# Notes on Modeling Long Sequences with Structured State Spaces

Compiled by D. Gueorguiev 3/26/2024

## Introductory Notes

A central goal of sequence modeling is designing a single principled model that can address sequence data across a range of modalities and tasks, particularly on long-range dependencies. Although conventional models including RNNs, CNNs, and Transformers have specialized variants for capturing long dependencies, they still struggle to scale very long sequences of 10000 or more steps.

Modeling sequences can be performed by simulating the fundamental state space model (SSM) , and showed that for appropriate choices of the state matrix A, this system could handle long-range dependencies mathematically and empirically. However, this method has prohibitive computation and memory requirements, rendering it infeasible as a general sequence modeling solution in its original form. Albert Gu *et al* propose in [3] the *Structured State Space sequence* (S4) model which is based for a new parametrization for the SSM and show that it can be computed more efficiently than conventional approaches for SSM. The key to the proposed technique involves conditioning *A* with a low-rank correction, allowing it to be diagonalized stably and reducing the SSM to the well-studied computation of Cauchy kernel.

## Appendix

### Support Vector Machines and Kernels

## Literature

[1] [HiPPO: Recurrent Memory with Optimal Polynomial Projections, A. Gu et al, Stanford U., 2020](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/state_space_models/HiPPO-Recurrent_Memory_with_Optimal_Polynomial_Projections_Gu_2020.pdf)

[2] [Combining Recurrent, Convolutional, and Continuous-time Models with Linear State-Space Layers, A. Gu et al, 2021](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/state_space_models/Combining_Recurrent_Convolutional_and_Continuous-time_Models_with_Linear_State-Space_Layers_Gu_2021.pdf)

[3] [Efficiently Modeling Long Sequences with Structured State Spaces, K. Goel, A. Gu et al, 2022](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/state_space_models/Efficiently_Modeling_Long_Sequences_with_Structured_State_Spaces_Gu_Stanford_2022.pdf)

[4] [Diagonal State Spaces are as Effective as Structured State Spaces, A. Gupta, A. Gu et al, 2022](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/state_space_models/Diagonal_State_Spaces_are_as_Effective_as_Structured_State_Spaces_Gupta_Gu_2022.pdf)

[5] [Hungry Hungry Hippos: Towards Language Modeling with State Space Models, D. Fu, T. Dao, 2023](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/state_space_models/Hungry_Hungry_Hippos-Towards_Language_Modeling_with_State_Space_Models_Fu_Dao_Stanford_2023.pdf)

[6] [Mamba: Linear-Time Sequence Modeling with Selective State Spaces, A. Gu, T. Dao, CMU, 2023](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/state_space_models/Mamba-Linear-Time_Sequence_Modeling_with_Selective_State_Spaces_Gu_CMU_2023.pdf.pdf)