1 Energy-based Model

Definition 1. *[Energy-based Model]*

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

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

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 θ.

Theorem 2. *[Universality]*

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

$$E_q(x) := -\ln q(x) + C, \tag{2}$$

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

$$q(x) = \frac{\exp(-E_q(x))}{\int_{\mathrm{supp}(q)} \mathrm{d}x' \exp(-E_q(x'))}. \tag{3}$$

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

Proof. Directly,

$$\begin{aligned} q(x) &= \frac{\exp(-E_q(x))}{\int_{\mathrm{supp}(q)} \mathrm{d}x' \exp(-E_q(x'))} \\ \{E_q := \cdots\} &= \frac{q(x)}{\int_{\mathrm{supp}(q)} \mathrm{d}x'\, q(x')} \\ \left\{\int_{\mathrm{supp}(q)} \mathrm{d}x'\, q(x') = 1\right\} &= q(x). \end{aligned}$$

□

Theorem 3. *[Maximum Entropy Principle]*

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

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

s.t. contrains

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

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 $p_D: \mathcal{M} \to \mathbb{R}$, define Lagrangian $L(\theta; p_D) := -\int_{\mathcal{M}} \mathrm{d}x\, p_D(x) \ln p_E(x;\theta)$. Then, the gradient of Lagrangian w.r.t. component θ^α is

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

or in more compact format,

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

2 Effective Theory

Definition 6. *[Effective Energy]*

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

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

Theorem 7. *[Effective Theory]*
Recall that $p_{E_{\text{eff}}}(v;\theta) := \int_{\mathcal{H}} \mathrm{d}h\, p(v,h;\theta)$. Then,

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

Lemma 8. *[Gradient of Effective Energy]*

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

Theorem 9. *[Learning Rule of Effective Theory]*
For any probability density $p_D: \mathcal{V}\to\mathbb{R}$, define Lagrangian $L(\theta;p_D) := -\int_{\mathcal{V}} \mathrm{d}v\, p_D(v) \ln p(v;\theta)$. Then, the gradient of Lagrangian w.r.t. component θ^α is

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

or in more compact format,

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

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 $D:=\{x_i|x_i\in\mathcal{M}, i=1,\ldots,N\}$, denote expectation as $\hat{x}$. Then a Boltzmann machine is defined by energy function

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

Remark 11. [MaxEnt Principle of BM]
Relating to MaxEnt principle, the observable that the model simulates is

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

for which it shall also simulate

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

Theorem 12. *[Activity Rule of BM]*
For $\forall\alpha$,

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

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

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

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

Thus

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

Theorem 13. *[Learning Rule of BM]*

$$\sum_x p_D(x)x^\mu = \sum_x p_E(x)x^\mu, \tag{16}$$

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[1]

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

Remark 15. [Relation with Boltzmann machine]

By replacements in Boltzmann machine,

$$x\to(v,h), \tag{18}$$

$$b\to(b,c), \tag{19}$$

and

$$W\to\begin{pmatrix}0 & U\\ U^T & 0\end{pmatrix}, \tag{20}$$

we obtain the restricted Boltzmann machine.

Theorem 16. *[Activity Rule of RBM]*

We have

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

1. We use latin letters for latent variables.

and

$$p(v_\alpha = 1 | v_{\backslash\alpha}, h_i) = \sigma\Big(\sum_i U_{\alpha i}\big(h^i - \hat{h}^i\big) + b_\alpha\Big). \tag{22}$$

Theorem 17. *[Effective Energy of RBM]*

We have

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

where soft-plus s is defined as

$$s(x) := \ln(1 + \mathrm{e}^x). \tag{24}$$

Proof. Directly,

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

Since

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

we have

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

The constant, which will be eliminated by Z, can be omitted. □

4 Perturbation Theory

4.1 Perturbation of Boltzmann Machine

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

4.1.1 0th-order

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

We have

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

where

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

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

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

□

Lemma 19. *We have*

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

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_α $V^\alpha := \sum_x p_0(x)\, x^\alpha$ is*

$$V^\alpha = \sigma^\alpha. \tag{28}$$

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*

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

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,

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

$$\sum_x p_0(x)(x - \sigma^\alpha)(x - \sigma^\beta) = \delta^{\alpha\beta} \sigma^\alpha (1 - \sigma^\alpha).$$

□

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*

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

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

$$V^{\alpha_1\cdots\alpha_4} = V_c^{\alpha_1\cdots\alpha_4} + \sum_{\textit{all pairs}} V^{\alpha_{m_1}\alpha_{m_2}} V^{\alpha_{n_1}\alpha_{n_2}}, \tag{31}$$

where "connected" part

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

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

4.1.2 1st-order

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

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

Proof. The gradient of loss gives

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

□

Theorem 25.

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

Proof. Directly,

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

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

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

we have

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

Then,

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

□

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

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

Proof. Directly,

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

□

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

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

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)=\cdots\} &\quad + \sum_x p_0(x)\Big[\frac{1}{2}W_{\alpha\beta}(x^\alpha-\sigma^\alpha)(x^\beta-\sigma^\beta) - \frac{1}{2}W_{\alpha\beta}V^{\alpha\beta}\Big](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}\} &\quad + \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}=\cdots\} &= \sum_x p_0(x)\Big(x^\mu-\sigma^\mu-\frac{1}{2}W_{\alpha\beta}V^{\alpha\beta\mu}\Big)\Big(x^\nu-\sigma^\nu-\frac{1}{2}W_{\alpha\beta}V^{\alpha\beta\nu}\Big) \\
\{\hat{x}^\alpha=\sigma^\alpha+\mathcal{O}(W)\} &\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) \\
\{\hat{x}^\alpha=\sigma^\alpha+\mathcal{O}(W)\} &\quad - \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) \\
\{V^{\mu\nu}=\cdots\} &= V^{\mu\nu} \\
\{\sigma^\mu=V^\mu=\cdots\} &\quad -0 \\
\{\sigma^\nu=V^\nu=\cdots\} &\quad -0 \\
\{V^{\alpha\beta\mu\nu}=\cdots\} &\quad + \frac{1}{2}W_{\alpha\beta}V^{\alpha\beta\mu\nu} \\
\{V^{\mu\nu}=\cdots\} &\quad - \frac{1}{2}W_{\alpha\beta}V^{\alpha\beta}V^{\mu\nu} \\
&= V^{\mu\nu} \\
\{V^{\alpha\beta\mu\nu}=V_c^{\alpha\beta\mu\nu}+\cdots\} &\quad + \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} \\
&= 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}) \\
\{\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 $\forall\mu$,

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

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

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

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

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 + \cdots\} &= \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} = \cdots\} &= \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)] &\quad + W_{\mu\mu}[\hat{x}^\mu(1 - \hat{x}^\mu)]^2\\
[\hat{x}^\alpha = \sigma^\alpha + \mathcal{O}(W)] &\quad + \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

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

□

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

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

For $\forall \mu$*,*

$$b_\mu = \hat{c}^\mu, \tag{40}$$

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

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

4.2 Perturbation of Restricted Boltzmann Machine

Lemma 30. *[Perturbation of RBM]*

We have

$$E_{\text{eff}}(v; U, b, c) = -\frac{1}{2} W^{\text{eff}}_{\alpha\beta} (v^\alpha - \hat{v}^\alpha)(v^\beta - \hat{v}^\beta) - b^{\text{eff}}_\alpha v^\alpha + \mathcal{O}(U^3 + c^3), \tag{42}$$

where

$$W^{\text{eff}}_{\alpha\beta} := \frac{1}{4} \sum_i U_{\alpha i} U_{\beta i}, \tag{43}$$

and

$$b^{\text{eff}}_\alpha := b_\alpha - \sum_i U_{\alpha i} v^\alpha \Big(\hat{h}^i - \frac{1}{2} - \frac{c_i}{4}\Big). \tag{44}$$

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

Proof. Recall that

$$E_{\text{eff}}(v; U, b, c) = \sum_\alpha \Big(\sum_i U_{\alpha i} v^\alpha \hat{h}^i - b_\alpha v^\alpha\Big) - \sum_i s\Big(\sum_\alpha U_{\alpha i} (v^\alpha - \hat{v}^\alpha) + c_i\Big), \tag{45}$$

where soft-plus s is defined as

$$s(x) := \ln(1 + \mathrm{e}^x). \tag{46}$$

Taylor expansion of soft-plus is

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

Thus

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

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

$$E_{\text{eff}}(v) = -\frac{1}{2}W_{\alpha\beta}^{\text{eff}}(v^\alpha - \hat{v}^\alpha)(v^\beta - \hat{v}^\beta) - b_\alpha^{\text{eff}} v^\alpha + \mathcal{O}(U^3 + c^3), \tag{47}$$

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

$$W_{\alpha\beta}^{\text{eff}} := \frac{1}{4}\sum_i U_{\alpha i} U_{\beta i}. \tag{48}$$

□

Theorem 31. *[Perturbation Equations of RBM]*
Up to $\mathcal{O}(U^3 + c^3)$, we have, for $\forall \mu$,

$$b_\mu - \sum_i U_{\mu i} v^\mu \left(\hat{h}^i - \frac{1}{2} - \frac{c_i}{4}\right) = \hat{c}^\mu - \frac{1}{4}\sum_i U_{\mu i} U_{\mu i}\left(\frac{1}{2} - \hat{x}^\mu\right), \tag{49}$$

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

$$\frac{1}{4}\sum_i U_{\mu i} U_{\nu i} = \frac{\hat{C}^{\mu\nu}}{\hat{x}^\mu(1-\hat{x}^\mu)\,\hat{x}^\nu(1-\hat{x}^\nu)}. \tag{50}$$

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 $\forall i = 1, \ldots, m$. Then the first equation reduce to, for $\forall \mu$,

$$b_\mu = \hat{c}^\mu - \frac{1}{4}\sum_i U_{\mu i} U_{\mu i}\left(\frac{1}{2} - \hat{x}^\mu\right). \tag{51}$$

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

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

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

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

□

Lemma 32. *[Positive Semi-definiteness of Covariance]*
Let X^μ, $\mu = 1, \ldots, N$ 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 33. *[Eigenvalues of Covariance]*[2]
Let $\{X^\mu | \mu = 1, \ldots, n\}$ random variables. Then we have:
$\exists \{a_{i\mu} \in \mathbb{R}, b_i \in \mathbb{R} | i = 1, \ldots, m, \mu = 1, \ldots, n\}$ s.t. for $\forall i$, $\sum_\nu a_{i\nu} X^\nu = b_i$, iff there exists m vanished eigenvalues in the covariance matrix of $\{X^\mu | \mu = 1, \ldots, n\}$.

Proof. Let $C^{\mu\nu} := \text{Cov}(X^\mu, X^\nu)$.

1. Proof of $\Rightarrow$

2. C.f. this question on stackexchange.com.

Directly,

$$\begin{aligned}\sum_\mu a_{i\mu} C^{\mu\nu} &= \sum_\mu a_{i\mu}\mathrm{Cov}(X^\mu, X^\nu)\\ &= \sum_\mu a_{i\mu}\mathbb{E}[(X^\mu - \mathbb{E}[X^\mu])(X^\nu - \mathbb{E}[X^\nu])]\\ &= \mathbb{E}\left[\left(\sum_\mu a_{i\mu} X^\mu - \mathbb{E}\left[\sum_\mu a_{i\mu} X^\mu\right]\right)(X^\nu - \mathbb{E}[X^\nu])\right]\\ &= \mathbb{E}[(b_i - b_i)(X^\nu - \mathbb{E}[X^\nu])]\\ &= 0.\end{aligned}$$

That is, a_i is an eigenvector of C with vanished eigenvalue.

2. Proof of $\Leftarrow$

From diagonalization $Q^T \Lambda Q = C$, where Q is orthogonal, we get $\Lambda = Q C Q^T$. On the other hand, let $Y := Q X$, we have

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

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

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

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

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

Then, if $\exists \lambda_i = 0$, then $\mathrm{Var}(Y_i) = 0$, implying $Y_i = \mathrm{Const} =: b_i$. Denote $a_{i\mu} := Q_{i\mu}$, then we find $a_{i\mu} X^\mu = b_i$. □

Theorem 34. *[Perturbation Solution of RBM]*

Let m is the number of independent variables in $\{X^\mu | \mu = 1, ..., n\}$, then we have a solution

1. *$W_{\mu\nu}$ has m positive eigenvalues and $n - m$ vanised ones. Let them be $\lambda_1, ..., \lambda_m$, with eigenvectors $u_1, ..., u_m$, Then, for $\forall \mu, i$,*

$$U_{\mu i} = 2\sqrt{\lambda_i} u_i^\mu. \tag{52}$$

2. *And, for $\forall \mu$,*

$$b^\mu = \sigma^{-1}\left(2\,\hat{x}^\mu - \frac{1}{2}\right). \tag{53}$$

3. *Perturbation demands*

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

Proof. Set

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

1. Recalling $\mathrm{Var}(X^\mu) = \hat{x}^\mu(1 - \hat{x}^\mu)$, we have, for $\forall \mu, \nu$,

$$W_{\mu\nu}^{\mathrm{eff}} = \frac{\mathrm{Cov}(X^\mu, X^\nu)}{\mathrm{Var}(X^\mu)\mathrm{Var}(X^\nu)}.$$

Then, by the lemma 32 and the lemma 33, we find that $W_{\mu\nu}$ is positive semi-definite, having m positive eigenvalues and $n - m$ vanised ones. Let $U_{\mu i} = 2\sqrt{\lambda_i} u_i^\mu$, we find

$$\begin{aligned}\frac{1}{4}\sum_i U_{\mu i} U_{\nu i} &= \sum_i \lambda_i u_i^\mu u_i^\nu\\ &= W_{\mu\nu}^{\mathrm{eff}}.\end{aligned}$$

Thus, the equations of U are satisfied.

2. From the equations of b,

$$\begin{aligned}\sigma(b_\mu) &= \sigma\left(\hat{c}^\mu - \frac{1}{4}\sum_i U_{\mu i} U_{\mu i}\left(\frac{1}{2} - \hat{x}^\mu\right)\right)\\ &= \sigma(\hat{c}^\mu) - \sigma'(\hat{c}^\mu)\frac{1}{4}\sum_i U_{\mu i} U_{\mu i}\left(\frac{1}{2} - \hat{x}^\mu\right)\\ \{\sigma'(\hat{c}^\mu) = \hat{x}^\mu(1 - \hat{x}^\mu)\} &= \sigma(\hat{c}^\mu) - \hat{x}^\mu(1 - \hat{x}^\mu)\frac{1}{4}\sum_i U_{\mu i} U_{\mu i}\left(\frac{1}{2} - \hat{x}^\mu\right)\\ \left\{W_{\mu\mu}^{\mathrm{eff}} = \frac{1}{4}\sum_i U_{\mu i} U_{\mu i}\right\} &= \hat{x}^\mu - \hat{x}^\mu(1 - \hat{x}^\mu) W_{\mu\mu}^{\mathrm{eff}}\left(\frac{1}{2} - \hat{x}^\mu\right)\\ \left\{W_{\mu\mu}^{\mathrm{eff}} = \frac{1}{\hat{x}^\mu(1 - \hat{x}^\mu)}\right\} &= \hat{x}^\mu - \left(\frac{1}{2} - \hat{x}^\mu\right)\\ &= 2\,\hat{x}^\mu - \frac{1}{2}.\end{aligned}$$

Thus

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

4.3 Validation of Perturbations

Remark 35. [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

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

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

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

Remark 36. [Validation of Perturbations 2]

For making the perturbation stated in theorem 34 valid, the dataset shall have the properties, for $\forall \alpha$,

$$\hat{x}^\alpha \approx 0.5 \tag{56}$$

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

$$\hat{C}^{\alpha\beta} \approx 0. \tag{57}$$

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

Definition 37. *[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 $Y_1, ..., Y_m$, s.t. for $\forall i$

$$p(y_i = 1 | x = 0) = \delta, \tag{58}$$

and

$$p(y_i = 1 | x = 1) = 1 - \delta. \tag{59}$$

Lemma 38. *We have, for $\forall i$,*

$$p(y_i = 1) = 0.5 + (2p - 1)(0.5 - \delta), \tag{60}$$

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

Theorem 39. *[Zoom-in Trick]*

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

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

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

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

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

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

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 $\delta \to 0.5$. In this situation, for $\forall (\alpha, i)$, $y^{(\alpha,i)} \sim \mathrm{Bernoulli}(0.5)$. Thus all independent, leading to $\hat{C}^{(\alpha,i)(\beta,j)} = 0$. □

Appendix A Perturbations by Temperature

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

Thus,

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

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

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

Thus,

$$W^{\text{eff}}_{\alpha\beta} \to \frac{\beta}{8}\sum_i U_{\alpha i} U_{\beta i},$$

and

$$b^{\text{eff}}_\alpha \to b_\alpha. \tag{63}$$

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