MEXT Grant-in-Aid for Scientific Research on Priority Areas
"Statistical Mechanical Approach to Probabilistic Information Processing"
Workshop on Mathematics of Statistical Inference (December 2004, Tohoku University, Sendai, Japan)

An LDPCC decoding algorithm based on Bowman-Levin approximation

— Contrast with BP and CCCP —

Masato Inoue [1], Miho Komiya [2] and Yoshiyuki Kabashima [3]

Department of Computational Intelligence and Systems Science,

Interdisciplinary Graduate School of Science and Engineering,

Tokyo Institute of Technology, Yokohama 226-8502, Japan

Belief propagation (BP) and the concave convex procedure (CCCP) are methods that can be used successfully to minimize the Bethe free energy of graphical models. The mechanisms of these methods, however, have not been well elucidated. We propose a method based on the Bowman-Levin (BL) approximation, which is another equation of the extrema of the Bethe free energy, so that we can compare the mechanisms of these methods and develop potentially better algorithms. We validated the BL algorithm through the decoding problem of low density parity check codes (LDPCC), and found that BL empirically ascends the Bethe free energy by means of simple iterated substitutions; BP and CCCP, on the other hand, descend the Bethe free energy. We also found that the gradient method causes BL to successfully converge.

Introduction

Recently, various statistical inference algorithms have become of interest in the field of large-scale information processing. Belief propagation (BP)[1] and the concave convex procedure (CCCP)[2] are among the most effective of these methods which minimize the Bethe free energy[3, 4]. BP and CCCP have both been used to successfully decode low density parity check codes (LDPCC) [5]. The mechanisms of these methods, however, have not yet been well elucidated. Specifically, we wonder why BP minimizes the Bethe free energy although this is not a *direct* aim of the BP method. We also wonder why CCCP successfully converges to a local minimum of the *true* Bethe free energy although CCCP descends the *modified* Bethe free energy.

In this paper, we introduce a new algorithm based on the Bowman-Levin (BL) equation[6], which minimizes the Bethe free energy. The Bowman-Levin equation is widely used in the field of statistical

[1]E-mail: inoue@sp.dis.titech.ac.jp
[2]E-mail: miho@sp.dis.titech.ac.jp
[3]E-mail: kaba@dis.titech.ac.jp

physics as a method that can find an extremum (saddle, local minimum, or local maximum) of the Bethe free energy. Through a comparison of BL with BP and CCCP, we hoped to elucidate the mechanisms of these methods.

LDPCC

The LDPCC decoding problem can be handled within a Bayesian framework. The prior probability of the codes, consisting of N binary spins ($\boldsymbol{x} \in \{+1, -1\}^N$), is defined as

$$P(\boldsymbol{x}) \propto \prod_{\mu}^{M} \left(1+\prod_{l\in\boldsymbol{\mu}} x_l \right), \tag{1}$$

where $\mu = 1, ..., M$ denotes the parity index and $\boldsymbol{\mu}$ denotes the spin indices involved in the μ-th parity. The proportion means the normalization of a probability function – i.e., the summation of $P(\boldsymbol{x})$ for all possible arguments $\boldsymbol{x}$ – should be 1. We consider a noisy channel with additive white Gaussian noise (AWGN); i.e., the distribution of the received codes is defined as

$$P(\boldsymbol{y}|\boldsymbol{x}) \propto \prod_{l}^{N} \exp\left(-\frac{(y_l - x_l)^2}{2\sigma^2} \right), \tag{2}$$

where σ^2 denotes the variance of the noise. The posterior probability of the sent code can then be expressed as

$$P(\boldsymbol{x}|\boldsymbol{y}) \propto \left[\prod_{\mu}^{M} \left(1+\prod_{l\in\boldsymbol{\mu}} x_l \right) \right] \left[\prod_{l}^{N} \exp\left(x_l \frac{y_l}{\sigma^2} \right) \right]. \tag{3}$$

Bethe free energy

The Bethe free energy can be interpreted as approximating the Kullback-Leibler (KL) divergence:

$$\mathrm{KL}(Q(\boldsymbol{x})||P(\boldsymbol{x})) \ \equiv \ \sum_{\boldsymbol{x}} Q(\boldsymbol{x}) \ln \frac{Q(\boldsymbol{x})}{P(\boldsymbol{x})} \tag{4}$$

where $P(\boldsymbol{x})$ denotes the posterior probability of codes (from here, we omit '$|\boldsymbol{y}$' for simplicity), and $Q(\boldsymbol{x})$ denotes a parameterized distribution. By improving the parameters of $Q(\boldsymbol{x})$, we minimize the KL divergence. Bethe approximation divides the posterior probability as

$$P(\boldsymbol{x}) \propto \left[\prod_{\mu}^{M} \tilde{P}_{\mu}(\boldsymbol{x}_{\boldsymbol{\mu}}) \right] \left[\prod_{l}^{N} \tilde{P}(x_l)^{1-|\boldsymbol{l}|} \right], \tag{5}$$

where $\boldsymbol{l}$ denotes the parity indices involving the l-th spin, while

$$\tilde{P}(\boldsymbol{x}_{\boldsymbol{\mu}}) \ \propto \ \left[1+\prod_{l\in\boldsymbol{\mu}} x_l \right] \left[\prod_{l}^{N} \tilde{P}(x_l) \right], \tag{6}$$

$$\tilde{P}(x_l) \ \propto \ \exp\left(x_l \frac{y_l}{\sigma^2} \right), \tag{7}$$

are the approximated marginal distributions. $Q(\boldsymbol{x})$ and $Q(\boldsymbol{x_\mu})$ are defined similarly,

$$Q(\boldsymbol{x}) \propto \left[\prod_{\mu}^{M} Q_\mu(\boldsymbol{x_\mu})\right]\left[\prod_{l}^{N} Q_l(x_l)^{1-|\boldsymbol{l}|}\right], \tag{8}$$

$$Q(\boldsymbol{x_\mu}) \propto \left[1+\prod_{l\in\boldsymbol{\mu}} x_l\right]\left[\prod_{l}^{N} Q_l(x_l)\right], \tag{9}$$

but are limited by an additional restriction:

$$\sum_{\boldsymbol{x}_{\boldsymbol{\mu}\setminus l}} Q(\boldsymbol{x_\mu}) = Q(x_l), \tag{10}$$

which demands consistency of the marginal distribution. Also, for $Q(\boldsymbol{x})$, $Q(\boldsymbol{x_\mu})$, and $Q(x_l)$, the conditions of the probability function must hold. Using these approximations, we obtain

$$\begin{aligned}
\mathrm{KL}(Q(\boldsymbol{x})||P(\boldsymbol{x})) &\simeq \sum_{\boldsymbol{x}} Q(\boldsymbol{x}) \ln \frac{\left[\prod_{\mu}^{M} Q_\mu(\boldsymbol{x_\mu})\right]\left[\prod_{l}^{N} Q_l(x_l)^{1-|\boldsymbol{l}|}\right]}{\left[\prod_{\mu}^{M} \tilde{P}_\mu(\boldsymbol{x_\mu})\right]\left[\prod_{l}^{N} \tilde{P}_l(x_l)^{1-|\boldsymbol{l}|}\right]} && (11)\\
&\simeq \sum_{\mu}^{M}\sum_{\boldsymbol{x_\mu}} Q_\mu(\boldsymbol{x_\mu}) \ln \frac{Q_\mu(\boldsymbol{x_\mu})}{\tilde{P}_\mu(\boldsymbol{x_\mu})} + \sum_{l}^{N}(1-|\boldsymbol{l}|)\sum_{x_l} Q_l(x_l)\ln\frac{Q_l(x_l)}{\tilde{P}_l(x_l)} && (12)\\
&= \sum_{\mu}^{M} \mathrm{KL}(Q_\mu(\boldsymbol{x_\mu})||\tilde{P}_\mu(\boldsymbol{x_\mu})) + \sum_{l}^{N}(1-|\boldsymbol{l}|)\,\mathrm{KL}(Q_l(x_l)||\tilde{P}_l(x_l)). && (13)
\end{aligned}$$

This is called the Bethe free energy.

Lagrange multipliers

To minimize the Bethe free energy, we used Lagrange multipliers $\lambda_{\mu l}(x_l)$. From here on, we write $\phi_\mu(\boldsymbol{x_\mu})$, $\psi_l(x_l)$, $b_\mu(\boldsymbol{x_\mu})$, and $q_l(x_l)$ instead of $\tilde{P}_\mu(\boldsymbol{x_\mu}), \tilde{P}_l(x_l), Q_\mu(\boldsymbol{x_\mu})$, and $Q_l(x_l)$, respectively. We also sometimes omit the arguments of these functions for simplicity. The objective function to minimize is now

$$G(\{b_\mu\},\{q_l\},\{\lambda_{\mu l}\}) \equiv \sum_{\mu}^{M}\sum_{\boldsymbol{x_\mu}} b_\mu \ln\frac{b_\mu}{\phi_\mu} + \sum_{l}^{N}(1-|\boldsymbol{l}|)\sum_{x_l} q_l \ln\frac{q_l}{\psi_l} + \sum_{\mu}^{M}\sum_{l\in\boldsymbol{\mu}}\sum_{x_l}\lambda_{\mu l}\left(\sum_{\boldsymbol{x}_{\boldsymbol{\mu}\setminus l}} b_\mu - q_l\right) \tag{14}$$

and we solve the following three equations:

$$0 = \frac{\partial G}{\partial b_\mu(\boldsymbol{x_\mu})} = \ln\frac{b_\mu}{\phi_\mu} + 1 + \sum_{l\in\boldsymbol{\mu}}\lambda_{\mu l}, \tag{15}$$

$$0 = \frac{\partial G}{\partial q_l(x_l)} = (1-|\boldsymbol{l}|)\left(\ln\frac{q_l}{\psi_l}+1\right) - \sum_{\mu\in\boldsymbol{l}}\lambda_{\mu l}, \tag{16}$$

$$0 = \frac{\partial G}{\partial \lambda_{\mu l}(x_l)} = \sum_{\boldsymbol{x}_{\boldsymbol{\mu}\setminus l}} b_\mu - q_l. \tag{17}$$

Using $x_l \in \{+1,-1\}$ and the normalization conditions of distribution functions, we can reduce q_l and $\lambda_{\mu l}$ to linear functions as

$$q_l(x_l) = \frac{1+x_l\tanh h_l}{2}, \tag{18}$$

$$\lambda_{\mu l}(x_l) = -x_l h_{\mu l}. \tag{19}$$

We also use $\hat{h}_{\mu l} \equiv h_{\mu l} + \frac{y_l}{\sigma^2}$.

Belief propagation (BP)

BP first applies the solutions of Eqs. (15) and (16) to G, and then solves Eq. (17), resulting in

$$\hat{h}_{\mu' l} + \operatorname{atanh} \prod_{l' \in \boldsymbol{\mu}' \setminus l} \tanh \hat{h}_{\mu' l'} \quad = \quad \frac{1}{1 - |\boldsymbol{l}|} \left(\frac{y_l}{\sigma^2} - \sum_{\mu'' \in \boldsymbol{l}} \hat{h}_{\mu'' l} \right) \tag{20}$$

This equation holds for any $\{l, \mu' \in \boldsymbol{l}\}$. BP replaces the left side of this equation with the average without μ:

$$\frac{1}{1 - |\boldsymbol{l}|} \sum_{\mu' \in \boldsymbol{l} \setminus \mu} \left(\hat{h}_{\mu' l} + \operatorname{atanh} \prod_{l' \in \boldsymbol{\mu}' \setminus l} \tanh \hat{h}_{\mu' l'} \right) \quad = \quad \frac{1}{1 - |\boldsymbol{l}|} \left(\frac{y_l}{\sigma^2} - \sum_{\mu'' \in \boldsymbol{l}} \hat{h}_{\mu'' l} \right) \tag{21}$$

We then obtain the decoding algorithm

$$\text{loop: } \hat{h}_{\mu l} \leftarrow \frac{y_l}{\sigma^2} + \sum_{\mu' \in \boldsymbol{l} \setminus \mu} \operatorname{atanh} \prod_{l' \in \boldsymbol{\mu}' \setminus l} \tanh \hat{h}_{\mu' l'}, \tag{22}$$

$$\text{result: } h_l \leftarrow \frac{y_l}{\sigma^2} + \sum_{\mu' \in \boldsymbol{l}} \operatorname{atanh} \prod_{l' \in \boldsymbol{\mu}' \setminus l} \tanh \hat{h}_{\mu' l'}. \tag{23}$$

We stop the iteration loop if the estimated sent code, $\hat{x}_l \equiv \operatorname{sign} h_l$, satisfies all parities, or the number of loops reaches an upper limit.

Concave convex procedure (CCCP)

CCCP uses a slightly modified Bethe free energy to guarantee the conversion to a local minimum. At (outer loop) step t, the subsequent parameters, $\{q^{t+1}\}$, are determined as extrema of the following modified free energy,

$$\tilde{G}^t \quad \equiv \quad G + \sum_{l}^{N} |\boldsymbol{l}| \sum_{x_l} q_l \ln \frac{q_l}{q_l^t} \tag{24}$$

$$= \quad G + \sum_{l}^{N} |\boldsymbol{l}| \, \mathrm{KL}(q_l(x_l) || q_l^t(x_l)). \tag{25}$$

This modification can be interpreted as limiting the distance from the current step to the next step. Similar to BP, CCCP solves Eqs. (15) and (16), and then solves Eq. (17). However, G is replaced with $\tilde{G}^t$, resulting in the following inner loop. After the conversion of this inner loop, the outer loop is performed to determine h_l^{t+1}.

$$\text{inner loop: } \hat{h}_{\mu l} \leftarrow \frac{1}{2} \left(\frac{y_l}{\sigma^2} + \sum_{\mu' \in \boldsymbol{l} \setminus \mu} (h_l^t - \hat{h}_{\mu' l}) + h_l^t - \operatorname{atanh} \prod_{l' \in \boldsymbol{\mu} \setminus l} \tanh \hat{h}_{\mu l'} \right), \tag{26}$$

$$\text{outer loop: } h_l^{t+1} \leftarrow \quad \frac{y_l}{\sigma^2} + \sum_{\mu' \in \boldsymbol{l}} (h_l^t - \hat{h}_{\mu' l}). \tag{27}$$

Bowman-Levin (BL)

There is much similarity among BP, CCCP, and BL. The BL strategy is also similar to that of BP, but solves Eqs. (15) and (17) first, resulting in the equations to determine $\{\hat{h}_{\mu l}\}$ using $\{h_l\}$. Unfortunately, these equations need some iteration to be solved:

$$\text{inner loop: } \hat{h}_{\mu l} \leftarrow h_l^t - \text{atanh} \prod_{l' \in \boldsymbol{\mu} \backslash l} \tanh \hat{h}_{\mu l'}. \tag{28}$$

Applying these results to G, we obtain

$$G = \sum_l^N \left\{ m_l \left(-\frac{y_l}{\sigma^2} + \sum_{\mu \in \boldsymbol{l}} \hat{h}_{\mu l} \right) + (1 - |\boldsymbol{l}|) \left(\ln \cosh \frac{y_l}{\sigma^2} + \frac{1+m_l}{2} \ln \frac{1+m_l}{2} + \frac{1-m_l}{2} \ln \frac{1-m_l}{2} \right) \right\}. \tag{29}$$

Then BL solves (16); i.e.,

$$0 \;=\; \frac{\partial G}{\partial m_l} = -\frac{y_l}{\sigma^2} + \sum_{\mu \in \boldsymbol{l}} \hat{h}_{\mu l} + (1 - |\boldsymbol{l}|) h_l, \tag{30}$$

where $m_l \equiv \tanh h_l$. Here, we assume

$$\frac{\partial \hat{h}_{\mu l}}{\partial m_l} = 0, \tag{31}$$

to reduce the calculation cost. We then obtain the following outer loop.

$$\text{outer loop: } h_l^{t+1} \leftarrow \frac{1}{1 - |\boldsymbol{l}|} \left(\frac{y_l}{\sigma^2} - \sum_{\mu' \in \boldsymbol{l}} \hat{h}_{\mu' l} \right). \tag{32}$$

After conversion of the inner loop, the outer loop is performed using the converged variables $\{\hat{h}_{\mu l}\}$.

The outer loop of the BL algorithm, however, cannot be solved through simple repeated substitutions; if we attempt this, the BL algorithm will empirically ascend the Bethe free energy. To avoid ascending the Bethe free energy, we employ a provisional method; i.e., the simple gradient method:

$$\text{outer loop: } h_l^{t+1} \leftarrow h_l^t - k \frac{\partial G}{\partial m_l}, \tag{33}$$

where k denotes a small positive step width. With the gradient method, the BL algorithm successfully converges. The BL performance with the gradient method was slightly worse than that of BP and noticeably worse than that of CCCP.

Various combinatorial methods with BL are considered possible; e.g., BL with the natural gradient method, the modified Bethe free energy (Eq. (24)) with a BL approach, and so on. By applying and comparing these methods, we not only gain a better understanding of how they work, but also can develop potentially better algorithms than BP or CCCP. Extension to problems other than LDPCC should also be possible.

Conclusion

The method we have proposed minimizes the Bethe free energy based on the Bowman-Levin (BL) equation. We have compared our BL algorithm to those of BP and CCCP with respect to the LDPCC decoding problem. The BL algorithm empirically ascends the Bethe free energy, while the BL algorithm combined with the gradient method descends it and converges. Thus, BL is useful not only in allowing us to compare its mechanisms with those of BP and CCCP, but also in helping to enable the development of potentially better algorithms.

Acknowledgements

This work was partially supported by a Grant-in-Aid for Scientific Research on Priority Areas No. 14084206.

References

[1] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, 1988.

[2] A.L. Yuille, Neural Comput., **14**, 1691-1722, 2002.

[3] Y. Kabashima and D. Saad, Europhys. Lett., **44**, 668-674, 1998.

[4] J.S. Yedidia, W.T. Freeman and Y. Weiss, *Understanding Belief Propagation and its Generalization*, http://www.merl.com, MITSUBISHI RESEACH LABORATORIES.

[5] T. Shibuya and K. Sakaniwa, IEICE Trans. on Fundamentals, **E86-A**(10), 2601-2606, 2003.

[6] D. R. Bowman and K. Levin, Phys. Rev. B, **25**, 3438, 1982.