Modelling of Random Processes using Orthonormal Bases

Brett Ninness[a] and Håkan Hjalmarsson[b]

[a]Centre for Industrial Control Science
and
Department of Electrical and Computer Engineering, University of Newcastle, AUSTRALIA.

[b]Department of Sensors, Signals and Systems
Royal Institute of Technology, Stockholm, SWEDEN.

Problem Description

- Observed data $\{y_k\}$:

$$\{y_k\} \text{ Stationary and Regular.}$$

$$\mathbf{E}\{y_k\} = 0$$

- Wold Decomposition

$$y_k = e_k + \sum_{n=1}^{\infty} h_n e_{k-n} = \underbrace{\left[1 + \sum_{n=1}^{\infty} h_n q^{-n}\right]}_{H(q)} e_k = H(q) e_k$$

$$\{e_k\} \text{ white with } \mathbf{E}\{e_k\} = 0, \mathbf{E}\left\{e_k^2\right\} = \sigma^2.$$

- Associated Spectral density $\Phi_y(\omega)$

$$\Phi_y(\omega) = \sigma^2 |H(e^{j\omega})|^2 = \sum_{\tau=-\infty}^{\infty} R_y(\tau) e^{j\omega\tau}.$$

$$H(z) = \exp\left\{\frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{z + e^{j\omega}}{z - e^{j\omega}} \log \Phi_y(\omega)\, \mathrm{d}\omega\right\}$$

- Problem - Find estimate of $\Phi_y(\omega)$ from observed data $\{y_k\}$.

Model Based Solutions

- Truncated approximation to $H^{-1}(z)$:

$$H^{-1}(z) \approx 1 - G(z,\theta),$$
$$G(z,\theta) \triangleq \sum_{n=1}^{p-1} \theta_n z^{-n}.$$

- Estimate $\widehat{\theta} = [\widehat{\theta}_0, \cdots, \widehat{\theta}_{p-1}]^T$ from $\{y_k\}$.
- Estimate spectrum as

$$\Phi_y(\omega, \widehat{\theta}) = \frac{\widehat{\sigma}^2}{|1 - G(e^{j\omega}, \widehat{\theta})|^2}.$$

- Find $\widehat{\theta}$ via least squares:

$$\widehat{\theta} = \arg\min_{\theta} \left\{ \frac{1}{N} \sum_{k=0}^{N-1} \varepsilon_k^2(\theta) \right\}$$
$$\varepsilon_k(\theta) = y_k - G(q,\theta) y_k.$$

- Efficient Solution:
  - Levinson/Schur Recursions.
  - Szegö orthogonal polynomial recursions.

Department of Electrical & Computer Engineering
The University of Newcastle, Australia

Model Based Solution - Extensions

- Suppose $H(z)$ has zeros near unit circle $\mathbf{T}$.
- Need large p for good approximation:

$$H^{-1}(z) \stackrel{?}{\approx} 1 - \sum_{n=1}^{p-1} \theta_n z^{-n} = 1 - G(z,\theta).$$

- Replace $\{z^{-n}\}$ with $\{\mathcal{B}_n(z)\}$:

$$G(z,\theta) = \sum_{n=0}^{p-1} \theta_n \mathcal{B}_n(z).$$

  where poles of $\{\mathcal{B}_n(z)\}$ are arbitrary - put near zeros of $H(z)$.

- Preserve Orthonormality:

$$\langle \mathcal{B}_m, \mathcal{B}_n \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{B}_m(e^{j\omega})\overline{\mathcal{B}_n(e^{j\omega})}\, \mathrm{d}\omega = \delta(n-m).$$

  – Improved Normal Equation Numerics.

  – Facilitates Theoretical Analysis.

Choices of Orthonormal $\{\mathcal{B}_k\}$

- FIR:

$$\mathcal{B}_n(q) = q^{-n}$$

- Laguerre:

$$\mathcal{B}_n(q) = \frac{\sqrt{1-\xi^2}}{(q-\xi)}\left(\frac{1-q\xi}{q-\xi}\right)^n \qquad ; |\xi| < 1.$$

- Kautz:

$$\frac{\sqrt{(1-\alpha^2)(1-\gamma^2)}}{q^2-\alpha(\gamma+1)q+\gamma}\left(\frac{\gamma q^2-\alpha(\gamma+1)q+1}{q^2-\alpha(\gamma+1)q+\gamma}\right)^{(n-1)/2}$$

$$\frac{\sqrt{(1-\gamma^2)}(q-\alpha)}{q^2-\alpha(\gamma+1)q+\gamma}\left(\frac{\gamma q^2-\alpha(\gamma+1)q+1}{q^2-\alpha(\gamma+1)q+\gamma}\right)^{n/2}$$

- Dutch Construction:

$$\left.\begin{array}{rcl} x_{k+1} & = & Ax_k + Bu_k \\ y_k & = & Cx_k + Du_k \end{array}\right\} \begin{array}{c} R(q) \text{ is all-pass} \\ \text{and balanced.} \end{array}$$

$$V(q) = (qI - A)^{-1}B$$

$$\mathcal{B}_n(q) = V_k(q)R^m(q)$$

Dutch Solution - Heuberger, Van den Hof

$$\left.\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k + Du_k \end{aligned}\right\} \begin{array}{c} R(q) \text{ is all-pass} \\ \text{and balanced.} \end{array}$$

Put $\{u_k\}$ as white noise with $\mathbf{E}\left\{u_k u_k^T\right\} = I$.
Then by Parseval's Theorem

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} V(e^{j\omega})V^\star(e^{j\omega})\,\mathrm{d}\omega = \underbrace{\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mathbf{E}\left\{x_k x_k^T\right\}}_{\triangleq P = I}.$$

Szegö Polynomial Interpretation

Choose poles $\{\xi_k\}$ in desired denominator

$$d_n(z) = \prod_{k=0}^{n-1}(z-\xi_k).$$

Then use Gram-Schmidt to orthonormalise $\{1, z, z^2, \cdots, z^m\}$ with respect to inner product

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega)\overline{g(\omega)}\frac{\mathrm{d}\omega}{|d_n(e^{j\omega})|^2}$$

Szegö Polynomial Interpretation

- Resultant polynomials $\{n_0(z), n_1(z), \cdots, n_m(z)\}$ called Szegö polynomials (1959).

- Co-efficients given by order m one step ahead optimal predictor of random process with spectral density

$$\Phi(\omega) = \frac{1}{|d_n(e^{j\omega})|^2}$$

Levinson Recursions, Lattice Filters.

- Put

$$V(z) = \left[\frac{n_0(z)}{d_n(z)}, \frac{n_1(z)}{d_n(z)}, \cdots, \frac{n_m(z)}{d_n(z)}\right]$$

-

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} n_\ell(e^{j\omega})\overline{n_r(e^{j\omega})}\frac{\mathrm{d}\omega}{|d_n(e^{j\omega})|^2} = \delta(\ell - r)$$

General Construction of Orthonormal Bases

- Poles required at $\{\xi_0, \xi_1, \cdots, \xi_n\}$:

$$\begin{aligned}
\mathcal{B}_0(q) &= \sqrt{1-\xi_0^2}\,\frac{q}{q-\xi_0} \\
\mathcal{B}_1(q) &= \sqrt{1-\xi_1^2}\,\frac{q(1-q\overline{\xi_0})}{(q-\xi_0)(q-\xi_1)} \\
\mathcal{B}_2(q) &= \sqrt{1-\xi_2^2}\,\frac{q(1-q\overline{\xi_0})(1-q\overline{\xi_1})}{(q-\xi_0)(q-\xi_1)(q-\xi_2)} \\
\vdots \quad & \vdots \quad \vdots \qquad \vdots \\
\mathcal{B}_n(q) &= \sqrt{1-\xi_n^2}\,\frac{q}{(q-\xi_n)}\prod_{k=0}^{n-1}\left(\frac{1-q\overline{\xi_k}}{q-\xi_k}\right).
\end{aligned}$$

- Injection of prior knowledge through choice of a **variety** of pole positions $\{\xi_k\}$.
- Derivation is very simple.
- History: complex rational approximation - Malmquist (1926), Walsh (1935), Linear prediction of stationary time series - Wiener (1949), Network Synthesis - Kautz (1956).

General Construction

- Select poles $\{\xi_0, \cdots \xi_n\}$ in $\mathcal{B}_n(z)$.
- Put non-minimum phase zeroes in $\mathcal{B}_n(z)$ to cancel poles in $\{\mathcal{B}_0, \mathcal{B}_1, \cdots, \mathcal{B}_{n-1}\}$.
- Resulting General Formulation

$$\mathcal{B}_n(z) = \left(\frac{\sqrt{1-|\xi_n|^2}}{z-\xi_n}\right)\prod_{k=0}^{n-1}\left(\frac{1-\overline{\xi_k}z}{z-\xi_k}\right)$$

- Complex poles - include in conjugate pairs

$$\begin{pmatrix} \mathcal{B}'_n \\ \mathcal{B}''_n \end{pmatrix} = \begin{pmatrix} c_0 & c_1 \\ c'_0 & c'_1 \end{pmatrix}\begin{pmatrix} \mathcal{B}_n \\ \mathcal{B}_{n+1} \end{pmatrix}$$

  - $\mathcal{B}_n(z)$ has a complex pole at ξ_n.
  - $\mathcal{B}_{n+1}(z)$ has complex poles at ξ_n and $\overline{\xi_n}$.

$$\mathcal{B}'_n(z) = \frac{\sqrt{1-|\xi_n|^2}(\beta z + \mu)}{z^2 + (\xi_n + \overline{\xi_n})z + |\xi_n|^2}\prod_{k=0}^{n-1}\left(\frac{1-\overline{\xi_k}z}{z-\xi_k}\right)$$

$$\mathcal{B}''_n(z) = \frac{\sqrt{1-|\xi_n|^2}(\beta' z + \mu')}{z^2 + (\xi_n + \overline{\xi_n})z + |\xi_n|^2}\prod_{k=0}^{n-1}\left(\frac{1-\overline{\xi_k}z}{z-\xi_k}\right)$$

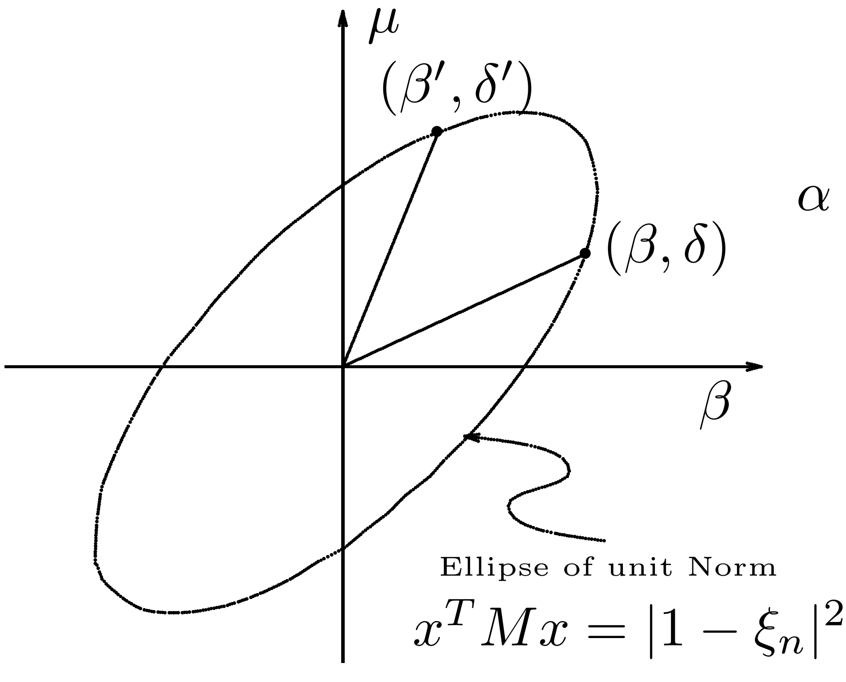
Choice of Numerator

Unit Norm constraint:r

$$(\beta,\mu)\underbrace{\begin{pmatrix} 1+|\xi_n|^2 & 2\mathrm{Re}\{\xi_n\} \\ 2\mathrm{Re}\{\xi_n\} & 1+|\xi_n|^2 \end{pmatrix}}_{M}\begin{pmatrix}\beta\\ \mu\end{pmatrix} = |1-\xi_n^2|^2$$

Orthogonality constraint - rotate by 90^o in normalised eigenspace of M:

$$\begin{pmatrix}\beta'\\ \mu'\end{pmatrix} = \frac{1}{\sqrt{1-\alpha^2}}\begin{pmatrix}\alpha & 1\\ -1 & -\alpha\end{pmatrix}\begin{pmatrix}\beta\\ \mu\end{pmatrix}$$



$$\alpha \triangleq \frac{\xi_n + \overline{\xi_n}}{1+|\xi_n|^2}$$

Department of Electrical & Computer Engineering
The University of Newcastle, Australia

Special Cases

- FIR: $\xi_k = 0$.
- Laguerre: $\xi_k = \xi$.
- Legendre:

$$\xi_k = \frac{2 - \alpha(2k+1)}{2 + \alpha(2k+1)}.$$

  Formed from classical $L_2([0,1])$ orthogonal Legendre polynomials.

- Kautz: Choose $\{\xi_k\}$ in complex conjugate pairs.
- Dutch 'Balanced Realisation' Construction: Repeat fixed set of poles $\{\xi_0, \cdots, \xi_m\}$ in cyclical manner.
- Can mix FIR, Laguerre, Legendre, Kautz or others as appropriate to prior knowledge.

Completeness

Completeness depends on how quickly $\{\xi_k\}$ approaches the stability boundary.

- $$\overline{\mathrm{Span}\{\mathcal{B}_k\}} = H_2(\mathbf{T}) \iff \sum_{k=0}^{\infty}(1 - |\xi_k|) = \infty$$

- $$\overline{\mathrm{Span}\{\mathcal{B}_k\}} = A(\mathbf{D}_\rho) \Leftarrow \sum_{k=0}^{\infty}(1 - |\xi_k|) = \infty$$

Christoffel-Darboux formula

Define the Blaschke product

$$\varphi_p(z) = \prod_{k=0}^{p-1} \frac{1 - \overline{\xi_k} z}{z - \xi_k}.$$

Then for $|\mu| < 1, |z| < 1$ the Reproducing Kernel of the orthonormal basis $\{\mathcal{B}_k\}$ is given by

$$\sum_{k=0}^{p-1} \overline{\mathcal{B}_k(\mu^{-1})}\mathcal{B}_k(z^{-1}) = \left\{ \frac{1 - \overline{\varphi_p(\mu^{-1})}\varphi_p(z^{-1})}{1 - \overline{\mu} z} \right\} \overline{\mu} z.$$

Numerical Properties

- Fixed Denominator Model Structure

$$\widehat{y}_{k|k-1}(\theta) = \left(\sum_{k=0}^{p-1} \theta_k \mathcal{B}_k(q)\right) u_k = \phi_k^T \theta$$

$$\phi_k^T = [\mathcal{B}_0(q)u_k, \cdots, \mathcal{B}_{p-1}(q)u_k]$$

$$\widehat{\theta} = \left(\underbrace{\frac{1}{N}\sum_{k=0}^{N-1} \phi_k \phi_k^T}_{R_p(N)} \right)^{-1} \frac{1}{N}\sum_{k=0}^{N-1} \phi_k y_k$$

- Numerical properties dependent on spectrum:

$$R_p(N) \xrightarrow{\text{a.s.}} R_\phi = \mathbf{E}\left\{\phi_k \phi_k^T\right\} \quad \text{as } N \to \infty.$$

$$\min_{\omega\in[-\pi,\pi]} \Phi_y(\omega) \leq \lambda(R_p) \leq \max_{\omega\in[-\pi,\pi]} \Phi_y(\omega).$$

- Define $R = \lim_{p\to\infty} R_p$. Then

$$\lambda(R) = \text{Range}\{\Phi_y(\omega)\}$$

Undermodelling Error

- Christoffel-Darboux formula is key - no isomorphism with FIR case possible since $\mathcal{B}_n\mathcal{B}_m \neq \mathcal{B}_{n+m}$. CF: $e^{j\omega n}e^{j\omega m} = e^{j\omega(m+n)}$.

- Suppose $G(z)$ has partial fraction expansion

$$G(z) = \sum_{i=0}^{r-1} \frac{\alpha_i}{z - \gamma_i}$$

where all the poles satisfy $|\gamma_i| < 1$. Put $\widehat{G}_p(z)$ as the best H_2 approximation to $G(z)$ with respect to the p basis functions $\{\mathcal{B}_0, \mathcal{B}_1, \cdots, \mathcal{B}_{p-1}\}$

$$\widehat{G}_p(z) = \sum_{k=0}^{p-1} \langle G, \mathcal{B}_k\rangle \, \mathcal{B}_k(z).$$

Then

$$|G(e^{j\omega}) - \widehat{G}_p(e^{j\omega})| < \sum_{i=0}^{r-1} \left| \frac{\alpha_i}{e^{j\omega} - \gamma_i} \right| \prod_{k=0}^{p-1} \left| \frac{\gamma_i - \xi_k}{1 - \overline{\xi_k}\gamma_i} \right|$$

- Convergence rapid for good choice of $\{\xi_k\}$.

Asymptotic Bias

$$\hat{\theta}_N \xrightarrow{a.s.} \theta_0 = \arg\min_{\theta} \left\{ \int_{-\pi}^{\pi} \Phi_\varepsilon(\omega,\theta)\,\mathrm{d}\omega \right\}.$$

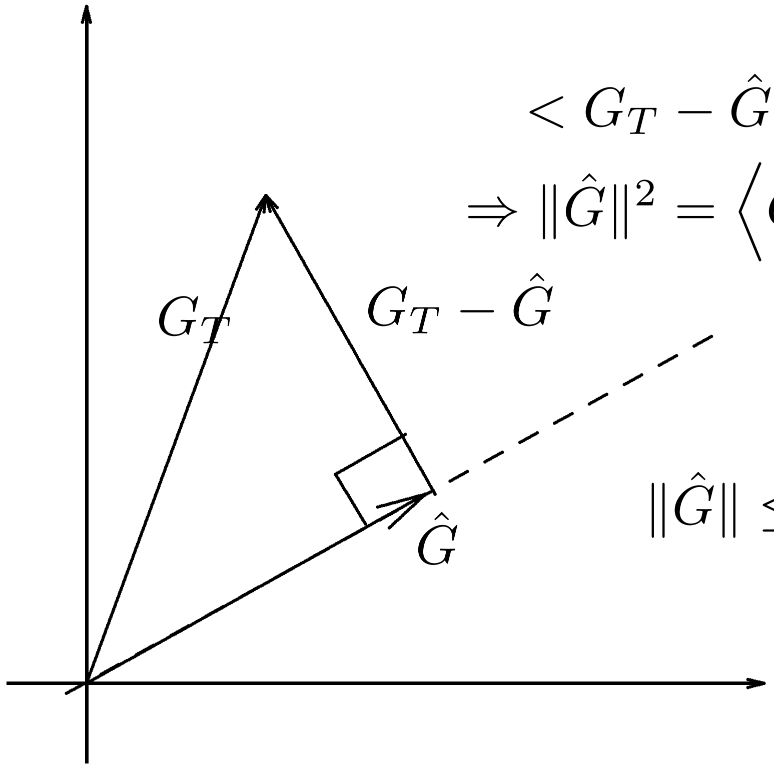
$$\Phi_\varepsilon(\omega,\theta) = \left| G(e^{j\omega}) - \sum_{k=0}^{p-1} \theta_k \mathcal{B}_k(e^{j\omega}) \right|^2 \Phi_y(\omega)$$

Projection Theorem$\Rightarrow$

Error orthogonal to subspace spanned by the $\{\mathcal{B}_k\}$.

$$< G_T - \hat{G}, \hat{G} >= 0$$

$$\Rightarrow \|\hat{G}\|^2 = \left\langle G_T, \hat{G} \right\rangle \underbrace{\leq \|G_T\|^2}_{\text{Cauchy-Schwarz}}$$



$$\|\hat{G}\| \leq \|G_T\|$$

$$\int_{-\pi}^{\pi} |G(e^{j\omega},\theta_0)|^2 \Phi_y(\omega)\,\mathrm{d}\omega \leq \int_{-\pi}^{\pi} |G(e^{j\omega})|^2 \Phi_y(\omega)\,\mathrm{d}\omega$$

Asymptotic Distribution - Main Result

Define

$$\gamma_p(\omega) = \sum_{k=0}^{p-1} |\mathcal{B}_k(e^{j\omega})|^2.$$

Then

$$\begin{bmatrix} \sqrt{\frac{N}{\gamma_p(\omega_1)}} & 0 \\ 0 & \sqrt{\frac{N}{\gamma_p(\omega_2)}} \end{bmatrix} \begin{bmatrix} \Phi_y(e^{j\omega_1}, \widehat{\theta}) - \Phi_y(e^{j\omega_1}, \theta_0) \\ \Phi_y(e^{j\omega_2}, \widehat{\theta}) - \Phi_y(e^{j\omega_2}, \theta_0) \end{bmatrix}$$

$$\xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_p(\omega_1, \omega_2)) \quad \text{as } N \to \infty$$

where for $\omega_1, \omega_2 \neq 0, \pi$

$$\lim_{p\to\infty} \Sigma_p(\omega_1, \omega_2) = 2 \begin{bmatrix} \Phi_y^2(\omega_1) & 0 \\ 0 & \Phi_y^2(\omega_2) \end{bmatrix}.$$

while for $\omega_1, \omega_2 = 0, \pi$

$$\lim_{p\to\infty} \Sigma_p(\omega_1, \omega_2) = 4 \begin{bmatrix} \Phi_y^2(\omega_1) & 0 \\ 0 & \Phi_y^2(\omega_2) \end{bmatrix}.$$

Interpretation

$$\mathrm{Var}\left\{\Phi_y(e^{j\omega},\widehat{\theta})\right\} \approx \frac{2}{N}\Phi_y^2(\omega)\sum_{k=0}^{p-1}|\mathcal{B}_k(e^{j\omega})|^2.$$

This unifies previously known results:

- FIR (Berk). Set $\xi_k = 0$:

$$\mathrm{Var}\left\{\Phi_y(e^{j\omega},\widehat{\theta})\right\} \approx \frac{2p}{N}\Phi_y^2(\omega)$$

- Laguerre (Wahlberg and Hannan). Set $\xi_k = \xi \in \mathbf{R}$:

$$\mathrm{Var}\left\{\Phi_y(e^{j\omega},\widehat{\theta})\right\} \approx \frac{2p}{N}\Phi_y^2(\omega)\frac{1-|\xi|^2}{|e^{j\omega}-\xi|^2}$$

Fundamental Properties and Limitations

- Orthonormal Bases as an **analysis** tool rather than an **implementation** tool.
- General orthonormal bases facilitate analysis of general fixed denominator model structure:

$$G(q,\beta) = \sum_{k=0}^{p-1} \beta_k q^k D_p^{-1}(q),$$

$$D_p(q) \triangleq \prod_{\ell=0}^{p-1} (q-\xi_\ell),$$

- Linearity $\Rightarrow$ estimates for general structure and orthonormal basis structure are identical:

$$G(q,\widehat{\theta}) = G(q,\widehat{\beta}).$$

- Orthonormal basis structure more tractable yet has identical properties. In particular:

$$\text{Var}\{\Phi_y(e^{j\omega},\widehat{\theta})\} \approx \frac{2}{N}\Phi_y^2(\omega)\sum_{k=0}^{p-1}|\mathcal{B}_k(e^{j\omega})|^2$$

Fundamental Properties and Limitations

- Fundamental role of Orthormal bases: They parameterise bias and variability for **any** fixed denominator structure.
  - Vital to have weak restrictions on choice of poles.
  - Eg - analysis with Laguerre pertains only to structures with one fixed pole.
- Fundamental limit of estimation accuracy.
  - Bias/Variance in choice of model order (well known).
  - Bias/Variance in choice of pole position (new result).

$$\text{Var} \propto \sum_{k=0}^{p-1} |\mathcal{B}_k(e^{j\omega})|^2, \quad \text{Bias} \propto \prod_{k=0}^{p-1} \left| \frac{\gamma_\ell^{ij} - \xi_k}{1 - \overline{\xi_k}\gamma_\ell^{ij}} \right|.$$

Simulation Example

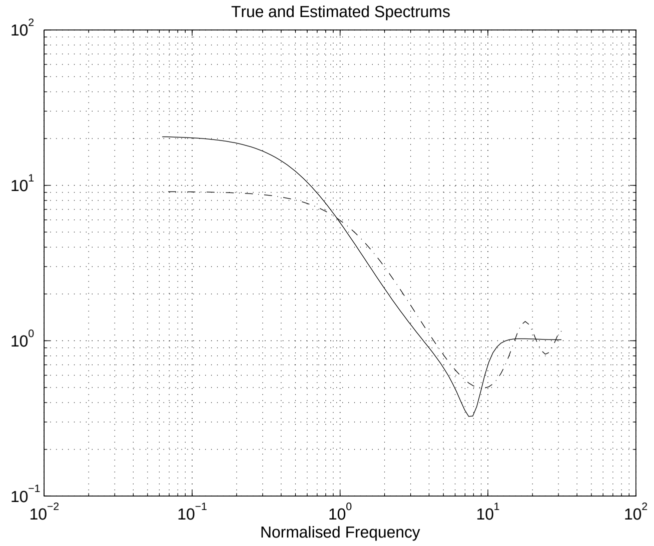
- $$H(z) = \frac{z^3 - 1.9235z^2 + 1.5910z - 0.5203}{z^3 - 1.9464z^2 + 1.5155z - 0.5368}.$$

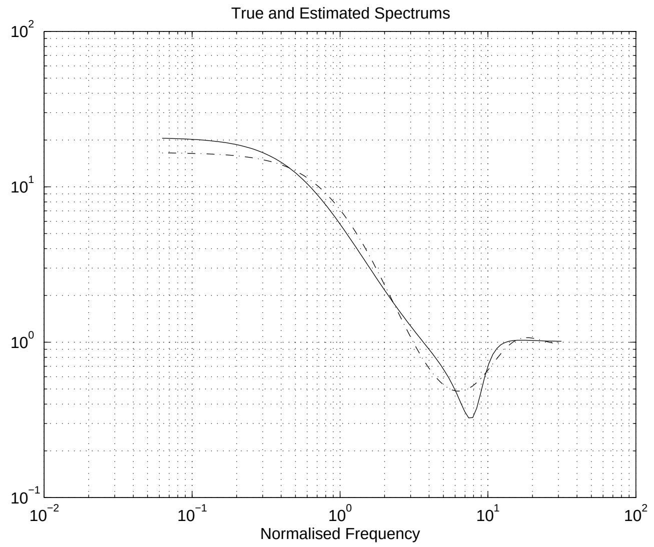
- Zeroes of $H(z)$ at

$$z = 0.7165, \quad z = 0.852e^{\pm j0.784}.$$

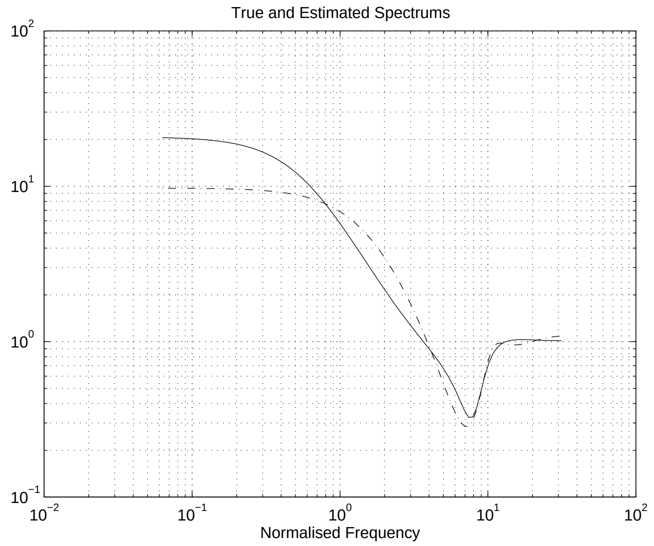
- $N = 1000$ samples of $\{y_k\}$ observed.

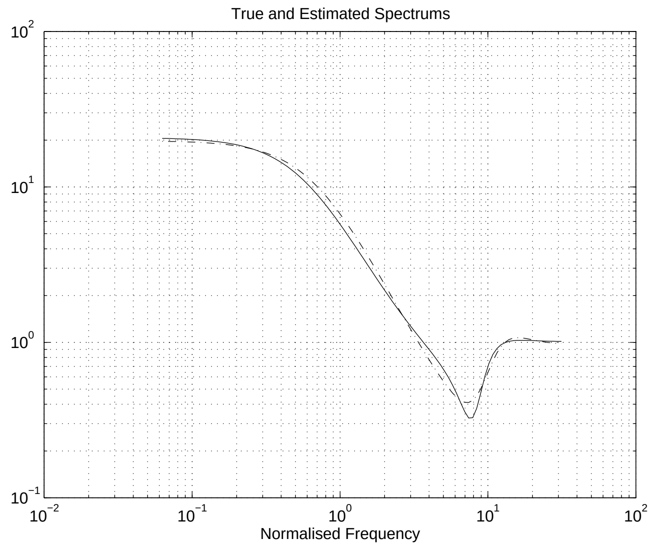
- Four different sets of pole choices:
  - FIR: $\{\xi_0, \cdots, \xi_3\} = \{0, 0, 0, 0\}$
  - Laguerre: $\{\xi_0, \cdots, \xi_3\} = \{0.5, 0.5, 0.5, 0.5\}$
  - Kautz:
    $\{\xi_0, \cdots, \xi_3\} = \{0.5 \pm j0.5, 0.5 \pm j0.5\}$
  - General:
    $\{\xi_0, \cdots, \xi_3\} = \{0.5, 0.5, 0.5 \pm j0.5\}$



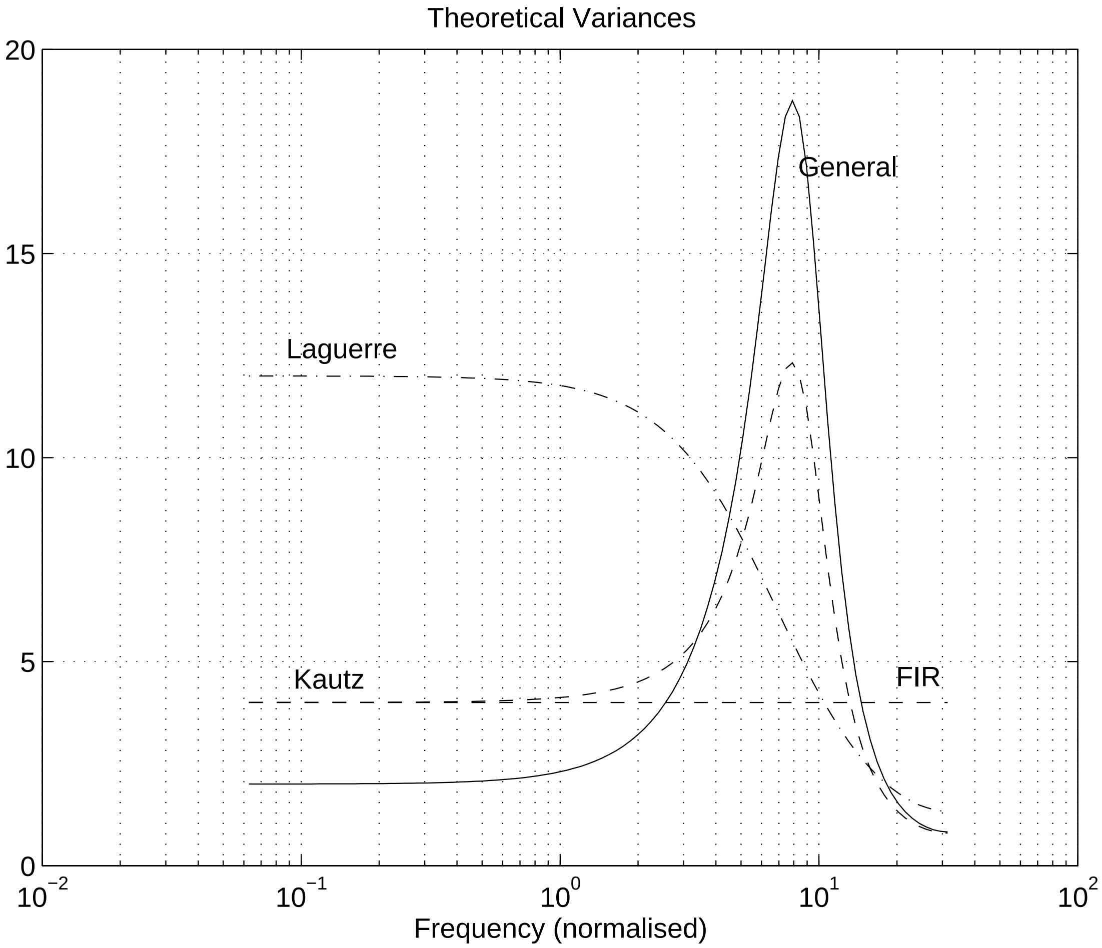


FIR and Laguerre Basis.





Kautz and General Basis.



Variability vs frequency