Statistical Mechanical Analysis of Low-Density Parity-Check Codes on General Markov Channel

Ryuhei Mori and Toshiyuki Tanaka

SITA2011
Iwate
30 November

Concept

It has been shown that

Large deviations theory (method of types) is useful for understanding the result of the replica method [Mori 2011].

In this work,

Large deviations theory (method of types) for Markov chain is applied to models including a Markov structure.

Types [Csiszár 1977]

- $\mathcal{X}$: a finite set
- $P_{\mathbf{x}}(a)$: the fraction of $a \in \mathcal{X}$ in $\mathbf{x} \in \mathcal{X}^N$

Example:
For $\mathcal{X} = \{a, b, c\}$, $\mathbf{x} = [a, a, b, a, c, c, a, b]$,

$$P_{\mathbf{x}}(a) = 4/8, \quad P_{\mathbf{x}}(b) = 2/8, \quad P_{\mathbf{x}}(c) = 2/8.$$

$$\mathcal{P}_N(\mathcal{X}) := \{P_{\mathbf{x}} \mid \mathbf{x} \in \mathcal{X}^N\}, \qquad |\mathcal{P}_N(\mathcal{X})| = \binom{N+|\mathcal{X}|-1}{|\mathcal{X}|-1}$$

Number of sequences of a particular type

$\mathcal{T}_P(N) := \{\mathbf{x} \in \mathcal{X}^N \mid P_{\mathbf{x}} = P\}$

$$|\mathcal{T}_P(N)| = \binom{N}{NP(a) \quad NP(b) \quad NP(c)} := \frac{N!}{(NP(a))!(NP(b))!(NP(c))!}$$

$$|\mathcal{T}_P(n)| \approx \exp\{NH(P)\}$$

Usage:

$$\sum_{\mathbf{x} \in \mathcal{X}^N} f(P_{\mathbf{x}}) = \sum_{P \in \mathcal{P}_N(\mathcal{X})} |T_P(N)| f(P)$$

$$\sum_{\mathbf{x} \in \{0,1\}^N} f(\mathbf{x}) = \sum_{i=0}^{N} \binom{N}{i} f(i)$$

Sanov's theorem

$$
\begin{aligned}
Q^N(\mathbf{x}) &= \prod_{a \in \mathcal{X}} Q(a)^{N P_{\mathbf{x}}(a)} \\
&= \exp\left\{ N \sum_{a \in \mathcal{X}} P_{\mathbf{x}}(a) \log Q(a) \right\} \\
&= \exp\{-N[H(P_{\mathbf{x}}) + D(P_{\mathbf{x}} \| Q)]\}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left\{N g(P_{X_1 X_2 \ldots X_N})\right\}\right] = \sum_{\boldsymbol{x} \in \mathcal{X}^N} Q^N(\boldsymbol{x}) \exp\left\{N g(P_{\boldsymbol{x}})\right\} \\
&= \sum_{P \in \mathcal{P}^N(\mathcal{X})} |\mathcal{T}_P(N)| \exp\{-N[H(P) + D(P \| Q)]\} \exp\left\{N g(P)\right\} \\
&\approx \sum_{P \in \mathcal{P}(\mathcal{X})} \exp\{-N(D(P \| Q) - g(P))\} \\
&\approx \sup_{P \in \mathcal{P}(\mathcal{X})} \exp\{-N(D(P \| Q) - g(P))\} \quad (\text{Laplace method})
\end{aligned}
$$

The second order types

- $\mathcal{X}$: a finite set
- $P_{\mathbf{x}}^{(2)}(a, b)$: the fraction of a pair of successive symbols $(a, b) \in \mathcal{X}^2$ in $\mathbf{x} \in \mathcal{X}^N$

Example:
For $\mathcal{X} = \{a, b, c\}$, $\mathbf{x} = [a, a, b, a, c, c, a, b]$

$P_{\mathbf{x}}^{(2)}(a, a) = 1/7$, $P_{\mathbf{x}}^{(2)}(a, b) = 2/7$, $P_{\mathbf{x}}^{(2)}(a, c) = 1/7$,
$P_{\mathbf{x}}^{(2)}(b, a) = 1/7$, $P_{\mathbf{x}}^{(2)}(b, b) = 0/7$, $P_{\mathbf{x}}^{(2)}(b, c) = 0/7$,
$P_{\mathbf{x}}^{(2)}(c, a) = 1/7$, $P_{\mathbf{x}}^{(2)}(c, b) = 0/7$, $P_{\mathbf{x}}^{(2)}(c, c) = 1/7$.

$$\mathcal{P}_N^{(2)}(\mathcal{X}) := \{P_{\mathbf{x}}^{(2)} \mid \mathbf{x} \in \mathcal{X}^N\}, \qquad |\mathcal{P}_N^{(2)}(\mathcal{X})| \sim d(|\mathcal{X}|) N^{|\mathcal{X}|^2 - |\mathcal{X}|}.$$

Number of sequence of particular type

$\mathcal{T}^{(\mathbf{2})}_{P_{X,Y}}(N) := \{\mathbf{x} \in \mathcal{X}^N \mid P^{(\mathbf{2})}_{\mathbf{x}} = P_{X,Y}\}$

$$|\mathcal{T}^{(\mathbf{2})}_{P_{X,Y}}(N)| = C \prod_{x\in\mathcal{X}} \binom{NP_X(x)}{\{NP_{X,Y}(x,y)\}_{y\in\mathcal{X}}}.$$

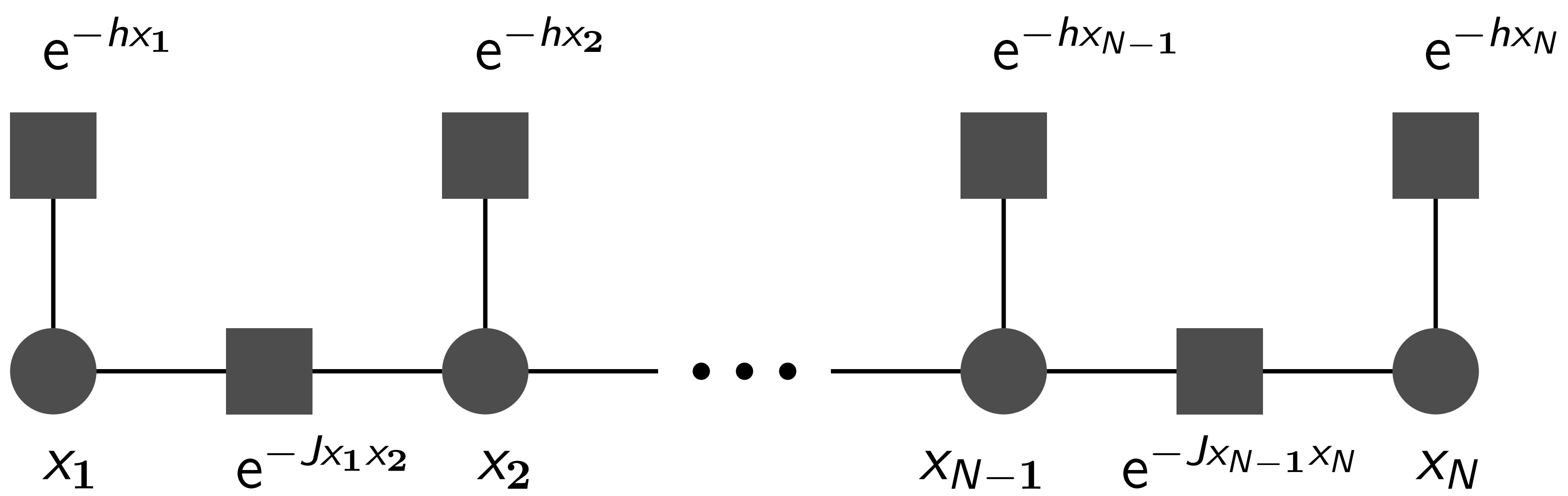
[Whittle 1955] [Billingsley 1961].

$$|\mathcal{T}^{n}_{P_{X,Y}}| \approx \exp\{NH(X \mid Y)\}, \qquad P_X \approx P_Y$$

One-dimensional Ising model

$$p(\mathbf{x}) := \frac{1}{Z(N)} \exp\left\{ -J \sum_{i=1}^{N-1} x_i x_{i+1} - h \sum_{i=1}^{N} x_i \right\}$$

$$Z(N) := \sum_{\mathbf{x} \in \{+1,-1\}^N} \exp\left\{ -J \sum_{i=1}^{N-1} x_i x_{i+1} - h \sum_{i=1}^{N} x_i \right\}$$



The method of transfer matrix

$$Z_N(x_1, x_N) := \sum_{\mathbf{x}\in\{+1,-1\}^N} \exp\left\{-J\sum_{i=1}^{N-1} x_i x_{i+1} - h\sum_{i=1}^{N} x_i\right\}.$$

$$\begin{bmatrix} Z_N(+1,+1) & Z_N(+1,-1) \\ Z_N(-1,+1) & Z_N(-1,-1) \end{bmatrix}$$

$$= \begin{bmatrix} Z_{N-1}(+1,+1) & Z_{N-1}(+1,-1) \\ Z_{N-1}(-1,+1) & Z_{N-1}(-1,-1) \end{bmatrix} \begin{bmatrix} \exp\{-J-h\} & \exp\{+J+h\} \\ \exp\{+J-h\} & \exp\{-J+h\} \end{bmatrix}$$

$$= \begin{bmatrix} \exp\{-h\} & 0 \\ 0 & \exp\{+h\} \end{bmatrix} \begin{bmatrix} \exp\{-J-h\} & \exp\{+J+h\} \\ \exp\{+J-h\} & \exp\{-J+h\} \end{bmatrix}^{N-1}$$

$$Z(N) = \sum_{(x_1,x_N)\in\mathcal{X}^2} Z_N(x_1, x_N) \sim \boldsymbol{\lambda}_{\mathbf{max}}^N.$$

The method of types for Markov chain

$$Z(N) = \sum_{\mathbf{x}\in\{+1,-1\}^N} \exp\left\{-J\sum_{i=1}^{N-1} x_i x_{i+1} - h\sum_{i=1}^{N} x_i\right\}$$

$$= \sum_{P_{S,T}\in\mathcal{P}_N^{(2)}} \left|\mathcal{T}^{(2)}_{P_{S,T}}(N)\right|$$

$$\cdot \exp\left\{-J\sum_{(s,t)\in\{+1,-1\}^2}(N-1)P_{S,T}(s,t)st - h\sum_{t\in\{+1,-1\}} NP_T(t)t\right\}$$

$$\lim_{N\to\infty}\frac{1}{N}\log Z(N) = \sup_{P_{ST},P_S=P_T}\{H(S\mid T) - J\langle ST\rangle - h\langle T\rangle\}$$

$$= \sup_{P_{ST},P_S=P_T}\{H(S,T) - H(T) - J\langle ST\rangle - h\langle T\rangle\}$$

The maximization problem can be solved by
the method of Lagrange multiplier.

Free energy of 1d Ising model 1/2

Lemma 1.

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N = \operatorname*{supextr}_{m_{\mathrm{LR}\to\mathrm{v}}} \{\log Z_{\mathrm{w}} - \log Z_{\mathrm{v}}\} .$$

where supextr *stands for supremum among saddle points.*

$$Z_{\mathrm{w}} = \sum_{(s,t)\in\{+1,-1\}^2} m_{\mathrm{LR}\to\mathrm{v}}(t) m_{\mathrm{LR}\to\mathrm{v}}(s) \exp\{-Jst - hs - ht\}$$

$$Z_{\mathrm{v}} = \sum_{t\in\{+1,-1\}} m_{\mathrm{LR}\to\mathrm{v}}(t)^2 \exp\{-ht\} .$$

The saddle point equation is

$$m_{\mathrm{LR}\to\mathrm{v}}(t) = \frac{1}{Z_{\mathrm{LR}\to\mathrm{v}}} \sum_{s\in\{+1,-1\}} m_{\mathrm{LR}\to\mathrm{v}}(s) \exp\{-Jst - hs\} .$$

This is the equation of belief propagation on the 1d Ising model of infinite size !

Free energy of 1d Ising model 2/2

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N = \log Z_{\mathsf{LR}\to\mathsf{v}}$$

where

$$m_{\mathsf{LR}\to\mathsf{v}}(t) = \frac{1}{Z_{\mathsf{LR}\to\mathsf{v}}} \sum_{(s,t)\in\{+1,-1\}^2} m_{\mathsf{LR}\to\mathsf{v}}(s) \exp\left\{-Jst - hs\right\}.$$

Here, $Z_{\mathsf{LR}\to\mathsf{v}}$ and $m_{\mathsf{LR}\to\mathsf{v}}$ are eigenvalue and eigenvector of

$$\begin{bmatrix} \exp\{-J-h\} & \exp\{+J+h\} \\ \exp\{+J-h\} & \exp\{-J+h\} \end{bmatrix}$$

which is the transfer matrix. Hence,

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N = \log \lambda_{\max}.$$

The method of types is useful for more complicated problems.

LDPC codes on memoryless channel

$$p(\mathbf{x} \mid \mathbf{y}) := \frac{1}{Z} \prod_a f(\mathbf{x}_{\partial a}) \prod_{i=1}^{N} W(y_i \mid x_i)$$

$$Z := \sum_{\mathbf{x} \in \mathcal{X}^N} \prod_a f(\mathbf{x}_{\partial a}) \prod_{i=1}^{N} W(y_i \mid x_i).$$

$$f(\mathbf{x}) := \mathbb{I}\left\{ \bigoplus_j x_j = 0 \right\}$$

$$p(\mathbf{y}) := \frac{1}{Z_0} \sum_{\mathbf{x} \in \mathcal{X}^N} \prod_a f(\mathbf{x}_{\partial a}) \prod_{i=1}^{N} W(y_i \mid x_i)$$

$$Z_0 := \sum_{\mathbf{x} \in \mathcal{X}^N} \prod_a f(\mathbf{x}_{\partial a}).$$

Conditional entropy and free energy

$$p(\mathbf{x} \mid \mathbf{y}) := \frac{1}{Z} \prod_a f(\mathbf{x}_{\partial a}) \prod_{i=1}^N W(y_i \mid x_i)$$

$$Z := \sum_{\mathbf{x} \in \mathcal{X}^N} \prod_a f(\mathbf{x}_{\partial a}) \prod_{i=1}^N W(y_i \mid x_i).$$

$$\mathbb{E}[H(X \mid Y)] = \mathbb{E}[\log Z] - \mathbb{E}[\log W(Y \mid X)]$$

Disordered system and replica method

$$\lim_{N\to\infty}\frac{1}{N}\mathbb{E}[\log Z] = \lim_{N\to\infty}\frac{1}{N}\left.\frac{\partial \log \mathbb{E}[Z^n]}{\partial n}\right|_{n=0}$$

$$= \lim_{N\to\infty}\frac{1}{N}\lim_{n\to 0}\frac{1}{n}\log\mathbb{E}[Z^n] \overbrace{=}^{?} \lim_{n\to 0}\frac{1}{n}\lim_{N\to\infty}\frac{1}{N}\log\mathbb{E}[Z^n]$$

For non-negative integer n,

$$Z^n = \left(\sum_{\boldsymbol{x}\in\mathcal{X}^N}\prod_a f(\boldsymbol{x}_{\partial a})\right)^n = \sum_{\boldsymbol{x}\in(\mathcal{X}^n)^N}\prod_a\left(\prod_{i=1}^n f(\boldsymbol{x}_{\partial a}^{(i)})\right).$$

Z^n can be regarded as a partition function of a new model in which

$$\mathcal{X} \longrightarrow \mathcal{X}^n$$

$$f(\boldsymbol{x}) \longrightarrow \prod_{i=1}^n f(\boldsymbol{x}^{(i)}).$$

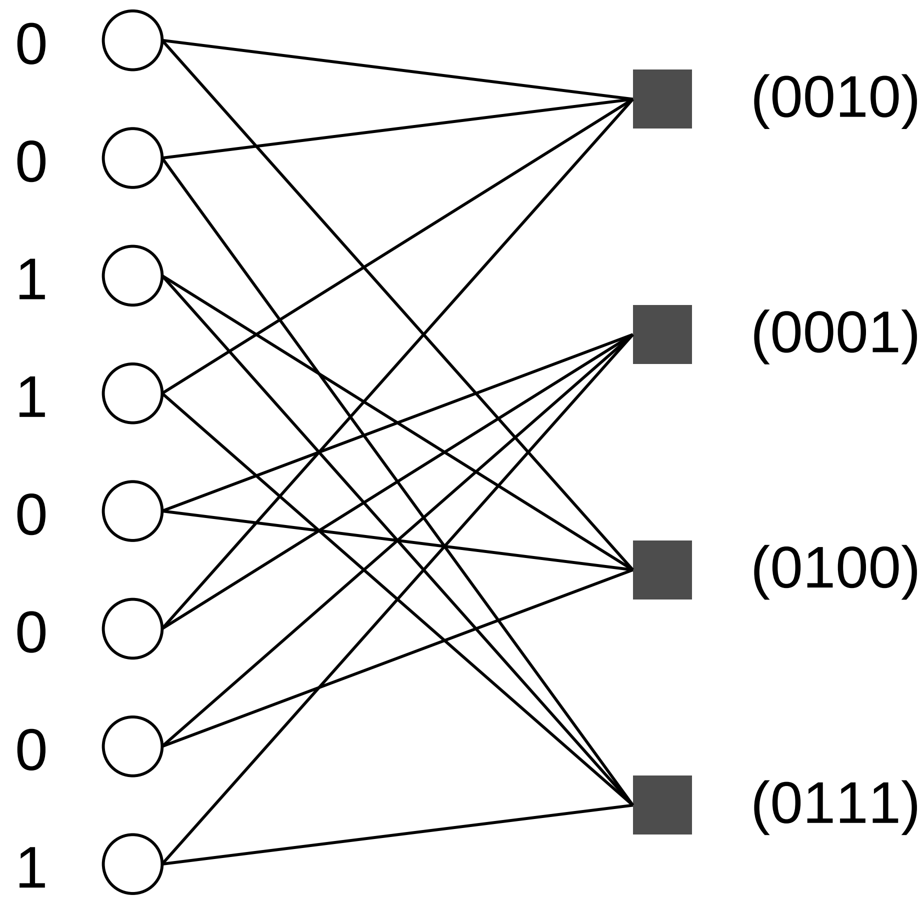
Types on factor graphs [Vontobel 2010]

$\boldsymbol{\nu}(x), x \in \mathcal{X}$: a type of variable nodes
$\boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{X}^r$: a type of factor nodes

$\mathcal{X} = \{0, 1\}$.

$\boldsymbol{\nu}(0) = 5/8$,

$\boldsymbol{\nu}(1) = 3/8$.



$\boldsymbol{\mu}(0010) = 1/4$,

$\boldsymbol{\mu}(0001) = 1/4$,

$\boldsymbol{\mu}(0100) = 1/4$,

$\boldsymbol{\mu}(0111) = 1/4$,

Otherwise,
$\boldsymbol{\mu}(\boldsymbol{x}) = 0$.

There is a constraint between $\boldsymbol{\nu}(x)$ and $\boldsymbol{\mu}(\boldsymbol{x})$.
More precisely, $\boldsymbol{\nu}(x)$ is uniquely determined from $\boldsymbol{\mu}(\boldsymbol{x})$.

Contribution of particular types to a partition function

$$Z = \sum_{\boldsymbol{x} \in \mathcal{X}^N} \prod_a f(\boldsymbol{x}_{\partial a})$$
$$= \sum_{\boldsymbol{\nu}, \boldsymbol{\mu}} N(\boldsymbol{\nu}, \boldsymbol{\mu}) \prod_{\boldsymbol{x} \in \mathcal{X}^r} f(\boldsymbol{x})^{\frac{\ell}{r} N \boldsymbol{\mu}(\boldsymbol{x})} =: \sum_{\boldsymbol{\nu}, \boldsymbol{\mu}} Z(\boldsymbol{\nu}, \boldsymbol{\mu}).$$

$$\mathbb{E}[N(\boldsymbol{\nu}, \boldsymbol{\mu})] = \binom{N}{\{N\boldsymbol{\nu}(x)\}_{x \in \mathcal{X}}} \binom{\frac{\ell}{r} N}{\{\frac{\ell}{r} N \boldsymbol{\mu}(\boldsymbol{x})\}_{\boldsymbol{x} \in \mathcal{X}^r}} \frac{\prod_{x \in \mathcal{X}} (N\boldsymbol{\nu}(x)\ell)!}{(N\ell)!}.$$

$$\lim_{N \to \infty} \frac{1}{N} \log \mathbb{E}[Z(\boldsymbol{\nu}, \boldsymbol{\mu})]$$
$$= \frac{\ell}{r} \mathcal{H}(\boldsymbol{\mu}) - (\ell - 1) \mathcal{H}(\boldsymbol{\nu}) + \frac{\ell}{r} \sum_{\boldsymbol{x} \in \mathcal{X}^r} \boldsymbol{\mu}(\boldsymbol{x}) \log f(\boldsymbol{x}).$$

Minus Bethe free energy of of mini (averaged) model [Mori 2011].

Free energy of LDPC codes on memoryless channel

$$\lim_{N\to\infty}\frac{1}{N}\log\mathbb{E}[Z^n] = \sup_{P_X,P_{U_1,\ldots,U_r}}\left\{\frac{l}{r}H(U_1,\ldots,U_r)-(l-1)H(X)\right.$$
$$\left.+\frac{l}{r}\left\langle\log\prod_{k=0}^{n}f(\mathbf{U}^{(k)})\right\rangle+\left\langle\log\left(\sum_{y\in\mathcal{Y}}\prod_{k=0}^{n}W(y\mid X^{(k)})\right)\right\rangle\right\}-R$$

Here, X and $U_1,\ldots,U_r$ are random variables on $\mathcal{X}^{n+1}$ satisfying

- X and U_K have the same distribution

where K denotes the uniform random variable on a set $\{1,\ldots,r\}$.

The saddle point equation for replica symmetric solution is equivalent to the density evolution of the belief propagation [Mori 2011].

LDPC codes on general Markov channel

$\mathcal{S}$: a set of states
$V(t \mid y, x, s)$: a transition probability for $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $s, t \in \mathcal{S}$

$$p(\mathbf{x} \mid \mathbf{y}) := \frac{1}{Z} \sum_{\mathbf{s} \in \mathcal{S}^N} \prod_a f(\mathbf{x}_{\partial a}) \prod_{i=1}^{N} W(y_i \mid x_i, s_i) V_0(s_1) \prod_{i=1}^{N-1} V(s_{i+1} \mid y_i, x_i, s_i)$$

$$Z := \sum_{\mathbf{x} \in \mathcal{X}^N} \sum_{\mathbf{s} \in \mathcal{S}^N} \prod_a f(\mathbf{x}_{\partial a}) \prod_{i=1}^{N} W(y_i \mid x_i, s_i) \cdot V_0(s_1) \prod_{i=1}^{N-1} V(s_{i+1} \mid y_i, x_i, s_i).$$

Free energy of LDPC codes on general Markov channel

Main result of this work

$$\lim_{N\to\infty}\frac{1}{N}\log\mathbb{E}[Z^n] = \sup\Bigg\{ H(X_1,S_1\mid X_2,S_2) - lH(X_1,S_1)$$

$$+\frac{l}{r}H(U_1,\dots,U_r,T_1,\dots,T_r)+\frac{l}{r}\left\langle \log\prod_{k=0}^{n} f(\mathbf{U}^{(k)})\right\rangle$$

$$+\left\langle \log\left(\sum_{y\in\mathcal{Y}}\prod_{k=0}^{n} W(y\mid X_2^{(k)},S_2^{(k)})V(S_1^{(k)}\mid y,X_2^{(k)},S_2^{(k)})\right)\right\rangle\Bigg\} - R.$$

- (X_1,S_1) and (X_2,S_2) have the same distribution
- (X_1,S_1) and (U_K,T_K) have the same distribution

where K denotes the uniform random variable on a set $\{1,\dots,r\}$.

The saddle point equation is equivalent to the density evolution of the belief propagation (joint iterative decoder).

The dicode erasure channel

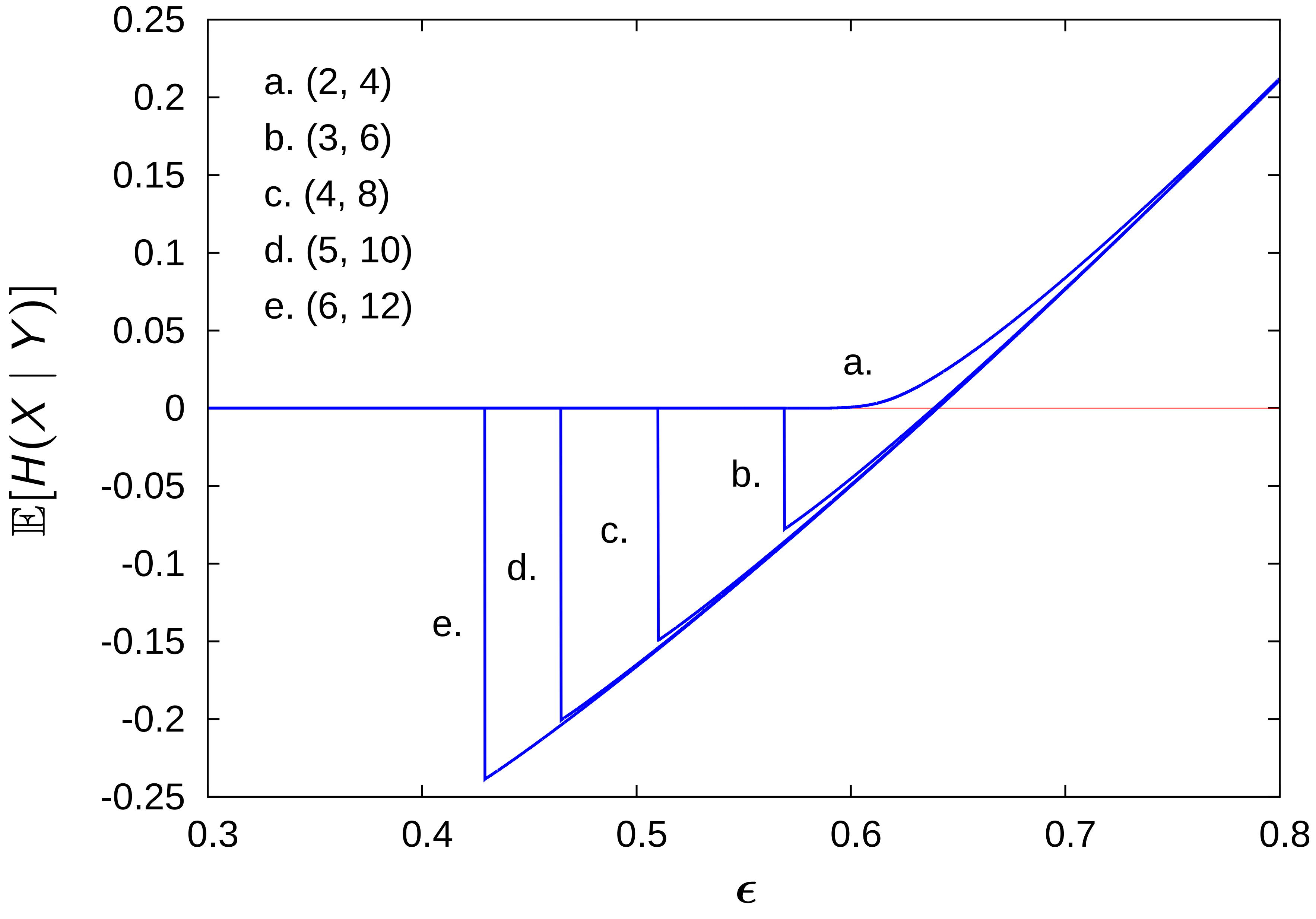
DEC(ϵ) is defined for $\mathcal{X} = \mathcal{S} = \{0, 1\}$, $\mathcal{Y} = \{-1, 0, +1, *\}$ as

$$W(y \mid x, s) = \begin{cases} 1 - \epsilon, & y = x - s \\ \epsilon, & y = * \end{cases}$$

$$V(s' \mid y, x, s) = 1, \qquad \text{for } s' = x.$$

The density evolution can be described by one parameter [Pfister and Siegel 2008].

Numerical calculation for the DEC(ϵ)



Summary, future works and open problem

Summary

- The method of types is useful for analysis of LDPC codes on memoryless channels (previous result)
- The method of types for Markov chain is useful for analysis of LDPC codes on Markov channels

Future works

- Analysis of IRA/ARA LDPC codes
- Compressed sensing of Markov source

Open problem

- Types for two-dimensional Markov chain e.g., two dimensional Ising model.