Statistical Inference and Loop Calculus in Physics, Computer and Information Sciences or Belief Propagation and Beyond

Michael Chertkov

Center for Nonlinear Studies & Theory Division, LANL

June 26, 2008 Landau-100, Chernogolovka

Outline

Example (1): Statistical Physics

Ising model $\sigma_i = \pm 1$

$$\mathcal{P}(\vec{\sigma}) = Z^{-1} \exp\left(\sum_{(i,j)} J_{ij}\sigma_i\sigma_j\right)$$

J_{ij} defines the graph (lattice)

- Ferromagnetic ($J_{ij} < 0$), Anti-ferromagnetic ($J_{ij} > 0$) or Frustrated/Glassy
- Magnetization (order parameter) and Ground State
- Thermodynamic Limit, $N \to \infty$
- Phase Transitions

Example (2): Information Theory, Machine Learning, etc

Probabilistic Reconstruction (Statistical Inference)

$\vec{\sigma}_{\text{orig}}$	$\Rightarrow$	$\vec{x}$	$\Rightarrow$	$\vec{\sigma}$
original *data* $\vec{\sigma}_{\text{orig}} \in \mathcal{C}$ codeword	noisy channel $\mathcal{P}(\vec{x}\|\vec{\sigma})$	corrupted data: log-likelihood magnetic field	statistical inference	possible preimage $\vec{\sigma} \in \mathcal{C}$

Maximum Likelihood [ground state]

Marginalization

$$\text{ML}(\vec{x}) = \arg\max_{\vec{\sigma}} \mathcal{P}(\vec{x}|\vec{\sigma})$$

$$\sigma_i^*(\vec{x}) = \arg\max_{\sigma_i} \sum_{\vec{\sigma}\setminus\sigma_i} \mathcal{P}(\vec{x}|\vec{\sigma})$$

e.g., decoding in forward error correction

Example (2): Information Theory, Machine Learning, etc

Probabilistic Reconstruction (Statistical Inference)

$\vec{\sigma}_{\text{orig}}$	$\Rightarrow$	$\vec{x}$	$\Rightarrow$	$\vec{\sigma}$
original *data* $\vec{\sigma}_{\text{orig}} \in \mathcal{C}$ codeword	noisy channel $\mathcal{P}(\vec{x}\|\vec{\sigma})$	corrupted data: log-likelihood magnetic field	statistical inference	possible preimage $\vec{\sigma} \in \mathcal{C}$

Maximum Likelihood [ground state]

Marginalization

$$\text{ML}(\vec{x}) = \arg\max_{\vec{\sigma}} \mathcal{P}(\vec{x}|\vec{\sigma})$$

$$\sigma_i^*(\vec{x}) = \arg\max_{\sigma_i} \sum_{\vec{\sigma}\setminus\sigma_i} \mathcal{P}(\vec{x}|\vec{\sigma})$$

e.g., decoding in forward error correction

Example (2): Information Theory, Machine Learning, etc

Probabilistic Reconstruction (Statistical Inference)

$\vec{\sigma}_{\rm orig}$	$\Rightarrow$	$\vec{x}$	$\Rightarrow$	$\vec{\sigma}$
original *data* $\vec{\sigma}_{\rm orig} \in \mathcal{C}$ codeword	noisy channel $\mathcal{P}(\vec{x}\vert\vec{\sigma})$	corrupted data: log-likelihood magnetic field	statistical inference	possible preimage $\vec{\sigma} \in \mathcal{C}$

Maximum Likelihood [ground state]

Marginalization

$$\mathrm{ML}(\vec{x}) = \arg\max_{\vec{\sigma}} \mathcal{P}(\vec{x}|\vec{\sigma})$$

$$\sigma_i^*(\vec{x}) = \arg\max_{\sigma_i} \sum_{\vec{\sigma}\setminus\sigma_i} \mathcal{P}(\vec{x}|\vec{\sigma})$$

e.g., decoding in forward error correction

Example (2): Information Theory, Machine Learning, etc

Probabilistic Reconstruction (Statistical Inference)

$\vec{\sigma}_{\text{orig}}$	$\Rightarrow$	$\vec{x}$	$\Rightarrow$	$\vec{\sigma}$
original *data* $\vec{\sigma}_{\text{orig}} \in \mathcal{C}$ codeword	noisy channel $\mathcal{P}(\vec{x}\vert\vec{\sigma})$	corrupted data: log-likelihood magnetic field	statistical inference	possible preimage $\vec{\sigma} \in \mathcal{C}$

Maximum Likelihood [ground state]

Marginalization

$$\text{ML}(\vec{x}) = \arg\max_{\vec{\sigma}} \mathcal{P}(\vec{x}|\vec{\sigma})$$

$$\sigma_i^*(\vec{x}) = \arg\max_{\sigma_i} \sum_{\vec{\sigma}\setminus\sigma_i} \mathcal{P}(\vec{x}|\vec{\sigma})$$

e.g., decoding in forward error correction

Example (3): Combinatorial Optimization, K-SAT

$$F(\vec{x}) = \begin{array}{c} (x_1 \vee x_2 \vee \bar{x}_3) \wedge \\ (x_5 \vee \bar{x}_1 \vee \bar{x}_4) \wedge \\ (x_2 \vee x_7 \vee x_3) \wedge \\ (\bar{x}_7 \vee x_5 \vee \bar{x}_5) \wedge \\ \cdots \end{array}$$

$1, 2, \cdots, N$ – *variables*

$F(\vec{x})$ is a conjunction of M clauses

$x_i = 0(\text{bad}), 1(\text{good})$

$\bar{x}_i$ is negation of x_i

$\vee = \text{OR} \quad \wedge = \text{AND}$

$\vec{x}$ is a "valid assignment" if $F(\vec{x}) = 1$

Probabilistic interpretation

$$P(\vec{x}) = Z^{-1} F(\vec{x}), \quad Z \equiv \sum_{\vec{x}} F(\vec{x})$$

- Finding a Valid Assignment, Counting Number of Assignments
- Graphical Representation, Sparseness
- Random, non-Random formulas
- SAT/UNSAT transition wrt $\alpha = M/N$, $M, N \to \infty$
- Focus on Algorithms for given finite instance

Complexity & Algorithms

- How many operations are required to evaluate a graphical model of size N?
- What is the exact algorithm with the least number of operations?
- If one is ready to trade optimality for efficiency, what is the best (or just good) approximate algorithm he/she can find for a given (small) number of operations?
- Given an approximate algorithm, how to decide if the algorithm is good or bad? What is the measure of success?
- How one can systematically improve an approximate algorithm?

- Linear (or Algebraic) in N is EASY, Exponential is DIFFICULT

Easy & Difficult Boolean Problems

EASY

- Any graphical problems on a tree (Bethe-Pieirls, dynamical programming, belief propagation, and other names)
- Ground State of a Rand. Field Ferrom. Ising model on any graph
- Partition function of a planar Ising model
- Finding if 2-SAT is satisfiable
- Decoding over Binary Erasure Channel = XOR-SAT
- Some network flow problems (max-flow, min-cut, shortest path, etc)
- Minimal Perfect Matching Problem
- Some special cases of Integer Programming (TUM)

Typical graphical problem, with loops and factor functions of a general position, is DIFFICULT

Easy & Difficult Boolean Problems

EASY

- Any graphical problems on a tree (Bethe-Pieirls, dynamical programming, belief propagation, and other names)
- Ground State of a Rand. Field Ferrom. Ising model on any graph
- Partition function of a planar Ising model
- Finding if 2-SAT is satisfiable
- Decoding over Binary Erasure Channel = XOR-SAT
- Some network flow problems (max-flow, min-cut, shortest path, etc)
- Minimal Perfect Matching Problem
- Some special cases of Integer Programming (TUM)

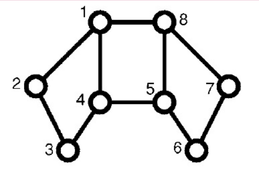
Typical graphical problem, with loops and factor functions of a general position, is DIFFICULT

Boolean Graphical Models = The Language

Forney style - variables on the edges

$$\mathcal{P}(\vec{\sigma}) = Z^{-1}\prod_a f_a(\vec{\sigma}_a)$$

$$\underbrace{Z = \sum_{\vec{\sigma}}\prod_a f_a(\vec{\sigma}_a)}_{\text{partition function}}$$



$$f_a \geq 0$$

$$\sigma_{ab} = \sigma_{ba} = \pm 1$$

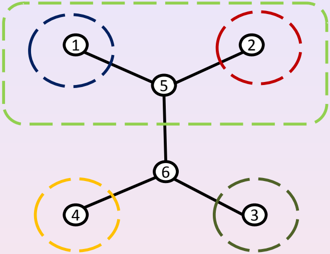
$$\vec{\sigma}_1 = (\sigma_{12}, \sigma_{14}, \sigma_{18})$$

$$\vec{\sigma}_2 = (\sigma_{12}, \sigma_{23})$$

Objects of Interest

- Most Probable Configuration = Maximum Likelihood = Ground State: $\arg\max \mathcal{P}(\vec{\sigma})$
- Marginal Probability: e.g. $\mathcal{P}(\sigma_{ab}) \equiv \sum_{\vec{\sigma}\setminus\sigma_{ab}} \mathcal{P}(\vec{\sigma})$
- Partition Function: Z

BP is Exact on a Tree   Bethe '35, Pieirls '36



$$Z_{15}(\sigma_{15}) = f_1(\sigma_{15}), \quad Z_{25}(\sigma_{25}) = f_2(\sigma_{25}),$$
$$Z_{36}(\sigma_{36}) = f_3(\sigma_{36}), \quad Z_{46}(\sigma_{46}) = f_4(\sigma_{46})$$
$$Z_{56}(\sigma_{56}) = \sum_{\vec{\sigma}_5 \setminus \sigma_{56}} f_5(\vec{\sigma}_5) Z_{15}(\sigma_{15}) Z_{25}(\sigma_{25})$$
$$Z = \sum_{\vec{\sigma}_6} f_6(\vec{\sigma}_6) Z_{36}(\sigma_{36}) Z_{46}(\sigma_{46}) Z_{56}(\sigma_{56})$$

$$Z_{ba}(\sigma_{ab}) = \sum_{\vec{\sigma}_a \setminus \sigma_{ab}} f_a(\vec{\sigma}_a) Z_{ac}(\sigma_{ac}) Z_{ad}(\sigma_{ad}) \Rightarrow Z_{ab}(\sigma_{ab}) = A_{ab} \exp(\eta_{ab}\sigma_{ab})$$

Belief Propagation Equations

$$\sum_{\vec{\sigma}_a} f_a(\vec{\sigma}_a) \exp\Big(\sum_{c \in a} \eta_{ac}\sigma_{ac}\Big) \left(\sigma_{ab} - \tanh\left(\eta_{ab} + \eta_{ba}\right)\right) = 0$$

Variational Method in Statistical Mechanics

$$P(\vec{\sigma}) = \frac{\prod_a f_a(\vec{\sigma}_a)}{Z}, \quad Z \equiv \sum_{\vec{\sigma}} \prod_a f_a(\vec{\sigma}_a)$$

Exact Variational Principle

J.W. Gibbs 1903 (or earlier)
also known as Kullback-Leibler (1951) in CS and IT

$$F\{b(\vec{\sigma})\} = -\sum_{\vec{\sigma}} b(\vec{\sigma}) \sum_a \ln f_a(\vec{\sigma}_a) + \sum_{\vec{\sigma}} b(\vec{\sigma}) \ln b(\vec{\sigma})$$

$$\left.\frac{\delta F}{\delta b(\vec{\sigma})}\right|_{b(\vec{\sigma})=p(\vec{\sigma})} = 0 \quad \text{under} \quad \sum_{\vec{\sigma}} b(\vec{\sigma}) = 1$$

Variational Ansatz

- Mean-Field: $p(\vec{\sigma}) \approx b(\vec{\sigma}) = \prod_{(a,b)} b_{ab}(\sigma_{ab})$
- Belief Propagation:

$$p(\vec{\sigma}) \approx b(\vec{\sigma}) = \frac{\prod_a b_a(\vec{\sigma}_a)}{\prod_{(a,b)} b_{ab}(\sigma_{ab})} \quad \text{(exact on a tree)}$$

$$\forall\, a;\ c \in a: \quad \sum_{\vec{\sigma}_a} b_a(\vec{\sigma}_a) = 1, \quad b_{ac}(\sigma_{ac}) = \sum_{\vec{\sigma}_a \setminus \sigma_{ac}} b_a(\vec{\sigma}_a)$$

Variational Method in Statistical Mechanics

$$P(\vec{\sigma}) = \frac{\prod_a f_a(\vec{\sigma}_a)}{Z}, \quad Z \equiv \sum_{\vec{\sigma}} \prod_a f_a(\vec{\sigma}_a)$$

Exact Variational Principe

J.W. Gibbs 1903 (or earlier)
also known as Kullback-Leibler (1951) in CS and IT

$$F\{b(\vec{\sigma})\} = -\sum_{\vec{\sigma}} b(\vec{\sigma}) \sum_a \ln f_a(\vec{\sigma}_a) + \sum_{\vec{\sigma}} b(\vec{\sigma}) \ln b(\vec{\sigma})$$

$$\left.\frac{\delta F}{\delta b(\vec{\sigma})}\right|_{b(\vec{\sigma})=p(\vec{\sigma})} = 0 \quad \text{under} \quad \sum_{\vec{\sigma}} b(\vec{\sigma}) = 1$$

Variational Ansatz

- Mean-Field: $p(\vec{\sigma}) \approx b(\vec{\sigma}) = \prod_{(a,b)} b_{ab}(\sigma_{ab})$
- Belief Propagation:

$$p(\vec{\sigma}) \approx b(\vec{\sigma}) = \frac{\prod_a b_a(\vec{\sigma}_a)}{\prod_{(a,b)} b_{ab}(\sigma_{ab})} \quad \text{(exact on a tree)}$$

$$\forall\ a;\ c \in a: \quad \sum_{\vec{\sigma}_a} b_a(\vec{\sigma}_a) = 1, \quad b_{ac}(\sigma_{ac}) = \sum_{\vec{\sigma}_a \setminus \sigma_{ac}} b_a(\vec{\sigma}_a)$$

Bethe Free Energy: variational approach

(Yedidia,Freeman,Weiss '01 - inspired by Bethe '35, Peierls '36)

$$F = \underbrace{-\sum_a \sum_{\vec\sigma_a} b_a(\vec\sigma_a)\ln f_a(\vec\sigma_a)}_{\text{self-energy}} + \underbrace{\sum_a \sum_{\vec\sigma_a} b_a(\vec\sigma_a)\ln b_a(\vec\sigma_a) - \sum_{(a,c)} b_{ac}(\sigma_{ac})\ln b_{ac}(\sigma_{ac})}_{\text{configurational entropy}}$$

$$\forall\ a;\ c\in a:\quad \sum_{\vec\sigma_a} b_a(\vec\sigma_a)=1,\quad b_{ac}(\sigma_{ac})=\sum_{\vec\sigma_a\setminus\sigma_{ac}} b_a(\vec\sigma_a)$$

$\Rightarrow$ Belief-Propagation Equations: $\left.\frac{\delta F}{\delta b}\right|_{\text{constr.}}=0$

Belief-Propagation as heuristics: iterative $\Rightarrow$ Gallager '61; MacKay '98

- Exact on a tree
- Trading optimality for reduction in complexity: $\sim 2^L \to \sim L$
- (BP = solving equations on the graph) $\neq$ (Message Passing = iterative BP)
- Convergence of MP to minimum of Bethe Free energy can be enforced
- $Z_{BP} \gtrless Z_{exact}$: BP ansatz in exact Gibbs Functional is not a truly variational substitution ($\sum_{\vec\sigma} b(\vec\sigma)=1$ is not guaranteed)

Bethe Free Energy: variational approach

(Yedidia,Freeman,Weiss '01 - inspired by Bethe '35, Peierls '36)

$$F = \underbrace{-\sum_a \sum_{\vec{\sigma}_a} b_a(\vec{\sigma}_a) \ln f_a(\vec{\sigma}_a)}_{\text{self-energy}} + \underbrace{\sum_a \sum_{\vec{\sigma}_a} b_a(\vec{\sigma}_a) \ln b_a(\vec{\sigma}_a) - \sum_{(a,c)} b_{ac}(\sigma_{ac}) \ln b_{ac}(\sigma_{ac})}_{\text{configurational entropy}}$$

$\forall\ a;\ c \in a: \quad \sum_{\vec{\sigma}_a} b_a(\vec{\sigma}_a) = 1, \quad b_{ac}(\sigma_{ac}) = \sum_{\vec{\sigma}_a \setminus \sigma_{ac}} b_a(\vec{\sigma}_a)$

$\Rightarrow$ Belief-Propagation Equations: $\left.\frac{\delta F}{\delta b}\right|_{\text{constr.}} = 0$

Belief-Propagation as heuristics: iterative $\Rightarrow$ Gallager '61; MacKay '98

- Exact on a tree
- Trading optimality for reduction in complexity: $\sim 2^L \to \sim L$
- (BP = solving equations on the graph) $\neq$ (Message Passing = iterative BP)
- Convergence of MP to minimum of Bethe Free energy can be enforced
- $Z_{BP} \gtrless Z_{\text{exact}}$: BP ansatz in exact Gibbs Functional is not a truly variational substitution ($\sum_{\vec{\sigma}} b(\vec{\sigma}) = 1$ is not guaranteed)

Linear Programming version of Belief Propagation

In the limit of large SNR, $\ln f_a \to \pm\infty$: BP→LP

$$\text{Minimize } F \approx E = -\sum_a \sum_{\vec{\sigma}_a} b_a(\vec{\sigma}_a) \ln f_a(\vec{\sigma}_a) = \text{self energy}$$

under set of linear constraints

LP decoding of LDPC codes — Feldman, Wainwright, Karger '03

- ML can be restated as an LP over a codeword polytope
- LP decoding is a "local codewords" relaxation of LP-ML
- Codeword convergence certificate
- Discrete and Nice for Analysis
- Large polytope $\{b_\alpha, b_i\} \Rightarrow$ Small polytope $\{b_i\}$

- Relation between BP and LP – Wainwright, Jordan '03

Outline

BP does not account for Loops

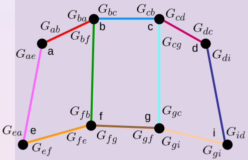
Questions:

- Is BP just a heuristic in a loopy case?
- Why does it (often) work so well?
- Does exact inference allow an expression in terms of BP?
- Can one correct BP systematically?

Gauge Transformations

Chertkov, Chernyak '06

Local Gauge, G, Transformations



$$Z = \sum_{\vec{\sigma}} \prod_a f_a(\vec{\sigma}_a), \;\; \vec{\sigma}_a = (\sigma_{ab}, \sigma_{ac}, \cdots)$$

$$\sigma_{ab} = \sigma_{ba} = \pm 1$$

$$f_a(\vec{\sigma}_a = (\sigma_{ab}, \cdots)) \to \sum_{\sigma'_{ab}} G_{ab}(\sigma_{ab}, \sigma'_{ab}) f_a(\sigma'_{ab}, \cdots)$$

$$\sum_{\sigma_{ab}} G_{ab}(\sigma_{ab}, \sigma') G_{ba}(\sigma_{ab}, \sigma'') = \delta(\sigma', \sigma'')$$

The partition function is invariant under any G-gauge!

$$Z = \sum_{\vec{\sigma}} \prod_a f_a(\vec{\sigma}_a) = \sum_{\vec{\sigma}} \prod_a \left(\sum_{\vec{\sigma}'_a} f_a(\vec{\sigma}'_a) \prod_{b \in a} G_{ab}(\sigma_{ab}, \sigma'_{ab}) \right)$$

Belief Propagation as a Gauge Fixing

Chertkov, Chernyak '06

$$Z=\sum_{\vec{\sigma}}\prod_a f_a(\vec{\sigma}_a)=\sum_{\boldsymbol{\sigma}}\prod_a\left(\sum_{\vec{\sigma}'_a} f_a(\vec{\sigma}'_a)\prod_{b\in a} G_{ab}(\sigma_{ab},\sigma'_{ab})\right)$$

$$Z=\underbrace{Z_0(G)}_{\substack{\text{ground state}\\ \vec{\sigma}=+\vec{1}}}+\underbrace{\sum_{\text{all possible colorings of the graph}} Z_c(G)}_{\vec{\sigma}\neq+\vec{1},\quad \text{excited states}}$$

Belief Propagation Gauge $\quad \forall a \ \&\ \forall b\in a:$

$$\sum_{\vec{\sigma}'_a} f_a(\vec{\sigma}')G^{(bp)}_{ab}(\sigma_{ab}=-1,\sigma'_{ab})\prod_{c\in a}^{c\neq b} G^{(bp)}_{ac}(+1,\sigma'_{ac})=0$$

No loose BLUE=colored edges at any vertex of the graph!

Belief Propagation as a Gauge Fixing (II)

$\forall a \;\&\; \forall b \in a:$

$$\left\{\begin{array}{l} \sum_{\vec{\sigma}'_a} f_a(\vec{\sigma}') G^{(bp)}_{ab}(-1,\sigma'_{ab}) \prod_{c\in a}^{c\neq b} G^{(bp)}_{ac}(+1,\sigma'_{ac}) = 0 \\ \sum_{\sigma_{ab}} G_{ab}(\sigma_{ab},\sigma') G_{ba}(\sigma_{ab},\sigma'') = \delta(\sigma',\sigma'') \end{array}\right. \Rightarrow \left\{\begin{array}{l} G^{(bp)}_{ba}(+1,\sigma'_{ab}) = \rho_a^{-1} \overbrace{\sum_{\vec{\sigma}'_a\setminus\sigma'_{ab}} f_a(\vec{\sigma}') \prod_{c\in a}^{c\neq b} G^{(bp)}_{ac}(+1,\sigma'_{ac})}^{sum-product} \\ \rho_a = \sum_{\vec{\sigma}'_a} f_a(\vec{\sigma}') \prod_{c\in a} G^{(bp)}_{ac}(+1,\sigma'_{ac}) \end{array}\right.$$

Belief Propagation in terms of Messages

$$G^{(bp)}_{ab}(+1,\sigma) = \frac{\exp(\sigma\eta_{ab})}{2\sqrt{\cosh(\eta_{ab}+\eta_{ba})}}, \quad G^{(bp)}_{ab}(-1,\sigma) = \sigma\frac{\exp(-\sigma\eta_{ba})}{2\sqrt{\cosh(\eta_{ab}+\eta_{ba})}} \Longrightarrow$$

$$\sum_{\vec{\sigma}_a\setminus\sigma_{ab}} f_a(\vec{\sigma}_a)\exp\left(\sum_{c\in a}\sigma_{ac}\eta_{ac}\right)(\sigma_{ab}-\tanh(\eta_{ab}+\eta_{ba})) = 0$$

$$b_a(\vec{\sigma}_a) = \frac{f_a(\vec{\sigma}_a)\exp\left(\sum_{b\in a}\sigma_{ab}\eta_{ab}\right)}{\sum_{\vec{\sigma}_a} f_a(\vec{\sigma}_a)\exp\left(\sum_{b\in a}\sigma_{ab}\eta_{ab}\right)}, \quad b_{ab}(\sigma) = \frac{\exp(\sigma(\eta_{ab}+\eta_{ba}))}{\sum_\sigma \exp(\sigma(\eta_{ab}+\eta_{ba}))}$$

Variational Principe and Gauge Fixing

$$Z = \underbrace{Z_0(G)}_{\vec{\sigma}=+\vec{1}} + \sum_{\vec{\sigma}\neq+\vec{1}} Z_c(G), \quad Z_0(G) \Rightarrow \underbrace{Z_0(\epsilon), \quad \epsilon_{ab}(\sigma_{ab}) = G_{ab}(+1,\sigma_{ab})}_{\text{depends only on the ground state gauges}}$$

Variational formulation of Belief Propagation

$$\left.\frac{\partial Z_0(\epsilon)}{\partial \epsilon_{ab}(\sigma_{ab})}\right|^{(bp)} = 0 \quad \Leftrightarrow \quad \text{Belief Propagation Equations}$$

$\mathcal{F}_0(\epsilon) = -\ln Z_0(\epsilon)$ – Bethe Free Energy at the BP-extremum

General Remarks on Gauge Fixing

- Related to the Re-parametrization Framework of Wainwright, Jaakkola and Willsky '03
- Generalizable to q-ary alphabet Chernyak, Chertkov '07
- ... suggests Loop Series for the Partition Function ⇒

Variational Principe and Gauge Fixing

$$Z = \underbrace{Z_0(G)}_{\vec{\sigma}=+\vec{1}} + \sum_{\vec{\sigma}\neq+\vec{1}} Z_c(G), \quad Z_0(G) \Rightarrow \underbrace{Z_0(\epsilon), \quad \epsilon_{ab}(\sigma_{ab}) = G_{ab}(+1,\sigma_{ab})}_{\text{depends only on the ground state gauges}}$$

Variational formulation of Belief Propagation

$$\left.\frac{\partial Z_0(\epsilon)}{\partial \epsilon_{ab}(\sigma_{ab})}\right|^{(bp)} = 0 \quad \Leftrightarrow \quad \text{Belief Propagation Equations}$$

$\mathcal{F}_0(\epsilon) = -\ln Z_0(\epsilon)$ – Bethe Free Energy at the BP-extremum

General Remarks on Gauge Fixing

- Related to the Re-parametrization Framework of Wainwright, Jaakkola and Willsky '03
- Generalizable to q-ary alphabet Chernyak, Chertkov '07
- ... suggests Loop Series for the Partition Function $\Rightarrow$

Loop Series:

Chertkov, Chernyak '06

Exact (!!) expression in terms of BP

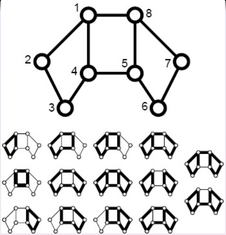
$$Z = \sum_{\vec{\sigma}_\sigma} \prod_a f_a(\vec{\sigma}_a) = Z_0 \left(1 + \sum_C r(C)\right)$$

$$r(C) = \frac{\prod_{a \in C} \mu_a}{\prod_{(ab) \in C} (1 - m_{ab}^2)} = \prod_{a \in C} \tilde{\mu}_a$$

$C \in$ Generalized Loops = Loops without loose ends

$$m_{ab} = \sum_{\vec{\sigma}_a} b_a^{(bp)}(\vec{\sigma}_a) \sigma_{ab}$$

$$\mu_a = \sum_{\vec{\sigma}_a} b_a^{(bp)}(\vec{\sigma}_a) \prod_{b \in a, C} (\sigma_{ab} - m_{ab})$$



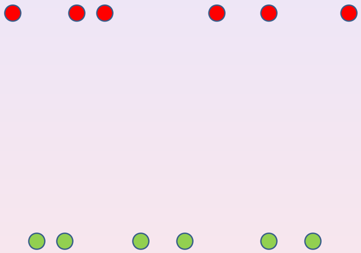
- The Loop Series is finite
- All terms in the series are calculated within BP
- BP is exact on a tree
- Different series for different BP-extrema
- Other (than BP) gauges lead to other representations
- Generalizes high-temperature expansion

Recent Developments (Loop Calculus & Related)

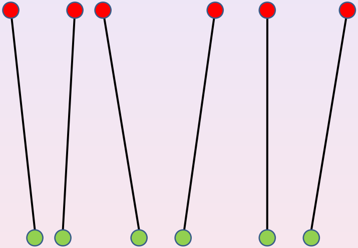
- Improving Decoding of LDPC codes in the Error-floor regime [MC, Chernyak '06]
- Loop Corrections for Approximate Inference on Factor Graphs [Gòmez, Mooij, Kappen '07]
- Low bound on Partition Function (Loop Series is positive): ordered state of FRFI [Sudderth, Wainwright, Willsky '08]
- BP is exact for some models with loops [Kolmogorov & Wainwright '05; Bayati, Shah, Sharma '06; MC '08]
- Loop Calculus for Planar Problems [MC, Chernyak, Teodorescu '08]
- Matching and Learning [MC, Kroc, Vergassola '08]

Outline

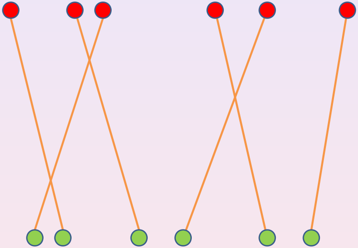
Two Snapshots



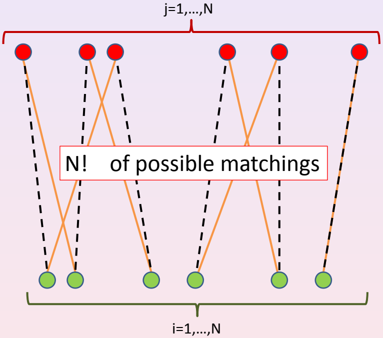
Two Snapshots



Two Snapshots



Two Snapshots



$d = 1$ advection and diffusion [example]

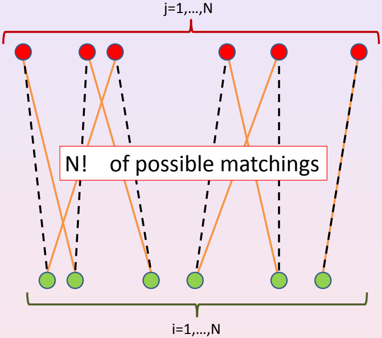
$$p_i^j = p(y^j|x_i) = \underbrace{\frac{\exp(-(y^j - e^S x_i)^2/(\kappa(e^{2S}-1))}{\sqrt{\pi\kappa(e^{2S}-1))}}}_{\text{probability of the i-j matching}}$$

$$p\left(\hat{\sigma}|\vec{x};\vec{y}\right) = Z^{-1}\prod_{(i,j)}\left(p_i^j\right)^{\sigma_i^j} F(\hat{\sigma})$$

$$F(\hat{\sigma}) \equiv \prod_i \delta\left(\sum_j \sigma_i^j, 1\right)\prod_j \delta\left(\sum_i \sigma_i^j, 1\right)$$

$$Z(\kappa, S) \equiv \sum_{\hat{\sigma}\equiv(0,1)^{N^2}}\prod_{(i,j)}\left(p_i^j\right)^{\sigma_i^j} F(\hat{\sigma})$$

Two Snapshots



$d = 1$ advection and diffusion [example]

$$p_i^j = p(y^j|x_i) = \underbrace{\frac{\exp(-(y^j - e^S x_i)^2/(\kappa(e^{2S}-1))}{\sqrt{\pi\kappa(e^{2S}-1))}}}_{\text{probability of the i-j matching}}$$

$$p\left(\hat{\sigma}|\vec{x};\vec{y}\right) = Z^{-1}\prod_{(i,j)}\left(p_i^j\right)^{\sigma_i^j} F(\hat{\sigma})$$

$$F(\hat{\sigma}) \equiv \prod_i \delta\left(\sum_j \sigma_i^j, 1\right)\prod_j \delta\left(\sum_i \sigma_i^j, 1\right)$$

$$Z(\kappa, S) \equiv \sum_{\hat{\sigma}\equiv(0,1)^{N^2}}\prod_{(i,j)}\left(p_i^j\right)^{\sigma_i^j} F(\hat{\sigma})$$

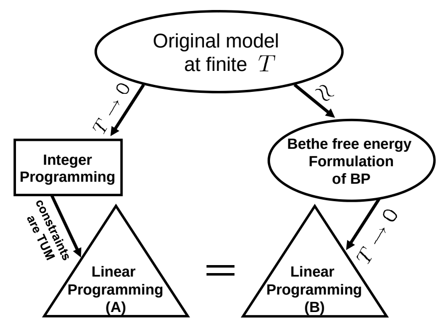
- Matching: $\text{argmax}_{\hat{\sigma}}\, p\left(\hat{\sigma}|\vec{x};\vec{y}\right)$ – [EASY, with BP→LP]
- "Learning": $\text{argmax}_{\kappa,S} Z(\kappa, S)$ – [DIFFICULT]

Can BELIEF PROPAGATION be exact
for some graphical models with LOOPS?

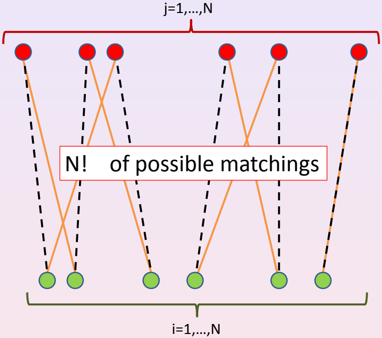
BP $\to$ LP is exact and easy: Bayati, Shah and Sharma '06

Maximum Weight Matching of a Bi-partite graph

BP $\to$ LP is exact [via Bethe Free Energy approach] MC'08



"Learning" the environment [MC, Kroc, Vergassola '08]



$d = 1$ advection and diffusion [example]

$$p_i^j = p(y^j|x_i) = \underbrace{\frac{\exp(-(y^j - e^S x_i)^2/(\kappa(e^{2S} - 1))}{\sqrt{\pi\kappa(e^{2S} - 1))}}}_{\text{probability of the i-j matching}}$$

$$p\left(\hat{\sigma}|\vec{x};\vec{y}\right) = Z^{-1}\prod_{(i,j)}\left(p_i^j\right)^{\sigma_i^j} F(\hat{\sigma})$$

$$F(\hat{\sigma}) \equiv \prod_i \delta\left(\sum_j \sigma_i^j, 1\right)\prod_j \delta\left(\sum_i \sigma_i^j, 1\right)$$

$$Z(\kappa, S) \equiv \sum_{\hat{\sigma}\equiv(0,1)^{N^2}} \prod_{(i,j)} \left(p_i^j\right)^{\sigma_i^j} F(\hat{\sigma})$$

$\text{argmax}_{\kappa,S} Z(\kappa, S)$

BP = bare approximation (heuristics)

+ Loop Series [corrections]

Loop Calculus for Matching

$$Z = Z_{\mathcal{BP}} * z, \quad z \equiv 1 + \sum_C r_C, \quad r_C = \left(\prod_{i\in C}(1-q_i)\right)\left(\prod_{j\in C}(1-q^j)\right)\prod_{(i,j)\in C}\frac{\beta_i^j}{1-\beta_i^j}$$

Mixed Derivative

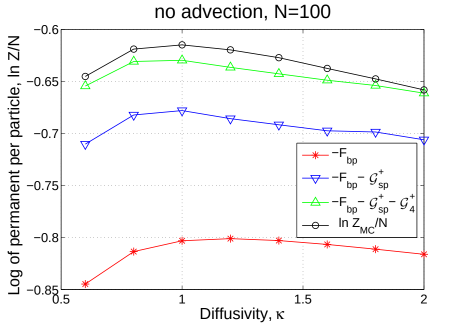
$$z = \left.\frac{\partial^{2N}\mathcal{Z}(\rho_1,\cdots,\rho_N,\rho^1,\cdots,\rho^N)}{\partial\rho_1\cdots\partial\rho_N\partial\rho^1\cdots\partial\rho^N}\right|_{\rho_1=\cdots=\rho_N=\rho^1\cdots=\rho^N=0}$$

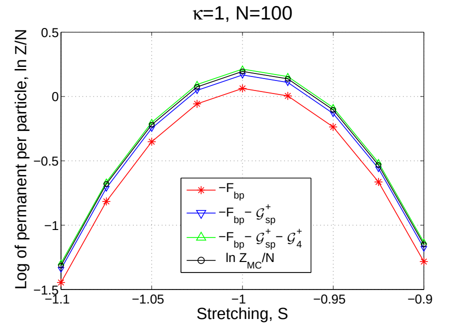
$$\mathcal{Z}(\vec{\rho}) \equiv \exp\left(\sum_i \rho_i + \sum_j \rho^j\right)\prod_{(i,j)}\left(1 + \frac{\beta_i^j}{(1-\beta_i^j)}\exp\left(-\rho_i - \rho^j\right)\right)$$

Cauchy Integral

$$z = \oint_{\Gamma_\rho} \exp\left(-\mathcal{G}(\vec{\rho})\right)\frac{\prod_i d\rho_i \prod_j d\rho^j}{(2\pi i)^{2N}}, \quad \mathcal{G}(\vec{\rho}) \equiv \sum_i 2\ln\rho_i + \sum_j 2\ln\rho^j - \ln\mathcal{Z}$$

"Learning": Cauchy-based heuristics vs FPRAS





BP, Loop Series = Mixed Derivatives = Cauchy Integral
≈ Saddle Point (heuristics) + Determinant + 4th order corr.

Fully Polynomial Randomized Algorithmic Scheme (Monte Carlo)

Shopping List for Matching & Learning +

- Analytical control of the saddle approximation quality
- Matching-Reconstruction as a Phase Transition
- $d = 2, 3$, realistic flows (multi scale)
- Multiple snapshots (a movie)
- ... technology for other reconstr. problems (e.g. phylogeny)

Future Challenges (Loop Calculus +)

- Disorder Average, Relation to Cavity, Replica Calculations
- Efficient Approximate Algorithms
- Relation & Complementarity to MCMC, Mixing Time
- Difficult "glassy" problems (physics), Satisfiability & Discrete Optimization (computer science, information theory)