A Novel Schur Decomposition-Based Weight Projection Method for Stable State-Space Neural-Network Architectures

Sergio Vanegas[1], Lasse Lensu[1], and Fredy Ruiz[2]

[1]Lappeenranta-Lahti University of Technology; [2]Politecnico di Milano

Code

LUT University · Politecnico Milano 1863

seed Systems and Engineering Science Doctorate programme

Co-funded by the European Union

Background

Recent developments in State-Space (SS) SysID theory have sparked renewed interest within the ML community. The SS framework models the sampled output $\boldsymbol{y}[k] \in \mathbb{R}^{n_y}, k \in \mathbb{N}_0$ of a real Linear Time-Invariant (LTI) system with exogenous input $\boldsymbol{u}[k] \in \mathbb{R}^{n_u}$ as a recurrence of the form

$$\begin{cases} \boldsymbol{x}[k+1] &= \boldsymbol{A}\boldsymbol{x}[k] + \boldsymbol{B}\boldsymbol{u}[k] \\ \boldsymbol{y}[k] &= \boldsymbol{C}\boldsymbol{x}[k] + \boldsymbol{D}\boldsymbol{u}[k] \end{cases} \quad s.t. \quad \begin{matrix} \boldsymbol{A} \in \mathbb{R}^{n_x \times n_x}, \ \boldsymbol{B} \in \mathbb{R}^{n_x \times n_u}, \\ \boldsymbol{C} \in \mathbb{R}^{n_y \times n_x}, \ \boldsymbol{D} \in \mathbb{R}^{n_y \times n_u}. \end{matrix} \tag{1}$$

- Albeit linear, the versatility and expressiveness of SS subsystems makes them ideal building blocks for modern sequence-modelling architectures [1], [2], [3].
- However, large architectures introduce overhead prohibitive for time-sensitive applications [4].
- As a result, backpropagation-ready stable matrix parameterizations [5] and weight regularization techniques [6] targeting smaller-scale nonlinear systems (e.g., those of Figure 1) have emerged.

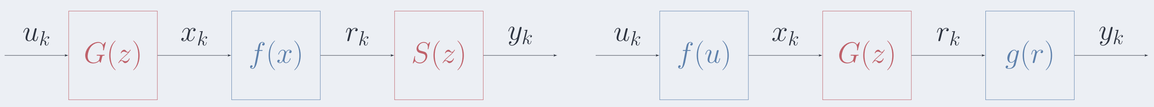


Fig. 1: Single-branch structures considered in Schoukens and Tiels [7]: Wiener-Hammerstein (WH, left) and Hammerstein-Wiener (HW, right) models. Dynamic LTI and static nonlinear blocks are outlined in blue and red, respectively.

Contributions	1. An innovative Schur-stable state-matrix weight-projection scheme; 2. An alternative stable state-matrix parameterization; 3. A comprehensive benchmark of the proposed methodologies.
Baselines	• System Identification Methods leveraging Backpropagation (SIMBa), from Di Natale et al. [5]. • Piecewise spectral-norm regularization, from Bemporad [6].

Nearest Ω-Stable Matrix Identification

Definition 1 (Schur Stability) *A system is deemed Schur-stable iff* $|\lambda_i| \leq 1 \quad \forall \lambda_i \in \Lambda_{\boldsymbol{A}}$, *where* $\Lambda_{\boldsymbol{A}} \in \mathbb{C}^{n_x}$ *denotes the spectrum of* $\boldsymbol{A}$.

Definition 2 (Nearest Ω-stable matrix [8]) *Let* Ω *be a non-empty closed subset of* $\mathbb{C}$, *and let*

$$S(\Omega, n, \mathbb{F}) \triangleq \left\{ \boldsymbol{X} \in \mathbb{F}^{n \times n} : \Lambda_{\boldsymbol{X}} \subseteq \Omega \right\} \subseteq \mathbb{F}^{n \times n} \tag{2}$$

be the set of $n \times n$ *matrices (with* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*) whose eigenvalues all belong to* Ω.
Given $\boldsymbol{A} \in \mathbb{F}^{n \times n}$, *the nearest* Ω*-stable matrix* $\hat{\boldsymbol{A}}$ *is defined as*

$$\hat{\boldsymbol{A}} \triangleq \underset{\boldsymbol{X} \in S(\Omega, n, \mathbb{F})}{\arg\min} \ \|\boldsymbol{A} - \boldsymbol{X}\|_F^2, \tag{3}$$

where $\|\boldsymbol{X}\|_F^2 \triangleq \sum_{i,j=1}^{n} |X_{ij}|^2 = \operatorname{tr}(\boldsymbol{X}^* \boldsymbol{X})$ *denotes the Frobenius norm.*

Corollary 3 *For the Schur Stability case,* $\Omega = \Omega_S \triangleq \{\lambda \in \mathbb{C} : |\lambda| \leq 1\}$.

Theorem 4 (Schur Decomposition [9]) *Given* $\boldsymbol{X} \in \mathbb{R}^{n \times n}$, *there exists a pair of matrices* $\boldsymbol{Z} \in O(n) \triangleq \{\boldsymbol{W} \in \mathbb{R}^{n \times n} : \boldsymbol{W}^\intercal \boldsymbol{W} = \boldsymbol{W}\boldsymbol{W}^\intercal = \boldsymbol{I}_n\}$ *real orthogonal and* $\boldsymbol{T} \in \mathbb{R}^{n \times n}$ *upper quasi-triangular defining the Schur decomposition; i.e., a pair such that*

$$\boldsymbol{Z}^\intercal \boldsymbol{X} \boldsymbol{Z} = \boldsymbol{T} = \begin{bmatrix} \boldsymbol{T}_{1,1} & \cdots & \boldsymbol{T}_{1,k} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{T}_{k,k} \end{bmatrix}, \tag{4}$$

where $\boldsymbol{T}_{i,i} \in \mathbb{R}^{1\times 1} \cup \mathbb{R}^{2\times 2}$ *and* $\left\lceil \frac{n}{2} \right\rceil \leq k \leq n$.

- The non-convexity of Ω_S makes the problem in Definition 2 notoriously hard to solve.
- Noferini and Poloni [8] provided a solution with such a guarantee by proposing a bilevel optimization over the real Schur decomposition of $\hat{\boldsymbol{A}}$ (see Theorem 4)
  1. The block-diagonal elements of $\boldsymbol{T}$ are projected to the nearest element in a set $\hat{\mathcal{S}}(\boldsymbol{T}_{i,i}) \subset \mathbb{R}^{2\times 2}$ of 15 elements at most;
  2. The orthogonal factor $\boldsymbol{Z}$ is optimized over $SO(2) \triangleq \{\boldsymbol{X} \in O(2) : \det(\boldsymbol{X}) = 1\}$.
- This method provides rigorous asymptotic-stability guarantees, but requires solving a computationally expensive, numerically sensitive surrogate optimization problem.

Truncated Schur-Stable Projection

Algorithm 1: Quasi-triangular Schur factor projection.

Input: Schur factor $\boldsymbol{T} \in \mathbb{R}^{n \times n}$ (quasi-triangular)
Output: Stable Schur factor $\hat{\boldsymbol{T}} \in \mathbb{R}^{n \times n}$
begin
  Initialize $i \leftarrow 1, \hat{\boldsymbol{T}} \leftarrow \boldsymbol{T}$;
  for $i \leftarrow 1$ *to* k **do** `// k from Theorem 4`
    if $\boldsymbol{T}_{i,i} \in \mathbb{R}^{1\times 1}$ **then**
      $\hat{\boldsymbol{T}}_{i,i} \leftarrow \boldsymbol{T}_{i,i} / \max\{1, |\boldsymbol{T}_{i,i}|\}$;
    else `// 2 × 2 block`
      Initialize $\boldsymbol{X} \leftarrow \boldsymbol{0}_{2\times 2}, d^2_{\min} \leftarrow \infty$;
      foreach $\boldsymbol{X'} \in \hat{\mathcal{S}}(\boldsymbol{T}_{i,i})$ **do**
        $d^2 \leftarrow \|\boldsymbol{T}_{i,i} - \boldsymbol{X'}\|_F^2$;
        if $d^2 \geq d^2_{\min}$ **then continue**;
        $\lambda_1, \lambda_2 \leftarrow \Lambda_{\boldsymbol{X'}}$;
        if $|\lambda_1| > 1$ **or** $|\lambda_2| > 1$ **then continue**;
        $d^2_{\min} \leftarrow d^2$; $\boldsymbol{X} \leftarrow \boldsymbol{X'}$
      $\hat{\boldsymbol{T}}_{i,i} \leftarrow \boldsymbol{X}$

- Search space restricted to stable matrices sharing the same orthogonal factor:
  1. The Schur decomposition $\boldsymbol{A} = \boldsymbol{Z}\boldsymbol{T}\boldsymbol{Z}^\intercal$ is calculated;
  2. The projection scheme in Algorithm 1 is applied to the quasi-triangular factor $\boldsymbol{T}$;
  3. The stable state-transition matrix is reconstructed as $\hat{\boldsymbol{A}} = \boldsymbol{Z}\hat{\boldsymbol{T}}\boldsymbol{Z}$.
- Alternatively, $\boldsymbol{A}$ can be parameterized directly through its Schur factors, exchanging the decomposition overhead for having to train twice the number of coefficients (orthogonal factor constrained via the Singular-Value Decomposition (SVD) method from Golub and Van Loan [10]).

Numerical Results

Dataset	Metric	Schur (proj.)	Schur (built)	SIMBa	Regularized
Silverbox	Test NMSE	(1.52 ± 7.86)e-2	(1.51 ± 7.87)e-2	$\mathbf{(1.50 \pm 7.97)e-2}$	(1.53 ± 7.49)e-2
	Epoch Time	(1.59 ± 0.02)e-1	(1.42 ± 0.01)e-1	$\mathbf{(1.39 \pm 0.06)e-1}$	(1.43 ± 0.03)e-1
	Epoch Count	8.36e3	**6.89e3**	1.40e4	1.05e4
CED	Test NMSE	$\mathbf{(3.30 \pm 7.40)e-2}$	(3.70 ± 8.79)e-2	(4.12 ± 9.92)e-2	(4.08 ± 8.49)e-2
	Epoch Time	$\mathbf{(6.14 \pm 0.27)e-3}$	(6.80 ± 0.16)e-3	(6.29 ± 0.15)e-3	(6.56 ± 0.10)e-3
	Epoch Count	1.14e4	**7.83e3**	2.65e4	1.75e4
EMPS	Test NMSE	(3.21 ± 3.24)e-3	$\mathbf{(3.07 \pm 4.51)e-3}$	(4.86 ± 5.78)e-3	(2.50 ± 5.18)e1
	Epoch Time	(7.49 ± 0.42)e-3	(7.46 ± 0.42)e-3	(6.51 ± 0.10)e-3	$\mathbf{(6.32 \pm 0.09)e-3}$
	Epoch Count	3.16e4	1.33e4	**1.31e4**	4.78e4
Industrial Robot	Test NMSE	(6.63 ± 12.3)e-1	$\mathbf{(5.76 \pm 10.9)e-1}$	(7.25 ± 17.5)e-1	(7.40 ± 17.5)e-1
	Epoch Time	$\mathbf{(1.78 \pm 0.01)e-1}$	(1.83 ± 0.01)e-1	(1.82 ± 0.04)e-1	(1.86 ± 0.02)e-1
	Epoch Count	1.25e4	**1.20e4**	2.13e4	2.95e4
Fine-Steering Mirror	Test NMSE	(4.76 ± 9.31)e-2	(3.96 ± 7.76)e-2	$\mathbf{(3.62 \pm 6.68)e-2}$	(3.99 ± 7.82)e-2
	Epoch Time	(3.03 ± 0.04)e-1	(3.02 ± 0.04)e-1	$\mathbf{(2.43 \pm 0.02)e-1}$	(3.03 ± 0.25)e-1
	Epoch Count	2.91e4	**2.33e4**	2.68e4	4.54e4

Table 1: Numerical results for the real-data benchmarks (HW models). Error uncertainty is defined as the standard deviation of the normalized squared differences. Median epoch time in seconds (uncertainty defined as half of the Inter-Quantile Range (IQR)). Epoch count reported for the best-performing model. Row-wise best is highlighted.

Conclusions

- This work provides a robust framework for stable SS stability using a direct projection method over the state-transition matrix and an alternative pre-factorized formulation.
- Both formulations consistently match the SotA in terms of accuracy, epoch time, and convergence rate, suggesting that the proposed Schur-based stabilization is not only theoretically sound but also computationally efficient for stacked-architecture SysID.
- Future research will focus on generalizing the projection algorithm to arbitrary upper bounds on eigenvalue magnitudes, achieving feature parity with SIMBa.

References

[1]A. Gu, K. Goel, and C. Re, "Efficiently Modeling Long Sequences with Structured State Spaces," in *International Conference on Learning Representations*, 2022. [Online]. Available: https://openreview.net/forum?id=uYLFoz1vlAC
[2]J. T. Smith, A. Warrington, and S. Linderman, "Simplified State Space Layers for Sequence Modeling," in *The Eleventh International Conference on Learning Representations*, 2023. [Online]. Available: https://openreview.net/forum?id=Ai8Hw3AXqks
[3]A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," in *First Conference on Language Modeling*, 2024. [Online]. Available: https://openreview.net/forum?id=tEYskw1VY2
[4]A. Salam, R. Mahmud, T. Islam, S. Mukta, and S. Shatabda, "A comprehensive survey on Mamba: Architectures, challenges, and opportunities," *Computer*, vol. 58, no. 8, pp. 64–76, 2025.
[5]L. Di Natale, M. Zakwan, P. Heer, G. Ferrari-Trecate, and C. N. Jones, "Simba: System identification methods leveraging backpropagation," *IEEE Transactions on Control Systems Technology*, 2024.
[6]A. Bemporad, "An L-BFGS-B approach for linear and nonlinear system identification under ℓ_1 and group-Lasso regularization," *IEEE Transactions on Automatic Control*, vol. 70, no. 7, pp. 4857–4864, 2025. DOI: 10.1109/TAC.2025.3541018
[7]M. Schoukens and K. Tiels, "Identification of block-oriented nonlinear systems starting from linear approximations: A survey," *Automatica*, vol. 85, pp. 272–292, 2017.
[8]V. Noferini and F. Poloni, "Nearest Ω-stable matrix via Riemannian optimization," *Numerische Mathematik*, vol. 148, no. 4, pp. 817–851, 2021.
[9]R. A. Horn and C. R. Johnson, "Matrix Analysis," in Cambridge University Press, 2012, ch. 2, p. 82.
[10]G. H. Golub and C. F. Van Loan, "Computing subspaces with the SVD," in *Matrix Computations*, 3rd. JHU press, 1996, pp. 601–606, ISBN: 978-0-8018-5414-9.

Funding Statement

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.
This programme has received funding from the European Union through Marie Skłodowska-Curie actions under project number 101081466: SEED - Systems and Engineering Science Doctorate.

LaTeX TikZposter