Krylov-based finite impulse response estimation

October 23, 2025

Fabio Matti (EPFL)

Joint work with Daniel Kressner (EPFL)
and Martin S. Andersen (DTU)

Chen Lujing, Chen Tianshi, Andersen Martin S. Fast kernel-based regularized system identification using Bayesian optimization // IEEE Trans. Automat. Contr. 2025a. 70, 11.

Chen Lujing, Chen Tianshi, Detha Utkarsh, Andersen Martin S. Towards scalable kernel-based regularized system identification // Proc. 62nd IEEE CDC. 2023. 1498–1504.

Chen Tianshi, Ohlsson Henrik, Ljung Lennart. On the estimation of transfer functions, regularizations and Gaussian processes-Revisited // Automatica. 2012. 48, 8. 1525–1535.

Chen Tyler, Hallman Eric. Krylov-aware stochastic trace estimation // SIAM J. Matrix Anal. Appl. 2023. 44, 3. 1218–1244.

Chen Tyler, Huber Caroline, Lin Ethan, Zaid Hajar. Preconditioning without a preconditioner // arXiv:2501.18717. 2025b.

M. F., Andersen Martin S., Kressner Daniel. Krylov-augmented kernel-based linear system identification // In prep. 2025.

Input $\mathbf{u} \in \mathbb{R}^n$

Output $\mathbf{y} \in \mathbb{R}^n$

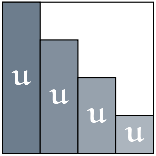
Finite impulse response (FIR) estimation

Input $\mathbf{u} \in \mathbb{R}^n$ FIR $\theta \in \mathbb{R}^m$ $\longrightarrow$ Output $\mathbf{y} \in \mathbb{R}^n$

Linear observation model [Chen et al., 2023]

$$\mathbf{y} = \mathbf{\Phi}\theta + \boldsymbol{e}$$

- $\mathbf{\Phi} = \mathsf{Toeplitz}(\mathbf{u})$ - input with offsets
- $\theta \sim \mathcal{N}(\mathbf{0}, \nu\mathbf{K}(\beta))$ - (unknown) impulse response
- $\mathbf{K}(\beta)$ - spline kernel (e.g. SS, TC/DC [Chen et al., 2012])
- $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ - noise

$\mathbf{\Phi} =$ 

Finite impulse response (FIR) estimation

Input $\mathbf{u} \in \mathbb{R}^n$

FIR $\theta \in \mathbb{R}^m$ $\longrightarrow$

Output $\mathbf{y} \in \mathbb{R}^n$

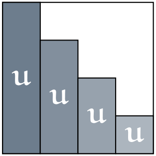
Linear observation model [Chen et al., 2023]

$$\mathbf{y} = \mathbf{\Phi}\theta + \mathbf{e}$$

- $\mathbf{\Phi} = \mathsf{Toeplitz}(\mathbf{u})$ - input with offsets
- $\theta \sim \mathcal{N}(\mathbf{0}, \nu\mathbf{K}(\beta))$ - (unknown) impulse response
- $\mathbf{K}(\beta)$ - spline kernel (e.g. SS, TC/DC [Chen et al., 2012])
- $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ - noise

Goal: determine "best" $\nu, \beta, \sigma^2 > 0$

$\mathbf{\Phi} =$



$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I} + \nu \underbrace{\mathbf{\Phi}\mathbf{K}(\beta)\mathbf{\Phi}^\top}_{=\mathbf{A}(\beta)\succcurlyeq \mathbf{0}})$$

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I} + \nu \underbrace{\mathbf{\Phi} \mathbf{K}(\beta) \mathbf{\Phi}^\top}_{=\mathbf{A}(\beta) \succcurlyeq 0})$$

Negative log-likelihood

With $\lambda = \sigma^2/\nu$ and analytical elimination of ν:

$$\mathcal{L}(\lambda, \beta) = \log(\underbrace{\mathbf{y}^\top (\lambda \mathbf{I} + \mathbf{A}(\beta))^{-1} \mathbf{y}}_{\text{inverse quadratic form (IQF)}}) + \underbrace{\mathrm{Tr}(\log(\lambda \mathbf{I} + \mathbf{A}(\beta)))}_{\text{trace of logarithm (TRACE)}}$$

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I} + \nu \underbrace{\mathbf{\Phi} \mathbf{K}(\beta) \mathbf{\Phi}^\top}_{= \mathbf{A}(\beta) \succcurlyeq 0})$$

Negative log-likelihood

With $\lambda = \sigma^2 / \nu$ and analytical elimination of ν:

$$\mathcal{L}(\lambda, \beta) = \log\Big(\underbrace{\mathbf{y}^\top (\lambda \mathbf{I} + \mathbf{A}(\beta))^{-1} \mathbf{y}}_{\text{inverse quadratic form (IQF)}} \Big) + \underbrace{\mathrm{Tr}(\log(\lambda \mathbf{I} + \mathbf{A}(\beta)))}_{\text{trace of logarithm (TRACE)}}$$

Minimize with Bayesian optimization:

- Evaluate $\mathcal{L}$ for many values of (λ, β)
- Minimum (λ^*, β^*) gives parameters most likely to produce the observed data under the assumed model

Direct method [Chen et al., 2023]

For fixed β and any λ

1. Compute eigendecomposition $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{V}^\top$
2. (IQF) $= \sum_{i=1}^{n} (\lambda + s_i)^{-1} (\boldsymbol{v}_i^\top \boldsymbol{y})^2$
3. (TRACE) $= \sum_{i=1}^{n} \log(\lambda + s_i)$

Direct method [Chen et al., 2023]

For fixed β and any λ

1. Compute eigendecomposition $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{V}^\top$
2. $(\text{IQF}) = \sum_{i=1}^{n} (\lambda + s_i)^{-1} (\boldsymbol{v}_i^\top \boldsymbol{y})^2$
3. $(\text{TRACE}) = \sum_{i=1}^{n} \log(\lambda + s_i)$

Indirect method [Chen et al., 2025a]

For fixed (λ, β)

1. Compute Nyström preconditioner $\mathbf{P}$
2. $(\text{IQF}) \approx \boldsymbol{y}^\top \,\mathsf{LSQR}(\lambda\mathbf{I} + \boldsymbol{A}, \boldsymbol{y}, \mathbf{P})$
3. $(\text{TRACE}) \approx \mathsf{Girard\text{-}Hutchinson}(\log(\lambda\mathbf{I} + \boldsymbol{A}))$

Direct method [Chen et al., 2023]

For fixed β and any λ

1. Compute eigendecomposition $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{V}^\top$
2. (IQF) $= \sum_{i=1}^{n} (\lambda + s_i)^{-1} (\boldsymbol{v}_i^\top \boldsymbol{y})^2$
3. (TRACE) $= \sum_{i=1}^{n} \log(\lambda + s_i)$

Indirect method [Chen et al., 2025a]

For fixed (λ, β)

1. Compute Nyström preconditioner $\mathbf{P}$
2. (IQF) $\approx \boldsymbol{y}^\top \mathsf{LSQR}(\lambda \mathbf{I} + \boldsymbol{A}, \boldsymbol{y}, \mathbf{P})$
3. (TRACE) $\approx \mathsf{Girard\text{-}Hutchinson}(\log(\lambda \mathbf{I} + \boldsymbol{A}))$

Method	scales well with n, m	fast $\lambda \to \mathcal{L}(\lambda, \beta)$
Direct	✗	✓
Indirect	✓	✗

Krylov method [M. et al., 2025]

Approximate in Krylov subspaces:

$$\mathcal{K}_k(\boldsymbol{A}, \mathbf{b}) = \mathsf{span}(\mathbf{b}, \boldsymbol{A}\mathbf{b}, \ldots, \boldsymbol{A}^{k-1}\mathbf{b})$$

Krylov method [M. et al., 2025]
Approximate in Krylov subspaces:

$$\mathcal{K}_k(\boldsymbol{A}, \mathbf{b}) = \mathsf{span}(\mathbf{b}, \boldsymbol{A}\mathbf{b}, \ldots, \boldsymbol{A}^{k-1}\mathbf{b})$$

Convenient property: Shift invariance

$$\mathcal{K}_k(\boldsymbol{A}, \mathbf{b}) = \mathcal{K}_k(\lambda\mathbf{I} + \boldsymbol{A}, \mathbf{b})$$

$\longrightarrow$ “fast $\lambda \to \mathcal{L}(\lambda, \beta)$”

Krylov method [M. et al., 2025]
Approximate in Krylov subspaces:

$$\mathcal{K}_k(\boldsymbol{A}, \mathbf{b}) = \mathsf{span}(\mathbf{b}, \boldsymbol{A}\mathbf{b}, \ldots, \boldsymbol{A}^{k-1}\mathbf{b})$$

Convenient property: Shift invariance

$$\mathcal{K}_k(\boldsymbol{A}, \mathbf{b}) = \mathcal{K}_k(\lambda \mathbf{I} + \boldsymbol{A}, \mathbf{b})$$

$\longrightarrow$ "fast $\lambda \to \mathcal{L}(\lambda, \beta)$"

Lanczos method: Accesses $\boldsymbol{A}$ with matrix-vector products
$\longrightarrow$ "scales well with n, m"

Krylov method [M. et al., 2025]
Approximate in Krylov subspaces:

$$\mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \mathsf{span}(\mathbf{b}, \mathbf{A}\mathbf{b}, \dots, \mathbf{A}^{k-1}\mathbf{b})$$

Convenient property: Shift invariance

$$\mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \mathcal{K}_k(\lambda \mathbf{I} + \mathbf{A}, \mathbf{b})$$

$\longrightarrow$ "fast $\lambda \to \mathcal{L}(\lambda, \beta)$"

Lanczos method: Accesses $\mathbf{A}$ with matrix-vector products
$\longrightarrow$ "scales well with n, m"

Method	scales well with n, m	fast $\lambda \to \mathcal{L}(\lambda, \beta)$
Krylov	✓	✓

Conjugate gradient approximation of (IQF) $= \mathbf{y}^\top(\lambda\mathbf{I} + \mathbf{A})^{-1}\mathbf{y}$

For fixed β

1. Compute orthonormal basis $\mathbf{Q}$ of $\mathcal{K}_k(\mathbf{A}, \mathbf{y})$
2. Compute eigendecomposition of $\mathbf{T} = \mathbf{Q}^\top\mathbf{A}\mathbf{Q} = \mathbf{V}\mathbf{S}\mathbf{V}^\top$
3. Let $\tilde{\mathbf{y}} = \mathbf{V}^\top\mathbf{Q}^\top\mathbf{y}$
4. (IQF) $\approx \mathbf{y}^\top\mathbf{Q}(\lambda\mathbf{I} + \mathbf{T})^{-1}\mathbf{Q}^\top\mathbf{y} = \sum_{i=1}^{k}(\lambda + s_i)^{-1}\tilde{y}_i^2$

Trace estimate of (TRACE) $= \mathsf{Tr}(\log(\lambda \mathbf{I} + \boldsymbol{A}))$
For fixed β

1. Compute orthonormal basis $\mathbf{Q}$ of $\mathcal{K}_k(\boldsymbol{A}, \boldsymbol{\Omega})$, $\boldsymbol{\Omega} \sim \mathcal{N}^{n \times n_\Omega}$
2. (TRACE) $\approx \mathsf{Tr}(\mathbf{Q}^\top \log(\lambda \mathbf{I} + \boldsymbol{A})\mathbf{Q})$ + "correction term"

Trace estimate of (TRACE) $= \mathsf{Tr}(\log(\lambda \mathbf{I} + \boldsymbol{A}))$

For fixed β

1. Compute orthonormal basis $\mathbf{Q}$ of $\mathcal{K}_k(\boldsymbol{A}, \boldsymbol{\Omega})$, $\boldsymbol{\Omega} \sim \mathcal{N}^{n \times n_\Omega}$
2. (TRACE) $\approx \mathsf{Tr}(\mathbf{Q}^\top \log(\lambda \mathbf{I} + \boldsymbol{A})\mathbf{Q}) +$ "correction term"

Quadratic form evaluation [Chen, Hallman, 2023]

To evaluate $\mathbf{Q}^\top \log(\lambda \mathbf{I} + \boldsymbol{A})\mathbf{Q}$, would usually compute

$$\mathcal{K}_l(\boldsymbol{A}, \mathbf{Q}) = \mathsf{span}(\mathbf{Q}, \boldsymbol{A}\mathbf{Q}, \ldots, \boldsymbol{A}^{l-1}\mathbf{Q})$$

Trace estimate of (TRACE) $= \operatorname{Tr}(\log(\lambda \mathbf{I} + \boldsymbol{A}))$

For fixed β

1. Compute orthonormal basis $\mathbf{Q}$ of $\mathcal{K}_k(\boldsymbol{A}, \boldsymbol{\Omega})$, $\boldsymbol{\Omega} \sim \mathcal{N}^{n \times n_\Omega}$
2. (TRACE) $\approx \operatorname{Tr}(\mathbf{Q}^\top \log(\lambda \mathbf{I} + \boldsymbol{A})\mathbf{Q})$ + "correction term"

Quadratic form evaluation [Chen, Hallman, 2023]

To evaluate $\mathbf{Q}^\top \log(\lambda \mathbf{I} + \boldsymbol{A})\mathbf{Q}$, would usually compute

$$\mathcal{K}_l(\boldsymbol{A}, \mathbf{Q}) = \operatorname{span}(\mathbf{Q}, \boldsymbol{A}\mathbf{Q}, \ldots, \boldsymbol{A}^{l-1}\mathbf{Q}) \quad \times \quad lkn_\Omega \text{ elements}$$

Trace estimate of (TRACE) $= \operatorname{Tr}(\log(\lambda \mathbf{I} + \boldsymbol{A}))$

For fixed β

1. Compute orthonormal basis $\mathbf{Q}$ of $\mathcal{K}_k(\boldsymbol{A}, \boldsymbol{\Omega})$, $\boldsymbol{\Omega} \sim \mathcal{N}^{n \times n_\Omega}$
2. (TRACE) $\approx \operatorname{Tr}(\mathbf{Q}^\top \log(\lambda \mathbf{I} + \boldsymbol{A})\mathbf{Q}) +$ "correction term"

Quadratic form evaluation [Chen, Hallman, 2023]

To evaluate $\mathbf{Q}^\top \log(\lambda \mathbf{I} + \boldsymbol{A})\mathbf{Q}$, would usually compute

$$\begin{aligned} \mathcal{K}_l(\boldsymbol{A}, \mathbf{Q}) &= \operatorname{span}(\mathbf{Q}, \boldsymbol{A}\mathbf{Q}, \ldots, \boldsymbol{A}^{l-1}\mathbf{Q}) \quad \times \quad lkn_\Omega \text{ elements} \\ &= \operatorname{span}(\boldsymbol{\Omega}, \boldsymbol{A}\boldsymbol{\Omega}, \ldots, \boldsymbol{A}^{l+k-2}\boldsymbol{\Omega}) \\ &= \mathcal{K}_{l+k-1}(\boldsymbol{A}, \boldsymbol{\Omega}) \quad \checkmark \quad (l+k-1)n_\Omega \text{ elements} \end{aligned}$$

$\implies$ "Krylov-aware trick"

Krylov-aware trace estimate of (TRACE) $= \mathrm{Tr}(\log(\lambda \mathbf{I} + \boldsymbol{A}))$

For fixed β

1. Compute orthonormal basis $\mathbf{Q}$ of $\mathcal{K}_k(\boldsymbol{A}, \boldsymbol{\Omega})$, $\boldsymbol{\Omega} \sim \mathcal{N}^{n \times n_\Omega}$

2.a. Extend orthonormal basis $\mathbf{Q}$ to $\boldsymbol{W}$ of $\mathcal{K}_{l+k-1}(\boldsymbol{A}, \boldsymbol{\Omega})$

2.b. Compute eigendecomposition of $\mathsf{T} = \boldsymbol{W}^\top \boldsymbol{A} \boldsymbol{W} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{V}^\top$

2.c. (TRACE) $\approx \mathrm{Tr}(\mathbf{Q}^\top \boldsymbol{W} \log(\lambda \mathbf{I} + \mathsf{T}) \boldsymbol{W}^\top \mathbf{Q})$
$= \sum_{i=1}^{k n_\Omega} \log(\lambda + s_i) \|\mathsf{V}(1 : k n_\Omega, i)\|_2^2$

Augmented Krylov subspace

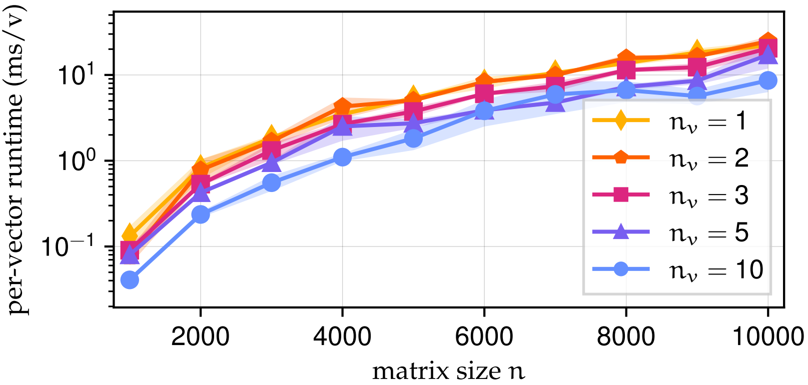
$$\mathcal{K}_k(\mathbf{A}, \mathbf{y}) + \mathcal{K}_k(\mathbf{A}, \boldsymbol{\Omega}) = \mathcal{K}_k(\mathbf{A}, [\mathbf{y}, \boldsymbol{\Omega}])$$

Augmented Krylov subspace

$$\mathcal{K}_k(\mathbf{A}, \mathbf{y}) + \mathcal{K}_k(\mathbf{A}, \mathbf{\Omega}) = \mathcal{K}_k(\mathbf{A}, [\mathbf{y}, \mathbf{\Omega}])$$

Can "bundle" matrix-vector products in a block

$$\mathbf{A}\mathbf{v}_1, \ldots, \mathbf{A}\mathbf{v}_{n_v} \quad \longrightarrow \quad \mathbf{A}[\mathbf{v}_1, \ldots, \mathbf{v}_{n_v}]$$



Standard result

$$\text{relative error of (IQF) approximation} \leqslant 4\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{2k}$$

$\kappa =$ condition number of $\lambda\mathbf{I} + \boldsymbol{A}$

Standard result

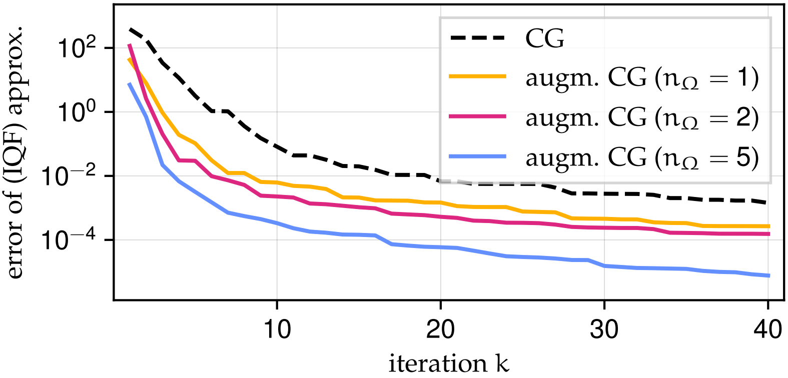
$$\text{relative error of (IQF) approximation} \leqslant 4\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{2k}$$

$\kappa =$ condition number of $\lambda\mathbf{I} + \mathbf{A}$

Theorem: Augmentation helps [M. et al., 2025]

Because $\mathcal{K}_k(\mathbf{A}, \mathbf{y}) \subseteq \mathcal{K}_k(\mathbf{A}, [\mathbf{y}, \mathbf{\Omega}])$

error of *augmented* (IQF) approx. $\leqslant$ error of (IQF) approx.


error of (IQF) approx.
10^{2}
10^{0}
10^{-2}
10^{-4}
10
20
30
40
iteration k
CG
augm. CG ($n_\Omega = 1$)
augm. CG ($n_\Omega = 2$)
augm. CG ($n_\Omega = 5$)

Theorem: Implicit preconditioning [M. et al., 2025]

Because $\mathcal{K}_k(\mathbf{P}^{-1}\mathbf{A}, \mathbf{P}^{-1}\mathbf{y}) \subseteq \mathcal{K}_{k+s}(\mathbf{A}, [\mathbf{y}, \mathbf{\Omega}])$ [Chen et al., 2025b]

$$\text{rel. error of } \textit{augmented} \text{ (IQF) approx.} \leqslant 4\left(\frac{\sqrt{\widetilde{\kappa}}-1}{\sqrt{\widetilde{\kappa}}+1}\right)^{2(k-s)}$$

$\widetilde{\kappa}$ = condition number of $\mathbf{P}^{-1}(\lambda\mathbf{I}+\mathbf{A})$ for any $\mathbf{P} = (\mathbf{I}+\mathbf{X})^{-1}$ with $\mathsf{range}(\mathbf{X}) \subseteq \mathcal{K}_{s+1}(\mathbf{A}, \mathbf{\Omega})$, $s \leqslant k$.

E.g. $\mathbf{P}$ = Nyström preconditioner [Chen et al., 2025b]

Test systems:

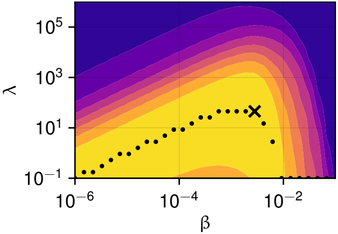
- $\mathbf{u} \in \mathbb{R}^{10^4}$ white Gaussian noise (input)
- $\mathbf{y} \in \mathbb{R}^{10^4}$ with transfer function $G(z) = (1 - 0.2z^{-1})^{-2}$ and add white Gaussian noise such that SNR = 10 (output)

Test systems:

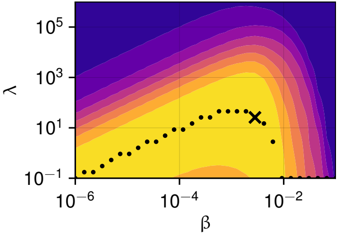
- $u \in \mathbb{R}^{10^4}$ white Gaussian noise (input)
- $y \in \mathbb{R}^{10^4}$ with transfer function $G(z) = (1 - 0.2z^{-1})^{-2}$ and add white Gaussian noise such that SNR = 10 (output)

Parameters of Krylov method:

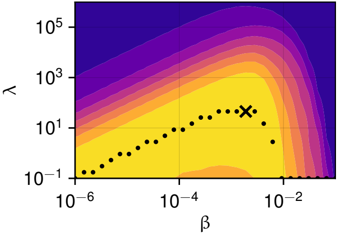
- $m = 2 \times 10^3$ (length of FIR)
- $k = 40$ (number of block Lanczos iterations)
- $n_\Omega = 1$ (augmentation size)



(A) Direct method.



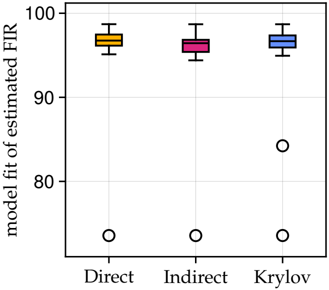
(B) Indirect method.



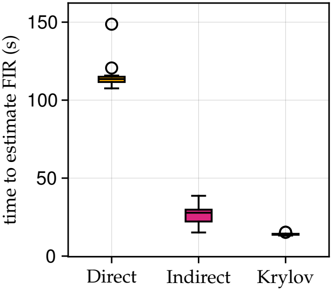
(C) Krylov method.

	runtime (s)
Direct method	67.24
Indirect method	404.62
Krylov method	9.31

(D) Runtimes of the methods.



(A) Model fit of the methods.



(B) Runtime of the methods.

The Krylov method exploits the shift-invariance and nestedness of Krylov subspaces to produce fast and accurate estimates of FIRs

Questions

We estimate

$$\mathsf{Tr}(\log(\lambda \mathbf{I} + \boldsymbol{A})) \approx \mathsf{Tr}(\mathbf{Q}^\top \log(\lambda \mathbf{I} + \boldsymbol{A})\mathbf{Q}^\top) + \text{“correction term”}$$

Correction term is trace estimate of residual

$$\begin{aligned} \mathsf{Tr}(\mathbf{B}) &= \mathsf{Tr}(\mathbf{Q}\mathbf{Q}^\top \mathbf{B}) + \mathsf{Tr}((\mathbf{I} - \mathbf{Q}\mathbf{Q}^\top)\mathbf{B}) \\ &= \mathsf{Tr}(\mathbf{Q}^\top \mathbf{B}\mathbf{Q}) + \mathsf{Tr}((\mathbf{I} - \mathbf{Q}\mathbf{Q}^\top)\mathbf{B}(\mathbf{I} - \mathbf{Q}\mathbf{Q}^\top)) \\ &\approx \mathsf{Tr}(\mathbf{Q}^\top \mathbf{B}\mathbf{Q}) + \underbrace{\frac{1}{n_\psi}\sum_{i=1}^{n_\psi} \boldsymbol{\psi}^\top \mathbf{B}\boldsymbol{\psi}}_{\text{“correction term”}} \end{aligned}$$

with $\boldsymbol{\psi} = (\mathbf{I} - \mathbf{Q}\mathbf{Q}^\top)\boldsymbol{z}$ for $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Conjecture: Augmentation helps

error of *augmented* (TRACE) approx. $\leqslant$ error of (TRACE) approx.

- Seems to hold numerically
- Theoretically not yet shown

Algorithm: Krylov-augmented FIR estim. [M. et al., 2025]

Input: Input signal $\mathbf{u} \in \mathbb{R}^n$ and output signal $\mathbf{y} \in \mathbb{R}^n$
Parameters: FIR length m, iterations k, augmentation size n_Ω
Output: Maximum likelihood estimate of FIR θ^*

Algorithm: Krylov-augmented FIR estim. [M. et al., 2025]

Input: Input signal $\mathbf{u} \in \mathbb{R}^n$ and output signal $\mathbf{y} \in \mathbb{R}^n$
Parameters: FIR length m, iterations k, augmentation size n_Ω
Output: Maximum likelihood estimate of FIR θ^*

Form Toepliz matrix $\mathbf{\Phi} = \text{Toepliz}(\mathbf{u})$

Algorithm: Krylov-augmented FIR estim. [M. et al., 2025]

Input: Input signal $\mathbf{u} \in \mathbb{R}^n$ and output signal $\mathbf{y} \in \mathbb{R}^n$
Parameters: FIR length m, iterations k, augmentation size n_Ω
Output: Maximum likelihood estimate of FIR θ^*

Form Toepliz matrix $\mathbf{\Phi} = \text{Toepliz}(\mathbf{u})$
Optimize in β by evaluating $\mathcal{L}(\lambda, \beta)$ with the Krylov method:

1. Sample Gaussian random matrix $\mathbf{\Omega} \in \mathbb{R}^{n \times n_\Omega}$
2. $\mathbf{Q}, \mathbf{T} = \text{Block-Lanczos}(\mathbf{\Phi}\mathbf{K}(\beta)\mathbf{\Phi}^\top, [\mathbf{y}, \mathbf{\Omega}], k)$
3. $\mathbf{s}, \mathbf{V} = \text{eig}(\mathbf{T})$
4. $\boldsymbol{w} = \|\mathbf{y}\|_2^2 \mathbf{V}(1,:)^2$ and $\widetilde{\boldsymbol{w}} = \sum_{j=1}^{kn_\Omega} \mathbf{V}(j,:)^2$
5. Minimize $\lambda \mapsto \log(\sum_i (\lambda + s_i)^{-1} w_i) + \sum_i \log(\lambda + s_i)\widetilde{w}_i$

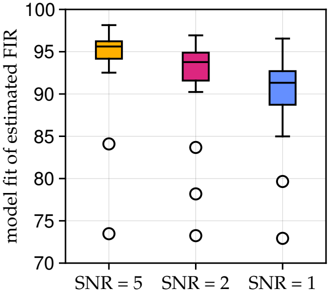
Algorithm: Krylov-augmented FIR estim. [M. et al., 2025]

Input: Input signal $\mathbf{u} \in \mathbb{R}^n$ and output signal $\mathbf{y} \in \mathbb{R}^n$
Parameters: FIR length m, iterations k, augmentation size n_Ω
Output: Maximum likelihood estimate of FIR θ^*

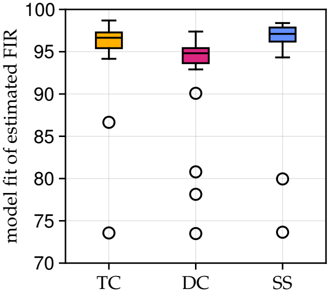
Form Toepliz matrix $\boldsymbol{\Phi} = \text{Toepliz}(\mathbf{u})$
Optimize in β by evaluating $\mathcal{L}(\lambda, \beta)$ with the Krylov method:

1. Sample Gaussian random matrix $\boldsymbol{\Omega} \in \mathbb{R}^{n \times n_\Omega}$
2. $\mathbf{Q}, \mathbf{T} = \text{Block-Lanczos}(\boldsymbol{\Phi}\mathbf{K}(\beta)\boldsymbol{\Phi}^\top, [\mathbf{y}, \boldsymbol{\Omega}], k)$
3. $\mathbf{s}, \mathbf{V} = \text{eig}(\mathbf{T})$
4. $\boldsymbol{w} = \|\mathbf{y}\|_2^2 \mathbf{V}(1,:)^2$ and $\widetilde{\boldsymbol{w}} = \sum_{j=1}^{kn_\Omega} \mathbf{V}(j,:)^2$
5. Minimize $\lambda \mapsto \log(\sum_i (\lambda + s_i)^{-1} w_i) + \sum_i \log(\lambda + s_i)\widetilde{w}_i$

Return $\theta^* = \mathbf{K}(\beta^*)\boldsymbol{\Phi}^\top \mathbf{Q}(\lambda^* \mathbf{I} + \mathbf{T})^{-1}\mathbf{Q}^\top \mathbf{y}$



(A) Smaller SNR.



(B) Different kernels.