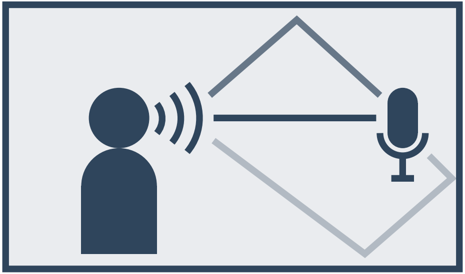
Linear system identification using augmented Krylov subspace methods

February 13, 2026

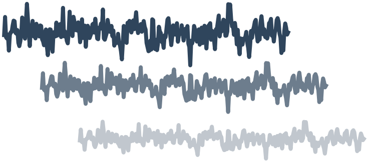
Fabio Matti (EPFL)

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

Example: Speech dereverberation



Input $\mathbf{u} \in \mathbb{R}^m$ (Speaker)



Output $\mathbf{y} \in \mathbb{R}^m$ (Microphone)

Finite impulse response (FIR) model

(unknown) impulse response

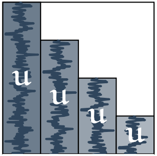
↓

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

↑

noise

$$\mathbf{\Phi} = \text{Toeplitz}(\mathbf{u}) =$$



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

Assumption:

- $\theta \sim \mathcal{N}(\mathbf{0}, \nu\mathbf{K}(\beta))$ - (unknown) impulse response
- $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ - noise

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

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

Maximum likelihood estimation

Negative log-likelihood of $\boldsymbol{y}$

With $\lambda = \sigma^2/\nu$:

$$\mathcal{L}(\lambda, \beta) = \log\Big(\underbrace{\boldsymbol{y}^\top (\lambda \mathbf{I} + \boldsymbol{A}(\beta))^{-1} \boldsymbol{y}}_{\text{inverse quadratic form (IQF)}} \Big) + \underbrace{\log(\det(\lambda \mathbf{I} + \boldsymbol{A}(\beta)))}_{\text{log-determinant (LOGDET)}}$$

for the symmetric positive semidefinite matrix $\boldsymbol{A}(\beta)$.

- Minimum (λ^*, β^*) gives "best" fitting parameters
- Optimizer needs to evaluate $\mathcal{L}$ for many pairs (λ, β)

Fast and scalable evaluation of $\mathcal{L}$

Existing methods for evaluating $\mathcal{L} = \text{IQF} + \text{LOGDET}$

Direct method [Chen et al., 2023]

For fixed β and any λ

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

Indirect method [Chen et al., 2025a]

For fixed (λ, β)

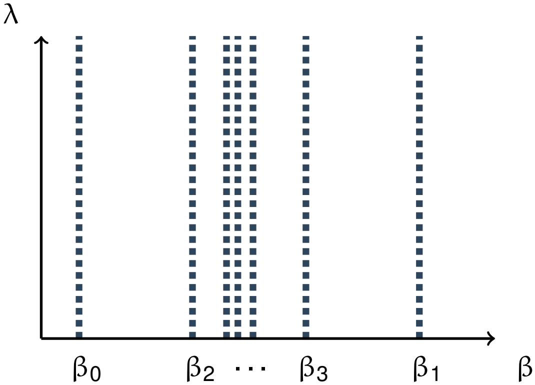
1. Compute Nyström preconditioner $\mathbf{P}$ for $\lambda \mathbf{I} + \boldsymbol{A}(\beta)$
2. (IQF) $\approx \boldsymbol{y}^\top \mathsf{LSQR}(\lambda \mathbf{I} + \boldsymbol{A}(\beta), \boldsymbol{y}, \mathbf{P})$
3. (LOGDET) $\approx \mathsf{Girard\text{-}Hutchinson}(\log(\lambda \mathbf{I} +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

Method	scales well with size of $\boldsymbol{A}(\beta)$	fast $\lambda \to \mathcal{L}(\lambda, \beta)$
Direct	✗	✓
Indirect	✓	✗

Approach: Fix β and use a Krylov subspace method for $\mathbf{A} = \mathbf{A}(\beta)$ to exploit the shift-invariance with λ:

$$\mathcal{K}_k(\lambda \mathbf{I} + \mathbf{A}, \mathbf{b}) = \mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \operatorname{span}\left\{\mathbf{b}, \mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, \ldots, \mathbf{A}^{k-1}\mathbf{b}\right\}$$

Optimization process:



$$\mathcal{L}(\lambda) = \log(\mathbf{y}^\top (\lambda \mathbf{I} + \mathbf{A})^{-1}\mathbf{y}) + \log(\det(\lambda \mathbf{I} + \mathbf{A}))$$

Let $\mathbf{Q}$ orthonormal basis of

$$\mathcal{K}_k(\mathbf{A}, \mathbf{y}) = \operatorname{span}\left\{\mathbf{y}, \mathbf{A}\mathbf{y}, \mathbf{A}^2\mathbf{y}, \ldots, \mathbf{A}^{k-1}\mathbf{y}\right\}$$

and $\mathbf{T} = \mathbf{Q}^\top \mathbf{A}\mathbf{Q}$.

Definition: Inverse quadratic form approximation

Standard approximation

$$\mathbf{y}^\top (\lambda \mathbf{I} + \mathbf{A})^{-1}\mathbf{y} \approx \mathbf{y}^\top \mathbf{Q}(\lambda \mathbf{I} + \mathbf{T})^{-1}\mathbf{Q}^\top \mathbf{y}$$
$$= \sum_{i=1}^{n} (\mathbf{Q}^\top \mathbf{y})_i^2 (\lambda + \lambda_i(\mathbf{T}))^{-1}$$

Computing the IQF (convergence)

Lemma: IQF approximation bound [M. et al., 2026]

$$\left| \frac{\mathbf{y}^\top (\lambda \mathbf{I} + \mathbf{A})^{-1} \mathbf{y} - \mathbf{y}^\top \mathbf{Q} (\lambda \mathbf{I} + \mathbf{T})^{-1} \mathbf{Q}^\top \mathbf{y}}{\mathbf{y}^\top (\lambda \mathbf{I} + \mathbf{A})^{-1} \mathbf{y}} \right| \leqslant 4 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^{2k}$$

where κ is the condition number of $\mathbf{A}$.

$$\mathcal{L}(\lambda, \beta) = \log(\mathbf{y}^\top(\lambda\mathbf{I} + \mathbf{A})^{-1}\mathbf{y}) + \log(\det(\lambda\mathbf{I} + \mathbf{A}))$$

Let $\mathbf{\Omega}$ Gaussian random matrix and $\mathbf{Q}$ orthonormal basis of

$$\mathcal{K}_k(\mathbf{A}, \mathbf{\Omega}) = \text{span}\left\{\mathbf{\Omega}, \mathbf{A}\mathbf{\Omega}, \mathbf{A}^2\mathbf{\Omega}, \ldots, \mathbf{A}^{k-1}\mathbf{\Omega}\right\}$$

and $\mathbf{T} = \mathbf{Q}^\top\mathbf{A}\mathbf{Q}$.

Definition: Log-det approximation [Li, Zhu, 2021]

Standard approximation

$$\begin{aligned}\log(\det(\lambda\mathbf{I} + \mathbf{A})) &\approx \log(\det(\lambda\mathbf{I} + \mathbf{Q}\mathbf{T}\mathbf{Q}^\top)) \\ &= (m-n)\log(\lambda) + \sum_{i=1}^{n}\log(\lambda + \lambda_i(\mathbf{T}))\end{aligned}$$

Lemma: LOGDET approximation bound [M. et al., 2026]

$$\left|\log(\det(\lambda\mathbf{I} + \boldsymbol{A})) - \log(\det(\lambda\mathbf{I} + \mathbf{QTQ}^\top))\right| \leqslant \sum_{i=n}^{m} \log(1 + \frac{\lambda_i(\boldsymbol{A})}{\lambda}) + \dots$$

What if we combine/augment the Krylov subspaces

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

and take $\widehat{\mathbf{Q}}$ as ONB of $\mathcal{K}_k(\mathbf{A}, [\mathbf{y}, \mathbf{\Omega}])$ and $\widehat{\mathbf{T}} = \widehat{\mathbf{Q}}^\top \mathbf{A} \widehat{\mathbf{Q}}$?

Benefits:

- Less matrix loads
- Simpler to implement
- Higher accuracy*

Analysis

Since $\mathcal{K}_k(\boldsymbol{A}, \mathbf{y}) \subseteq \mathcal{K}_k(\boldsymbol{A}, [\mathbf{y}, \boldsymbol{\Omega}]) \implies$ "augmented" estimator $\mathbf{y}^\top \widehat{\mathbf{Q}}(\lambda \mathbf{I} + \widehat{\mathbf{T}})^{-1} \widehat{\mathbf{Q}}^\top \mathbf{y}$ is always better [M. et al., 2026]

Since $\mathcal{K}_k(\boldsymbol{A}, \boldsymbol{\Omega}) \subseteq \mathcal{K}_k(\boldsymbol{A}, [\mathbf{y}, \boldsymbol{\Omega}]) \implies$ "augmented" estimator $\mathsf{log}(\mathsf{det}(\lambda \mathbf{I} + \widehat{\mathbf{Q}} \widehat{\mathbf{T}} \widehat{\mathbf{Q}}^\top))$ is always better [M. et al., 2026]

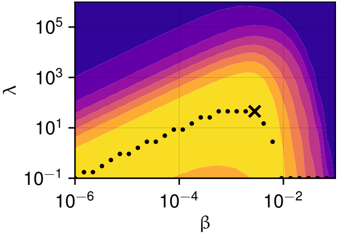
Improved convergence with augmentation

$$\left| \frac{\mathbf{y}^\top (\lambda \mathbf{I} + \mathbf{A})^{-1} \mathbf{y} - \mathbf{y}^\top \widehat{\mathbf{Q}} (\lambda \mathbf{I} + \widehat{\mathbf{T}})^{-1} \widehat{\mathbf{Q}}^\top \mathbf{y}}{\mathbf{y}^\top (\lambda \mathbf{I} + \mathbf{A})^{-1} \mathbf{y}} \right| \leqslant 4 \left(\frac{\sqrt{\widehat{\kappa}} - 1}{\sqrt{\widehat{\kappa}} + 1} \right)^{2k}$$

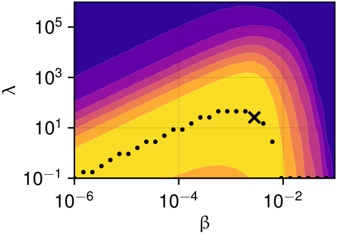
where $\widehat{\kappa}$ is the condition number of $\mathbf{P}^{-1}\mathbf{A}$ for any $\mathbf{P} = (\mathbf{I} + \mathbf{X})^{-1}$ with $\mathsf{range}(\mathbf{X}) \subseteq \mathcal{K}_{s+1}(\mathbf{A}, \mathbf{\Omega})$, $s \leqslant k$ [Chen et al., 2025b].

Example: $\mathbf{P}$ is Nyström preconditioner [Frangella et al., 2023].

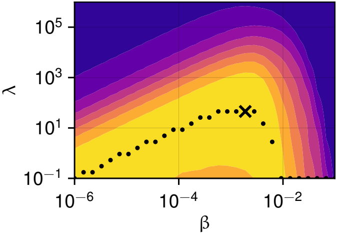
Experiments



(A) Direct method.



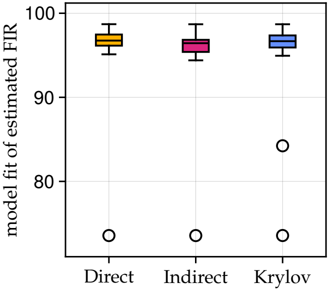
(B) Indirect method.



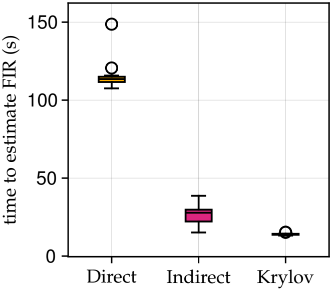
(C) Krylov method.

	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.

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.

Frangella Zachary, Tropp Joel A., Udell Madeleine. Randomized Nyström preconditioning // SIAM Journal on Matrix Analysis and Applications. 2023. 44, 2. 718–752.

Li Hanyu, Zhu Yuanyang. Randomized block Krylov subspace methods for trace and log-determinant estimators // BIT Numerical Mathematics. 2021. 61, 3. 911–939.

M. F., Andersen Martin S., Kressner Daniel. Krylov-augmented kernel-based linear system identification // In prep. 2026.