Adapting unconstrained spiking neural networks to explore effects of time discretization on network properties

Research Question: Correlation between step size and accuracy for real world task

Zhejia Hu [145] Nergis Tömen [24] Aurora Micheli [24] Lilika Markatou [34]

[1]Author [2]Supervisor [3]Examiner [4]Delft University of Technology [5]Student number: 5444055

Leaky Integrate-And-Fire Model

The Leaky Integrate-and-Fire (LIF) model reflects the integration process of neuronal dynamics in a network. This model describes the evolution of the *membrane potential*, denoted as u. The membrane potential dynamics are influenced by the input current, i, and are parameterized by the membrane resistance, R, and the membrane capacitance $\tau_m \frac{\mathrm{d}u(t)}{\mathrm{d}t} + u(t) = R\,i(t)$.

The dynamics of the incoming synaptic current for a neuron with indices (l,k) are governed by the weighted sum of the presynaptic spike trains impinging upon its dendritic tree. The change in the synaptic current $\frac{\mathrm{d}i^{(l,k)}(t)}{\mathrm{d}t}$ at time t is determined by the dot product between the vector of learnable synaptic weights and the vector of spike trains $\tau_s \frac{\mathrm{d}i^{(l,k)}(t)}{\mathrm{d}t} + i^{(l,k)}(t) = \mathbf{w}^{(l,k)} \cdot \mathbf{s}^{(l-1)}(t)$.

The neuron's temporal response $u(t)$ to the incoming spike train has closed-form solution, which could be calculated by using Laplace Transform

$$u^{(l,k)}(t) = \text{Intialisation} + R\sum_n \left[\frac{\mathbf{w}_n^{(l,k)}}{\tau_m - \tau_s}\left(\exp\left(-\frac{t-t_n}{\tau_m}\right) - \exp\left(-\frac{t-t_n}{\tau_s}\right)\right)\right] \quad (1)$$

Time Discretization: Numerical Methods Perspective

Time is discretization is achieved employing difference equations to approximate the original differential equations. Backward Euler Method is a first-order implicit integration scheme with the same truncation error of $\mathcal{O}(\Delta^2 t)$: $\frac{\mathrm{d}f(t_k)}{\mathrm{d}t} \approx \frac{f(t_k) - f(t_k - \Delta t)}{\Delta t}$ (corresponding z-transform is $\frac{z-1}{z\,\Delta t}$). but inherently numerical stable. For the neuron system, it could be derived that

$$\begin{bmatrix} u(t_k) \\ i(t_k) \end{bmatrix} = \begin{bmatrix} 1 + \frac{\Delta t}{\tau_m} & -\frac{R\,\Delta t}{\tau_m} \\ 0 & 1 + \frac{\Delta t}{\tau_s} \end{bmatrix}^{-1} \left(\begin{bmatrix} u(t_k - \Delta t) \\ i(t_k - \Delta t) \end{bmatrix} + \frac{\Delta t}{\tau_s} \begin{bmatrix} 0 \\ \mathbf{w} \cdot \mathbf{s}(t_k) \end{bmatrix} \right) \quad (2)$$

We calculate the Lipschitz constant L that satisfies the following condition for any two states $\mathbf{n}_1(t) := [u_1(t)\ i_1(t)]^{\mathrm{T}}$ and $\mathbf{n}_2(t) := [u_2(t)\ i_2(t)]^{\mathrm{T}}$ (define $\Delta u := u_1 - u_2$ and $\Delta i := i_1 - i_2$)

$$\|\dot{\mathbf{n}}_1 - \dot{\mathbf{n}}_2\|_2 \le \sqrt{\max\left\{\frac{1+R}{\tau_m^2}, \frac{R^2+R}{\tau_m^2} + \frac{1}{\tau_s^2}\right\}}\,\|\mathbf{n}_1 - \mathbf{n}_2\|_2 \quad (3)$$

Moreover there exists a finite real number M such that $|u''(t)| \le M$. Therefore, the global truncation error of Euler's method at time t_k is constrained by an upper bound of $\frac{M\,\Delta t}{2\,L}(\exp(L\,t_k) - 1)$.

Population Dynamics in Spiking Neural Network

Let $p := p(t, u, i)$ denotes the probability density function of the system's state. We derive the corresponding Fokker-Planck equation and identify the drift coefficient $\lambda(t)\,w$ and the diffusion coefficient $\frac{\lambda(t)\,w^2}{2\,\tau_s}$ associated with $I(t)$. This formulation leads to the conclusion that the variables $(U(t),\ I(t))$ jointly form a bivariate Markov process

$$\frac{\partial}{\partial t}p = \frac{\partial}{\partial u}\left[\left(\frac{u}{\tau_m} - \frac{Ri}{\tau_s}\right)p\right] + \frac{\partial}{\partial i}\left[\frac{i - \lambda(t)\,w}{\tau_s}p + \frac{1}{2}\frac{\lambda(t)\,w^2}{\tau_s^2}\frac{\partial}{\partial i}p\right] \quad (4)$$

References

[1] Florian Bacho and Dominique Chu.
Exploring tradeoffs in spiking neural networks, 2023.

Time Discretization: Signal Processing Perspective

An alternative approach to discretize the original ODE involves sampling and scaling the original voltage signal: $u[n] := \Delta t\ u(n\,\Delta t)$, where the scaling factor preserves the signal's energy. This sampling procedure employs a sampling frequency $\Omega_s := \frac{2\,\pi}{\Delta t}$ and can be represented as a mapping from the s-domain to the z-domain via the transformation $z := \exp(s\,\Delta t)$

$$H_{\text{continuous}}(s) = \frac{R}{(\tau_m\,s + 1)\,(\tau_s\,s + 1)} \quad (5)$$

$$H_{\text{discrete}}(z) = \frac{R\,\Delta t}{\tau_m - \tau_s}\left[\frac{1}{1 - \exp\left(-\frac{\Delta t}{\tau_m}\right)z^{-1}} - \frac{1}{1 - \exp\left(-\frac{\Delta t}{\tau_s}\right)z^{-1}}\right] \quad (6)$$

The region of convergence for this transformation is defined by $|z| > \exp\left(-\frac{\Delta t}{\tau_m}\right)$. For $\omega \in [-\pi,\ \pi]$, the corresponding sampled continuous time frequency range is $\Omega \in \left[-\frac{\pi}{\Delta t},\ \frac{\pi}{\Delta t}\right]$. Frequencies beyond this range induce aliasing effects.

Network Architecture

The state of a neuron, denoted as (l,k), comprises its membrane potential and current, represented by the vector $\mathbf{n}^{(l,k)} := \left[u^{(l,k)},\ i^{(l,k)}\right]$. Spike times are enumerated for each discrete time step in SNN and a spike is emitted when the neuron's membrane potential reaches a predefined threshold value. The Backward Euler Discretization method is employed in the implementation, as it can be interpreted as a multi-layer Recurrent Neural Network.

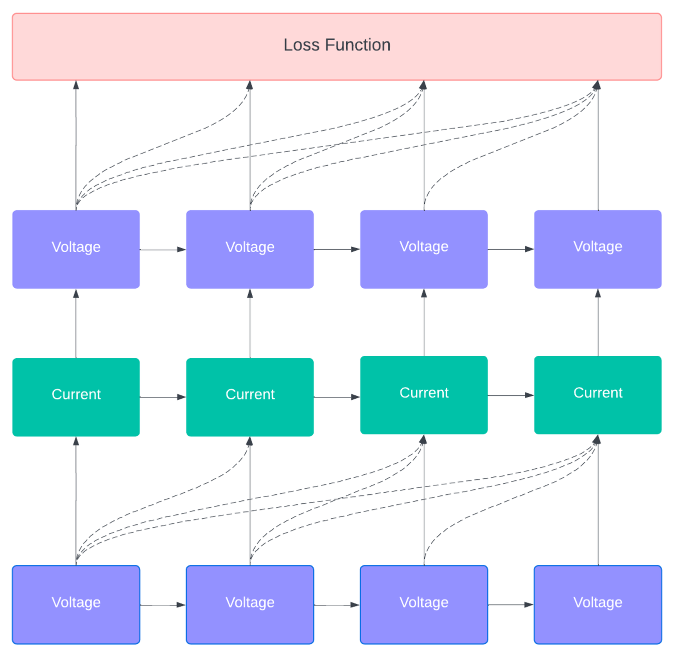


Figure 1. Network Architecture

For backward pass, we adopt an event-driven learning approach, propagating gradient information to previous layers through spike time [1]. The term $\frac{\partial i^{(l)}[t_m]}{\partial u^{(l-1)}[t_k]}$, contributing to error backpropagation between different layers, can be estimated through dynamic programming, accounting for both inter-neuron and intra-neuron dependencies (t_l is defined as the immediate spike time following t_k)

$$\frac{\partial i^{(l)}[t_m]}{\partial u^{(l-1)}[t_k]} = \frac{\partial i^{(l)}[t_m]}{\partial t_k}\frac{\partial t_k}{\partial u^{(l-1)}[t_k]} + \frac{\partial i^{(l)}[t_m]}{\partial u^{(l-1)}[t_l]}\frac{\partial u^{(l-1)}[t_l]}{\partial t_k}\frac{\partial t_k}{\partial u^{(l-1)}[t_k]} \quad (7)$$

Experimental Setup and Accuracy

MNIST handwritten digit dataset includes $60,000$ training samples and $10,000$ testing samples, each of which is a single-channel image with dimensions of 28×28 pixels. The batch size is 50 and the training process spans 4 epochs with the simulation performing over 50 time steps. Poisson input spikes are generated using a rate encoding scheme. The shared hyper-parameters are $(\tau_m,\ \tau_s,\ R) = (0.8,\ 0.6,\ 1.2)$ and $(v_{\text{thres}},\ v_{\text{reset}}) = (1.5,\ 1)$.

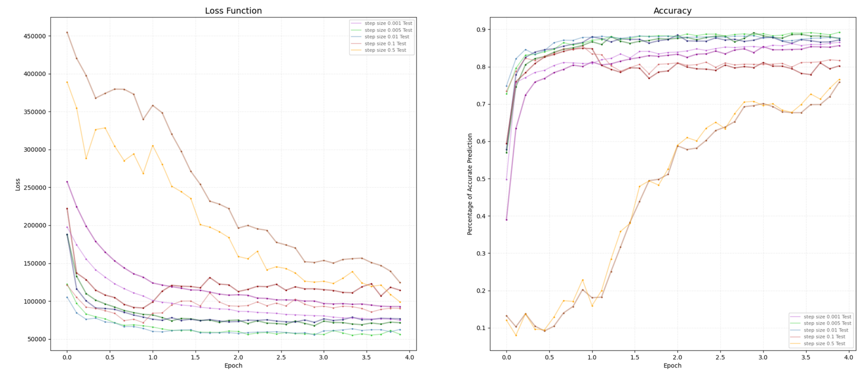
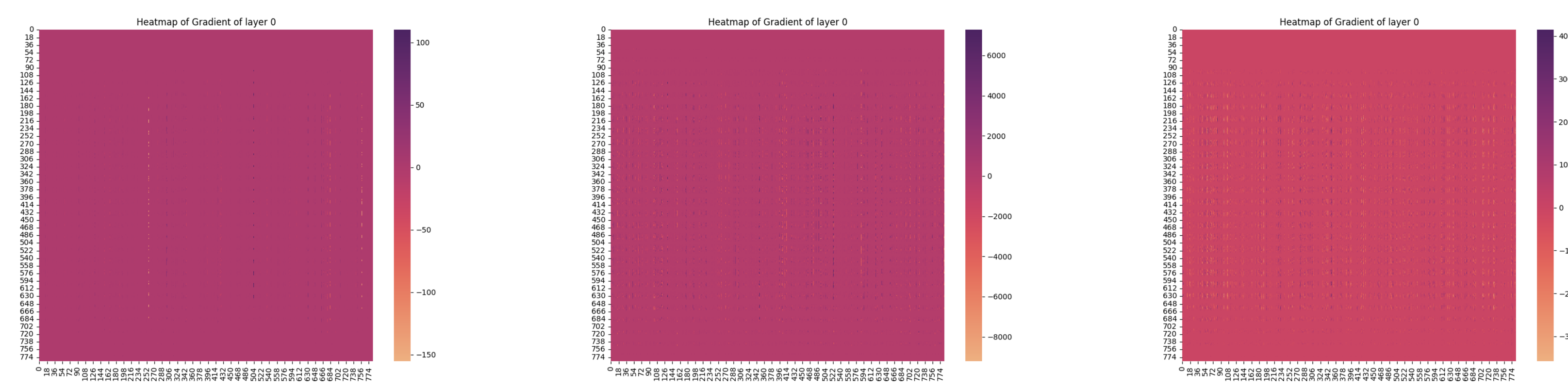


Figure 2. Accuracy and Losses for different step sizes

.

Optimization and Spectral Analysis

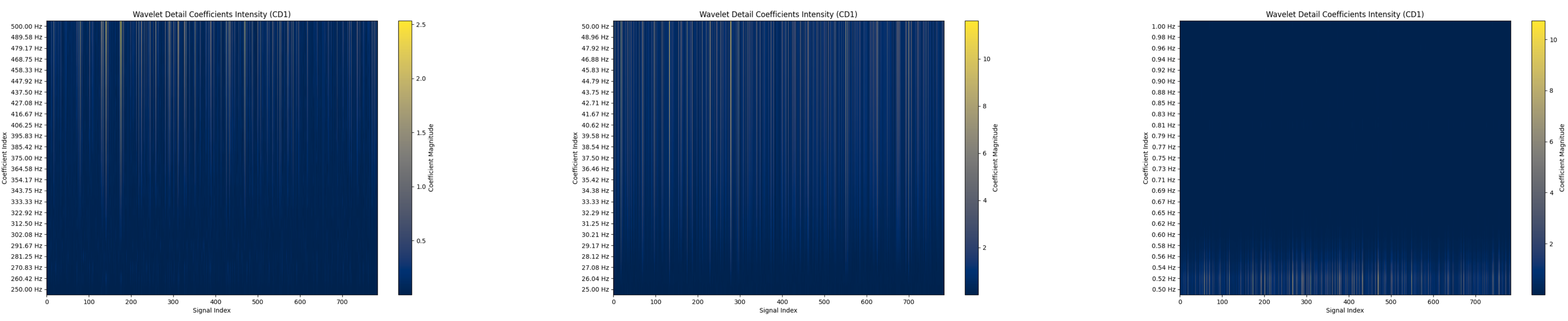
As the step size increases, the gradient values exhibit significant growth, particularly in the hidden layers of the network. This phenomenon exacerbates the gradient explosion problem.



(a) Step size 0.001 (b) Step size 0.01 (c) Step size 0.1

Figure 3. Layer 0 Weight Gradient

Spectral analysis of voltage signals reveals that smaller step sizes result in higher magnitudes of high-frequency components in the voltage signals, while larger step sizes lead to increased magnitudes in low-frequency components.



(a) Step size 0.001 (b) Step size 0.01 (c) Step size 0.5

Figure 4. Spectral Analysis of the Hidden Layer Current Output: Detailed Coefficients