

## 1 Energy-based Model

**Definition 1.** *[Energy-based Model]*

Let  $\mathcal{M}$  a measure space, and  $E: \mathbb{R}^m \rightarrow (\mathcal{M} \rightarrow \mathbb{R})$ . Then define probabilistic model based on  $E$  as

$$p_E(x; \theta) = \frac{\exp(-E(x; \theta))}{\int_{\mathcal{M}} dx' \exp(-E(x'; \theta))}, \quad (1)$$

where  $\theta \in \mathbb{R}^m$  and  $x \in \mathcal{M}$ .

We call this an energy-based model, where  $E(\cdot; \theta)$  is called a energy function parameterized by  $\theta$ .

**Theorem 2.** *[Universality]*

For any probability density  $q: \mathcal{M} \rightarrow \mathbb{R}$  and for  $\forall C \in \mathbb{R}$ , define, for  $\forall x \in \text{supp}(q)$ ,

$$E_q(x) := -\ln q(x) + C, \quad (2)$$

then, for  $\forall x \in \text{supp}(q)$ ,

$$q(x) = \frac{\exp(-E_q(x))}{\int_{\text{supp}(q)} dx' \exp(-E_q(x'))}. \quad (3)$$

That is, for any probability density, there exists an energy function (up to constant) that can describe the probability density.

**Proof.** Directly,

$$\begin{aligned} q(x) &= \frac{\exp(-E_q(x))}{\int_{\text{supp}(q)} dx' \exp(-E_q(x'))} \\ \{E_q := \dots\} &= \frac{q(x)}{\int_{\text{supp}(q)} dx' q(x')} \\ \left\{ \int_{\text{supp}(q)} dx' q(x') = 1 \right\} &= q(x). \end{aligned}$$

□

**Theorem 3.** *[Maximum Entropy Principle]*

For any probability density  $p_D: \mathcal{M} \rightarrow \mathbb{R}$ , we have

$$p_E(x) = \text{argmax}_p H[X], \quad (4)$$

s.t. constrains

$$\mathbb{E}_{x \sim p_D} \left[ \frac{\partial E}{\partial \theta^\alpha}(x; \theta) \right] = \mathbb{E}_{x \sim p} \left[ \frac{\partial E}{\partial \theta^\alpha}(x; \theta) \right] \quad (5)$$

are satisfied.

**Theorem 4.** *[Activity Rule]*

The local maximum of  $p_E(\cdot; \theta)$  is the local minimum of  $E(\cdot; \theta)$ , and vice versa.

**Theorem 5.** *[Learning Rule]*

For any probability density  $p_D: \mathcal{M} \rightarrow \mathbb{R}$ , define Lagrangian  $L(\theta; p_D) := -\int_{\mathcal{M}} dx p_D(x) \ln p_E(x; \theta)$ . Then, the gradient of Lagrangian w.r.t. component  $\theta^\alpha$  is

$$\frac{\partial L}{\partial \theta^\alpha}(\theta; p_D) = \int_{\mathcal{M}} dx p_D(x) \frac{\partial E}{\partial \theta^\alpha}(x; \theta) - \int_{\mathcal{M}} dx p_E(x; \theta) \frac{\partial E}{\partial \theta^\alpha}(x; \theta), \quad (6)$$

or in more compact format,

$$\frac{\partial L}{\partial \theta^\alpha}(\theta; p_D) = \mathbb{E}_{x \sim p_D} \left[ \frac{\partial E}{\partial \theta^\alpha}(x; \theta) \right] - \mathbb{E}_{x \sim p_E(x; \theta)} \left[ \frac{\partial E}{\partial \theta^\alpha}(x; \theta) \right]. \quad (7)$$

## 2 Effective Theory

**Definition 6.** *[Effective Energy]*

Suppose exists  $(\mathcal{V}, \mathcal{H})$ , s.t.  $\mathcal{M} = \mathcal{V} \oplus \mathcal{H}$ . Re-denote  $E(x; \theta) \rightarrow E(v, h; \theta)$  and  $p_E(x; \theta) \rightarrow p_E(v, h; \theta)$ . Then, define effective energy  $E_{\text{eff}}: \mathcal{V} \rightarrow \mathbb{R}$  as

$$E_{\text{eff}}(v; \theta) := -\ln \int_{\mathcal{H}} dh \exp(-E(v, h; \theta)). \quad (8)$$

**Theorem 7.** [Effective Theory]

Recall that  $p_{E_{\text{eff}}}(v; \theta) := \int_{\mathcal{H}} dh p(v, h; \theta)$ . Then,

$$p_{E_{\text{eff}}}(v; \theta) = \frac{\exp(-E_{\text{eff}}(v; \theta))}{\int_{\mathcal{V}} dv' \exp(-E_{\text{eff}}(v'; \theta))}. \quad (9)$$

**Lemma 8.** [Gradient of Effective Energy]

$$\frac{\partial E_{\text{eff}}}{\partial \theta^\alpha}(v, \theta) = \int_{\mathcal{H}} dh p(h|v; \theta) \frac{\partial E}{\partial \theta^\alpha}(v, h; \theta). \quad (10)$$

**Theorem 9.** [Learning Rule of Effective Theory]

For any probability density  $p_D: \mathcal{V} \rightarrow \mathbb{R}$ , define Lagrangian  $L(\theta; p_D) := -\int_{\mathcal{V}} dv p_D(v) \ln p(v; \theta)$ . Then, the gradient of Lagrangian w.r.t. component  $\theta^\alpha$  is

$$\frac{\partial L}{\partial \theta^\alpha}(\theta; p_D) = \int_{\mathcal{V}} dv \int_{\mathcal{H}} dh p_D(v) p(h|v; \theta) \frac{\partial E}{\partial \theta^\alpha}(v, h; \theta) - \int_{\mathcal{V}} dv \int_{\mathcal{H}} dh p(v, h; \theta) \frac{\partial E}{\partial \theta^\alpha}(v, h; \theta),$$

or in more compact format,

$$\frac{\partial L}{\partial \theta^\alpha}(\theta; p_D) = \mathbb{E}_{v \sim p_D, h \sim p_E(h|v; \theta)} \left[ \frac{\partial E}{\partial \theta^\alpha}(v, h; \theta) \right] - \mathbb{E}_{v, h \sim p_E(v, h; \theta)} \left[ \frac{\partial E}{\partial \theta^\alpha}(v, h; \theta) \right]. \quad (11)$$

## 3 Examples

### 3.1 Boltzmann Machine

**Definition 10.** [Boltzmann Machine]

Let  $\mathcal{M} = \{0, 1\}^n$ ,  $W \in \mathbb{R}^{(n \times n)}$  being symmetric,  $b \in \mathbb{R}^n$ ,  $\theta := (W, b)$ . Given dataset  $D := \{x_i | x_i \in \mathcal{M}, i = 1, \dots, N\}$ , denote expectation as  $\hat{x}$ . Then a Boltzmann machine is defined by energy function

$$E(x; W, b) := -\frac{1}{2} \sum_{\alpha, \beta} W_{\alpha\beta} (x^\alpha - \hat{x}^\alpha) (x^\beta - \hat{x}^\beta) - \sum_{\alpha} b_{\alpha} x^{\alpha}. \quad (12)$$

**Remark 11.** [MaxEnt Principle of BM]

Relating to MaxEnt principle, the observable that the model simulates is

$$\forall (\alpha, \beta), \mathbb{E}_{x \sim P_D} [(x^\alpha - \hat{x}^\alpha)(x^\beta - \hat{x}^\beta)], \quad (13)$$

for which it shall also simulate

$$\forall \alpha, \mathbb{E}_{x \sim P_D} [\hat{x}^\alpha]. \quad (14)$$

**Theorem 12.** [Activity Rule of BM]

For  $\forall \alpha$ ,

$$p_E(x_\alpha = 1 | x_{\setminus \alpha}) = \sigma \left( \sum_{\alpha \neq \beta} W_{\alpha\beta} (x^\beta - \hat{x}^\beta) + c_\alpha \right), \quad (15)$$

where  $c_\alpha := b_\alpha + (1/2 - \hat{x}^\alpha) W_{\alpha\alpha}$ . The sigmoid function  $\sigma := 1/(1 + e^{-x})$ . This relation is held for arbitrary replacement of the vector  $\hat{x}$ .

**Proof.** Directly, for  $\forall \gamma$ ,

$$\begin{aligned}
& \ln p(x_\gamma = 1 | x_{\setminus \gamma}) - \ln p(x_\gamma = 0 | x_{\setminus \gamma}) \\
[\alpha = \beta = \gamma] &= \frac{1}{2} W_{\gamma\gamma} (1 - \hat{x}^\gamma)^2 - \frac{1}{2} W_{\gamma\gamma} (-\hat{x}^\gamma)^2 \\
[\alpha \neq \gamma, \beta = \gamma] &+ \frac{1}{2} \sum_{\alpha \neq \gamma} W_{\alpha\gamma} (x^\alpha - \hat{x}^\alpha) (1 - \hat{x}^\gamma) - \frac{1}{2} \sum_{\alpha \neq \gamma} W_{\alpha\gamma} (x^\alpha - \hat{x}^\alpha) (-\hat{x}^\gamma) \\
[\alpha = \gamma, \beta \neq \gamma] &+ \frac{1}{2} \sum_{\beta \neq \gamma} W_{\gamma\beta} (1 - \hat{x}^\gamma) (x^\beta - \hat{x}^\beta) - \frac{1}{2} \sum_{\beta \neq \gamma} W_{\gamma\beta} (-\hat{x}^\gamma) (x^\beta - \hat{x}^\beta) \\
[\alpha, \beta \neq \gamma] &+ \frac{1}{2} \sum_{\alpha, \beta \neq \gamma} W_{\alpha\beta} (x^\alpha - \hat{x}^\alpha) (x^\beta - \hat{x}^\beta) - \frac{1}{2} \sum_{\alpha, \beta \neq \gamma} W_{\alpha\beta} (x^\alpha - \hat{x}^\alpha) (x^\beta - \hat{x}^\beta) \\
[\alpha = \gamma] &+ b^\gamma - 0 \\
[\alpha \neq \gamma] &+ \sum_{\alpha \neq \gamma} b_\gamma x^\gamma - \sum_{\alpha \neq \gamma} b_\gamma x^\gamma \\
&= \frac{1}{2} W_{\gamma\gamma} - W_{\gamma\gamma} \hat{x}^\gamma \\
&+ \frac{1}{2} \sum_{\alpha \neq \gamma} W_{\alpha\gamma} (x^\alpha - \hat{x}^\alpha) \\
&+ \frac{1}{2} \sum_{\beta \neq \gamma} W_{\gamma\beta} (x^\beta - \hat{x}^\beta) \\
&+ 0 \\
&+ b_\gamma \\
&+ 0 \\
&= \left( \frac{1}{2} - \hat{x}^\gamma \right) W_{\gamma\gamma} + \sum_{\alpha \neq \gamma} W_{(\gamma\alpha)} (x^\alpha - \hat{x}^\alpha) + b_\gamma
\end{aligned}$$

Thus

$$p(x_\gamma = 1 | x_{\setminus \gamma}) = \sigma \left[ \sum_{\alpha \neq \gamma} \frac{1}{2} (W_{\alpha\gamma} + W_{\gamma\alpha}) (x^\alpha - \hat{x}^\alpha) + \left( b_\gamma + \left( \frac{1}{2} - \hat{x}^\gamma \right) W_{\gamma\gamma} \right) \right]. \quad \square$$

**Theorem 13.** *[Learning Rule of BM]*

$$\sum_x p_D(x) x^\mu = \sum_x p_E(x) x^\mu, \quad (16)$$

and

$$\sum_x p_D(x) (x^\mu - \hat{x}^\mu) (x^\nu - \hat{x}^\nu) = \sum_x p_E(x) (x^\mu - \hat{x}^\mu) (x^\nu - \hat{x}^\nu).$$

### 3.2 Restricted Boltzmann Machine

**Definition 14.** *[Restricted Boltzmann Machine]*

Let  $\mathcal{V} = \{0, 1\}^{m_1}$  and  $\mathcal{H} = \{0, 1\}^{m_2}$ ,  $\mathcal{M} = \mathcal{V} \times \mathcal{H}$ . Let  $U \in \mathbb{R}^{(m_1 \times m_2)}$ ,  $b \in \mathbb{R}^{m_1}$ ,  $c \in \mathbb{R}^{m_2}$ . Then a restricted Boltzmann machine is defined by energy function<sup>1</sup>

$$E(v, h; U, b, c) := - \sum_{\alpha, i} U_{\alpha i} (v^\alpha - \hat{v}^\alpha) (h^i - \hat{h}^i) - \sum_{\alpha} b_{\alpha} v^{\alpha} - \sum_i c_i h^i. \quad (17)$$

**Remark 15.** *[Relation with Boltzmann machine]*

---

1. We use latin letters for latent variables.

By replacements in Boltzmann machine,

$$x \rightarrow (v, h), \quad (18)$$

$$b \rightarrow (b, c), \quad (19)$$

and

$$W \rightarrow \begin{pmatrix} 0 & U \\ U^T & 0 \end{pmatrix}, \quad (20)$$

we obtain the restricted Boltzmann machine.

**Theorem 16.** *[Activity Rule of RBM]*

*We have*

$$p(h_i = 1 | v_\alpha, h_{\setminus i}) = \sigma \left( \sum_{\alpha} U_{\alpha i} (v^\alpha - \hat{v}^\alpha) + c_i \right), \quad (21)$$

and

$$p(v_\alpha = 1 | v_{\setminus \alpha}, h_i) = \sigma \left( \sum_i U_{\alpha i} (h^i - \hat{h}^i) + b_\alpha \right). \quad (22)$$

**Theorem 17.** *[Effective Energy of RBM]*

*We have*

$$E_{\text{eff}}(v; U, b, c) = \sum_{\alpha} \left( \sum_i U_{\alpha i} v^\alpha \hat{h}^i - b_\alpha v^\alpha \right) - \sum_i s \left( \sum_{\alpha} U_{\alpha i} (v^\alpha - \hat{v}^\alpha) + c_i \right), \quad (23)$$

where soft-plus  $s$  is defined as

$$s(x) := \ln(1 + e^x). \quad (24)$$

**Proof.** Directly,

$$\begin{aligned} E_{\text{eff}}(v) &= -\ln \left( \prod_i \sum_{h^i=0,1} \right) \exp(-E(v, h)) \\ \{\text{Definition}\} &= -\ln \left( \prod_i \sum_{h^i=0,1} \right) \exp \left( \sum_{\alpha, i} U_{\alpha \beta} (v^\alpha - \hat{v}^\alpha) (h^i - \hat{h}^i) + \sum_{\alpha} b_{\alpha} v^\alpha + \sum_i c_i h^i \right) \\ \{\text{Extract } bv\} &= -\sum_{\alpha} b_{\alpha} v^\alpha - \ln \left( \prod_i \sum_{h^i=0,1} \right) \exp \left[ \sum_{\alpha, i} U_{\alpha \beta} (v^\alpha - \hat{v}^\alpha) (h^i - \hat{h}^i) + \sum_i c_i h^i \right] \\ \{\text{Combine}\} &= -\sum_{\alpha} b_{\alpha} v^\alpha - \ln \left( \prod_i \sum_{h^i=0,1} \right) \exp \left[ \sum_i \left( \sum_{\alpha} U_{\alpha i} (v^\alpha - \hat{v}^\alpha) \right) (h^i - \hat{h}^i) + \sum_i c_i h^i \right] \\ \{\exp \sum = \prod \exp\} &= -\sum_{\alpha} b_{\alpha} v^\alpha - \ln \prod_i \left[ \sum_{h^i=0,1} \exp \left( \sum_{\alpha} U_{\alpha i} (v^\alpha - \hat{v}^\alpha) (h^i - \hat{h}^i) + c_i h^i \right) \right] \\ \{\ln \prod = \sum \ln\} &= -\sum_{\alpha} b_{\alpha} v^\alpha - \sum_i \ln \sum_{h^i=0,1} \exp \left( \sum_{\alpha} U_{\alpha i} (v^\alpha - \hat{v}^\alpha) (h^i - \hat{h}^i) + c_i h^i \right). \end{aligned}$$

Since

$$\begin{aligned} & \sum_{h^i=0,1} \exp \left( \sum_{\alpha} U_{\alpha i} (v^\alpha - \hat{v}^\alpha) (h^i - \hat{h}^i) + c_i h^i \right) \\ &= \exp \left( \sum_{\alpha} U_{\alpha i} (v^\alpha - \hat{v}^\alpha) (1 - \hat{h}^i) + c_i \right) + \exp \left( \sum_{\alpha} U_{\alpha i} (v^\alpha - \hat{v}^\alpha) (-\hat{h}^i) \right) \\ \{\text{Extract}\} &= \exp \left( \sum_{\alpha} U_{\alpha i} (v^\alpha - \hat{v}^\alpha) (-\hat{h}^i) \right) \left[ \exp \left( \sum_{\alpha} U_{\alpha i} (v^\alpha - \hat{v}^\alpha) + c_i \right) + 1 \right], \end{aligned}$$

we have

$$\begin{aligned}
E_{\text{eff}}(v) & \\
\{\text{Previous}\} &= -\sum_{\alpha} b_{\alpha} v^{\alpha} - \sum_i \ln \sum_{h^i=0,1} \exp\left(\sum_{\alpha} U_{\alpha i} (v^{\alpha} - \hat{v}^{\alpha})(h^i - \hat{h}^i) + c_i h^i\right) \\
\{\text{Plugin}\} &= -\sum_{\alpha} b_{\alpha} v^{\alpha} + \sum_i \sum_{\alpha} U_{\alpha i} (v^{\alpha} - \hat{v}^{\alpha}) \hat{h}^i \\
&\quad - \sum_i \ln \left[ \exp\left(\sum_{\alpha} U_{\alpha i} (v^{\alpha} - \hat{v}^{\alpha}) + c_i\right) + 1 \right] \\
\{s(x) := \dots\} &= -\sum_{\alpha} b_{\alpha} v^{\alpha} + \sum_{\alpha, i} U_{\alpha i} (v^{\alpha} - \hat{v}^{\alpha}) \hat{h}^i - \sum_i s\left(\sum_{\alpha} U_{\alpha i} (v^{\alpha} - \hat{v}^{\alpha}) + c_i\right) \\
\{\text{Extract Const}\} &= -\sum_{\alpha} b_{\alpha} v^{\alpha} + \sum_{\alpha, i} U_{\alpha i} v^{\alpha} \hat{h}^i - \sum_i s\left(\sum_{\alpha} U_{\alpha i} (v^{\alpha} - \hat{v}^{\alpha}) + c_i\right) + \text{Const} \\
\{\text{Combine}\} &= \sum_{\alpha} \left( \sum_i U_{\alpha i} v^{\alpha} \hat{h}^i - b_{\alpha} v^{\alpha} \right) - \sum_i s\left(\sum_{\alpha} U_{\alpha i} (v^{\alpha} - \hat{v}^{\alpha}) + c_i\right) + \text{Const}.
\end{aligned}$$

The constant, which will be eliminated by  $Z$ , can be omitted.  $\square$

## 4 Perturbation Theory

### 4.1 Perturbation of Boltzmann Machine

Define  $p_i(x)$  by Taylor expansion  $p_E(x) = p_0(x) + p_1(x) + \dots + p_n(x) + \mathcal{O}(W^{n+1})$ . Denote  $\sigma_{\alpha} := \sigma(b_{\alpha})$ .

#### 4.1.1 0th-order

**Lemma 18.** [0th-order of Boltzmann Machine]

We have

$$p_0(x) = \prod_{\alpha} p_{\alpha}(x^{\alpha}), \quad (25)$$

where

$$p_{\alpha}(x) := \frac{\exp(b_{\alpha} x)}{1 + \exp(b_{\alpha})}. \quad (26)$$

**Proof.** Since  $E_0(x; W, b) := -\sum_{\alpha} b_{\alpha} x^{\alpha}$ ,

$$\begin{aligned}
p_0(x) &= \frac{\exp(\sum_{\alpha} b_{\alpha} x^{\alpha})}{\sum_{x'^1 \in \{0,1\}} \dots \sum_{x'^n \in \{0,1\}} \exp(\sum_{\alpha} b_{\alpha} x'^{\alpha})} \\
\{\exp \sum = \prod \exp\} &= \prod_{\alpha} \frac{\exp(b_{\alpha} x^{\alpha})}{\sum_{x'^{\alpha} \in \{0,1\}} \exp(b_{\alpha} x'^{\alpha})} \\
&= \prod_{\alpha} \frac{\exp(b_{\alpha} x^{\alpha})}{1 + \exp(b_{\alpha})} \\
&= \prod_{\alpha} p_{\alpha}(x).
\end{aligned}$$

$\square$

**Lemma 19.** We have

$$\frac{\partial p_{\alpha}}{\partial b_{\alpha}}(x) = p_{\alpha}(x)(x - \sigma_{\alpha}). \quad (27)$$

**Proof.** Directly,

$$\begin{aligned}\frac{\partial}{\partial b_\alpha} p_\alpha(x) &= \frac{\partial}{\partial b_\alpha} \frac{\exp(b_\alpha x)}{1 + \exp(b_\alpha)} \\ &= \frac{\exp(b_\alpha x)x}{1 + \exp(b_\alpha)} - \frac{\exp(b_\alpha x)[\exp(b_\alpha)]}{[1 + \exp(b_\alpha)]^2} \\ &= \frac{\exp(b_\alpha x)}{1 + \exp(b_\alpha)} \left[ x - \frac{\exp(b_\alpha)}{1 + \exp(b_\alpha)} \right] \\ &= p_\alpha(x)(x - \sigma(b_\alpha)).\end{aligned}$$

□

**Lemma 20.** For  $\forall \alpha$ , the mean of  $p_\alpha$   $V^\alpha := \sum_x p_0(x) x^\alpha$  is

$$V^\alpha = \sigma^\alpha. \quad (28)$$

**Proof.** Since  $(\partial p_\alpha / \partial b_\alpha)(x) = p_\alpha(x)(x - \sigma(b_\alpha))$ ,

$$\begin{aligned}\sum_x \frac{\partial}{\partial b_\alpha} p_\alpha(x) &= \sum_x p_\alpha(x)x - \sum_x p_\alpha(x)\sigma(b_\alpha) \\ \frac{\partial}{\partial b_\alpha} \sum_x p_\alpha(x) &= \sum_x p_\alpha(x)x - \left( \sum_x p_\alpha(x) \right) \sigma(b_\alpha) \\ 0 &= \sum_x p_\alpha(x)x - \sigma(b_\alpha).\end{aligned}$$

□

**Lemma 21.** Variance  $V^{\alpha_1 \alpha_2} := \sum_x p_0(x) (x - \sigma^{\alpha_1})(x - \sigma^{\alpha_2}) = \sum_x p_0(x) \prod_{i=1}^2 (x - \sigma^{\alpha_i})$  is

$$V^{\alpha_1 \alpha_2} = \delta^{\alpha_1 \alpha_2} \sigma^{\alpha_1} (1 - \sigma^{\alpha_1}). \quad (29)$$

**Proof.** Since  $(\partial p_\alpha / \partial b_\alpha)(x) = p_\alpha(x)(x - \sigma(b_\alpha))$ ,

$$\begin{aligned}\frac{\partial^2 p_0}{\partial b_\beta \partial b_\alpha}(x) &= \frac{\partial}{\partial b_\beta} [p_0(x)(x - \sigma^\alpha)] \\ &= p_0(x)(x - \sigma^\alpha)(x - \sigma^\beta) - \delta^{\alpha\beta} p_0(x) \sigma^\alpha (1 - \sigma^\alpha).\end{aligned}$$

Thus,

$$\begin{aligned}\sum_x \frac{\partial^2 p_0}{\partial b_\beta \partial b_\alpha}(x) &= \sum_x p_0(x)(x - \sigma^\alpha)(x - \sigma^\beta) - \sum_x \delta^{\alpha\beta} p_0(x) \sigma^\alpha (1 - \sigma^\alpha) \\ 0 &= \sum_x p_0(x)(x - \sigma^\alpha)(x - \sigma^\beta) - \delta^{\alpha\beta} \sigma^\alpha (1 - \sigma^\alpha) \\ \sum_x p_0(x)(x - \sigma^\alpha)(x - \sigma^\beta) &= \delta^{\alpha\beta} \sigma^\alpha (1 - \sigma^\alpha).\end{aligned}$$

□

**Lemma 22.** 3-momentum  $V^{\alpha_1 \alpha_2 \alpha_3} := \sum_x p_0(x) \prod_{i=1}^3 (x - \sigma^{\alpha_i})$  is

$$V^{\alpha_1 \alpha_2 \alpha_3} = \delta^{\alpha_1 \alpha_2 \alpha_3} \sigma^{\alpha_1} (1 - \sigma^{\alpha_1}) (1 - 2\sigma^{\alpha_1}). \quad (30)$$

**Lemma 23.** 4-momentum  $V^{\alpha_1 \dots \alpha_4} := \sum_x p_0(x) \prod_{i=1}^4 (x - \sigma^{\alpha_i})$  is

$$V^{\alpha_1 \dots \alpha_4} = V_c^{\alpha_1 \dots \alpha_4} + \sum_{\text{all pairs}} V^{\alpha_{m_1} \alpha_{m_2}} V^{\alpha_{n_1} \alpha_{n_2}}, \quad (31)$$

where “connected” part

$$V_c^{\alpha_1 \dots \alpha_4} := \delta^{\alpha_1 \dots \alpha_4} \sigma^{\alpha_1} (1 - \sigma^{\alpha_1}) [1 - 6\sigma^{\alpha_1} + 6(\sigma^{\alpha_1})^2], \quad (32)$$

and  $(m_1, m_2), (n_1, n_2)$  runs over all (three) pairs.

#### 4.1.2 1st-order

**Lemma 24.** For  $\forall \alpha$ ,

$$\hat{x}^\alpha = \sigma^\alpha + \mathcal{O}(W). \quad (33)$$

**Proof.** The gradient of loss gives

$$\begin{aligned}\sum_x p_D(x)x^\alpha &= \hat{x}^\alpha = \sum_x p_E(x)x^\alpha \\ \{\text{Taylor expand}\} &= \sum_x p_0(x)x^\alpha + \mathcal{O}(W) \\ \left\{ \sum_x p_0(x)x^\alpha = \sigma^\alpha \right\} &= \sigma^\alpha + \mathcal{O}(W).\end{aligned}$$

□

**Theorem 25.**

$$\frac{p_1(x)}{p_0(x)} = \frac{1}{2}W_{\alpha\beta}(x^\alpha - \sigma^\alpha)(x^\beta - \sigma^\beta) - \frac{1}{2}W_{\alpha\beta}V^{\alpha\beta}. \quad (34)$$

**Proof.** Directly,

$$\begin{aligned}p_E(x) &= \frac{\exp(b_\alpha x^\alpha + \frac{1}{2}W_{\alpha\beta}(x^\alpha - \hat{x}^\alpha)(x^\beta - \hat{x}^\beta))}{Z} \\ \{\text{Extract } b_\alpha x^\alpha\} &= \frac{\exp(b_\alpha x^\alpha) \exp(\frac{1}{2}W_{\alpha\beta}(x^\alpha - \hat{x}^\alpha)(x^\beta - \hat{x}^\beta))}{Z} \\ \{\text{Expand to } \mathcal{O}(W)\} &= \frac{\exp(b_\alpha x^\alpha) \{1 + \frac{1}{2}W_{\alpha\beta}(x^\alpha - \hat{x}^\alpha)(x^\beta - \hat{x}^\beta) + \dots\}}{Z_0(1 + Z_1 + \dots)} \\ \{p_0(x) = \dots\} &= p_0(x) \frac{\{1 + \frac{1}{2}W_{\alpha\beta}(x^\alpha - \hat{x}^\alpha)(x^\beta - \hat{x}^\beta) + \dots\}}{1 + Z_1 + \dots} \\ \left\{ \frac{1}{1+\epsilon} \sim 1 - \epsilon \right\} &= p_0(x) \left\{ 1 + \frac{1}{2}W_{\alpha\beta}(x^\alpha - \hat{x}^\alpha)(x^\beta - \hat{x}^\beta) + \dots \right\} \{1 - Z_1 + \dots\} \\ \{\text{Expand}\} &= p_0(x) \left\{ 1 + \frac{1}{2}W_{\alpha\beta}(x^\alpha - \hat{x}^\alpha)(x^\beta - \hat{x}^\beta) - Z_1 + \dots \right\} \\ &=: p_0(x) + p_1(x) + \dots\end{aligned}$$

Thus

$$\begin{aligned}\frac{p_1(x)}{p_0(x)} &= \frac{1}{2}W_{\alpha\beta}(x^\alpha - \hat{x}^\alpha)(x^\beta - \hat{x}^\beta) - Z_1 \\ \{\hat{x}^\alpha = \sigma^\alpha + \mathcal{O}(W)\} &= \frac{1}{2}W_{\alpha\beta}(x^\alpha - \sigma^\alpha)(x^\beta - \sigma^\beta) - Z_1.\end{aligned}$$

Now we compute  $Z_1$ . Since

$$\begin{aligned}1 &= \sum_x p_E(x) = \sum_x p_0(x) \left\{ 1 + \frac{1}{2}W_{\alpha\beta}(x^\alpha - \sigma^\alpha)(x^\beta - \sigma^\beta) - Z_1 \right\} \\ \left\{ \sum_x p_0(x) = 1 \right\} &= 1 + \frac{1}{2}W_{\alpha\beta} \left[ \sum_x p_0(x)(x^\alpha - \sigma^\alpha)(x^\beta - \sigma^\beta) \right] - Z_1 \\ \{V^{\alpha\beta} := \dots\} &= 1 + \frac{1}{2}W_{\alpha\beta}V^{\alpha\beta} - Z_1\end{aligned}$$

we have

$$Z_1 = \frac{1}{2}W_{\alpha\beta}V^{\alpha\beta}.$$

Then,

$$\begin{aligned}\frac{p_1(x)}{p_0(x)} &= \frac{1}{2}W_{\alpha\beta}(x^\alpha - \sigma^\alpha)(x^\beta - \sigma^\beta) - Z_1 \\ \{Z_1 = \dots\} &= \frac{1}{2}W_{\alpha\beta}(x^\alpha - \sigma^\alpha)(x^\beta - \sigma^\beta) - \frac{1}{2}W_{\alpha\beta}V^{\alpha\beta}.\end{aligned}$$

□

**Lemma 26.** Up to  $\mathcal{O}(W)$ , for  $\forall \gamma$ ,

$$\sum_x p_E(x)x^\gamma = V^\gamma + \frac{1}{2}W_{\alpha\beta}V^{\alpha\beta\gamma}. \quad (35)$$

**Proof.** Directly,

$$\begin{aligned}
\sum_x p_E(x) x^\gamma &= \sum_x p_0(x) x^\gamma + \sum_x p_1(x) x^\gamma \\
\{p_1(x) = \dots\} &= \sum_x p_0(x) x^\gamma + \sum_x p_0(x) \left[ \frac{1}{2} W_{\alpha\beta} (x^\alpha - \sigma^\alpha) (x^\beta - \sigma^\beta) - \frac{1}{2} W_{\alpha\alpha} \sigma^\alpha (1 - \sigma^\alpha) \right] x^\gamma \\
\{\text{Expand}\} &= \sum_x p_0(x) x^\gamma \\
&\quad + \frac{1}{2} W_{\alpha\beta} \sum_x p_0(x) (x^\alpha - \sigma^\alpha) (x^\beta - \sigma^\beta) x^\gamma \\
&\quad - \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta} \sum_x p_0(x) x^\gamma \\
&= \sum_x p_0(x) x^\gamma \\
\{\text{Combine}\} &+ \frac{1}{2} W_{\alpha\beta} \sum_x p_0(x) (x^\alpha - \sigma^\alpha) (x^\beta - \sigma^\beta) (x^\gamma - \sigma^\gamma) + \frac{1}{2} W_{\alpha\beta} \sum_x p_0(x) (x^\alpha - \sigma^\alpha) (x^\beta - \sigma^\beta) \sigma^\gamma \\
&\quad - \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta} \sum_x p_0(x) x^\gamma \\
&= V^\gamma \\
&\quad + \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta\gamma} + \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta} \sigma^\gamma \\
\{V^\gamma = \sigma^\gamma\} &- \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta} \sigma^\gamma \\
&= V^\gamma + \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta\gamma}.
\end{aligned}$$

□

**Lemma 27.** Up to  $\mathcal{O}(W)$ , for  $\forall(\mu, \nu)$ ,

$$\sum_x p_E(x) (x^\mu - \hat{x}^\mu) (x^\nu - \hat{x}^\nu) = V^{\mu\nu} + W_{(\alpha\beta)} V^{\alpha\mu} V^{\beta\nu} + \frac{1}{2} W_{\alpha\beta} V_c^{\alpha\beta\mu\nu}. \quad (36)$$

**Proof.** Directly,

$$\begin{aligned}
&\sum_x p_E(x) (x^\mu - \hat{x}^\mu) (x^\nu - \hat{x}^\nu) \\
\{p_E = p_0 + p_1\} &= \sum_x p_0(x) (x^\mu - \hat{x}^\mu) (x^\nu - \hat{x}^\nu) + \sum_x p_1(x) (x^\mu - \hat{x}^\mu) (x^\nu - \hat{x}^\nu) \\
&= \sum_x p_0(x) (x^\mu - \hat{x}^\mu) (x^\nu - \hat{x}^\nu) \\
\{p_1(x) = \dots\} &+ \sum_x p_0(x) \left[ \frac{1}{2} W_{\alpha\beta} (x^\alpha - \sigma^\alpha) (x^\beta - \sigma^\beta) - \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta} \right] (x^\mu - \hat{x}^\mu) (x^\nu - \hat{x}^\nu) \\
&= \sum_x p_0(x) (x^\mu - \hat{x}^\mu) (x^\nu - \hat{x}^\nu) \\
\{\text{Expand}\} &+ \frac{1}{2} W_{\alpha\beta} \sum_x p_0(x) (x^\alpha - \sigma^\alpha) (x^\beta - \sigma^\beta) (x^\mu - \hat{x}^\mu) (x^\nu - \hat{x}^\nu) \\
&\quad - \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta} \sum_x p_0(x) (x^\mu - \hat{x}^\mu) (x^\nu - \hat{x}^\nu) \\
\{\hat{x} = \dots\} &= \sum_x p_0(x) \left( x^\mu - \sigma^\mu - \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta\mu} \right) \left( x^\nu - \sigma^\nu - \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta\nu} \right) \\
\{\hat{x}^\alpha = \sigma^\alpha + \mathcal{O}(W)\} &+ \frac{1}{2} W_{\alpha\beta} \sum_x p_0(x) (x^\alpha - \sigma^\alpha) (x^\beta - \sigma^\beta) (x^\mu - \sigma^\mu) (x^\nu - \sigma^\nu)
\end{aligned}$$



$$\begin{aligned}
& \{\hat{x}^\alpha = \sigma^\alpha + \mathcal{O}(W)\} - \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta} \sum_x p_0(x) (x^\mu - \sigma^\mu) (x^\nu - \sigma^\nu) \\
& [\text{Expand}] = \sum_x p_0(x) (x^\mu - \sigma^\mu) (x^\nu - \sigma^\nu) \\
& \quad - \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta\nu} \sum_x p_0(x) (x^\mu - \sigma^\mu) \\
& \quad - \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta\mu} \sum_x p_0(x) (x^\nu - \sigma^\nu) \\
& \quad + \frac{1}{2} W_{\alpha\beta} \sum_x p_0(x) (x^\alpha - \sigma^\alpha) (x^\beta - \sigma^\beta) (x^\mu - \sigma^\mu) (x^\nu - \sigma^\nu) \\
& \quad - \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta} \sum_x p_0(x) (x^\mu - \sigma^\mu) (x^\nu - \sigma^\nu) \\
& \quad \{V^{\mu\nu} = \dots\} = V^{\mu\nu} \\
& \quad \{\sigma^\mu = V^\mu = \dots\} - 0 \\
& \quad \{\sigma^\nu = V^\nu = \dots\} - 0 \\
& \quad \{V^{\alpha\beta\mu\nu} = \dots\} + \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta\mu\nu} \\
& \quad \{V^{\mu\nu} = \dots\} - \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta} V^{\mu\nu} \\
& \quad = V^{\mu\nu} \\
& \quad \{V^{\alpha\beta\mu\nu} = V_c^{\alpha\beta\mu\nu} + \dots\} + \frac{1}{2} W_{\alpha\beta} (V_c^{\alpha\beta\mu\nu} + V^{\alpha\beta} V^{\mu\nu} + V^{\alpha\mu} V^{\beta\nu} + V^{\alpha\nu} V^{\beta\mu}) \\
& \quad - \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta} V^{\mu\nu} \\
& \quad = V^{\mu\nu} + \frac{1}{2} W_{\alpha\beta} (V_c^{\alpha\beta\mu\nu} + V^{\alpha\mu} V^{\beta\nu} + V^{\alpha\nu} V^{\beta\mu}) \\
& \quad \{\text{Combine}\} = V^{\mu\nu} + W_{(\alpha\beta)} V^{\alpha\mu} V^{\beta\nu} + \frac{1}{2} W_{\alpha\beta} V_c^{\alpha\beta\mu\nu}.
\end{aligned}$$

□

**Theorem 28.** *[Perturbation Solutions of BM]*

Define  $\hat{c}^\mu := \sigma^{-1}(\hat{x}^\mu)$  and  $\hat{C}^{\mu\nu} := \sum_x p_D(x) (x^\mu - \hat{x}^\mu) (x^\nu - \hat{x}^\nu)$ . Then, up to  $\mathcal{O}(W^2)$ , for  $\forall \mu$ ,

$$\hat{C}^{\mu\mu} = \hat{x}^\mu (1 - \hat{x}^\mu), \quad (37)$$

$$b_\mu = \hat{c}^\mu - W_{\mu\mu} \left( \frac{1}{2} - \hat{x}^\mu \right); \quad (38)$$

and for  $\forall \mu, \nu$  with  $\mu \neq \nu$ ,

$$W_{\mu\nu} = \frac{\hat{C}^{\mu\nu}}{\hat{x}^\mu (1 - \hat{x}^\mu) \hat{x}^\nu (1 - \hat{x}^\nu)}. \quad (39)$$

**Proof.** Here we prove the second declaration.

When  $\mu \neq \nu$ , we have

$$\begin{aligned}
\hat{C}^{\mu\nu} &= \sum_x p_E(x) (x^\mu - \hat{x}^\mu) (x^\nu - \hat{x}^\nu) \\
\{V^{\mu\nu} \propto \delta^{\mu\nu}\} &= W_{(\alpha\beta)} V^{\alpha\mu} V^{\beta\nu} \\
\{W \text{ symmetric}\} &= W_{\alpha\beta} V^{\alpha\mu} V^{\beta\nu} \\
\{V^{\alpha_1\alpha_2} = \delta^{\alpha_1\alpha_2} \sigma^{\alpha_1} (1 - \sigma^{\alpha_1})\} &= W_{\mu\nu} \sigma^\mu (1 - \sigma^\mu) \sigma^\nu (1 - \sigma^\nu) \\
\{\hat{x}^\alpha = \sigma^\alpha + \mathcal{O}(W)\} &= W_{\mu\nu} \hat{x}^\mu (1 - \hat{x}^\mu) \hat{x}^\nu (1 - \hat{x}^\nu)
\end{aligned}$$

thus, for  $\forall \mu \neq \nu$ ,

$$W_{\mu\nu} = \frac{\hat{C}^{\mu\nu}}{\hat{x}^\mu (1 - \hat{x}^\mu) \hat{x}^\nu (1 - \hat{x}^\nu)}.$$

And for  $\mu = \nu$ ,

$$\begin{aligned}
\hat{C}^{\mu\mu} &= \sum_x p_E(x) (x^\mu - \hat{x}^\mu) (x^\mu - \hat{x}^\mu) \\
\{W_{\mu\nu} \text{ symmetric}\} &= V^{\mu\mu} + W_{\alpha\beta} V^{\alpha\mu} V^{\beta\mu} + \frac{1}{2} W_{\alpha\beta} V_c^{\alpha\beta\mu\mu} \\
&= \sigma^\mu (1 - \sigma^\mu) \\
&\quad + W_{\alpha\beta} \delta^{\alpha\mu} \delta^{\beta\mu} [\sigma^\mu (1 - \sigma^\mu)]^2 \\
&\quad + \frac{1}{2} W_{\alpha\beta} \delta^{\alpha\beta\mu\mu} \sigma^\mu (1 - \sigma^\mu) [1 - 6\sigma^\mu + 6(\sigma^\mu)^2] \\
&= \sigma^\mu (1 - \sigma^\mu) \\
&\quad + W_{\mu\mu} [\sigma^\mu (1 - \sigma^\mu)]^2 \\
&\quad + \frac{1}{2} W_{\mu\mu} \sigma^\mu (1 - \sigma^\mu) [1 - 6\sigma^\mu + 6(\sigma^\mu)^2] \\
\{\hat{x} = \sigma + \dots\} &= \left( \hat{x}^\mu - \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta\mu} \right) \left( 1 - \hat{x}^\mu + \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta\mu} \right) \\
&\quad + W_{\mu\mu} [\sigma^\mu (1 - \sigma^\mu)]^2 \\
&\quad + \frac{1}{2} W_{\mu\mu} \sigma^\mu (1 - \sigma^\mu) [1 - 6\sigma^\mu + 6(\sigma^\mu)^2] \\
\{\text{Expand}\} &= \hat{x}^\mu (1 - \hat{x}^\mu) + W_{\alpha\beta} V^{\alpha\beta\mu} \left( \hat{x}^\mu - \frac{1}{2} \right) \\
&\quad + W_{\mu\mu} [\sigma^\mu (1 - \sigma^\mu)]^2 \\
&\quad + \frac{1}{2} W_{\mu\mu} \sigma^\mu (1 - \sigma^\mu) [1 - 6\sigma^\mu + 6(\sigma^\mu)^2] \\
\{V^{\alpha\beta\mu} = \dots\} &= \hat{x}^\mu (1 - \hat{x}^\mu) + W_{\mu\mu} \sigma^\mu (1 - \sigma^\mu) (1 - 2\sigma^\mu) \left( \hat{x}^\mu - \frac{1}{2} \right) \\
&\quad + W_{\mu\mu} [\sigma^\mu (1 - \sigma^\mu)]^2 \\
&\quad + \frac{1}{2} W_{\mu\mu} \sigma^\mu (1 - \sigma^\mu) [1 - 6\sigma^\mu + 6(\sigma^\mu)^2] \\
&= \hat{x}^\mu (1 - \hat{x}^\mu) + W_{\mu\mu} \hat{x}^\mu (1 - \hat{x}^\mu) (1 - 2\hat{x}^\mu) \left( \hat{x}^\mu - \frac{1}{2} \right) \\
[\hat{x}^\alpha = \sigma^\alpha + \mathcal{O}(W)] &+ W_{\mu\mu} [\hat{x}^\mu (1 - \hat{x}^\mu)]^2 \\
[\hat{x}^\alpha = \sigma^\alpha + \mathcal{O}(W)] &+ \frac{1}{2} W_{\mu\mu} \hat{x}^\mu (1 - \hat{x}^\mu) [1 - 6\hat{x}^\mu + 6(\hat{x}^\mu)^2] \\
\{\text{Combine}\} &= \hat{x}^\mu (1 - \hat{x}^\mu) \\
&\quad + W_{\mu\mu} \hat{x}^\mu (1 - \hat{x}^\mu) \times \\
&\quad \times \left\{ (1 - 2\hat{x}^\mu) \left( \hat{x}^\mu - \frac{1}{2} \right) + \hat{x}^\mu (1 - \hat{x}^\mu) + \frac{1}{2} [1 - 6\hat{x}^\mu + 6(\hat{x}^\mu)^2] \right\} \\
\{\text{Simplify}\} &= \hat{x}^\mu (1 - \hat{x}^\mu),
\end{aligned}$$

Thus,

$$\hat{C}^{\mu\nu} = \hat{x}^\mu (1 - \hat{x}^\mu) + \mathcal{O}(W^2).$$

Finally, we have, for  $\forall \mu$ ,

$$\begin{aligned}
\hat{x}^\mu &= V^\mu + \frac{1}{2} W_{\alpha\beta} V^{\alpha\beta\mu} \\
&= \sigma^\mu + \frac{1}{2} W_{\alpha\beta} \delta^{\alpha\beta\mu} \sigma^\alpha (1 - \sigma^\alpha) (1 - 2\sigma^\alpha) \\
&= \sigma^\mu + W_{\mu\mu} \sigma^\mu (1 - \sigma^\mu) \left( \frac{1}{2} - \sigma^\mu \right). \\
\{\hat{x}^\alpha = \sigma^\alpha + \mathcal{O}(W)\} &= \sigma^\mu + W_{\mu\mu} \hat{x}^\mu (1 - \hat{x}^\mu) \left( \frac{1}{2} - \hat{x}^\mu \right)
\end{aligned}$$

Thus

$$\sigma^\mu = \hat{x}^\mu - W_{\mu\mu} \hat{x}^\mu (1 - \hat{x}^\mu) \left( \frac{1}{2} - \hat{x}^\mu \right).$$

Since  $\sigma^\mu := \sigma(b_\mu)$  and  $\sigma'(\hat{c}^\mu) = \sigma(\hat{c}^\mu)(1 - \sigma(\hat{c}^\mu)) = \hat{x}^\mu(1 - \hat{x}^\mu)$ , we have

$$\begin{aligned} \sigma(b_\mu) &= \sigma(\hat{c}^\mu) - \sigma'(\hat{c}^\mu) W_{\mu\mu} \left( \frac{1}{2} - \hat{x}^\mu \right) \\ &= \sigma \left( \hat{c}^\mu - W_{\mu\mu} \left( \frac{1}{2} - \hat{x}^\mu \right) \right) + \mathcal{O}(W^2). \end{aligned}$$

Thus

$$b_\mu = \hat{c}^\mu - W_{\mu\mu} \left( \frac{1}{2} - \hat{x}^\mu \right).$$

□

**Corollary 29.** *[Solution without Self-interaction]*

If set, for  $\forall \mu, W_{\mu\mu} = 0$ , then, up to  $\mathcal{O}(W)$ , we have the perurbation solution of Boltzmann machine as follow.

For  $\forall \mu$ ,

$$b_\mu = \hat{c}^\mu, \quad (40)$$

and for  $\forall \mu, \nu$  with  $\mu \neq \nu$ ,

$$W_{\mu\nu} = \frac{\hat{C}^{\mu\nu}}{\hat{x}^\mu(1 - \hat{x}^\mu) \hat{x}^\nu(1 - \hat{x}^\nu)}. \quad (41)$$

**Lemma 30.** Let  $X^\mu, \mu = 1, \dots, N$  random variables. Then we have matrix

$$\frac{\text{Cov}(X^\mu, X^\nu)}{\text{Var}(X^\mu) \text{Var}(X^\nu)}$$

positive semi-definite.

**Proof.** Directly, define  $Z^\mu := X^\mu / \text{Var}[X^\mu]$ . Then, we have

$$\mathbb{E}[Z^\mu] = \frac{\mathbb{E}[X^\mu]}{\text{Var}[X^\mu]}.$$

Then,

$$\begin{aligned} \frac{\text{Cov}(X^\mu, X^\nu)}{\text{Var}(X^\mu) \text{Var}(X^\nu)} &= \frac{\mathbb{E}[(X^\mu - \mathbb{E}[X^\mu])(X^\nu - \mathbb{E}[X^\nu])]}{\text{Var}(X^\mu) \text{Var}(X^\nu)} \\ &= \mathbb{E} \left[ \frac{(X^\mu - \mathbb{E}[X^\mu])}{\text{Var}(X^\mu)} \frac{(X^\nu - \mathbb{E}[X^\nu])}{\text{Var}(X^\nu)} \right] \\ &= \mathbb{E}[(Z^\mu - \mathbb{E}[Z^\mu])(Z^\nu - \mathbb{E}[Z^\nu])] \\ &= \text{Cov}(Z^\mu, Z^\nu), \end{aligned}$$

which, as a covariance matrix, is positive semi-definite. □

**Lemma 31.** *[Eigenvalues of Covariance]<sup>2</sup>*

Let  $\{X^\mu | \mu = 1, \dots, N\}$  random variables. Let the  $\lambda_1, \dots, \lambda_N$  the eigenvalues of  $\text{Cov}(X^\mu, X^\nu)$ , with eigenvectors  $u_1, \dots, u_N$ . Let orthogonal matrix  $Q := (u_1, \dots, u_N)$ , and  $Y := QX$ . Then we have, for  $\forall \mu$ ,

$$\text{Var}(Y_\mu) = \lambda_\mu,$$

and for  $\forall \mu, \nu$  with  $\mu \neq \nu$ ,

$$\text{Cov}(Y_\mu, Y_\nu) = 0.$$

---

2. C.f. [this question](#) on stackexchange.com.

**Proof.** Denote  $C^{\mu\nu} := \text{Cov}(X^\mu, X^\nu)$ . From  $Q^T \Lambda Q = C$ , we get  $\Lambda = Q C Q^T$ . On the other hand,

$$\begin{aligned} \text{Cov}(Y_\mu, Y_\nu) &= \mathbb{E}[(Y_\mu - \mathbb{E}[Y_\mu])(Y_\nu - \mathbb{E}[Y_\nu])] \\ &= \mathbb{E}[(Q_{\mu\alpha}X^\alpha - \mathbb{E}[Q_{\mu\alpha}X^\alpha])(Q_{\nu\beta}X^\beta - \mathbb{E}[Q_{\nu\beta}X^\beta])] \\ &= Q_{\mu\alpha}\mathbb{E}[(X^\alpha - \mathbb{E}[X^\alpha])(X^\beta - \mathbb{E}[X^\beta])]Q_{\nu\beta} \\ &= Q_{\mu\alpha}C_{\alpha\beta}Q_{\nu\beta} \\ &= Q C Q^T. \end{aligned}$$

Thus, we get  $\Lambda = \text{Cov}(Y_\mu, Y_\nu)$ . We conclude that, for  $\forall \mu$ ,

$$\lambda_\mu = \text{Cov}(Y_\mu, Y_\mu) = \text{Var}(Y_\mu),$$

and for  $\forall \mu, \nu$  with  $\mu \neq \nu$ ,

$$\text{Cov}(Y_\mu, Y_\nu) = 0. \quad \square$$

**Corollary 32.** *[Solution with Positive Semi-definiteness]*

1. If set, for  $\forall \mu$ ,

$$W_{\mu\mu} = \frac{1}{\hat{x}^\mu(1 - \hat{x}^\mu)}, \quad (42)$$

then  $W_{\mu\nu}$  is positive semi-defined.

2. In this case, we find, for  $\forall \mu$ ,

$$\sigma^\mu = 2\hat{x}^\mu - \frac{1}{2}. \quad (43)$$

In addition, we shall check whether  $\sigma^\mu \in (0, 1)$  or not.

3. Perturbation demands

$$\left| \hat{x}^\mu - \frac{1}{2} \right| \ll \hat{x}^\mu.$$

**Proof.** Here we prove the declarations one by one.

1. Directly,

$$\begin{aligned} W_{\mu\mu} &= \frac{1}{\hat{x}^\mu(1 - \hat{x}^\mu)} \\ \{\hat{C}^{\mu\nu} = \hat{x}^\mu(1 - \hat{x}^\mu) + \mathcal{O}(W^2)\} &= \frac{\hat{C}^{\mu\mu}}{[\hat{x}^\mu(1 - \hat{x}^\mu)]^2}. \end{aligned}$$

Together with  $W_{\mu \neq \nu}$ , we find, for  $\forall (\mu, \nu)$ ,

$$W_{\mu\nu} = \frac{\text{Cov}[X^\mu, X^\nu]}{\text{Var}[X^\mu] \text{Var}[X^\nu]},$$

where we used  $\text{Var}[X^\mu] = \hat{x}^\mu(1 - \hat{x}^\mu)$ . This is positive semi-definite.

2. In this case,

$$\begin{aligned} \sigma^\mu &= \hat{x}^\mu - W_{\mu\mu}\hat{x}^\mu(1 - \hat{x}^\mu)\left(\frac{1}{2} - \hat{x}^\mu\right) \\ \{W_{\mu\mu} = \dots\} &= \hat{x}^\mu - \frac{1}{\hat{x}^\mu(1 - \hat{x}^\mu)}\hat{x}^\mu(1 - \hat{x}^\mu)\left(\frac{1}{2} - \hat{x}^\mu\right) \\ &= \hat{x}^\mu - \left(\frac{1}{2} - \hat{x}^\mu\right) \\ &= 2\hat{x}^\mu - \frac{1}{2}. \end{aligned}$$

3. Since perturbation demands

$$|\sigma^\mu - \hat{x}^\mu| \ll \hat{x}^\mu,$$

we get

$$\left| \hat{x}^\mu - \frac{1}{2} \right| \ll \hat{x}^\mu. \quad \square$$

## 4.2 Perturbation of Restricted Boltzmann Machine

**Lemma 33.** *[Perturbation of RBM]*

For  $\forall i$ , let  $\hat{h}^i \equiv 1/2$  and  $c_i \equiv 0$ , then we have

$$E_{\text{eff}}(v; U, b, c) = -\frac{1}{2} W_{\alpha\beta}^{\text{eff}} (v^\alpha - \hat{v}^\alpha)(v^\beta - \hat{v}^\beta) - b_\alpha^{\text{eff}} v^\alpha + \mathcal{O}(U^3 + c^3), \quad (44)$$

where

$$W_{\alpha\beta}^{\text{eff}} := \frac{1}{4} \sum_i U_{\alpha i} U_{\beta i}, \quad (45)$$

and

$$b_\alpha^{\text{eff}} := b_\alpha - \sum_i U_{\alpha i} v^\alpha \left( \hat{h}^i - \frac{1}{2} - \frac{c_i}{4} \right). \quad (46)$$

That is, restricted Boltzmann machine reduces to a Boltzmann machine.

**Proof.** Recall that

$$E_{\text{eff}}(v; U, b, c) = \sum_\alpha \left( \sum_i U_{\alpha i} v^\alpha \hat{h}^i - b_\alpha v^\alpha \right) - \sum_i s \left( \sum_\alpha U_{\alpha i} (v^\alpha - \hat{v}^\alpha) + c_i \right), \quad (47)$$

where soft-plus  $s$  is defined as

$$s(x) := \ln(1 + e^x). \quad (48)$$

Taylor expansion of soft-plus is

$$s(x) = 0 + \frac{x}{2} + \frac{x^2}{8} + \mathcal{O}(x^3).$$

Thus

$$\begin{aligned} E_{\text{eff}}(v) &= \sum_\alpha \left( \sum_i U_{\alpha i} v^\alpha \hat{h}^i - b_\alpha v^\alpha \right) \\ \{\text{Taylor expand}\} &- \frac{1}{2} \sum_i \left[ \sum_\alpha U_{\alpha i} (v^\alpha - \hat{v}^\alpha) + c_i \right] - \frac{1}{8} \sum_i \left[ \sum_\alpha U_{\alpha i} (v^\alpha - \hat{v}^\alpha) + c_i \right]^2 \\ &+ \mathcal{O}(U^3 + c^3) \\ \{\text{Expand}\} &= \sum_{\alpha, i} U_{\alpha i} v^\alpha \hat{h}^i - \sum_\alpha b_\alpha v^\alpha \\ &- \frac{1}{2} \sum_{\alpha, i} U_{\alpha i} (v^\alpha - \hat{v}^\alpha) - \frac{1}{2} \sum_i c_i \\ &- \frac{1}{8} \sum_{\alpha, \beta} \left( \sum_i U_{\alpha i} U_{\beta i} \right) (v^\alpha - \hat{v}^\alpha)(v^\beta - \hat{v}^\beta) - \frac{1}{4} \sum_{\alpha, i} U_{\alpha i} (v^\alpha - \hat{v}^\alpha) c_i - \frac{1}{8} \sum_i c_i^2 \\ &+ \mathcal{O}(U^3 + c^3) \\ \left[ \propto \sum_{\alpha, i} U_{\alpha i} v^\alpha \right] &= \sum_{\alpha, i} U_{\alpha i} v^\alpha \left( \hat{h}^i - \frac{1}{2} - \frac{c_i}{4} \right) \\ &- \sum_\alpha b_\alpha v^\alpha \\ &- \frac{1}{8} \sum_{\alpha, \beta} \left( \sum_i U_{\alpha i} U_{\beta i} \right) (v^\alpha - \hat{v}^\alpha)(v^\beta - \hat{v}^\beta) \\ [\text{Without } v] + \text{Const} &+ \mathcal{O}(U^3 + c^3) \\ &= -\frac{1}{2} \sum_{\alpha, \beta} \left[ \frac{1}{4} \left( \sum_i U_{\alpha i} U_{\beta i} \right) \right] (v^\alpha - \hat{v}^\alpha)(v^\beta - \hat{v}^\beta) \\ &- \sum_\alpha \left[ b_\alpha - \sum_i U_{\alpha i} v^\alpha \left( \hat{h}^i - \frac{1}{2} - \frac{c_i}{4} \right) \right] v^\alpha \\ &+ \text{Const} \\ &+ \mathcal{O}(U^3 + c^3). \end{aligned}$$

Omitting the constant, which will be eliminated by  $Z$ , we have

$$E_{\text{eff}}(v) = -\frac{1}{2}W_{\alpha\beta}^{\text{eff}}(v^\alpha - \hat{v}^\alpha)(v^\beta - \hat{v}^\beta) - b_\alpha^{\text{eff}}v^\alpha + \mathcal{O}(U^3 + c^3), \quad (49)$$

where

$$b_\alpha^{\text{eff}} := b_\alpha - \sum_i U_{\alpha i} v^\alpha \left( \hat{h}^i - \frac{1}{2} - \frac{c_i}{4} \right),$$

and

$$W_{\alpha\beta}^{\text{eff}} := \frac{1}{4} \sum_i U_{\alpha i} U_{\beta i}. \quad (50) \quad \square$$

**Theorem 34.** [Perturbation Equations of RBM]

Up to  $\mathcal{O}(U^3 + c^3)$ , we have, for  $\forall \mu$ ,

$$b_\mu - \sum_i U_{\mu i} v^\mu \left( \hat{h}^i - \frac{1}{2} - \frac{c_i}{4} \right) = \hat{c}^\mu - \frac{1}{4} \sum_i U_{\mu i} U_{\mu i} \left( \frac{1}{2} - \hat{x}^\mu \right), \quad (51)$$

and for  $\forall \mu, \nu$  with  $\mu \neq \nu$ ,

$$\frac{1}{4} \sum_i U_{\mu i} U_{\nu i} = \frac{\hat{C}^{\mu\nu}}{\hat{x}^\mu(1 - \hat{x}^\mu) \hat{x}^\nu(1 - \hat{x}^\nu)}. \quad (52)$$

The  $\hat{h}^i$  and  $c_i$  are free parameters, and the general setting is  $\hat{h}^i = 1/2$  and  $c_i = 0$  for  $\forall i = 1, \dots, m$ . Then the first equation reduce to, for  $\forall \mu$ ,

$$b_\mu = \hat{c}^\mu - \frac{1}{4} \sum_i U_{\mu i} U_{\mu i} \left( \frac{1}{2} - \hat{x}^\mu \right). \quad (53)$$

**Proof.** By the perturbation solution of BM, for  $\forall \mu$ ,

$$b_\alpha^{\text{eff}} = \hat{c}^\mu - W_{\mu\mu}^{\text{eff}} \left( \frac{1}{2} - \hat{x}^\mu \right) \\ \left\{ W_{\alpha\beta}^{\text{eff}} := \frac{1}{4} \sum_i U_{\alpha i} U_{\beta i} \right\} = \hat{c}^\mu - \frac{1}{4} \sum_i U_{\mu i} U_{\mu i} \left( \frac{1}{2} - \hat{x}^\mu \right),$$

and for  $\forall \mu, \nu$  with  $\mu \neq \nu$ ,

$$W_{\mu\nu}^{\text{eff}} = \frac{\hat{C}^{\mu\nu}}{\hat{x}^\mu(1 - \hat{x}^\mu) \hat{x}^\nu(1 - \hat{x}^\nu)} \\ \{\text{Definition}\} = \frac{1}{4} \sum_i U_{\mu i} U_{\nu i}.$$

$\square$

**Remark 35.** [Perturbation Solution of RBM]

1. This perturbation equations of RBM may not have a solution. Indeed, if  $i = 1$  only, then  $U_{\alpha i} \rightarrow U_{\alpha 1} =: u_\alpha$ . We have, for  $\forall \mu, \nu$  with  $\mu \neq \nu$ ,

$$\frac{1}{4} u_\mu u_\nu = \frac{\hat{C}^{\mu\nu}}{\hat{x}^\mu(1 - \hat{x}^\mu) \hat{x}^\nu(1 - \hat{x}^\nu)}.$$

There are  $n(n-1)/2$  equations but with  $n$  d.o.f.s, thus cannot be solved.

2. There is inner (gauge) symmetry. Indeed, if let  $\hat{h}^i = 1/2$  and  $c_i = 0$  for  $\forall i = 1, \dots, m$ , then for any orthogonal  $m \times m$  matrix  $E$ ,  $U \rightarrow UE$  leaves the equations invariant.

### 4.3 Validation of Perturbations

**Remark 36.** [Validation of Perturbations 1]

Based on the dimension analysis, it's suspected that the condition of validation of perturbation solution in the corollary 29 is

$$W_{\mu\nu} = \frac{\text{Cov}(X^\mu, X^\nu)}{\text{Var}(X^\mu)\text{Var}(X^\nu)} \ll \frac{1}{\sqrt{\text{Var}(X^\mu)\text{Var}(X^\nu)}}. \quad (54)$$

That is, the Pearson coefficients is tiny: for  $\forall \mu, \nu$  with  $\mu \neq \nu$ ,

$$\frac{\text{Cov}(X^\mu, X^\nu)}{\sqrt{\text{Var}(X^\mu)\text{Var}(X^\nu)}} \ll 1 \quad (55)$$

**Remark 37.** [Validation of Perturbations 2]

For making the perturbation stated in corollary 32 valid, the dataset shall have the properties, for  $\forall \alpha$ ,

$$\hat{x}^\alpha \approx 0.5 \quad (56)$$

and for  $\forall \alpha, \beta$  with  $\alpha \neq \beta$ ,

$$\hat{C}^{\alpha\beta} \approx 0. \quad (57)$$

Given a dataset of  $X^a$ , we construct a “soften version” of it,  $Y^a$ , s.t. this  $Y^a$  satisfies these properties.

**Definition 38.** [Zoom-in Trick]

Given Bernoulli random variable  $X$ , and a parameter  $\delta \in [0, 0.5)$ , we duplicate it to i.i.d. Bernoulli random variables  $Y_1, \dots, Y_m$ , s.t. for  $\forall i$

$$p(y_i = 1 | x = 0) = \delta, \quad (58)$$

and

$$p(y_i = 1 | x = 1) = 1 - \delta. \quad (59)$$

**Lemma 39.** We have, for  $\forall i$ ,

$$p(y_i = 1) = 0.5 + (2p - 1)(0.5 - \delta), \quad (60)$$

where  $p := p(x = 1)$ .

**Theorem 40.** [Zoom-in Trick]

Let  $\epsilon := 0.5 - \delta > 0$ . We have, for  $\forall(\alpha, i)$ ,

$$\lim_{\epsilon \rightarrow 0} \hat{y}^{(\alpha, i)} = 0,$$

and for  $\forall(\alpha, i), (\beta, j)$  with  $(\alpha, i) \neq (\beta, j)$ ,

$$\lim_{\epsilon \rightarrow 0} \hat{C}^{(\alpha, i)(\beta, j)} = 0.$$

Specifically for the first limit, we have  $\hat{y}^{(\alpha, i)} \sim \mathcal{N}(\mu, \sigma)$  where

$$\mu := 0.5 + (2\hat{x}^\alpha - 1)(0.5 - \delta), \quad (61)$$

and

$$\sigma := \sqrt{\frac{0.25 - [(2\hat{x}^\alpha - 1)(0.5 - \delta)]^2}{N}},$$

with  $N$  the data-size.

**Proof.** The first limit can be derived from the  $\hat{y}^{(\alpha, i)} \sim \mathcal{N}(\mu, \sigma)$ .

The second limit can be proved by considering the limit case, where  $\delta \rightarrow 0.5$ . In this situation, for  $\forall(\alpha, i)$ ,  $y^{(\alpha, i)} \sim \text{Bernoulli}(0.5)$ . Thus all independent, leading to  $\hat{C}^{(\alpha, i)(\beta, j)} = 0$ .  $\square$

## Appendix A Perturbations by Temperature

Let  $\beta := 1/T$ . Then inserting temperature is replacements  $U \rightarrow \beta U$ ,  $b \rightarrow \beta b$ ,  $c \rightarrow \beta c$ , and  $E_{\text{eff}}(v) \rightarrow -\beta^{-1}E_{\text{eff}}(v)$ .

Thus,

$$\begin{aligned}
E_{\text{eff}}(v; \beta) &= \sum_{\alpha} \left( \sum_i U_{\alpha i} v^{\alpha} \hat{h}^i - b_{\alpha} v^{\alpha} \right) - \beta^{-1} \sum_i s \left( \beta \sum_{\alpha} U_{\alpha i} (v^{\alpha} - \hat{v}^{\alpha}) + \beta c_i \right) \\
&= \sum_{\alpha} \left( \sum_i U_{\alpha i} v^{\alpha} \hat{h}^i - b_{\alpha} v^{\alpha} \right) \\
[\text{Taylor expand}] &- \frac{1}{2} \sum_i \left[ \sum_{\alpha} U_{\alpha i} (v^{\alpha} - \hat{v}^{\alpha}) + c_i \right] - \frac{\beta}{8} \sum_i \left[ \sum_{\alpha} U_{\alpha i} (v^{\alpha} - \hat{v}^{\alpha}) + c_i \right]^2 + \mathcal{O}(\beta^2) \\
&= \sum_{\alpha, i} U_{\alpha i} v^{\alpha} \hat{h}^i - \sum_{\alpha} b_{\alpha} v^{\alpha} \\
&\quad - \frac{1}{2} \sum_{\alpha, i} U_{\alpha i} (v^{\alpha} - \hat{v}^{\alpha}) \\
&\quad - \frac{\beta}{8} \sum_i \sum_{\alpha, \beta} U_{\alpha i} (v^{\alpha} - \hat{v}^{\alpha}) U_{\beta i} (v^{\beta} - \hat{v}^{\beta}) - \frac{\beta}{4} \sum_{\alpha, i} U_{\alpha i} (v^{\alpha} - \hat{v}^{\alpha}) c_i \\
&\quad + \text{Const} \\
&\quad + \mathcal{O}(\beta^2) \\
&= \sum_{\alpha, i} U_{\alpha i} (v^{\alpha} - \hat{v}^{\alpha}) \left( \hat{h}^i - \frac{\beta}{4} c_i - \frac{1}{2} \right) \\
&\quad - \frac{\beta}{8} \sum_i \sum_{\alpha, \beta} U_{\alpha i} (v^{\alpha} - \hat{v}^{\alpha}) U_{\beta i} (v^{\beta} - \hat{v}^{\beta}) - \sum_{\alpha} b_{\alpha} v^{\alpha} \\
&\quad + \text{Const} \\
&\quad + \mathcal{O}(\beta^2).
\end{aligned}$$

Let  $\hat{h}^i \equiv 1/2$  and  $c_i \equiv 0$  for  $\forall i$ , and omit the constant, then

$$E_{\text{eff}}(v; \beta) = - \sum_{\alpha, \beta} \left( \frac{\beta}{8} \sum_i U_{\alpha i} U_{\beta i} \right) (v^{\alpha} - \hat{v}^{\alpha}) (v^{\beta} - \hat{v}^{\beta}) - \sum_{\alpha} b_{\alpha} v^{\alpha} + \mathcal{O}(\beta^2). \quad (62)$$

Thus,

$$W_{\alpha\beta}^{\text{eff}} \rightarrow \frac{\beta}{8} \sum_i U_{\alpha i} U_{\beta i},$$

and

$$b_{\alpha}^{\text{eff}} \rightarrow b_{\alpha}. \quad (63)$$

$$\frac{p_1(x)}{p_0(x)} = \beta E(x) - \beta \sum_{\alpha} \left( \frac{W_{\alpha\alpha}}{4} + \frac{b_{\alpha}}{2} \right) + \mathcal{O}(\beta^2).$$