Exactness of Belief Propagation for Some Graphical Models with Loops

Michael Chertkov

Center for Nonlinear Studies & Theory Division, LANL

Jan 29, 2008 ITA 2008, UCSD

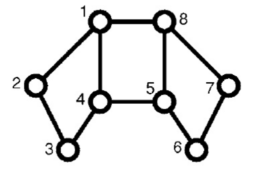
http://arxiv.org/abs/0801.0341

Boolean Graphical Models = The Language

Forney style - variables on the edges

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

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



$f_a \geq 0$

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

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

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

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

Example (from Statistical Physics)

Ising model $\sigma_i = \pm 1$

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

J_{ij} defines the graph (lattice)

Graphical Representation

Variables are usually associated with vertexes ... but transformation to the Forney-style graph (variables on the edges) is straightforward

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

Complexity & Algorithms

- How many operations are required?
- 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?

- Linear (or Algebraic) in N is EASY, Exponential is DIFFICULT
- Typical graphical problem with loops is DIFFICULT
- BELIEF PROPAGATION is a heuristics ignoring Loops

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:

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

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

Bethe Free Energy: variational approach
(Yedidia,Freeman,Weiss '01 - inspired by Bethe '35, Peierls '36)

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

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

$\Rightarrow$ 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: $\sim 2^L \to \sim L$
- (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

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

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

Tree reweighted BP of Kolmogorov & Wainwright '05

At $T \to 0$ BP solves the Ferromagnetic Random Field Ising model exactly on any graph!

Another Easy Example with Loops: Bayati, Shah and Sharma '06

Maximum Weight Matching of a Bi-partite graph

Our task for today:

Give simple and "almost universal" proof
for BP exactness in these special cases

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

Tree reweighted BP of Kolmogorov & Wainwright '05

At $T \to 0$ BP solves the Ferromagnetic Random Field Ising model
exactly on any graph!

Another Easy Example with Loops: Bayati, Shah and Sharma '06

Maximum Weight Matching of a Bi-partite graph

Our task for today:

Give simple and "almost universal" proof
for BP exactness in these special cases

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

Tree reweighted BP of Kolmogorov & Wainwright '05

At $T \to 0$ BP solves the Ferromagnetic Random Field Ising model
exactly on any graph!

Another Easy Example with Loops: Bayati, Shah and Sharma '06

Maximum Weight Matching of a Bi-partite graph

Our task for today:

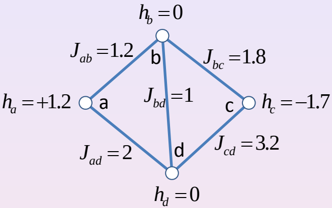
Give simple and "almost universal" proof
for BP exactness in these special cases

Ferromagnetic Random-Field Ising Model

$$p(\vec{\sigma}) = Z^{-1} \exp\left(\frac{1}{2T}\sum_{(i,j)} J_{ij}\sigma_i\sigma_j + \frac{1}{T}\sum_i h_i\sigma_i\right)$$

$J_{ij} \geq 0, \;\; h_i \gtrless 0$

(i,j) are edges on an undirected graph $\mathcal{G}$

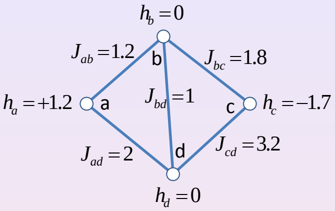


Ground State, $T \to 0$

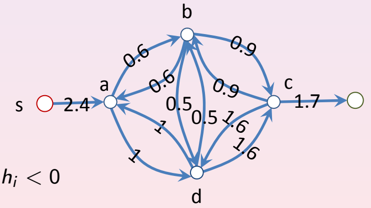
$$\min_{\boldsymbol{\sigma}}\left(-\frac{1}{2}\sum_{(i,j)\in\mathcal{G}} J_{ij}\sigma_i\sigma_j - \sum_{i\in\mathcal{G}} h_i\sigma_i\right)\Bigg|_{\forall\ i\in\mathcal{G}:\quad \sigma_i=\pm 1}$$

Undirected ⇒ Directed ⇒ (s-t)-Extended

$$\min_{\sigma}\left(-\frac{1}{2}\sum_{(i,j)\in\mathcal{G}} J_{ij}\sigma_i\sigma_j-\sum_{i\in\mathcal{G}}h_i\sigma_i\right)\Bigg|_{\forall\ i\in\mathcal{G}:\quad \sigma_i=\pm 1}$$



$$\min_{\sigma}\left(-\frac{1}{2}\sum_{(i,j)\in\mathcal{G}'_d} J_{i\to j}\sigma_i\sigma_j\right)\Bigg|_{\forall\ i\in\mathcal{G}'_d:\quad \sigma_i=\pm 1;\sigma_s=+1;\sigma_t=-1}$$

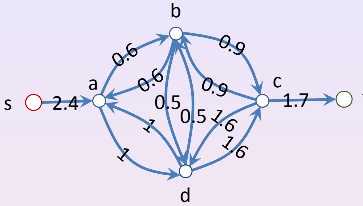


directed: $J_{i\to j}=J_{j\to i}=J_{ij}/2$

(s-t) extended: $J_{s\to i}=2h_i,\ \text{if}\ h_i>0\quad J_{i\to t}=2|h_i|,\ \text{if}\ h_i<0$

From Vertexes to Edges

$$\min_{\sigma}\left(-\frac{1}{2}\sum_{(i,j)\in\mathcal{G}'_d} J_{i\to j}\sigma_i\sigma_j\right)\Bigg|_{\forall\ i\in\mathcal{G}'_d:\quad \sigma_i=\pm1;\sigma_s=+1;\sigma_t=-1}$$



Integer Linear Programming

$$\eta_{i\to j}=\begin{cases}1, & \sigma_i=1,\sigma_j=-1\\ 0, & \text{otherwise}\end{cases}\qquad p_i=(1-\sigma_i)/2=0,1$$

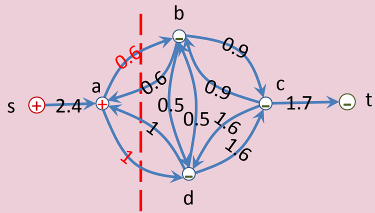
$$\sigma_i\sigma_j+\sigma_j\sigma_i=2-4(\eta_{i\to j}+\eta_{j\to i}),\quad \sigma_s\sigma_i=1-2\eta_{s\to i},\quad \sigma_i\sigma_t=1-2\eta_{i\to t}$$

$$-\frac{1}{2}\sum_{(i\to j)\in\mathcal{G}'_d} J_{i\to j}+\min_{\{\eta,p\}}\sum_{(i\to j)\in\mathcal{G}'_d} J_{i\to j}\eta_{i\to j}\Bigg|\ \begin{array}{c}\forall i\in\mathcal{G}'_d,\ p_i=0,1;\ \ p_s=0,\ \ p_t=1\\ \forall(i\to j)\in\mathcal{G}'_d:\\ p_i-p_j+\eta_{i\to j}=0,1\end{array}$$

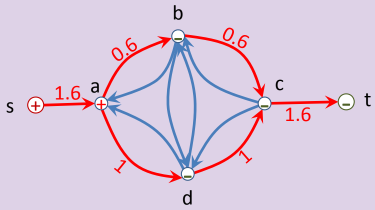
FRFI=Min-Cut=Max-Flow

$$-\frac{1}{2}\sum_{(i\to j)\in\mathcal{G}'_d} J_{i\to j} + \min_{\{\eta,p\}} \sum_{(i\to j)\in\mathcal{G}'_d} J_{i\to j}\eta_{i\to j}\Bigg| \begin{array}{c} \forall i\in\mathcal{G}'_d,\, p_i=0,1;\;\; p_s=0,\;\; p_t=1 \\ \forall (i\to j)\in\mathcal{G}'_d: \\ p_i-p_j+\eta_{i\to j}=0,1 \end{array}$$

Min-Cut



Max-Flow

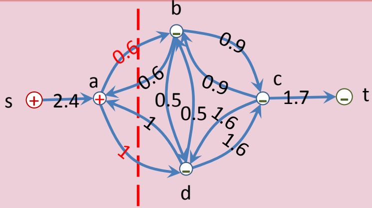


A.K. Hartman & H. Rieger, *Optimization Algorithms in Physics*, Wiley-VCH, 2002, and references therein

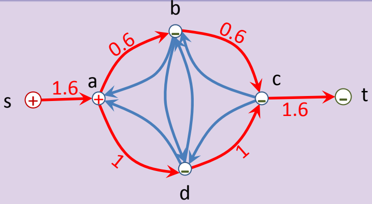
Back to Undirected Graph

$$-\frac{1}{2}\sum_{(i\to j)\in\mathcal{G}'_d} J_{i\to j} + \min_{\{\eta,p\}} \sum_{(i\to j)\in\mathcal{G}'_d} J_{i\to j}\eta_{i\to j}\Big| \quad \begin{array}{c}\forall i\in\mathcal{G}'_d,\, p_i=0,1;\;\; p_s=0,\;\; p_t=1\\ \forall(i\to j)\in\mathcal{G}'_d:\;\; p_i-p_j+\eta_{i\to j}=0,1\end{array}$$

Min-Cut



Max-Flow



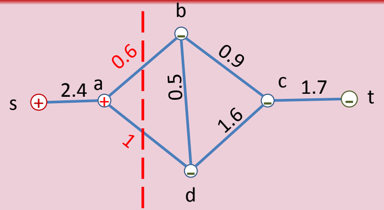
$$-\frac{1}{2}\sum_{(i,j)\in\mathcal{G}'} J_{ij} + \min_{\{\eta,p\}} \sum_{(i,j)\in\mathcal{G}'} J_{ij}\eta_{ij}\Big| \quad \begin{array}{c}\forall i\in\mathcal{G}',\, p_i=0,1;\;\; p_s=0,\;\; p_t=1\\ \forall(i,j)\in\mathcal{G}':\;\; p_i-p_j+\eta_{ij}=0,1\end{array}$$

$$J_{si}=2h_i,\quad \text{if}\quad h_i>0 \qquad J_{it}=2|h_i|,\quad \text{if}\quad h_i<0$$

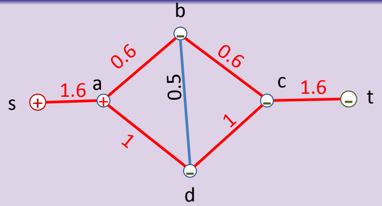
Back to Undirected Graph

$$-\frac{1}{2}\sum_{(i\to j)\in\mathcal{G}'_d} J_{i\to j} + \min_{\{\eta,p\}} \sum_{(i\to j)\in\mathcal{G}'_d} J_{i\to j}\eta_{i\to j}\Big| \quad \begin{array}{c}\forall i\in\mathcal{G}'_d,\, p_i=0,1;\;\; p_s=0,\;\; p_t=1\\ \forall (i\to j)\in\mathcal{G}'_d:\;\; p_i-p_j+\eta_{i\to j}=0,1\end{array}$$

Min-Cut



Max-Flow



$$-\frac{1}{2}\sum_{(i,j)\in\mathcal{G}'} J_{ij} + \min_{\{\eta,p\}} \sum_{(i,j)\in\mathcal{G}'} J_{ij}\eta_{ij}\Big| \quad \begin{array}{c}\forall i\in\mathcal{G}',\, p_i=0,1;\;\; p_s=0,\;\; p_t=1\\ \forall (i,j)\in\mathcal{G}':\;\; p_i-p_j+\eta_{ij}=0,1\end{array}$$

$$J_{si}=2h_i,\quad \text{if}\quad h_i>0 \qquad J_{it}=2|h_i|,\quad \text{if}\quad h_i<0$$

FRFI/Min-Cut/Max-Flow is EASY

- Many network algorithms. See e.g. T.H. Cormen, et al, *Introduction to Algorithms*, MIT-Press (2001)
- Reduction to Linear Programming. See e.g. H. Papadimitriou, I. Steiglitz, *Combinatorial Optimization: Alg. and Complexity*, Dover (1998)

Relaxation of Min-Cut Integer LP to respective LP is exact

$$-\frac{1}{2}\sum_{(i,j)\in\mathcal{G}'} J_{ij} + \min_{\{\eta,p\}} \sum_{(i,j)\in\mathcal{G}'} J_{ij}\eta_{ij}\Big|_{\substack{p_s=0,\ p_t=1;\ \forall i\in\mathcal{G}',\, p_i=0,1 \\ \forall (i,j)\in\mathcal{G}'\,:\ p_i-p_j+\eta_{ij}=0,1}}$$

- Matrix of LP constraints is Totally Uni-Modular (TUM)
- Min-Cut LP and Max-Flow LP are Dual

FRFI/Min-Cut/Max-Flow is EASY

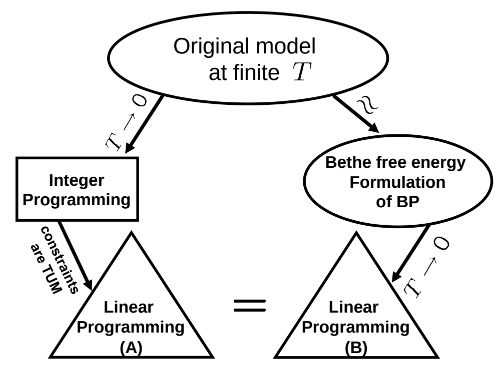
- Many network algorithms. See e.g. T.H. Cormen, et al, *Introduction to Algorithms*, MIT-Press (2001)
- Reduction to Linear Programming. See e.g. H. Papadimitriou, I. Steiglitz, *Combinatorial Optimization: Alg. and Complexity*, Dover (1998)

Relaxation of Min-Cut Integer LP to respective LP is exact

$$-\frac{1}{2}\sum_{(i,j)\in\mathcal{G}'} J_{ij} + \min_{\{\eta,p\}} \sum_{(ij)\in\mathcal{G}'_d} J_{ij}\eta_{ij}\Big| \begin{array}{l} p_s=0,\ p_t=1;\ \forall i\in\mathcal{G}', p_i=[0,1] \\ \forall(i,j)\in\mathcal{G}':\ p_i-p_j+\eta_{ij}=[0,1] \end{array}$$

- Matrix of LP constraints is Totally Uni-Modular (TUM)
- Min-Cut LP and Max-Flow LP are Dual

Proof of the BP-exactness via the Bethe Free energy approach



Chertkov '08

Bethe Free Energy for FRFI

At any Temperature

Minimize the Free Energy :

$$\mathcal{F} = E - TS, \quad E = -\sum_{(i,j)}\sum_{\sigma_i,\sigma_j} b_{ij}(\sigma_i,\sigma_j)\frac{J_{ij}}{2}\sigma_i\sigma_j - \sum_i\sum_{\sigma_i} b_i(\sigma_i)h_i\sigma_i$$

$$S = \sum_{(i,j)}\sum_{\sigma_i,\sigma_j} b_{ij}(\sigma_i,\sigma_j)\ln b_{(i,j)}(\sigma_i,\sigma_j) - \sum_i\sum_{\sigma_i} b_i(\sigma_i)\ln b_i(\sigma_i)$$

$$\forall i \ \ \& \ \ \forall j \in i: \quad b_i(\sigma_i) = \sum_{\sigma_j} b_{ij}(\sigma_i,\sigma_j), \quad \forall i: \quad \sum_{\sigma_i} b_i(\sigma_i) = 1$$

$T \to 0 \quad \Rightarrow \quad$ Linear Programming

Minimize the Self Energy :

$$E = -\sum_{(i,j)}\sum_{\sigma_i,\sigma_j} b_{ij}(\sigma_i,\sigma_j)\frac{J_{ij}}{2}\sigma_i\sigma_j - \sum_i\sum_{\sigma_i} b_i(\sigma_i)h_i\sigma_i$$

$$\forall i \ \ \& \ \ \forall j \in i: \quad b_i(\sigma_i) = \sum_{\sigma_j} b_{ij}(\sigma_i,\sigma_j), \quad \forall i: \quad \sum_{\sigma_i} b_i(\sigma_i) = 1$$

Linear Programming (B) for FRFI

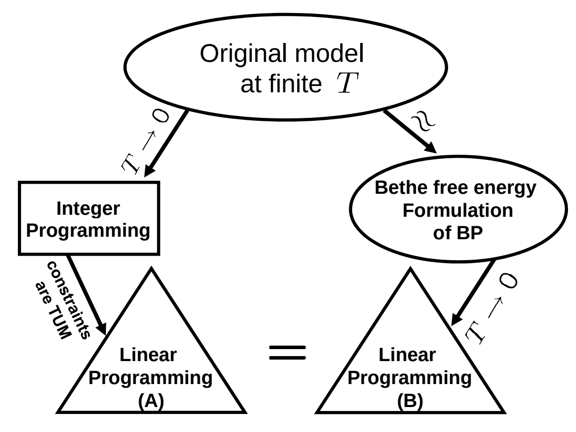
(s-t) modification: $\begin{cases} J_{si} = 2h_i & b_{si}(\sigma_s,\sigma_i) = b_i(\sigma_i)\delta(\sigma_s,+1) & h_i > 0 \\ J_{it} = 2|h_i| & b_{it}(\sigma_i,\sigma_t) = b_i(\sigma_i)\delta(\sigma_t,-1) & h_i < 0 \end{cases}$

$$\min_{\{b_i;b_{ij}\}} \left(- \sum_{(i,j)\in\mathcal{G}'} \sum_{\sigma_i,\sigma_j} b_{ij}(\sigma_i,\sigma_j) \frac{J_{ij}}{2} \sigma_i \sigma_j \right) \Bigg| \begin{array}{c} \forall i \in \mathcal{G}' \ \ \& \ \ \forall j \in i: \quad b_i(\sigma_i) = \sum_{\sigma_j} b_{ij}(\sigma_i,\sigma_j) \\ \forall i \in \mathcal{G}': \quad \sum_{\sigma_i} b_i(\sigma_i) = 1 \\ b_s(+) = 1 \ \ \& \ \ b_d(-) = 1 \end{array}.$$

$$\{b\} \to \{\mu,\pi\}: \quad \begin{cases} \mu_{ij} \equiv b_{ij}(+,-) + b_{ij}(-,+) = 1 - b_{ij}(+,+) - b_{ij}(-,-), & \forall (i,j) \in \mathcal{G}' \\ \pi_i = b_i(-) = b_{ij}(-,+) + b_{ij}(-,-), & \forall i \in \mathcal{G}' \end{cases}$$

$$-\frac{1}{2} \sum_{(i,j)\in\mathcal{G}'} J_{ij} + \min_{\{\mu,\pi\}} \sum_{(i,j)\in\mathcal{G}'} J_{ij}\mu_{ij} \Bigg| \begin{array}{c} \forall (i,j) \in \mathcal{G}': \quad \pi_i - \pi_j + \mu_{ij} \ge 0 \\ \forall (i,j) \in \mathcal{G}': \quad 1 \ge \pi_i, \mu_{ij} \ge 0 \\ \pi_s = 0, \quad \pi_t = 1 \end{array}$$

FRFI at $T = 0$ is solved exactly by BP



FRFI at $T=0$ is solved exactly by BP

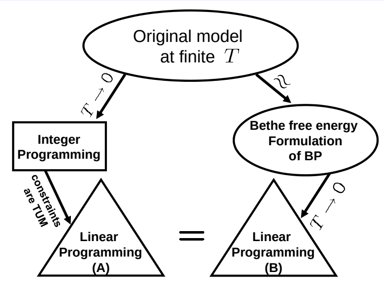
LP(A)

$$-\frac{1}{2}\sum_{(i,j)\in\mathcal{G}'} J_{ij} + \min_{\{\eta,p\}} \sum_{(ij)\in\mathcal{G}'} J_{ij}\eta_{ij}\Big| \begin{array}{c} p_s=0,\ \ p_t=1;\ \ \forall i\in\mathcal{G}', p_i=[0,1] \\ \forall (i,j)\in\mathcal{G}':\ \ p_i-p_j+\eta_{ij}=[0,1] \end{array}$$

LP(B)

$$-\frac{1}{2}\sum_{(i,j)\in\mathcal{G}'} J_{ij} + \min_{\{\mu,\pi\}} \sum_{(i,j)\in\mathcal{G}'} J_{ij}\mu_{ij}\Bigg| \begin{array}{c} \forall (i,j)\in\mathcal{G}':\quad \pi_i-\pi_j+\mu_{ij}\geq 0 \\ \forall (i,j)\in\mathcal{G}':\quad 1\geq \pi_i,\mu_{ij}\geq 0 \\ \pi_s=0,\quad \pi_t=1 \end{array}$$

LP(A)=LP(B)



The scheme also works for $T \to 0$ of

$$p(\boldsymbol{\sigma}) = Z^{-1} \exp\left(-T^{-1} \sum_i h_i \sigma_i\right) \prod_\alpha \delta\left(\sum_i J_{\alpha i} \sigma_i, m_\alpha\right).$$

where $\hat{J}$ is a Totally Uni-Modular matrix

Summary

- Loops are not always bad for BP!
- Hint (1): Check for TUM (IP=LP)
- Hint (2): Bethe Free Energy, BP $\rightarrow$ LP

Future Challenges

- Other problems easy for BP ... in spite of loops
- Finite (small) temperature. Perturbative exploration. Approximate Algorithms. Loop Calculus [Chertkov, Chernyak '06].
- "Almost" TUM problems (small duality gap). E.g. weakly frustrated spin-glass. Perturbative, Loop Calculus, Algorithms.
- ...

Thank You!