

and

$$\Gamma_p = \begin{bmatrix} E & F^T C^T & \cdots & F^T (A^T)^{p-2} C^T \\ CF & E & \cdots & \vdots \\ \vdots & \vdots & \ddots & F^T C^T \\ CA^{p-2}F & \cdots & CF & E \end{bmatrix}$$

By repeating for θ_k, $k = 1, 2, \cdots, N$, we get

$$\mathcal{S}_{\mathrm{C}} = \mathcal{O}_p \mathcal{X}_{\mathrm{C}} + \Gamma_p \mathcal{W}_{\mathrm{C}} \tag{12}$$

where $z_k = e^{j\theta_k}$, $k = 1, 2, \cdots, N$ and

$$\mathcal{S}_{\mathrm{C}} = \frac{1}{\sqrt{N}} \begin{bmatrix} S(z_1) & \cdots & S(z_N) \\ e^{j\theta_1} S(z_1) & \cdots & e^{j\theta_N} S(z_N) \\ \vdots & \ddots & \vdots \\ e^{j(p-1)\theta_1} S(z_1) & \cdots & e^{j(p-1)\theta_N} S(z_N) \end{bmatrix},$$

$$\mathcal{W}_{\mathrm{C}} = \frac{1}{\sqrt{N}} \begin{bmatrix} I_m & \cdots & I_m \\ e^{j\theta_1} I_m & \cdots & e^{j\theta_N} I_m \\ \vdots & \ddots & \vdots \\ e^{j(p-1)\theta_1} I_m & \cdots & e^{j(p-1)\theta_N} I_m \end{bmatrix},$$

$$\mathcal{X}_{\mathrm{C}} = \frac{1}{\sqrt{N}} \begin{bmatrix} X_{\mathrm{C}}^{\mathrm{c}}(\theta_1) & \cdots & X_{\mathrm{C}}^{\mathrm{c}}(\theta_N) \\ X_{\mathrm{C}}^{\mathrm{ac,p}}(\theta_1) & \cdots & X_{\mathrm{C}}^{\mathrm{ac,p}}(\theta_N) \end{bmatrix}.$$

Outline

From (7) and (12),

$$\widehat{\mathcal{S}}_{\mathrm{C}} = \mathcal{O}_p \mathcal{X}_{\mathrm{C}} + \Gamma_p \mathcal{W}_{\mathrm{C}} + \mathcal{N}_{\mathrm{C}} \tag{13}$$

where

$$\widehat{\mathcal{S}}_{\mathrm{C}} = \frac{1}{\sqrt{N}} \begin{bmatrix} S_1 & \cdots & S_N \\ e^{j\theta_1} S_1 & \cdots & e^{j\theta_N} S_N \\ \vdots & \vdots & \ddots \quad \vdots \\ e^{j(p-1)\theta_1} S_1 & \cdots & e^{j(p-1)\theta_N} S_N \end{bmatrix},$$

$$\mathcal{N}_{\mathrm{C}} = \frac{1}{\sqrt{N}} \begin{bmatrix} \eta_1 & \cdots & \eta_N \\ e^{j\theta_1} \eta_1 & \cdots & e^{j\theta_N} \eta_N \\ \vdots & \ddots & \vdots \\ e^{j(p-1)\theta_1} \eta_1 & \cdots & e^{j(p-1)\theta_N} \eta_N \end{bmatrix}.$$

Since $\mathcal{O}_p$ is a real matrix and we are interested in the real range space, we convert (13) into a relation involving only real valued matrices:

$$\begin{aligned} \widehat{\mathcal{S}} &= \mathcal{O}_p\mathcal{X} + \Gamma_p\mathcal{W} + \mathcal{N} \\ &= \mathcal{S} + \mathcal{N} \end{aligned}$$

where

$$\begin{aligned} \widehat{\mathcal{S}} &= \left[\operatorname{Re}\widehat{\mathcal{S}}_C \;\; \operatorname{Im}\widehat{\mathcal{S}}_C\right], \\ \mathcal{S} &= \left[\operatorname{Re}\mathcal{S}_C \;\; \operatorname{Im}\mathcal{S}_C\right], \\ \mathcal{X} &= \left[\operatorname{Re}\mathcal{X}_C \;\; \operatorname{Im}\mathcal{X}_C\right], \\ \mathcal{W} &= \left[\operatorname{Re}\mathcal{W}_C \;\; \operatorname{Im}\mathcal{W}_C\right] \\ \mathcal{N} &= \left[\operatorname{Re}\mathcal{N}_C \;\; \operatorname{Im}\mathcal{N}_C\right]. \end{aligned}$$

Let $\mathcal{W}^{\perp}$ be the projection matrix onto the null space of $\mathcal{W}$:

$$\mathcal{W}^{\perp} \triangleq I_{2mN} - \mathcal{W}^{H}(\mathcal{W}\mathcal{W}^{H})^{-1}\mathcal{W}.$$

The term $\Gamma_p\mathcal{W}$ is cancelled when multiplied from right by $\mathcal{W}^{\perp}$:

$$\begin{aligned}\widehat{\mathcal{S}}\mathcal{W}^{\perp} &= \mathcal{O}_p\mathcal{X}\mathcal{W}^{\perp} + \mathcal{N}\mathcal{W}^{\perp} \\ &= \mathcal{S}\mathcal{W}^{\perp} + \mathcal{N}\mathcal{W}^{\perp}.\end{aligned}$$

- The range space of $\mathcal{S}\mathcal{W}^{\perp}$ equals the range space of $\mathcal{O}_p$ unless rank cancellations occur. A sufficient condition for the range spaces to be equal is that the intersection between the row spaces of $\mathcal{W}$ and $\mathcal{X}$ is empty.

Lemma Let $N \geq (p/2) + n + 1$. Suppose that the frequencies θ_k are distinct and z_k is not an eigenvalue of A for all k. Then

$$\operatorname{rank} \begin{bmatrix} \mathcal{W} \\ \mathcal{X} \end{bmatrix} = pm + 2n \Longleftrightarrow (A, B, C, D) \text{ minimal.}$$

- If the frequencies are distinct, the number of data satisfies $N \geq (p/2) + n + 1$, and (A, B, C, D) is minimal, then the two row spaces of $\mathcal{W}$ and $\mathcal{X}$ do not intersect and the range space of $\mathcal{S}\mathcal{W}^{\perp}$ coincides with the range space of $\mathcal{O}_p$.

In De Moor:93, it was shown that by using the SVD of $\widehat{\mathcal{S}}\mathcal{W}^{\perp}$, the $2n$ left singular vectors corresponding to the $2n$ largest singular values form a strongly consistent estimate of $\text{range}(\mathcal{S}\mathcal{W}^{\perp})$ if the following conditions hold w.p.1

$$\text{(i)} \qquad \lim_{N\to\infty} \mathcal{S}\mathcal{W}^{\perp}(\mathcal{N}\mathcal{W}^{\perp})^{T} = 0; \tag{14}$$

$$\text{(ii)} \qquad \lim_{N\to\infty} \mathcal{N}\mathcal{W}^{\perp}(\mathcal{N}\mathcal{W}^{\perp})^{T} = \alpha I_{pm} \tag{15}$$

for some scalar $\alpha \geq 0$.

In McKelvey *etal.*:1996, it was shown under the noise assumptions that (14) holds and

$$\lim_{N\to\infty} \mathcal{N}\mathcal{W}^{\perp}(\mathcal{N}\mathcal{W}^{\perp})^{T} = \mathcal{K}\mathcal{K}^{T}, \qquad \text{w.p.1.}$$

where $\mathcal{K} \in \mathbf{R}^{pm \times pm}$ is a matrix defined by

$$\mathcal{K}\mathcal{K}^T = \mathrm{Re}\left(\mathcal{W}_C \mathcal{R} \mathcal{W}_C^H\right),$$

$$\mathcal{R} = \begin{bmatrix} \mathcal{R}_1 & 0 & \cdots & 0 \\ 0 & \mathcal{R}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathcal{R}_N \end{bmatrix}.$$

The matrix $\mathcal{K}$ can be found by a Cholesky decomposition. Thus, from (14) we have the weighted version

$$\mathcal{K}^{-1}\widehat{\mathcal{S}}\mathcal{W}^{\perp} = \mathcal{K}^{-1}\mathcal{S}\mathcal{W}^{\perp} + \mathcal{K}^{-1}\mathcal{N}\mathcal{W}^{\perp} \tag{16}$$

satisfying (14) and (15) with $\alpha = 1$.

- Hence, the $2n$ left singular vectors corresponding to the $2n$ largest singular values of $\mathcal{K}^{-1}\widehat{\mathcal{S}}\mathcal{W}^{\perp}$ will form a strongly consistent estimate of the range space of $\mathcal{K}^{-1}\mathcal{S}\mathcal{W}^{\perp}$ which equals to the range space of $\mathcal{K}^{-1}\mathcal{O}_p$.

A numerically efficient way of forming $\widehat{\mathcal{S}}\mathcal{W}^{\perp}$ is to use the QR-factorization:

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

A simple derivation yields

$$\widehat{\mathcal{S}}\mathcal{W}^{\perp} = R_{22}Q_2^T$$

and it suffices to use R_{22} since Q_2^T is a matrix of full rank.

Thus, the $2n$ left singular vectors corresponding to the $2n$ largest singular values of $\mathcal{K}^{-1}\widehat{\mathcal{S}}\mathcal{W}^{\perp}$ are obtained from the SVD:

$$\mathcal{K}^{-1}R_{22} = \begin{bmatrix} \widehat{U}_{2n} & \widetilde{U} \end{bmatrix} \begin{bmatrix} \widehat{\Sigma}_{2n} & 0 \\ 0 & \widetilde{\Sigma} \end{bmatrix} \begin{bmatrix} \widehat{V}_{2n} \\ \widetilde{V} \end{bmatrix}$$

where this decomposition is partitioned such that $\widehat{\Sigma}_{2n}$ contains the $2n$ largest singular values.

- Our consistency analysis has shown that

$$\lim_{N\to\infty} \mathcal{K}\widehat{U}_{2n} = \mathcal{O}_p T, \qquad \text{w.p.1}$$

for some nonsingular matrix T. In the calculation of $\widehat{U}_{2n}$, $2n$ elements with fixed indices can be chosen freely subject to the constraint that magnitudes are not greater than unity. Thus, by fixing values of those elements for all N, $\widehat{U}_{2n}$ converges to a matrix denoted by U_{2n} w.p.1 as $N \to \infty$:

$$\mathcal{K}U_{2n} = \mathcal{O}_p T. \tag{17}$$

- This asymptotic formula (in the number of data) will be the key in the development of our algorithm.

Lemma Let S_k, $k = 1, \cdots, N$ be noise-free samples of the power spectrum of a discrete-time system of order n satisfying Assumption 1 at N distinct frequencies θ_k. Furthermore, let $N \geq (p/2) + n + 1$ and $\mathcal{K} \in \mathbf{R}^{pm \times pm}$ be any nonsingular matrix. Then, for some nonsingular T

$$\mathcal{K}\widehat{U}_{2n} = \mathcal{O}_p T.$$

- Thus, the equations derived from the asymptotic formula are also valid for a finite number of data.

Let J_{u} and J_{d} be the upward and downward shift matrices defined by

$$J_{\mathrm{u}}\mathcal{O}_p \triangleq \begin{bmatrix} CA & F^T(A^T)^{p-2} \\ \vdots & \vdots \\ CA^{p-2} & F^TA^T \\ CA^{p-1} & F^T \end{bmatrix}$$

$$J_{\mathrm{d}}\mathcal{O}_p \triangleq \begin{bmatrix} C & F^T(A^T)^{p-1} \\ \vdots & \vdots \\ CA^{p-2} & F^TA^T \end{bmatrix}$$

Then,

$$J_{\mathrm{u}}\mathcal{O}_p = J_{\mathrm{d}}\mathcal{O}_pA'$$

where

$$A' = \begin{bmatrix} A & 0 \\ 0 & (A^T)^{-1} \end{bmatrix}.$$

Hence,

$$A' = (J_{\mathrm{d}}\mathcal{O}_p)^{\dagger} J_{\mathrm{u}}\mathcal{O}_p = TA''T^{-1}$$

where

$$A'' \stackrel{\Delta}{=} (J_{\mathrm{d}}\mathcal{K}U_{2n})^{\dagger} J_{\mathrm{u}}\mathcal{K}U_{2n}.$$

- A' and A'' are similar matrices, *i.e.,* they have the same Jordan blocks in their Jordan canonical representations.

Likewise,

$$C' \stackrel{\Delta}{=} [C \;\; F^T(A^T)^{p-1}] = J_f \mathcal{O}_p = C'' T^{-1}$$

where

$$\begin{aligned} J_f &= [I_m \;\; 0_{m\times(p-1)}], \\ C'' &= J_f \mathcal{K} U_{2n}. \end{aligned}$$

Put A'' into the following Jordan canonical form:

$$A'' = [\Pi_c \;\; \Pi_{ac}] \begin{bmatrix} \Sigma_c & 0 \\ 0 & \Sigma_{ac} \end{bmatrix} [\Pi_c \;\; \Pi_{ac}]^{-1} \tag{18}$$

where the eigenvalues of Σ_c lie inside the unit circle.

- Since $H(z)$ is invariant to similarity transformations, let

$$A = \Sigma_c \tag{19}$$

- The canonical form is invariant to ordering of eigenvalues as long as the eigenvalues and the corresponding eigenvectors of Σ_c are permuted accordingly, in complex pairs.
- From the similarity of A'' to A', we may also let

$$\Sigma_{ac} = (\Sigma_c^T)^{-1}. \tag{20}$$

imposing of course a certain structure on T!

Let

$$\Pi = [\Pi_c \;\; \Pi_{ac}] \, .$$

Then,

$$A' = TA''T^{-1} = \Pi^{-1} A'' \Pi.$$

The relations among Σ, Π, and T are captured in the following lemma. Recall that A has distinct eigenvalues.

Lemma Consider the Jordan canonical form of A'' given by (18) where A and Σ_{ac} satisfy (19) and (20). Then, Σ_c is a block diagonal matrix:

$$\Sigma_c = \begin{bmatrix} \Sigma_1 & 0 & \cdots & 0 \\ 0 & \Sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_k \end{bmatrix}, \qquad \Sigma_i \in \mathbf{R}^{n_i \times n_i}$$

where $n_i \in \{1, 2\}$, $\nu_i \neq 0$, and

$$\Sigma_i \triangleq \begin{cases} \mu_i, & \text{if } n_i = 1 \\ \begin{bmatrix} \mu_i & \nu_i \\ -\nu_i & \mu_i \end{bmatrix}, & \text{if } n_i = 2. \end{cases}$$

Also, Σ_{ac} is a block diagonal matrix with block types and sizes compatible with Σ_c.

For some Λ_c and Λ_{ac} compatible with Σ_c, the following holds

$$\Pi = T^{-1}\Lambda$$

where

$$\Lambda \triangleq \begin{bmatrix} \Lambda_c & 0 \\ 0 & \Lambda_{ac} \end{bmatrix}.$$

Let X and Y be two block diagonal matrices with block sizes and types compatible with Σ_c, then X^T, XY and X^{-1} are also compatible with Σ_c and $XY = YX$.

Now, multiplying (17) from right by Π, we get

$$\mathcal{K}U_{2n}\Pi = \mathcal{O}_p\Lambda.$$

Hence,

$$\mathcal{K}U_{2n}\Pi_{\mathrm{c}} = \begin{bmatrix} C\Lambda_{\mathrm{c}} \\ \vdots \\ C\Sigma_{\mathrm{c}}^{p-2}\Lambda_{\mathrm{c}} \\ C\Sigma_{\mathrm{c}}^{p-1}\Lambda_{\mathrm{c}} \end{bmatrix},$$

$$\mathcal{K}U_{2n}\Pi_{\mathrm{ac}} = \begin{bmatrix} F^T(\Sigma_{\mathrm{c}}^T)^{p-1}\Lambda_{\mathrm{ac}} \\ \vdots \\ F^T\Sigma_{\mathrm{c}}^T\Lambda_{\mathrm{ac}} \\ F^T\Lambda_{\mathrm{ac}} \end{bmatrix}.$$

Thus,

$$C\Lambda_{\mathrm{c}} = J_{\mathrm{f}}\mathcal{K}U_{2n}\Pi_{\mathrm{c}}, \qquad F^T\Lambda_{\mathrm{ac}} = J_{\mathrm{l}}\mathcal{K}U_{2n}\Pi_{\mathrm{ac}}$$

where

$$J_{\mathrm{l}} \triangleq [0_{m\times(p-1)} \;\; I_m].$$

- The problem of finding C, F, and E is now reduced to estimating E, Λ_{c}, and Λ_{ac} from the spectral data.

From the lemma, $S(z)$ can be written as

$$S(z) = E + \chi(z)\,Z + Z^T\,\chi^T(z^{-1})$$

where

$$\begin{aligned}\chi(z) &= J_{\mathrm{f}}\mathcal{K}U_{2n}\Pi_{\mathrm{c}}\left(zI_n-\Sigma_{\mathrm{c}}\right)^{-1},\\ Z &= \Lambda_{\mathrm{c}}^{-1}F.\end{aligned}$$

Thus, E and Z can be estimated from the data by solving the following linear least-squares problem:

$$\begin{aligned}E^{\sharp},Z^{\sharp} = \arg\min_{\check{E},\check{Z}}\sum_{k=1}^{N}\|\mathcal{R}_k^{-\frac{1}{2}}(\chi(z_k)\check{Z}+\check{Z}^T\chi^T(z_k^{-1})\\ +\check{E}-S(z_k))\|_F^2.\end{aligned}$$

- The formula (21) is non-asymptotic in N though asymptotic quantities are used in it. However, it suggests a scheme to consistently estimate the parameters A, C, E, and F.

- When the spectrum samples are noise-free, we can replace U_{2n} with $\widehat{U}_{2n}$.

Lemma Let $S(z)$ be the power spectrum of a discrete-time system of order n satistying Assumption 1. If $N \geq (p/2) + n + 1$, then $E^\sharp = E, \qquad Z^\sharp = Z$.

- Once we find Z, we calculate C and F as follows

$$C = J_{\mathrm{f}} \mathcal{K} U_{2n} \Pi_{\mathrm{c}}, \qquad F = Z$$

due to the fact that $H(z)$ is invariant to post-multiplication of C by Λ_{c} and pre-multiplication of F by $\Lambda_{\mathrm{c}}^{-1}$ since $\Lambda_{\mathrm{c}}^{-1}(zI_n - \Sigma_{\mathrm{c}})^{-1}\Lambda_{\mathrm{c}} = (zI_n - \Sigma_{\mathrm{c}})^{-1}$.

Determination of B and D:

- Solve the following *Riccati* equation for P:

$$P = APA^T + (F - APC^T)(E - CPC^T)^{-1}(F - APC^T)^T.$$

- Compute B and D as follows

$$\begin{aligned} B &= (F - APC^T)(E - CPC^T)^{-\frac{1}{2}}; \\ D &= (E - CPC^T)^{\frac{1}{2}}. \end{aligned}$$

Let

$$\widetilde{A} \triangleq (J_{\mathrm{d}} \mathcal{K} \widehat{U}_{2n})^{\dagger} J_{\mathrm{u}} \mathcal{K} \widehat{U}_{2n}$$

and put $\widetilde{A}$ into the Jordan canonical form:

$$\widetilde{A} \triangleq \begin{bmatrix} \widehat{\Pi}_{\mathrm{c}} & \widehat{\Pi}_{\mathrm{ac}} \end{bmatrix} \begin{bmatrix} \widehat{\Sigma}_{\mathrm{c}} & 0 \\ 0 & \widehat{\Sigma}_{\mathrm{ac}} \end{bmatrix} \begin{bmatrix} \widehat{\Pi}_{\mathrm{c}} & \widehat{\Pi}_{\mathrm{ac}} \end{bmatrix}^{-1}$$

where the eigenvalues of $\widehat{\Sigma}_{\mathrm{c}}$ lie inside the unit circle. Let

$$\begin{aligned} \widehat{A} &\triangleq \widehat{\Sigma}_{\mathrm{c}}, \\ \widehat{C} &\triangleq J_{\mathrm{f}} \mathcal{K} \widehat{U}_{2n} \widehat{\Pi}_{\mathrm{c}}. \end{aligned}$$

Then,

$$\lim_{N \to \infty} \widetilde{A} = A'', \qquad \text{w.p.1.}$$

We can freely choose $2n$ elements of $\widehat{\Pi}_c$ and $\widehat{\Pi}_{ac}$ with fixed indices subject to the constraint that magnitudes are not greater than unity. Then, by fixing values of those elements equal to the values of the corresponding elements in Π_c and Π_{ac} for all N,

$$\lim_{N\to\infty} \widehat{\Sigma}_c = \Sigma_c \quad \text{and} \quad \lim_{N\to\infty} \widehat{\Sigma}_{ac} = \Sigma_{ac}, \qquad \text{w.p.1;}$$

$$\lim_{N\to\infty} \widehat{\Pi}_c = \Pi_c \quad \text{and} \quad \lim_{N\to\infty} \widehat{\Pi}_{ac} = \Pi_{ac}, \qquad \text{w.p.1.}$$

Let

$$\widehat{\chi}(z) \stackrel{\Delta}{=} \widehat{C}(zI_n - \widehat{\Sigma}_c)^{-1}.$$

Then, as $N \to \infty$

$$\lim_{N\to\infty} \|\widehat{\chi} - \chi\|_\infty, \qquad \text{w.p.1.}$$

- The uniform convergence is due to the fact that the spectral radius of the limit matrix Σ_c is less than one.

The estimates of E and F are obtained by solving the following linear least-squares problem:

$$\widehat{E}, \widehat{F} \quad \triangleq \quad \arg\min_{\check{E},\check{F}} \sum_{k=1}^{N} \| \mathcal{R}_k^{-1/2} (\widehat{\chi}(z_k)\,\check{F} + \check{F}^T\,\widehat{\chi}^T(z_k^{-1}) + \check{E} - S_k) \|_F^2$$

Outline

Algorithm 1 Subspace algorithm with nonuniformly spaced spectrum samples:

- 1. Given the data S_k, θ_k, and the covariance data $\mathcal{R}_k$, form the matrices $\mathcal{S}$, $\mathcal{W}_{\mathrm{C}}$, $\mathcal{W}$, and $\mathcal{K}$.
- 2. Calculate the QR-factorization.
- 3. Calculate the SVD.
- 4. Determine the system order n by inspecting the singular values and partition the SVD such that $\widehat{\Sigma}_{2n}$ contains the $2n$ largest singular values.
- 5. Calculate $\widetilde{A}$.
- 6. Block-diagonalize $\widetilde{A}$.

- 7. Calculate $\widehat{C}$.
- 8. Solve the least-squares problem for $\widehat{E}$ and $\widehat{F}$ where $\widehat{\chi}$.
- 9. Solve the Riccati equation for $\widehat{P}$:

$$\begin{aligned}\widehat{P} &= \widehat{A}\widehat{P}\widehat{A}^T + (\widehat{F} - \widehat{A}\widehat{P}\widehat{C}^T)(\widehat{E} - \widehat{C}\widehat{P}\widehat{C}^T)^{-1} \\ &\quad \cdot(\widehat{F} - \widehat{A}\widehat{P}\widehat{C}^T)^T\end{aligned}$$

and calculate $\widehat{B}$ and $\widehat{D}$ from

$$\begin{aligned}\widehat{B} &= (\widehat{F} - \widehat{A}\widehat{P}\widehat{C}^T)(\widehat{E} - \widehat{C}\widehat{P}\widehat{C}^T)^{-\frac{1}{2}}; \\ \widehat{D} &= (\widehat{E} - \widehat{C}\widehat{P}\widehat{C}^T)^{\frac{1}{2}}.\end{aligned}$$

- 10. Calculate $\widehat{G}(z)$ and $\widehat{S}_N(z)$.

Outline

Theorem 2 Consider Algorithm 1 with N noise-free samples of the power spectrum of a discrete-time system of order n satisfying Assumption 1 at N distinct frequencies θ_k. Let $\mathcal{K} \in \mathbf{R}^{pm \times pm}$ be any nonsingular matrix. If $N \geq (p/2) + n + 1$, then Algorithm 1 is correct.

Theorem 3 Consider Algorithm 1 with corrupted measurements of the power spectrum of a discrete-time system of order n satisfying Assumption 1 where the corruptions and the frequencies satisfy the assumptions (4), (5), and (6). Then, Algorithm 1 is strongly consistent.

- The algorithm in Van Overschee *etal.*:1997 is a special case of Algorithm 1. The only difference between the algorithms is the choice of the annihilator $\mathcal{W}^{\perp}$.
- The power spectrum estimated by the above algorithm may not satisfy the positivity requirement due to noise and undermodeling. If a positive definite solution the Riccati equation fails to exist, then the spectral factor can not be computed. Two methods enforcing the positivity condition outlined in VanOverschee *etal.*:1997 can be integrated into Algorithm 1 without modification.

Outline

The true system $G(z) = C(zI_4 - A)^{-1}B + D$:

$$A = \begin{bmatrix} 0.8876 & 0.4494 & 0 & 0 \\ -0.4494 & 0.7978 & 0 & 0 \\ 0 & 0 & -0.6129 & 0.0645 \\ 0 & 0 & -6.4516 & -0.7419 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.2247 \\ 0.8989 \\ 0.0323 \\ 0.1290 \end{bmatrix}$$

$$C = [0.4719 \;\; 0.1124 \;\; 9.6774 \;\; 1.6129]$$

$$D = 0.9626.$$

N noisy samples of $S(z)$:

$$S_k = S(e^{j\theta_k}) + \widetilde{S}(e^{j\theta_k})\nu_k, \qquad k = 1, \cdots, N$$

where the noise term $\widetilde{S}(e^{j\theta_k})\nu_k$ is composed of a noise transfer function $\widetilde{S}(z)$, given by a second-order state-space model:

$$\widetilde{S}(z) = \widetilde{C}(zI_2 - \widetilde{A})^{-1}\widetilde{B} + \widetilde{D}$$

with

$$\begin{aligned} \widetilde{A} &= \begin{bmatrix} 0.6296 & 0.0741 \\ -7.4074 & 0.4815 \end{bmatrix}, \quad \widetilde{B} = \begin{bmatrix} 0.04 \\ 0.9 \end{bmatrix} \\ \widetilde{C} &= [1.6300 \;\; 0.0740], \quad \widetilde{D} = 0.2 \end{aligned}$$

ν_k: i.i.d. complex, zero-mean, unit-variance random variables.

- The variance of the noise process at each frequency equals $\mathcal{R}_k = |\widetilde{S}(z_k)|^2$.

Frequencies randomly and independently picked from

$$\left[\frac{(2k-1)\pi}{2N}, \frac{(2k+1)\pi}{2N}\right], \qquad k = 1, \cdots, N.$$

Thus, each θ_k has a uniform distribution.

Consistency properties of Algorithm 1

Monte Carlo simulations performed, to estimate the power spectrum given the samples S_k using different noise realizations of ν_k.

For $N = 400$ and fixed frequencies, 100 different noise realizations were generated, and Algorithm 1 with $p = 50$ estimated 100 models.

To assess the quality of the resulting model, both the (measured) supremum norm

$$\|\widehat{S}_n - S\|_{m,\infty} \stackrel{\Delta}{=} \max_{1 \le k \le N} |\widehat{S}_N(z_k) - S(z_k)|$$

and the (measured) $\mathcal{H}_2$ norm

$$\|\widehat{S}_n - S\|_{m,2} \triangleq \left(\frac{1}{N} \sum_{k=1}^{N} |\widehat{S}_N(z_k) - S(z_k)|^2 \right)^{\frac{1}{2}}$$

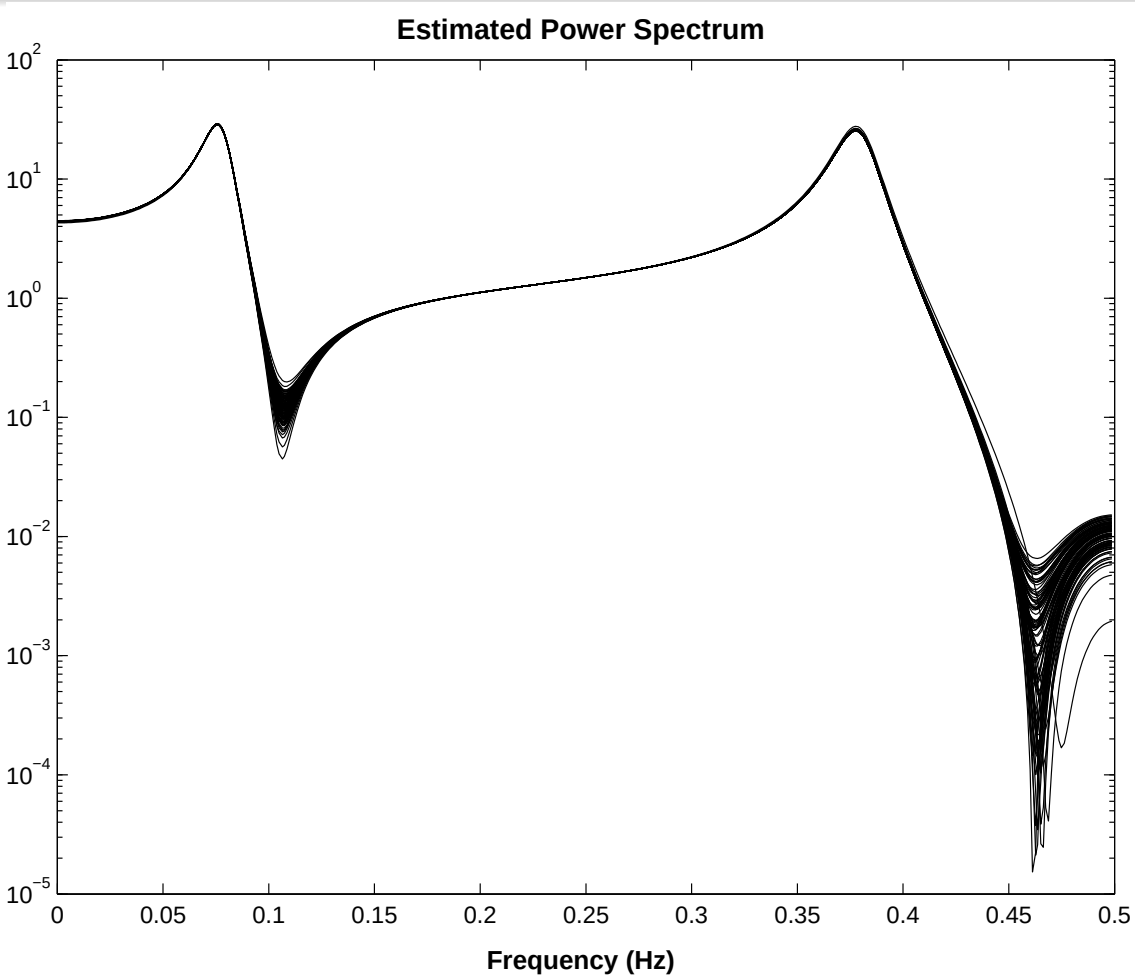
of the estimation error were determined for each estimated model and averaged over the 100 estimated models.

Results:

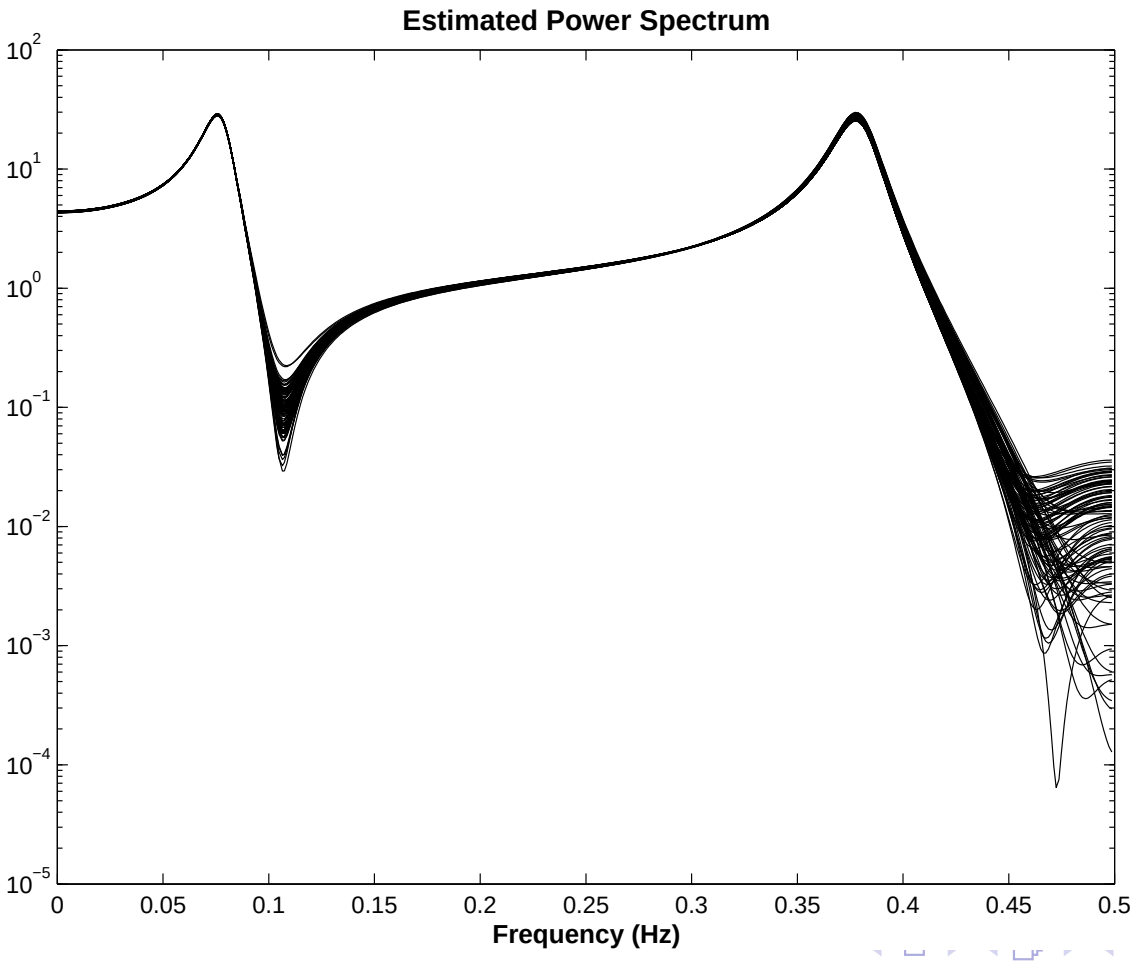
1. $\|\widehat{S}_n - S\|_{m,2} = 0.3307$ and $\|\widehat{S}_n - S\|_{m,\infty} = 2.4208$ using the covariance information.
2. $\|\widehat{S}_n - S\|_{m,2} = 0.4500$ and $\|\widehat{S}_n - S\|_{m,\infty} = 2.6240$ without using the covariance information, *i.e.*, $\mathcal{R}_k = 1$ for all k.

Using noise covariance information reduces $\mathcal{H}_2$ error by 35%!

The results from Monte Carlo simulations for the 100 estimated models using the covariance information.



The results from Monte Carlo simulations for the 100 estimated models without using the covariance information.



Outline

We consider one practical application of Algorithm 1. Provided that occasional large irregularities such as potholes are removed from the analysis,the road surface may be described as a realization of a stationary random process.

- Then, the response of a vehicle traversing a road can be determined by accepted techniques of the theory of random vibration. If the road surface is further assumed to be homogenous and isotrophic, a road profile can be completely described by a single power spectral density evaluated from any single track.

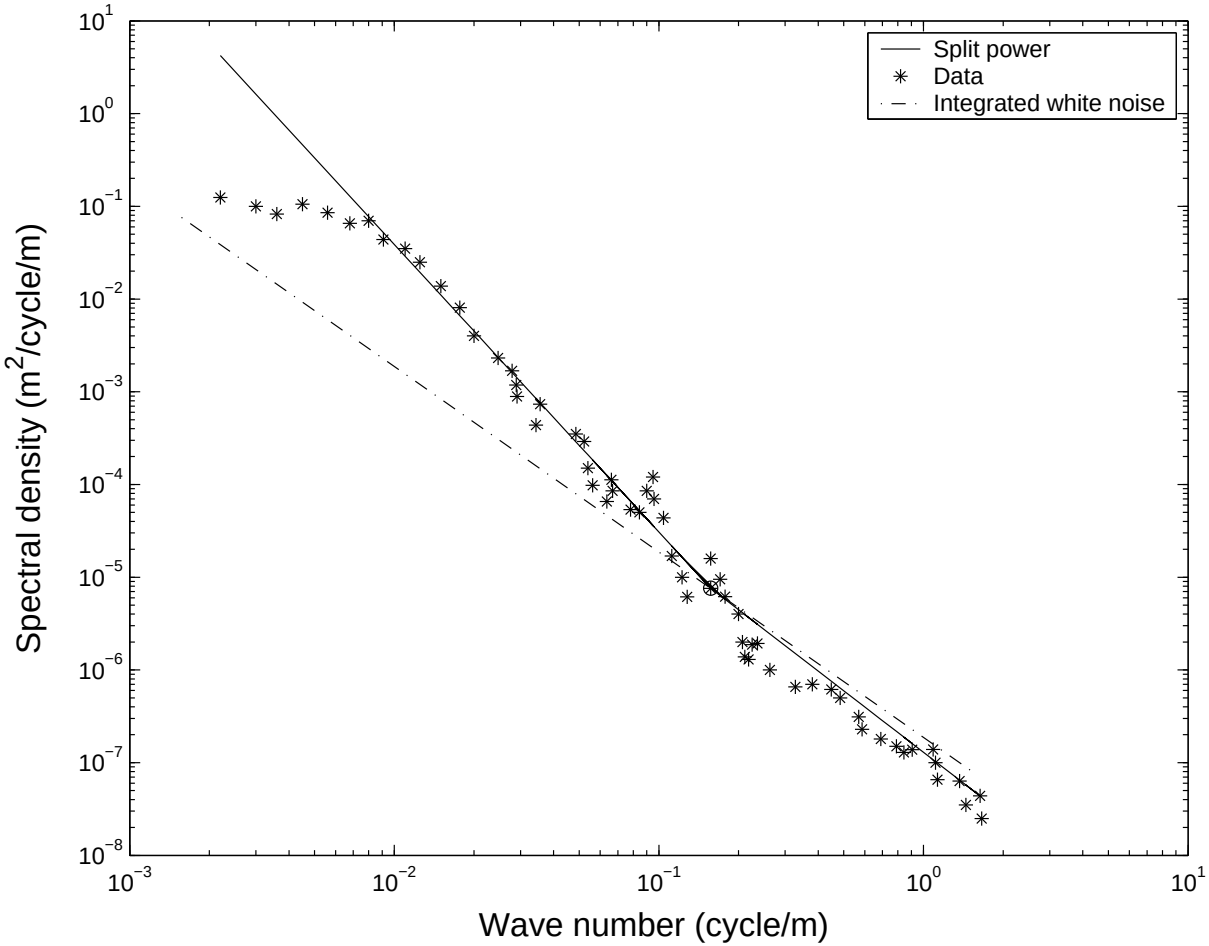
In the figure, plotted are the spectral density of a typical road and its split power law approximation:

$$\widehat{S}(j2\pi\widetilde{n}) = \begin{cases} C|\widetilde{n}/\widetilde{n}_0|^{-2\delta_1}, & 0 < |\widetilde{n}| < \widetilde{n}_0; \\ C|\widetilde{n}/\widetilde{n}_0|^{-2\delta_2}, & \widetilde{n}_0 \leq |\widetilde{n}| < \infty \end{cases}$$

obtained by trial and error for $\widetilde{n}_0 = 0.15708$ cycles/m, $\delta_1 = 1.6$, $\delta_2 = 1.1$, and $C = 0.76 \times 10^{-5}$ and the integrated white noise approximation to the data: $C|\widetilde{n}/\widetilde{n}_0|^{-2}$.

- Clearly, the fit by the integrated white-noise modeling is rather poor; in particular at the frequencies below $\widetilde{n}_0$.
- The split power approximation can not be generated by shape filters; hence, not suitable for simulating the response of vehicle. Besides, it is unbounded at $\widetilde{n} = 0$.

The road power spectrum Dodds and Robson:73 and its approximate modeling by the split power law and the integrated white noise.



Objective Find a low order shape filter whose output spectrum matches the spectral data as closely as possible.

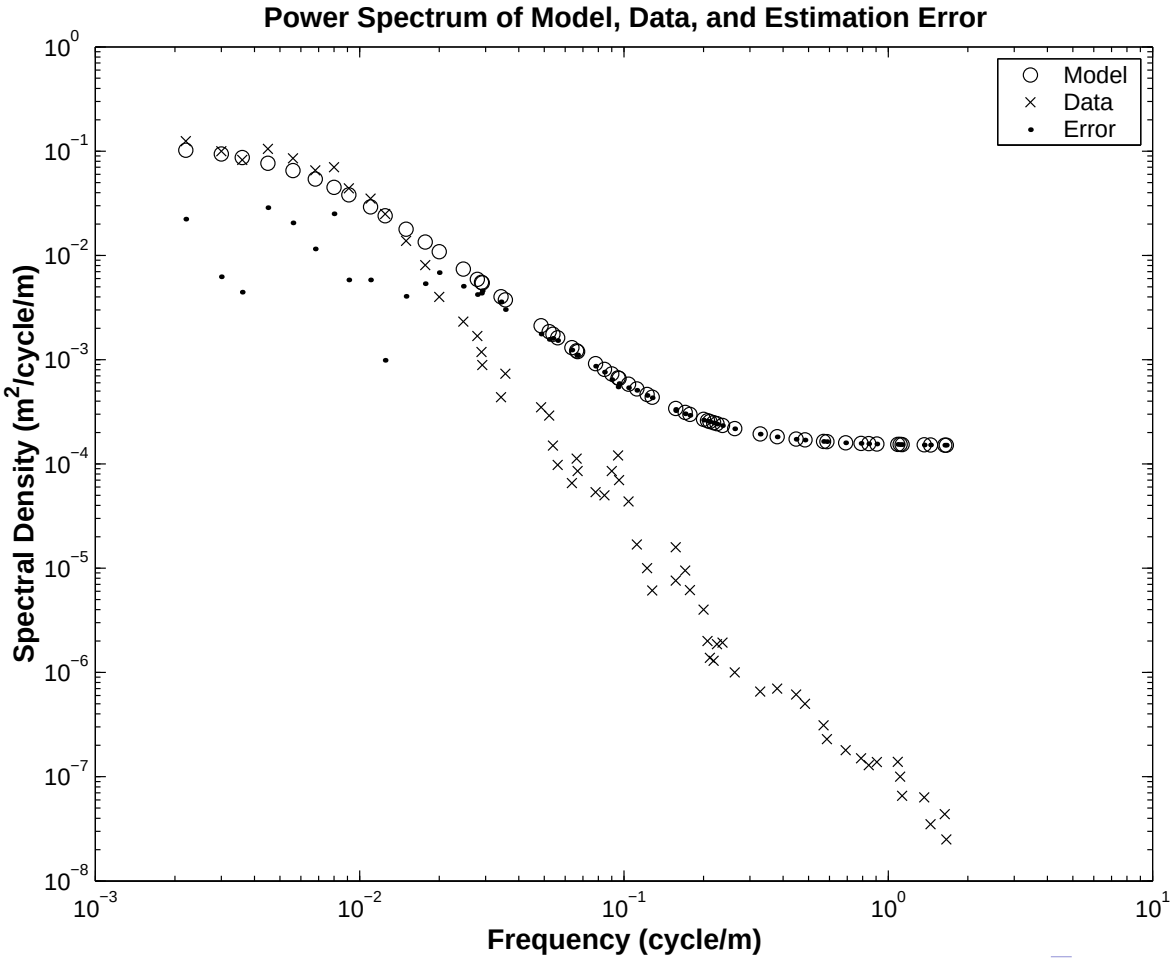
- The continuous-time estimation problem is converted to a discrete one by using the bilinear map:

$$s = \psi(z) = \lambda \frac{z-1}{z+1} \qquad (\lambda > 0).$$

- $N = 63$, $\lambda = 0.2$, $p = 32$, $n = 1$, $\mathcal{R}_N = I_N$ (chosen values).
- The continuous-time spectral factor:

$$\widehat{G}_N(s) = 0.0122 \, \frac{s + 1.1154}{s + 0.0404}.$$

The spectral data and its modeling by a rational model of order one produced by Algorithm 1 with $\mathcal{R} = I_N$.

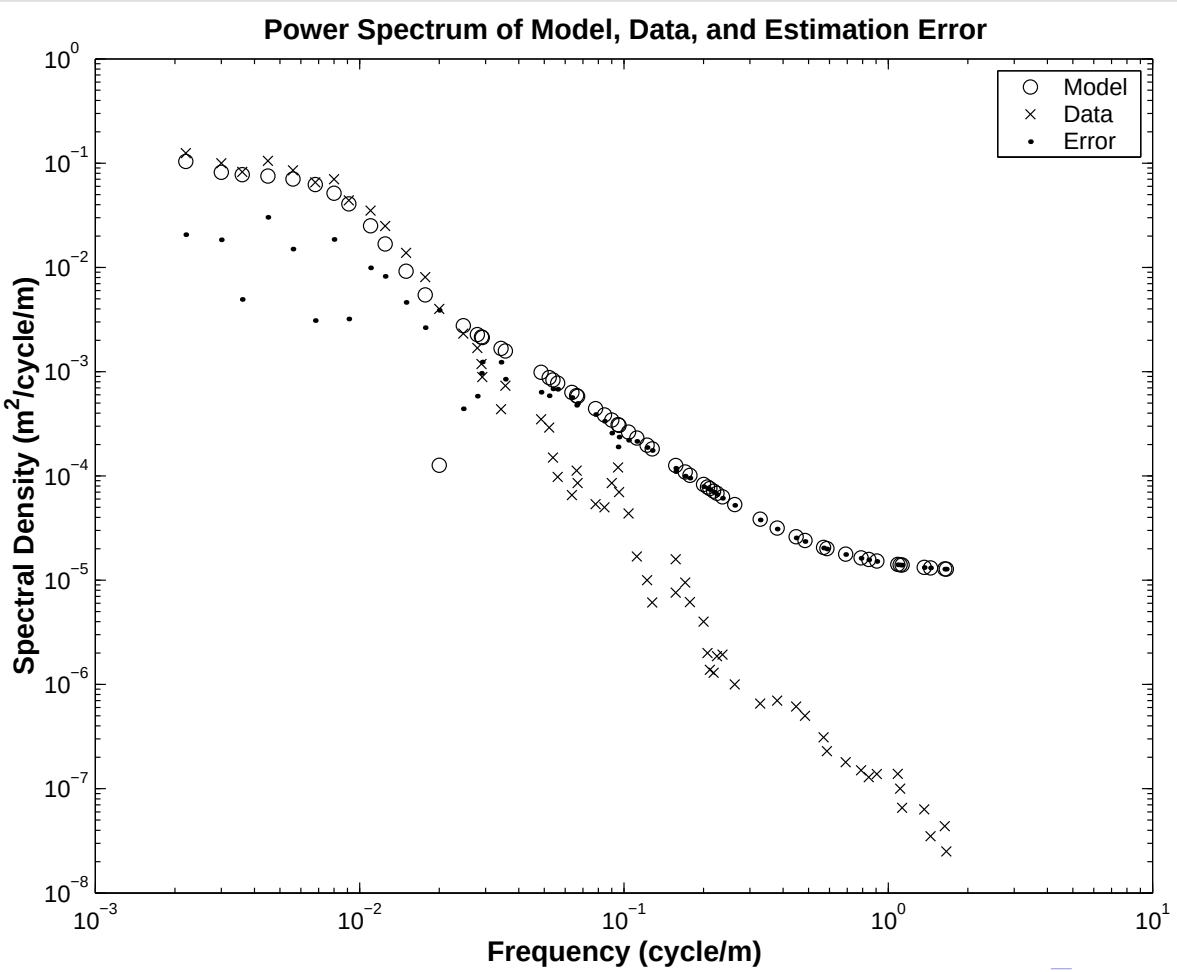


- The first order rational filter is accurate up to 0.02 cycles/m.
- The power spectrum of this model is not integrable ! A convergence factor may be introduced.

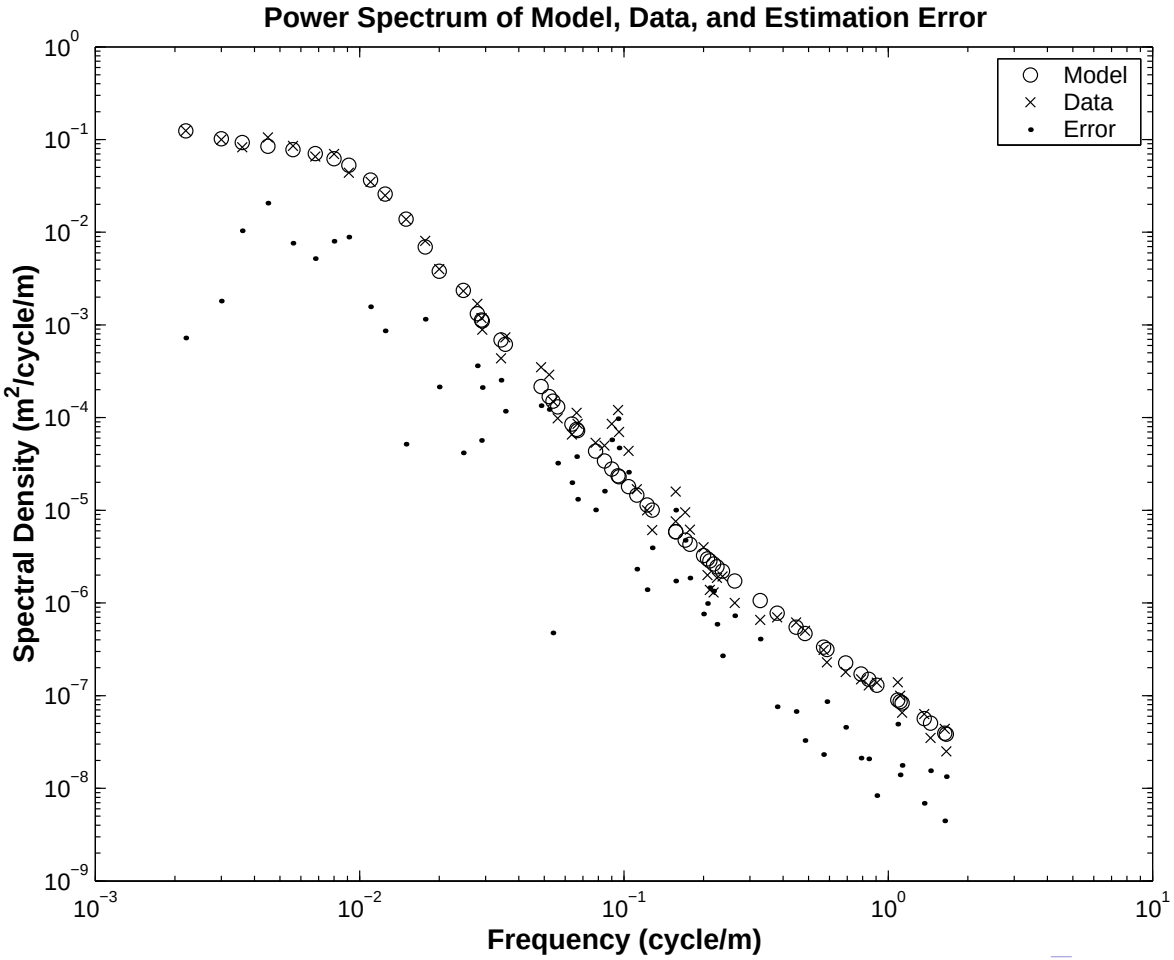
Next tried are the higher model orders with $\mathcal{R}$ as a design variable. In the figures, the output spectra and the estimation errors are compared with the road data for $n = 7$, $p = 32$, and the two cases $\mathcal{R} = I_N$ and $\mathcal{R}_k = S_k$.

- Clear improvement in the high frequency rolling by weighting.

The spectral data and its modeling by a rational model of order 7 produced by Algorithm 1 with $\mathcal{R} = I_N$.



The spectral data and its modeling by a rational model of order 7 produced by Algorithm 1 with $\mathcal{R}_k = I_N$.



- We presented a strongly consistent subspace algorithm for the identification of square multi-input/multi-output, discrete-time, linear-time invariant systems from nonuniformly spaced power spectrum measurements.
- The algorithm was illustrated with one practical example that solves a stochastic road modeling problem.