



Gauge Transformations and Loop Calculus: General Theory & Application to Permanents

Misha Chertkov

T-Division LANL and New Mexico Consortium

Nov 2, 2011

Princeton Workshop

Outline

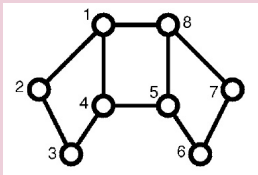
- 1 Graphical Models. Free Energy Functional. Belief Propagation.
 - Graphical Models
 - Complexity. Easy vs Difficult.
 - Bethe Free Energy & Belief Propagation (approx)
- 2 Gauge Transformations & Loop Series (Exact with BP)
 - Loops ... Questions
 - Gauge Transformations
 - Loop Series
- 3 The Story of Permanent: Learning Flows. $BP+ \rightarrow GBP+$.
 - Particle Tracking (Fluid Mechanics): Learning the Flow
 - BP and Loop Series (re-summed) for Permanents
 - Generalized BP. $perm=GBP^*perm$.

Graphical Models = The Language

Forney style - Variables on Edges

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

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



$$f_a \geq 0$$

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

Boolean, for example

$$\sigma_1 = (\sigma_{12}, \sigma_{14}, \sigma_{18})$$

$$\sigma_2 = (\sigma_{12}, \sigma_{23})$$

Objects of Interest

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

Easy & Difficult Problems

EASY [polynomial time]

- Any graphical problems **on a tree** (Bethe-Peierls, Dynamical Programming, BP, TAP and other names)
- Ground State of a Rand. Field Ferrom. Ising model on any graph
- Partition function of a planar Ising model or Dimer 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**

Exact Free Energy: from Discrete to Continuous

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

Exact Variational Principle

J.W. Gibbs 1903 (or earlier)

also known as Kullback-Leibler (1951) in CS and IT

$$F\{b(\boldsymbol{\sigma})\} = - \sum_{\boldsymbol{\sigma}} b(\boldsymbol{\sigma}) \sum_a \ln f_a(\boldsymbol{\sigma}_a) + \sum_{\boldsymbol{\sigma}} b(\boldsymbol{\sigma}) \ln b(\boldsymbol{\sigma})$$
$$-\log Z = \min_{\{b\}} F\{b(\boldsymbol{\sigma})\} \Big|_{\sum_{\boldsymbol{\sigma}} b(\boldsymbol{\sigma})=1}$$

“Variational” Ansatz

- Mean-Field: $p(\boldsymbol{\sigma}) \approx b(\boldsymbol{\sigma}) = \prod_{(a,b)} b_{ab}(\boldsymbol{\sigma}_{ab})$ exact in asymptotic
- Belief Propagation:

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

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

Bethe Free Energy: variational approach

(Yedidia, Freeman, Weiss '01 -

inspired by Bethe '35, Peierls '36)

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

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

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

Belief-Propagation as an approximation: iterative \Rightarrow Gallager '61; MacKay '98

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

Outline

- 1 Graphical Models. Free Energy Functional. Belief Propagation.
 - Graphical Models
 - Complexity. Easy vs Difficult.
 - Bethe Free Energy & Belief Propagation (approx)
- 2 Gauge Transformations & Loop Series (Exact with BP)
 - Loops ... Questions
 - Gauge Transformations
 - Loop Series
- 3 The Story of Permanent: Learning Flows. BP+ \rightarrow GBP+.
 - Particle Tracking (Fluid Mechanics): Learning the Flow
 - BP and Loop Series (re-summed) for Permanents
 - Generalized BP. perm=GBP*perm.



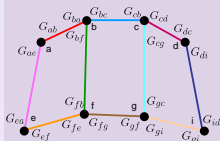
BP does not account for Loops

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?

Local Gauge, G , Transformations

Chertkov, Chernyak '06



$$Z = \sum_{\sigma} \prod_a f_a(\sigma_a), \quad \sigma_a = (\sigma_{ab}, \sigma_{ac}, \dots)$$

$\sigma_{ab} = \sigma_{ba} = \pm 1$; allows generalization to q -ary case

$$f_a(\sigma_a = (\sigma_{ab}, \dots)) \rightarrow \sum_{\sigma'_{ab}} G_{ab}(\sigma_{ab}, \sigma'_{ab}) f_a(\sigma'_{ab}, \dots)$$

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

The G -gauge Invariance!

[global]

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

Reminiscent of

- Re-parametrizations [pairwise models] of Wainwright, Jaakola, Willsky '03 [term by term]
- Holographic Transformations of Valiant '03 [global]

Belief Propagation as a Gauge Fixing

Chertkov, Chernyak '06

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

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

Belief Propagation Gauge

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

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

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

Belief Propagation as a Gauge Fixing (II)

$\forall a \text{ \& \; } \forall b \in a :$

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

Belief Propagation in terms of Messages

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

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

$$b_a^{(G)}(\sigma_a) = \frac{f_a(\sigma_a) \prod_{b \sim a} G_{ab}(0, \sigma_{ab})}{\sum_{\sigma_a} f_a(\sigma_a) \prod_{b \sim a} G_{ab}(0, \sigma_{ab})}, \quad b_{ab}^{(G)}(\sigma_{ab}) = G_{ab}(0, \sigma_{ab}) G_{ba}(0, \sigma_{ab})$$

Variational Principle and Gauge Fixing

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

“Variational” formulation of Belief Propagation

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

$\mathcal{F}_0(\epsilon) = -\ln Z_0(\epsilon)$ is the **Bethe Free Energy**
of Yedidia, Freeman, Weiss '01

Third Way

[Wainwright, Jaakola, Willsky '03]

Re-parametrization = Substitution

$$\forall \sigma : \prod_a f_a(\sigma_a) = Z^{(G)}(\mathbf{0}) \left(\prod_{\{a,b\}} b_{ab}^{(G)}(\sigma_{ab}) \right) \prod_a \left(\frac{b_a^{(G)}(\sigma_a)}{\prod_{b \sim a} b_{ab}^{(G)}(\sigma_{ab})} \right)$$

$$Z = \sum_{\sigma} \prod_a f_a(\sigma_a) = Z^{(G)}(\mathbf{0}) \mathbb{E}_{\text{edge}} \left[\prod_a \left(\frac{b_a^{(G)}(\sigma_a)}{\prod_{b \sim a} b_{ab}^{(G)}(\sigma_{ab})} \right) \right]$$
$$b_{\text{edge}}^{(G)}(\sigma) = \prod_{\{a,b\}} b_{ab}^{(G)}(\sigma_{ab})$$

From General Reparametrization/Gauge to BP

$$\forall a, \forall b \sim a, \forall \sigma_{ab} : \sum_{\sigma_a \setminus \sigma_{ab}} b_a^{(\text{BP})}(\sigma_a) = b_{ab}^{(\text{BP})}(\sigma_{ab})$$

my interpretation of [Sudderth, Wainwright, Willsky '09]

General Remarks

- Gauge Fixing and Re-parametrization are equivalent over the binary MRF (variables on vertices, pair-wise interaction)
- q -ary generalizations are different
- ... suggest **Loop Series** for the Partition Function \Rightarrow

Loop Series:

Chertkov, Chernyak '06

Exact (!!) expression in terms of BP

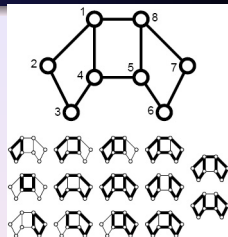
$$Z = \sum_{\sigma_{\sigma}} \prod_a f_a(\sigma_a) = Z_0 \left(1 + \sum_C r(C) \right)$$

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

$C \in$ **Generalized Loops** = sub-graphs without loose ends

$$m_{ab} = \sum_{\sigma_a} b_a^{(bp)}(\sigma_a) \sigma_{ab}$$

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



- The **Loop Series** is finite
- All terms in the series are calculated **within BP**
- BP is exact on a tree
- BP is a **Gauge fixing** condition [Valiant talk]. Other choices of Gauges would lead to different representations.
- LS is a generalization of "high-temperature" expansion [Jerrum talk].

Have been busy since ...

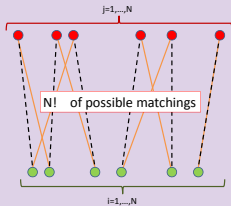
- Heuristics based on Loop Series for Improving BP (Graphical Codes) [MC, Chernyak '06-'08]
- Loop Tower (Loop Series beyond binary) [Chernyak, MC '07]
- Estimating Corrections to BP via Loop Series. Counting Independent Sets [Chandrasekaran, MC, Gamarnik, Shah, Shin '09]
- There are other interesting choices of Gauges. Matching and Fermion Models (new relations). [Chernyak, MC '08]
- **Planar and Surface GM which are Easy** [Chernyak, MC '08 + Chernyak, MC, Loebel in progress]
- Orbit Products for Gaussian GM ($\det = \text{BP} \cdot \det$) [Johnson, Chernyak, MC '09-]
- **The Story of Permanent** [MC, Kroc, Krzakala, Vergassola, Zdeborova '08-'11] + [Watanabe, MC '10] + [A.Yedidia, MC '11]

Outline

- 1 Graphical Models. Free Energy Functional. Belief Propagation.
 - Graphical Models
 - Complexity. Easy vs Difficult.
 - Bethe Free Energy & Belief Propagation (approx)
- 2 Gauge Transformations & Loop Series (Exact with BP)
 - Loops ... Questions
 - Gauge Transformations
 - Loop Series
- 3 The Story of Permanent: Learning Flows. BP+ \rightarrow GBP+.
 - Particle Tracking (Fluid Mechanics): Learning the Flow
 - BP and Loop Series (re-summed) for Permanents
 - Generalized BP. perm=GBP*perm.

Learning via Statistical Inference

Two images



And after all we actually don't need matching.
Our goal is to LEARN THE FLOW.

Particle Image Velocimetry & Lagrangian Particle Tracking [standard solution]

- Take snapshots often = Avoid trajectory overlap
- Consequence = A lot of data
- Gigabit/s to monitor a two-dimensional slice of a 10cm^3 experimental cell with a pixel size of 0.1mm and exposition time of 1ms
- Still need to "learn" velocity (diffusion) from matching

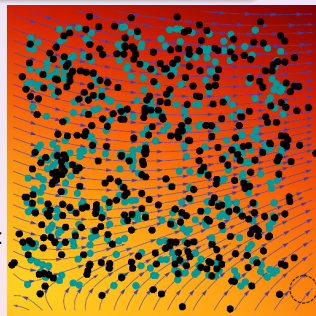
New twist

- Take fewer snapshots = Let particles overlap
- Put extra efforts into Learning/Inference
- Use the turbulence/physics community knowledge of Lagrangian evolution
- Focus on learning (rather than matching)

Inference & Learning by Passing Messages Between Images

- PNAS 10.1073/pnas.0910994107, MC, L.Kroc, F. Krzakala, L. Zdeborova, M. Vergassola

- Observe N identical particles in consequential images, $\mathbf{x} = (x_a^{(i)} | a = 1, \dots, N; i = 1, 2)$
- Have a low parametric model for the probability of a particle to travel from $x_a^{(1)}$ to $x_b^{(2)}$:
$$p_{ab} = f(x_a^{(1)}, x_b^{(2)}; \lambda)$$
- **Maximum Likelihood estimator is the permanent:**
$$\mathcal{P}(\mathbf{x}|\lambda) = \sum_{\sigma} \mathcal{P}(\sigma, \mathbf{x}|\lambda) = \text{perm}(p)$$
- Optimal Learning: $\arg \max_{\lambda} \text{perm}(p)$



- Max-Product is exact for max perfect matching [Bayati, Shah, Sharma 2006]
- Other encouraging experiments with BP for permanents [Huang, Jebara 2009]

Permanent as Counting over a Graphical Model

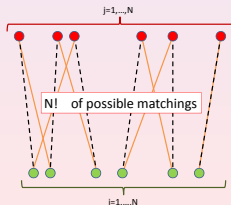
$$\text{perm}(\mu) = \underbrace{\sum_{\pi \in S_N} \mu_{a\pi(a)}}_{\text{sum over permutations}} = \overbrace{\sum_{\sigma} \prod_{a \in \mathcal{G}_0^{(N \times N)}} f_a^{\text{dimer}}(\sigma_a)}^{\text{partition function of a GM}}$$

$$\forall a \in \mathcal{G}_0^{(N \times N)} : f_a^{\text{dimer}}(\sigma_a) = \begin{cases} \sum_{b \sim a} \mu_{ab} \sigma_{ab}, & \sum_{b \sim a} \sigma_{ab} = 1; \\ 0, & \sum_{b \sim a} \sigma_{ab} \neq 1; \end{cases}$$

$$\sigma = (\sigma_{ab} = 0, 1 | \{a, b\} \in \mathcal{G}^{(N \times N)})$$

$$\forall a \in \mathcal{G}_0^{(N \times N)} : \sigma_a = (\sigma_{ab} = 0, 1 | b \sim a)$$

$\mathcal{G}^{(N \times N)} = (\mathcal{G}_0^{(N \times N)}, \mathcal{G}_1^{(N \times N)})$ is the bi-partite graph



Exact Free Energy for Permanents

Belief of a Perfect Matching expressed via Marginal Beliefs

$$b(\sigma) = \begin{cases} \prod_{a=1, \dots, N} \beta_{a\pi_a}, & \pi \in S_N, \forall a: \sigma_{a\pi_a} = \sum_{b \sim a} \sigma_{ab} = 1; \\ 0, & \text{otherwise;} \end{cases}$$

$$\mathcal{F}_{\text{perm}}(\beta) = \underbrace{- \sum_{\{a,b\}} \beta_{ab} \log(\mu_{ab})}_{\text{self energy}} + \overbrace{\sum_{\pi \in S_N} \left(\prod_a \beta_{a\pi_a} \right) \log \left(\prod_a \beta_{a\pi_a} \right)}^{\text{- entropy}}$$

$$\beta \in \mathcal{D}_{\text{perm}} = \underbrace{\left(\beta = (\beta_{ab} \in [0; 1] \mid \{a, b\} \in \mathcal{G}_0^{(N \times N)}) \mid \forall a: \sum_{b \sim a} \beta_{ab} = 1 \right)}_{\text{non-negative double stochastic}}$$

Bethe Free Energy for Permanents

$$\mathcal{F}_{\text{perm}}^{(\text{BP})}(\beta) = - \sum_{\{a,b\}} \beta_{ab} \log(\mu_{ab}) + \sum_{\{a,b\}} (\beta_{ab} \log(\beta_{ab}) - (1 - \beta_{ab}) \log(1 - \beta_{ab}))$$

$$\mathcal{D}_{\text{perm}} = \left(\beta = (\beta_{ab} \in [0; 1] | \{a, b\} \in \mathcal{G}_0^{(N \times N)}) \mid \forall a : \sum_{b \sim a} \beta_{ab} = 1 \right)$$

Convexity of the Bethe Free Energy

[Vontobel 2011]

$\mathcal{F}_{\text{perm}}^{(\text{BP})}(\beta)$ is convex over $\mathcal{D}_{\text{perm}}$

Belief Propagation Equations

$$\forall a : \sum_{b \sim a} \beta_{ab}^{(\text{BP})} = 1, \quad \forall \{a, b\} : \beta_{ab}^{(\text{BP})} (1 - \beta_{ab}^{(\text{BP})}) = \frac{\mu_{ab}}{u_a u_b}$$

- u_a are potentials=Lagrangian multipliers

Loop Series for Permanents (in three ways)

- assuming $\beta^{(\text{BP})}$ lies in the interior of $\mathcal{D}_{\text{perm}}$

sum over generalized loops

$$\frac{\text{perm}(\mu)}{Z_{\text{perm}}^{(\text{BP})}} = 1 + \sum_{\alpha \in \Upsilon_{\text{perm}}} \left(\prod_{a \in \alpha_0} (1 - q_a(\alpha)) \right) \left(\prod_{\{a,b\} \in \alpha_1} \frac{\beta_{ab}^{(\text{BP})}}{1 - \beta_{ab}^{(\text{BP})}} \right)$$

mixed derivative

$$\frac{\text{perm}(\mu)}{Z_{\text{perm}}^{(\text{BP})}} = \prod_a \frac{\partial}{\partial \eta_a} \left(e^{\sum_a \eta_a} \left(\prod_{\{a,b\}} \left(1 + e^{-\eta_a - \eta_b} \frac{\beta_{ab}^{(\text{BP})}}{1 - \beta_{ab}^{(\text{BP})}} \right) \right) \right) \Big|_{\eta \rightarrow 0}$$

perm=BP*perm

[Y. Watanabe, MC '09]

$$\frac{\text{perm}(\mu)}{Z_{\text{perm}}^{(\text{BP})}} = \frac{\text{perm}(\beta^{(\text{BP})} .* (1 - \beta^{(\text{BP})}))}{\prod_{\{a,b\}} (1 - \beta_{ab}^{(\text{BP})})}$$

- follows directly from the BP Eqs.
- can also be derived from LS

Generalized BP for Permanents

[A.Yedidia and MC 2011]

Bethe Free Energy for permanents $\gamma \in [-1; 1]$

$$\mathcal{F}_{\text{perm}}^{(\gamma)}(\beta) = -\sum_{\{a,b\}} \beta_{ab} \log \mu_{ab} + \sum_{\{a,b\}} (\beta_{ab} \log(\beta_{ab}) + \gamma(1 - \beta_{ab}) \log(1 - \beta_{ab}))$$

- $\beta^{(\gamma)} = \arg \min_{\beta \in \mathcal{D}_{\text{perm}}} \mathcal{F}_{\text{perm}}^{(\gamma)}(\beta)$
- BP: $\gamma = -1$; MF: $\gamma = 1$
- GBP of [Yedidia, Freeman, Weiss 2005] for permanents
- only modify coefficients, not regions

Generalized BP equations: β is non-negative double stochastic +

$$\forall \{a, b\} : \frac{\beta_{ab}^{(\gamma)}}{(1 - \beta_{ab}^{(\gamma)})^\gamma} = \frac{\mu_{ab}}{u_a u_b}$$

- $\gamma > 0$: $\beta^{(\gamma)}$ is in the interior of $\mathcal{D}_{\text{perm}}$
- $\gamma \leq 0$: $\beta^{(\gamma)}$ may be on the boundary of $\mathcal{D}_{\text{perm}}$
- $\beta^{(\gamma)}$ is on the boundary \Rightarrow the sub-set of $\beta_{ab} = 1$ forms a partial perfect matching
- $\mu \rightarrow \mu^{1/T}$, $T > 0 \Rightarrow \exists T_c > 0$ such that at $0 < T < T_c$, $\beta^{(\gamma)}$ is a perfect matching (full or partial)
- $\mathcal{F}_{\text{perm}}^{(\gamma)}(\beta^{(\gamma)})$ is continuous in γ , decreases monotonically with γ increase

perm=GBP*perm and Bounds

[A. Yedidia and MC 2011]

perm=GBP*perm: assuming $\beta^{(\gamma)}$ lies in the interior of $\mathcal{D}_{\text{perm}}$

$$\frac{\text{perm}(\mu)}{Z_{\text{perm}}^{(\gamma)}} = \text{perm} \left(\beta^{(\gamma)} ./ (1 - \beta^{(\gamma)})^\gamma \right) \left(\prod_{\{a,b\}} (1 - \beta_{ab}^{(\gamma)})^\gamma \right) \bullet \text{ follows directly from the GBP Eqs.}$$

Sandwich Inequality

$$Z_{\text{perm}}^{(\text{BP})} = Z_{\text{perm}}^{(\gamma=-1)} \leq \text{perm}(\mu) \leq Z_{\text{perm}}^{(\gamma=0)}$$

- $\text{perm}(\mu) \geq Z_{\text{perm}}^{(\gamma=-1)}$ \Leftarrow Gurvits 2011. For $\beta^{(\gamma)}$ in the interior, follows from perm=BP*perm + [Schrijver 1998]: $\text{perm}(\beta .* (1 - \beta)) \leq \prod_{\{a,b\}} (1 - \beta_{ab})$
- $\text{perm}(\mu) \leq Z_{\text{perm}}^{(\gamma=0)}$ follows from the perm=GBP*perm at $\gamma = 0$ + $\text{perm}(\beta) \leq 1$ if β is non-negative, double-stochastic.

Upper Bound Conjecture by Gurvits 2011

- (translated to our language) $Z_{\text{perm}}^{(\gamma=-1/2)}$ gives the upper-bound, achieved at μ which is block-diagonal = 2×2 blocks of 1/2
- The conjecture was confirmed in [our simulations](#), even though for many random ensembles the $Z_{\text{perm}}^{(\gamma=-1/2)}$ is a serious overshoot

Conclusions and Path Forward

Conclusions

- Gauge Invariance is an Important Principle - leads to (a) theory guarantees, and (b) better algorithms.
- Loop Calculus gives Natural Explanation to BP success in many loopy cases.
- Further Progress can be achieved (via BP and more generally gauge approach) making use of specific features of GM. The Story of Permanents is a good example of the power of the approach.

Path Forward

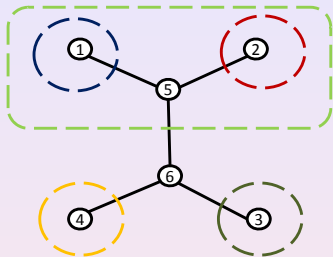
- More general matchings (bi-matchings, paths, clusters, etc)
- Other re-summation techniques, hybrid methods
- Learning (mixed optimization, counting +)
- Adversarial settings (min-max or min-sum)

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 Polynomial) in N is **EASY**, Exponential is **DIFFICULT**

BP is Exact on a Tree

Bethe '35, Peierls '36



$$Z_{51}(\sigma_{51}) = f_1(\sigma_{51}), \quad Z_{52}(\sigma_{52}) = f_2(\sigma_{52}),$$

$$Z_{63}(\sigma_{63}) = f_3(\sigma_{63}), \quad Z_{64}(\sigma_{64}) = f_4(\sigma_{64})$$

$$Z_{65}(\sigma_{56}) = \sum_{\sigma_5 \setminus \sigma_{56}} f_5(\sigma_5) Z_{51}(\sigma_{51}) Z_{52}(\sigma_{52})$$

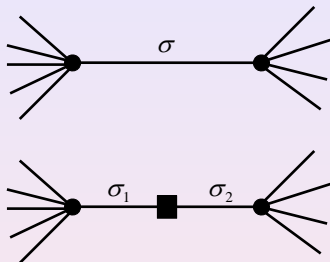
$$Z = \sum_{\sigma_6} f_6(\sigma_6) Z_{63}(\sigma_{63}) Z_{64}(\sigma_{64}) Z_{65}(\sigma_{65})$$

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

Belief Propagation Equations

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

e.g. Thouless-Anderson-Palmer (1977) Eqs.



$$\forall \sigma_1, \sigma_2 = \pm 1, \quad \forall \eta_1, \eta_2 \in \mathcal{C}$$

$$\begin{aligned} \delta(\sigma_1, \sigma_2) &= \frac{1 + \sigma_1 \sigma_2}{2} \text{ [substitution]} \\ &= \frac{\exp(\eta_1 \sigma_1 + \eta_2 \sigma_2)}{2 \cosh(\eta_1 + \eta_2)} (1 + \sigma_1 \sigma_2 \exp(-(\eta_1 + \eta_2)(\sigma_1 + \sigma_2))) \end{aligned}$$

Dance in Turbulence [movie]

Learn the flow from tracking particles

Direct Derivation of LS for Permanent from BP-Eqs.

$$\forall a : \sum_{b \sim a} \beta_{ab}^{(\text{BP})} = 1 \quad (*), \quad \forall \{a, b\} : \beta_{ab}^{(\text{BP})} (1 - \beta_{ab}^{(\text{BP})}) = \frac{\mu_{ab}}{u_a u_b} \quad (**)$$

$$\sum_{\{a,b\}} \beta_{ab}^{(\text{BP})} \log(\text{Eq.}(**)) \Rightarrow$$

$$\sum_{\{a,b\}} \beta_{ab}^{(\text{BP})} \log(u_a u_b) = \sum_a \log(u_a)$$

$$= \sum_{\{a,b\}} \left(\beta_{ab}^{(\text{BP})} \log(\mu_{ab} / \beta_{ab}^{(\text{BP})}) - \beta_{ab}^{(\text{BP})} \log(1 - \beta_{ab}^{(\text{BP})}) \right)$$

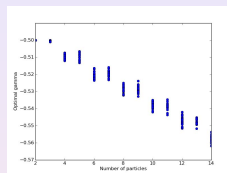
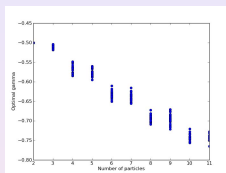
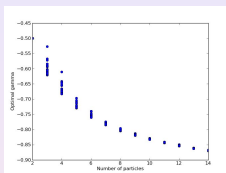
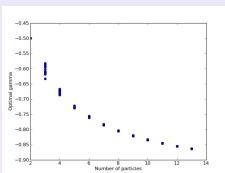
$$= \log(Z_{\text{perm}}^{(\text{BP})}) - \sum_{\{a,b\}} \log(1 - \beta_{ab}^{(\text{BP})})$$

$$\text{perm}(\text{Eq.}(**)) \Rightarrow$$

$$\text{perm}(\mu) = \text{perm}(\beta^{(\text{BP})} \cdot * (1 - \beta^{(\text{BP})})) \prod_a u_a$$

Combining \Rightarrow

$$\frac{\text{perm}(\mu)}{Z_{\text{perm}}^{(\text{BP})}} = \frac{\text{perm}(\beta^{(\text{BP})} \cdot * (1 - \beta^{(\text{BP})}))}{\prod_{\{a,b\}} (1 - \beta_{ab}^{(\text{BP})})},$$



- $[0; 1]$ uniform, "exp -1 " (exponential distributed), $[0; 1/20]$ -shifted and $[0; 1/2000]$ -shifted (random around the "Gurvits" matrix $= 2 \times 2$ blocks of $1/2$) ensembles. Full (small) matrix.
- Compared with exact (Zero-suppressed Decision Diagrams)
- Optimal γ vs $\#$ of particles (size of the matrix) is shown

◀ Perm=GBP*Perm