My research

Jack Raymond

University of Rome, La Sapienza

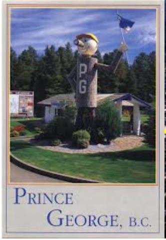
jack.raymond@physics.org

Dwave: Seminar

November 8

Overview

Who am I



Discuss today:

* Variational inference and approximation in graphical models.

Reasons I find Statistical Physics inspiring:

LDPC, kSAT, compressed sensing : where are the hard problems, why does local search fail or slow down, can nucleation help avoid metastability? phase characteristics and transitions.

My research: Edinburgh University; Martin Evans

Self-propelled particles (non-equilibrium physics on a 1d lattice).

My research: Aston University (NCRG); David Saad

* Code Division Multiple Access (analysis of protocols for multi-user access channels).

  Noisy computation (robustness of computation on random directed acyclic graphs).

  1-in-k SAT on random graphs (a boolean constraint satisfaction problem).

My research: Hong Kong UST; Michael Wong

Typical case hard optimization (next nearest neighbor Ising models on random graphs).

Compressed sensing (extracting a sparse signal of N components from $M < N$ measurements).

My research: University of Rome; Federico Ricci-Tersenghi

* Loop-correction algorithms (improvements to tree-like iterative algorithms).
* Linear response for inverse problems (improvements to simple matrix inverse methods).
  Compressed sensing.
  Typical case hard optimization (vertex cover/ resource allocation).

Short summary

I have been to a lot of photogenic places...

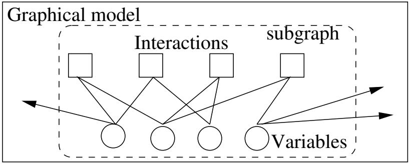
I like optimization problems on random graphs.

Next The MARG problem, graphical model representations.

The probability for N interacting variables x can be described:

$$P(x) = \frac{1}{Z} \prod_a \psi_a(x)$$

A factor graph representation is very general:



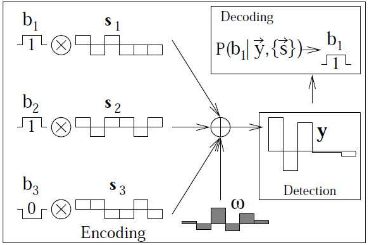
The marginalization and maximization problems:

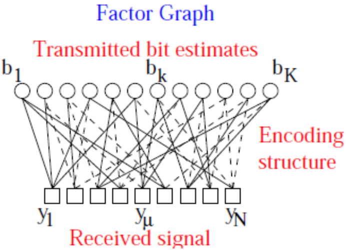
MARG Approximate $P(x_i)$,$P(\{x_i, x_j\})$ and other marginals efficiently.

MAX Approximate $\text{argmax}_x P(x)$.

Inference (marg. / max) is NP-hard in the (interesting regimes) of the problems presented.

e.g. Code Division Multiple Access





Given signal reconstruct best source (MAX), or best marginal estimates (MARG).

$$P(x|y) \propto \underbrace{\prod_{\mu} \exp\left[-\frac{1}{2\sigma^2}(y_\mu - \sum_i A_{\mu i} x_i)\right]}_{factorized\ likelihood} \underbrace{\prod_i [\delta(x_i - 1) + \delta(x_i + 1)]}_{factorized\ prior} \quad (1)$$

Working from the free energy for graphical models

The free energy is

$$F = -\log \sum_x \prod_a \psi_a(x_a)$$

The sum over N states should not be done naively:

1. Transfer-Matrix method (Junction Tree algorithm), or Markov Chain Monte Carlo (MCMC).

If too slow:

2. Advanced or simple mean-field methods (variational methods).

Graphical models, a restricted example

essential exponential family of probabilities; discrete states.

digestible Pairwise interactions; spin variables $x_i = \pm 1$, $\beta = 1$.

$$P(x) = \exp\left[\sum_{ij} J_{ij} x_i x_j + \sum_i H_i x_i + F(J,H)\right] .$$

The free energy ensures normalization

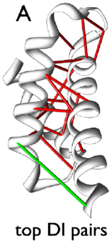
$$F(J,H) = -\log\left[\sum_x \exp\left(\sum_{ij} J_{ij} x_i x_j + \sum_i H_i x_i\right)\right] .$$

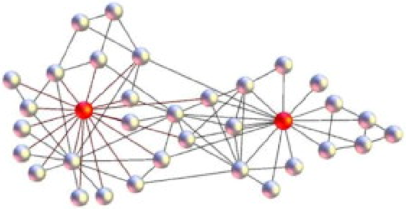
Graphical models, a restricted example

$$P(x) = \exp(\sum_{ij} J_{ij} x_i x_j + \sum_i H_i x_i + F(J, H)) \, . \qquad (2)$$

To do:

1. (Direct) Infer marginal distributions $E[x_i]$ and $E[x_i x_j]$ given couplings J and fields H.
2. (Inverse) Infer couplings and fields from M data samples $\{x^1, \ldots, x^M\}$ or statistics $E[x_i]$, $E[x_i x_j]$.





Short summary

Graphical models can represent a range of disordered problems, revealing features through graph theoretic properties.

The MARG problem is to approximate $P(x_i)$ from an intractable distribution $P(x)$.

Next Solving the MARG problem from the variational free energy or by local iterative schemes.

Relation between the free energy and marginals

The free energy is the moment generating function, the marginals are simple functions of the moments

$$F(H, J) = -\log\left[\sum_x \exp\left(\sum_{ij} J_{ij}x_i x_j + \sum_i H_i x_i\right)\right] \ . \tag{3}$$

Magnetizations (1st order cumulants):

$$E[x_i] = -\frac{\partial}{\partial H_i} F \ ; \tag{4}$$

Connected Correlations (2nd order cumulants, responses):

$$E[x_i x_j] - E[x_i]E[x_j] = -\frac{\partial^2}{\partial H_i \partial H_j} F \ ; \ . \tag{5}$$

VP From a tractable family of probability models $Q(x)$, select a model minimizing the Kullback-Leibler Divergence to $P(x)$.

$$KL(Q||P) = \sum Q(x) \log \left(\frac{Q(x)}{P(x)} \right) \geq 0 \quad (6)$$

The variational free energy is an upper bound to $F(H, J)$

$$F_{H,J}(Q) = E_Q[\log Q(x)] - \sum_{ij} J_{ij} E_Q[x_i x_j] - \sum_i H_i E_Q[x_i] \quad (7)$$

H, J are the quenched parameters, $\{Q(x) : \sum Q(x) = 1\}$ is variational and easily marginalized.

Local iteration schemes

LI Define local probability approximations (e.g. magnetization $E[x_i]$), pass information locally to refine the estimates under consistency constraints.

NMF Mean-field argument (exact for weak correlations): all variables have a magnetization, observing neighbor states and local interactions one can reestimate:

$$M_i = \tanh(H_i + \sum_j J_{ij} M_j) \tag{8}$$

BP Belief propagation (exact for locally tree like graphs): variables have a set of conditional magnetizations one for each outward neighbor, which is the magnetization excluding the effect of that neighbor ($J_{ij} = 0$). These conditional magnetizations are updated

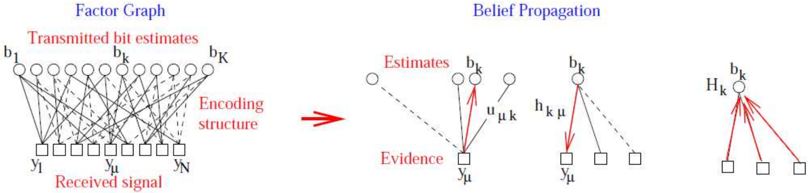
$$m_{i \to j} = \tanh\left\{H_i + \sum_{k \backslash j} \mathrm{atanh}[\tanh(J_{ik}) m_{k \to i}]\right\} \tag{9}$$

Local iteration versus variational schemes

- Mean field iterative scheme = local minima of Mean field free energy.
- Belief propagation iterative scheme (Gallagher, Bethe-Peiels, . . .) = local minima of Bethe free energy (Yedidia 2005).
- Expectation propagation schemes (Minka) = Weighted sum of Gaussian and Marginal free energies (Wainwright and Jordan, 2008).

Most common iterative schemes represent specifiic ways to minimze non-convex free energies.

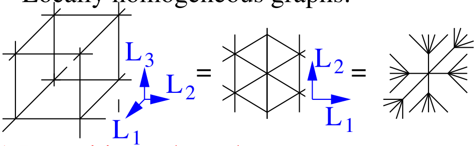
Belief propagation (Bethe free energy): it does work



The MARG problem is solved in practice by BP for $N \sim O(100), O(1000)$, in CDMA, LDPC (channel coding), compressed sensing, community detection models. Large, disordered "locally tree like" or "fully connected" networks.

Belief propagation (Bethe free energy): it doesn't work

Locally homogeneous graphs:



1) Insensitive to boundary
2) Insensitive to dimensionality

For locally identical problems coupling J, field H, connectivity k:

MF

$$M = \tanh\{H + kJM\} \tag{10}$$

BP

$$m_{\rightarrow} = \tanh\{H + (k-1)\text{atanh}[\tanh(J)m_{\rightarrow}]\} \tag{11}$$

Corrections for short and long loops, sensitivity to dimensionality and boundary.

*__Short loops__: Cluster Variational Method (Kikuchi,51)
pay an exponential price in cluster width, loops can remain but on larger regions.
Long loops: Loop calculus (Chertkov and Chernyak,05)
many interesting graphs have a factorial number of loops, how to select a subset?
something else...:
Moment matching: Minka and Qi (2004), Opper and Winther (2005).
*Linear Response: Kappen et al. (1998), Welling and Teh (2004), Montanari and Rizzo (2005), Yasuda et al (2013), Opper and Winther (2001)
Variable Clamping: Eaton et al. (2008), Mooij et al (2005)

How to calculate linear response

Using an unconstrained representation of the free energy e.g.

$$b_i(x_i) = \frac{1 + (M_i^* + \delta M_i)x_i}{2} \quad (12)$$

Free energy about the fixed point (by Taylor expansion)

$$F_{model,H,J}(M) = F_{model}(M^*) + 0 + \frac{1}{2}\delta M^T \left.\frac{\partial^2 F_{model}}{\partial M \partial M^T}\right|_{M^*} \delta M \quad (13)$$

A small change in H, will lead to a correspondingly small change in the magnetization

$$F_{model,H+\delta H,J}(M^* + \delta M) = F_{model}(M^*) + \delta H^T \left.\frac{\partial F_{model}}{\partial H}\right|_{M^*} + \delta H^T \left.\frac{\partial^2 F_{model}}{\partial H \partial M^T}\right|_{M^*} \delta M + \frac{1}{2}\delta M^T \left.\frac{\partial^2 F_{model}}{\partial M \partial M^T}\right|_{M^*} \delta M \quad (14)$$

Need to minimize in the variational parameter (the perturbation δM)

How to calculate linear response (continued..)

The new saddle-points

$$
\begin{aligned}
0 &= \delta H^T \left.\frac{\partial^2 F_{model}}{\partial H \partial M^T}\right|_{M^*} + \delta M^T \left.\frac{\partial^2 F_{model}}{\partial M \partial M^T}\right|_{M^*} \\
\left[\frac{\partial M}{\partial H}\right]^{-1} &= \left.\frac{\partial^2 F_{model}}{\partial M \partial M^T}\right|_{M^*}
\end{aligned}
$$

Inverse correlation matrix is a simple function of the Hessian. Infact

$$
\left[\frac{\partial M}{\partial H}\right]^{-1} = \frac{\partial^2 F(M^*, C^*)}{\partial M \partial M^T} - \frac{\partial^2 F(M^*, C^*)}{\partial M \partial C^T} \left[\frac{\partial^2 F(M^*, C^*)}{\partial C \partial C^T}\right]^{-1} \frac{\partial^2 F(M^*, C^*)}{\partial C \partial M^T}
$$

In the case of Bethe and NMF, the estimates obtained depend on the loops. Note we can also use the expression to estimate couplings:

$$
\left(\left[C^{data}\right]^{-1}\right)_{i,j} = \frac{\partial^2 F(M^{data}, C^{data})}{\partial M_i \partial M_j} - \ldots = -J_{i,j} + \left[\frac{\partial^2 S(M^*, C^*)}{\partial M_i \partial M_j} - \ldots\right]
$$

Short summary

Choose a variational form Q and minimize.

Possibly improve the estimation of correlations with linear response.

Belief Propagation values for M^*, C^* are ignorant of loops; linear response dM/dH can see loops, but is also inexact.

The linear response expression can be inverted to get J as simple functions of M and C (from data).

Next Recent work with Federico Ricci-Tersenghi; making variational approximations self-consistent with linear response constraints

A simple idea

1 Clamp out some small subset of troublesome variables.

2 Use cluster methods to coarse grain the problem (deal with short loops), use expansion methods for important long loops.

3 Calculate message perturbations (linear response) to yield extra information (connected correlations).

3(NEW) Require consistency between the message information and the linear response information.

J. Raymond and F. Ricci-Tersenghi, "Mean-field method with correlations determined by linear response," *Phys. Rev. E*, vol. 87, p. 052111, 2013.

Why should the constraints help?

A variational approach based on trial probability distribution Q

$$F(Q) = \sum [Q(x) \log Q(x)] - \sum_{i,j} J_{i,j} \sum [Q(x) x_i x_j] - \sum_i H_i \sum [Q(x) x_i] \quad (15)$$

1 Constraints on the physical quantity make the method consistent (in its predictions).

2 In a good variational approximations $Q^* \sim P + \epsilon \delta P$: 1st order derivative errors $O(\epsilon)$, n^{th} order derivative errors $O(\epsilon^n)$. (e.g. Opper, 2003).

3 First derivative conditions introduce local consistency. The Hessian related constraints introduces a global consistency (Welling and Teh, 2004).

NEW Fix parameters by second order constraints, reduce errors and include extra graph wide consistency.

Two ways to get the self response

At the minima of the variational free energy

$$1 - \left[\frac{\partial}{\partial H_i} F_{model} \right]^2 = 1 - (M_i)^2 \ .$$

If we were using the exact free energy, this would equal the response

$$\frac{\partial M_i}{\partial H_i} = \frac{\partial^2}{\partial H_i \partial H_i} F_{model} \ .$$

but this is violated in NMF, Bethe and other approximations. we introduce a constraint.

Variational approach implemented

Reasonable trial distribution (e.g. NMF, independent variables)

$$Q(x) = \prod_{i=1}^{N} Q_i(x_i) = \prod_{i=1}^{N} \frac{1 + M_i x_i}{2} \ . \tag{16}$$

Kullback-Leibler divergence (relative entropy) is:

$$D(Q||P) = -\sum_{ij} J_{ij} E_Q[x_i x_j] - \sum_i H_i E_Q[x_i] + E_Q[\log(Q)] - F(J,H) \geq 0 \ . \tag{17}$$

E_Q is the expectation with respect to the trial distribution. The model free energy is

$$F_{model}(M) = E_{model}(M) - S_{model}(M) \geq F \ . \tag{18}$$

STD Minimize w.r.t M. $\partial_M F = 0$

NEW Minimize subject to $(1 - M_i^2) = [\partial_H^2 F(M)]_{i,i}$.

Diagonal and Off-diagonal consistency

Beliefs can always be represented by magnetization (M) and correlations (C) in a consistent way:

$$b_i(x_i) = \frac{1 + M_i x_i}{2} \qquad b_{ij}(x_i) = b_i(x_i) b_j(x_j) + \frac{C_{ij} x_i x_j}{4}$$

On-diagonal constraint: Lagrange multiplier λ_i

$$\left[1 - \left(\frac{\partial F_{model}}{\partial H_i} \right)^2 = \right] \; 1 - (M_i)^2 \left[= \frac{\partial^2 F_{model}}{\partial H_i^2} \right]$$

Off-diagonal constraint: Lagrange multiplier $\lambda_{i,j}$

$$\left[\frac{\partial F_{model}}{\partial J_{ij}} - \frac{\partial F_{model}}{\partial H_i} \frac{\partial F_{model}}{\partial H_j} = \right] \; C_{ij} \; \left[= \frac{\partial^2 F_{model}}{\partial H_i \partial H_j} \right]$$

NB: derivatives evaluated at M^*, C^*

To minimize the *constrained* variational free energy

$$
\begin{aligned}
\partial_M F(M, C, \lambda) &= 0 \qquad \text{saddle-point} \\
\partial_C F(M, C, \lambda) &= 0 \qquad \text{saddle-point} \\
\partial^2_{H_i, H_i} F(b, \lambda) &= 1 - M_i^2 \qquad \text{on-diagonal consistency} \\
\partial^2_{H_i, H_j} F(b, \lambda) &= C_{i,j} \qquad \text{off-diagonal consistency}
\end{aligned}
$$

Normally we fix M, C by the first two conditions, with $\lambda = 0$. Now we introduce new variables λ and new constraints, and solve jointly.

Standard

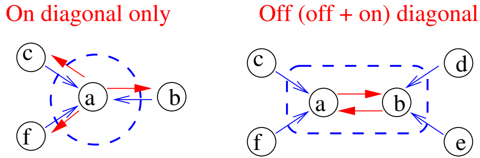
$$m_{i\to j} = \tanh\left[H_i + \sum_{k=neigh.\setminus j} \operatorname{atanh}\left(\tanh(J_{ki})m_{k\to i}\right)\right]$$

vs New (new effective field, effective coupling)

$$m_{i\to j} = \tanh\Bigg\{\left[H_i + \lambda_i M_i + \sum_{k=neigh} \lambda_{ik} M_k\right] + \sum_{k=neigh.\setminus j} \operatorname{atanh}\left(\tanh(J_{ki} - \lambda_{ki})m_{k\to i}\right)\Bigg\}$$

λ are Lagrange multipliers for the constraints, M are magnetizations to be fixed self-consistently.
$O(N)$ algorithm on sparse graph.
To calculate all linear response we simply perturb these messages $O(N^2)$ algorithm.

Schematically



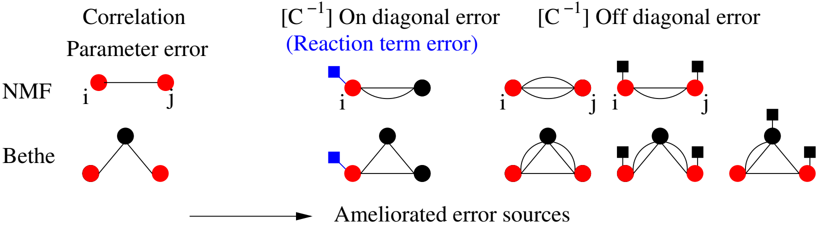
With only on-diagonal constraints same as I-SUSP algorithm (Yasuda et al., 2013).

Take simultaneously messages $\{u_{\rightarrow cav.}\}$ and responses $\{u_{\rightarrow cav.,z}\}$ incident on the edge/vertex region.

Calculate consistently M,C and λ on the region, generate new internal messages.

One choice among many available from variational formulation.

Weak coupling/ high temperature result



Magnetization and Connected Correlation errors by expansion about independent variables.

- NMF errors: $O(J^2, \beta^3)$.
- λ-NMF errors: mag. $O(J^3, \beta^4)$ and c.c. $O(J^2, \beta^3)$.
- Bethe errors: $O(J^3, \beta^4)$.
- λ-Bethe errors: $O(J^4, \beta^5)$.

Opper Winther (2001): (Ising spins, equivalent to adaptive-TAP)

Recovery and extension of adaptive-TAP framework (Opper and Winther, 2001)

$$M_i = \tanh\left(H_i + \sum_{j(\neq i)} J_{ij} M_j + \lambda_i M_i\right) , \qquad (19)$$

$$(1 - M_i^2) = \left[\left(Diag[\frac{1}{1 - M_i^2} - \lambda_i] - J\right)^{-1}\right]_{i,i} . \qquad (20)$$

λ_i is the Onsager reaction term, we needn't know the distribution of J in order for its calculation.
Exact in (advanced) "mean-field" models (includes Sherrington Kirkpatrick model, Hopfield model, fully connected disordered models).

Short summary

We shift the point of evaluation for the free energy from the minima, to a new minima consistent with linear response.

We gain an order of magnitude in estimate quality, adding constraints transforms a bad standard approximation (NMF) into an optimal choice (TAP) for a broad range of models.

We can also formulate "message passing like" algorithms.

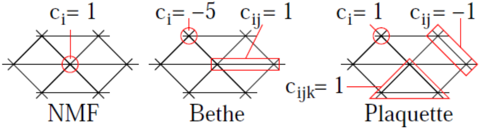
Next Brief introduction of a more advance variational method, some applications.

The cluster variational (region based) approximations

Empirically, the entropy on large regions $\{\beta\}$ can be unde rstood in terms of contributions from smaller constitutent parts The unexplained part is $\tilde{S}$ and typically decays exponentially with cluster size.

$$S = \sum_\alpha \tilde{S}_\alpha + \sum_\beta \tilde{S}_\beta = \sum_\alpha \tilde{S}_\alpha + correlation/loop \ \ correction.$$

We truncate the Mobius transform of the entropy (An,88).
e.g. $\{\alpha\} = \{i\}$ NMF, $\alpha = \{(i,j)\} \cup \{i\}$ Bethe.
The model parameters of the approximation are beliefs b (marginal probabilities)



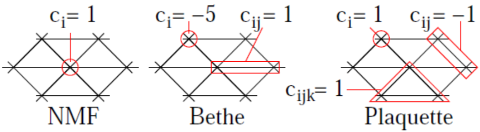
$$S_{model} = \sum_\alpha c_\alpha S_\alpha$$

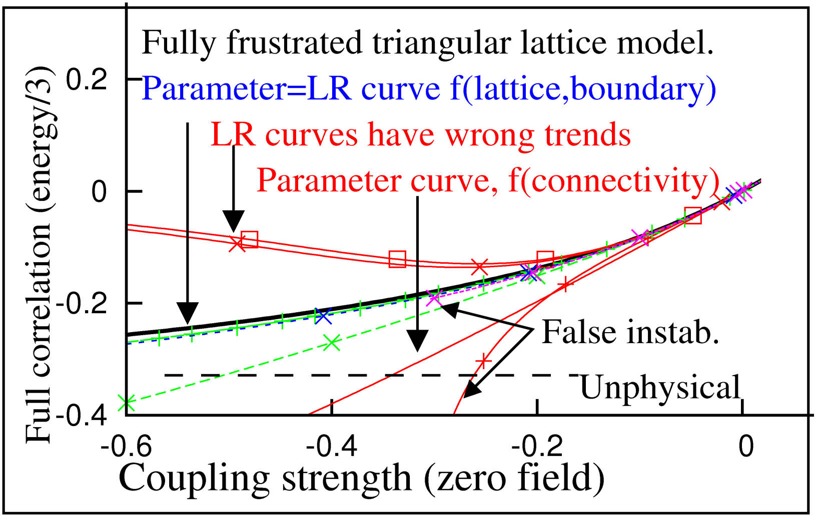
The cluster variational (region based) approximations

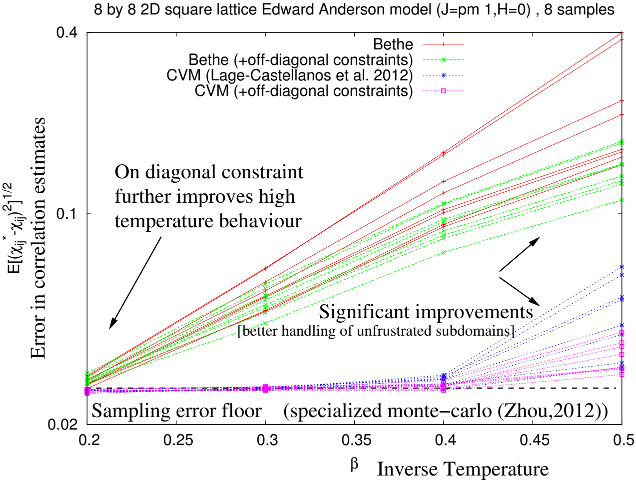
$$\tilde{S}_i(b_i) = -\sum_{x_i} b_i(x_i) \log b_i(x_i)$$

$$\tilde{S}_{ij}(b_{ij}, b_i, b_j) = -\sum_{x_i, x_j} b_{ij}(x_i, x_j) \log \frac{b_{ij}(x_i, x_j)}{b_i(x_i) b_j(x_j)}$$

$$\tilde{S}_{ijk}(\{b.\}) = -\sum b_{ijk}(x_i, x_j, x_k) \log \frac{b_{ijk}(x_i, x_j, x_k) b_i(x_i) b_j(x_j) b_k(x_k)}{b_{ij}(x_i, x_j) b_{jk}(x_i, x_j) b_{ik}(x_i, x_j)}$$



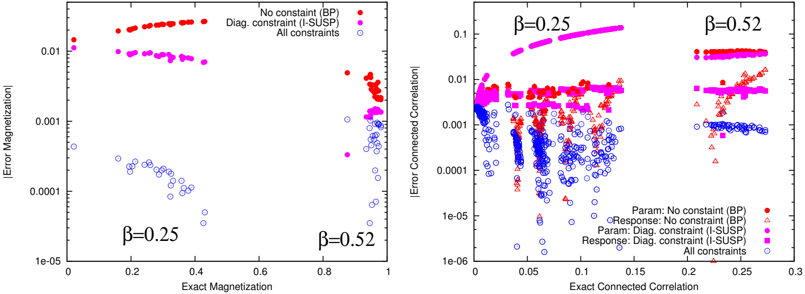

Fully frustrated triangular lattice model.
Parameter=LR curve f(lattice,boundary)
LR curves have wrong trends
Parameter curve, f(connectivity)
False instab.
Unphysical
Full correlation (energy/3)
0.2
0
-0.2
-0.4
-0.6
-0.4
-0.2
0
Coupling strength (zero field)



Disordered and frustrated model, complicated phase space.

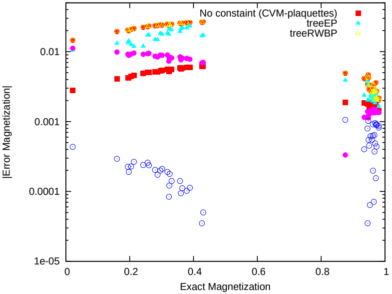
Disordered models, e.g. Random field Ising model

2D square lattice random field Ising model: J=1, H iid Gaussian variables N(0.2,0.5), L=5



- For models dominated by a single pure state (weakly correlated) convergence is robust. e.g. small and large β (if unique ground state).
- Adding constraints boosts performance significantly

and other algorithms?

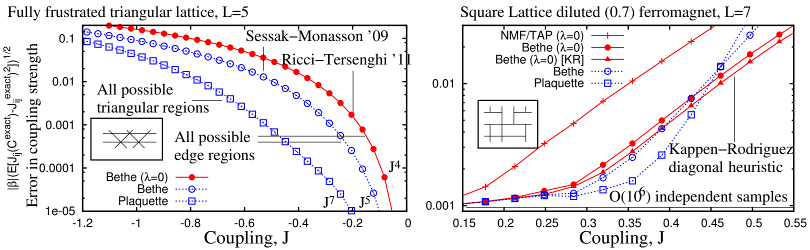


libDAI C++ library of Mooij (standard implementations)

Tree EP (Qi and Minka 2005): A structured moment matching algorithm.

Gen. BP (Yedidia et al.2005, Heskes et al. 2003) : A Kikuchi free energy (plaquette regions).

TRW BP (Wainwright et al., Wiegerinck et al. 2003): A convexified form of the Bethe approximation.



Form:

$$J_{i,j} = -[C^{-1}_{data}]_{i,j}$$

$$J_{i,j} = -[C^{-1}_{data}]_{i,j} + \Lambda^{B}_{i,j}(M_{data}, C^{*}) \; ; \; -[C^{-1}_{data}]_{i,j} + \Lambda^{B}_{i,j}(M_{data}, C_{data})$$

Similar expressions for magnetization.

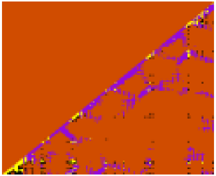
- Provably improved scaling for weak-coupling limit.
- From realistic data: Improvement in an intermediate range of temperatures
- Works well away from phase transitions.

Inverse examples: Protein

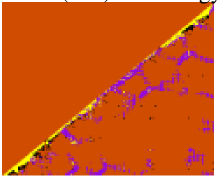
Residue contact maps for protein (length 161 residues [21 states])

Comparison of largest inferred couplings and true structure

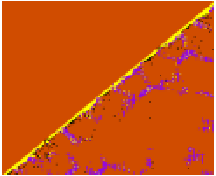
Mutual information



Bethe (like) free energy



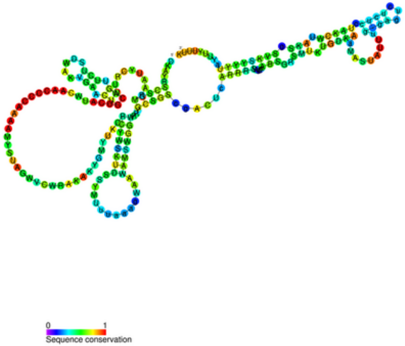
NMF free energy



Blue is true contact structure (well understood special case), [Morcosa et al PNAS 2011]
Yellow indicate true positives (on largest inferred couplings), grey false positives

The sequence of residues (each one of 21 states) in a protein of length O(100) are correlated through evolutionary function.
If residues fold together they coevolve (show strong correlation), we can use a pairwise model to infer functional relationships (i.e. spatial proximity in folding).

Inverse examples: RNA families (telemorase)



State of the art for proteins and RNA (Morcos 2009), simple inverse; we can add Λ^V to determine tertiary structure?

$$J_{i,j} = -[C_{data}^{-1}]_{i,j} + \Lambda^V(M, C)$$

Thanks for attention

- A consistent variational approximation, using information implicit to the approximation.
- **weak-coupling limit validation**.
- Significant results: **Montanari-Rizzo approximation for loop corrections**; **adaptive-TAP**; **Sessak-Monasson inverse Ising expression**; and extensions.
- Variational framework: algorithmic flexibility and expansion methods available.
- With distributed message passing $O(N^2)$ frameworks (and proven I-SUSP, Yasuda et al.)

J. Raymond and F. Ricci-Tersenghi, "Mean-field method with correlations determined by linear response," *Phys. Rev. E*, vol. 87, p. 052111, 2013.

http://chimera.roma1.infn.it/JACK jack.raymond@physics.org

References

J. Raymond and F. Ricci-Tersenghi, "Mean-field method with correlations determined by linear response," *Phys. Rev. E*, vol. 87, p. 052111, 2013.

G. An, "A note on the cluster variation method," *Journal of Statistical Physics*, vol. 52, pp. 727–734, 1988.

Alessandro Pelizzola.
Cluster variation method in statistical physics and probabilistic graphical models.
J. Phys. A, 38(33):R309, 2005.

J.S. Yedidia, W.T. Freeman, and Y. Weiss.
Constructing free-energy approximations and generalized belief propagation algorithms.
Information Theory, IEEE Transactions on, 51(7):2282–2312, 2005.

M. Welling and Y. Teh, "Linear response algorithms for approximate inference in graphical models," *Neural Comput.*, vol. 16, pp. 197–221, 2004.

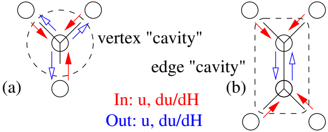
M. Yasuda and K. Tanaka, “Susceptibility propagation by using diagonal consistency,” *Phys. Rev. E*, vol. 87, p. 012134, Jan 2013.

M. Opper and O. Winther, “Adaptive and self-averaging thouless-anderson-palmer mean-field theory for probabilistic modeling,” *Phys. Rev. E*, vol. 64, p. 056131, 2001.

A. Lage-Castellanos, R. Mulet, F. Ricci-Tersenghi, and T. Rizzo.
Replica cluster variational method: the replica symmetric solution for the 2d random bond ising model.
Jour. of Physics A, 46(13):135001, 2013.

A. Montanari and T. Rizzo.
How to compute loop corrections to the Bethe approximation.
J. Stat. Mech., page P10011, 2005.

V. Sessak and R. Monasson.
Small-correlation expansions for the inverse ising problem.
J. Phys. A, 42(5):055001, 2009.



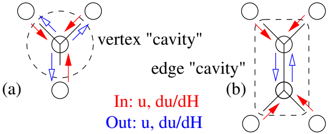
A message passing scheme uses $u_{\rightarrow}$ and $u_{\rightarrow,z} = \partial u / \partial H_z$ akin to Belief Propagation and Susceptibility Propagation.
Knowing $\{M, \lambda\}$ we have BP/SP subject to modified fields and couplings

$$\tilde{H}_i = H_i + \lambda_i M_i + \sum_j \lambda_{ij} M_j \; ; \qquad \tilde{J}_{ij} = J_{ij} - \lambda_{ij} \; .$$

Scheme (a) for on-diagonal constaints (Yasuda et al. I-SUSP)

$$M_i = \tanh(H_i + \lambda_i M_i + \sum_j u^t_{j\rightarrow i})$$

$$\lambda_i = -\frac{1}{1 - M_i^2} \sum_j u^t_{j\rightarrow i,i}$$



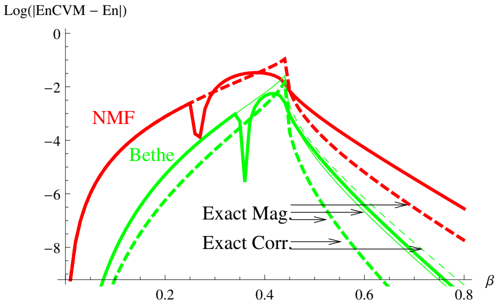
Scheme (b) for on and off-diagonal constraints ($i \leftrightarrow j$):

$$M_i = \tanh\left(H_i + \lambda_i M_i + \lambda_{ij} M_j + \sum_{k\setminus j} u^t_{k\to i} + \hat{u}_{j\to i}\right) .$$

$$\lambda_i = -\left[2\lambda_{ij} + \hat{u}'_{j\to i}\lambda_j\right]\hat{u}'_{j\to i} - \frac{1}{(1-M_i^2)}\left[\sum_{k\setminus j} u_{k\to i,i} + \hat{u}'_{j\to i}\sum_{k\setminus i} u_{k\to j,i}\right] .$$

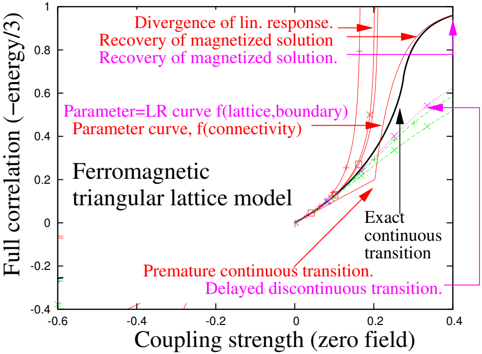
$$\lambda_{ij} = -\left[\lambda_i + \frac{(1-M_j^2)}{(1-M_i^2)}\lambda_j + \hat{u}'_{i\to j}\lambda_{ij}\right]\hat{u}'_{i\to j} - \frac{1}{(1-M_i^2)}\left[\sum_{k\setminus j} u_{k\to j,i} + \hat{u}'_{i\to j}\sum_{k\setminus j} u_{k\to i,i}\right] .$$

$$\hat{u}_{j\to i} = \operatorname{atanh}\left[\tanh(\tilde{J}_{ij})\tanh(H_j + \lambda_j M_j + \lambda_{ij} M_i + \sum_{k\setminus i} u^t_{k\to j})\right] \; ; \qquad \hat{u}'_{j\to i} = \frac{\partial t_{j\to i}}{\partial H_j} .$$



Comparison of entropy approximation error when entering exact or partially exact marginal information (relative to standard maximum entropy [solid thick line]). A residual error remains because of the entropy truncation.

Asymptotic phenomena continued..



Interesting phenomena (i.e. outstanding challenges)

- Absence of solutions in some models at low temperature (e.g. 2D ferromagnetic lattices)
- Absence of continuous transitions where expected
- Absence of message-passing scheme meeting "message passing paradigm"
- Non-convexity (concavity/convexity) of free energy (the feasibility of finding solutions is in general unclear)

The main problem with the algorithm remains convergence and complexity... being addressed in libdai "double loop" formulation.

Abstract

Variational free energies when minimized can make inconsistent predictions of physical quantities, as defined by first and second derivative identities. Minimizing subject to consistency of these derivatives can be shown to improve approximation. The constraints allow a connection with a number of disparate results: From a constrained naive mean-field free energy we recover a type of expectation-propagation algorithm. From a constrained Bethe approximation, applied to the inverse problem of coupling estimation, we recover and extend the Sessak-Monasson expression.

The End