A geometric approach to variance analysis in system identification

Jonas Mårtensson and Håkan Hjalmarsson

School of Electrical Engineering, KTH, Stockholm

16th ERNSI Workshop, 1-3 October 2007

Outline

Outline

Intro: *The* Asymptotic Covariance Matrix

We all know that under certain conditions

$$\sqrt{N}\left(\hat{\theta}_N - \theta_o\right) \in \mathrm{As}\mathcal{N}\left(0, P\right)$$

where P also is the *asymptotic covariance matrix*

$$P = \mathrm{AsCov}\,\hat{\theta}_N \triangleq \lim_{N\to\infty} N \cdot \mathbf{E}\left[(\hat{\theta}_N - \mathbf{E}\hat{\theta}_N)^T \overline{(\hat{\theta}_N - \mathbf{E}\hat{\theta}_N)}\right]$$

(we assume vectors to be row vectors.)

Intro: Pandora's (Black) Box

An enormous amount of model error information hidden in P:

Intro: Pandora's (Black) Box

An enormous amount of model error information hidden in P:

- Structural
  - Model structure (ARX, ARMAX, Box-Jenkins, non-linear,...)
  - Model order
  - Open vs closed loop
  - Input channels and input excitation
  - Sensor channels
  - Noise

Intro: Pandora's (Black) Box

An enormous amount of model error information hidden in P:

- Structural
  - Model structure (ARX, ARMAX, Box-Jenkins, non-linear,...)
  - Model order
  - Open vs closed loop
  - Input channels and input excitation
  - Sensor channels
  - Noise
- System properties
  - Frequency function
  - Impulse response coefficients
  - Gains
  - Poles and zeros
  - Control applications

Intro: Pandora's (Black) Box

An enormous amount of model error information hidden in P:

- Structural
  - Model structure (ARX, ARMAX, Box-Jenkins, non-linear,...)
  - Model order
  - Open vs closed loop
  - Input channels and input excitation
  - Sensor channels
  - Noise
- System properties
  - Frequency function
  - Impulse response coefficients
  - Gains
  - Poles and zeros
  - Control applications

Many have contributed to revealing the secrets of P!

Intro: This talk

Yet another attempt to decipher P,

Intro: This talk

Yet another attempt to decipher P,

but through a backdoor:

- Geometrical interpretation

Outline

The structure of P

A non-singlular P can always be written as

$$P = [\langle \Psi, \Psi \rangle]^{-1}$$

The structure of P

A non-singlular P can always be written as

$$P = [\langle \Psi, \Psi \rangle]^{-1}$$

where $\Psi : \mathbb{C} \to \mathbb{C}^{n \times m}$, for some integer $m > 0$ depending on the model structure,

The structure of P

A non-singlular P can always be written as

$$P = [\langle \Psi, \Psi \rangle]^{-1}$$

where $\Psi : \mathbb{C} \to \mathbb{C}^{n \times m}$, for some integer $m > 0$ depending on the model structure,
and where

$$\langle \Psi, \Psi \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Psi(e^{j\omega}) \Psi^*(e^{j\omega}) \mathrm{d}\omega$$

(inner products between the rows of Ψ)

The structure of P: Example 1 - FIR models

$$y_t = \sum_{k=1}^{n} \theta_k \, u_{t-k} + e_t = \theta \varphi_t^T + e_t$$

The structure of P: Example 1 - FIR models

$$y_t = \sum_{k=1}^{n} \theta_k \, u_{t-k} + e_t = \theta \varphi_t^T + e_t$$

$$P = \lambda_o \left[\mathbf{E} \left[\varphi_t^T \varphi_t \right] \right]^{-1}$$

The structure of P: Example 1 - FIR models

$$y_t = \sum_{k=1}^{n} \theta_k \, u_{t-k} + e_t = \theta \varphi_t^T + e_t$$

$$P = \lambda_o \left[\mathbf{E} \left[\varphi_t^T \varphi_t \right] \right]^{-1}$$

$$\mathbf{E} \left[\varphi_t^T \varphi_t \right] = \begin{bmatrix} r_0 & r_1 & \dots & r_{n-1} \\ r_1 & r_0 & \dots & r_{n-2} \\ \vdots & & & \vdots \\ r_{n-1} & r_{n-2} & \dots & r_0 \end{bmatrix}$$

The structure of P: Example 1 - FIR models

$$y_t = \sum_{k=1}^{n} \theta_k \; u_{t-k} + e_t = \theta \varphi_t^T + e_t$$

$$P = \lambda_o \left[\mathbf{E} \left[\varphi_t^T \varphi_t \right] \right]^{-1}$$

$$\mathbf{E} \left[\varphi_t^T \varphi_t \right] = \begin{bmatrix} r_0 & r_1 & \dots & r_{n-1} \\ r_1 & r_0 & \dots & r_{n-2} \\ \vdots & & & \vdots \\ r_{n-1} & r_{n-2} & \dots & r_0 \end{bmatrix}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_n(e^{j\omega}) \Phi_u(\omega) \Gamma_n^*(e^{j\omega})$$

where $\Gamma_n(q) = \begin{bmatrix} q^{-1} & \dots & q^{-n} \end{bmatrix}^T$

The structure of P: Example 1 - FIR models

$$y_t = \sum_{k=1}^{n} \theta_k \, u_{t-k} + e_t = \theta \varphi_t^T + e_t$$

$$P = \lambda_o \left[\mathbf{E} \left[\varphi_t^T \varphi_t \right] \right]^{-1}$$

$$\mathbf{E} \left[\varphi_t^T \varphi_t \right] = \begin{bmatrix} r_0 & r_1 & \dots & r_{n-1} \\ r_1 & r_0 & \dots & r_{n-2} \\ \vdots & & & \vdots \\ r_{n-1} & r_{n-2} & \dots & r_0 \end{bmatrix}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_n(e^{j\omega}) \Phi_u(\omega) \Gamma_n^*(e^{j\omega})$$

where $\Gamma_n(q) = \begin{bmatrix} q^{-1} & \dots & q^{-n} \end{bmatrix}^T$

$$\Rightarrow \; P = \langle \Gamma_n \Phi_u^{1/2} / \sqrt{\lambda_o}, \Gamma_n \Phi_u^{1/2} / \sqrt{\lambda_o} \rangle = \langle \Psi, \Psi \rangle$$

The structure of P: Example 2 - NFIR models

$$y_t = \sum_{k=1}^{n} \theta_k \, u_{t-k} + \sum_{k=n+1}^{m} \theta_k \, (u_{t-(k-n)})^2 + e_t$$

The structure of P: Example 2 - NFIR models

$$y_t = \sum_{k=1}^{n} \theta_k \, u_{t-k} + \sum_{k=n+1}^{m} \theta_k \, \left(u_{t-(k-n)}\right)^2 + e_t = \theta \varphi_t^T + e_t$$

The structure of P: Example 2 - NFIR models

$$y_t = \sum_{k=1}^{n} \theta_k \, u_{t-k} + \sum_{k=n+1}^{m} \theta_k \, (u_{t-(k-n)})^2 + e_t = \theta \varphi_t^T + e_t$$

$$\begin{aligned} \varphi_t &= [u_{t-1}, \ldots, u_{t-n}, (u_{t-1})^2, \ldots, (u_{t-m})^2] \\ &= (M(q) z_t)^T \end{aligned}$$

where

$$M(q) = \begin{bmatrix} \Gamma_n(q) & 0 \\ 0 & \Gamma_m(q) \end{bmatrix}, \quad z_t = \begin{bmatrix} u_t \\ (u_t)^2 \end{bmatrix}$$

The structure of P: Example 2 - NFIR models

$$y_t = \sum_{k=1}^{n} \theta_k \, u_{t-k} + \sum_{k=n+1}^{m} \theta_k \, (u_{t-(k-n)})^2 + e_t = \theta \varphi_t^T + e_t$$

$$\begin{aligned} \varphi_t &= [u_{t-1}, \ldots, u_{t-n}, (u_{t-1})^2, \ldots, (u_{t-m})^2] \\ &= (M(q) z_t)^T \end{aligned}$$

where

$$M(q) = \begin{bmatrix} \Gamma_n(q) & 0 \\ 0 & \Gamma_m(q) \end{bmatrix}, \quad z_t = \begin{bmatrix} u_t \\ (u_t)^2 \end{bmatrix}$$

$$\Rightarrow \quad \Psi = \frac{1}{\sqrt{\lambda_o}} M \Phi_z^{1/2}.$$

where $\Phi_z^{1/2}$ is a Cholesky factor of Φ_z, the spectrum of z.

The structure of P: Other quantities

Typically not the parameters θ directly we are interested in, but rather some function

$J(\theta) : \mathbb{R}^{1\times n} \longrightarrow \mathbb{C}^{1\times q}$:

The structure of P: Other quantities

Typically not the parameters θ directly we are interested in, but rather some function

$J(\theta) : \mathbb{R}^{1\times n} \to \mathbb{C}^{1\times q}$:

- frequency function
- impulse response
- system gain

The structure of P: Other quantities

Typically not the parameters θ directly we are interested in, but rather some function

$J(\theta) : \mathbb{R}^{1\times n} \to \mathbb{C}^{1\times q}$:

- frequency function
- impulse response
- system gain

$$\text{AsCov } J(\hat{\theta}_N) = \Lambda^T \left[\langle \Psi, \Psi \rangle\right]^{-1} \overline{\Lambda}$$

where

$$\Lambda \triangleq J'(\theta_o) \in \mathbb{C}^{n\times q}$$

Outline

A refresher on orthogonal projection: Scalar case

A refresher on orthogonal projection: Scalar case

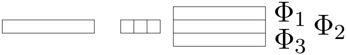
Least squares estimation

$$Y = \theta\Phi + E$$

A refresher on orthogonal projection: Scalar case

Least squares estimation

$$Y = \theta\Phi + E$$

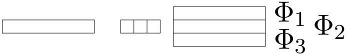


(Y and θ are row vectors)

A refresher on orthogonal projection: Scalar case

Least squares estimation

$$Y = \theta \Phi + E$$



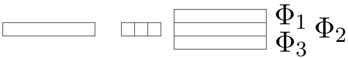
(Y and θ are row vectors)

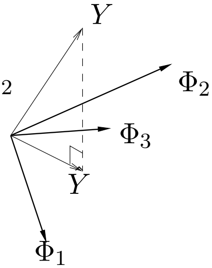
$$\hat{Y} = Y \Phi^T \left[\Phi \Phi^T \right]^{-1} \Phi$$

A refresher on orthogonal projection: Scalar case

Least squares estimation

$$Y = \theta\Phi + E$$





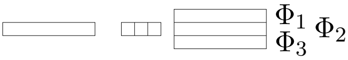
(Y and θ are row vectors)

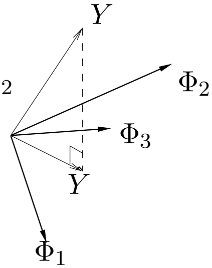
$$\hat{Y} = Y\Phi^T \left[\Phi\Phi^T\right]^{-1} \Phi$$

A refresher on orthogonal projection: Scalar case

Least squares estimation

$$Y = \theta\Phi + E$$





(Y and θ are row vectors)

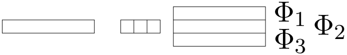
$$\hat{Y} = Y\Phi^T \left[\Phi\Phi^T\right]^{-1} \Phi$$

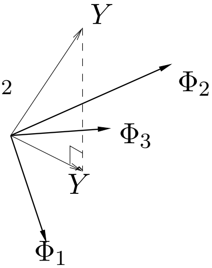
General case:
Let the rows of $X = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}^T$ span a (closed) subspace $\mathcal{S}_X$ to a Hilbert space $\mathcal{H}$.

A refresher on orthogonal projection: Scalar case

Least squares estimation

$$Y = \theta\Phi + E$$





(Y and θ are row vectors)

$$\hat{Y} = Y\Phi^T \left[\Phi\Phi^T\right]^{-1} \Phi$$

General case:
Let the rows of $X = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}^T$ span a (closed) subspace $\mathcal{S}_X$ to a Hilbert space $\mathcal{H}$.
Then the projection of $f \in \mathcal{H}$ on $\mathcal{S}_X$ is given by

$$\hat{f} = \langle f, X \rangle \left[\langle X, X \rangle\right]^{-1} X$$

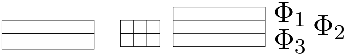
A refresher on orthogonal projection: MV-case

Multi-output least squares estimation

$$Y = \theta\Phi + E$$

Multi-output least squares estimation

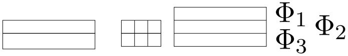
$$Y = \theta\Phi + E$$



A refresher on orthogonal projection: MV-case

Multi-output least squares estimation

$$Y = \theta\Phi + E$$

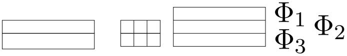


$$\hat{Y} = Y\Phi^T \left[\Phi\Phi^T\right]^{-1} \Phi$$

A refresher on orthogonal projection: MV-case

Multi-output least squares estimation

$$Y = \theta\Phi + E$$



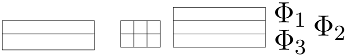
$$\hat{Y} = Y\Phi^T \left[\Phi\Phi^T\right]^{-1} \Phi$$

Same equation as before. Each row of Y projected on the row space of Φ.

A refresher on orthogonal projection: MV-case

Multi-output least squares estimation

$$Y = \theta\Phi + E$$



$$\hat{Y} = Y\Phi^T \left[\Phi\Phi^T\right]^{-1} \Phi$$

Same equation as before. Each row of Y projected on the row space of Φ.

General case:
The projection of $f_i \in \mathcal{H}$, $i = 1, \ldots, n$ on $\mathcal{S}_X$ is given by the rows of

$$\hat{f} = \langle f, X \rangle \left[\langle X, X \rangle\right]^{-1} X$$

where $f = \begin{bmatrix} f_1 & f_2 & \ldots & f_n \end{bmatrix}^T$

A refresher on orthogonal projection: The norm

$$\hat{f} = \langle f, X \rangle \left[\langle X, X \rangle \right]^{-1} X$$

A refresher on orthogonal projection: The norm

$$\hat{f} = \langle f, X \rangle \left[\langle X, X \rangle \right]^{-1} X$$

The “norm” of the projection:

$$\langle \hat{f}, \hat{f} \rangle = \langle f, X \rangle \left[\langle X, X \rangle \right]^{-1} \langle X, f \rangle$$

A refresher on orthogonal projection

$$\langle \hat{f}, \hat{f} \rangle = \langle f, X \rangle \left[\langle X, X \rangle \right]^{-1} \langle X, f \rangle$$

A refresher on orthogonal projection

$$\langle \hat{f}, \hat{f} \rangle = \langle f, X \rangle \left[\langle X, X \rangle \right]^{-1} \langle X, f \rangle$$

Hmmm, didn’t I just see something similar a minute ago????

A refresher on orthogonal projection

$$\langle \hat{f}, \hat{f} \rangle = \langle f, X \rangle \left[\langle X, X \rangle \right]^{-1} \langle X, f \rangle$$

Hmmm, didn't I just see something similar a minute ago????

$$\text{AsCov } J(\hat{\theta}_N) = \Lambda^T \left[\langle \Psi, \Psi \rangle \right]^{-1} \overline{\Lambda}$$

A refresher on orthogonal projection

$$\langle \hat{f}, \hat{f} \rangle = \langle f, X \rangle \left[\langle X, X \rangle \right]^{-1} \langle X, f \rangle$$

Hmmm, didn't I just see something similar a minute ago????

$$\mathrm{AsCov}\ J(\hat{\theta}_N) = \Lambda^T \left[\langle \Psi, \Psi \rangle \right]^{-1} \overline{\Lambda}$$

What if $\Lambda = \langle \Psi, \gamma \rangle$ for some function γ?

A refresher on orthogonal projection

$$\langle \hat{f}, \hat{f} \rangle = \langle f, X \rangle \left[\langle X, X \rangle \right]^{-1} \langle X, f \rangle$$

Hmmm, didn't I just see something similar a minute ago????

$$\text{AsCov } J(\hat{\theta}_N) = \Lambda^T \left[\langle \Psi, \Psi \rangle \right]^{-1} \overline{\Lambda}$$

What if $\Lambda = \langle \Psi, \gamma \rangle$ for some function γ?

$$\text{AsCov } J(\hat{\theta}_N) = \langle \hat{\gamma}, \hat{\gamma} \rangle = \langle \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\}, \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\} \rangle$$

A refresher on orthogonal projection

$$\langle \hat{f}, \hat{f} \rangle = \langle f, X \rangle \left[\langle X, X \rangle \right]^{-1} \langle X, f \rangle$$

Hmmm, didn't I just see something similar a minute ago????

$$\text{AsCov } J(\hat{\theta}_N) = \Lambda^T \left[\langle \Psi, \Psi \rangle \right]^{-1} \overline{\Lambda}$$

What if $\Lambda = \langle \Psi, \gamma \rangle$ for some function γ?

$$\text{AsCov } J(\hat{\theta}_N) = \langle \hat{\gamma}, \hat{\gamma} \rangle = \langle \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\}, \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\} \rangle$$

Take home message

Scalar quantities: The asymptotic variance is the squared norm of γ projected onto the subspace spanned by the rows of Ψ.

Outline

A geometric interpretation of P

$$\text{AsCov } J(\hat{\theta}_N) = \langle \hat{\gamma}, \hat{\gamma} \rangle = \langle \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\}, \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\} \rangle$$

Scalar quantities: The asymptotic variance is the squared norm of γ projected onto the subspace spanned by the rows of Ψ.

Why is this useful?

A geometric interpretation of P

$$\text{AsCov } J(\hat{\theta}_N) = \langle \hat{\gamma}, \hat{\gamma} \rangle = \langle \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\}, \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\} \rangle$$

Scalar quantities: The asymptotic variance is the squared norm of γ projected onto the subspace spanned by the rows of Ψ.

Why is this useful?

- Often γ can be chosen so that it only depends on the quantity $J(\theta)$ of interest

A geometric interpretation of P

$$\text{AsCov } J(\hat{\theta}_N) = \langle \hat{\gamma}, \hat{\gamma} \rangle = \langle \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\}, \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\} \rangle$$

Scalar quantities: The asymptotic variance is the squared norm of γ projected onto the subspace spanned by the rows of Ψ.

Why is this useful?

- Often γ can be chosen so that it only depends on the quantity $J(\theta)$ of interest
- The influence on Ψ from model structure, model order, experimental conditions often simple to establish.

A geometric interpretation of P

$$\text{AsCov } J(\hat{\theta}_N) = \langle \hat{\gamma}, \hat{\gamma} \rangle = \langle \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\}, \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\} \rangle$$

Scalar quantities: The asymptotic variance is the squared norm of γ projected onto the subspace spanned by the rows of Ψ.

Why is this useful?

- Often γ can be chosen so that it only depends on the quantity $J(\theta)$ of interest
- The influence on Ψ from model structure, model order, experimental conditions often simple to establish.

Decoupling!

A geometric interpretation of P

$$\text{AsCov}\ J(\hat{\theta}_N) = \langle \hat{\gamma}, \hat{\gamma} \rangle = \langle \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\}, \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\} \rangle$$

Scalar quantities: The asymptotic variance is the squared norm of γ projected onto the subspace spanned by the rows of Ψ.

Why is this useful?

- Often γ can be chosen so that it only depends on the quantity $J(\theta)$ of interest
- The influence on Ψ from model structure, model order, experimental conditions often simple to establish.

Decoupling!

The main contribution in this talk: A methodology to analyze the asymptotic covariance matrix.

Examples will be used to illustrate this.

Outline

Example: FIR system

$$
\begin{aligned}
y_t &= \sum_{k=1}^{n} \theta_k \, u_{t-k} + e_t = \theta \varphi_t^T + e_t \\
P &= \left[\langle \Psi, \Psi \rangle\right]^{-1} \\
\Psi &= \frac{1}{\sqrt{\lambda_o}} \, \Gamma_n \Phi_u^{1/2}
\end{aligned}
$$

Structural results: Adding parameters

Example: FIR system

$$y_t = \sum_{k=1}^{n} \theta_k \, u_{t-k} + e_t = \theta \varphi_t^T + e_t$$

$$P = \left[\langle \Psi, \Psi \rangle\right]^{-1}$$

$$\Psi = \frac{1}{\sqrt{\lambda_o}} \, \Gamma_n \Phi_u^{1/2}$$

True order $n = n_o$, $J = J(\theta_1, \ldots, \theta_{n_o})$.

Structural results: Adding parameters

Example: FIR system

$$y_t = \sum_{k=1}^{n} \theta_k \, u_{t-k} + e_t = \theta \varphi_t^T + e_t$$
$$P = [\langle \Psi, \Psi \rangle]^{-1}$$
$$\Psi = \frac{1}{\sqrt{\lambda_o}} \, \Gamma_n \Phi_u^{1/2}$$

True order $n = n_o$, $J = J(\theta_1, \ldots, \theta_{n_o})$.
What happens with the accuracy if we over-model ($n > n_o$)?

Adding parameters: FIR example

$$\text{AsCov } J(\hat{\theta}_N) = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}^* \left[\langle \Omega, \Omega \rangle\right]^{-1} \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$$

Adding parameters: FIR example

$$\text{AsCov } J(\hat{\theta}_N) = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}^* [\langle \Omega, \Omega \rangle]^{-1} \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$$

where

$$\Omega = \Gamma_n \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} = \begin{bmatrix} \Gamma_{n_o} \\ \Gamma_{n_o,n} \end{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}}$$

Adding parameters: FIR example

$$\text{AsCov } J(\hat{\theta}_N) = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}^* [\langle \Omega, \Omega \rangle]^{-1} \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$$

where

$$\Omega = \Gamma_n \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} = \begin{bmatrix} \Gamma_{n_o} \\ \Gamma_{n_o,n} \end{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}}$$

where

$$\Gamma_{m,n}(z) = \begin{bmatrix} z^{-(m+1)} & \dots & z^{-n} \end{bmatrix}^T.$$

Adding parameters: FIR example

$$\text{AsCov } J(\hat{\theta}_N) = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}^* [\langle \Omega, \Omega \rangle]^{-1} \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$$

where

$$\Omega = \Gamma_n \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} = \begin{bmatrix} \Gamma_{n_o} \\ \Gamma_{n_o,n} \end{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}}$$

where

$$\Gamma_{m,n}(z) = \begin{bmatrix} z^{-(m+1)} & \dots & z^{-n} \end{bmatrix}^T .$$

The question is now how large $\Lambda^* [\langle \Psi, \Psi \rangle]^{-1} \Lambda$ is in comparison with the expression above.

Structural results: Adding parameters

Geometrical result

Let $\mathcal{X}$ and $\mathcal{Y}$ be two subspaces of $\mathcal{L}_2^m$ such that $\mathcal{X} \subseteq \mathcal{Y} \subseteq \mathcal{L}_2^m$ and let $\gamma \in \mathcal{L}_2^{q \times m}$. It then holds that

$$\begin{aligned} &\langle \mathrm{Proj}_{\mathcal{Y}}\{\gamma\}, \mathrm{Proj}_{\mathcal{Y}}\{\gamma\} \rangle - \langle \mathrm{Proj}_{\mathcal{X}}\{\gamma\}, \mathrm{Proj}_{\mathcal{X}}\{\gamma\} \rangle \\ &= \langle \mathrm{Proj}_{\mathcal{X}^\perp(\mathcal{Y})}\{\gamma\}, \mathrm{Proj}_{\mathcal{X}^\perp(\mathcal{Y})}\{\gamma\} \rangle \geq 0 \end{aligned}$$

($\mathcal{X}^\perp(\mathcal{Y})$ denotes the orthogonal complement of $\mathcal{X}$ in $\mathcal{Y}$)

Structural results: Adding parameters

Geometrical result

Let $\mathcal{X}$ and $\mathcal{Y}$ be two subspaces of $\mathcal{L}_2^m$ such that $\mathcal{X} \subseteq \mathcal{Y} \subseteq \mathcal{L}_2^m$ and let $\gamma \in \mathcal{L}_2^{q \times m}$. It then holds that

$$\begin{aligned} &\langle \mathrm{Proj}_{\mathcal{Y}}\{\gamma\}, \mathrm{Proj}_{\mathcal{Y}}\{\gamma\} \rangle - \langle \mathrm{Proj}_{\mathcal{X}}\{\gamma\}, \mathrm{Proj}_{\mathcal{X}}\{\gamma\} \rangle \\ &= \langle \mathrm{Proj}_{\mathcal{X}^\perp(\mathcal{Y})}\{\gamma\}, \mathrm{Proj}_{\mathcal{X}^\perp(\mathcal{Y})}\{\gamma\} \rangle \geq 0 \end{aligned}$$

($\mathcal{X}^\perp(\mathcal{Y})$ denotes the orthogonal complement of $\mathcal{X}$ in $\mathcal{Y}$)

- The asymptotic covariance can never decrease when parameters are added

Structural results: Adding parameters

Geometrical result

Let $\mathcal{X}$ and $\mathcal{Y}$ be two subspaces of $\mathcal{L}_2^m$ such that $\mathcal{X} \subseteq \mathcal{Y} \subseteq \mathcal{L}_2^m$ and let $\gamma \in \mathcal{L}_2^{q \times m}$. It then holds that

$$\begin{aligned}
&\langle \mathrm{Proj}_{\mathcal{Y}}\{\gamma\}, \mathrm{Proj}_{\mathcal{Y}}\{\gamma\} \rangle - \langle \mathrm{Proj}_{\mathcal{X}}\{\gamma\}, \mathrm{Proj}_{\mathcal{X}}\{\gamma\} \rangle \\
&= \langle \mathrm{Proj}_{\mathcal{X}^\perp(\mathcal{Y})}\{\gamma\}, \mathrm{Proj}_{\mathcal{X}^\perp(\mathcal{Y})}\{\gamma\} \rangle \geq 0
\end{aligned}$$

($\mathcal{X}^\perp(\mathcal{Y})$ denotes the orthogonal complement of $\mathcal{X}$ in $\mathcal{Y}$)

- The asymptotic covariance can never decrease when parameters are added
- Above expression provides a quantification for the increase

Structural results: Adding parameters

Geometrical result

Let $\mathcal{X}$ and $\mathcal{Y}$ be two subspaces of $\mathcal{L}_2^m$ such that $\mathcal{X} \subseteq \mathcal{Y} \subseteq \mathcal{L}_2^m$ and let $\gamma \in \mathcal{L}_2^{q\times m}$. It then holds that

$$\begin{aligned} &\langle \text{Proj}_{\mathcal{Y}}\{\gamma\}, \text{Proj}_{\mathcal{Y}}\{\gamma\}\rangle - \langle \text{Proj}_{\mathcal{X}}\{\gamma\}, \text{Proj}_{\mathcal{X}}\{\gamma\}\rangle \\ &= \langle \text{Proj}_{\mathcal{X}^\perp(\mathcal{Y})}\{\gamma\}, \text{Proj}_{\mathcal{X}^\perp(\mathcal{Y})}\{\gamma\}\rangle \geq 0 \end{aligned}$$

($\mathcal{X}^\perp(\mathcal{Y})$ denotes the orthogonal complement of $\mathcal{X}$ in $\mathcal{Y}$)

- The asymptotic covariance can never decrease when parameters are added
- Above expression provides a quantification for the increase
- No increase if γ orthogonal to the orthogonal complement of X in $\mathcal{Y}$

Adding parameters: FIR example

$$\text{AsCov } J(\hat{\theta}_N) = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}^* [\langle \Omega, \Omega \rangle]^{-1} \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$$

where $\Omega = \Gamma_n \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} = \begin{bmatrix} \Gamma_{n_o} \\ \Gamma_{n_o,n} \end{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} \triangleq \begin{bmatrix} \Psi \\ \Phi \end{bmatrix}$

Adding parameters: FIR example

$$\text{AsCov } J(\hat{\theta}_N) = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}^* [\langle \Omega, \Omega \rangle]^{-1} \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$$

$$\text{where } \Omega = \Gamma_n \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} = \begin{bmatrix} \Gamma_{n_o} \\ \Gamma_{n_o,n} \end{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} \triangleq \begin{bmatrix} \Psi \\ \Phi \end{bmatrix}$$

$\mathcal{X} = \mathcal{S}_\Psi \subseteq \mathcal{S}_\Omega = \mathcal{Y}$.

Adding parameters: FIR example

$$\text{AsCov } J(\hat{\theta}_N) = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}^* [\langle \Omega, \Omega \rangle]^{-1} \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$$

$$\text{where } \Omega = \Gamma_n \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} = \begin{bmatrix} \Gamma_{n_o} \\ \Gamma_{n_o,n} \end{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} \triangleq \begin{bmatrix} \Psi \\ \Phi \end{bmatrix}$$

$\mathcal{X} = \mathcal{S}_\Psi \subseteq \mathcal{S}_\Omega = \mathcal{Y}$.

A property of a suitable γ: $\langle \Omega, \gamma \rangle = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$ i.e. $\gamma \perp \mathcal{S}_\Phi$.

Adding parameters: FIR example

$$\text{AsCov } J(\hat{\theta}_N) = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}^* [\langle \Omega, \Omega \rangle]^{-1} \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$$

$$\text{where } \Omega = \Gamma_n \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} = \begin{bmatrix} \Gamma_{n_o} \\ \Gamma_{n_o,n} \end{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} \triangleq \begin{bmatrix} \Psi \\ \Phi \end{bmatrix}$$

$\mathcal{X} = \mathcal{S}_\Psi \subseteq \mathcal{S}_\Omega = \mathcal{Y}$.

A property of a suitable γ: $\langle \Omega, \gamma \rangle = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$ i.e. $\gamma \perp \mathcal{S}_\Phi$.

- No increase:

Adding parameters: FIR example

$$\text{AsCov } J(\hat{\theta}_N) = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}^* [\langle \Omega, \Omega \rangle]^{-1} \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$$

where $\Omega = \Gamma_n \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} = \begin{bmatrix} \Gamma_{n_o} \\ \Gamma_{n_o,n} \end{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} \triangleq \begin{bmatrix} \Psi \\ \Phi \end{bmatrix}$

$\mathcal{X} = \mathcal{S}_\Psi \subseteq \mathcal{S}_\Omega = \mathcal{Y}$.

A property of a suitable γ: $\langle \Omega, \gamma \rangle = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$ i.e. $\gamma \perp \mathcal{S}_\Phi$.

- No increase: $\langle \Phi, \Psi \rangle = 0$

Adding parameters: FIR example

$$\text{AsCov } J(\hat{\theta}_N) = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}^* [\langle \Omega, \Omega \rangle]^{-1} \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$$

where $\Omega = \Gamma_n \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} = \begin{bmatrix} \Gamma_{n_o} \\ \Gamma_{n_o,n} \end{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} \triangleq \begin{bmatrix} \Psi \\ \Phi \end{bmatrix}$

$\mathcal{X} = \mathcal{S}_\Psi \subseteq \mathcal{S}_\Omega = \mathcal{Y}$.

A property of a suitable γ: $\langle \Omega, \gamma \rangle = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$ i.e. $\gamma \perp \mathcal{S}_\Phi$.

- No increase: $\langle \Phi, \Psi \rangle = 0 \Leftrightarrow \mathcal{S}_\Phi \perp \mathcal{S}_\Psi$

Adding parameters: FIR example

$$\text{AsCov } J(\hat{\theta}_N) = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}^* [\langle \Omega, \Omega \rangle]^{-1} \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$$

$$\text{where } \Omega = \Gamma_n \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} = \begin{bmatrix} \Gamma_{n_o} \\ \Gamma_{n_o,n} \end{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} \triangleq \begin{bmatrix} \Psi \\ \Phi \end{bmatrix}$$

$\mathcal{X} = \mathcal{S}_\Psi \subseteq \mathcal{S}_\Omega = \mathcal{Y}$.

A property of a suitable γ: $\langle \Omega, \gamma \rangle = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$ i.e. $\gamma \perp \mathcal{S}_\Phi$.

- No increase: $\langle \Phi, \Psi \rangle = 0 \Leftrightarrow \mathcal{S}_\Phi \perp \mathcal{S}_\Psi \Rightarrow \mathcal{S}_\Psi^\perp(\mathcal{S}_\Omega) = \mathcal{S}_\Phi$

Adding parameters: FIR example

$$\text{AsCov } J(\hat{\theta}_N) = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}^* [\langle \Omega, \Omega \rangle]^{-1} \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$$

where $\Omega = \Gamma_n \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} = \begin{bmatrix} \Gamma_{n_o} \\ \Gamma_{n_o,n} \end{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} \triangleq \begin{bmatrix} \Psi \\ \Phi \end{bmatrix}$

$\mathcal{X} = \mathcal{S}_\Psi \subseteq \mathcal{S}_\Omega = \mathcal{Y}$.

A property of a suitable γ: $\langle \Omega, \gamma \rangle = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$ i.e. $\gamma \perp \mathcal{S}_\Phi$.

- No increase: $\langle \Phi, \Psi \rangle = 0 \Leftrightarrow \mathcal{S}_\Phi \perp \mathcal{S}_\Psi \Rightarrow$ $\mathcal{S}_\Psi^\perp(\mathcal{S}_\Omega) = \mathcal{S}_\Phi \Rightarrow \text{Proj}_{\mathcal{S}_\Psi^\perp(\mathcal{S}_\Omega)}\{\gamma\} = 0$

Adding parameters: FIR example

$$\text{AsCov } J(\hat{\theta}_N) = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}^* [\langle \Omega, \Omega \rangle]^{-1} \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$$

where $\Omega = \Gamma_n \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} = \begin{bmatrix} \Gamma_{n_o} \\ \Gamma_{n_o,n} \end{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} \triangleq \begin{bmatrix} \Psi \\ \Phi \end{bmatrix}$

$\mathcal{X} = \mathcal{S}_\Psi \subseteq \mathcal{S}_\Omega = \mathcal{Y}$.

A property of a suitable γ: $\langle \Omega, \gamma \rangle = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$ i.e. $\gamma \perp \mathcal{S}_\Phi$.

- No increase: $\langle \Phi, \Psi \rangle = 0 \Leftrightarrow \mathcal{S}_\Phi \perp \mathcal{S}_\Psi \Rightarrow$ $\mathcal{S}_\Psi^\perp(\mathcal{S}_\Omega) = \mathcal{S}_\Phi \Rightarrow \text{Proj}_{\mathcal{S}_\Psi^\perp(\mathcal{S}_\Omega)}\{\gamma\} = 0$ ($\Leftarrow$ also holds)

Adding parameters: FIR example

$$\text{AsCov } J(\hat{\theta}_N) = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}^* [\langle \Omega, \Omega \rangle]^{-1} \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$$

$$\text{where } \Omega = \Gamma_n \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} = \begin{bmatrix} \Gamma_{n_o} \\ \Gamma_{n_o,n} \end{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} \triangleq \begin{bmatrix} \Psi \\ \Phi \end{bmatrix}$$

$\mathcal{X} = \mathcal{S}_\Psi \subseteq \mathcal{S}_\Omega = \mathcal{Y}$.

A property of a suitable γ: $\langle \Omega, \gamma \rangle = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$ i.e. $\gamma \perp \mathcal{S}_\Phi$.

- No increase: $\langle \Phi, \Psi \rangle = 0 \Leftrightarrow \mathcal{S}_\Phi \perp \mathcal{S}_\Psi \Rightarrow$ $\mathcal{S}_\Psi^\perp(\mathcal{S}_\Omega) = \mathcal{S}_\Phi \Rightarrow \text{Proj}_{\mathcal{S}_\Psi^\perp(\mathcal{S}_\Omega)}\{\gamma\} = 0$ ($\Leftarrow$ also holds)
- Strict increase:

Adding parameters: FIR example

$$\text{AsCov } J(\hat{\theta}_N) = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}^* [\langle \Omega, \Omega \rangle]^{-1} \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$$

$$\text{where } \Omega = \Gamma_n \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} = \begin{bmatrix} \Gamma_{n_o} \\ \Gamma_{n_o,n} \end{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} \triangleq \begin{bmatrix} \Psi \\ \Phi \end{bmatrix}$$

$\mathcal{X} = \mathcal{S}_\Psi \subseteq \mathcal{S}_\Omega = \mathcal{Y}$.

A property of a suitable γ: $\langle \Omega, \gamma \rangle = \begin{bmatrix} \Lambda \\ 0_{(n-n_o)\times q} \end{bmatrix}$ i.e. $\gamma \perp \mathcal{S}_\Phi$.

- No increase: $\langle \Phi, \Psi \rangle = 0 \Leftrightarrow \mathcal{S}_\Phi \perp \mathcal{S}_\Psi \Rightarrow$ $\mathcal{S}_\Psi^\perp(\mathcal{S}_\Omega) = \mathcal{S}_\Phi \Rightarrow \text{Proj}_{\mathcal{S}_\Psi^\perp(\mathcal{S}_\Omega)}\{\gamma\} = 0$ ($\Leftarrow$ also holds)
- Strict increase: $\langle \Phi, \Psi \rangle$ full column rank.

Structural results: Adding parameters - FIR example

- No increase: If and only if $\langle \Phi, \Psi \rangle = 0$
- Strict increase: $\langle \Phi, \Psi \rangle$ full column rank.

Structural results: Adding parameters - FIR example

- No increase: If and only if $\langle \Phi, \Psi \rangle = 0$
- Strict increase: $\langle \Phi, \Psi \rangle$ full column rank.

FIR example:

$$\begin{aligned}
\langle \Phi, \Psi \rangle =& \langle \Gamma_{n_o,n} \Phi_u^{1/2}, \Gamma_{n_o} \Phi_u^{1/2} \rangle \\
&= R_{n_o,n} \triangleq \begin{bmatrix} r_{n_o} & r_{n_o-1} & \cdots & r_1 \\ r_{n_o+1} & r_{n_o} & \cdots & r_2 \\ \vdots & & & \vdots \\ r_{n-1} & r_{n-2} & \cdots & r_{n-n_o} \end{bmatrix},
\end{aligned}$$

Structural results: Adding parameters - FIR example

- No increase: If and only if $\langle \Phi, \Psi \rangle = 0$
- Strict increase: $\langle \Phi, \Psi \rangle$ full column rank.

FIR example:

$$\begin{aligned} \langle \Phi, \Psi \rangle &= \langle \Gamma_{n_o,n} \Phi_u^{1/2}, \Gamma_{n_o} \Phi_u^{1/2} \rangle \\ &= R_{n_o,n} \triangleq \begin{bmatrix} r_{n_o} & r_{n_o-1} & \cdots & r_1 \\ r_{n_o+1} & r_{n_o} & \cdots & r_2 \\ \vdots & & & \vdots \\ r_{n-1} & r_{n-2} & \cdots & r_{n-n_o} \end{bmatrix}, \end{aligned}$$

- No increase: If and only if $r_1 = \ldots = r_{n-1} = 0$

Structural results: Adding parameters - FIR example

- No increase: If and only if $\langle \Phi, \Psi \rangle = 0$
- Strict increase: $\langle \Phi, \Psi \rangle$ full column rank.

FIR example:

$$\begin{aligned} \langle \Phi, \Psi \rangle =& \langle \Gamma_{n_o,n} \Phi_u^{1/2}, \Gamma_{n_o} \Phi_u^{1/2} \rangle \\ =& R_{n_o,n} \triangleq \begin{bmatrix} r_{n_o} & r_{n_o-1} & \cdots & r_1 \\ r_{n_o+1} & r_{n_o} & \cdots & r_2 \\ \vdots & & & \vdots \\ r_{n-1} & r_{n-2} & \cdots & r_{n-n_o} \end{bmatrix}, \end{aligned}$$

- No increase: If and only if $r_1 = \ldots = r_{n-1} = 0$
- Strict increase: If and only if $R_{n_o,n}$ has full column rank ($n \geq 2n_o$ necessary condition)

Structural results: Adding parameters

- No increase: If and only if $\langle \Phi, \Psi \rangle = 0$
- Strict increase: $\langle \Phi, \Psi \rangle$ full column rank.

Derivation not tied to the FIR example.

Structural results: Adding parameters

- No increase: If and only if $\langle \Phi, \Psi \rangle = 0$
- Strict increase: $\langle \Phi, \Psi \rangle$ full column rank.

Derivation not tied to the FIR example.

Generalization immediate!

Structural results: Adding parameters - NFIR

$$y_t = \sum_{k=1}^{n} \theta_k \, u_{t-k} + \sum_{k=n+1}^{m} \theta_k \, (u_{t-(k-n)})^2 + e_t = \theta \varphi_t^T + e_t$$

$$\Psi = \frac{1}{\sqrt{\lambda_o}} M \Phi_z^{1/2}.$$

where $\Phi_z^{1/2}$ is a Cholesky factor of Φ_z, the spectrum of $z = [u_t \; u_t^2]^T$.

Structural results: Adding parameters - NFIR

$$y_t = \sum_{k=1}^{n} \theta_k \, u_{t-k} + \sum_{k=n+1}^{m} \theta_k \, (u_{t-(k-n)})^2 + e_t = \theta \varphi_t^T + e_t$$

$$\Psi = \frac{1}{\sqrt{\lambda_o}} M \Phi_z^{1/2}.$$

where $\Phi_z^{1/2}$ is a Cholesky factor of Φ_z, the spectrum of $z = [u_t \; u_t^2]^T$.
White input with $\mathbf{E}[u_t^3] \neq 0$:

Structural results: Adding parameters - NFIR

$$y_t = \sum_{k=1}^{n} \theta_k \; u_{t-k} + \sum_{k=n+1}^{m} \theta_k \; (u_{t-(k-n)})^2 + e_t = \theta \varphi_t^T + e_t$$

$$\Psi = \frac{1}{\sqrt{\lambda_o}} M \Phi_z^{1/2}.$$

where $\Phi_z^{1/2}$ is a Cholesky factor of Φ_z, the spectrum of $z = [u_t \; u_t^2]^T$.
White input with $\mathbf{E}[u_t^3] \neq 0$:

$$\Psi \triangleq \begin{bmatrix} \sqrt{\mathbf{E}[u_t^2]} \; \Gamma_n & 0 \\ \frac{\mathbf{E}[u_t^3]}{\sqrt{\mathbf{E}[u_t^2]}} \; \Gamma_n & \sqrt{\mathbf{E}[u_t^4] - \frac{\mathbf{E}^2[u_t^3]}{\mathbf{E}[u_t^2]}} \; \Gamma_n \end{bmatrix}$$

Structural results: Adding parameters - NFIR

$$y_t = \sum_{k=1}^{n} \theta_k \; u_{t-k} + \sum_{k=n+1}^{m} \theta_k \; (u_{t-(k-n)})^2 + e_t = \theta \varphi_t^T + e_t$$

$$\Psi = \frac{1}{\sqrt{\lambda_o}} M \Phi_z^{1/2}.$$

where $\Phi_z^{1/2}$ is a Cholesky factor of Φ_z, the spectrum of $z = [u_t \; u_t^2]^T$.
White input with $\mathbf{E}[u_t^3] \neq 0$:

$$\Psi \triangleq \begin{bmatrix} \sqrt{\mathbf{E}[u_t^2]} \; \Gamma_n & 0 \\ \frac{\mathbf{E}[u_t^3]}{\sqrt{\mathbf{E}[u_t^2]}} \; \Gamma_n & \sqrt{\mathbf{E}[u_t^4] - \frac{\mathbf{E}^2[u_t^3]}{\mathbf{E}[u_t^2]}} \; \Gamma_n \end{bmatrix}$$

Compare the asymptotic covariance of the first $2n$ parameters when estimating $m = 2n$ and $m > 2n$ non-linear parameters.

Structural results: Adding parameters - NFIR

$$\Psi \triangleq \begin{bmatrix} \sqrt{\mathbf{E}[u_t^2]}\,\Gamma_n & 0 \\ \frac{\mathbf{E}[u_t^3]}{\sqrt{\mathbf{E}[u_t^2]}}\,\Gamma_n & \sqrt{\mathbf{E}[u_t^4] - \frac{\mathbf{E}^2[u_t^3]}{\mathbf{E}[u_t^2]}}\,\Gamma_n \end{bmatrix}$$

Structural results: Adding parameters - NFIR

$$\Psi \triangleq \begin{bmatrix} \sqrt{\mathbf{E}[u_t^2]}\;\Gamma_n & 0 \\ \frac{\mathbf{E}[u_t^3]}{\sqrt{\mathbf{E}[u_t^2]}}\;\Gamma_n & \sqrt{\mathbf{E}[u_t^4] - \frac{\mathbf{E}^2[u_t^3]}{\mathbf{E}[u_t^2]}}\;\Gamma_n \end{bmatrix}$$

$$\Phi \triangleq \begin{bmatrix} \frac{\mathbf{E}[u_t^3]}{\sqrt{\mathbf{E}[u_t^2]}}\;\Gamma_{n,m} & \sqrt{\mathbf{E}[u_t^4] - \frac{\mathbf{E}^2[u_t^3]}{\mathbf{E}[u_t^2]}}\;\Gamma_{n,m} \end{bmatrix}$$

Structural results: Adding parameters - NFIR

$$\Psi \triangleq \begin{bmatrix} \sqrt{\mathbf{E}[u_t^2]}\,\Gamma_n & 0 \\ \frac{\mathbf{E}[u_t^3]}{\sqrt{\mathbf{E}[u_t^2]}}\,\Gamma_n & \sqrt{\mathbf{E}[u_t^4] - \frac{\mathbf{E}^2[u_t^3]}{\mathbf{E}[u_t^2]}}\,\Gamma_n \end{bmatrix}$$

$$\Phi \triangleq \begin{bmatrix} \frac{\mathbf{E}[u_t^3]}{\sqrt{\mathbf{E}[u_t^2]}}\,\Gamma_{n,m} & \sqrt{\mathbf{E}[u_t^4] - \frac{\mathbf{E}^2[u_t^3]}{\mathbf{E}[u_t^2]}}\,\Gamma_{n,m} \end{bmatrix}$$

- $\mathcal{S}_\Psi \perp \mathcal{S}_\Phi \Rightarrow$ No increase in the asymptotic covariance of the linear parameters and the first n non-linear parameters from additional non-linear parameters.

Structural results: Adding parameters - NFIR

$$\Psi \triangleq \begin{bmatrix} \sqrt{\mathbf{E}[u_t^2]}\,\Gamma_n & 0 \\ \frac{\mathbf{E}[u_t^3]}{\sqrt{\mathbf{E}[u_t^2]}}\,\Gamma_n & \sqrt{\mathbf{E}[u_t^4] - \frac{\mathbf{E}^2[u_t^3]}{\mathbf{E}[u_t^2]}}\,\Gamma_n \end{bmatrix}$$

$$\Phi \triangleq \begin{bmatrix} \frac{\mathbf{E}[u_t^3]}{\sqrt{\mathbf{E}[u_t^2]}}\,\Gamma_{n,m} & \sqrt{\mathbf{E}[u_t^4] - \frac{\mathbf{E}^2[u_t^3]}{\mathbf{E}[u_t^2]}}\,\Gamma_{n,m} \end{bmatrix}$$

- $\mathcal{S}_\Psi \perp \mathcal{S}_\Phi \Rightarrow$ No increase in the asymptotic covariance of the linear parameters and the first n non-linear parameters from additional non-linear parameters.
- Can also easily show that increase is strict for the linear parameters when more than n non-linear parameters are estimated.

Structural results: Adding sources

Structural results: Adding sources

FIR system with second input:

$$y_t = \sum_{k=1}^{n} \theta_k \left(u_{t-k} + w_{t-k}\right) + \sum_{k=1}^{m} \theta_{n+k}\, w_{t-k} + e_t$$

Structural results: Adding sources

FIR system with second input:

$$y_t = \sum_{k=1}^{n} \theta_k \left(u_{t-k} + w_{t-k}\right) + \sum_{k=1}^{m} \theta_{n+k}\, w_{t-k} + e_t$$

This system can be written in linear regression form using

$$\varphi_t^T = \begin{bmatrix} \Gamma_n & \Gamma_n \\ 0_{m\times 1} & \Gamma_m \end{bmatrix} \begin{bmatrix} u_t \\ w_t \end{bmatrix}$$

Structural results: Adding sources

FIR system with second input:

$$y_t = \sum_{k=1}^{n} \theta_k \, (u_{t-k} + w_{t-k}) + \sum_{k=1}^{m} \theta_{n+k} \, w_{t-k} + e_t$$

This system can be written in linear regression form using

$$\varphi_t^T = \begin{bmatrix} \Gamma_n & \Gamma_n \\ 0_{m\times 1} & \Gamma_m \end{bmatrix} \begin{bmatrix} u_t \\ w_t \end{bmatrix}$$

$$\Psi = \Psi_e \triangleq \begin{bmatrix} \Gamma_n & \Gamma_n \\ 0_{m\times 1} & \Gamma_m \end{bmatrix} \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} & 0 \\ 0 & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}} \end{bmatrix} = \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} \Gamma_n & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}} \Gamma_n \\ 0_{m\times 1} & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}} \Gamma_m \end{bmatrix}$$

Structural results: Adding sources

FIR system with second input:

$$y_t = \sum_{k=1}^{n} \theta_k \, (u_{t-k} + w_{t-k}) + \sum_{k=1}^{m} \theta_{n+k} \, w_{t-k} + e_t$$

This system can be written in linear regression form using

$$\varphi_t^T = \begin{bmatrix} \Gamma_n & \Gamma_n \\ 0_{m\times 1} & \Gamma_m \end{bmatrix} \begin{bmatrix} u_t \\ w_t \end{bmatrix}$$

$$\Psi = \Psi_e \triangleq \begin{bmatrix} \Gamma_n & \Gamma_n \\ 0_{m\times 1} & \Gamma_m \end{bmatrix} \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} & 0 \\ 0 & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}} \end{bmatrix} = \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} \Gamma_n & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}} \Gamma_n \\ 0_{m\times 1} & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}} \Gamma_m \end{bmatrix}$$

Second column of Ψ due to the new input w, but also m new rows due to the additional parameters $\theta_{n+1}, \ldots, \theta_{n+m}$ that need to be estimated.

Structural results: Adding sources

FIR system with second input:

$$y_t = \sum_{k=1}^{n} \theta_k \left(u_{t-k} + w_{t-k}\right) + \sum_{k=1}^{m} \theta_{n+k}\, w_{t-k} + e_t$$

$$\Psi_e = \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}}\, \Gamma_n & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\, \Gamma_n \\ 0_{m\times 1} & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\, \Gamma_m \end{bmatrix}$$

Structural results: Adding sources

FIR system with second input:

$$y_t = \sum_{k=1}^{n} \theta_k \left(u_{t-k} + w_{t-k}\right) + \sum_{k=1}^{m} \theta_{n+k}\, w_{t-k} + e_t$$

$$\Psi_e = \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}}\,\Gamma_n & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\,\Gamma_n \\ 0_{m\times 1} & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\,\Gamma_m \end{bmatrix}$$

We're only interested in $\theta_1, \ldots, \theta_n$.

Structural results: Adding sources

FIR system with second input:

$$y_t = \sum_{k=1}^{n} \theta_k \left(u_{t-k} + w_{t-k}\right) + \sum_{k=1}^{m} \theta_{n+k}\, w_{t-k} + e_t$$

$$\Psi_e = \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}}\, \Gamma_n & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\, \Gamma_n \\ 0_{m\times 1} & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\, \Gamma_m \end{bmatrix}$$

We're only interested in $\theta_1, \ldots, \theta_n$.
Will using w increase the accuracy?

Structural results: Adding sources

FIR system with second input:

$$y_t = \sum_{k=1}^{n} \theta_k \, (u_{t-k} + w_{t-k}) + \sum_{k=1}^{m} \theta_{n+k} \, w_{t-k} + e_t$$

$$\Psi_e = \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} \, \Gamma_n & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}} \, \Gamma_n \\ 0_{m \times 1} & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}} \, \Gamma_m \end{bmatrix}$$

We're only interested in $\theta_1, \ldots, \theta_n$.
Will using w increase the accuracy?

+ SNR in first terms increase

Structural results: Adding sources

FIR system with second input:

$$y_t = \sum_{k=1}^{n} \theta_k \left(u_{t-k} + w_{t-k}\right) + \sum_{k=1}^{m} \theta_{n+k}\, w_{t-k} + e_t$$

$$\Psi_e = \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}}\,\Gamma_n & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\,\Gamma_n \\ 0_{m\times 1} & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\,\Gamma_m \end{bmatrix}$$

We're only interested in $\theta_1, \ldots, \theta_n$.
Will using w increase the accuracy?

+ SNR in first terms increase
- m additional parameters has to be estimated

Structural results: Geometric analysis

Let $\Psi \in \mathcal{L}_2^{n\times m}$, let $\Lambda \in \mathbb{C}^{n\times q}$ and

$$\Psi_e = \begin{bmatrix} \Psi & \psi \\ 0_{\delta n\times m} & \eta \end{bmatrix}$$

Structural results: Geometric analysis

Let $\Psi \in \mathcal{L}_2^{n\times m}$, let $\Lambda \in \mathbb{C}^{n\times q}$ and

$$\Psi_e = \begin{bmatrix} \Psi & \psi \\ 0_{\delta n\times m} & \eta \end{bmatrix}$$

Assume that $\langle \Psi, \Psi \rangle > 0$ and $\langle \Psi_e, \Psi_e \rangle > 0$ and define

$$X \triangleq \Lambda^* \left[\langle \Psi, \Psi \rangle\right]^{-1} \Lambda - \begin{bmatrix} \Lambda \\ 0_{\delta n\times q} \end{bmatrix}^* \left[\langle \Psi_e, \Psi_e \rangle\right]^{-1} \begin{bmatrix} \Lambda \\ 0_{\delta n\times q} \end{bmatrix}$$

Structural results: Geometric analysis

Let $\Psi \in \mathcal{L}_2^{n\times m}$, let $\Lambda \in \mathbb{C}^{n\times q}$ and

$$\Psi_e = \begin{bmatrix} \Psi & \psi \\ 0_{\delta n\times m} & \eta \end{bmatrix}$$

Assume that $\langle \Psi, \Psi \rangle > 0$ and $\langle \Psi_e, \Psi_e \rangle > 0$ and define

$$X \triangleq \Lambda^* \left[\langle \Psi, \Psi \rangle\right]^{-1} \Lambda - \begin{bmatrix} \Lambda \\ 0_{\delta n\times q} \end{bmatrix}^* \left[\langle \Psi_e, \Psi_e \rangle\right]^{-1} \begin{bmatrix} \Lambda \\ 0_{\delta n\times q} \end{bmatrix}$$

Theorem

$$X = \langle \mathrm{Proj}_{\mathcal{S}^{\perp}_{\Psi_e}(\mathcal{S}_\Omega)}\{\gamma\}, \mathrm{Proj}_{\mathcal{S}^{\perp}_{\Psi_e}(\mathcal{S}_\Omega)}\{\gamma\} \rangle \geq 0$$

where $\gamma \triangleq \Lambda^* \left[\langle \Phi, \Phi \rangle\right]^{-1} \Phi$, $\Phi \triangleq \begin{bmatrix} \Psi & 0_{n\times \delta m} \end{bmatrix}$, $\Omega \triangleq \begin{bmatrix} \Psi_e \\ \Phi \end{bmatrix}$

Structural results: Adding sources

$$\Psi_e = \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}}\ \Gamma_n & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\ \Gamma_n \\ 0_{m\times 1} & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\ \Gamma_m \end{bmatrix}$$

Structural results: Adding sources

$$\Psi_e = \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}}\,\Gamma_n & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\,\Gamma_n \\ 0_{m\times 1} & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\,\Gamma_m \end{bmatrix} = \begin{bmatrix} \Psi & \psi \\ 0_{\delta n\times m} & \eta \end{bmatrix}$$

Structural results: Adding sources

$$\Psi_e = \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}}\Gamma_n & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\Gamma_n \\ 0_{m\times 1} & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\Gamma_m \end{bmatrix} = \begin{bmatrix} \Psi & \psi \\ 0_{\delta n \times m} & \eta \end{bmatrix}$$

Difference given by $\langle \mathrm{Proj}_{\mathcal{S}^{\perp}_{\Psi_e}(\mathcal{S}_\Omega)}\{\gamma\}, \mathrm{Proj}_{\mathcal{S}^{\perp}_{\Psi_e}(\mathcal{S}_\Omega)}\{\gamma\}\rangle$

$$\Omega = \begin{bmatrix} \Psi_e \\ \Phi \end{bmatrix}$$

$$\Phi = \begin{bmatrix} \Psi & 0 \end{bmatrix}$$

Structural results: Adding sources

$$\Psi_e = \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}} \Gamma_n & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}} \Gamma_n \\ 0_{m\times 1} & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}} \Gamma_m \end{bmatrix} = \begin{bmatrix} \Psi & \psi \\ 0_{\delta n \times m} & \eta \end{bmatrix}$$

Difference given by $\langle \mathrm{Proj}_{\mathcal{S}^{\perp}_{\Psi_e}(\mathcal{S}_\Omega)}\{\gamma\}, \mathrm{Proj}_{\mathcal{S}^{\perp}_{\Psi_e}(\mathcal{S}_\Omega)}\{\gamma\}\rangle$

$$\Omega = \begin{bmatrix} \Psi_e \\ \Phi \end{bmatrix}$$

$$\Phi = \begin{bmatrix} \Psi & 0 \end{bmatrix}$$

Suppose $m \geq n$ (i.e. at least as many parameters associated with w as with u needs to be estimated):

Structural results: Adding sources

$$\Psi_e = \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}}\Gamma_n & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\Gamma_n \\ 0_{m\times 1} & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\Gamma_m \end{bmatrix} = \begin{bmatrix} \Psi & \psi \\ 0_{\delta n\times m} & \eta \end{bmatrix}$$

Difference given by $\langle \mathrm{Proj}_{\mathcal{S}^{\perp}_{\Psi_e}(\mathcal{S}_\Omega)}\{\gamma\}, \mathrm{Proj}_{\mathcal{S}^{\perp}_{\Psi_e}(\mathcal{S}_\Omega)}\{\gamma\}\rangle$

$$\Omega = \begin{bmatrix} \Psi_e \\ \Phi \end{bmatrix}$$

$$\Phi = \begin{bmatrix} \Psi & 0 \end{bmatrix}$$

Suppose $m \geq n$ (i.e. at least as many parameters associated with w as with u needs to be estimated): $\Rightarrow \mathcal{S}_\Phi \subseteq \mathcal{S}_{\Psi_e}$

Structural results: Adding sources

$$\Psi_e = \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}}\Gamma_n & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\Gamma_n \\ 0_{m\times 1} & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\Gamma_m \end{bmatrix} = \begin{bmatrix} \Psi & \psi \\ 0_{\delta n\times m} & \eta \end{bmatrix}$$

Difference given by $\langle \mathrm{Proj}_{\mathcal{S}^{\perp}_{\Psi_e}(\mathcal{S}_\Omega)}\{\gamma\}, \mathrm{Proj}_{\mathcal{S}^{\perp}_{\Psi_e}(\mathcal{S}_\Omega)}\{\gamma\}\rangle$

$$\Omega = \begin{bmatrix} \Psi_e \\ \Phi \end{bmatrix}$$

$$\Phi = \begin{bmatrix} \Psi & 0 \end{bmatrix}$$

Suppose $m \geq n$ (i.e. at least as many parameters associated with w as with u needs to be estimated): $\Rightarrow \mathcal{S}_\Phi \subseteq \mathcal{S}_{\Psi_e} \Rightarrow \mathcal{S}_\Omega = \mathcal{S}_{\Psi_e}$

Structural results: Adding sources

$$\Psi_e = \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}}\Gamma_n & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\Gamma_n \\ 0_{m\times 1} & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\Gamma_m \end{bmatrix} = \begin{bmatrix} \Psi & \psi \\ 0_{\delta n\times m} & \eta \end{bmatrix}$$

Difference given by $\langle \mathrm{Proj}_{\mathcal{S}^{\perp}_{\Psi_e}(\mathcal{S}_\Omega)}\{\gamma\}, \mathrm{Proj}_{\mathcal{S}^{\perp}_{\Psi_e}(\mathcal{S}_\Omega)}\{\gamma\}\rangle$

$$\Omega = \begin{bmatrix} \Psi_e \\ \Phi \end{bmatrix}$$

$$\Phi = \begin{bmatrix} \Psi & 0 \end{bmatrix}$$

Suppose $m \geq n$ (i.e. at least as many parameters associated with w as with u needs to be estimated): $\Rightarrow \mathcal{S}_\Phi \subseteq \mathcal{S}_{\Psi_e} \Rightarrow \mathcal{S}_\Omega = \mathcal{S}_{\Psi_e} \Rightarrow \mathcal{S}^{\perp}_{\Psi_e}(\mathcal{S}_\Omega)$ is empty

Structural results: Adding sources

$$\Psi_e = \begin{bmatrix} \frac{\Phi_u^{1/2}}{\sqrt{\lambda_o}}\Gamma_n & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\Gamma_n \\ 0_{m\times 1} & \frac{\Phi_w^{1/2}}{\sqrt{\lambda_o}}\Gamma_m \end{bmatrix} = \begin{bmatrix} \Psi & \psi \\ 0_{\delta n\times m} & \eta \end{bmatrix}$$

Difference given by $\langle \mathrm{Proj}_{\mathcal{S}_{\Psi_e}^{\perp}(\mathcal{S}_\Omega)}\{\gamma\}, \mathrm{Proj}_{\mathcal{S}_{\Psi_e}^{\perp}(\mathcal{S}_\Omega)}\{\gamma\}\rangle$

$$\Omega = \begin{bmatrix} \Psi_e \\ \Phi \end{bmatrix}$$

$$\Phi = \begin{bmatrix} \Psi & 0 \end{bmatrix}$$

Suppose $m \geq n$ (i.e. at least as many parameters associated with w as with u needs to be estimated): $\Rightarrow \mathcal{S}_\Phi \subseteq \mathcal{S}_{\Psi_e} \Rightarrow \mathcal{S}_\Omega = \mathcal{S}_{\Psi_e} \Rightarrow \mathcal{S}_{\Psi_e}^{\perp}(\mathcal{S}_\Omega)$ is empty$\Rightarrow$
No improvement (regardless of the power of w).

Structural results: Adding sensors

Structural results: Adding sensors

OE system (e_t is WGN with variance λ_e)

$$y_t = G(q, \theta)u_t + e_t$$

Structural results: Adding sensors

OE system (e_t is WGN with variance λ_e)

$$y_t = G(q, \theta)u_t + e_t$$

Investment issue ($$$$$): Will buying an additional, very accurate, sensor

$$z_t = M^o(q)G^o(q)u_t + w_t$$

where $\{w_t\}$ is WGN with variance $\lambda_w << \lambda_e$, increase the accuracy of the estimate of θ?

Structural results: Adding sensors

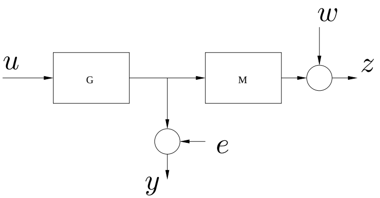
OE system (e_t is WGN with variance λ_e)

$$y_t = G(q, \theta)u_t + e_t$$

Investment issue ($$$$$): Will buying an additional, very accurate, sensor

$$z_t = M^o(q)G^o(q)u_t + w_t$$

where $\{w_t\}$ is WGN with variance $\lambda_w << \lambda_e$, increase the accuracy of the estimate of θ?



Structural results: Adding sensors

Sensor dynamics unknown and needs to be estimated:

$$M(q, \beta), \qquad \text{where } \beta \in \mathbb{R}^{1 \times n_\beta}.$$

Structural results: Adding sensors

Sensor dynamics unknown and needs to be estimated:

$$M(q,\beta), \qquad \text{where } \beta \in \mathbb{R}^{1\times n_\beta}.$$

ML estimation with two sensors $\Rightarrow$

$$\Psi = \Psi_2 = \begin{bmatrix} \frac{R}{\sqrt{\lambda_e}} G'(\theta^o) & \frac{R}{\sqrt{\lambda_w}} M^o G'(\theta^o) \\ 0_{n_\beta \times 1} & \frac{R}{\sqrt{\lambda_w}} G^o M'(\beta^o) \end{bmatrix}$$

Structural results: Adding sensors

Sensor dynamics unknown and needs to be estimated:

$$M(q, \beta), \qquad \text{where } \beta \in \mathbb{R}^{1\times n_\beta}.$$

ML estimation with two sensors $\Rightarrow$

$$\Psi = \Psi_2 = \begin{bmatrix} \frac{R}{\sqrt{\lambda_e}} G'(\theta^o) & \frac{R}{\sqrt{\lambda_w}} M^o G'(\theta^o) \\ 0_{n_\beta \times 1} & \frac{R}{\sqrt{\lambda_w}} G^o M'(\beta^o) \end{bmatrix}$$

No improvement whatsoever if $\frac{1}{\sqrt{\lambda_w}} M^o G'(\theta^o)$ spanned by rows of $\frac{1}{\sqrt{\lambda_w}} G^o M'(\beta^o)$.

Structural results: Adding sensors

Example: $\mathcal{L}_2$-gain estimation

Structural results: Adding sensors

Example: $\mathcal{L}_2$-gain estimation

System G and sensor M of FIR type.

Structural results: Adding sensors

Example: $\mathcal{L}_2$-gain estimation

System G and sensor M of FIR type.

$J(\theta) = \|\theta\|$

Structural results: Adding sensors

Example: $\mathcal{L}_2$-gain estimation

System G and sensor M of FIR type.

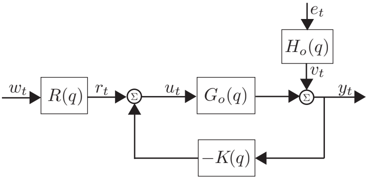
$$J(\theta) = \|\theta\|$$

There is no sensor that can improve the accuracy of J, even if the sensor only contains one single unknown parameter and the sensor is virtually noise free.

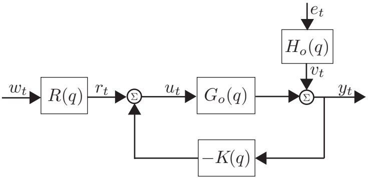
Outline

Properties of SISO LTI systems

Properties of SISO LTI systems



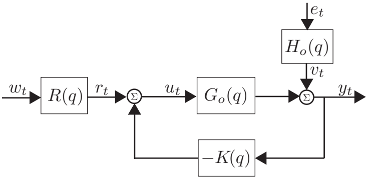
Properties of SISO LTI systems



$$y(t) = \begin{bmatrix} G(q,\theta) & H(q,\theta) \end{bmatrix} \begin{bmatrix} u(t) \\ e(t) \end{bmatrix} = T(q,\theta)\chi(t)$$

$$\chi(t) = \begin{bmatrix} S(q)R(q) & -C(q)S(q)H(q) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r(t) \\ e(t) \end{bmatrix} = U(q)\xi(t).$$

Properties of SISO LTI systems



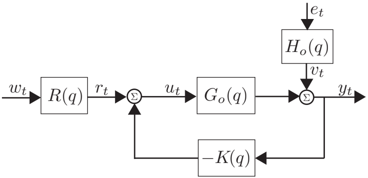
$$y(t) = \begin{bmatrix} G(q,\theta) & H(q,\theta) \end{bmatrix} \begin{bmatrix} u(t) \\ e(t) \end{bmatrix} = T(q,\theta)\chi(t)$$

$$\chi(t) = \begin{bmatrix} S(q)R(q) & -C(q)S(q)H(q) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r(t) \\ e(t) \end{bmatrix} = U(q)\xi(t).$$

Closed loop: $y(t) = T(q,\theta)U(q)\xi(t)$

Prediction error gradient: $\Psi(z) = \frac{1}{\sqrt{\lambda_o}H_o(z)} T'(z,\theta^o)U_o(z)\Phi_\xi^{1/2}$

Properties of SISO LTI systems



$$y(t) = \begin{bmatrix} G(q,\theta) & H(q,\theta) \end{bmatrix} \begin{bmatrix} u(t) \\ e(t) \end{bmatrix} = T(q,\theta)\chi(t)$$

$$\chi(t) = \begin{bmatrix} S(q)R(q) & -C(q)S(q)H(q) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r(t) \\ e(t) \end{bmatrix} = U(q)\xi(t).$$

Closed loop: $y(t) = T(q,\theta)U(q)\xi(t)$

Prediction error gradient: $\Psi(z) = \dfrac{1}{\sqrt{\lambda_o} H_o(z)} T'(z,\theta^o) U_o(z) \Phi_\xi^{1/2}$

Open loop: $\Psi(z) = \begin{bmatrix} \frac{R(z)}{\sqrt{\lambda_o}{}_o(z)} G'(z,\theta^o) & \frac{1}{H_o(z)} H'(z,\theta^o) \end{bmatrix}$

Properties of SISO linear systems

Which quantities are difficult to estimate?

- Frequency functions
- Impulse response coefficients
- System gains (e.g. $\mathcal{L}_2$)
- Zeros and poles
- Control design performance
-

Rules of thumbs??

Properties of SISO linear systems

Basic idea:

- Parametrize quantity of interest J in terms of the system impulse response (could also be done using the frequency function)
- Use re-parametrization together with geometrical approach

Properties of SISO linear systems

Theorem

The asymptotic variance of an estimated quantity $J(G(\cdot))$ is given by

$$\operatorname{AsVar} \hat{J} = \lambda_0 \left\| \Pr_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

Properties of SISO linear systems

Theorem

The asymptotic variance of an estimated quantity $J(G(\cdot))$ is given by

$$\operatorname{AsVar} \hat{J} = \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

Is this really better???

The variance expression 1(3)

$$\operatorname{AsVar} \hat{J} = \lambda_0 \left\| \Pr_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

The variance expression 1(3)

$$\operatorname{AsVar}\hat{J} = \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

Definition

$$\nabla J(z) = \sum_{k=1}^{\infty} \overline{\frac{\mathrm{d}J(g)}{\mathrm{d}g_k}} z^{-k}$$

The variance expression 1(3)

$$\text{AsVar}\,\hat{J} = \lambda_0 \left\| \text{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

Definition

$$\nabla J(z) = \sum_{k=1}^{\infty} \overline{\frac{\mathrm{d}J(g)}{\mathrm{d}g_k}} z^{-k}$$

Examples: $J(\cdot)$ is

- the $\mathcal{L}_2$-gain $\Rightarrow \nabla J(z) = 2G(z)$
- a real nmp-zero $z_k \Rightarrow \nabla J(z) \propto \frac{1}{z - z_k^{-1}}$
- an impulse response coefficient $g_k \Rightarrow \nabla J(z) = z^{-k}$

The variance expression 2(3)

$$\operatorname{AsVar} \hat{J} = \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

The variance expression 2(3)

$$\text{AsVar}\,\hat{J} = \lambda_0 \left\| \text{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

$\frac{H(z)}{S(z)R(z)}$ is the spectral factor of $\frac{\Phi_v(z)}{\Phi_u^r(z)}$

The variance expression 3(3)

$$\text{AsVar}\,\hat{J} = \lambda_0 \left\| \text{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

The variance expression 3(3)

$$\mathrm{AsVar}\,\hat{J} = \lambda_0 \left\| \mathrm{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

Definition

The space $\mathcal{S}_\Psi$ is defined as the linear span of the rows of the predictor gradient Ψ. The space can be written as

$$\mathcal{S}_\Psi = \mathrm{span}\{H^{-1}(z)T'(z)U(z)\}$$

The variance expression 3(3)

$$\mathrm{AsVar}\,\hat{J} = \lambda_0 \left\| \mathrm{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

Definition

The space $\mathcal{S}_\Psi$ is defined as the linear span of the rows of the predictor gradient Ψ. The space can be written as

$$\mathcal{S}_\Psi = \mathrm{span}\{H^{-1}(z)T'(z)U(z)\}$$

Example:

For an FIR model structure in open loop the space is

$$\mathcal{S}_\Psi = \mathrm{span}\left\{ \begin{bmatrix} R(z)z^{-1} & 0 \end{bmatrix}, \cdots, \begin{bmatrix} R(z)z^{-n} & 0 \end{bmatrix} \right\}$$

where $\Phi_u(z) = R(z)R(z^{-1})$.

Properties of SISO LTI systems

$$\operatorname{AsVar}\hat{J} = \lambda_0 \left\| \Pr_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

What does the expression tell us?

Properties of SISO LTI systems

$$\operatorname{AsVar} \hat{J} = \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \left[\frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} \quad 0 \right] \right) \right\|^2$$

What does the expression tell us?
How the asymptotic variance depends on

- Input signal
- Feedback control
- Model structure
- The property itself $J(\cdot)$

Properties of SISO LTI systems

$$\operatorname{AsVar} \hat{J} = \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

What does the expression tell us?
How the asymptotic variance depends on

- Input signal
- Feedback control
- Model structure
- The property itself $J(\cdot)$

Influence decoupled!

Properties of SISO LTI systems: Upper bounds

$$\operatorname{AsVar} \hat{J} = \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

Properties of SISO LTI systems: Upper bounds

$$\operatorname{AsVar} \hat{J} = \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

Still think it's too messy?

Properties of SISO LTI systems: Upper bounds

$$\operatorname{AsVar} \hat{J} = \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

Still think it's too messy?

Then forget about the projection:

$$\operatorname{AsVar} \hat{J} \leq \lambda_0 \left\| \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

Properties of SISO LTI systems: Upper bounds

$$\operatorname{AsVar} \hat{J} = \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

Still think it's too messy?

Then forget about the projection:

$$\operatorname{AsVar} \hat{J} \leq \lambda_0 \left\| \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

$$\operatorname{AsVar} \hat{J} \leq \|\nabla J\|^2_{\Phi_v/\Phi_u^r}$$

Independent of the model structure!

Properties of SISO LTI systems: Examples

$$\operatorname{AsVar} \hat{J} \leq \|\nabla J\|^2_{\Phi_v/\Phi_u^r}$$

Properties of SISO LTI systems: Examples

$$\mathrm{AsVar}\, \hat{J} \leq \|\nabla J\|^2_{\Phi_v/\Phi_u^r}$$

Quantity	Upper bound
impulse response coef.	$\|1\|^2_{\frac{\Phi_v}{\Phi_u^r}}$
L_2-gain	$\frac{4}{N}\|G\|^2_{\frac{\Phi_v}{\Phi_u^r}}$
real nmp-zero	$\frac{1}{\lvert\widetilde{G}_o(z_o)\rvert^2}\left\|\frac{1}{1-z_o^{-1}z^{-1}}\right\|^2_{\frac{\Phi_v}{\Phi_u^r}}$

$$(\widetilde{G}_o(z) = G_o(z)/(1 - z_o z^{-1}))$$

Properties of LTI systems: Frequency function

For an estimate of the frequency function $\sum_{k=1}^{\infty} g_k e^{-j\omega k}$ we have

$$\frac{dJ}{dg_k} = e^{-j\omega k}$$

Properties of LTI systems: Frequency function

For an estimate of the frequency function $\sum_{k=1}^{\infty} g_k e^{-j\omega k}$ we have

$$\frac{dJ}{dg_k} = e^{-j\omega k}$$

and the expression reduces to

$$\operatorname{AsCov} T(e^{j\omega}, \hat{\theta}_N) = W^T(e^{j\omega}) \sum_{k=1}^{n} \mathcal{B}_k^T(e^{j\omega}) \overline{\mathcal{B}_k(e^{j\omega}) W(e^{j\omega})}$$

$$\text{where } W(z) \triangleq \sqrt{\lambda_0} H_o(z) \Phi_\xi^{-1/2} U_o^{-1}(z)$$

and where $\{\mathcal{B}_k\}_{k=1}^{n}$ is any orthonormal basis for the space $\mathcal{S}_\Psi$.

Properties of LTI systems: Frequency function

For an estimate of the frequency function $\sum_{k=1}^{\infty} g_k e^{-j\omega k}$ we have

$$\frac{dJ}{dg_k} = e^{-j\omega k}$$

and the expression reduces to

$$\mathrm{AsCov}\, T(e^{j\omega}, \hat{\theta}_N) = W^T(e^{j\omega}) \sum_{k=1}^{n} \mathcal{B}_k^T(e^{j\omega}) \overline{\mathcal{B}_k(e^{j\omega}) W(e^{j\omega})}$$

where $W(z) \triangleq \sqrt{\lambda_0} H_o(z) \Phi_\xi^{-1/2} U_o^{-1}(z)$

and where $\{\mathcal{B}_k\}_{k=1}^{n}$ is any orthonormal basis for the space $\mathcal{S}_\Psi$.

Same expression as given by Xie& Ljung (2001), Ninness&Hjalmarsson(2003,2004)

Properties of SISO LTI systems: MV control

Minimum phase system with 1 time delay $\Rightarrow$ MV-controller

$$K_{\mathsf{MV}}(G_o, H_o) = \frac{H_o - 1}{G_o},$$

Properties of SISO LTI systems: MV control

Minimum phase system with 1 time delay $\Rightarrow$ MV-controller

$$K_{\mathsf{MV}}(G_o, H_o) = \frac{H_o - 1}{G_o},$$

Performance of any controller

$$V(K) \triangleq \mathbf{E}\left[(y_t - e_t)^2\right]$$

Properties of SISO LTI systems: MV control

Minimum phase system with 1 time delay $\Rightarrow$ MV-controller

$$K_{\mathsf{MV}}(G_o, H_o) = \frac{H_o - 1}{G_o},$$

Performance of any controller

$$V(K) \triangleq \mathbf{E}\left[(y_t - e_t)^2\right]$$

Certainty equivalence control

$$K_{\mathsf{MV}}(G(q, \hat{\theta}_N), H(q, \hat{\theta}_N))$$

Properties of SISO LTI systems: MV control

Minimum phase system with 1 time delay $\Rightarrow$ MV-controller

$$K_{\mathsf{MV}}(G_o, H_o) = \frac{H_o - 1}{G_o},$$

Performance of any controller

$$V(K) \triangleq \mathbf{E}\left[(y_t - e_t)^2\right]$$

Certainty equivalence control

$$K_{\mathsf{MV}}(G(q, \hat{\theta}_N), H(q, \hat{\theta}_N))$$

What can we say about

$$J(\hat{\theta}_N) \triangleq V(K_{\mathsf{MV}}(G(q, \hat{\theta}_N), H(q, \hat{\theta}_N))) \qquad ???$$

Properties of SISO LTI systems: Example 2 - Minimum variance control

An old problem:

Gevers and Ljung (1986):

Properties of SISO LTI systems: Example 2 - Minimum variance control

An old problem:

Gevers and Ljung (1986):

$$\delta J \triangleq \lim_{N\to\infty} N \cdot \mathbf{E} J(\hat{\theta}_N) = \|X_o\|^2_{\mathrm{AsCov}\, T(\cdot, \hat{\theta}_N)}$$

$$X_o = \frac{\sqrt{\lambda_o}}{H_o} \begin{bmatrix} \frac{H_o - 1}{G_o} & -1 \end{bmatrix}$$

Properties of SISO LTI systems: Example 2 - Minimum variance control

An old problem:
Gevers and Ljung (1986):

$$\delta J \triangleq \lim_{N\to\infty} N \cdot \mathbf{E} J(\hat{\theta}_N) = \|X_o\|^2_{\mathrm{AsCov}\, T(\cdot,\hat{\theta}_N)}$$
$$X_o = \frac{\sqrt{\lambda_o}}{H_o}\begin{bmatrix}\frac{H_o-1}{G_o} & -1\end{bmatrix}$$

High model order approximation of $\mathrm{AsCov}\, T(e^{j\omega},\hat{\theta}_N)$ yields

$$\delta J \approx \delta J_{ho} \triangleq m\|X_o\|^2_{\Phi_v \Phi_\chi^{-T}} \quad (m \text{ is the model order})$$

Properties of SISO LTI systems: Example 2 - Minimum variance control

An old problem:
Gevers and Ljung (1986):

$$\delta J \triangleq \lim_{N\to\infty} N \cdot \mathbf{E} J(\hat{\theta}_N) = \|X_o\|^2_{\mathrm{AsCov}\, T(\cdot,\hat{\theta}_N)}$$

$$X_o = \frac{\sqrt{\lambda_o}}{H_o} \begin{bmatrix} \frac{H_o - 1}{G_o} & -1 \end{bmatrix}$$

High model order approximation of $\mathrm{AsCov}\, T(e^{j\omega}, \hat{\theta}_N)$ yields

$$\delta J \approx \delta J_{ho} \triangleq m \|X_o\|^2_{\Phi_v \Phi_\chi^{-T}} \quad (m \text{ is the model order})$$

Optimization of δJ_{ho} yields that closed loop control with MV-controller optimal. Reference signal excitation plays no role.

Properties of SISO LTI systems: Example 2 - Minimum variance control

An old problem:
Gevers and Ljung (1986):

$$\delta J \triangleq \lim_{N\to\infty} N \cdot \mathbf{E} J(\hat{\theta}_N) = \|X_o\|^2_{\operatorname{AsCov} T(\cdot,\hat{\theta}_N)}$$

$$X_o = \frac{\sqrt{\lambda_o}}{H_o}\begin{bmatrix} \frac{H_o - 1}{G_o} & -1 \end{bmatrix}$$

High model order approximation of $\operatorname{AsCov} T(e^{j\omega}, \hat{\theta}_N)$ yields

$$\delta J \approx \delta J_{ho} \triangleq m\|X_o\|^2_{\Phi_v \Phi_\chi^{-T}} \quad (m \text{ is the model order})$$

Optimization of δJ_{ho} yields that closed loop control with MV-controller optimal. Reference signal excitation plays no role.
Minimum cost: $\delta J_{ho} = \lambda_o m$

Properties of SISO LTI systems: MV control

$$\delta J = \|X_o\|^2_{\mathrm{AsCov}\, T(\cdot, \hat{\theta}_N)}, \quad X_o = \frac{\sqrt{\lambda_o}}{H_o} \begin{bmatrix} \frac{H_o - 1}{G_o} & -1 \end{bmatrix}$$

Properties of SISO LTI systems: MV control

$$\delta J = \|X_o\|^2_{\mathrm{AsCov}\, T(\cdot,\hat{\theta}_N)}, \quad X_o = \frac{\sqrt{\lambda_o}}{H_o}\begin{bmatrix}\frac{H_o-1}{G_o} & -1\end{bmatrix}$$

Exact expr.: $\mathrm{AsCov}\, T(e^{j\omega},\hat{\theta}_N) = W^T(e^{j\omega})\sum_{k=1}^{n}\mathcal{B}_k^T(e^{j\omega})\overline{\mathcal{B}_k(e^{j\omega})W(e^{j\omega})}$

where $W(z) \triangleq \sqrt{\lambda_0}H_o(z)\Phi_\xi^{-1/2}U_o^{-1}(z)$

Properties of SISO LTI systems: MV control

$$\delta J = \|X_o\|^2_{\mathrm{AsCov}\,T(\cdot,\hat{\theta}_N)}, \quad X_o = \frac{\sqrt{\lambda_o}}{H_o}\begin{bmatrix}\frac{H_o-1}{G_o} & -1\end{bmatrix}$$

Exact expr.: $\mathrm{AsCov}\,T(e^{j\omega},\hat{\theta}_N) = W^T(e^{j\omega})\sum_{k=1}^{n}\mathcal{B}_k^T(e^{j\omega})\overline{\mathcal{B}_k(e^{j\omega})W(e^{j\omega})}$

where $W(z) \triangleq \sqrt{\lambda_0}H_o(z)\Phi_\xi^{-1/2}U_o^{-1}(z)$

Write $\mathcal{B}_k = [\mathcal{B}_k^G,\ \mathcal{B}_k^H]$. Then simple algebra gives

$$\delta J = \lambda_o\sum_{k=1}^{n}\left\|Z_o\,\mathcal{B}_k^G\right\|^2 + \lambda_o\sum_{k=1}^{n}\left\|\mathcal{B}_k^H\right\|^2$$

where $Z_o = \frac{1}{S_oR}\left(\frac{H_o-1}{G_o} - KS_oH_o\right)\sqrt{\lambda_o}$

Minimum variance control: A global lower bound

Minimum variance control: A global lower bound

$$\delta J = \lambda_o \sum_{k=1}^{n} \left\| Z_o \, \mathcal{B}_k^G \right\|^2 + \lambda_o \sum_{k=1}^{n} \left\| \mathcal{B}_k^H \right\|^2$$

Minimum variance control: A global lower bound

$$\delta J = \lambda_o \sum_{k=1}^{n} \left\| Z_o \, \mathcal{B}_k^G \right\|^2 + \lambda_o \sum_{k=1}^{n} \left\| \mathcal{B}_k^H \right\|^2$$

Let n_H number of parameters that only influence noise model H.

Minimum variance control: A global lower bound

$$\delta J = \lambda_o \sum_{k=1}^{n} \left\| Z_o \, \mathcal{B}_k^G \right\|^2 + \lambda_o \sum_{k=1}^{n} \left\| \mathcal{B}_k^H \right\|^2$$

Let n_H number of parameters that only influence noise model H.
$\Rightarrow n_H$ rows of Ψ of the form $[0, \; X]$

Minimum variance control: A global lower bound

$$\delta J = \lambda_o \sum_{k=1}^{n} \left\| Z_o \, \mathcal{B}_k^G \right\|^2 + \lambda_o \sum_{k=1}^{n} \left\| \mathcal{B}_k^H \right\|^2$$

Let n_H number of parameters that only influence noise model H.
$\Rightarrow n_H$ rows of Ψ of the form $[0, \; X]$
$\Rightarrow n_H$ basis functions $\mathcal{B}_k$ of the form $\mathcal{B}_k = [0, \; \mathcal{B}_k^H]$.

Minimum variance control: A global lower bound

$$\delta J = \lambda_o \sum_{k=1}^{n} \left\| Z_o \, \mathcal{B}_k^G \right\|^2 + \lambda_o \sum_{k=1}^{n} \left\| \mathcal{B}_k^H \right\|^2$$

Let n_H number of parameters that only influence noise model H.
$\Rightarrow n_H$ rows of Ψ of the form $[0, \ X]$
$\Rightarrow n_H$ basis functions $\mathcal{B}_k$ of the form $\mathcal{B}_k = [0, \ \mathcal{B}_k^H]$.
But then

$$\delta J = \lambda_o \sum_{k=n_H+1}^{n} \left\| Z_o \, \mathcal{B}_k^G \right\|^2 + \lambda_o \sum_{k=1}^{n} \left\| \mathcal{B}_k^H \right\|^2 \geq \lambda_o \sum_{k=1}^{n_H} \left\| \mathcal{B}_k^H \right\|^2 = \lambda_o n_H$$

Minimum variance control: A global lower bound

$$\delta J = \lambda_o \sum_{k=1}^{n} \left\| Z_o \, \mathcal{B}_k^G \right\|^2 + \lambda_o \sum_{k=1}^{n} \left\| \mathcal{B}_k^H \right\|^2$$

Let n_H number of parameters that only influence noise model H.
$\Rightarrow n_H$ rows of Ψ of the form $[0, \; X]$
$\Rightarrow n_H$ basis functions $\mathcal{B}_k$ of the form $\mathcal{B}_k = [0, \; \mathcal{B}_k^H]$.
But then

$$\delta J = \lambda_o \sum_{k=n_H+1}^{n} \left\| Z_o \, \mathcal{B}_k^G \right\|^2 + \lambda_o \sum_{k=1}^{n} \left\| \mathcal{B}_k^H \right\|^2 \geq \lambda_o \sum_{k=1}^{n_H} \left\| \mathcal{B}_k^H \right\|^2 = \lambda_o n_H$$

since $1 = \|\mathcal{B}_k\|^2 = \|\mathcal{B}_k^G\|^2 + \|\mathcal{B}_k^H\|^2$.

Minimum variance control: A global upper bound

Minimum variance control: A global upper bound

$$\delta J = \lambda_o \sum_{k=1}^{n} \left\| Z_o \, \mathcal{B}_k^G \right\|^2 + \lambda_o \sum_{k=1}^{n} \left\| \mathcal{B}_k^H \right\|^2$$

where $Z_o = \frac{1}{S_o R} \left(\frac{H_o - 1}{G_o} - K S_o H_o \right) \sqrt{\lambda_o}$

Minimum variance control: A global upper bound

$$\delta J = \lambda_o \sum_{k=1}^{n} \left\| Z_o \, \mathcal{B}_k^G \right\|^2 + \lambda_o \sum_{k=1}^{n} \left\| \mathcal{B}_k^H \right\|^2$$

where $Z_o = \frac{1}{S_o R} \left(\frac{H_o - 1}{G_o} - K S_o H_o \right) \sqrt{\lambda_o}$

Let K be the minimum variance controller

Minimum variance control: A global upper bound

$$\delta J = \lambda_o \sum_{k=1}^{n} \left\| Z_o \, \mathcal{B}_k^G \right\|^2 + \lambda_o \sum_{k=1}^{n} \left\| \mathcal{B}_k^H \right\|^2$$

where $Z_o = \dfrac{1}{S_o R} \left(\dfrac{H_o - 1}{G_o} - K S_o H_o \right) \sqrt{\lambda_o}$

Let K be the minimum variance controller $\Rightarrow Z_o = 0$

Minimum variance control: A global upper bound

$$\delta J = \lambda_o \sum_{k=1}^{n} \left\| Z_o \, \mathcal{B}_k^G \right\|^2 + \lambda_o \sum_{k=1}^{n} \left\| \mathcal{B}_k^H \right\|^2$$

where $Z_o = \dfrac{1}{S_o R} \left(\dfrac{H_o - 1}{G_o} - K S_o H_o \right) \sqrt{\lambda_o}$

Let K be the minimum variance controller $\Rightarrow Z_o = 0$

$$\Rightarrow \quad \delta J = \lambda_o \sum_{k=1}^{n} \left\| \mathcal{B}_k^H \right\|^2 \leq \lambda_o n$$

Minimum variance control: ARMAX models

Minimum variance control: ARMAX models

$$A(q)y_t = B(q)u_t + C(q)e_t$$
$$B(q) = b_1q^{-1} + \ldots + b_{n_b}q^{-n_b},\ A(q) = 1 + a_1q^{-1} + \ldots + a_{n_a}q^{-n_a},$$

Minimum variance control: ARMAX models

$$A(q)y_t = B(q)u_t + C(q)e_t$$

$$B(q) = b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b}, \; A(q) = 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a},$$

i) using minimum variance control in the identification experiment results in a cost δJ that satisfies the bounds

$$\lambda_o n_c \leq \delta J \leq \lambda_o (n_a + n_b + n_c - \min(n_a, n_c))$$

Minimum variance control: ARMAX models

$$A(q)y_t = B(q)u_t + C(q)e_t$$
$$B(q) = b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b},\ A(q) = 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a},$$

i) using minimum variance control in the identification experiment results in a cost δJ that satisfies the bounds

$$\lambda_o n_c \leq \delta J \leq \lambda_o (n_a + n_b + n_c - \min(n_a, n_c))$$

ii) minimum variance control is optimal and yields the minimum achievable cost $\delta J = \lambda_o\, n_c$ when

$$n_c \geq n_a, \quad \text{and} \quad C_o - A_o = B_o X$$

for some polynomial X.

Minimum variance control: ARMAX models

$$A(q)y_t = B(q)u_t + C(q)e_t$$
$$B(q) = b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b},\; A(q) = 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a},$$

i) using minimum variance control in the identification experiment results in a cost δJ that satisfies the bounds

$$\lambda_o n_c \leq \delta J \leq \lambda_o (n_a + n_b + n_c - \min(n_a, n_c))$$

ii) minimum variance control is optimal and yields the minimum achievable cost $\delta J = \lambda_o\, n_c$ when

$$n_c \geq n_a, \quad \text{and} \quad C_o - A_o = B_o X$$

for some polynomial X.

Result ii) is a generalization of Theorem 1 in Hildebrand&Solari 2007 which covers the case B_o constant and $n_c \geq n_a$.

Is open loop or closed loop identification better when the input spectrum due to the reference signal is fix?

Properties of SISO LTI systems: Open vs closed

Is open loop or closed loop identification better when the input spectrum due to the reference signal is fix?

$$\operatorname{AsVar} \hat{J} = \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

Properties of SISO LTI systems: Open vs closed

Is open loop or closed loop identification better when the input spectrum due to the reference signal is fix?

$$\operatorname{AsVar} \hat{J} = \lambda_0 \left\| \Pr_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

$$|RS_o|^2 \text{ fix } \Rightarrow$$

$$\operatorname{AsVar} \hat{J} = \lambda_0 \left\| \Pr_{\mathcal{S}_\Psi} \left(\begin{bmatrix} f \\ 0 \end{bmatrix} \right) \right\|^2$$

where f is fix.

Properties of SISO LTI systems: Open vs closed

Is open loop or closed loop identification better when the input spectrum due to the reference signal is fix?

$$\operatorname{AsVar} \hat{J} = \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2$$

$$|RS_o|^2 \text{ fix } \Rightarrow$$

$$\operatorname{AsVar} \hat{J} = \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi} \left(\begin{bmatrix} f \\ 0 \end{bmatrix} \right) \right\|^2$$

where f is fix.

$\Rightarrow$ Open or closed loop operation only affects the subspace $\mathcal{S}_\Psi$

Box Jenkins models

$$\mathcal{S}_{\Psi_{ol}} = \mathrm{Span}\left\{ \begin{bmatrix}\alpha_1\\0\end{bmatrix}^T, \cdots, \begin{bmatrix}\alpha_{n_\theta}\\0\end{bmatrix}^T, \begin{bmatrix}0\\\gamma_1\end{bmatrix}^T, \cdots, \begin{bmatrix}0\\\gamma_{n_\eta}\end{bmatrix}^T \right\}$$

$$\mathcal{S}_{\Psi_{cl}} = \mathrm{Span}\left\{ \begin{bmatrix}\alpha_1\\\beta_1\end{bmatrix}^T, \cdots, \begin{bmatrix}\alpha_{n_\theta}\\\beta_{n_\theta}\end{bmatrix}^T, \begin{bmatrix}0\\\gamma_1\end{bmatrix}^T, \cdots, \begin{bmatrix}0\\\gamma_{n_\eta}\end{bmatrix}^T \right\}$$

Properties of SISO LTI systems: Open vs closed

Box Jenkins models

$$\mathcal{S}_{\Psi_{ol}} = \text{Span}\left\{ \begin{bmatrix} \alpha_1 \\ 0 \end{bmatrix}^T, \cdots, \begin{bmatrix} \alpha_{n_\theta} \\ 0 \end{bmatrix}^T, \begin{bmatrix} 0 \\ \gamma_1 \end{bmatrix}^T, \cdots, \begin{bmatrix} 0 \\ \gamma_{n_\eta} \end{bmatrix}^T \right\}$$

$$\mathcal{S}_{\Psi_{cl}} = \text{Span}\left\{ \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix}^T, \cdots, \begin{bmatrix} \alpha_{n_\theta} \\ \beta_{n_\theta} \end{bmatrix}^T, \begin{bmatrix} 0 \\ \gamma_1 \end{bmatrix}^T, \cdots, \begin{bmatrix} 0 \\ \gamma_{n_\eta} \end{bmatrix}^T \right\}$$

But $\mathcal{S}_{\Psi_{cl}} \subseteq \mathcal{S}_{\Psi_+} = \mathcal{S}_{\Psi_{ol}} \oplus \text{Span}\left\{ \begin{bmatrix} 0 \\ \beta_1 \end{bmatrix}^T, \cdots, \begin{bmatrix} 0 \\ \beta_{n_\theta} \end{bmatrix}^T \right\}$

Properties of SISO LTI systems: Open vs closed

Box Jenkins models

$$\mathcal{S}_{\Psi_{ol}} = \text{Span}\left\{ \begin{bmatrix} \alpha_1 \\ 0 \end{bmatrix}^T, \cdots, \begin{bmatrix} \alpha_{n_\theta} \\ 0 \end{bmatrix}^T, \begin{bmatrix} 0 \\ \gamma_1 \end{bmatrix}^T, \cdots, \begin{bmatrix} 0 \\ \gamma_{n_\eta} \end{bmatrix}^T \right\}$$

$$\mathcal{S}_{\Psi_{cl}} = \text{Span}\left\{ \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix}^T, \cdots, \begin{bmatrix} \alpha_{n_\theta} \\ \beta_{n_\theta} \end{bmatrix}^T, \begin{bmatrix} 0 \\ \gamma_1 \end{bmatrix}^T, \cdots, \begin{bmatrix} 0 \\ \gamma_{n_\eta} \end{bmatrix}^T \right\}$$

But $\mathcal{S}_{\Psi_{cl}} \subseteq \mathcal{S}_{\Psi_+} = \mathcal{S}_{\Psi_{ol}} \oplus \text{Span}\left\{ \begin{bmatrix} 0 \\ \beta_1 \end{bmatrix}^T, \cdots, \begin{bmatrix} 0 \\ \beta_{n_\theta} \end{bmatrix}^T \right\}$

and $\text{Proj}_{\mathcal{S}_{\Psi_{ol}}} \left\{ \begin{bmatrix} f & 0 \end{bmatrix} \right\} = \text{Proj}_{\mathcal{S}_{\Psi_+}} \left\{ \begin{bmatrix} f & 0 \end{bmatrix} \right\}$

Properties of SISO LTI systems: Open vs closed

Box Jenkins models

$$\mathcal{S}_{\Psi_{ol}} = \text{Span}\left\{ \begin{bmatrix} \alpha_1 \\ 0 \end{bmatrix}^T, \dots, \begin{bmatrix} \alpha_{n_\theta} \\ 0 \end{bmatrix}^T, \begin{bmatrix} 0 \\ \gamma_1 \end{bmatrix}^T, \dots, \begin{bmatrix} 0 \\ \gamma_{n_\eta} \end{bmatrix}^T \right\}$$

$$\mathcal{S}_{\Psi_{cl}} = \text{Span}\left\{ \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix}^T, \dots, \begin{bmatrix} \alpha_{n_\theta} \\ \beta_{n_\theta} \end{bmatrix}^T, \begin{bmatrix} 0 \\ \gamma_1 \end{bmatrix}^T, \dots, \begin{bmatrix} 0 \\ \gamma_{n_\eta} \end{bmatrix}^T \right\}$$

But $\mathcal{S}_{\Psi_{cl}} \subseteq \mathcal{S}_{\Psi_+} = \mathcal{S}_{\Psi_{ol}} \oplus \text{Span}\left\{ \begin{bmatrix} 0 \\ \beta_1 \end{bmatrix}^T, \dots, \begin{bmatrix} 0 \\ \beta_{n_\theta} \end{bmatrix}^T \right\}$

and $\text{Proj}_{\mathcal{S}_{\Psi_{ol}}}\left\{ \begin{bmatrix} f & 0 \end{bmatrix} \right\} = \text{Proj}_{\mathcal{S}_{\Psi_+}}\left\{ \begin{bmatrix} f & 0 \end{bmatrix} \right\}$

$$\text{AsVar}\, J(\hat{\theta}_{cl}) \leq \text{AsVar}\, J(\hat{\theta}_{ol})$$

c.f. Bombois et al and Aguero et al.

Outline

1. Introduction
2. The structure of the asymptotic covariance matrix P
3. A refresher on orthogonal projection
4. A geometric interpretation of P
5. Structural results
6. Analysis of properties of SISO LTI systems
7. **A non-linear sample**
8. Input design
9. Summary

A nonlinear sample: Hammerstein systems

A nonlinear sample: Hammerstein systems

$$y_t = G(q, \beta)x_t(\alpha) + e_t, \quad x_t(\alpha) = X(u_t, \alpha)$$

A nonlinear sample: Hammerstein systems

$$y_t = G(q, \beta)x_t(\alpha) + e_t, \quad x_t(\alpha) = X(u_t, \alpha)$$

Assumptions

- $G(z, \beta_o)$ is in the span of $G'(z, \beta_o)$.
- $x_t(\alpha_o)$ is separable and $X(x, \alpha_o)$ is invertible wrt x.

A nonlinear sample: Hammerstein systems

$$y_t = G(q,\beta)x_t(\alpha) + e_t, \quad x_t(\alpha) = X(u_t,\alpha)$$

Assumptions

- $G(z,\beta_o)$ is in the span of $G'(z,\beta_o)$.
- $x_t(\alpha_o)$ is separable and $X(x,\alpha_o)$ is invertible wrt x.

$$\operatorname{AsVar} G(e^{j\omega_o},\hat{\beta}_N) = \lambda_o \frac{\sum_{k=1}^{n_\beta} |\mathcal{B}_k^x(e^{j\omega_o})|^2}{\Phi_x(e^{j\omega})} + \frac{1}{4}\left(\frac{\mathrm{d}}{\mathrm{d}\alpha}\log\Phi_x(e^{j\omega_o})\right)^T P_\alpha \left(\frac{\mathrm{d}}{\mathrm{d}\alpha}\log\Phi_x(e^{j\omega_o})\right)$$

where $\{\mathcal{B}_k^x\}_{k=1}^{n_\beta}$ is any orthonormal basis for the row space of $R_x G'(\beta_o)$ where R_x is the stable minimum phase spectral factor of Φ_x.

A nonlinear sample: Hammerstein systems

$$y_t = G(q,\beta)x_t(\alpha) + e_t, \quad x_t(\alpha) = X(u_t,\alpha)$$

Assumptions

- $G(z,\beta_o)$ is in the span of $G'(z,\beta_o)$.
- $x_t(\alpha_o)$ is separable and $X(x,\alpha_o)$ is invertible wrt x.

$$\operatorname{AsVar} G(e^{j\omega_o},\hat{\beta}_N) = \lambda_o \frac{\sum_{k=1}^{n_\beta} |\mathcal{B}_k^x(e^{j\omega_o})|^2}{\Phi_x(e^{j\omega})} + \frac{1}{4}\left(\frac{\mathrm{d}}{\mathrm{d}\alpha}\log\Phi_x(e^{j\omega_o})\right)^T P_\alpha \left(\frac{\mathrm{d}}{\mathrm{d}\alpha}\log\Phi_x(e^{j\omega_o})\right)$$

where $\{\mathcal{B}_k^x\}_{k=1}^{n_\beta}$ is any orthonormal basis for the row space of $R_x G'(\beta_o)$ where R_x is the stable minimum phase spectral factor of Φ_x.

Generalization of Ninness and Gibson 2002.

Outline

Open loop input design

$$\min_{\Phi_u} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_u(e^{j\omega})}{\Phi_v(e^{j\omega})} \, d\omega$$

$$\text{s.t. } \operatorname{AsVar} J(\hat{\theta}_N) \leq \frac{1}{\alpha}$$

(α is the precision)

Open loop input design: FIR example (again)

$$\begin{aligned}\operatorname{AsVar}\hat{J} &= \lambda_0 \left\| \Pr_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2 \\ &= \lambda_0 \left\| \Pr_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{1}{R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2\end{aligned}$$

Open loop input design: FIR example (again)

$$\begin{aligned}\operatorname{AsVar}\hat{J} &= \lambda_0 \left\| \Pr_{\mathcal{S}_\Psi}\left(\nabla J(z)\begin{bmatrix}\frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0\end{bmatrix}\right)\right\|^2 \\ &= \lambda_0 \left\| \Pr_{\mathcal{S}_\Psi}\left(\nabla J(z)\begin{bmatrix}\frac{1}{R(z^{-1})} & 0\end{bmatrix}\right)\right\|^2\end{aligned}$$

$$\mathcal{S}_\Psi = \operatorname{span}\left\{\begin{bmatrix}R(z)z^{-1} & 0\end{bmatrix}, \cdots, \begin{bmatrix}R(z)z^{-n} & 0\end{bmatrix}\right\}$$

where $\Phi_u(z) = R(z)R(z^{-1})$.

Open loop input design: FIR example (again)

$$\begin{aligned}\operatorname{AsVar} \hat{J} &= \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2 \\ &= \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{1}{R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2\end{aligned}$$

$$\mathcal{S}_\Psi = \operatorname{span}\left\{ \begin{bmatrix} R(z)z^{-1} & 0 \end{bmatrix}, \cdots, \begin{bmatrix} R(z)z^{-n} & 0 \end{bmatrix} \right\}$$

where $\Phi_u(z) = R(z)R(z^{-1})$.

Recall: $\nabla J(z) = \sum_{k=1}^{\infty} \overline{\frac{\mathrm{d}J(g)}{\mathrm{d}g_k}} z^{-k} = \sum_{k=1}^{\infty} \lambda_k z^{-k}$ ($\lambda_k \in \mathbb{R}$ assumed)

Open loop input design: FIR example (again)

$$\begin{aligned}\operatorname{AsVar}\hat{J} &= \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi}\left(\nabla J(z)\begin{bmatrix}\frac{H(z^{-1})}{S(z^{-1})R(z^{-1})} & 0\end{bmatrix}\right)\right\|^2 \\ &= \lambda_0 \left\| \operatorname{Pr}_{\mathcal{S}_\Psi}\left(\nabla J(z)\begin{bmatrix}\frac{1}{R(z^{-1})} & 0\end{bmatrix}\right)\right\|^2\end{aligned}$$

$$\mathcal{S}_\Psi = \operatorname{span}\left\{\begin{bmatrix}R(z)z^{-1} & 0\end{bmatrix}, \cdots, \begin{bmatrix}R(z)z^{-n} & 0\end{bmatrix}\right\}$$

where $\Phi_u(z) = R(z)R(z^{-1})$.

Recall: $\nabla J(z) = \sum_{k=1}^{\infty} \overline{\frac{\mathrm{d}J(g)}{\mathrm{d}g_k}} z^{-k} = \sum_{k=1}^{\infty} \lambda_k z^{-k}$ ($\lambda_k \in \mathbb{R}$ assumed)

Assume $\sum_{k=-\infty}^{\infty} \lambda_{|k|+1} z^{-k} = Q(z)Q(z^{-1})$

Open loop input design: FIR example (again)

Pick $\Phi_u(z) = \lambda_o \alpha \lambda_1 Q(z) Q(z^{-1})$

Open loop input design: FIR example (again)

Pick $\Phi_u(z) = \lambda_o \alpha \lambda_1 Q(z) Q(z^{-1})$

$$\begin{aligned}
\mathrm{AsVar}\,\hat{J} &= \lambda_0 \left\| \mathrm{Pr}_{\mathcal{S}_\Psi} \left(\nabla J(z) \begin{bmatrix} \frac{1}{R(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2 \\
&= \frac{1}{\alpha\lambda_1} \left\| \mathrm{Pr}_{\mathcal{S}_\Psi} \left(z^{-1} \sum_{k=0}^{\infty} \lambda_{k+1} z^{-k} \begin{bmatrix} \frac{1}{Q(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2 \\
&= \frac{1}{\alpha\lambda_1} \left\| \mathrm{Pr}_{\mathcal{S}_\Psi} \left(z^{-1} \sum_{k=-\infty}^{\infty} \lambda_{k+1} z^{-k} \begin{bmatrix} \frac{1}{Q(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2 \\
&= \frac{1}{\alpha\lambda_1} \left\| \mathrm{Pr}_{\mathcal{S}_\Psi} \left(z^{-1} Q(z) Q(z^{-1}) \begin{bmatrix} \frac{1}{Q(z^{-1})} & 0 \end{bmatrix} \right) \right\|^2 \\
&= \frac{1}{\alpha\lambda_1} \left\| \mathrm{Pr}_{\mathcal{S}_\Psi} \left(\begin{bmatrix} z^{-1} Q(z) & 0 \end{bmatrix} \right) \right\|^2 \\
&= \frac{1}{\alpha\lambda_1} \left\| \begin{bmatrix} z^{-1} Q(z) & 0 \end{bmatrix} \right\|^2 = \frac{1}{\alpha\lambda_1} \left\| Q(z) \right\|^2 = \frac{1}{\alpha}
\end{aligned}$$

FIR example: Summary

FIR example: Summary

Assumption: $\sum_{k=-\infty}^{\infty} \lambda_{|k|+1} z^{-k} = Q(z)Q(z^{-1})$

FIR example: Summary

Assumption: $\sum_{k=-\infty}^{\infty} \lambda_{|k|+1} z^{-k} = Q(z)Q(z^{-1})$

$$\Phi_u(z) = \lambda_o \alpha \lambda_1 Q(z)Q(z^{-1}) \Rightarrow \text{AsVar}\, J(\hat{\theta}_N) \leq \frac{1}{\alpha}$$

FIR example: Summary

Assumption: $\sum_{k=-\infty}^{\infty} \lambda_{|k|+1} z^{-k} = Q(z)Q(z^{-1})$

$$\Phi_u(z) = \lambda_o \alpha \lambda_1 Q(z)Q(z^{-1}) \Rightarrow \mathrm{AsVar}\, J(\hat{\theta}_N) \leq \frac{1}{\alpha}$$

With slightly more effort can show that the above spectrum solves

$$\min_{\Phi_u} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_u(e^{j\omega})}{\Phi_v(e^{j\omega})}\, d\omega, \qquad \text{s.t. } \mathrm{AsVar}\, J(\hat{\theta}_N) \leq \frac{1}{\alpha}$$

FIR example: Summary

Assumption: $\sum_{k=-\infty}^{\infty} \lambda_{|k|+1} z^{-k} = Q(z)Q(z^{-1})$

$$\Phi_u(z) = \lambda_o \alpha \lambda_1 Q(z)Q(z^{-1}) \Rightarrow \text{AsVar}\, J(\hat{\theta}_N) \leq \frac{1}{\alpha}$$

With slightly more effort can show that the above spectrum solves

$$\min_{\Phi_u} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_u(e^{j\omega})}{\Phi_v(e^{j\omega})}\, d\omega, \qquad \text{s.t. } \text{AsVar}\, J(\hat{\theta}_N) \leq \frac{1}{\alpha}$$

- Minimum cost $= \alpha \lambda_1^2$

FIR example: Summary

Assumption: $\sum_{k=-\infty}^{\infty} \lambda_{|k|+1} z^{-k} = Q(z)Q(z^{-1})$

$$\Phi_u(z) = \lambda_o \alpha \lambda_1 Q(z)Q(z^{-1}) \Rightarrow \mathrm{AsVar}\, J(\hat{\theta}_N) \leq \tfrac{1}{\alpha}$$

With slightly more effort can show that the above spectrum solves

$$\min_{\Phi_u} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_u(e^{j\omega})}{\Phi_v(e^{j\omega})}\, d\omega, \qquad \text{s.t. } \mathrm{AsVar}\, J(\hat{\theta}_N) \leq \frac{1}{\alpha}$$

- Minimum cost $= \alpha \lambda_1^2$
- All problems with same dJ/dg_1 have same cost!

FIR example: Summary

Assumption: $\sum_{k=-\infty}^{\infty} \lambda_{|k|+1} z^{-k} = Q(z)Q(z^{-1})$

$\Phi_u(z) = \lambda_o \alpha \lambda_1 Q(z)Q(z^{-1}) \Rightarrow \mathrm{AsVar}\, J(\hat{\theta}_N) \leq \frac{1}{\alpha}$

With slightly more effort can show that the above spectrum solves

$$\min_{\Phi_u} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_u(e^{j\omega})}{\Phi_v(e^{j\omega})}\, d\omega, \qquad \text{s.t. } \mathrm{AsVar}\, J(\hat{\theta}_N) \leq \frac{1}{\alpha}$$

- Minimum cost $= \alpha \lambda_1^2$
- All problems with same dJ/dg_1 have same cost!
- nmp-zero estimates: same cost regardless of model/system order (robustness)!

FIR example: Summary

Assumption: $\sum_{k=-\infty}^{\infty} \lambda_{|k|+1} z^{-k} = Q(z)Q(z^{-1})$

$$\Phi_u(z) = \lambda_o \alpha \lambda_1 Q(z)Q(z^{-1}) \Rightarrow \mathrm{AsVar}\, J(\hat{\theta}_N) \leq \frac{1}{\alpha}$$

With slightly more effort can show that the above spectrum solves

$$\min_{\Phi_u} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_u(e^{j\omega})}{\Phi_v(e^{j\omega})}\, d\omega, \qquad \text{s.t. } \mathrm{AsVar}\, J(\hat{\theta}_N) \leq \frac{1}{\alpha}$$

- Minimum cost $= \alpha \lambda_1^2$
- All problems with same dJ/dg_1 have same cost!
- nmp-zero estimates: same cost regardless of model/system order (robustness)!
- Little or no bias error

FIR example: Summary

Assumption: $\sum_{k=-\infty}^{\infty} \lambda_{|k|+1} z^{-k} = Q(z)Q(z^{-1})$

$$\Phi_u(z) = \lambda_o \alpha \lambda_1 Q(z)Q(z^{-1}) \Rightarrow \operatorname{AsVar} J(\hat{\theta}_N) \leq \frac{1}{\alpha}$$

With slightly more effort can show that the above spectrum solves

$$\min_{\Phi_u} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_u(e^{j\omega})}{\Phi_v(e^{j\omega})} \, d\omega, \qquad \text{s.t. } \operatorname{AsVar} J(\hat{\theta}_N) \leq \frac{1}{\alpha}$$

- Minimum cost $= \alpha \lambda_1^2$
- All problems with same dJ/dg_1 have same cost!
- nmp-zero estimates: same cost regardless of model/system order (robustness)!
- Little or no bias error
- Generalization to other model structures

FIR example: Summary

Assumption: $\sum_{k=-\infty}^{\infty} \lambda_{|k|+1} z^{-k} = Q(z)Q(z^{-1})$

$$\Phi_u(z) = \lambda_o \alpha \lambda_1 Q(z)Q(z^{-1}) \Rightarrow \mathrm{AsVar}\, J(\hat{\theta}_N) \leq \frac{1}{\alpha}$$

With slightly more effort can show that the above spectrum solves

$$\min_{\Phi_u} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_u(e^{j\omega})}{\Phi_v(e^{j\omega})}\, d\omega, \qquad \text{s.t. } \mathrm{AsVar}\, J(\hat{\theta}_N) \leq \frac{1}{\alpha}$$

- Minimum cost $= \alpha \lambda_1^2$
- All problems with same dJ/dg_1 have same cost!
- nmp-zero estimates: same cost regardless of model/system order (robustness)!
- Little or no bias error
- Generalization to other model structures
- Generalization to identification of a frequency band in poster session on Monday (Rojas, Barenthin and Hjalmarsson)

Outline

Summary

Summary

Geometric analysis of the asymptotic variance (assuming only variance errors).

$$\text{AsCov } J(\hat{\theta}_N) = \langle \hat{\gamma}, \hat{\gamma} \rangle = \langle \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\} \rangle$$

Summary

Geometric analysis of the asymptotic variance (assuming only variance errors).

$$\text{AsCov } J(\hat{\theta}_N) = \langle \hat{\gamma}, \hat{\gamma} \rangle = \langle \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\} \rangle$$

Different quantities decoupled!

Summary

Geometric analysis of the asymptotic variance (assuming only variance errors).

$$\mathrm{AsCov}\ J(\hat{\theta}_N) = \langle \hat{\gamma}, \hat{\gamma} \rangle = \langle \mathrm{Proj}_{\mathcal{S}_\Psi}\{\gamma\} \rangle$$

Different quantities decoupled!

- Structural analysis (we never had to compute γ explicitly)
  - Model order
  - Adding excitation sources
  - Adding sensors

Summary

Geometric analysis of the asymptotic variance (assuming only variance errors).

$$\text{AsCov } J(\hat{\theta}_N) = \langle \hat{\gamma}, \hat{\gamma} \rangle = \langle \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\} \rangle$$

Different quantities decoupled!

- Structural analysis (we never had to compute γ explicitly)
  - Model order
  - Adding excitation sources
  - Adding sensors
- Linear systems
  - Explicit expression
  - Rules of thumbs ($\|\nabla J\|^2_{\Phi_v/\Phi_u^r}$)
  - Minimum variance control
  - Closed loop vs open loop identification

Summary

Geometric analysis of the asymptotic variance (assuming only variance errors).

$$\text{AsCov}\ J(\hat{\theta}_N) = \langle \hat{\gamma}, \hat{\gamma} \rangle = \langle \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\} \rangle$$

Different quantities decoupled!

- Structural analysis (we never had to compute γ explicitly)
  - Model order
  - Adding excitation sources
  - Adding sensors
- Linear systems
  - Explicit expression
  - Rules of thumbs ($\|\nabla J\|^2_{\Phi_v/\Phi_u^r}$)
  - Minimum variance control
  - Closed loop vs open loop identification
- Non-linear systems

Summary

Geometric analysis of the asymptotic variance (assuming only variance errors).

$$\text{AsCov } J(\hat{\theta}_N) = \langle \hat{\gamma}, \hat{\gamma} \rangle = \langle \text{Proj}_{\mathcal{S}_\Psi}\{\gamma\} \rangle$$

Different quantities decoupled!

- Structural analysis (we never had to compute γ explicitly)
    - Model order
    - Adding excitation sources
    - Adding sensors
- Linear systems
    - Explicit expression
    - Rules of thumbs ($\|\nabla J\|^2_{\Phi_v/\Phi_u^r}$)
    - Minimum variance control
    - Closed loop vs open loop identification
- Non-linear systems
- Input design

Summary

Main contribution

The methodology!