CM-Mamba: Multimodal Contrastive Mamba for Time Series Forecasting

Time + Recurrence Plots + Contrastive Alignment

Leandro Stival

Wageningen University & Research

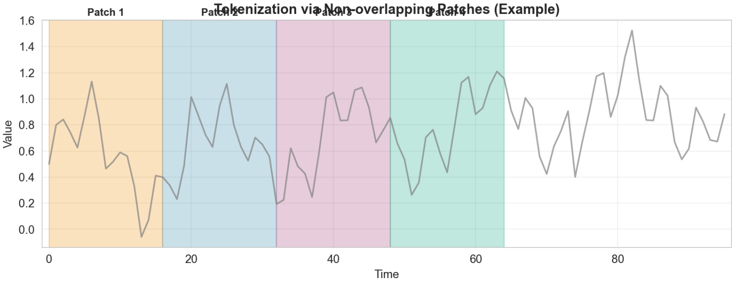
January 21, 2026

Overview

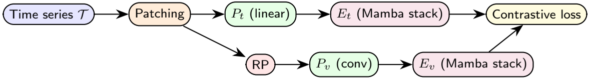
1. Motivation & Idea
2. Recurrence Plots: Visual View
3. Projectors: From Raw Tokens to Embeddings
4. Inside a Mamba Block (Feature Walkthrough)
5. Contrastive Alignment (Temporal vs Visual)
6. HiPPO Initialization (Numeric + Visual)
7. Wrap-up

Problem: Why CM-Mamba?

- Mamba/SSMs are efficient for long horizons, but can miss fine-grained local patterns (lossy fixed-state memory).
- Recurrence plots (RP) make local dynamics explicit by turning a 1D patch into a 2D structure.
- CM-Mamba aligns temporal and visual views with contrastive learning (no attention blocks added).



CM-Mamba: Pipeline Overview



Token Shapes & Similarity

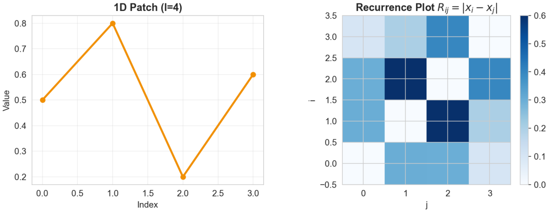
- Temporal tokens $x^t \in \mathbb{R}^{P \times l}$ and visual tokens $x^v \in \mathbb{R}^{P \times l \times l}$.
- Encoders output normalized embeddings: z^t, z^v; similarity $S_{ij} = (z_i^t)^\top z_j^v / \tau$.

Recurrence Plot: Numerical Example

Patch (toy, $l = 4$):

$$x = [0.5,\ 0.8,\ 0.2,\ 0.6]$$

$$R_{ij} = |x_i - x_j| = \begin{pmatrix} 0.0 & 0.3 & 0.3 & 0.1 \\ 0.3 & 0.0 & 0.6 & 0.2 \\ 0.3 & 0.6 & 0.0 & 0.4 \\ 0.1 & 0.2 & 0.4 & 0.0 \end{pmatrix}$$



This 2D structure makes short-term dynamics explicit.

Temporal Projector P_t: Numeric + Visual

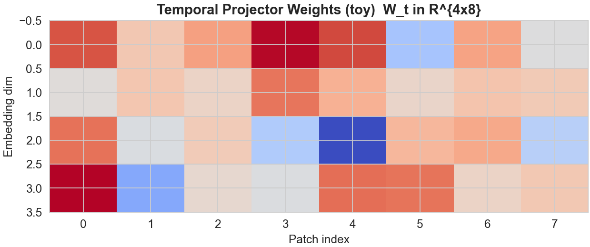
Given a patch $x \in \mathbb{R}^l$, a toy linear projector outputs $e^t = W_t x + b_t$.

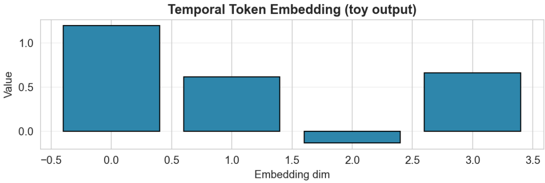
Toy patch ($l = 8$):

$$x = [0.500,\ 0.800,\ 0.842,\ 0.742,\ 0.626,\ 0.872,\ 1.132,\ 0.852]$$

$$e^t = W_t x + b_t \in \mathbb{R}^4, \quad b_t = [-0.044,\ -0.099,\ -0.017,\ 0.008]$$

$$e^t = [1.193,\ 0.612,\ -0.135,\ 0.660]$$

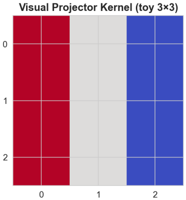


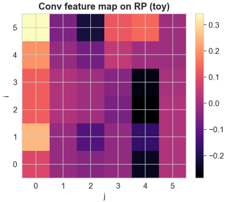


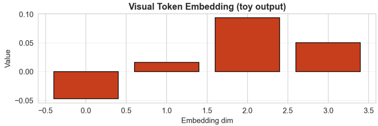
Visual Projector P_v: RP $\rightarrow Conv \rightarrow Embedding(intuition)$

We build the visual token by convolving the recurrence plot:

$$x \in \mathbb{R}^l \Rightarrow RP \in \mathbb{R}^{l \times l} \Rightarrow \text{Conv}(RP) \Rightarrow e^v \in \mathbb{R}^d$$







Visual Projector P_v: Toy numeric example

Toy output embedding $(d = 4)$:

$$e^v = W_v \, \mathrm{vec}(\mathrm{Conv}(RP)) + b_v, \quad b_v = [-0.018,\ -0.069,\ -0.032,\ -0.111]$$

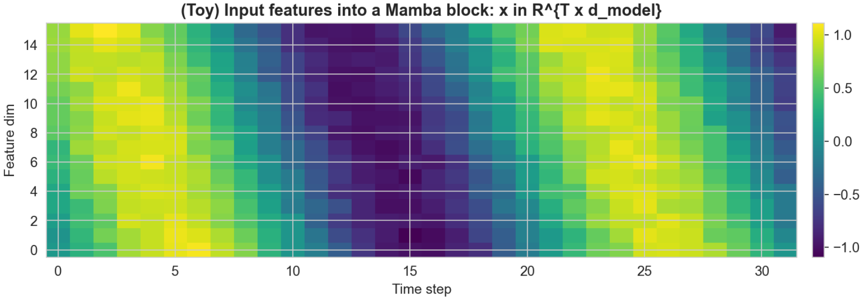
$$e^v = [-0.047,\ 0.016,\ 0.094,\ 0.051]$$

Mamba Block: What features are produced?

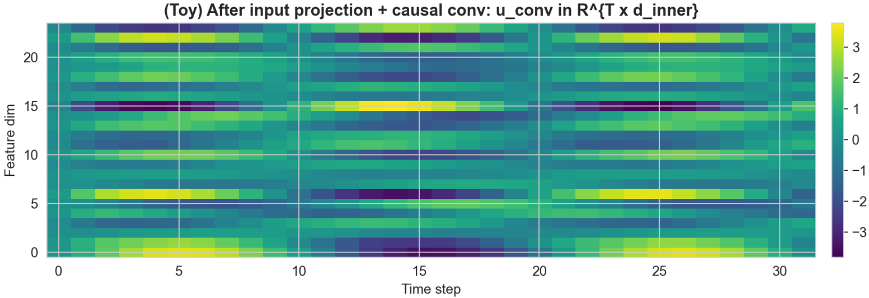
We visualize a *toy* Mamba block (small dims, fixed seed) to show how features evolve.

- Input features: $x \in \mathbb{R}^{T \times d_{model}}$
- After projection + causal conv: $u_{conv} \in \mathbb{R}^{T \times d_{inner}}$
- Selective scan produces an SSM-mixed signal (then gated).
- Output is projected back + residual.

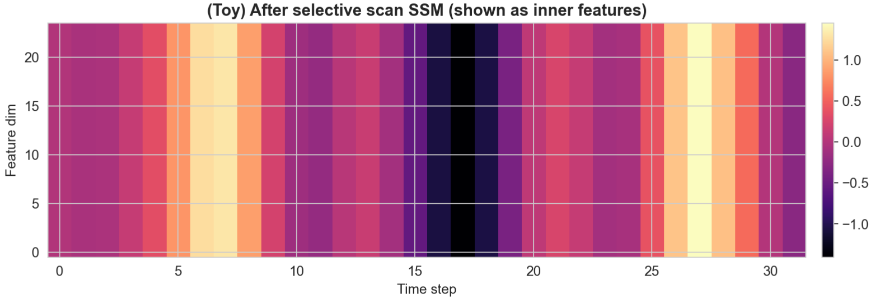
Toy Mamba Block: Input features



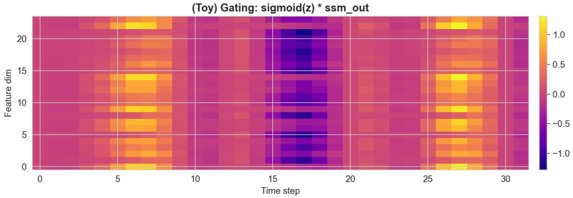
Toy Mamba Block: After causal conv

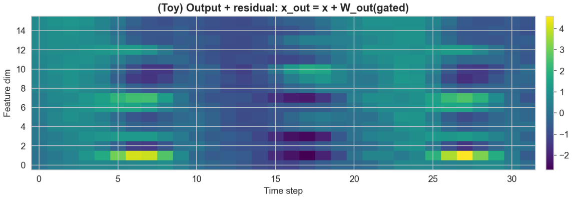


Toy Mamba Block: After selective scan (SSM mixing)



Toy Mamba Block: Gating + output + residual





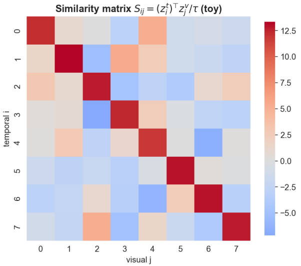
Contrastive Similarity: Visual + Numeric Example

We normalize embeddings and compute:

$$S_{ij} = \frac{(z_i^t)^\top z_j^v}{\tau}, \quad \tau = 0.07$$

Toy 4×4 slice of S:

$$S \approx \begin{pmatrix} 12.14 & 0.99 & 0.02 & -2.90 \\ 1.41 & 13.34 & -5.25 & 5.46 \\ 2.98 & 1.01 & 12.71 & -4.72 \\ 0.41 & 0.77 & -7.13 & 12.40 \end{pmatrix}$$



HiPPO-LegS: Numeric Example (N=4)

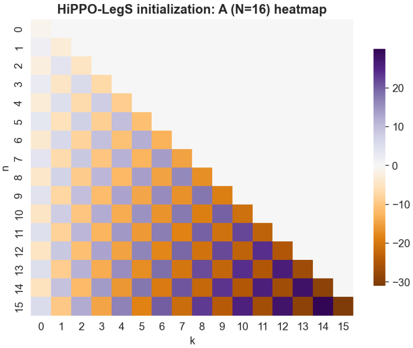
Using the paper's initialization:

$$A_{n,k} = -(2n+1)\,\delta_{n,k} + (-1)^{n-k+1}\sqrt{(2n+1)(2k+1)}\,\mathbb{I}_{k<n}$$

$$A = \begin{pmatrix} -1.000 & 0.000 & 0.000 & 0.000 \\ 1.732 & -3.000 & 0.000 & 0.000 \\ -2.236 & 3.873 & -5.000 & 0.000 \\ 2.646 & -4.583 & 5.916 & -7.000 \end{pmatrix}$$

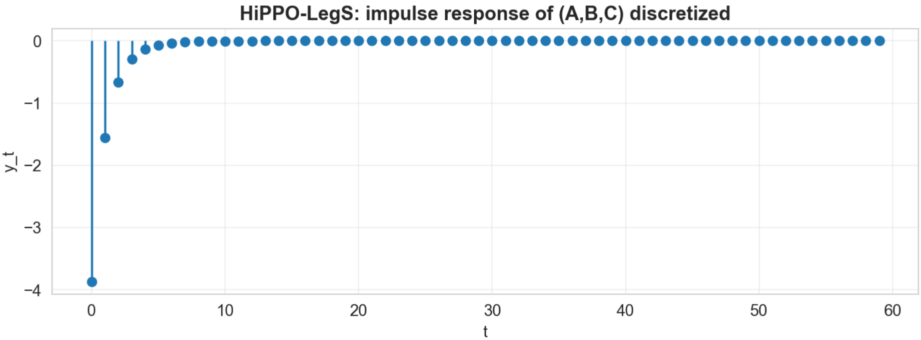
$$B = [1.000,\ 1.732,\ 2.236,\ 2.646], \quad C = [1.000,\ -1.732,\ 2.236,\ -2.646]$$

HiPPO-LegS: Visual intuition (A structure)



Heatmap of A (N=16)

HiPPO-LegS: Visual intuition (memory dynamics)



Discretized impulse response: memory dynamics

Takeaways

- CM-Mamba preserves Mamba's efficiency while injecting local structure via recurrence plots.
- The python-generated feature heatmaps show how a Mamba block transforms features step by step.
- HiPPO-LegS provides a principled initialization for long-range memory; its structure is visible in A.

Questions?