THE UNIVERSITY OF ARIZONA

Los Alamos NATIONAL LABORATORY

WAYNE STATE UNIVERSITY

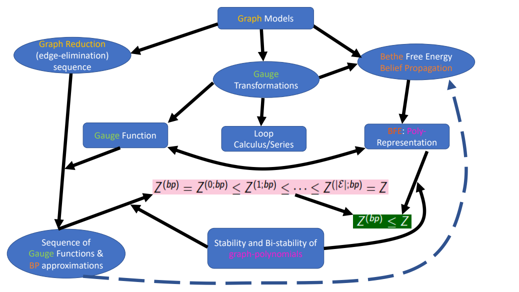
Gauges, Loops, and Polynomials: for Partition Functions of Graphical Models

Misha Chertkov [a,b],
Vladimir Chernyak [c] and Yury Maximov [b]

[a] Program in Applied Mathematics @ University of Arizona
[b] Theory Division @ Los Alamos National Laboratory
[c] Chemistry @ Wayne State University

03/21/19, Simons Institute, Berkeley

Outline

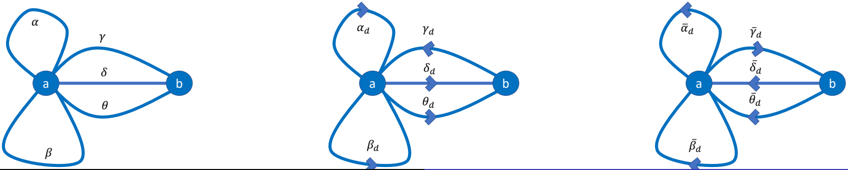


1. Intro - I: Graphical Models
   - Graphical Models: What?
   - Graphical Models: Why?
2. Intro-II: Bethe Free Energy
   - Variations: Exact $\Rightarrow$ Approximate
   - Poly- Representation for BFE & BP
   - Belief Polytope. Forcing BP into Interior.
3. Gauge T-, Graph R- & BP
   - Gauge Transformation, Gauge Function & Loop Series
   - Gauge Function $\Rightarrow$ Partition Function via Graph/Algebraic R-
   - Algebraic/Graph Reduction with BP- ... Exact & Approximate
4. Bi-Stability via BP & Gauges
   - BP vs Exact
   - Sequence of BP Low Bounds

What: Multi-Factor Graphical Model

- Multi-variate Probability Distribution: $p(\sigma) \doteq \frac{f(\sigma)}{Z}$
  - State vector: $\sigma = \{0,1\}^{|\mathcal{E}|}$
- Factorized according to the undirected graph, $(\mathcal{G},\mathcal{E})$
  - components of σ reside on edges: $\forall \alpha \in \mathcal{E}: \ \sigma_\alpha = \{0,1\}$
  - factors reside on nodes: $f(\sigma) \doteq \prod_{a\in\mathcal{V}} f_a(\sigma_a)$
  - $\forall a \in \mathcal{V}: \quad \sigma_a \in \{0,1\}^{e_{\rm d}(a)}$, e.g. $e_d(b) = (\gamma_d, \bar\delta_d, \bar\theta_d)$.
  - $(\mathcal{G},\mathcal{E})$ includes self-loops & multi-edges

Focus on computing Partition Function: $Z \doteq \sum_\sigma f(\sigma)$



Why: GM for Optimization, Inference & Learning

Efficient Optimization & Inference

- Optimization = Maximal Likelihood
- Inference
  - **Partition Function**
  - Marginal Probabilities
  - Sampling

Efficient Learning = reconstruct GM from samples

- Vuffray, Misra, Lokhov, MC (2016-) – not in this talk

Structure-specific OIL Applications

- Engineered Systems
  - Energy (Power, Natural Gas, Heating) Networks
- Physical Media
  - Fluid Mechanics

Exact Variational Principe for Partition Function

Gibbs ($\leq$ 1902)-Kullblack-Leibler (1951)

- $\boldsymbol{b} = (0 \leq \boldsymbol{b}(\sigma) \leq 1 | \forall \sigma)$ - "belief" for each state
- $$Z = \max_{\boldsymbol{b}} \prod_\sigma \left(\frac{\prod_a f_a(\sigma_a)}{\boldsymbol{b}(\sigma)} \right)^{\boldsymbol{b}(\sigma)} \Bigg| \begin{array}{l} \forall \sigma : \quad \boldsymbol{b}(\sigma) \geq 0 \\ \sum_\sigma \boldsymbol{b}(\sigma) = 1 \end{array}$$
  - Convex optimization over exponentially large space
  - Discrete $\to$ continuous

Bethe- (BP-) Variational Principe for Partition Function

Yedidia-Freeman-Weiss (2005)

- Exact over tree-graphs

$$\boldsymbol{b}(\sigma) \approx \frac{\prod_a b_a(\sigma_a)}{\prod_{\alpha\in\mathcal{E}} b_\alpha(\sigma_\alpha)} \text{ s.t.}$$

$$\forall a \in \mathcal{V},\ \forall \sigma_a \in S_a:\ \ b_a(\sigma_a) \doteq \sum_{\sigma\setminus\sigma_a} \boldsymbol{b}(\sigma)$$

$$\forall \alpha \in \mathcal{E},\ \forall \sigma_\alpha = \{0,1\}:\ \ b_\alpha(\sigma_\alpha) = \sum_{\sigma\setminus\sigma_\alpha} \boldsymbol{b}(\sigma)$$

  - Dynamic Programming $\Rightarrow$ Belief Propagation
    Bethe (1935), Peierls (1935), Gallager (1961), Pearl (1988)
- Ansatz in general. Substitute into Gibbs-Kubblack-Leibler $\Rightarrow$

Bethe- (BP-) Variational Principe for Partition Function

Yedidia-Freeman-Weiss (2005)

$$\min_b(-\log \mathcal{Z}(b)) \text{ s.t. } \begin{cases} \forall a\in\mathcal{V} & \forall \alpha\in e_{\rm d}(a),\ \forall\tau\in\{0,1\}: & b_\alpha(\tau)=\sum_{s\in S_a}^{s_\alpha=\tau} b_a(s) \\ & \forall s\in S_a: & b_a(s)\geq 0 \\ & & \sum_{s\in S_a} b_a(s)=1 \end{cases}.$$

where $\mathcal{Z}(b)$ is the Belief Propagation Partition Function (BP-PF)

$$\mathcal{Z}(b)\doteq\left(\prod_{a\in\mathcal{V}}\prod_{\sigma_a\in S_a}\left(\frac{f_a(\sigma_a)}{b_a(\sigma_a)}\right)^{b_a(\sigma_a)}\right)\left(\prod_{\alpha\in\mathcal{E}}\prod_{\sigma_\alpha\in\{0,1\}}(b_\alpha(\sigma_\alpha))^{b_\alpha(\sigma_\alpha)}\right),$$

of the vector b of marginal beliefs associated with edges and nodes

$$b=(b_\alpha(\sigma_\alpha)|\,\alpha\in\mathcal{E},\ \sigma_\alpha\in\{0,1\})\oplus(b_a(\sigma_a)|\,a\in\mathcal{V},\ \sigma_a\in S_a).$$

Poly- (Dual-) Representation (for Bethe Appr. of Z)

Anari & Oveis Gharan 2017, Straczak & Vishnoi 2017

$$\sup_{\beta\in[0;1]^{\mathcal{E}}} \min_{x\in\mathbb{R}_+^{\mathcal{E}_\mathrm{d}}} \mathcal{L}(\beta,x)$$

$$\mathcal{L}(\beta,x) \doteq \left(\prod_{\alpha\in\mathcal{E}} \beta_\alpha^{\beta_\alpha}(1-\beta_\alpha)^{1-\beta_\alpha}\right)\prod_{a\in\mathcal{V}} \frac{h_a(x_a)}{\prod\limits_{\alpha\in\mathcal{E}_d(a)} x_\alpha^{\beta_\alpha}}$$

$$\forall a\in\mathcal{V}: \quad h_a(x_a) \doteq \sum_{s\in S_a} f_a(s) \prod_{\alpha\in e_\mathrm{d}(a)} x_\alpha^{s_\alpha}$$

- Generalizes Gurvits (2011 – permanent)

Graphical Model Examples: Hard- & Soft-

Hard Example ($\forall a: \quad \exists \sigma_a$ s.t. $f_a(\sigma_a) = 0$): Perfect Matching

- $\forall a: \quad f_a(\sigma_a) = \begin{cases} \prod\limits_{\alpha \in e_d(a)} (\mu_\alpha)^{\sigma_\alpha}, & \sum_{\alpha \in e_d(a)} \sigma_\alpha = 1 \\ 0, & \text{otherwise} \end{cases}$
- Bethe (Free Energy) optimization is convex (Vontobel 2010)
- BP solution may be on the BP-polytope boundary (at low "temperature") (Watanabe, MC 2009)
- In the case of bi-partite graph Bethe appr. gives low bound for Z (Gurvits 2011)

Graphical Model Examples: Hard- & Soft-

Soft Example ($= \forall a: \quad f_a(\sigma_a) > 0$): "soft" Perfect Matching

- $\forall a: \quad f_a(\sigma_a) = \begin{cases} \prod\limits_{\alpha \in e_d(a)} (\mu_\alpha)^{\sigma_\alpha}, & \sum_{\alpha \in e_d(a)} \sigma_\alpha = 1 \\ \epsilon > 0, & \text{otherwise} \end{cases}$

For any Soft GM

- **Theorem:** Solution of Bethe optimization is always in the interior of BP polytope (MC, VC, YM 2019)

Graphical Model Examples: Hard- & Soft-

Soft Example ($= \forall a: \quad f_a(\sigma_a) > 0$): "soft" Perfect Matching

- $\forall a: \quad f_a(\sigma_a) = \begin{cases} \prod\limits_{\alpha \in e_d(a)} (\mu_\alpha)^{\sigma_\alpha}, & \sum_{\alpha \in e_d(a)} \sigma_\alpha = 1 \\ \epsilon > 0, & \text{otherwise} \end{cases}$

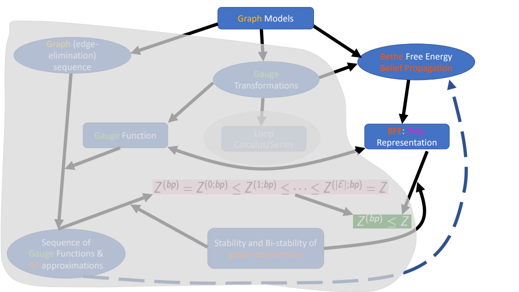
For any Soft GM

- **Theorem:** Solution of Bethe optimization is always in the interior of BP polytope (MC, VC, YM 2019)

will assume small-ϵ regularization for the remainder of the talk

- BP ($=$ solution of Bethe optimization) is in the interior

Outline


Graph Models
Graph (edge-elimination) sequence
Bethe Free Energy
Belief Propagation
Gauge Transformations
Gauge Function
Loop Calculus/Series
BFE: Poly-Representation
$Z^{(bp)} = Z^{(0;bp)} \leq Z^{(1;bp)} \leq \cdots \leq Z^{(|\mathcal{E}|;bp)} = Z$
$Z^{(bp)} \leq Z$
Sequence of Gauge Functions & BP approximations
Stability and Bi-stability of graph polynomials

Gauge Transformation [MC, Chernyak 2006]

Gauge-T= multi-linear transformation of factors:

- $\forall a, \quad \forall \sigma_a: \quad f_a(\sigma_a) \to \sum_{\varsigma_a} f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} G_{\alpha_d}(\sigma_\alpha, \varsigma_{\alpha_d})$
- which **keeps Z invariant**
  - $\forall G: \quad Z = \sum_\sigma \prod_{a \in \mathcal{V}} f_a(\sigma_a) = \sum_\sigma z(\sigma|G)$
  - $z(\sigma|G) \doteq \prod_a f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} G_{\alpha_d}(\sigma_\alpha, \varsigma_{\alpha_d})$
  - $\varsigma_a \doteq (\varsigma_{\alpha_d} = 0, 1 | \alpha_d \in e_d(a))$

related approaches

- Reparametrization (Wainwright, Jaakola, Willsky 2003)
- Hollographic transformation/algorithm (Valliant 2004)

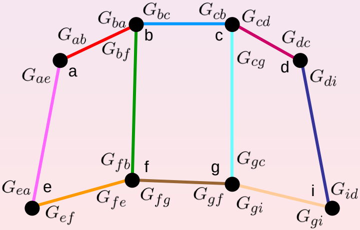
Gauge Transformation [MC, Chernyak 2006]

Gauge-T= multi-linear transformation of factors:

- $\forall a, \quad \forall \sigma_a: \quad f_a(\sigma_a) \to \sum_{\varsigma_a} f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} G_{\alpha_d}(\sigma_\alpha, \varsigma_{\alpha_d})$
- which **keeps Z invariant**
  - $\forall G: \quad Z = \sum_\sigma \prod_{a \in \mathcal{V}} f_a(\sigma_a) = \sum_\sigma z(\sigma|G)$
  - $z(\sigma|G) \doteq \prod_a f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} G_{\alpha_d}(\sigma_\alpha, \varsigma_{\alpha_d})$
  - $\varsigma_a \doteq (\varsigma_{\alpha_d} = 0, 1 | \alpha_d \in e_d(a))$

Orthogonality of GT

- $\boldsymbol{G}$ are (2×2) matrices
  two per age
- $\forall \alpha \in \mathcal{E}, \quad \boldsymbol{G}_{\alpha_d}^T * \boldsymbol{G}_{\bar{\alpha}_d} = 1$
  guarantees invariance of Z



Gauge Transformation [MC, Chernyak 2006]

Gauge-T= multi-linear transformation of factors:

- $\forall a, \quad \forall \sigma_a: \quad f_a(\sigma_a) \to \sum_{\varsigma_a} f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} G_{\alpha_d}(\sigma_\alpha, \varsigma_{\alpha_d})$
- which **keeps Z invariant**
  - $\forall G: \quad Z = \sum_\sigma \prod_{a \in \mathcal{V}} f_a(\sigma_a) = \sum_\sigma z(\sigma|G)$
  - $z(\sigma|G) \doteq \prod_a f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} G_{\alpha_d}(\sigma_\alpha, \varsigma_{\alpha_d})$
  - $\varsigma_a \doteq (\varsigma_{\alpha_d} = 0, 1 | \alpha_d \in e_d(a))$

Polynomial Representation of Gauges

- $\forall \alpha_d \in \mathcal{E}_d: \quad G_{\alpha_d} =$
$$\frac{1}{(x_{\alpha_d} x_{\bar\alpha_d})^{1/4}\sqrt{1+x_{\alpha_d}x_{\bar\alpha_d}}}\begin{pmatrix} \sqrt{x_{\bar\alpha_d}} & x_{\alpha_d}\sqrt{x_{\bar\alpha_d}} \\ -x_{\bar\alpha_d}\sqrt{x_{\alpha_d}} & \sqrt{x_{\alpha_d}} \end{pmatrix}$$
- $x \doteq (x_{\alpha_d} > 0 | \alpha_d \in \mathcal{E}_d)$ is positive component-wise

Gauge Transformation [MC, Chernyak 2006]

Gauge-T= multi-linear transformation of factors:

- $\forall a, \quad \forall \sigma_a: \quad f_a(\sigma_a) \to \sum_{\varsigma_a} f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} G_{\alpha_d}(\sigma_\alpha, \varsigma_{\alpha_d})$
- which **keeps Z invariant**
  - $\forall G: \quad Z = \sum_\sigma \prod_{a \in \mathcal{V}} f_a(\sigma_a) = \sum_\sigma z(\sigma|G)$
  - $z(\sigma|G) \doteq \prod_a f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} G_{\alpha_d}(\sigma_\alpha, \varsigma_{\alpha_d})$
  - $\varsigma_a \doteq (\varsigma_{\alpha_d} = 0, 1 | \alpha_d \in e_d(a))$

Theorem: Any extremum of BFE (a BP solution) is a stationary point of $z(x)$. Minimum of BFE is achieved at the extremum maximizing $z(x)$.

Extremum of BFE: $\quad \forall a \in \mathcal{V}, \quad \forall \alpha \in \mathcal{E}_{\mathrm{d}}(a): \partial_{x_\alpha} z(x)|_{x=x^{(\mathrm{bp})}} = 0$

Gauge Function: $\quad z(x) \doteq \frac{\prod_{a \in \mathcal{V}} h_a(x_a)}{\prod_{\alpha \in \mathcal{E}} (1 + x_{\alpha_d} x_{\bar{\alpha}_d})} = \frac{h(x)}{\prod_{\alpha \in \mathcal{E}} (1 + x_{\alpha_d} x_{\bar{\alpha}_d})}$

Gauge Transformation [MC, Chernyak 2006]

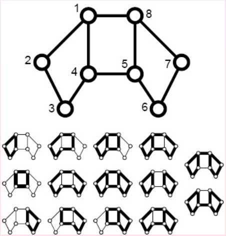
Gauge-T= multi-linear transformation of factors:

- $\forall a, \quad \forall \sigma_a: \quad f_a(\sigma_a) \to \sum_{\varsigma_a} f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} G_{\alpha_d}(\sigma_\alpha, \varsigma_{\alpha_d})$
- which **keeps Z invariant**
  - $\forall G: \quad Z = \sum_\sigma \prod_{a\in\mathcal{V}} f_a(\sigma_a) = \sum_\sigma z(\sigma|G)$
  - $z(\sigma|G) \doteq \prod_a f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} G_{\alpha_d}(\sigma_\alpha, \varsigma_{\alpha_d})$
  - $\varsigma_a \doteq (\varsigma_{\alpha_d} = 0, 1 | \alpha_d \in e_d(a))$

Loop Series/Calculus

$$Z = \sum_\sigma z(\sigma|x^{(bp)}), \quad \forall \sigma: \quad z(\sigma|x^{(bp)}) = Z^{(bp)} \frac{\prod\limits_a \mu_a^{(bp)}}{\prod\limits_\alpha \beta_\alpha^{(bp)}(1-\beta_\alpha^{(bp)})}$$

$$\forall a: \quad \mu_a \doteq \frac{\sum_{\varsigma_a} f_a(\varsigma_a) \prod_\alpha \left((x_\alpha^{(bp)})^{\varsigma_\alpha} (\varsigma_\alpha - \mu_\alpha)^{\sigma_\alpha}\right)}{\sum_{\varsigma_a} f_a(\varsigma_a) \prod_{\alpha\in e(\sigma)(a)} (x_\alpha^{(bp)})^{\varsigma_\alpha}}$$



Gauge Transformation [MC, Chernyak 2006]

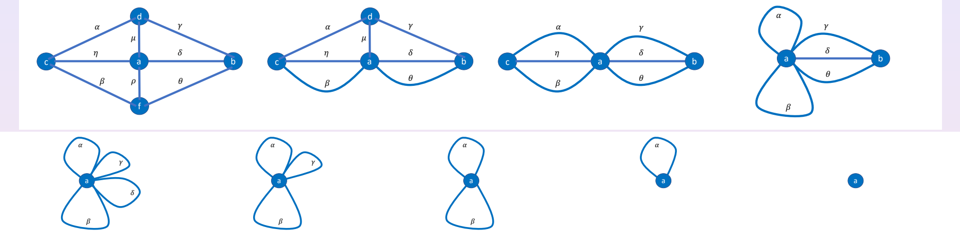
Gauge-T= multi-linear transformation of factors:

- $\forall a, \quad \forall \sigma_a: \quad f_a(\sigma_a) \to \sum_{\varsigma_a} f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} G_{\alpha_d}(\sigma_\alpha, \varsigma_{\alpha_d})$
- which **keeps Z invariant**
  - $\forall G: \quad Z = \sum_\sigma \prod_{a \in \mathcal{V}} f_a(\sigma_a) = \sum_\sigma z(\sigma|G)$
  - $z(\sigma|G) \doteq \prod_a f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} G_{\alpha_d}(\sigma_\alpha, \varsigma_{\alpha_d})$
  - $\varsigma_a \doteq (\varsigma_{\alpha_d} = 0, 1 | \alpha_d \in e_d(a))$

Further (finite graph) uses of "Beyond BP" = Loop Series & Gauge T-

- Minimum Weight Perfect Matching (optimization) – Blossom BP [Sungsoo Ahn, Sejun Park, Jinwoo Shin & MC NIPS 2015]
- LS maps GM to another GM (possibly with negative weights). The new GM can be sampled with MCMC. [Sungsoo Ahn, Jinwoo Shin, MC NIPS 2016]
- BP can be mixed with Mean Field – Gauging Variational Inference [Sungsoo Ahn, Jinwoo Shin & MC NIPS 2017]

Partition Function via Graph-R = Elimination of Edges



- Fix edge-elimination order
- Sum over edge variables sequentially
- Get a sequence of GMs
  - node-degree & complexity grows (exponentially)

Proposition:
Partition Functions in Graph-R sequence are identical/invariant

From Gauge-F to Partition Function

Gauge Function

$$z(x) \doteq \frac{\prod\limits_{a\in\mathcal{V}} h_a(x_a)}{\prod\limits_{\alpha\in\mathcal{E}} \left(1+x_{\alpha_d} x_{\bar{\alpha}_d}\right)}$$

Theorem: The Algebraic (Differentiate & Marginalize) -T applied to the Gauge-F results in Z

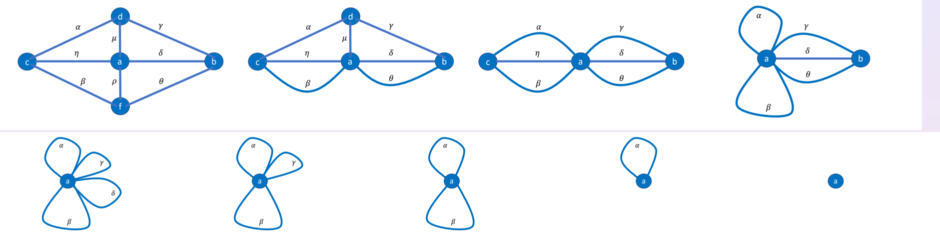
$$m=0: \quad x^{(0)} \doteq x, \quad \mathcal{Z}^{(0)}(x) \doteq z(x)$$

$$m=1,\cdots,|\mathcal{E}|: \quad x^{(m)} \doteq x^{(m-1)} \setminus \{x_{\alpha_d^{(m)}}, x_{\bar{\alpha}_d^{(m)}}\}$$

$$\mathcal{Z}^{(m)}(x^{(m)}) \doteq \left(1+\partial_{x_{\alpha_d^{(m)}}} \partial_{x_{\bar{\alpha}_d^{(m)}}}\right) \left(\left(1+x_{\alpha_d^{(m)}} x_{\bar{\alpha}_d^{(m)}}\right) \mathcal{Z}^{(m-1)}\right)\Big|_{x_{\alpha_d^{(m)}}=x_{\bar{\alpha}_d^{(m)}}=0}$$

$$m=|\mathcal{E}|: \quad x^{(|\mathcal{E}|)} = \emptyset, \quad \mathcal{Z}^{(|\mathcal{E}|)} = Z$$

Graph-R is equivalent to Algebraic-R



Theorem: Algebraic (diff.+marg.) -R = Graph (edge elimination)-T

$$\mathcal{Z}^{(m)}(x^{(m)}) = \frac{h^{(m)}(x^{(m)})}{\prod_{\alpha\in\mathcal{E}^{(m)}}\left(1 + x_{\alpha_d} x_{\bar{\alpha}_d}\right)}$$

$$h^{(m)}(x^{(m)}) \doteq \prod_{a\in\mathcal{V}^{(m)}}\left(\sum_{\varsigma_a} f_a^{(m)}(\varsigma_a) \prod_{\alpha_d\in e_d(a)} x_{\alpha_d}^{\varsigma_{\alpha_d}}\right)$$

Edge Elimination with Belief Propagation

Theorem: Any extremum of BFE (a BP solution) is a stationary point of $z(x)$. Minimum of BFE is achieved at the extremum maximizing $z(x)$.

Extremum of BFE: $\forall a \in \mathcal{V}, \quad \forall \alpha \in \mathcal{E}_{\rm d}(a) : \partial_{x_\alpha} z(x)|_{x=x^{(\rm bp)}} = 0$

Gauge Function: $z(x) \doteq \frac{\prod\limits_{a\in\mathcal{V}} h_a(x_a)}{\prod\limits_{\alpha\in\mathcal{E}} \left(1+x_{\alpha_d} x_{\bar{\alpha}_d}\right)} = \frac{h(x)}{\prod\limits_{\alpha\in\mathcal{E}} \left(1+x_{\alpha_d} x_{\bar{\alpha}_d}\right)}$

Edge Elimination with Belief Propagation

Theorem: Any extremum of BFE (a BP solution) is a stationary point of $z(x)$. Minimum of BFE is achieved at the extremum maximizing $z(x)$.

Extremum of BFE: $\forall a \in \mathcal{V}, \quad \forall \alpha \in \mathcal{E}_{\rm d}(a): \partial_{x_\alpha} z(x)|_{x=x^{(\rm bp)}} = 0$

Gauge Function: $z(x) \doteq \frac{\prod\limits_{a\in\mathcal{V}} h_a(x_a)}{\prod\limits_{\alpha\in\mathcal{E}} \left(1+x_{\alpha_d}x_{\bar{\alpha}_d}\right)} = \frac{h(x)}{\prod\limits_{\alpha\in\mathcal{E}} \left(1+x_{\alpha_d}x_{\bar{\alpha}_d}\right)}$

Let us "eliminate" a single edge

$$h(x) \doteq h^{(0,0)} + h^{(1,0)}x_{\alpha_d} + h^{(0,1)}x_{\bar{\alpha}_d} + h^{(1,1)}x_{\alpha_d}x_{\bar{\alpha}_d}$$

$$\partial_{x_{\alpha_d}}\partial_{x_{\bar{\alpha}_d}} \frac{h(x)}{1+x_{\alpha_d}x_{\bar{\alpha}_d}} = 0\Big|_{x^{(\alpha-bp)}} \quad \Rightarrow$$

Edge Elimination with Belief Propagation

Theorem: Any extremum of BFE (a BP solution) is a stationary point of $z(x)$. Minimum of BFE is achieved at the extremum maximizing $z(x)$.

Extremum of BFE: $\forall a \in \mathcal{V}, \quad \forall \alpha \in \mathcal{E}_{\mathrm{d}}(a) : \partial_{x_\alpha} z(x)|_{x=x^{(\mathrm{bp})}} = 0$

Gauge Function: $z(x) \doteq \frac{\prod_{a\in\mathcal{V}} h_a(x_a)}{\prod_{\alpha\in\mathcal{E}} \left(1+x_{\alpha_d} x_{\bar\alpha_d}\right)} = \frac{h(x)}{\prod_{\alpha\in\mathcal{E}} \left(1+x_{\alpha_d} x_{\bar\alpha_d}\right)}$

Let us "eliminate" a single edge

$$h(x) \doteq h^{(0,0)} + h^{(1,0)} x_{\alpha_d} + h^{(0,1)} x_{\bar\alpha_d} + h^{(1,1)} x_{\alpha_d} x_{\bar\alpha_d}$$

$$\partial_{x_{\alpha_d}} \partial_{x_{\bar\alpha_d}} \frac{h(x)}{1+x_{\alpha_d} x_{\bar\alpha_d}} = 0 \Big|_{x^{(\alpha-bp)}} \quad \Rightarrow$$

$$2x_{\alpha_d}^{(\alpha-\mathrm{bp})} h^{(1,0)} = 2x_{\bar\alpha_d}^{(\alpha-\mathrm{bp})} h^{(0,1)} = h^{(1,1)} - h^{(0,0)} + \sqrt{\left(h^{(1,1)} - h^{(0,0)}\right)^2 + 4h^{(0,1)} h^{(1,0)}}$$

Edge Elimination with Belief Propagation

Theorem: Any extremum of BFE (a BP solution) is a stationary point of $z(x)$. Minimum of BFE is achieved at the extremum maximizing $z(x)$.

Extremum of BFE: $\quad \forall a \in \mathcal{V}, \quad \forall \alpha \in \mathcal{E}_{\rm d}(a) : \partial_{x_\alpha} z(x)|_{x=x^{(\rm bp)}} = 0$

Gauge Function: $\quad z(x) \doteq \frac{\prod\limits_{a\in\mathcal{V}} h_a(x_a)}{\prod\limits_{\alpha\in\mathcal{E}}\left(1+x_{\alpha_d}x_{\bar\alpha_d}\right)} = \frac{h(x)}{\prod\limits_{\alpha\in\mathcal{E}}\left(1+x_{\alpha_d}x_{\bar\alpha_d}\right)}$

Let us "eliminate" a single edge

$$h(x) \doteq h^{(0,0)} + h^{(1,0)}x_{\alpha_d} + h^{(0,1)}x_{\bar\alpha_d} + h^{(1,1)}x_{\alpha_d}x_{\bar\alpha_d}$$

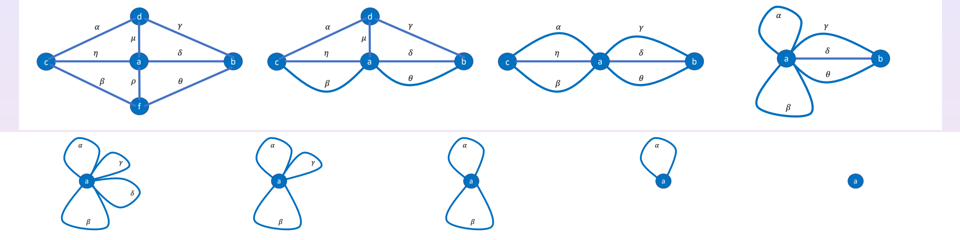
$$\partial_{x_{\alpha_d}}\partial_{x_{\bar\alpha_d}} \frac{h(x)}{1+x_{\alpha_d}x_{\bar\alpha_d}} = 0\Big|_{x^{(\alpha-bp)}} \quad \Rightarrow$$

$$2x_{\alpha_d}^{(\alpha-\rm bp)}h^{(1,0)} = 2x_{\bar\alpha_d}^{(\alpha-\rm bp)}h^{(0,1)} = h^{(1,1)} - h^{(0,0)} + \sqrt{\left(h^{(1,1)}-h^{(0,0)}\right)^2 + 4h^{(0,1)}h^{(1,0)}}$$

If α is a normal edge **BP Graph Reduction (edge contraction) is exact**

$$x_{\alpha_d}^{(\alpha-\rm bp)} = \frac{h^{(1,1)}}{h^{(1,0)}}, \quad x_{\bar\alpha_d}^{(\alpha-\rm bp)} = \frac{h^{(1,1)}}{h^{(0,1)}}, \quad h^{(\alpha-\rm bp)} = h^{(1,1)} + h^{(0,0)}$$

Graph Reduction with BP to Bouquet (exact)

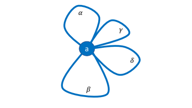


Introduce BP at each step in a Graph-R sequence

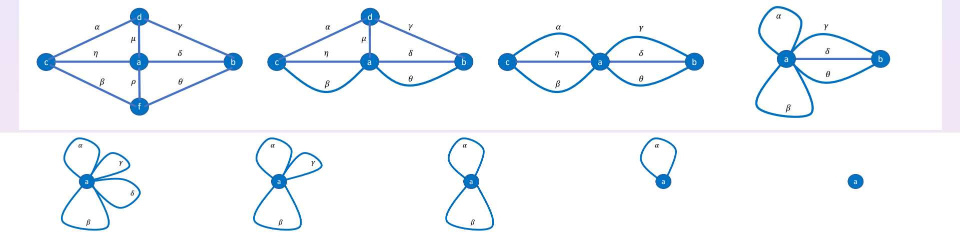
- **Theorem**: BP Graph-R of a **normal edge** is exact

Definition: Bouquet-graph: contract all normal edges

- **Corollary**: sequence of BP Graph-R leading to bouquet-graph is exact

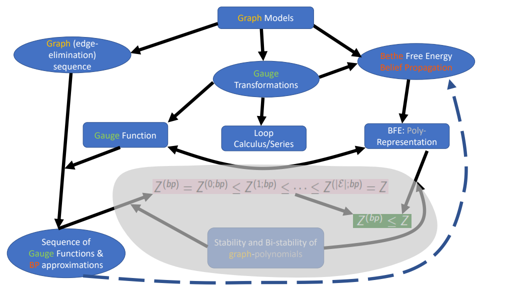


Graph Reduction with BP

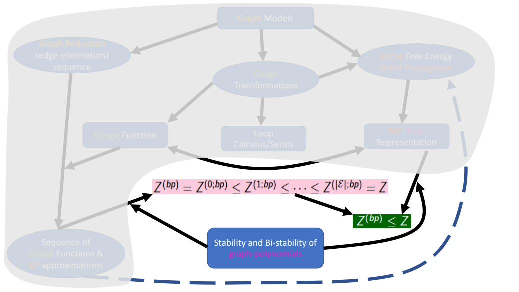


- BP-sequence (in general) is not monotonic.
- Complexity increases with contraction.
- Can be stopped/truncated at any step
  ⇒ an approximate elimination

No assumptions about GM so far – fully general



Outline


Graph Models
Graph Reduction (edge-elimination) sequence
Gauge Transformations
Bethe Free Energy Belief Propagation
Gauge Function
Loop Calculus/Series
BP: Poly Representation
$Z^{(bp)} = Z^{(0;bp)} \leq Z^{(1;bp)} \leq \cdots \leq Z^{(|\mathcal{E}|;bp)} = Z$
$Z^{(bp)} \leq Z$
Stability and Bi-stability of graph-polynomials
Sequence of Gauge Functions & BP approximations

BP vs Exact for an Edge-elimination step

$$\forall a \in \mathcal{V}, \quad \forall \alpha \in \mathcal{E}_{\mathrm{d}}(a): \; \partial_{x_\alpha} z(x)\big|_{x=x^{(\mathrm{bp})}} = 0$$

$$z(x) \doteq \frac{\prod_{a\in\mathcal{V}} h_a(x_a)}{\prod_{\alpha\in\mathcal{E}} \left(1+x_{\alpha_d} x_{\bar{\alpha}_d}\right)} = \frac{h(x)}{\prod_{\alpha\in\mathcal{E}} \left(1+x_{\alpha_d} x_{\bar{\alpha}_d}\right)}$$

$$h(x) \doteq h^{(0,0)} + h^{(1,0)} x_{\alpha_d} + h^{(0,1)} x_{\bar{\alpha}_d} + h^{(1,1)} x_{\alpha_d} x_{\bar{\alpha}_d}$$

BP vs Exact for an Edge-elimination step

$$\forall a \in \mathcal{V}, \quad \forall \alpha \in \mathcal{E}_{\rm d}(a): \partial_{x_\alpha} z(x)|_{x=x^{(\rm bp)}} = 0$$

$$z(x) \doteq \frac{\prod\limits_{a\in\mathcal{V}} h_a(x_a)}{\prod\limits_{\alpha\in\mathcal{E}} (1+x_{\alpha_d} x_{\bar\alpha_d})} = \frac{h(x)}{\prod\limits_{\alpha\in\mathcal{E}} (1+x_{\alpha_d} x_{\bar\alpha_d})}$$

$$h(x) \doteq h^{(0,0)} + h^{(1,0)} x_{\alpha_d} + h^{(0,1)} x_{\bar\alpha_d} + h^{(1,1)} x_{\alpha_d} x_{\bar\alpha_d}$$

Lemma: BP Reduction vs Exact Reduction (Differential Version)

$$\forall x^{(1)} = x \setminus \{x_{\alpha_d}, x_{\bar\alpha_d}\}: \left.\frac{h(x)}{1+x_{\alpha_d} x_{\bar\alpha_d}}\right|_{\substack{x_{\alpha_d} = x_{\alpha_d}^{(\alpha-\rm bp)} \\ x_{\bar\alpha_d} = x_{\bar\alpha_d}^{(\alpha-bp)}}} \leq h^{(1,1)}(x^{(1)}) + h^{(0,0)}(x^{(1)})$$

holds if $\quad h^{(0,1)}(x^{(1)}) h^{(1,0)}(x^{(1)}) \leq h^{(0,0)}(x^{(1)}) h^{(1,1)}(x^{(1)})$

BP vs Exact for an Edge-elimination step

Lemma: BP Reduction vs Exact Reduction (Differential Version)

$$\forall x^{(1)} = x \setminus \{x_{\alpha_d}, x_{\bar{\alpha}_d}\} : \left.\frac{h(x)}{1+x_{\alpha_d} x_{\bar{\alpha}_d}}\right|_{\substack{x_{\alpha_d} = x_{\alpha_d}^{(\alpha-\mathrm{bp})} \\ x_{\bar{\alpha}_d} = x_{\bar{\alpha}_d}^{(\alpha-bp)}}} \leq h^{(1,1)}(x^{(1)}) + h^{(0,0)}(x^{(1)})$$

holds if $\quad h^{(0,1)}(x^{(1)}) h^{(1,0)}(x^{(1)}) \leq h^{(0,0)}(x^{(1)}) h^{(1,1)}(x^{(1)})$

Lemma: BP Reduction vs Exact Reduction (Variational Version) [Anari & Oveis Gharan 2017] & [Straczak & Vishnoi 2017]

$\forall x^{(1)} : \quad h^{(0,1)}(x^{(1)}) h^{(1,0)}(x^{(1)}) \leq h^{(0,0)}(x^{(1)}) h^{(1,1)}(x^{(1)})$ guarantees that

$$\forall \beta_\alpha \in [0;1], \quad (\beta_\alpha)^{\beta_\alpha} (1-\beta_\alpha)^{1-\beta_\alpha} \inf_{x_{\alpha_d}, x_{\bar{\alpha}_d}} \frac{h(x)}{(x_{\alpha_d} x_{\bar{\alpha}_d})^{\beta_\alpha}} \leq h^{(1,1)}(x^{(1)}) + h^{(0,0)}(x^{(1)})$$

When $Z^{(bp)}$ is a low bound?

Definition: [Real-Stability (RS) and Bi-Stability (BS)]

- A nonzero polynomial, $g(x) \in \mathbb{R}[x_1, \cdots, x_N]$, with real coefficients is RS if none of its roots $z = (z_1, \cdots, z_N) \in \mathbb{C}^N$ satisfies: $\text{Im}(z_i) > 0$ for every $i = 1, \cdots, N$.
- A polynomial $h(x_{\alpha_d^{(1)}}, x^{(1)}_{\bar{\alpha}_d}; x_{\alpha_d^{(2)}}, x^{(2)}_{\bar{\alpha}_d} \cdots)$ is BS if $h(x_{\alpha_d^{(1)}}, -x^{(1)}_{\bar{\alpha}_d}; x_{\alpha_d^{(2)}}, -x^{(2)}_{\bar{\alpha}_d} \cdots)$ is Real Stable.

Theorem: Monotonicity of BP-estimations (for contraction sequence)

If $\forall a \in \mathcal{V},\ h_a(x_a)$ is bi-stable then

- each polynomial in the contraction (=differentiate+marginalize) sequence, $h^{(m)}(x^{(m)})$, is stable
- Value of BP estimation for Z does not decrease with elimination: $Z^{(bp)} = Z^{(0;bp)} \leq Z^{(1;bp)} \leq \cdots \leq Z^{(|\mathcal{E}|;bp)} = Z$.

When $Z^{(bp)}$ is a low bound?

Theorem: Monotonicity of BP-estimations (for contraction sequence)

If $\forall a \in \mathcal{V}, \; h_a(x_a)$ is bi-stable then

- each polynomial in the contraction (=differentiate+marginalize) sequence, $h^{(m)}(x^{(m)})$, is stable
- Value of BP estimation for Z does not decrease with elimination: $Z^{(bp)} = Z^{(0;bp)} \leq Z^{(1;bp)} \leq \cdots \leq Z^{(|\mathcal{E}|;bp)} = Z$.

$Z^{(bp)} \leq Z$

- via poly-methods
  - Gurvits (2011) - permanents
  - Straczak & Vishnoi 2017 - bi-partite stable
  - Anari & Oveis Gharan 2017 - bi-stable

When $Z^{(bp)}$ is a low bound?

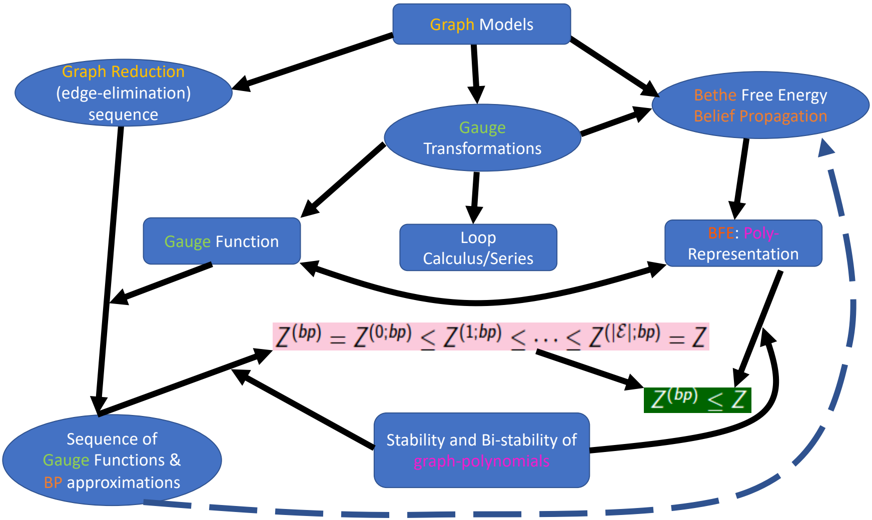
$Z^{(bp)} \leq Z$

- Attractive (ferro) Ising model

- via Loop Series, with restrictions (all loop terms are positive) Sudderth, Wainwright, Willsky 2007
- via Graph Covers (Vontobel) with no restrictions Ruozi 2012

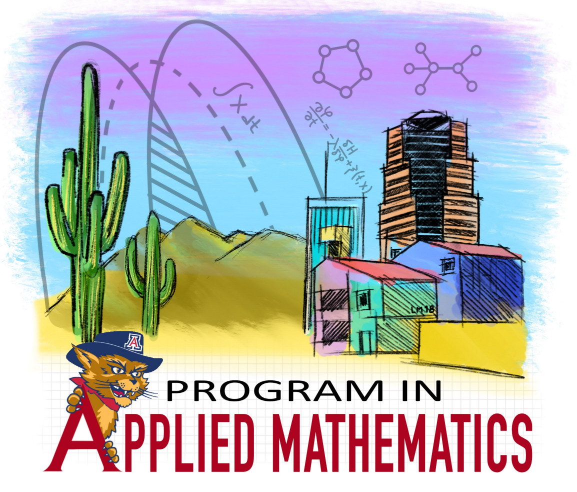
another approach – not related to the poly- considerations

Summary


Graph Models
Graph Reduction (edge-elimination) sequence
Gauge Transformations
Bethe Free Energy Belief Propagation
Gauge Function
Loop Calculus/Series
BFE: Poly-Representation
$Z^{(bp)} = Z^{(0;bp)} \leq Z^{(1;bp)} \leq \cdots \leq Z^{(|\mathcal{E}|;bp)} = Z$
$Z^{(bp)} \leq Z$
Sequence of Gauge Functions & BP approximations
Stability and Bi-stability of graph-polynomials

Path Forward

- Mixing the "differentiate and marginalize" approach with the **Loop Calculus**/Series
- Generalization to **higher alphabets** (homogeneous polynomials)
- **Tighter** Approximations (mix of gauge/BP- and graph-transformations)
  - theory & heuristics
- From **global to local** stability – broader class of "tractable" polynomials
- Sequential **Elimination**, e.g. extending
  - Gauged Mini-Bucket Elimination for Approximate Inference [Sungsoo Ahn, Jinwoo Shin, Adrian Weller & MC 2018]
  - Bucket renormalization for approximate inference [Sungsoo Ahn, Jinwoo Shin, Adrian Weller & MC 2018]



Thank You !