Convergent Message-Passing Algorithms for Inference over General Graphs with Convex Free Energies

Tamir Hazan, Amnon Shashua

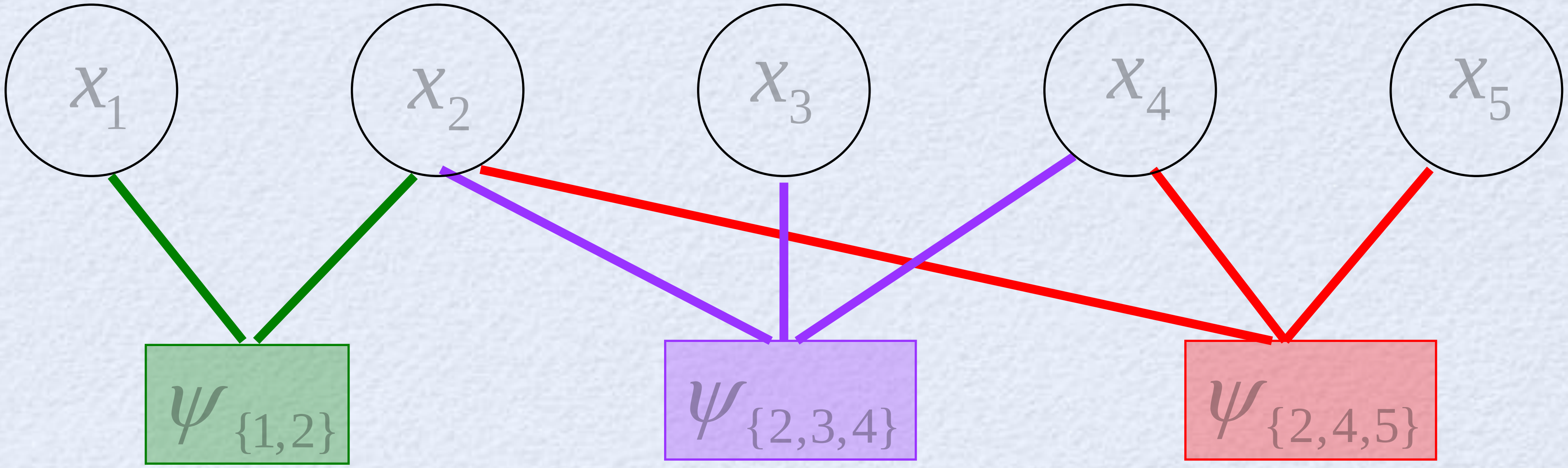
School of Computer Science and Engineering
Hebrew University of Jerusalem

Graphical Models - Background

Input: $p(x_1,...,x_n) \propto \prod \psi_\alpha(x_\alpha)$ **Output:** $p(x_i) = ?$

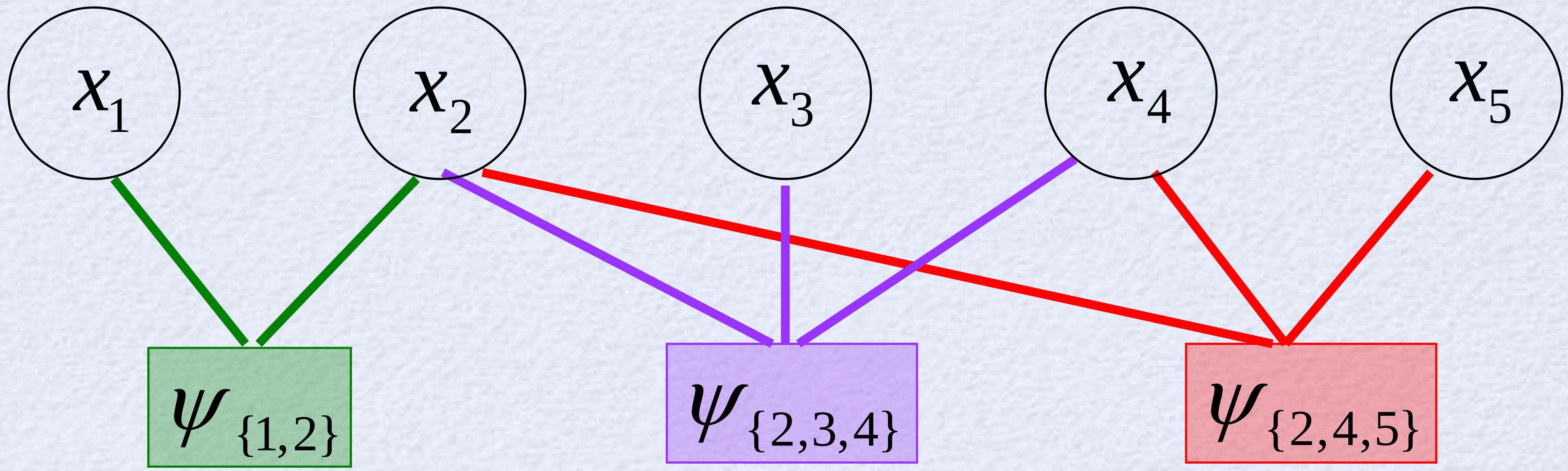
Graphical Models - Background

Input: $p(x_1,\ldots,x_n) \propto \prod \psi_\alpha(x_\alpha)$ **Output:** $p(x_i) = ?$



Graphical Models - Background

Input: $p(x_1,\ldots,x_n) \propto \prod \psi_\alpha(x_\alpha)$ **Output:** $p(x_i) = ?$



Belief Propagation:

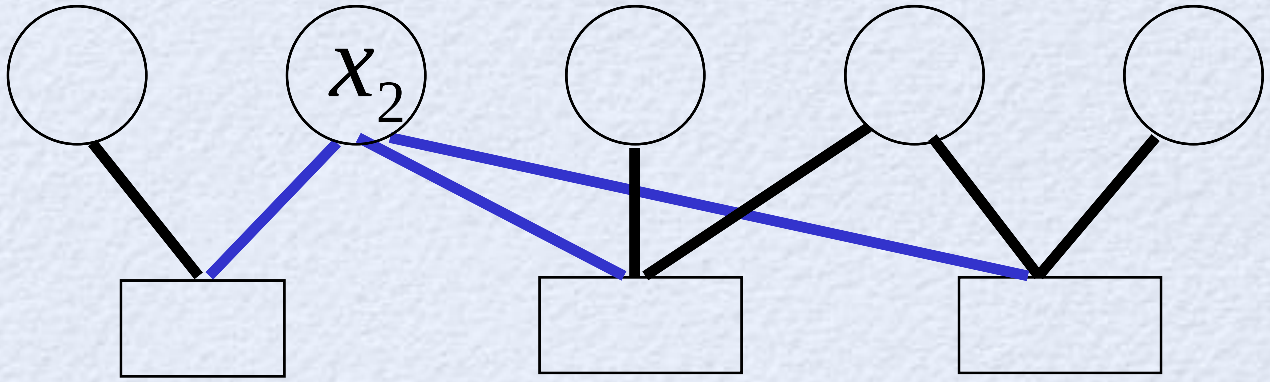
- $n_{i\to a}(x_i) = \prod_{c\in N(i)\setminus a} m_{c\to i}(x_i)$
- $m_{a\to i}(x_i) := \sum_{\mathbf{x}_a \setminus x_i} \psi_a(\mathbf{x}_a) \prod_{j\in N(a)\setminus i} n_{j\to a}(x_j)$

BP - Variational Methods

Bethe free energy: Approximating the marginals $b_\alpha(x_\alpha), b_i(x_i)$

$$\min_{b_\alpha, b_i} \quad -\sum_\alpha \sum_{x_\alpha} b_\alpha(x_\alpha) \ln \psi_\alpha(x_\alpha) - \sum_\alpha H(b_\alpha) - \sum_i (1-d_i) H(b_i)$$

$$d_2 = 3$$

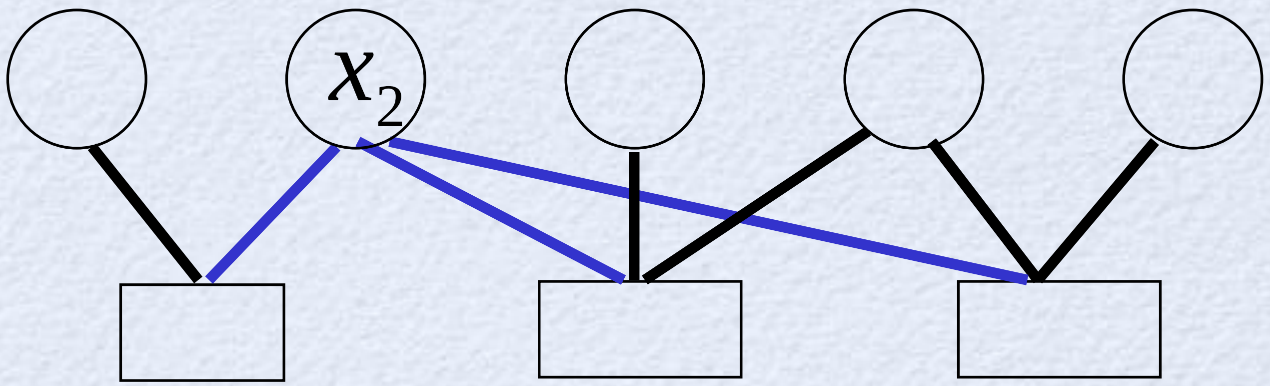


BP - Variational Methods

Bethe free energy: Approximating the marginals $b_\alpha(x_\alpha), b_i(x_i)$

$$\min_{b_\alpha, b_i} \quad -\sum_\alpha \sum_{x_\alpha} b_\alpha(x_\alpha) \ln \psi_\alpha(x_\alpha) - \sum_\alpha H(b_\alpha) - \sum_i (1-d_i) H(b_i)$$

$$d_2 = 3$$



- Stationary Bethe free energy = BP fixed points (Yedidia, Freeman, Weiss ’01)
- When the factor graph has no cycles Bethe is convex over the marginalization constraints $b_i(x_i) = \sum_{x_\alpha \setminus x_i} b_\alpha(x_\alpha)$ and BP is exact.
- Factor graph has cycles: Bethe is non-convex and BP might not converge.

BP - Variational Methods

TRW free energy (Wainwright, Jaakkola, Willsky ’02):

$$\min \quad -\sum_{\alpha}\sum_{x_\alpha} b_\alpha(x_\alpha)\ln\psi_\alpha(x_\alpha) - \sum_{\alpha}\bar{c}_\alpha H(b_\alpha) - \sum_{i}\bar{c}_i H(b_i)$$

$\bar{c}_\alpha =$ Weighted number of spanning trees through an edge α

$$\bar{c}_i = 1 - \sum_{\alpha\in N(i)}\bar{c}_\alpha$$

- TRW free energy is convex over the marginalization constraints.
- Convergent message passing algorithm for cliques of size 2 (Globerson & Jaakkola, ’07)

Convex Free Energies

$$\min \quad -\sum_\alpha \sum_{x_\alpha} b_\alpha(x_\alpha) \ln \psi_\alpha(x_\alpha) - \sum_\alpha \bar{c}_\alpha H(b_\alpha) - \sum_i \bar{c}_i H(b_i)$$

Claim: (Pakzad and Anantharam ‘02, Heskes ’04, Weiss et al. ’07)

“convex free energy” is convex (over the marginalization constraints) for all factor graphs if there exists $c_\alpha, c_i, c_{i\alpha} \geq 0$ such that

$$\bar{c}_\alpha = c_\alpha + \sum_{i \in N(\alpha)} c_{i\alpha} \quad \text{and} \quad \bar{c}_i = c_i - \sum_{\alpha \in N(i)} c_{i\alpha}$$

Convex Free Energies

$$\min \quad -\sum_\alpha \sum_{x_\alpha} b_\alpha(x_\alpha) \ln \psi_\alpha(x_\alpha) - \sum_\alpha \bar{c}_\alpha H(b_\alpha) - \sum_i \bar{c}_i H(b_i)$$

Claim: (Pakzad and Anantharam ‘02, Heskes ’04, Weiss et al. ’06)

“convex free energy” is convex (over the marginalization constraints) for all factor graphs if there exists $c_\alpha, c_i, c_{i\alpha} \geq 0$ such that

$$\bar{c}_\alpha = c_\alpha + \sum_{i \in N(\alpha)} c_{i\alpha} \quad \text{and} \quad \bar{c}_i = c_i - \sum_{\alpha \in N(i)} c_{i\alpha}$$

$$\Downarrow$$

$$\min \quad -\sum_\alpha \sum_{x_\alpha} \ln \psi_a(x_a) b_\alpha(x_\alpha) - \sum_\alpha c_\alpha H(b_\alpha) - \sum_i c_i H(b_i) - \sum_{i, \alpha \in N(i)} c_{i\alpha}(H(b_\alpha) - H(b_i))$$

Background - Summary

Bethe free energy

non-convex for general graphs

↕

Belief Propagation

TRW free energy

specific convexification of free energy

↕

Globerson Jaakkola '07 message passing algorithm for |α|=2

convex free energy

general form of convexification with parameters

$c_\alpha, c_i, c_{i\alpha} \geq 0$

↕

?

message passing algorithm for |α|≥2

Contributions:

1) Convergent message passing algorithm for convex free energy

2) Heuristic for choosing good convex free energy.

Convex Belief Propagation

$$\min_{b \in R^m} f(b) + \sum_{i=1}^{n} h_i(b)$$

strictly convex & differentiable

Strictly convex & proper ($h_i(b) = \infty$ for some b)

Convex Belief Propagation

$$\min_{b \in R^m} f(b) + \sum_{i=1}^{n} h_i(b)$$

$f(b)$: strictly convex & differentiable

$h_i(b)$: Strictly convex & proper ($h_i(b) = \infty$ for some b)

$$\min \left(-\sum_{\alpha}\sum_{x_\alpha} \ln \psi_a(x_a) b_\alpha(x_\alpha) - \sum_{\alpha} c_\alpha H(b_\alpha) \right) + \sum_{i} \left(-c_i H(b_i) - \sum_{\alpha \in N(i)} c_{i\alpha} (H(b_\alpha) - H(b_i)) \right)$$

$\forall$i marginals of x_i agree

Convex Belief Propagation

$$\min_{b \in R^m} f(b) + \sum_{i=1}^{n} h_i(b)$$

$f(b)$: strictly convex & differentiable

$h_i(b)$: Strictly convex & proper ($h_i(b) = \infty$ for some b)

$$\min \left(-\sum_{\alpha} \sum_{x_\alpha} \ln \psi_a(x_a) b_\alpha(x_\alpha) - \sum_{\alpha} c_\alpha H(b_\alpha) \right) + \sum_{i} \left(-c_i H(b_i) - \sum_{\alpha \in N(i)} c_{i\alpha} (H(b_\alpha) - H(b_i)) \right)$$

$\forall$i marginals of x_i agree

$$h_i(b) = \begin{cases} -c_i H(b_i) - \sum c_{i\alpha} (H(b_\alpha) - H(b_i)) & \text{Whenever marginals of } x_i \text{ agree} \\ \infty & \text{Otherwise} \end{cases}$$

Convex Message Passing

$$\min_b \; f(b) + \sum h_i(b)$$

For t=1,2,...

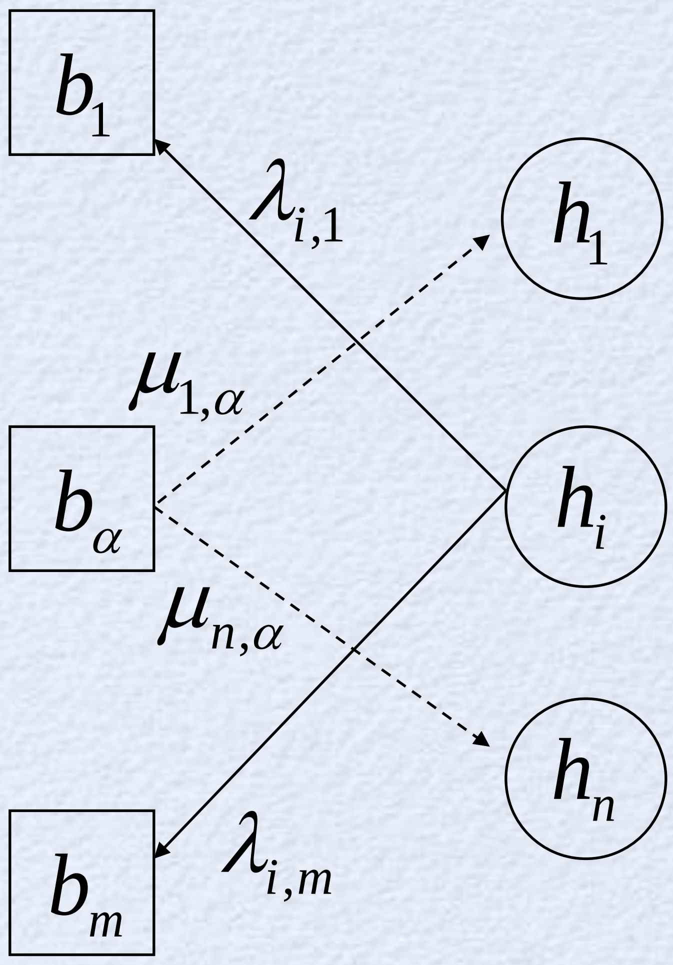
For i=1,2,...,n

$$\mu_i \leftarrow \sum_{j \neq i} \lambda_j$$

$$b^* \leftarrow \underset{b \in domain(h_i)}{\arg\min} \left\{ f(b) + b^T \mu_i + h_i(b) \right\}$$

$$\lambda_i \leftarrow -\mu_i - \nabla f(b^*)$$

Output: b^*



* We also have a similar parallel message passing algorithm

Convex Message Passing

$$\min_b f(b) + \sum h_i(b)$$

For t=1,2,...

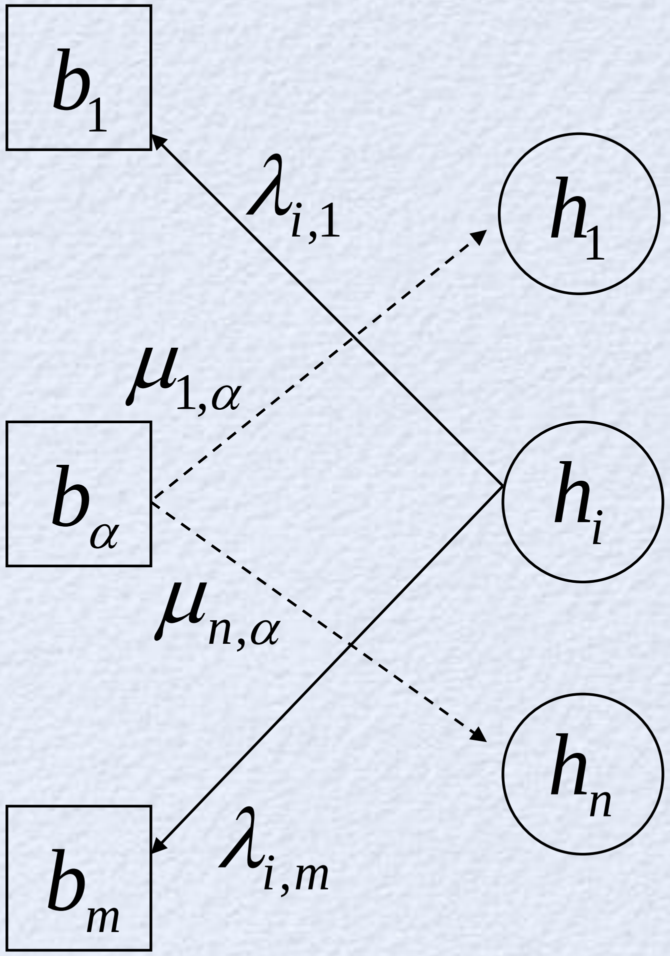
For i=1,2,...,n

$$\mu_i \leftarrow \sum_{j \neq i} \lambda_j$$

$$b^* \leftarrow \arg\min_{b \in domain(h_i)} \left\{ f(b) + b^T \mu_i + h_i(b) \right\}$$

$$\lambda_i \leftarrow -\mu_i - \nabla f(b^*)$$

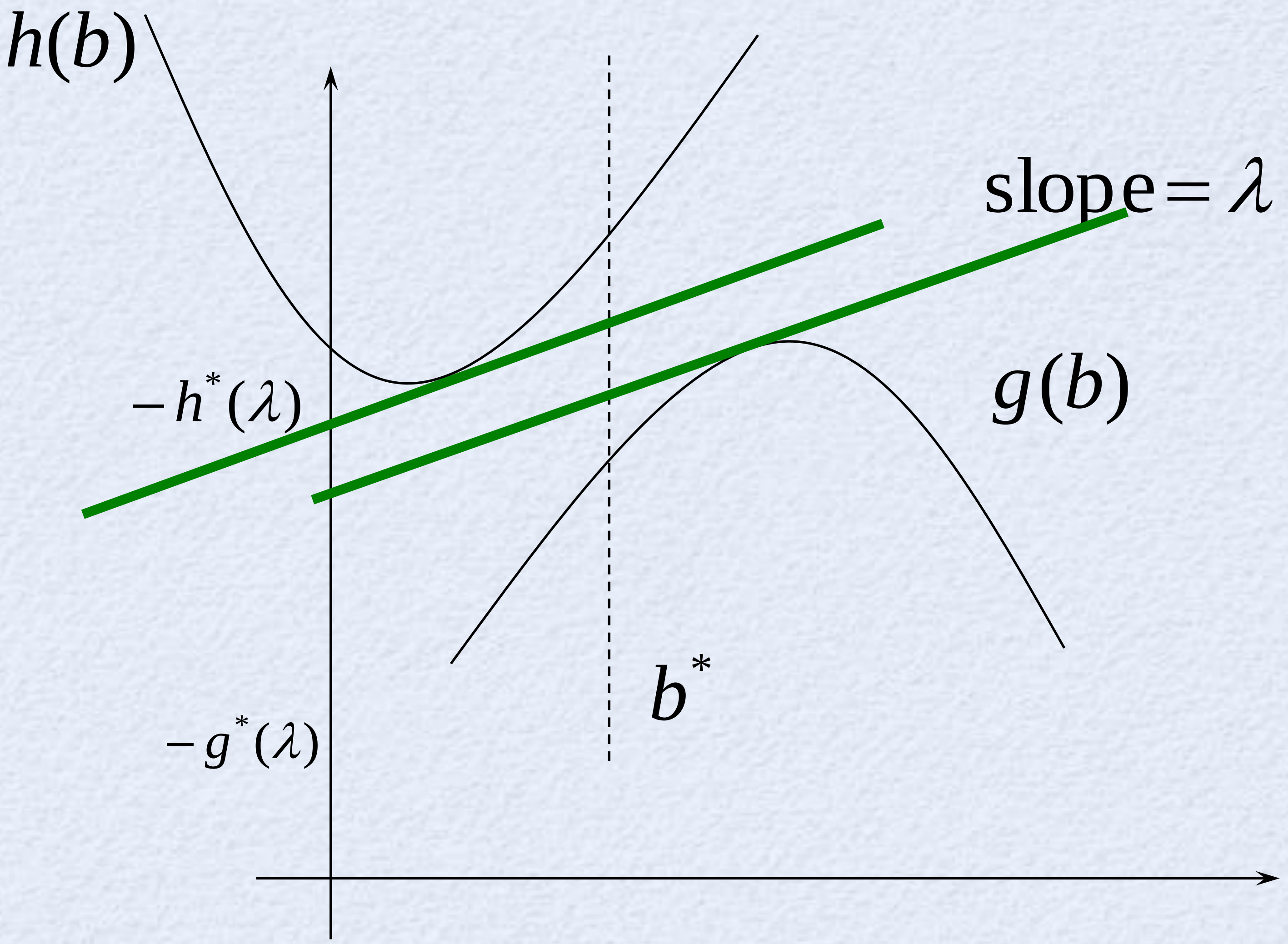
Output: b^*



* We also have a similar parallel message passing algorithm

Primer on Fenchel Duality

Primal $h(b) - g(b)$ **Dual** $g^*(\lambda) - h^*(\lambda)$

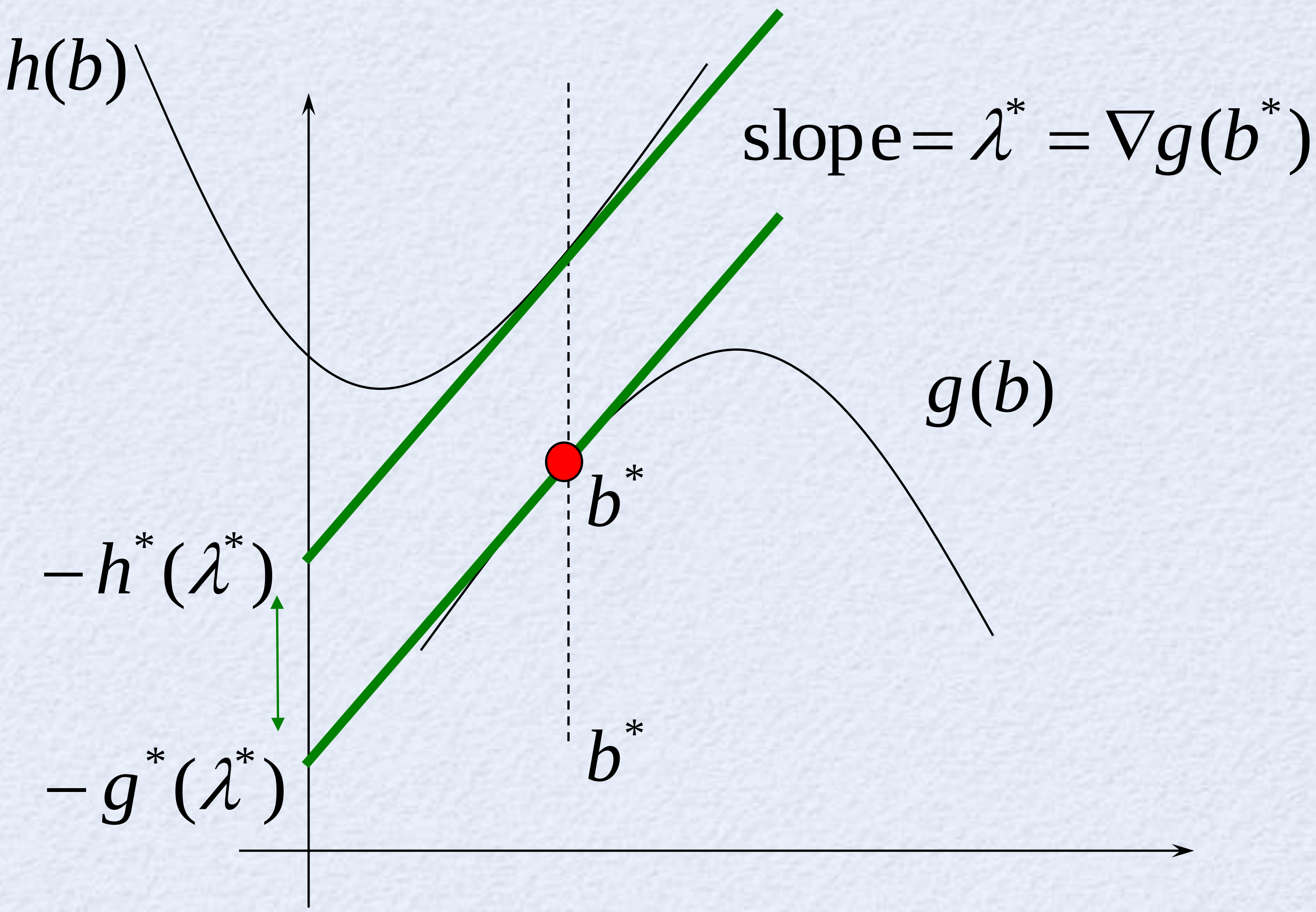


$h^*(\lambda) = \max_b \left\{ \lambda^T x - h(b) \right\}$ convex conjugate of *h(b)*

$g^*(\lambda) = \min_b \left\{ \lambda^T b - g(b) \right\}$ concave conjugate of *g(b)*

Primer on Fenchel Duality

$$\min_b h(b) - g(b) = \max_\lambda g^*(\lambda) - h^*(\lambda), \quad and \quad \lambda^* = \nabla g(b^*)$$



More Conveniently: $\min_b g(b) + h(b) = \max_\lambda \left\{ \min_b \left(g(b) + b^T \lambda \right) - h^*(\lambda) \right\}$

$$\lambda^* = -\nabla g(b^*)$$

Sequential message-passing

Fenchel duality theorem

$$\min_b g(b) + h(b) = \max_\lambda \left\{ \min_b \left(g(b) + b^T \lambda \right) - h^*(\lambda) \right\} \qquad \lambda^* = -\nabla g(b^*)$$

Sequential message-passing

Fenchel duality theorem

$$\min_b g(b) + h(b) = \max_\lambda \left\{ \min_b \left(g(b) + b^T \lambda \right) - h^*(\lambda) \right\} \qquad \lambda^* = -\nabla g(b^*)$$

Generalized Fenchel duality theorem

$$\min f(b) + \sum_{i=1}^n h_i(b) = \max_{\lambda_1,\ldots,\lambda_k} \left\{ \min_b \left[f(b) + b^T (\sum_{i=1}^n \lambda_i) \right] - \sum_{i=1}^n h_i^*(\lambda_i) \right\}$$

Sequential message-passing

Fenchel duality theorem

$$\min_b g(b) + h(b) = \max_\lambda \left\{ \min_b \left(g(b) + b^T \lambda \right) - h^*(\lambda) \right\} \qquad \lambda^* = -\nabla g(b^*)$$

Generalized Fenchel duality theorem

$$\min f(b) + \sum_{i=1}^n h_i(b) = \max_{\lambda_1,\ldots,\lambda_k} \left\{ \min_b \left[f(b) + b^T (\sum_{i=1}^n \lambda_i) \right] - \sum_{i=1}^n h_i^*(\lambda_i) \right\}$$

Block update approach: Optimize λ_i Set $\mu_i \leftarrow \sum_{j \neq i} \lambda_j$

$$\max_{\lambda_i} \left\{ \min_b \left(f(b) + b^T \mu_i + b^T \lambda_i \right) - h_i^*(\lambda_i) \right\}$$

Sequential message-passing

Fenchel duality theorem

$$\min_b g(b) + h(b) = \max_\lambda \left\{ \min_b \left(g(b) + b^T \lambda \right) - h^*(\lambda) \right\} \qquad \lambda^* = -\nabla g(b^*)$$

Generalized Fenchel duality theorem

$$\min f(b) + \sum_{i=1}^n h_i(b) = \max_{\lambda_1,\ldots,\lambda_k} \left\{ \min_b \left[f(b) + b^T \left(\sum_{i=1}^n \lambda_i \right) \right] - \sum_{i=1}^n h_i^*(\lambda_i) \right\}$$

Block update approach: Optimize λ_i Set $\mu_i \leftarrow \sum_{j \neq i} \lambda_j$

$$\max_{\lambda_i} \left\{ \min_b \left(f(b) + b^T \mu_i + b^T \lambda_i \right) - h_i^*(\lambda_i) \right\}$$

Sequential message-passing

Fenchel duality theorem

$$\min_b g(b) + h(b) = \max_\lambda \left\{ \min_b \left(g(b) + b^T \lambda \right) - h^*(\lambda) \right\}$$

$$\lambda^* = -\nabla g(b^*)$$

Generalized Fenchel duality theorem

$$\min f(b) + \sum_{i=1}^n h_i(b) = \max_{\lambda_1,\ldots,\lambda_k} \left\{ \min_b \left[f(b) + b^T \left(\sum_{i=1}^n \lambda_i \right) \right] - \sum_{i=1}^n h_i^*(\lambda_i) \right\}$$

Block update approach: Optimize λ_i Set $\mu_i \leftarrow \sum_{j \neq i} \lambda_j$

$$\max_{\lambda_i} \left\{ \min_b \left(f(b) + b^T \mu_i + b^T \lambda_i \right) - h_i^*(\lambda_i) \right\} = \min_b f(b) + b^T \mu + h_i(b)$$

$$\lambda_i \leftarrow -\nabla f(b^*) - \mu_i$$

Sequential message-passing

Fenchel duality theorem

$$\min_b g(b) + h(b) = \max_\lambda \left\{ \min_b \left(g(b) + b^T \lambda \right) - h^*(\lambda) \right\} \qquad \lambda^* = -\nabla g(b^*)$$

Generalized Fenchel duality theorem

$$\min \; f(b) + \sum_{i=1}^n h_i(b) = \max_{\lambda_1,\ldots,\lambda_k} \left\{ \min_b \left[f(b) + b^T (\sum_{i=1}^n \lambda_i) \right] - \sum_{i=1}^n h_i^*(\lambda_i) \right\}$$

Algorithm: Repeat until convergence

Block update approach: Optimize λ_i Set $\mu_i \leftarrow \sum_{j \neq i} \lambda_j$

$$\max_{\lambda_i} \left\{ \min_b \left(f(b) + b^T \mu_i + b^T \lambda_i \right) - h_i^*(\lambda_i) \right\} = \min_b \; f(b) + b^T \mu + h_i(b)$$

$$\lambda_i \leftarrow -\nabla f(b^*) - \mu_i$$

Output: b^*

Sequential message-passing

$$\min_b \; f(b) + \sum h_i(b)$$

$$\min_{b \in \bigcap C_i} f(b) \quad \Longleftarrow \quad h_i(b) = \begin{cases} 0 & \text{if} \quad b \in C_i \\ \infty & \text{otherwise} \end{cases}$$

Bregman's successive projection algorithm (Hildreth, Dykstra, Csiszar, Tseng)

For t=1,2,...

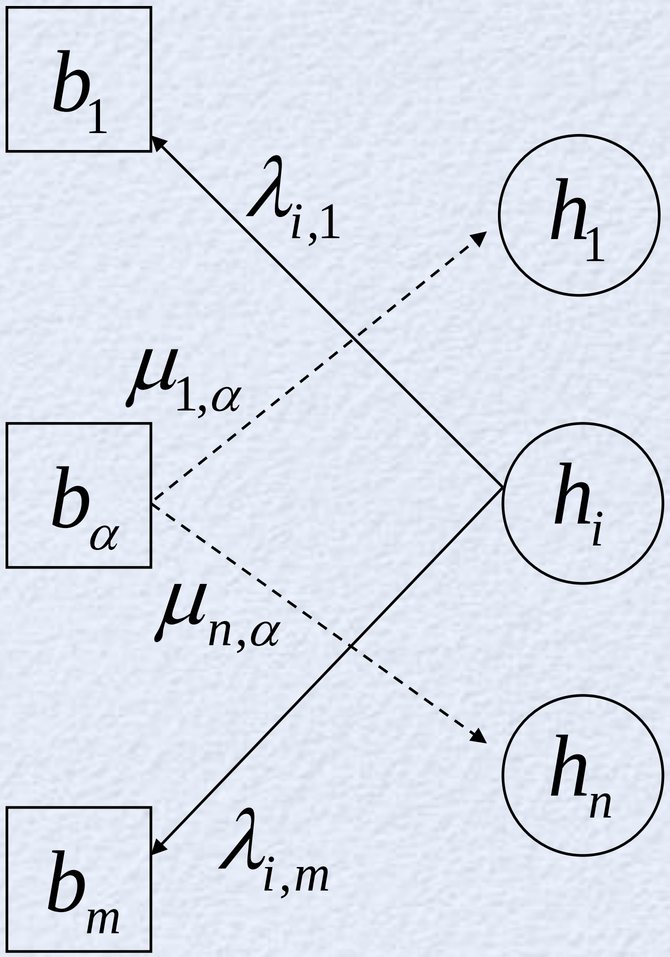
For i=1,2,...,n

$$\mu_i \leftarrow \sum_{j \neq i} \lambda_j$$

$$b^* \leftarrow \underset{b \in domain(h_i)}{\arg\min} \left\{ f(b) + b^T \mu_i + h_i(b) \right\}$$

$$\lambda_i \leftarrow -\mu_i - \nabla f(b^*)$$

Output: b^*



Convex Belief Propagation

$$\min \left(-\sum_{\alpha} \sum_{x_\alpha} \ln \psi_a(x_a) b_\alpha(x_\alpha) - \sum_{\alpha} c_\alpha H(b_\alpha) \right) + \sum_{i} \left(-c_i H(b_i) - \sum_{\alpha \in N(i)} c_{i\alpha} (H(b_\alpha) - H(b_i)) \right)$$

$\forall$i marginals of x_i agree

$$\downarrow$$

$$\min_{b \in R^m} f(b) + \sum_{i=1}^{n} h_i(b)$$

For t=1,2,...

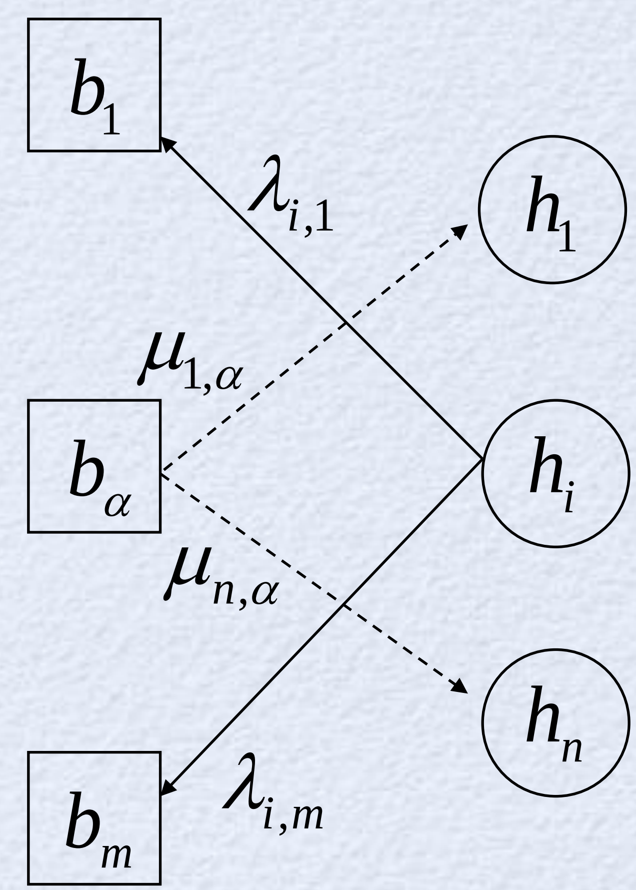
For i=1,2,...,n

$$\mu_i \leftarrow \sum_{j \neq i} \lambda_j$$

$$b^* \leftarrow \underset{b \in domain(h_i)}{\arg\min} \left\{ f(b) + b^T \mu_i + h_i(b) \right\}$$

$$\lambda_i \leftarrow -\mu_i - \nabla f(b^*)$$

Output: b^*



Convex Belief Propagation

Closed form solution for $b^* \leftarrow \underset{b \in domain(h_i)}{\arg\min} \left\{ f(b) + b^T \mu_i + h_i(b) \right\}$

Our algorithm:

$$m_{a \to i}(x_i) := \sum_{\mathbf{x}_a \setminus x_i} \left(\psi_a(\mathbf{x}_a) \prod_{j \in N(a) \setminus i} n_{j \to a}(x_\alpha) \right)^{1/\hat{c}_{i\alpha}}$$

$$b_i(x_i) \propto \prod_{\alpha \in N(i)} m_{\alpha \to i}^{\hat{c}_{i\alpha}/\hat{c}_i}(x_i)$$

$$n_{i \to \alpha}(x_\alpha) = \left(\psi_\alpha(x_\alpha) \prod_{j \in N(\alpha) \setminus i} n_{j \to \alpha}(x_\alpha) \right)^{-\frac{c_{i\alpha}}{\hat{c}_{i\alpha}}} \left(\frac{b_i(x_i)}{m_{\alpha \to i}(x_i)} \right)^{c_\alpha}$$

Belief propagation:

- $m_{a \to i}(x_i) := \sum_{\mathbf{x}_a \setminus x_i} \psi_a(\mathbf{x}_a) \prod_{j \in N(a) \setminus i} n_{j \to a}(x_j)$
- $b_i(x_i) \propto \prod_{a \in N(i)} m_{a \to i}(x_i)$
- $n_{i \to a}(x_i) = \dfrac{b_i(x_i)}{m_{\alpha \to i}(x_i)}$

Convex Belief Propagation

Closed form solution for $b^* \leftarrow \underset{b \in domain(h_i)}{\arg\min} \left\{ f(b) + b^T \mu_i + h_i(b) \right\}$

Our algorithm:

$$m_{a \to i}(x_i) := \sum_{\mathbf{x}_a \setminus x_i} \left(\psi_a(\mathbf{x}_a) \prod_{j \in N(a) \setminus i} n_{j \to a}(x_\alpha) \right)^{1/\hat{c}_{i\alpha}}$$

$$b_i(x_i) \propto \prod_{\alpha \in N(i)} m_{\alpha \to i}^{\hat{c}_{i\alpha} / \hat{c}_i}(x_i)$$

$$n_{i \to \alpha}(x_\alpha) = \left(\psi_\alpha(x_\alpha) \prod_{j \in N(\alpha) \setminus i} n_{j \to \alpha}(x_\alpha) \right)^{-\frac{c_{i\alpha}}{\hat{c}_{i\alpha}}} \left(\frac{b_i(x_i)}{m_{\alpha \to i}(x_i)} \right)^{c_\alpha}$$

Belief propagation:

- $m_{a \to i}(x_i) := \sum_{\mathbf{x}_a \setminus x_i} \psi_a(\mathbf{x}_a) \prod_{j \in N(a) \setminus i} n_{j \to a}(x_j)$
- $n_{i \to a}(x_i) = \frac{b_i(x_i)}{m_{\alpha \to i}(x_i)}$
- $b_i(x_i) \propto \prod_{a \in N(i)} m_{a \to i}(x_i)$

Convex Belief Propagation

Interesting Notes:

1) $\min_b f(b) + \sum_i h_i(b)$ when f(),h_i() are non-convex if the algorithm converges then it is to a stationary point.

2) Set $c_\alpha = 1, c_i = 1 - d_i, c_{i\alpha} = 0,$ (Bethe free energy, non-convex!) our message passing algorithm = belief propagation.

3) For general $c_\alpha, c_i = 1 - \sum c_\alpha,$ and $c_{i\alpha} = 0,$ (non-convex!) our algorithm

- $m_{a\to i}(x_i) := \sum_{\mathbf{x}_a \setminus x_i} \psi_a^{1/c_\alpha}(\mathbf{x}_a) \prod_{j\in N(a)\setminus i} n_{j\to a}(x_j)$
- $n_{i\to a}(x_i) := \prod_{c\in N(i)\setminus a} m_{\beta\to i}^{c_\beta}(x_i)$

Open question =? “Fractional BP” (Wiegerinck, Heskes ‘03)

Determine Convex Energy

$$\min \quad -\sum_{\alpha}\sum_{x_\alpha} \ln \psi_a(x_a) b_\alpha(x_\alpha) - \sum_{\alpha} \bar{c}_\alpha H(b_\alpha) - \sum_{i} \bar{c}_i H(b_i)$$

Heuristic: Find convex energy while $\bar{c}_\alpha$ is close to 1 as possible

$$\bar{c}_i = 1 - \sum_{\alpha \in N(i)} \bar{c}_\alpha$$

Motivation:

$\bar{c}_\alpha = 1$ - Bethe approximation

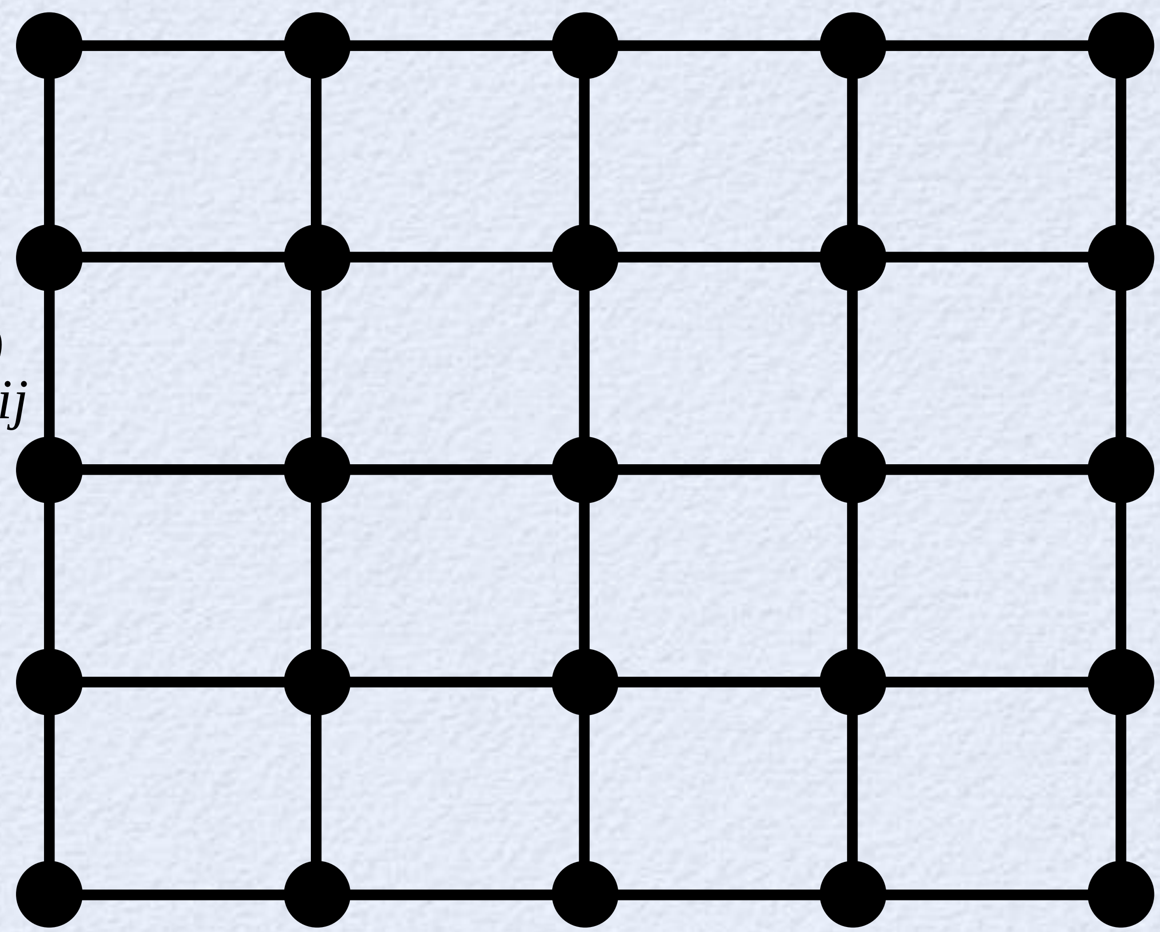
Experiments – Ising Model

$n \times n$ grid $\Longrightarrow$ n^2 variables

Every variable has two values $x_i \in \{-1, 1\}$

field θ_i

interaction θ_{ij}



$\theta_{ij} > 0$ attractive

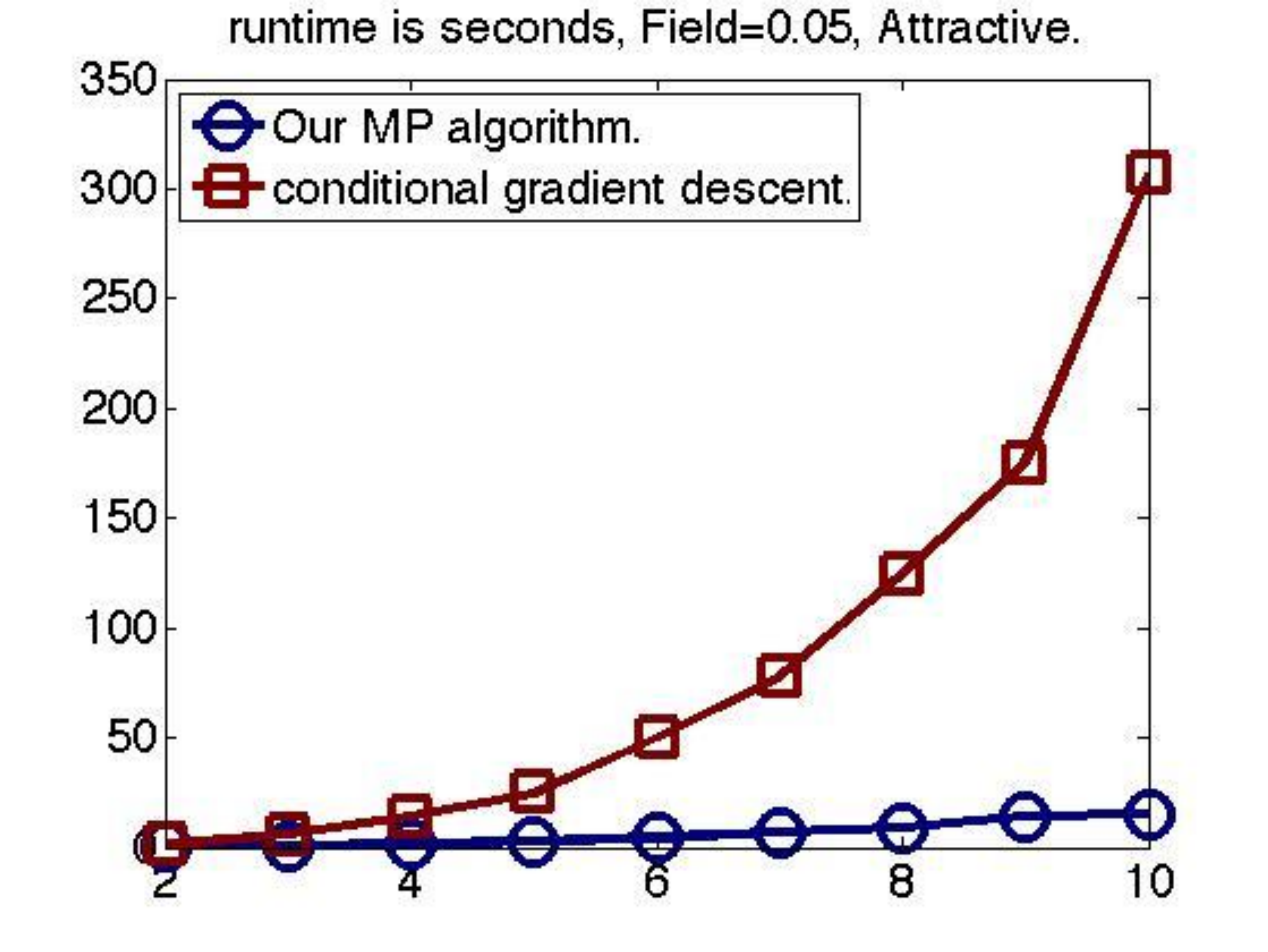
$\theta_{ij} \neq 0$ mixed

$$\Phi_i(x_i) = \exp(\theta_i x_i)$$
$$\psi_{ij}(x_i, x_j) = \exp(\theta_{ij} x_i x_j)$$

$$p(x_1, \ldots, x_n) \propto \prod_{i,j \in E} \psi_{ij}(x_i, x_j) \prod_i \Phi_i(x_i)$$

Experiments – Efficiency

Seconds

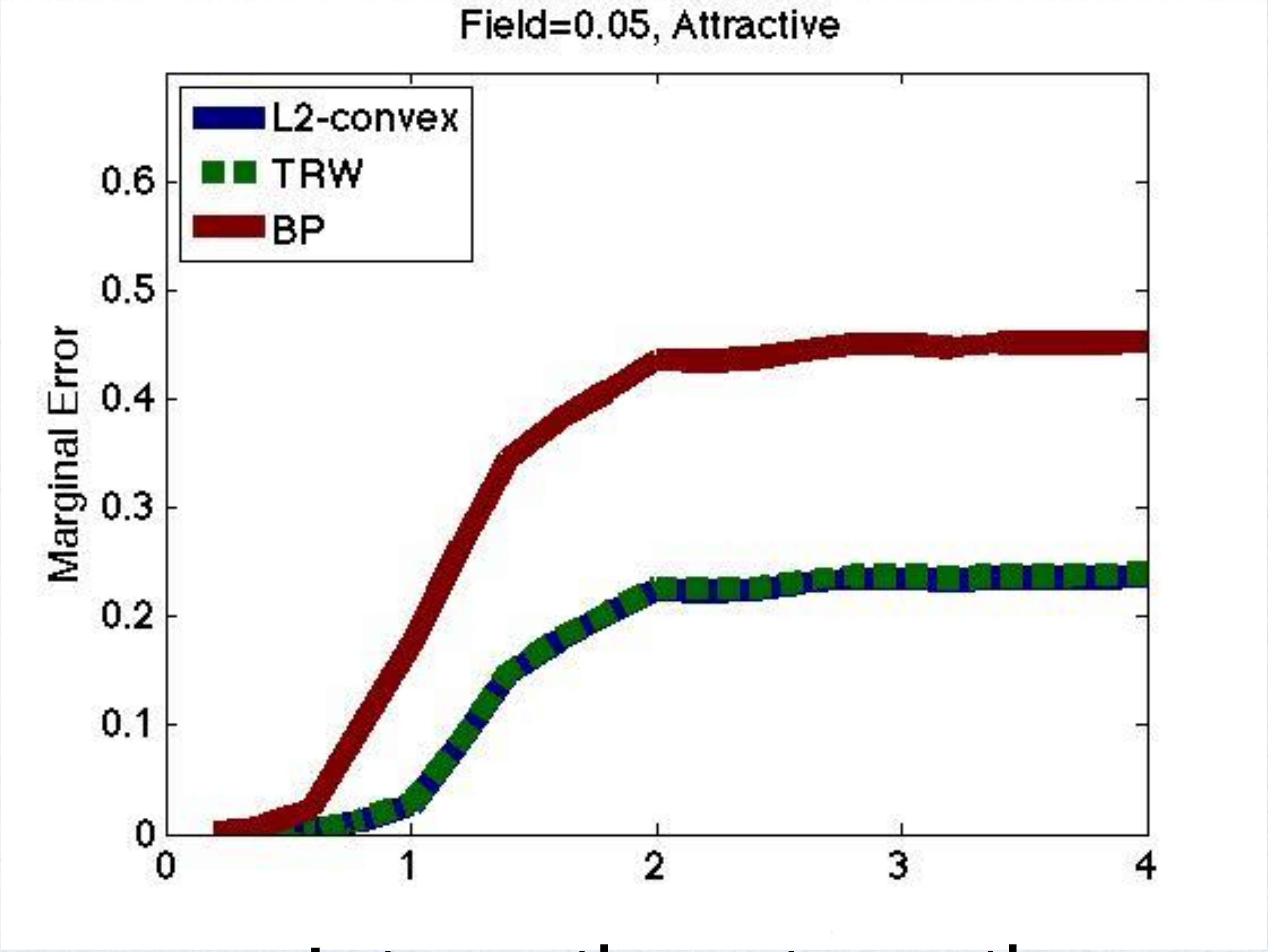


Grid size

Approximating Marginals – Ising model

convex free energies Vs. Bethe free energy

Marginal Error

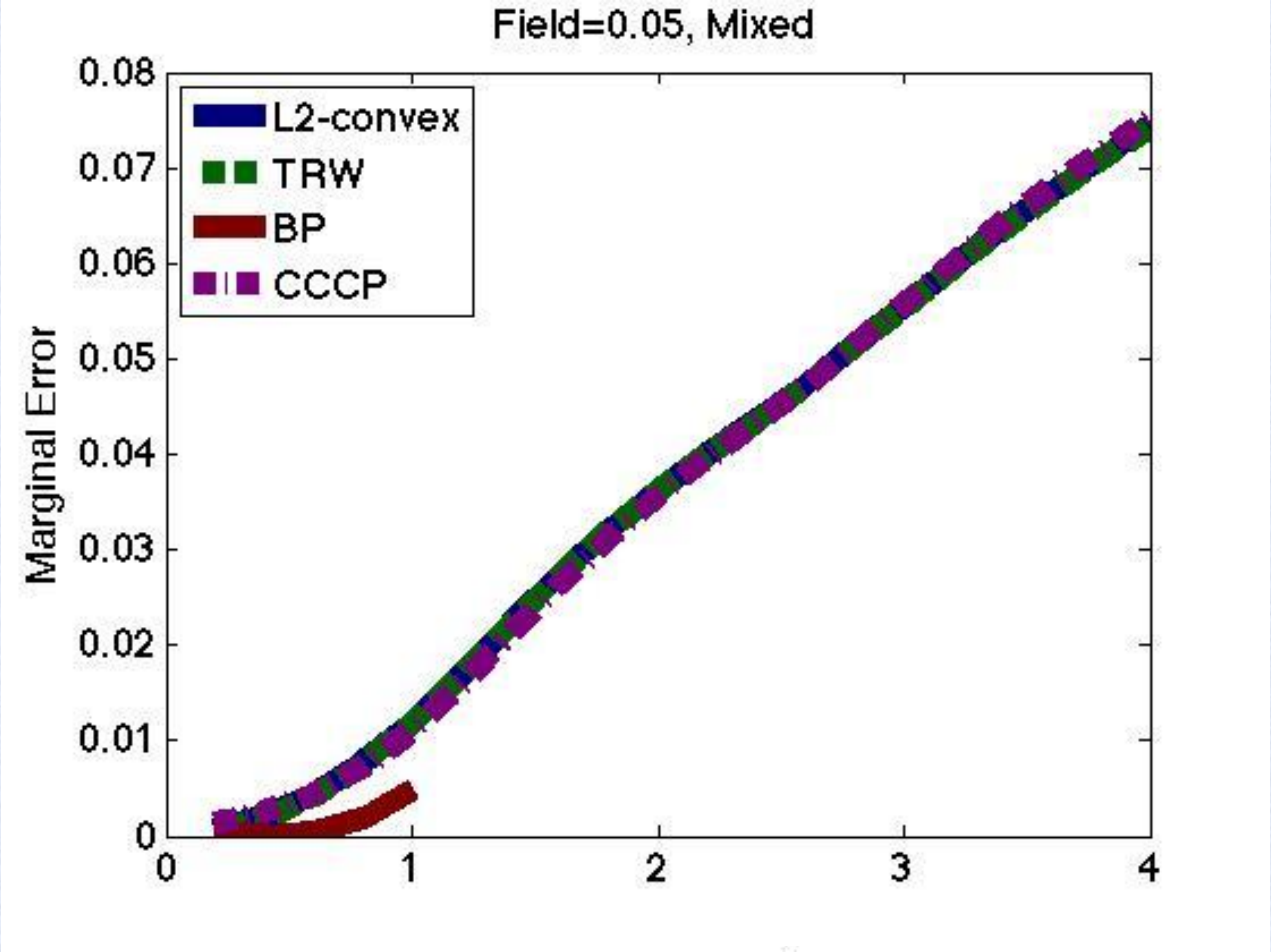


Interaction strength

Approximating Marginals – Ising model

convex free energies Vs. Bethe free energy

Marginal Error

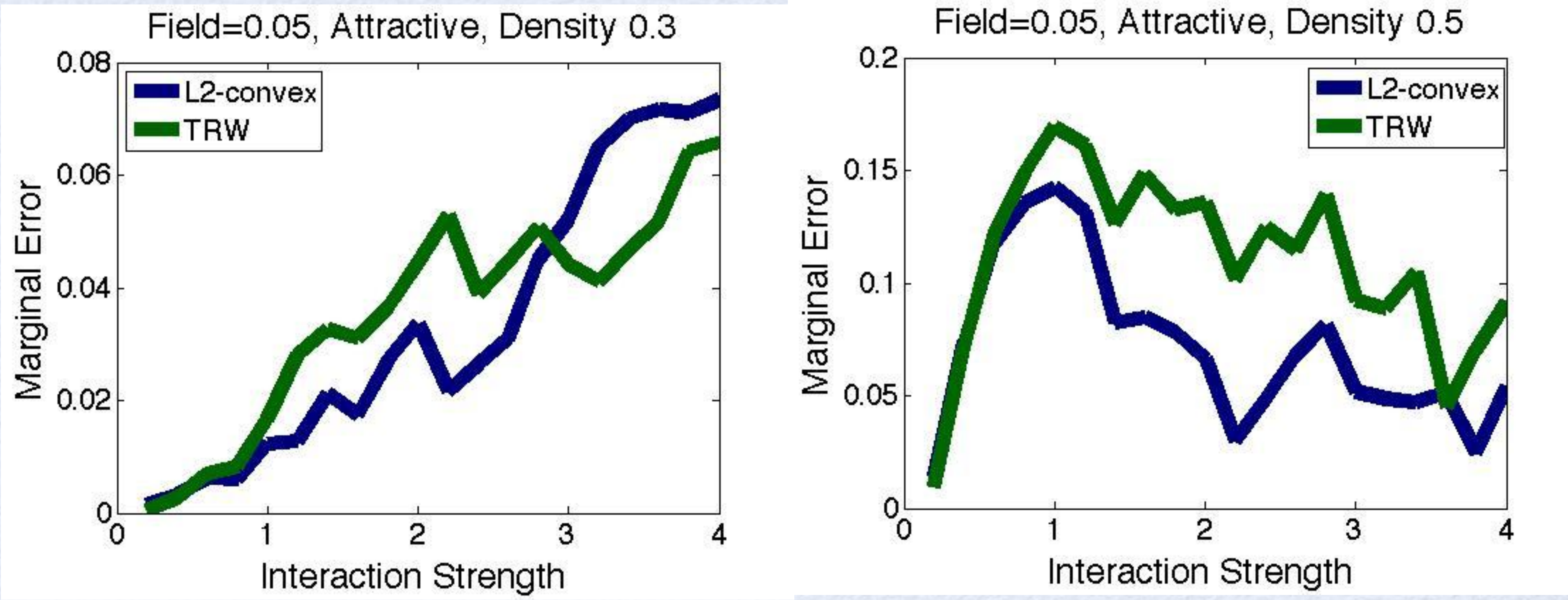


Interaction strength

Approximating Marginals – Random Graph

TRW free energy Vs. L2-convex free energy

10 vertices. Edge density 30% (almost tree), 50% (far from tree)



Summary

- Novel perspective of message passing as primal-dual optimization for the general class of $\min_b f(b) + \sum_i h_i(b)$
- Convex free energy mapped to $\min_b f(b) + \sum_i h_i(b)$
- Heuristic for setting convex free energy by “convexifying” Bethe free energy
- Algorithm applies to general clique sizes; BP is a particular case of our algorithm
- Region Graphs energies (Kikuchi)
- Future work: Theoretical foundation on our heuristic for setting the convex free energy.