Deep Learning - Additional Homework Assignment

Derivation from Continuous to Discrete Parameters Mamba Eq.(4)

Name: Chu, Po-Jui   Student ID: 11028141

Date Submitted: April 1, 2025

1 Introduction and Notation

We start with a standard continuous-time linear time-invariant (LTI) state-space model given by

$$\frac{d\boldsymbol{h}(t)}{dt} = \boldsymbol{A}\,\boldsymbol{h}(t) + \boldsymbol{B}\,\boldsymbol{x}(t), \tag{1}$$

where:

- $\boldsymbol{h}(t) \in \mathbb{R}^n$ is the state (or hidden state) vector at time t, which compresses all past information.
- $\boldsymbol{x}(t) \in \mathbb{R}^d$ is the input signal at time t (e.g., tokens in a text, samples in audio, etc.).
- $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{B} \in \mathbb{R}^{n\times d}$ are the continuous parameters. Matrix $\boldsymbol{A}$ governs the intrinsic dynamics of the system, while $\boldsymbol{B}$ determines how the input affects the state.
- Δ (or Δt) denotes the sampling step size.

Our goal is to discretize the continuous system so that the discrete update takes the recursive form

$$\boldsymbol{h}_t = \boldsymbol{A}_d\,\boldsymbol{h}_{t-1} + \boldsymbol{B}_d\,\boldsymbol{x}_t, \tag{2}$$

and, in particular, to derive the following formulas (which correspond to Eq.(4) in the Mamba paper):

$$\bar{\boldsymbol{A}} = \exp\bigl(\Delta\,\boldsymbol{A}\bigr), \tag{3}$$

$$\bar{\boldsymbol{B}} = \bigl(\Delta\,\boldsymbol{A}\bigr)^{-1}\Bigl[\exp\bigl(\Delta\,\boldsymbol{A}\bigr) - \boldsymbol{I}\Bigr]\Delta\,\boldsymbol{B}. \tag{4}$$

Here, we use $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{B}}$ to denote the discrete-time parameters. The notation $\Delta\,\boldsymbol{A}$ (and similarly $\Delta\,\boldsymbol{B}$) emphasizes the role of the sampling step.

Reference: Gu, A., & Dao, T. (2023). *Mamba: Linear-time sequence modeling with selective state spaces.* arXiv preprint arXiv:2312.00752.

2 Step-by-Step Derivation

2.1 Step 1: Analytical Solution of the Continuous Model

For the continuous system in (1), consider the interval $[t, t+\Delta]$. The solution to the differential equation (using the method of variation of constants, or Duhamel's principle) is given by

$$\boldsymbol{h}(t+\Delta) = \exp\bigl(\boldsymbol{A}\Delta\bigr)\boldsymbol{h}(t) + \int_0^{\Delta} \exp\bigl[\boldsymbol{A}(\Delta-\tau)\bigr]\boldsymbol{B}\,\boldsymbol{x}(t+\tau)\,d\tau. \tag{5}$$

In (5):

- The first term, $\exp\bigl(\boldsymbol{A}\Delta\bigr)\boldsymbol{h}(t)$, represents the natural evolution of the state in the absence of input.
- The integral term accumulates the effect of the input over the interval $[t, t+\Delta]$, weighted by the system dynamics.

2.2 Step 2: Zero-Order Hold (ZOH) Assumption

In digital systems, it is common to assume that the input remains constant over each sampling interval. That is, we assume:

$$\boldsymbol{x}(t+\tau) = \boldsymbol{x}(t) \quad \text{for } \tau \in [0, \Delta]. \tag{6}$$

Under this zero-order hold (ZOH) assumption, the integral in (5) simplifies to:

$$\int_0^{\Delta} \exp\bigl[\boldsymbol{A}(\Delta-\tau)\bigr]\boldsymbol{B}\,\boldsymbol{x}(t)\,d\tau = \left(\int_0^{\Delta} \exp\bigl[\boldsymbol{A}(\Delta-\tau)\bigr]\,d\tau\right)\boldsymbol{B}\,\boldsymbol{x}(t). \tag{7}$$

2.3 Step 3: Change of Variables

Let $u = \Delta - \tau$. Then when $\tau = 0$, we have $u = \Delta$, and when $\tau = \Delta$, $u = 0$. Rewriting the integral, we obtain:

$$\int_0^{\Delta} \exp\bigl[\boldsymbol{A}(\Delta-\tau)\bigr]\,d\tau = \int_0^{\Delta} \exp\bigl(\boldsymbol{A}u\bigr)\,du. \tag{8}$$

2.4 Step 4: Evaluating the Matrix Exponential Integral

Assuming that $\boldsymbol{A}$ is invertible, the integral in (8) has the closed-form solution:

$$\int_0^{\Delta} \exp\bigl(\boldsymbol{A}u\bigr)\,du = \boldsymbol{A}^{-1}\Bigl[\exp\bigl(\boldsymbol{A}\Delta\bigr) - \boldsymbol{I}\Bigr], \tag{9}$$

which is the matrix analogue of the scalar result:

$$\int_0^{\Delta} e^{au}\,du = \frac{e^{a\Delta}-1}{a}.$$

2.5 Step 5: Discrete-Time State Update

Substitute the result from (9) back into (7) to obtain the discrete-time update:

$$\boldsymbol{h}(t+\Delta) = \exp\big(\boldsymbol{A}\Delta\big)\boldsymbol{h}(t) + \boldsymbol{A}^{-1}\Big[\exp\big(\boldsymbol{A}\Delta\big) - \boldsymbol{I}\Big]\boldsymbol{B}\,\boldsymbol{x}(t). \tag{10}$$

This expression shows that the discrete-time parameters are:

$$\boldsymbol{A}_d = \exp\big(\boldsymbol{A}\Delta\big), \tag{11}$$

$$\boldsymbol{B}_d = \boldsymbol{A}^{-1}\Big[\exp\big(\boldsymbol{A}\Delta\big) - \boldsymbol{I}\Big]\boldsymbol{B}. \tag{12}$$

2.6 Step 6: Introducing the Step-Size Scaling

In the Mamba paper, the continuous parameters are scaled by the step size Δ. In other words, we consider $\Delta\boldsymbol{A}$ and $\Delta\boldsymbol{B}$ instead of $\boldsymbol{A}$ and $\boldsymbol{B}$ alone. Thus, the discrete parameters are rewritten as:

$$\bar{\boldsymbol{A}} = \exp\big(\Delta\,\boldsymbol{A}\big), \tag{13}$$

$$\bar{\boldsymbol{B}} = \big(\Delta\,\boldsymbol{A}\big)^{-1}\Big[\exp\big(\Delta\,\boldsymbol{A}\big) - \boldsymbol{I}\Big]\Delta\,\boldsymbol{B}, \tag{14}$$

which is exactly Eq.(4) in the Mamba paper.

3 Summary

Starting from the continuous-time state-space model

$$\frac{d\boldsymbol{h}(t)}{dt} = \boldsymbol{A}\,\boldsymbol{h}(t) + \boldsymbol{B}\,\boldsymbol{x}(t),$$

we derived its analytical solution over the interval $[t, t+\Delta]$ using Duhamel's principle. By assuming a zero-order hold (ZOH) for the input $\boldsymbol{x}(t)$, we simplified the integral term. A change of variable allowed us to express the integral in a standard closed form:

$$\int_0^{\Delta} \exp\big(\boldsymbol{A}u\big)\,du = \boldsymbol{A}^{-1}\Big[\exp\big(\boldsymbol{A}\Delta\big) - \boldsymbol{I}\Big].$$

Substituting back, the discrete update becomes

$$\boldsymbol{h}(t+\Delta) = \exp\big(\boldsymbol{A}\Delta\big)\boldsymbol{h}(t) + \boldsymbol{A}^{-1}\Big[\exp\big(\boldsymbol{A}\Delta\big) - \boldsymbol{I}\Big]\boldsymbol{B}\,\boldsymbol{x}(t).$$

Finally, by incorporating the sampling step size (writing $\Delta\boldsymbol{A}$ and $\Delta\boldsymbol{B}$), we obtain

$$\bar{\boldsymbol{A}} = \exp\big(\Delta\,\boldsymbol{A}\big),$$

$$\bar{\boldsymbol{B}} = \big(\Delta\,\boldsymbol{A}\big)^{-1}\Big[\exp\big(\Delta\,\boldsymbol{A}\big) - \boldsymbol{I}\Big]\Delta\,\boldsymbol{B},$$

which is the desired discrete representation.

This derivation clearly shows how the continuous parameters are transformed into discrete parameters using the ZOH assumption, preserving the dynamics of the continuous system in a discrete framework.