

1 Hopfield Network

1.1 Discrete-time Hopfield Network

1.1.1 Definition

Definition 1. *[Discrete-time Hopfield Network]*

Let $t \in \mathbb{N}$ and $x \in \{-1, +1\}^d$, $W \in \mathbb{R}^d \times \mathbb{R}^d$ with $W_{\alpha\beta} = W_{\beta\alpha}$ and $W_{\alpha\alpha} = 0$, and $b \in \mathbb{R}^d$. Define discrete-time dynamics

$$x^\alpha(t+1) = \text{sign}(W_{\alpha\beta} x^\beta(t) + b^\alpha).$$

The (x, W, b) is called a discrete-time Hopfield network.

1.1.2 Convergence

Lemma 2. Let (x, W, b) a discrete-time Hopfield network. Define $\mathcal{E}(x) := -(1/2)W_{\alpha\beta}x^\alpha x^\beta - b_\alpha x^\alpha$. Then $\mathcal{E}(x(t+1)) - \mathcal{E}(x(t)) \leq 0$.

Proof. Consider async-updation of Hopfield network, that is, change the component at dimension $\hat{\alpha}$, i.e. $x'_{\hat{\alpha}} = \text{sign}[W_{\hat{\alpha}\beta}x^\beta + b_{\hat{\alpha}}]$, then

$$\begin{aligned} \mathcal{E}(x') - \mathcal{E}(x) &= -\frac{1}{2}W_{\alpha\beta}x'^\alpha x'^\beta - b_\alpha x'^\alpha + \frac{1}{2}W_{\alpha\beta}x^\alpha x^\beta + b_\alpha x^\alpha \\ &= -2(x'^{\hat{\alpha}} - x^{\hat{\alpha}})(W_{\hat{\alpha}\beta}x^\beta + b_{\hat{\alpha}}), \end{aligned}$$

which employs conditions $W_{\alpha\beta} = W_{\beta\alpha}$ and $W_{\alpha\alpha} = 0$. Next, we prove that, combining with $x'_{\hat{\alpha}} = \text{sign}[W_{\hat{\alpha}\beta}x^\beta + b_{\hat{\alpha}}]$, this implies $\mathcal{E}(x') - \mathcal{E}(x) \leq 0$.

If $(x'^{\hat{\alpha}} - x^{\hat{\alpha}}) > 0$, then $x'^{\hat{\alpha}} = 1$ and $x^{\hat{\alpha}} = -1$. Since $x'_{\hat{\alpha}} = \text{sign}[W_{\hat{\alpha}\beta}x^\beta + b_{\hat{\alpha}}]$, $W_{\hat{\alpha}\beta}x^\beta + b_{\hat{\alpha}} > 0$. Then $\mathcal{E}(x') - \mathcal{E}(x) < 0$. Contrarily, if $(x'^{\hat{\alpha}} - x^{\hat{\alpha}}) < 0$, then $x'^{\hat{\alpha}} = -1$ and $x^{\hat{\alpha}} = 1$, implying $W_{\hat{\alpha}\beta}x^\beta + b_{\hat{\alpha}} < 0$. Also $\mathcal{E}(x') - \mathcal{E}(x) < 0$. Otherwise, $\mathcal{E}(x') - \mathcal{E}(x) = 0$. So, we conclude $\mathcal{E}(x') - \mathcal{E}(x) \leq 0$. \square

Theorem 3. *[Convergence of Discrete-time Hopfield Network]* Let (x, W, b) a discrete-time Hopfield network. Then any trajectory obeying the update rule will converge either to a fixed point or a limit circle.

Proof. Since the states of the network are finite, the \mathcal{E} is lower bounded. \square

1.2 Continuous-time Hopfield Network

1.2.1 Definition

Definition 4. *[Continuous-time Hopfield Network]*

Let $t \in [0, +\infty)$ and $x \in [-1, +1]^d$, $W \in \mathbb{R}^d \times \mathbb{R}^d$ with $W_{\alpha\beta} = W_{\beta\alpha}$, and $b \in \mathbb{R}^d$. Define dynamics

$$\tau \frac{dx^\alpha}{dt}(t) = -x^\alpha(t) + f(W_{\alpha\beta}x^\beta(t) + b^\alpha),$$

where $\tau \in (0, +\infty)$ a constant and $f: \mathbb{R} \rightarrow [-1, 1]$ being increasing. The $(x, W, b; \tau, f)$ is called a continuous-time Hopfield network.

Remark 5. With

$$\tau \frac{x^\alpha(t + \Delta t) - x^\alpha(t)}{\Delta t} = -x^\alpha(t) + f(W^\alpha_\beta x^\beta(t) + b^\alpha).$$

Setting $\Delta t = \tau$ gives and $f(\cdot) = \text{sign}(\cdot)$ gives

$$x^\alpha(t + \tau) = \text{sign}(W^\alpha_\beta x^\beta(t) + b^\alpha),$$

which is the same as the discrete-time Hopfield network.

1.2.2 Convergence

Lemma 6. Let $(x, W, b; \tau, f)$ a continuous-time Hopfield network. Define $a^\alpha := W^\alpha_\beta x^\beta + b^\alpha$ and $y^\alpha := f(a^\alpha)$, then

$$\mathcal{E}(y) := -\frac{1}{2} W_{\alpha\beta} y^\alpha y^\beta - b_\alpha y^\alpha + \sum_\alpha \int^{y^\alpha} f^{-1}(y^\alpha) dy^\alpha.$$

Then $\mathcal{E}(y(x(t + dt))) - \mathcal{E}(y(x(t))) \leq 0$.

Proof. The dynamics of a^α is

$$\begin{aligned} \tau \frac{da^\alpha}{dt} &= \tau W^\alpha_\beta \frac{dx^\beta}{dt} \\ &= W^\alpha_\beta [-x^\beta(t) + f(a^\beta)] \\ &= -(W^\alpha_\beta x^\beta(t) + b^\alpha) + b^\alpha + W^\alpha_\beta y^\beta \\ &= W^\alpha_\beta y^\beta + b^\alpha - a^\alpha. \end{aligned}$$

Since W is symmetric, we have $\partial \mathcal{E} / \partial y^\alpha = -W_{\alpha\beta} y^\beta - b_\alpha + f^{-1}(y_\alpha)$. Then

$$\begin{aligned} \frac{d\mathcal{E}}{dt} &= \frac{dy^\alpha}{dt} (-W_{\alpha\beta} y^\beta - b_\alpha + f^{-1}(y_\alpha)) \\ &= \frac{dy^\alpha}{dt} (-W_{\alpha\beta} y^\beta - b_\alpha + a_\alpha) \\ &= -\frac{dy^\alpha}{dt} (W_{\alpha\beta} y^\beta + b_\alpha - a_\alpha) \end{aligned}$$

Notice that, the second term of rhs is exactly the dynamics of a_α , then

$$\begin{aligned} \frac{d\mathcal{E}}{dt} &= -\tau \frac{dy^\alpha}{dt} \frac{da_\alpha}{dt} \\ &= -\tau \frac{dy^\alpha}{da^\alpha} \left(\frac{da^\alpha}{dt} \frac{da_\alpha}{dt} \right) \\ &= -\tau f'(a^\alpha) \left(\frac{da^\alpha}{dt} \frac{da_\alpha}{dt} \right). \end{aligned}$$

Since f is increasing and $\tau > 0$, $d\mathcal{E}/dt \leq 0$. □

Remark 7. The condition $W_{\alpha\alpha} = 0$ for $\forall \alpha$ is not essential for this lemma. Indeed, this condition is absent in the proof. This differs from the case of discrete-time.

Theorem 8. [Convergence of Continuous-time Hopfield Network] Let $(x, W, b; \tau, f)$ a continuous-time Hopfield network. Then any trajectory along the dynamics will converge either to a fixed point or a limit circle.

Proof. The function $E := \mathcal{E} \circ y$ is lower bounded since y , i.e. function $f: \mathbb{R} \rightarrow [-1, 1]$, is bounded. This E is a Lyapunov function for the continuous-time Hopfield network. \square

1.2.3 Learning Rule

Corollary 9. Let $(x, W, b; \tau, f)$ a continuous-time Hopfield network. And $D := \{x_n | x_n \in \mathbb{R}^d, n = 1, \dots, N\}$ a dataset¹. If add constraint $W_{\alpha\alpha} = 0$ for $\forall \alpha$, then we can train the Hopfield network by seeking a proper parameters (W, b) , s.t. its stable points cover the dataset as much as possible, by²

Algorithm 1

```
W, b = init_W, init_b # e.g. by Glorot initializer
for step in range(max_step):
    for x in dataset:
        y = f(W @ x + b)
        loss = norm(x - y)
        optimizer.minimize(objective=loss, variables=(W, b))
        W = set_zero_diag(symmetrize(W))
```

Proof. For $\forall x_n \in D$, we try to find (W, b) , s.t. $dx/dt = 0$ at x_n , i.e.

$$x_n^\alpha = f(W_{\alpha\beta} x_n^\beta + b^\alpha).$$

When $W_{\alpha\alpha} = 0$ for $\forall \alpha$, $f(W_{\alpha\beta} x_n^\beta + b^\alpha)$ thus has no information of x^α , it has to predict the x^α by the interaction between x^α and the other x 's components. \square

Remark 10. This algorithm is equivalent to

Algorithm 2

```
dt = ... # e.g. 0.1
W, b = init_W, init_b
for step in range(max_step):
    for x in dataset:
        # that is, compute x(dt), with x(0) = x
        y = ode_solve(f=lambda t, x: -x + f(W @ x + b), t0=0, t1=dt, x0=x)
        loss = norm(x - y)
        optimizer.minimize(objective=loss, variables=(W, b))
        W = set_zero_diag(symmetrize(W))
```

Indeed, trying to reach $y = x$ within a small interval will force x to be a fixed point.

1.2.4 Relation to Auto-encoder

Notice that at fixed point x_* , $x_*^\alpha = f(W_{\alpha\beta} x_*^\beta + b^\alpha)$, which is a single-layer auto-encoder. The learning rule is also simply the learning rule of single-layer auto-encoder.

1.2.5 Stability of Fixed Points

We study the stability of fixed points. Let $z^\alpha := W_{\alpha\beta} x^\beta + b^\alpha$. Jacobian

$$\begin{aligned} J^\alpha_\beta &= \frac{\partial}{\partial x^\beta} (-x^\alpha + f(z^\alpha)) \\ &= -\delta^\alpha_\beta + f'(z^\alpha) W^\alpha_\beta. \end{aligned}$$

1. We use Greek alphabet for component in \mathbb{R}^d and Lattin alphabet for element in dataset.

2. This algorithm generalizes the algorithm 42.9 of Mackay.

If $f(x) = \tanh(x)$, and at fixed point,

$$\begin{aligned} J^\alpha_\beta &= -\delta^\alpha_\beta + \frac{1}{2}(1 - f^2(z^\alpha))W^\alpha_\beta \\ &= -\delta^\alpha_\beta + \frac{1}{2}(1 - x^\alpha_\star)(1 + x^\alpha_\star)W^\alpha_\beta. \end{aligned}$$

The eigen-value of J , $\lambda_J =: -1 + \lambda$, have

$$\det\left(\frac{1}{2}(1 - x^\alpha_\star)(1 + x^\alpha_\star)W^\alpha_\beta - \lambda\delta^\alpha_\beta\right) = 0$$

For instance, if $x^\alpha_\star \rightarrow \pm 1$ for $\forall \alpha$, that is $\|x^\alpha_\star - 1\| \ll 1$, then, because of the linearity of this equation, we will have $\lambda \ll 1$. In this case, $\lambda_J \approx -1 < 0$, indicating the stability of the fixed point x_\star .

2 Variations

2.1 Dense Associative Memories

Theorem 11. Let $v \in \mathbb{R}^d$, $F \in C^1(\mathbb{R}^n, \mathbb{R})$, $W \in \mathbb{R}^n \times \mathbb{R}^d$, $b \in \mathbb{R}^n$, and $\tau > 0$. Define the dynamics

$$\tau \frac{dx}{dt} = -\nabla E(x) = -x + W^T \cdot \nabla F(W \cdot x + b) + v.$$

If $\nabla F(\cdot)$ is bounded, i.e. $\exists K > 0$ s.t. $\max_{x \in \mathbb{R}^n} \{\|\nabla F(x)\|\} < K$, then any trajectory along the dynamics will converge either to a fixed point or a limit circle.

Proof. Let $E(x) := \frac{1}{2}x_\alpha x^\alpha - v_\alpha x^\alpha - F(W^\alpha_\beta x^\beta + b^\alpha)$, then $\tau dx/dt = -\nabla E(x)$. The $-x$ term will dominate the $W^T \cdot \nabla F(W \cdot x + b)$ term for $\|x\| > K\|W\|$, thus converges. So E is a Lyapunov function of the dynamics. \square

Example 12. Let $F(x) := \sum_\alpha \int^{x^\alpha} \sigma(s) ds$, where σ is sigmoid function. Then

$$\tau \frac{dx}{dt} = -x + W^T \cdot \sigma(W \cdot x + b) + v.$$

This coincides with the form in ref 1.

Example 13. Let $F(x) := \beta^{-1} \ln(\beta \sum_\alpha e^{x^\alpha})$, $b = 0$, and $v = 0$, then

$$\tau \frac{dx}{dt} = -x + W^T \cdot \text{softmax}(\beta W \cdot x).$$

This coincides with the form in ref 2.

Example 14. Let $v_i := W_{i,\cdot}$, i.e. the i th row of the matrix W . Assume $\|v_i\| = 1$ for $\forall i = 1, \dots, n$. Let $F(x) := \beta^{-1} \ln(\beta \sum_\alpha e^{x^\alpha})$, and $v = 0$, then

$$\tau \frac{dx^\alpha}{dt} = -x^\alpha + \sum_i p_i v_i^\alpha,$$

where $z_i := v_i \cdot x + b_i$ and then $p^i := \exp(\beta z^i) / \sum_j \exp(\beta z^j)$. The $\{(p_i, v_i) | i = 1, \dots, n\}$ forms a categorical distribution.

Lemma 15. Assume example 14. The Jacobian of the dynamics is

$$J^{\alpha\beta}(x) = -\delta^{\alpha\beta} + \text{Cov}_{p(x)}(v^\alpha, v^\beta),$$

where $\text{Cov}_p(\cdot, \cdot)$ denotes the covariance given distribution p .

Proof. Directly,

$$\begin{aligned} J^{\alpha\beta} &\equiv \frac{\partial}{\partial x_\beta} \left(-x^\alpha + \sum_i v_i^\alpha p_i \right) \\ &= -\delta^{\alpha\beta} + \sum_{i,j} v_i^\alpha \frac{\partial p_i}{\partial z^j} \frac{\partial z^j}{\partial x_\beta} \\ &= -\delta^{\alpha\beta} + \beta \sum_{i,j} v_i^\alpha v_j^\beta (p_i \delta_{i,j} - p_i p_j) \\ &= -\delta^{\alpha\beta} + \beta \sum_i p_i v_i^\alpha v_i^\beta - \beta \left(\sum_i p_i v_i^\alpha \right) \left(\sum_j p_j v_j^\beta \right) \\ &= -\delta^{\alpha\beta} + \beta \mathbb{E}(v^\alpha v^\beta) - \beta \mathbb{E}(v^\alpha) \mathbb{E}(v^\beta) \\ &= -\delta^{\alpha\beta} + \beta \text{Cov}_p(v^\alpha, v^\beta). \end{aligned}$$

And notice that the only variable that depends on x is p . So we insert x and gain the result. \square

For instance, at fixed point $x = v_1$, $p = (1, 0, \dots, 0)$. $\text{Cov}_{p(v_1)}(v^\alpha, v^\beta) = v_1^\alpha v_1^\beta - v_1^\alpha v_1^\beta = 0$. So $J^{\alpha\beta} = -\delta^{\alpha\beta}$ is negative defined, indicating that the fixed point is stable.

2.2 Cellular Automa

TODO

3 References

1. [On autoencoder scoring.](#)
2. [Hopfield networks is All You Need.](#)