Infinite-Dimensional HiPPO Provides an Explicit Formula for LSSLs

Atsushi Takabatake
Department of Mathematics
Kobe University
1-1 Rokkodai-cho, Nada-ku, Kobe, 657-8501
244s008s@stu.kobe-u.ac.jp

Takaharu Yaguchi
Department of Mathematics
Kobe University
RIKEN AIP
1-1 Rokkodai-cho, Nada-ku, Kobe, 657-8501
yaguchi@pearl.kobe-u.ac.jp

Abstract

Recently, models for sequence processing, such as large-scale language models, have become increasingly important. Among these, the Linear State Space Layer (LSSL) has been proposed as a fast sequence-processing model. It is also known that using the HiPPO matrix in these LSSLs improves their performance. In this paper, we extend the HiPPO matrix to an operator on function spaces. Furthermore, we show that the resulting infinite-dimensional LSSL with HiPPO admits an explicit solution.

1 Introduction

Recently, large language models based on deep learning have attracted considerable attention. These models treat language data as time-series data and are typically trained using deep learning layers designed for sequence processing. A key requirement for such sequence-processing layers is the ability to capture long-range dependencies. The most prominent example is the Attention layer used in Transformers [14]. However, it is well known that the computational cost of the attention mechanism grows quadratically with the sequence length. As an alternative approach, methods based on State Space Models (SSMs) have been proposed. When used as layers in deep learning, these are often referred to as Linear State Space Layer (LSSL) [5]. Simply parameterizing LSSL with arbitrary matrices does not yield good performance; instead, they require initialization with specific matrices. One such example is the HiPPO (High-order Polynomial Projection Operators) matrix, which has been shown to improve the performance of LSSL [4]. In fact, a variety of SSM-based architectures, including the S4 model [4], Mamba [2], as well as other recent works [6, 13], also incorporate the HiPPO matrix. In this paper, we propose the Infinite-dimensional HiPPO, an operator on function spaces corresponding to the HiPPO matrices, by interpreting the LSSLs that use HiPPO matrices as discretizations of differential equations on function spaces. Furthermore, by exploiting the fact that the infinite-dimensional HiPPO can be regarded as a weighted integral operator, we show that the HiPPO-LSSL in infinite dimensions admits an explicit representation. In addition, we present supplementary experiments to verify that the explicit solution achieves comparable performance to existing HiPPO-based methods.

2 Background

LSSL: Linear State Space Layer A State Space Model (SSM) is defined as follows [8, 3, 4, 6, 12].

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MATH-AI.

Definition 1. Let $\mathbb{B}_{\rm in}, \mathbb{B}_{\rm state}, \mathbb{B}_{\rm out}$ be Banach spaces. The transformation $\Phi : \mathrm{Dom}\,(\Phi) \to \mathbb{B}_{\rm out}^{[0,T]}$, defined as follows, is called a (Linear) State Space Model. Here, $\mathrm{Dom}\,(\Phi)$ is a subset of $\mathbb{B}_{\rm in}^{[0,T]}$.

$$\Phi(u) = Ch,\ \frac{\partial h}{\partial t} = Ah + Bu,\ h(0) = 0 \tag{1}$$

where $A \in \mathcal{L}\left(\mathbb{B}_{\rm state}, \mathbb{B}_{\rm state}\right), B \in \mathcal{L}\left(\mathbb{B}_{\rm in}, \mathbb{B}_{\rm state}\right), C \in \mathcal{L}\left(\mathbb{B}_{\rm state}, \mathbb{B}_{\rm out}\right)$, and h denotes the solution of the differential equation corresponding to a given u.

In particular, when the Banach spaces in this equation are Euclidean spaces and the time direction t is discretized, the resulting system becomes a layer referred to as a Linear State Space Layer (LSSL) [5]. In this paper, we consider infinite-dimensional LSSLs as infinite-dimensional SSMs, and discuss them without distinction.

HiPPO: High-order Polynomial Projection Operators

Definition 2. The matrices $A^{\rm hippo} \in \mathbb{R}^{N\times N}, B^{\rm hippo} \in \mathbb{R}^{N\times 1}$ defined below, are called HiPPO (High-order Polynomial Projection Operator) matrices [3].

$$A^{\rm hippo}_{ij} = -\begin{cases} \sqrt{2i+1}\sqrt{2j+1} & (i > j) \\ i+1 & (i = j) \\ 0 & \end{cases},\ B^{\rm hippo}_{i1} = \sqrt{2i+1}\ \ (i, j = 0, \dots, N-1) \tag{2}$$

It is (experimentally) known that using these matrices as the A and B of an LSSL, either as initial parameter values or as fixed values, improves the performance of the LSSL [4]. Note that in this case, we take $\mathbb{B}_{\rm in} = \mathbb{R}$, $\mathbb{B}_{\rm state} = \mathbb{R}^N$, and $\mathbb{B}_{\rm out} = \mathbb{R}$; that is, the input and output functions are one-dimensional.

3 Infinite Dimensional HiPPO

Since $A^{\rm hippo}$ is a lower-triangular matrix, this matrix can be regarded as a discretization of an integral operator. The following theorem formalizes this observation. All proofs of theorems in this paper are provided in Appendix A.

Theorem 1. The differential equation of LSSL in a finite-dimensional space using the HiPPO matrices

$$\frac{\partial h}{\partial t} = A^{\rm hippo} h + B^{\rm hippo} u \tag{3}$$

is identical to the discretization of the differential equation in function space

$$\begin{aligned} \frac{\partial \tilde{h}}{\partial t}(t)(x) = &-\sqrt{2x+1}\int_0^x \sqrt{2\xi+1}\tilde{h}(t)(\xi)d\xi - \frac{1}{2}\tilde{h}(t)(x) \\ &+ \sqrt{2x+1}\left(-\frac{1}{2}\int_0^t e^{-(t-s)}u(s)ds + u(t)\right) \end{aligned} \tag{4}$$

when discretized in the ξ-direction by numerical integration using the trapezoidal rule. Here, the step size is taken as $d\xi \approx \Delta\xi = 1$, so that $h_i \approx \tilde{h}(i)$.

The differential equation (4) can be viewed as a state space model on an infinite-dimensional Hilbert space. Specifically, it can be rewritten in state-space model form.

For $\chi \in L^2([0,N],\mathbb{C})$, we define the operators $\mathcal{F}_\chi : L^2([0,N],\mathbb{C}) \to L^2([0,N],\mathbb{C}), \mathcal{G}_\chi : \mathbb{C} \to L^2([0,N],\mathbb{C})$ by

$$[\mathcal{F}_{\chi,\omega} f](s) = -\chi(x)\int_0^x \overline{\chi(\xi)} f(\xi)d\xi - \omega f,\ [\mathcal{G}_\chi a](s) = a\chi(x). \tag{5}$$

We set the input function as

$$v(t) = -\frac{1}{2}\int_0^t e^{-(t-s)}u(s)ds + u(t), \tag{6}$$

and the differential equation (4) can be written using $\chi_0(x) = \sqrt{2x+1}$ as

$$\frac{\partial \tilde{h}}{\partial t}(t) = \mathcal{F}_{\chi_0, \frac{1}{2}}[\tilde{h}(t)] + \mathcal{G}_{\chi_0}[v(t)]. \tag{7}$$

In particular, the above representation constitutes an infinite-dimensional state space model. Accordingly, we take $\mathcal{F}_{\chi,\omega}$ and $\mathcal{G}_\chi$ as the definition of the infinite-dimensional HiPPO.

Definition 3. The operators $\mathcal{F}_{\chi,\omega}$ and $\mathcal{G}_\chi$ are called the infinite-dimensional HiPPO.

Noting that infinite-dimensional HiPPO are defined for general χ and ω. In the next section, we derive the solution of LSSL using the infinite-dimensional HiPPO for the general χ and ω.

4 The explicit representation of LSSL with Infinite-dimensional HiPPO

First, we consider the implicit solution of the SSM. Next, we explicitly compute this implicit solution in the case of the infinite-dimensional HiPPO.

4.1 An implicit representation of LSSL

The SSM is described using the solution of a differential equation. By applying results from semi-group theory to the SSM, we can see that the following holds.

Proposition 1. Consider the SSM (1) on a Hilbert space. Suppose that u is Lipschitz continuous on $t \in [0, T]$, and that A and B are bounded linear operators. Then there exists a unique strong solution, which admits the following representation.

$$h(t) = \int_0^t \exp\left((t-s)A\right)B[u](s)ds \tag{8}$$

We apply Proposition 1 to the SSM with the infinite-dimensional HiPPO, given by

$$\Phi(v) = \mathcal{C}\tilde{h},\ \frac{\partial \tilde{h}}{\partial t}(t) = \mathcal{F}_{\chi,\omega}[\tilde{h}(t)] + \mathcal{G}_\chi[v(t)],\ h(0) = 0, \tag{9}$$

where $\mathcal{C} \in \mathscr{B}\left(L^2([0,N]), \mathbb{C}\right)$. Noting that, by the Riesz representation theorem, there exists $c \in L^2([0,N])$ such that $\mathcal{C}f = \langle c, f\rangle_{L^2([0,N])}$, and applying Proposition 1, we obtain that the LSSL using the infinite-dimensional HiPPO admits the following representation:

$$\Phi(v)(t) = \int_0^N c(x)\left[\int_0^t \exp\left((t-s)\mathcal{F}_{\chi,\omega}\right)\mathcal{G}_\chi[v(s)]ds\right](x)dx \tag{10}$$

4.2 A explicit representation of LSSL with Infinite-dimensional HiPPO

By computing the exponential of the infinite-dimensional HiPPO appearing in the above equation, we obtain an integral representation of the LSSL using the infinite-dimensional HiPPO.

Theorem 2. Suppose that v is Lipschitz continuous. Then, the strong solution of the LSSL using the infinite-dimensional HiPPO can be expressed as

$$y(t) = \int_0^N \int_0^t c(x)\chi(x)J_0\left(2\sqrt{t-\tau}\sqrt{\int_0^x |\chi(s)|^2\,\mathrm{d}s}\right)e^{-\omega(t-\tau)}v(\tau)\mathrm{d}\tau\mathrm{d}x \tag{11}$$

In particular, by setting $c(x) = \hat{c}\left(\sqrt{\int_0^x |\chi(s)|^2\,\mathrm{d}s}\right)\overline{\chi(x)}, \theta = \sqrt{\int_0^N |\chi(s)|^2\,\mathrm{d}s}$, we have

$$y(t) = \int_0^t 2\mathcal{H}_0\left[\hat{c}\cdot \mathbb{1}_{[0,\theta]}\right](2\sqrt{t-\tau})\cdot e^{-\omega(t-\tau)}\cdot v(\tau)\mathrm{d}\tau. \tag{12}$$

Here, J_ν denotes the Bessel function of order ν, and $\mathcal{H}_\nu$ denotes the Hankel transform of order ν. That is,

$$\left[\mathcal{H}_\nu f\right](t) = \int_0^\infty f(s)J_\nu(st)s\mathrm{d}s. \tag{13}$$

This representation is explicit in the sense that it is expressed entirely in terms of integrals, without using the exponential of the operator.

The above theorem follows from the fact that the infinite-dimensional HiPPO can be regarded as a weighted integral operator. By generalizing Cauchy ’ s formula for repeated integration to weighted integrals, one can compute powers of the infinite-dimensional HiPPO and thereby obtain its exponential. This discussion is difficult in finite dimensions and highlights one of the advantages of interpreting the HiPPO in an infinite-dimensional setting.

Remark 1. Focusing on equation (12), we see that χ only affects the norm θ of χ and the parameter function $\hat{c}$. In other words, from the infinite-dimensional perspective, when c is treated as a parameter, the degrees of freedom of χ are absorbed into those of the parameters and do not carry any essential significance.

5 Experiments

The main objective of this paper is to derive the explicit solution described above. However, in order to examine the differences between this explicit solution and the LSSL using the existing HiPPO, we also conducted supplementary numerical experiments.

5.1 Verification approach

The derived explicit solution can be seen as providing an explicit expression for the SSM convolution kernel[4]. Therefore, in the S4 model, the convolution kernel can be replaced with the kernel obtained from the explicit solution, namely,

$$K(t) = \int_0^N c(x)\chi(x)J_0\left(2\sqrt{t}\sqrt{\int_0^x |\chi(s)|^2\,\mathrm{d}s}\right)e^{-\omega t}\mathrm{d}x. \tag{14}$$

Based on this, we conducted experiments comparing a layer using the above kernel with a standard S4 layer. Note that in the original S4 model, the parts corresponding to A and B are not learned in the replaced kernel, and therefore they were fixed during the experiments. For the task, we used sequential MNIST. Detailed information regarding hyperparameters and other experimental settings is provided in Appendix B.

5.2 Result

As a result, the learning curves are shown in Figure 1, and the accuracy and validation loss at the end of training are presented in Table 1.

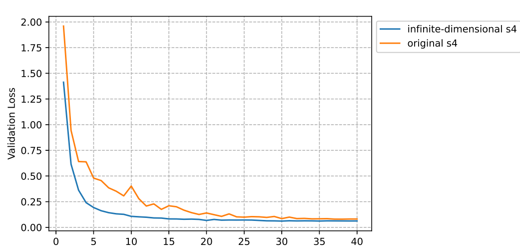

Figure 1: Validation loss curves during training on the sequential MNIST.

Table 1: Validation losses at the end of training

	Accuracy	Validation loss
original s4	0.9737	0.0813
inf-dim'l s4	0.9808	0.0624

From these results, we can confirm that using the convolution kernel derived from the explicit solution achieves performance comparable to that of the original S4 model.

6 Conclusion

In this paper, we proposed the infinite-dimensional HiPPO derived from finite-dimensional LSSLs using the HiPPO matrix, and showed that LSSLs based on this operator can be expressed as an explicit solution. We also conducted supplementary numerical experiments, demonstrating that using this explicit solution yields results comparable to existing models. Compared with the representation of LSSLs using the conventional HiPPO matrix, this approach provides a formulation that is easier to analyze theoretically and interpret conceptually. Building on this, the explicit solution opens up possibilities for further developments, such as improved computational efficiency and the application from the function space $L^2([0, N])$ to other spaces, enabling new forms of HiPPO.

Acknowledgments and Disclosure of Funding

Funding in direct support of this work: JST CREST Grant Number JPMJCR1914 and JPMJCR24Q5, JSPS KAKENHI Grant Number 25K15148, JST ASPIRE JPMJAP2329 and NIFS Collaborative Research NIFS25KISC015.

References

[1] Yann N. Dauphin, Angela Fan, Michael Auli, and David Grangier. Language modeling with gated convolutional networks. *CoRR*, abs/1612.08083, 2016.

[2] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling*, 2024.

[3] Albert Gu, Tri Dao, Stefano Ermon, Atri Rudra, and Christopher Ré. Hippo: Recurrent memory with optimal polynomial projections. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, 2020.

[4] Albert Gu, Karan Goel, and Christopher Re. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations*, 2022.

[5] Albert Gu, Isys Johnson, Karan Goel, Khaled Kamal Saab, Tri Dao, Atri Rudra, and Christopher Re. Combining recurrent, convolutional, and continuous-time models with linear state space layers. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021.

[6] Ramin M. Hasani, Mathias Lechner, Tsun-Hsuan Wang, Makram Chahine, Alexander Amini, and Daniela Rus. Liquid structural state-space models. *CoRR*, abs/2209.12951, 2022.

[7] Dan Hendrycks and Kevin Gimpel. Bridging nonlinearities and stochastic regularizers with gaussian error linear units, 2017.

[8] Baras John, S., Brockett Roger, W., and Fuhrmann Paul, A. *State-space models for infinite-dimensional systems*, volume 19 of *IEEE Transactions on Automatic Control*. 1974.

[9] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[10] Frank Olver, Daniel Lozier, Ronald Boisvert, and Charles Clark. *The NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, NY, 2010-05-12 00:05:00 2010.

[11] Amnon Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*, volume 44 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1983.

[12] T. Konstantin Rusch and Daniela Rus. Oscillatory state-space models. In *The Thirteenth International Conference on Learning Representations*, 2025.

[13] Jimmy T.H. Smith, Andrew Warrington, and Scott Linderman. Simplified state space layers for sequence modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

[14] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

A Proofs

A.1 The Proof of Proposition. 1

Proposition 1 can be proved by applying the following well-known results from semigroup theory.

Theorem 3 ([11], section 4-2, Cor 2.11)**.** Let A be the generator of a semigroup $\{T(t)\}_{t\geq 0}$ on a reflexive Banach space, and let the initial value be $h_0 \in \mathcal{D}(A)$. If f is Lipschitz continuous on $[0, T]$, then the strong solution of

$$\begin{aligned} \frac{dh}{dt} &= Ah + f \\ h(0) &= h_0 \end{aligned} \tag{15}$$

can be uniquely expressed for $t \in [0, T]$ as

$$h(t) = T(t)h_0 + \int_0^t T(t-s)f(s)ds. \tag{16}$$

Proposition 1. Consider the SSM (1) on a Hilbert space. Suppose that u is Lipschitz continuous on $t \in [0, T]$, and that A and B are bounded linear operators. Then there exists a unique strong solution, which admits the following representation.

$$h(t) = \int_0^t \exp\left((t-s)A\right)B[u](s)ds \tag{8}$$

Proof. A Hilbert space is a reflexive Banach space. Next, if A is a bounded linear operator, then exp $\exp(tA)$ can be defined as a bounded linear operator, and thus forms a semigroup with generator A. It also holds that the initial value $0 \in \mathcal{D}(A)$. Therefore, it suffices to check the Lipschitz continuity of Bu. Noting that B is bounded, i.e. $\|B\| < \infty$, and that u is assumed to be Lipschitz continuous, we have

$$\begin{aligned} \|Bu(t) - Bu(s)\| &\leq \|B\| \, \|u(t) - u(s)\| \\ &\leq L\|B\||t-s| \end{aligned} \tag{17}$$

for any s, t. Hence, Bu is Lipschitz continuous. □

A.2 The Proof of Theorem. 1

Theorem 1. The differential equation of LSSL in a finite-dimensional space using the HiPPO matrices

$$\frac{\partial h}{\partial t} = A^{\text{hippo}}h + B^{\text{hippo}}u \tag{3}$$

is identical to the discretization of the differential equation in function space

$$\begin{aligned} \frac{\partial \tilde{h}}{\partial t}(t)(x) = &-\sqrt{2x+1}\int_0^x \sqrt{2\xi+1}\tilde{h}(t)(\xi)d\xi - \frac{1}{2}\tilde{h}(t)(x) \\ &+ \sqrt{2x+1}\left(-\frac{1}{2}\int_0^t e^{-(t-s)}u(s)ds + u(t)\right) \end{aligned} \tag{4}$$

when discretized in the ξ-direction by numerical integration using the trapezoidal rule. Here, the step size is taken as $d\xi \approx \Delta\xi = 1$, so that $h_i \approx \tilde{h}(i)$.

Proof. By expressing equation (3) componentwise, we obtain, for $i > 0$, the following after some rearrangement:

$$\begin{aligned}\frac{\partial h_i}{\partial t} =& -\sqrt{2i+1}h_0 - \sqrt{2i+1}\sum_{l=1}^{i-1}\sqrt{2l+1}h_l - (i+1)h_i + \sqrt{2i+1}u(t)\\ =& -\sqrt{2i+1}\cdot\frac{h_0}{2} - \sqrt{2i+1}\sum_{l=1}^{i-1}\sqrt{2l+1}h_l - \left(i+\frac{1}{2}\right)h_i - \frac{h_i}{2}\\ & -\sqrt{2i+1}\cdot\frac{h_0}{2} + \sqrt{2i+1}u\\ =& \sqrt{2i+1}\cdot\frac{h_0}{2} + \sqrt{2i+1}\sum_{l=1}^{i-1}\sqrt{2l+1}h_l - (2i+1)\frac{h_i}{2} - \frac{h_i}{2}\\ & -\sqrt{2i+1}\cdot\frac{h_0}{2} + \sqrt{2i+1}u\\ =& -\sqrt{2i+1}\left(\frac{h_0}{2} + \sum_{l=1}^{i-1}\sqrt{2l+1}h_l + \sqrt{2i+1}\frac{h_i}{2}\right) - \frac{h_i}{2}\\ & -\sqrt{2i+1}\cdot\frac{h_0}{2} + \sqrt{2i+1}u\end{aligned} \tag{18}$$

For $i = 0$, we have

$$\frac{\partial h_0}{\partial t} = -h_0 + u(t), \tag{19}$$

which can be solved as

$$h_0(t) = \int_0^t e^{-(t-s)}u(s)\mathrm{d}s. \tag{20}$$

Substituting this into equation (18), we obtain

$$\begin{aligned}\frac{\partial h_i}{\partial t}(t) =& -\sqrt{2i+1}\left(\frac{h_0(t)}{2} + \sum_{l=1}^{i-1}\sqrt{2l+1}h_l(t) + \sqrt{2i+1}\frac{h_i(t)}{2}\right) - \frac{h_i(t)}{2}\\ & -\sqrt{2i+1}\cdot\frac{h_0(t)}{2} + \sqrt{2i+1}u(t)\\ =& -\sqrt{2i+1}\left(\frac{h_0(t)}{2} + \sum_{l=1}^{i-1}\sqrt{2l+1}h_l(t) + \sqrt{2i+1}\frac{h_i(t)}{2}\right) - \frac{h_i(t)}{2}\\ & -\sqrt{2i+1}\cdot\frac{1}{2}\int_0^t e^{-(t-s)}u(s)\mathrm{d}s + \sqrt{2i+1}u(t)\\ =& -\sqrt{2i+1}\left(\frac{h_0(t)}{2} + \sum_{l=1}^{i-1}\sqrt{2l+1}h_l(t) + \sqrt{2i+1}\frac{h_i(t)}{2}\right) - \frac{h_i(t)}{2}\\ & +\sqrt{2i+1}\left(-\frac{1}{2}\int_0^t e^{-(t-s)}u(s)\mathrm{d}s + u(t)\right).\end{aligned} \tag{21}$$

This can be interpreted as a discretization arising from the numerical integration of equation (4), where the first term corresponds to the trapezoidal rule. □

A.3 The Proof of Theorem. 2

To compute the exponential of the Infinite-dimensional HiPPO, we consider the following operator.

Lemma 1. Let $\varphi, \psi \in L^2([a,b], \mathbb{C})$, and define

$$\mathcal{F}[g](x) := \varphi(x)\int_a^x g(\xi)\psi(\xi)\mathrm{d}\xi. \tag{22}$$

Then $\mathcal{F}$ is a bounded linear operator on $L^2([a,b], \mathbb{C})$.

Proof. First, for $g \in L^2([a,b],\mathbb{C})$, we show that $\mathcal{F}[g] \in L^2([a,b],\mathbb{C})$. Let

$$f(x) := \int_a^x g(\xi)\psi(\xi)\, d\xi. \tag{23}$$

Since f is continuous in x, we have $f \in L^\infty([a,b],\mathbb{C})$. Thus, $\mathcal{F}[g] = \varphi f \in L^2([a,b],\mathbb{C})$.

$$f(x) := \int_a^x g(\xi)\psi(\xi)\mathrm{d}\xi \tag{24}$$

Next, the boundedness follows from the triangle inequality and Holder's inequality:

$$\begin{aligned}
\|\mathcal{F}[g]\|_{L^2}^2 &= \int_a^b \left|\varphi(x)\int_a^x g(\xi)\psi(\xi)\mathrm{d}\xi\right|^2 dx \\
&= \int_a^b |\varphi(x)|^2 \left|\int_a^x g(\xi)\psi(\xi)\mathrm{d}\xi\right|^2 dx \\
&\leq \int_a^b |\varphi(x)|^2 \left(\int_a^x |g(\xi)\psi(\xi)|\,\mathrm{d}\xi\right)^2 dx \\
&\leq \int_a^b |\varphi(x)|^2 \left(\int_a^b |g(\xi)\psi(\xi)|\,\mathrm{d}\xi\right)^2 dx \\
&\leq \int_a^b |\varphi(x)|^2\, dx\, (\|g\|_{L^2}\|\psi\|_{L^2})^2 \\
&\leq \|\varphi\|_{L^2}^2 \|\psi\|_{L^2}^2 \|g\|_{L^2}^2
\end{aligned} \tag{25}$$

Hence, we have $\|\mathcal{F}[g]\|_{L^2} \leq \|\varphi\|_{L^2}\|\psi\|_{L^2}\|g\|_{L^2}$ which shows that $\mathcal{F}$ is a bounded operator. □

This $\mathcal{F}$ can be viewed as a weighted integral operator. From this, using the following lemma, which is an extension of the formula for iterated integrals, we can obtain an integral representation for the powers of $\mathcal{F}$.

Lemma 2. For any $n \geq 1$, the n-th power of $\mathcal{F}$ satisfies

$$\mathcal{F}^n[g](x) = \frac{\varphi(x)}{(n-1)!}\int_a^x g(\xi)\psi(\xi)\left(\int_\xi^x \varphi(s)\psi(s)\mathrm{d}s\right)^{n-1}\mathrm{d}\xi. \tag{26}$$

Proof. Since $\varphi,\psi,g \in L^2([a,b],\mathbb{C})$, it follows that $\varphi\psi, \varphi g \in L^1([a,b],\mathbb{C})$. Hence, by the Fubini–Tonelli theorem, the function

$$p(s_1,...,s_n) := \left(\prod_{i=0}^{n-1}\varphi(s_i)\psi(s_i)\right)\psi(s_n)g(s_n) \tag{27}$$

is integrable on $[a,b]^n$.

From this, it follows that p is also integrable over the set

$$\{(s_1,...,s_n) \mid a \leq s_n \leq s_{n-1} \leq \cdots \leq s_1 \leq x\} \tag{28}$$

and by Fubini's theorem, we have

$$
\begin{aligned}
&\mathcal{F}^n[g](x) \\
&= \varphi(x)\int_a^x \psi(s_1)\varphi(s_1)\int_a^{s_1}\psi(s_2)\dots\varphi(s_{n-1})\int_a^{s_{n-1}}\psi(s_n)g(s_n)\mathrm{d}s_n\mathrm{d}s_{n-1}\dots\mathrm{d}s_2\mathrm{d}s_1 \\
&= \varphi(x)\int_a^x\int_a^{s_1}\cdots\int_a^{s_{n-1}}\left(\prod_{i=0}^{n-1}\varphi(s_i)\psi(s_i)\right)\psi(s_n)g(s_n)\mathrm{d}s_n\mathrm{d}s_{n-1}\dots\mathrm{d}s_2\mathrm{d}s_1 \\
&= \varphi(x)\int_{a\le s_n\le s_{n-1}\le\cdots\le s_1\le x}\left(\prod_{i=0}^{n-1}\varphi(s_i)\psi(s_i)\right)\psi(s_n)g(s_n)\mathrm{d}s_n\mathrm{d}s_{n-1}\cdots\mathrm{d}s_2\mathrm{d}s_1 \\
&= \varphi(x)\int_a^x\psi(s_n)g(s_n)\left(\int_{s_n\le s_{n-1}\le\cdots\le s_1\le x}\left(\prod_{i=0}^{n-1}\varphi(s_i)\psi(s_i)\right)\mathrm{d}s_{n-1}\cdots\mathrm{d}s_2\mathrm{d}s_1\right)\mathrm{d}s_n \\
&= \varphi(x)\int_a^x\psi(\xi)g(\xi)\left(\int_{\xi\le s_{n-1}\le\cdots\le s_1\le x}\left(\prod_{i=0}^{n-1}\varphi(s_i)\psi(s_i)\right)\mathrm{d}s_{n-1}\cdots\mathrm{d}s_2\mathrm{d}s_1\right)\mathrm{d}\xi.
\end{aligned}
\tag{29}
$$

We now compute the multiple integral

$$
I_{\xi,x} := \int_{\xi\le s_{n-1}\le\cdots\le s_1\le x}\left(\prod_{i=0}^{n-1}\varphi(s_i)\psi(s_i)\right)\mathrm{d}s_{n-1}\cdots\mathrm{d}s_2\mathrm{d}s_1. \tag{30}
$$

Using the fact that for any permutation $\sigma\in S_{n-1}$,

$$
\prod_{i=0}^{n-1}\varphi(s_i)\psi(s_i) = \prod_{i=0}^{n-1}\varphi\left(s_{\sigma(i)}\right)\psi\left(s_{\sigma(i)}\right) \tag{31}
$$

holds, we obtain

$$
\begin{aligned}
&\int_{[\xi,x]^{n-1}}\left(\prod_{i=0}^{n-1}\varphi(s_i)\psi(s_i)\right)\mathrm{d}s_{n-1}\cdots\mathrm{d}s_2\mathrm{d}s_1 \\
&= \sum_{\sigma\in S_{n-1}}\int_{\xi\le s_{\sigma(n-1)}\le\cdots\le s_{\sigma(1)}\le x}\left(\prod_{i=0}^{n-1}\varphi(s_i)\psi(s_i)\right)\mathrm{d}s_{n-1}\cdots\mathrm{d}s_2\mathrm{d}s_1 \\
&= \sum_{\sigma\in S_{n-1}}\int_{\xi\le s_{\sigma(n-1)}\le\cdots\le s_{\sigma(1)}\le x}\left(\prod_{i=0}^{n-1}\varphi\left(s_{\sigma(i)}\right)\psi\left(s_{\sigma(i)}\right)\right)\mathrm{d}s_{n-1}\cdots\mathrm{d}s_2\mathrm{d}s_1 \\
&= \sum_{\sigma\in S_{n-1}}\int_{\xi\le s_{n-1}\le\cdots\le s_1\le x}\left(\prod_{i=0}^{n-1}\varphi(s_i)\psi(s_i)\right)\mathrm{d}s_{n-1}\cdots\mathrm{d}s_2\mathrm{d}s_1 \\
&= |S_{n-1}|\int_{\xi\le s_{n-1}\le\cdots\le s_1\le x}\left(\prod_{i=0}^{n-1}\varphi(s_i)\psi(s_i)\right)\mathrm{d}s_{n-1}\cdots\mathrm{d}s_2\mathrm{d}s_1 \\
&= (n-1)!\int_{\xi\le s_{n-1}\le\cdots\le s_1\le x}\left(\prod_{i=0}^{n-1}\varphi(s_i)\psi(s_i)\right)\mathrm{d}s_{n-1}\cdots\mathrm{d}s_2\mathrm{d}s_1 \\
&= (n-1)!I_{\xi,x}.
\end{aligned}
\tag{32}
$$

On the other hand, by separating the variables, we have

$$
\begin{aligned}
\int_{[\xi,x]^{n-1}}\left(\prod_{i=0}^{n-1}\varphi(s_i)\psi(s_i)\right)\mathrm{d}s_{n-1}\cdots\mathrm{d}s_2\mathrm{d}s_1 &= \int_\xi^x\varphi(s_{n-1})\psi(s_{n-1})\mathrm{d}s_{n-1}\cdots\int_\xi^x\varphi(s_1)\psi(s_1)\mathrm{d}s_1 \\
&= \left(\int_\xi^x\varphi(s)\psi(s)\mathrm{d}s\right)^{n-1}.
\end{aligned}
\tag{33}
$$

Therefore, we have

$$I_{\xi,x} = \frac{1}{(n-1)!}\left(\int_{\xi}^{x} \varphi(s)\psi(s)\mathrm{d}s\right)^{n-1} \tag{34}$$

Substituting this into equation (29) yields the result.

□

Next, we present a functional identity that will be used in the proof of Proposition 2.

Lemma 3. Let

$$I_{\alpha}(x) = \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(k+\alpha+1)}\left(\frac{x}{2}\right)^{2k+\alpha} \tag{35}$$

i.e., I_α is the modified Bessel function of the first kind, and define

$$H(z) := \sum_{k=0}^{\infty} \frac{z^k}{k!(k+1)!}. \tag{36}$$

Then, for $z \in \mathbb{C} \setminus \{0\}$,

$$H(z) = \frac{I_1\left(2\sqrt{z}\right)}{\sqrt{z}} \tag{37}$$

holds, where

$$\sqrt{z} = \sqrt{|z|}e^{\frac{i}{2}\arg z}. \tag{38}$$

Proof. This can be verified by direct computation.

$$\begin{aligned}\sum_{k=0}^{\infty} \frac{z^k}{k!(k+1)!} &= \sum_{k=0}^{\infty} \frac{1}{k!(k+1)!}\left(\frac{2\sqrt{z}}{2}\right)^{2k} \\ &= \sum_{k=0}^{\infty} \frac{1}{k!(k+1)!\sqrt{z}}\left(\frac{2\sqrt{z}}{2}\right)^{2k+1} \\ &= \frac{I_1\left(2\sqrt{z}\right)}{\sqrt{z}}\end{aligned} \tag{39}$$

□

Remark 2. Note that since $H(|z|) \leq \exp(|z|)$, the radius of convergence of $H(z)$ is infinite.

Using these lemmas, we compute the exponential of the weighted integral operator $\mathcal{F}$

Lemma 4. For $c \in \mathbb{C}$, it holds that

$$\exp\left(c\mathcal{F}\right)[g](x) = g(x) + \varphi(x)\int_a^x g(\xi)\psi(\xi)\frac{cI_1\left(2\sqrt{c\int_\xi^x \varphi(s)\psi(s)\mathrm{d}s}\right)}{\sqrt{c\int_\xi^x \varphi(s)\psi(s)\mathrm{d}s}}\mathrm{d}\xi. \tag{40}$$

Proof. From Lemma 2, the following equation holds.

$$\begin{aligned}
\exp\left(c\mathcal{F}\right)[g](x) &= \sum_{k=0}^{\infty}\frac{\left(c\mathcal{F}_a\right)^k[g](x)}{k!} \\
&= g(x) + \sum_{k=1}^{\infty}\frac{\varphi(x)c^k}{k!\,(k-1)!}\int_a^x g(\xi)\psi(\xi)\left(\int_\xi^x \varphi(s)\psi(s)\mathrm{d}s\right)^{k-1}\mathrm{d}\xi \\
&= g(x) + \varphi(x)\int_a^x g(\xi)\psi(\xi)\left(\sum_{k=1}^{\infty}\frac{c^k}{k!\,(k-1)!}\left(\int_\xi^x \varphi(s)\psi(s)\mathrm{d}s\right)^{k-1}\right)\mathrm{d}\xi \\
&= g(x) + \varphi(x)\int_a^x g(\xi)\psi(\xi)\left(\sum_{k=0}^{\infty}\frac{c^{k+1}}{(k+1)!k!}\left(\int_\xi^x \varphi(s)\psi(s)\mathrm{d}s\right)^{k}\right)\mathrm{d}\xi \\
&= g(x) + \varphi(x)\int_a^x g(\xi)\psi(\xi)\left(c\sum_{k=0}^{\infty}\frac{1}{(k+1)!k!}\left(c\int_\xi^x \varphi(s)\psi(s)\mathrm{d}s\right)^{k}\right)\mathrm{d}\xi \\
&= g(x) + \varphi(x)\int_a^x g(\xi)\psi(\xi)cH\left(c\int_\xi^x \varphi(s)\psi(s)\mathrm{d}s\right)\mathrm{d}\xi \\
&= g(x) + \varphi(x)\int_a^x g(\xi)\psi(\xi)\frac{cI_1\left(2\sqrt{c\int_\xi^x \varphi(s)\psi(s)\mathrm{d}s}\right)}{\sqrt{c\int_\xi^x \varphi(s)\psi(s)\mathrm{d}s}}\mathrm{d}\xi
\end{aligned} \tag{41}$$

Here, the interchange of the integral and the series follows from the bounded convergence theorem, noting that $H(z)$ is continuous around $z = 0$.

□

In particular, using this lemma, we can compute the exponential of the infinite-dimensional HiPPO operator. Hereafter, we set $a = 0$ and $b = N$.

Proposition 2. For $c \in \mathbb{R}_{>0}$, it holds that

$$\exp\left(c\mathcal{F}_{\chi,\omega}\right)[g](x) = \exp\left(-c\omega\right)\left(g(x) + \chi(x)\int_0^x g(\xi)\overline{\chi(\xi)}\frac{\sqrt{c}J_1\left(2\sqrt{c}\sqrt{\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}\right)}{\sqrt{\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}}\mathrm{d}\xi\right). \tag{42}$$

In particular, when $g = \chi$,

$$\exp\left(c\mathcal{F}_{\chi,\omega}\right)[\chi](x) = \exp\left(-c\omega\right)\chi(x)J_0\left(2\sqrt{c}\sqrt{\int_a^x |\chi(s)|^2\,\mathrm{d}s}\right) \tag{43}$$

holds.

Proof. Define $\mathcal{E}[g](x) := \chi(x)\int_0^x \overline{\chi(\xi)}g(\xi)d\xi$. Then, we have

$$\begin{aligned}
c\mathcal{F}_{\chi,\omega}[g] &= -c\chi(x)\int_0^x \overline{\chi(\xi)}g(\xi)d\xi - c\omega g \\
&= \left(-c\mathcal{E} - c\omega\mathcal{I}\right)[g]
\end{aligned} \tag{44}$$

where $\mathcal{I}$ is the identity operator. Since $-c\mathcal{E}$ and $-c\omega\mathcal{I}$ commute as operators,

$$\exp\left(c\mathcal{F}_{\chi,\omega}\right) = \exp\left(-c\omega\mathcal{I}\right)\exp\left(-c\mathcal{E}\right) \tag{45}$$

holds. Noting that $\mathcal{E}$ denotes the weighted integral operator with $\varphi = \chi$ and $\psi = \overline{\chi}$, from Proposition 4, it holds that

$$\exp\left(-c\mathcal{E}\right)[g](x) = g(x) + \chi(x)\int_0^x g(\xi)\overline{\chi(\xi)}\frac{-cI_1\left(2\sqrt{-c\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}\right)}{\sqrt{-c\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}}\mathrm{d}\xi. \tag{46}$$

Using the relation between the modified Bessel function and the Bessel function: $iI_1(iz) = J_1(z)$, and rearranging, we obtain

$$\begin{aligned}\exp\left(-c\mathcal{E}\right)[g](x) &= g(x) + \chi(x)\int_0^x g(\xi)\overline{\chi(\xi)}\frac{-cI_1\left(2\sqrt{-c\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}\right)}{\sqrt{-c\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}}\mathrm{d}\xi \\ &= g(x) + \chi(x)\int_0^x g(\xi)\overline{\chi(\xi)}\frac{i\sqrt{c}I_1\left(2i\sqrt{c}\sqrt{\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}\right)}{\sqrt{\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}}\mathrm{d}\xi \\ &= g(x) + \chi(x)\int_0^x g(\xi)\overline{\chi(\xi)}\frac{\sqrt{c}J_1\left(2\sqrt{c\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}\right)}{\sqrt{\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}}\mathrm{d}\xi.\end{aligned} \tag{47}$$

Hence,

$$\begin{aligned}\exp\left(c\mathcal{F}_{\chi,\omega}\right)[g](x) &= \exp\left(-c\omega\mathcal{I}\right)\left[\exp\left(-c\mathcal{E}\right)[g]\right](x) \\ &= e^{-c\omega}\left(g(x) + \chi(x)\int_0^x g(\xi)\overline{\chi(\xi)}\frac{\sqrt{c}J_1\left(2\sqrt{c\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}\right)}{\sqrt{\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}}\mathrm{d}\xi\right)\end{aligned} \tag{48}$$

holds. In particular, when $g = \chi$, it hold that

$$\begin{aligned}\exp\left(c\mathcal{F}_{\chi,\omega}\right)[\chi](x) &= e^{-c\omega}\left(\chi(x) + \chi(x)\int_0^x \chi(\xi)\overline{\chi(\xi)}\frac{\sqrt{c}J_1\left(2\sqrt{c}\sqrt{\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}\right)}{\sqrt{\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}}\mathrm{d}\xi\right) \\ &= e^{-c\omega}\left(\chi(x) + \chi(x)\int_0^x |\chi(\xi)|^2\,\frac{\sqrt{c}J_1\left(2\sqrt{c}\sqrt{\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}\right)}{\sqrt{\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}}\mathrm{d}\xi\right)\end{aligned} \tag{49}$$

Here, by making the change of variables $y = 2\sqrt{c}\sqrt{\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}$, and noting

$$\mathrm{d}y = -\frac{\sqrt{c}\,|\chi(\xi)|^2}{\sqrt{\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}}\mathrm{d}\xi \tag{50}$$

it holds that

$$\begin{aligned}\exp\left(c\mathcal{F}_{\chi,\omega}\right)[\chi](x) &= e^{-c\omega}\left(\chi(x) + \chi(x)\int_0^x |\chi(\xi)|^2\,\frac{\sqrt{c}J_1\left(2\sqrt{c}\sqrt{\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}\right)}{\sqrt{\int_\xi^x |\chi(s)|^2\,\mathrm{d}s}}\mathrm{d}\xi\right) \\ &= e^{-c\omega}\left(\chi(x) - \chi(x)\int_{2\sqrt{c}\sqrt{\int_0^x |\chi(s)|^2\mathrm{d}s}}^0 J_1\left(y\right)\mathrm{d}y\right) \\ &= e^{-c\omega}\left(\chi(x) + \chi(x)\int_0^{2\sqrt{c}\sqrt{\int_0^x |\chi(s)|^2\mathrm{d}s}} J_1\left(y\right)\mathrm{d}y\right).\end{aligned} \tag{51}$$

Thus, using the formula for the Bessel function[10]:

$$\frac{dJ_0(z)}{dz} = J_1(z), \tag{52}$$

we obtain

$$
\begin{aligned}
\exp\left(c\mathcal{F}_{\chi,\omega}\right)[\chi](x) &= e^{-c\omega}\left(\chi(x)+\chi(x)\int_0^{2\sqrt{c}\sqrt{\int_0^x|\chi(s)|^2\mathrm{d}s}} J_1\left(y\right)\mathrm{d}y\right)\\
&= e^{-c\omega}\left(\chi(x)+\chi(x)\left(J_0\left(2\sqrt{c}\sqrt{\int_0^x|\chi(s)|^2\,\mathrm{d}s}\right)-J_0(0)\right)\right)\\
&= e^{-c\omega}\left(\chi(x)+\chi(x)\left(J_0\left(2\sqrt{c}\sqrt{\int_0^x|\chi(s)|^2\,\mathrm{d}s}\right)-1\right)\right)\\
&= e^{-c\omega}\chi(x)J_0\left(2\sqrt{c}\sqrt{\int_0^x|\chi(s)|^2\,\mathrm{d}s}\right)
\end{aligned}
\tag{53}
$$

□

Remark 3. In the above proof, note that even without assuming continuity of $\chi \in L^2([0,N])$, a general result from Lebesgue integration theory ensures that $\mathrm{d}y$ can be expressed using equation (50) for almost every ξ, and hence the final expression holds.

The following theorem immediately follows from this proposition.

Theorem 2. *Suppose that v is Lipschitz continuous. Then, the strong solution of the LSSL using the infinite-dimensional HiPPO can be expressed as*

$$
y(t)=\int_0^N\int_0^t c(x)\chi(x)J_0\left(2\sqrt{t-\tau}\sqrt{\int_0^x|\chi(s)|^2\,\mathrm{d}s}\right)e^{-\omega(t-\tau)}v(\tau)\mathrm{d}\tau\mathrm{d}x \tag{11}
$$

In particular, by setting $c(x)=\hat{c}\left(\sqrt{\int_0^x|\chi(s)|^2\,\mathrm{d}s}\right)\overline{\chi(x)}, \theta=\sqrt{\int_0^N|\chi(s)|^2\,\mathrm{d}s}$, we have

$$
y(t)=\int_0^t 2\mathcal{H}_0\left[\hat{c}\cdot\mathbb{1}_{[0,\theta]}\right](2\sqrt{t-\tau})\cdot e^{-\omega(t-\tau)}\cdot v(\tau)\mathrm{d}\tau. \tag{12}
$$

Proof. The solution of LSSL using the Infinite-dimensional HiPPO can be expressed as

$$
y(t)=\int_0^N c(x)\left[\int_0^t \exp\left((t-s)\mathcal{F}_{\chi,\omega}\right)\mathcal{G}_\chi[v(s)]\mathrm{d}s\right](x)\mathrm{d}x. \tag{54}
$$

By substituting the result of Proposition 2 into this, we obtain

$$
y(t)=\int_0^N\int_0^t c(x)\chi(x)J_0\left(2\sqrt{t-\tau}\sqrt{\int_0^x|\chi(s)|^2\,\mathrm{d}s}\right)e^{-\omega(t-\tau)}v(\tau)\mathrm{d}\tau\mathrm{d}x. \tag{55}
$$

By substituting $c(x)=\hat{c}\left(\sqrt{\int_0^x|\chi(s)|^2\,ds}\right)\overline{\chi(x)}$, we can write

$$
y(t)=\int_0^N|\chi(x)|^2\int_0^t\hat{c}\left(\sqrt{\int_0^x|\chi(s)|^2\,\mathrm{d}s}\right)J_0\left(2\sqrt{t-\tau}\sqrt{\int_0^x|\chi(s)|^2\,\mathrm{d}s}\right)e^{-\omega(t-\tau)}v(\tau)\mathrm{d}\tau\mathrm{d}x. \tag{56}
$$

By making the change of variables $\xi=\sqrt{\int_0^x|\chi(s)|^2\,\mathrm{d}s}$, we have

$$
|\chi(x)|^2\,\mathrm{d}x = 2\xi\mathrm{d}\xi, \tag{57}
$$

and thus

$$
\begin{aligned}
y(t) & \\
&= \int_0^t \left(\int_0^N |\chi(x)|^2 \hat{c} \left(\sqrt{\int_0^x |\chi(s)|^2 \,\mathrm{d}s} \right) J_0 \left(2\sqrt{t-\tau} \sqrt{\int_0^x |\chi(s)|^2 \,\mathrm{d}s} \right) \mathrm{d}x \right) e^{-\omega(t-\tau)} v(\tau) \mathrm{d}\tau \\
&= \int_0^t \left(\int_0^{\sqrt{\int_0^N |\chi(s)|^2 \mathrm{d}s}} 2\xi \hat{c}(\xi) J_0 \left(2\sqrt{t-\tau}\xi \right) \mathrm{d}\xi \right) e^{-\omega(t-\tau)} v(\tau) \mathrm{d}\tau \\
&= \int_0^t \left(\int_0^\theta 2\xi \hat{c}(\xi) J_0 \left(2\sqrt{t-\tau}\xi \right) \mathrm{d}\xi \right) e^{-\omega(t-\tau)} v(\tau) \mathrm{d}\tau \\
&= \int_0^t 2\mathcal{H}_0 \left[\hat{c} \mathbb{1}_{[0,\theta]} \right] \left(2\sqrt{t-\tau} \right) e^{-\omega(t-\tau)} v(\tau) \mathrm{d}\tau.
\end{aligned}
\tag{58}
$$

□

B Details of Experiments

We conducted an experiment to compare the differences by replacing the SSM convolution kernel of the S4 model with one computed using the closed-form solution. The details are described below.

B.1 Architecture

Using $\chi(x) = \sqrt{2x+1}$ derived from the existing HiPPO matrice, continuous convolution kernel is expressed as

$$
K(t) = \int_0^N c(x)\sqrt{2x+1} J_0 \left(2\sqrt{t}(x^2+x) \right) e^{-\omega t} \mathrm{d}x. \tag{59}
$$

Based on this equation, We conducted experiments using the S4 model [4] and a modified version in which the SSM convolution kernel of the S4 model was replaced with the following kernel:

$$
K_l = \sum_{n=0}^{N-1} c_n \left(2n\Delta s + 1 \right) J_0 \left(2\sqrt{l\Delta t} \sqrt{(n\Delta s)^2 + n\Delta s} \right) \Delta t \Delta s e^{-\omega l \Delta t} \tag{60}
$$

Here, $c = (c_k)_{k=0}^{N-1}$ and Δt are learnable parameters, whereas $\omega = \frac{1}{2}$ and $\Delta s = 1$ are fixed. In addition, we multiplied the convolution kernel by

$$
V_{l'} = \mathrm{FFT} \left(\left(-\frac{1}{2} e^{-k\Delta t} \Delta t \right)_{k=0}^{L-1} \right)_{l'} + 1. \tag{61}
$$

This corresponds to replacing

$$
v(t) = -\frac{1}{2} \int_0^t e^{-(t-s)} u(s) ds + u(t). \tag{62}
$$

B.2 Details

The overall model consists of an input layer, intermediate layers, and an output layer. The input layer was implemented as a linear layer, while the output layer consists of a temporal averaging layer followed by a linear layer. The intermediate layer was a single-layer S4 model with a skip connection. A batch normalization layer was applied before the S4 layer, the activation function of the S4 output was the Gaussian Error Linear Unit (GELU)[7], and after the skip connection, a Gated Linear Unit (GLU)[1] was used. We set the model dimension to $d_{model} = 64$, and the dimension of the state space to $N = 64$, and initialized the parameters as follows: $C_n \sim N(0,1), \log(\Delta t) \sim U(\log(0.1), \log(0.001))$. We used the Adam optimizer [9], together with a cosine annealing scheduler with warmup. The number of warmup steps was set to 1200. For the experimental environment, we used PyTorch as the library, and conducted training on an NVIDIA GeForce RTX 3060 Laptop GPU.