# Convergent Message-Passing Algorithms for LP-relaxations and Approximated Inference

**Tamir Hazan**

School of Computer Science and Engineering
Hebrew University of Jerusalem

Joint work: **Amnon Shashua**

# Markov Random Fields

$$p(x_1, \ldots, x_n) \propto \prod_\alpha \psi_\alpha(x_\alpha)$$

$$x_\alpha \subset \{x_1, \ldots, x_n\}$$

The inference task (NP-hard problems in general):

marginal probability

$$p(x_i) = \sum_{x \setminus x_i} p(x_1, \ldots, x_n)$$

Maximum a-posteriori

$$\arg\max_x \; p(x_1, \ldots, x_n)$$

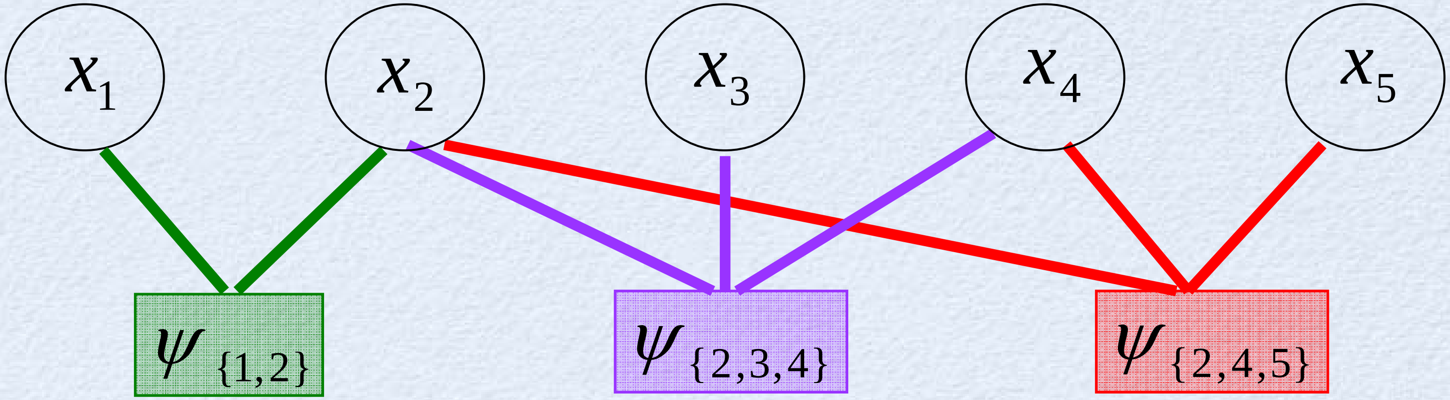
# Applications

- Error Correcting Codes (Gallager 1963, Feldman et. al. 2003)
- Medical Diagnosis (Jaakkola, Jordan 1999)
- Super Resolution (Freeman et. al. 2000)
- Stereo vision (Tappen, Freeman 2003, Meltzer et. al. 2005)
- Protein Folding (Yanover, Weiss 2003)
- Clustering (Shental et. al. 2003)

Image editing (Cho et. al. 2008)

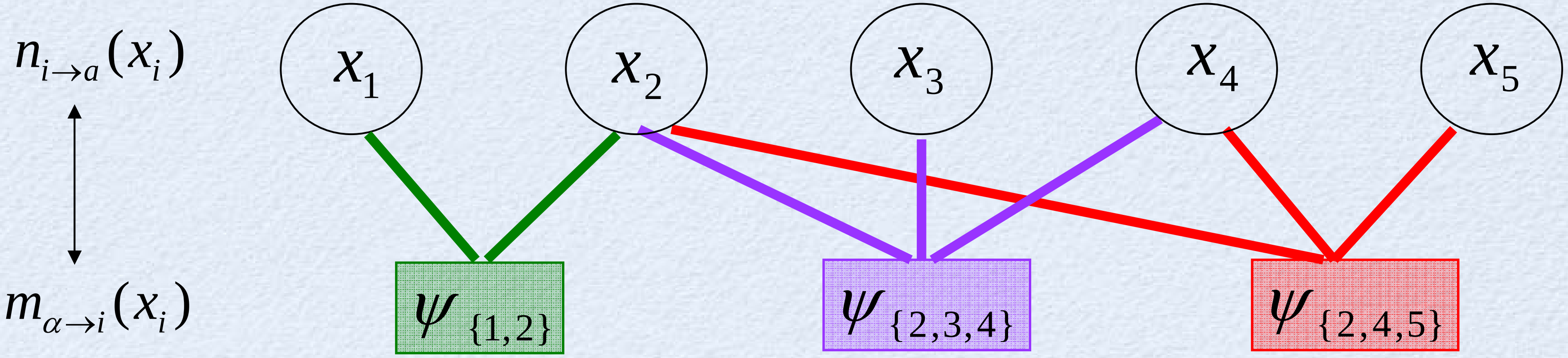
# Graphical Models - Background

$$p(x_1,\ldots,x_n) \propto \prod \psi_\alpha(x_\alpha)$$



# Graphical Models - Background

$$p(x_1,\ldots,x_n) \propto \prod \psi_\alpha(x_\alpha) \longrightarrow \sum_{x\setminus x_i} p(x) = ?$$



**Belief Propagation:**

- $n_{i\to a}(x_i) = \prod_{c\in N(i)\setminus a} m_{c\to i}(x_i)$

sum-product

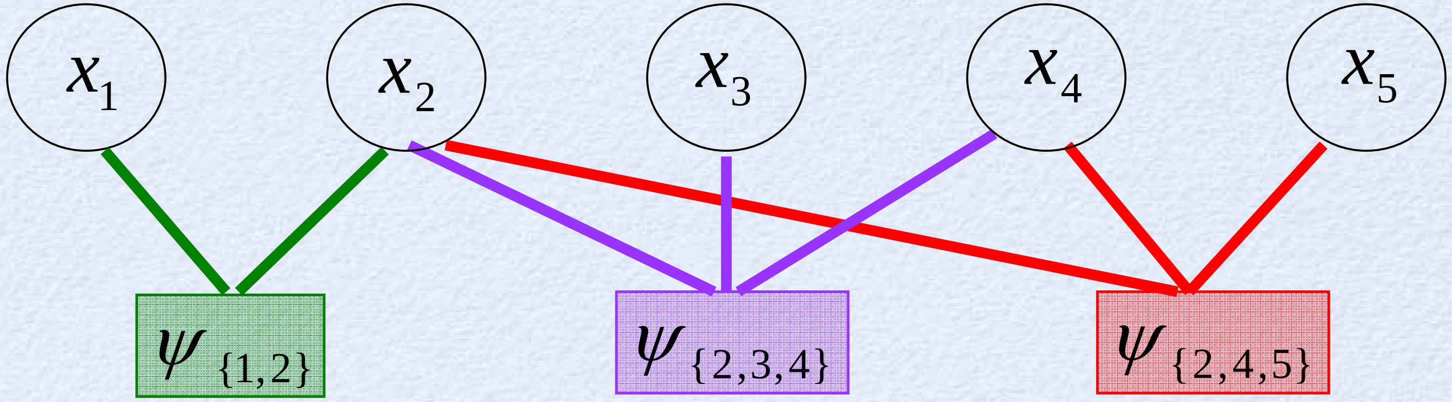
- $m_{a\to i}(x_i) = \sum_{\mathbf{x}_a\setminus x_i} \psi_a(\mathbf{x}_a) \prod_{j\in N(a)\setminus i} n_{j\to a}(x_j)$

# Graphical Models - Background

$$p(x_1,\ldots,x_n) \propto \prod \psi_\alpha(x_\alpha)$$

$$\sum_{x\setminus x_i} p(x) = ?$$

$$\arg\max p(x)$$



**Belief Propagation:**

- $n_{i\to a}(x_i) = \prod_{c\in N(i)\setminus a} m_{c\to i}(x_i)$

sum-product

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

max-product

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

# Variational Methods

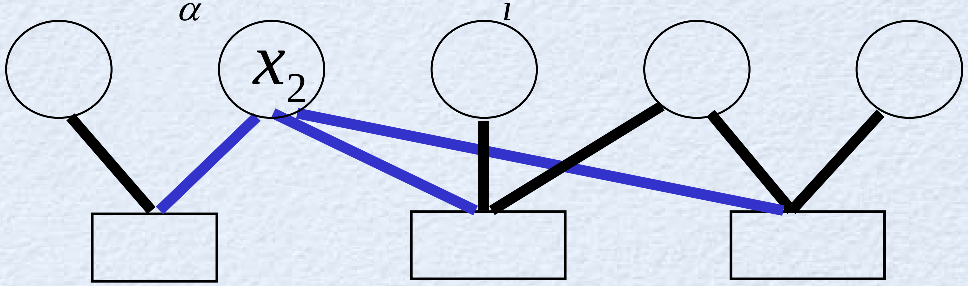
**Approximating marginal probabilities:** $b_\alpha(x_\alpha), b_i(x_i)$

For notational convenience $\psi_\alpha(x_\alpha) = \exp(-\theta_\alpha(x_\alpha))$

**Bethe free energy**

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

$$d_2 = 3$$



# Variational Methods

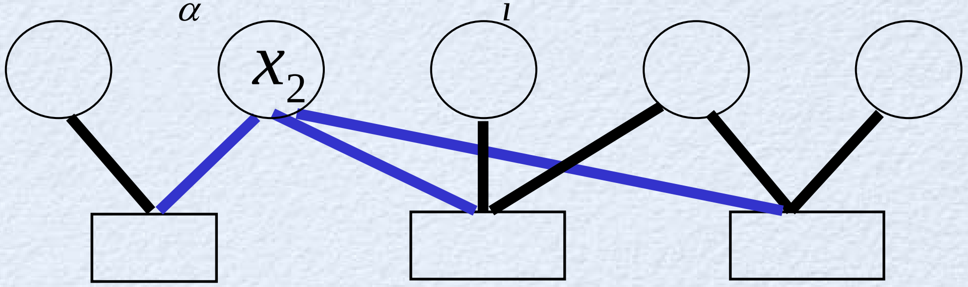
**Approximating marginal probabilities:** $b_\alpha(x_\alpha), b_i(x_i)$

For notational convenience $\psi_\alpha(x_\alpha) = \exp\left(-\theta_\alpha(x_\alpha)\right)$

**Bethe free energy**

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

$$d_2 = 3$$



- Stationary Bethe free energy = sum-product fixed points (Yedidia et.al ’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 sum-product is exact.
- Factor graph has cycles: Bethe is non-convex and alg might not converge.

# Variational Methods

**Approximating marginal probabilities:**

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

$$\min \quad \sum_{\alpha} \sum_{x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \sum_{\alpha} c_\alpha H(b_\alpha) - \sum_{i} c_i H(b_i)$$

$c_\alpha =$ Weighted number of spanning trees through an edge α

$$c_i = 1 - \sum_{\alpha \in N(i)} c_\alpha$$

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

# Variational Methods

## Approximating marginal probabilities:

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

$$\min \quad \sum_{\alpha} \sum_{x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \sum_{\alpha} c_\alpha H(b_\alpha) - \sum_{i} c_i H(b_i)$$

$c_\alpha =$ Weighted number of spanning trees through an edge α

$$c_i = 1 - \sum_{\alpha \in N(i)} c_\alpha$$

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

**Approximated free energy**

$$\min \quad \sum_{\alpha} \sum_{x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \sum_{\alpha} c_\alpha H(b_\alpha) - \sum_{i} c_i H(b_i)$$

# Variational Methods

**Maximum a-Poteriori (MAP):** $p(x) \propto \exp\left(\sum \theta_\alpha(x_\alpha)\right)$

$$\arg\max_{x}\; p(x) = \arg\min_{x} \left\{\sum_\alpha \theta_\alpha(x_\alpha)\right\}$$

# Variational Methods

**MAP:**

$$\arg\max_{x}\; p(x) = \arg\min_{x} \left\{ \sum_{\alpha} \theta_{\alpha}(x_{\alpha}) \right\}$$

**Integer Program:**

$$\arg\min_{b_{\alpha} \in \{0,1\},\, \sum b_{\alpha}(x_{\alpha})=1} \; \sum_{\alpha, x_{\alpha}} b_{\alpha}(x_{\alpha}) \theta_{\alpha}(x_{\alpha}) \quad \text{s.t.} \quad \sum_{x_{\alpha} \setminus x_i} b_{\alpha}(x_{\alpha}) = b_i(x_i)$$

# Variational Methods

**MAP:**

$$\arg\max_{x}\ p(x) = \arg\min_{x} \left\{ \sum\nolimits_{\alpha} \theta_{\alpha}(x_{\alpha}) \right\}$$

**Integer Program:**

$$\arg\min_{b_{\alpha}\in\{0,1\},\ \sum b_{\alpha}(x_{\alpha})=1} \sum_{\alpha,x_{\alpha}} b_{\alpha}(x_{\alpha})\theta_{\alpha}(x_{\alpha}) \quad \text{s.t.} \quad \sum_{x_{\alpha}\setminus x_i} b_{\alpha}(x_{\alpha}) = b_i(x_i)$$

**Linear Program Relaxation (Wainwright, Jaakkola, Willsky '05):**

$$\arg\min_{b_{\alpha}\geq 0,\ \sum b_{\alpha}(x_{\alpha})=1} \sum_{\alpha,x_{\alpha}} b_{\alpha}(x_{\alpha})\theta_{\alpha}(x_{\alpha}) \quad \text{s.t.} \quad \sum_{x_{\alpha}\setminus x_i} b_{\alpha}(x_{\alpha}) = b_i(x_i)$$

# Related Work

$$\arg\min_{b_\alpha \geq 0, \sum b_\alpha(x_\alpha)=1} \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) \quad \text{s.t.} \quad \sum_{x_\alpha \setminus x_i} b_\alpha(x_\alpha) = b_i(x_i)$$

- Dual block ascent (Globerson, Jaakkola, 2007)
- Proximal minimizations (Ravikumar, Agrawal, Wainwright, 2008)

$$b^{(t)} = \arg\min \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) + D_f(b \| b^{(t-1)})$$

- Perturbation methods (Weiss, Yanover, Meltzer, 2007)

$$\arg\min \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\tilde{H}$$

approximated free energy –

message-passing algorithms: sum-product (ε=1)

# Related Work

$$\arg\min_{b_\alpha \geq 0, \sum b_\alpha(x_\alpha)=1} \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\tilde{H} \quad \text{s.t.} \quad \sum_{x_\alpha \setminus x_i} b_\alpha(x_\alpha) = b_i(x_i)$$

**Advantages:**

For ε→0 the argument b* is optimal (Mangasarian ’79)

If $-\tilde{H}$ is strictly convex there is a unique optimum

# Related Work

$$\underset{b_\alpha \geq 0, \sum b_\alpha(x_\alpha)=1}{\arg\min} \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\tilde{H} \quad \text{s.t.} \quad \sum_{x_\alpha \setminus x_i} b_\alpha(x_\alpha) = b_i(x_i)$$

**Advantages:**

For ε→0 the argument b* is optimal (Mangasarian ’79)

If $-\tilde{H}$ is strictly convex there is a unique optimum

**Disadvantages:** $\varepsilon \cdot \arg\min \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \frac{\theta_\alpha(x_\alpha)}{\varepsilon} - \tilde{H}$

Numerical instability

Message-passing algorithms with potentials $\psi_\alpha(x_\alpha)^{1/\varepsilon}$

Undefined for ε=0

# Related Work

$$\arg\min_{b_\alpha \ge 0, \sum b_\alpha(x_\alpha)=1} \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\tilde{H} \quad \text{s.t.} \quad \sum_{x_\alpha \setminus x_i} b_\alpha(x_\alpha) = b_i(x_i)$$

**Advantages:**

For ε→0 the argument b* is optimal (Mangasarian ’79)

If $-\tilde{H}$ is strictly convex there is a unique optimum

**Disadvantages:** $\varepsilon \cdot \arg\min \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\frac{\theta_\alpha(x_\alpha)}{\varepsilon} - \tilde{H}$

Numerical instability

Message-passing algorithms with potentials $\psi_\alpha(x_\alpha)^{1/\varepsilon}$

Undefined for ε=0

**Weiss et. al :** Use (convex) Max-Product. No convergence guarantees.

# Our work

marginal probability

$$\min \sum_{\alpha,x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \tilde{H}$$

Maximum a-posteriori

$$\min \sum_{\alpha,x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha)$$

ε=1

perturbation

$$\min \sum_{\alpha,x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\tilde{H}$$

# Our work

marginal probability

$$\min \sum_{\alpha,x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \tilde{H}$$

Maximum a-posteriori

$$\min \sum_{\alpha,x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha)$$

ε=1

perturbation

**Contributions:**

$$\min \sum_{\alpha,x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\tilde{H}$$

- Message-passing algorithms with (1/ε)-norm. If $-\tilde{H}$ is convex: attains the optimum
- For ε→0 the solution approaches the optimal solution of LP-relaxation

# Our work

marginal probability

$$\min \sum_{\alpha,x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \tilde{H}$$

Maximum a-posteriori

$$\min \sum_{\alpha,x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha)$$

ε=1

perturbation

**Contributions:**

$$\min \sum_{\alpha,x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\tilde{H}$$

ε=1

For ε=1 it is a sum-product type. If $-\tilde{H}$ is convex it recovers the optimum.
If $\tilde{H}$ = Bethe entropy → sum-product alg

- Message-passing algorithms with (1/ε)-norm. If $\tilde{H}$ is convex: attains the optimum
- For ε→0 the solution approaches the optimal solution of LP-relaxation

ε=0

For ε=0 it is a max-product type. For some convex $-\tilde{H}$ the algorithm converges.
If $\tilde{H}$ = Bethe entropy → max-product alg

# Convex Message Passing

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

f(),$h_i$() are convex and proper (can obtain the value $\infty$)

f() is strictly convex and $h_i$() are continuous (in their domain)

```
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: $b^*$
```

Related work: Von-Neumann, Hildreth, Bregman, Csiszar, Dykstra, Han, Tseng

For formal derivation using Fenchel duality: Hazan-Shashua UAI08.

# Convex Message Passing

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

f(),h$_i$() are convex and proper (can get the value of infinity)

f() is strictly convex and h$_i$() are continuous (in their domain)

```
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: $b^*$
```

Analytic solution for approximated free enrgies (efficient)

# Non-Convex Message Passing

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

**Claim:** Assume f() has invertible derivative and $h_i$() is differentiable over its affine domain. Then if the algorithm converges it reaches a stationary point of the primal program

```
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: $b^*$
```

# Convex Belief Propagation

$$\min_{b \in R^m} \underbrace{f(b)}_{\text{strictly convex}} + \sum_{i=1}^{n} \underbrace{h_i(b)}_{\text{convex \& proper ( } h_i(b) = \infty \text{ for some b)}}$$

Hazan-Shashua UAI08

? ↑

$$\arg\min \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon \tilde{H}$$

$$\text{s.t.} \quad \sum_{x_\alpha} b_\alpha(x_\alpha) = 1, \quad \sum_{x_\alpha \setminus x_i} b_\alpha(x_\alpha) = b_i(x_i)$$

# Convex Belief Propagation

$$\min_{b\in R^m} \underbrace{f(b)}_{\text{strictly convex}} + \sum_{i=1}^{n} \underbrace{h_i(b)}_{\text{convex \& proper ( } h_i(b)=\infty \text{ for some b)}}$$

? ↑

$$\arg\min \sum_{\alpha,x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\left(\sum_\alpha c_\alpha H(b_\alpha) + \sum_i c_i H(b_i) + \sum_{i,\alpha\in N(i)} c_{i\alpha}\left(H(b_\alpha) - H(b_i)\right)\right)$$

$\forall i$ marginals of $x_i$ agree

Convex if $c_\alpha, c_i, c_{i\alpha} \geq 0$ (Pakzad and Anantharam ‘02, Heskes ’04, Weiss et al. ’07)

# Convex Belief Propagation

$$\min_{b\in R^m} \underbrace{f(b)}_{\text{strictly convex}} + \sum_{i=1}^{n} \underbrace{h_i(b)}_{\text{convex \& proper ( } h_i(b)=\infty \text{ for some b)}}$$

$$\arg\min \sum_{\alpha,x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\sum_{\alpha} c_\alpha H(b_\alpha) + \sum_i \left( -\varepsilon\, c_i H(b_i) - \sum_{\alpha\in N(i)} \varepsilon\, c_{i\alpha}(H(b_\alpha) - H(b_i)) \right)$$

$\forall$i marginals of $x_i$ agree

Convex if $c_\alpha, c_i, c_{i\alpha} \geq 0$ (Pakzad and Anantharam ‘02, Heskes ’04, Weiss et al. ’07)

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

# Convex Belief Propagation

**After mapping the algorithm takes the form**

$$\boxed{\begin{aligned}&\mu_i \leftarrow \sum\nolimits_{j \neq i} \lambda_j\\ &b^* \leftarrow \arg\min \left\{ f(b) + b^T \mu_i + h_i(b) \right\}\\ &\lambda_i \leftarrow -\mu_i - \nabla f(b^*)\end{aligned}}$$

For i=1,...,n

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

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

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

For ε→0 Numerically unstable! Same situation as Weiss et. al.

# Convex Belief Propagation

**Coping with numerical instability:**

**Change variables** $b_i(x_i) \leftarrow b_i^{\varepsilon}(x_i), \quad m_{i\to\alpha}(x_i) \leftarrow m_{i\to\alpha}^{\varepsilon}(x_i)$

For i=1,...,n

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

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

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

$$\|z\|_{1/\varepsilon} = \left(\sum_j z_i^{1/\varepsilon}\right)^{\varepsilon}$$ for z≥0

$$\left(\frac{b_i^{\varepsilon}(x_i)}{m^{\varepsilon}_{\alpha\to i}(x_i)}\right)$$

# Norm-Product Propagation

**The Norm-Product algorithm:**

For i=1,...,n

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

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

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

The (1/ε)-norm is bounded for every ε≥1, and well-defined for ε=0 ( $\|z\|_\infty=\max|z_i|$ ). Stability depends on the accuracy of norm computation (in Matlab ε can be as low as$10^{-15}$)

# Norm-Product Propagation

**The Norm-Product algorithm:**

$$\arg\min \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\tilde{H}$$

- Message-passing algorithms with (1/ε)-norm. If $-\tilde{H}$ is convex: attains the optimum

**To be shown:**

- For ε=1 it is a sum-product type. If $-\tilde{H}$ is convex it recovers the optimum. If $\tilde{H}$ = Bethe entropy → sum-product algorithm
- For ε=0 it is a max-product type. For some convex $-\tilde{H}$ there are convergence guarantees. If $\tilde{H}$ = Bethe entropy → max-product algorithm
- For ε→0 and convex $-\tilde{H}$ approaches the optimal solution of LP-relaxation

# Norm-Product Propagation

**The Norm-Product algorithm without conditional entropies:**

$$\arg\min \quad \sum_{\alpha,x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\left(\sum_\alpha c_\alpha H(b_\alpha) - \sum_i c_i H(b_i)\right)$$

For i=1,...,n

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

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

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

# Norm-Product Propagation

**The Norm-Product algorithm without conditional entropies:**

$$\arg\min \quad \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\left(\sum_\alpha c_\alpha H(b_\alpha) - \sum_i c_i H(b_i)\right)$$

For i=1,...,n

$$\forall \alpha \in N(i) \qquad m_{a\to i}(x_i) = \left\| \psi_a^{1/c_\alpha}(\mathbf{x}_a) \prod_{j \in N(a)\setminus i} n_{j\to a}^{1/c_\alpha}(x_j) \right\|_{1/\varepsilon}$$

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

$$\forall \alpha \in N(i) \qquad n_{i\to\alpha}(x_i) = \left(\frac{b_i(x_i)}{m_{\alpha\to i}(x_i)}\right)^{c_\alpha}$$

# Norm-Product Propagation

**The Norm-Product algorithm with Bethe entropy (powers = 1):**

$$\arg\min \quad \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\left(\sum_\alpha H(b_\alpha) - \sum_i (1 - d_i)H(b_i)\right)$$

For i=1,...,n

$$\forall \alpha \in N(i) \qquad m_{a \to i}(x_i) = \left\| \psi_a(\mathbf{x}_a) \prod_{j \in N(a) \setminus i} n_{j \to a}(x_j) \right\|_{1/\varepsilon}$$

For ε=1 sum-product
For ε=0 max-product

$$b_i(x_i) \propto \prod_{\alpha \in N(i)} m_{\alpha \to i}(x_i)$$

$$\forall \alpha \in N(i) \qquad n_{i \to \alpha}(x_i) = \frac{b_i(x_i)}{m_{\alpha \to i}(x_i)}$$

# Max-Product Propagation

**The Norm-Product algorithm with ε=0 (max-product type) and without conditional entropies:**

$$\arg\min \sum_{\alpha,x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\left(\sum_\alpha c_\alpha H(b_\alpha) - \sum_i c_i H(b_i)\right)$$

For i=1,...,n

$$\forall \alpha \in N(i) \qquad m_{a\to i}(x_i) = \max_{x_\alpha \setminus x_i} \psi_a^{1/c_\alpha}(\mathbf{x}_a) \prod_{j\in N(a)\setminus i} n_{j\to a}^{1/c_\alpha}(x_j)$$

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

$$\forall \alpha \in N(i) \qquad n_{i\to\alpha}(x_i) \propto \left(\frac{b_i(x_i)}{m_{\alpha\to i}(x_i)}\right)^{c_\alpha}$$

**<u>Theorem:</u>** If $c_i$, $c_\alpha$>0 the algorithms performs dual block ascent on the LP-dual, therefore converges (not necessarily optimal)

# LP-Relaxation Bound

**Linear Program Relaxation:**

$$b^* = \underset{b_\alpha \geq 0, \sum b_\alpha(x_\alpha)=1}{\arg\min} \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) \quad \text{s.t.} \quad \sum_{x_\alpha \setminus x_i} b_\alpha(x_\alpha) = b_i(x_i)$$

**Convex Norm-Product:**

$$b_\varepsilon = \underset{b_\alpha \geq 0, \sum b_\alpha(x_\alpha)=1}{\arg\min} \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\tilde{H} \quad \text{s.t.} \quad \sum_{x_\alpha \setminus x_i} b_\alpha(x_\alpha) = b_i(x_i)$$

**Claim:** $0 \leq \theta^T b_\varepsilon - \theta^T b^* \leq \varepsilon\left( \sum_\alpha c_\alpha \ln|X_\alpha| + \sum_i c_i \ln|X_i| \right)$

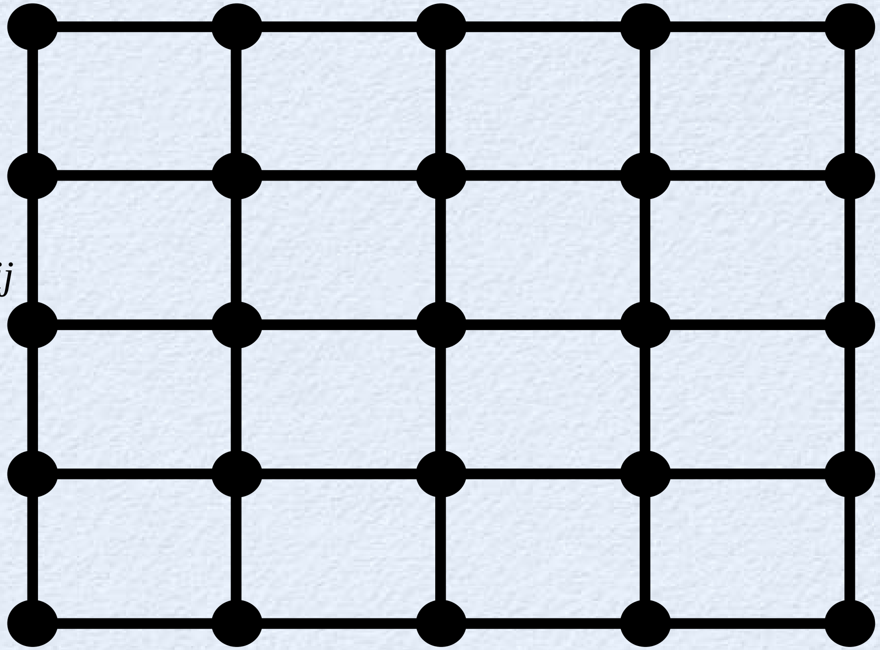
**Illustration:**

Say we want to be 0.01 close to the LP-relaxation of a problem with 100 factors and two possible assignments and $c_\alpha=1$: Set $\varepsilon=10^{-4}$

# Experiments – Ising Model

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

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



field $\theta_i$

interaction $\theta_{ij}$

$\theta_{ij} > 0$ attractive

$\theta_{ij} \neq 0$ mixed

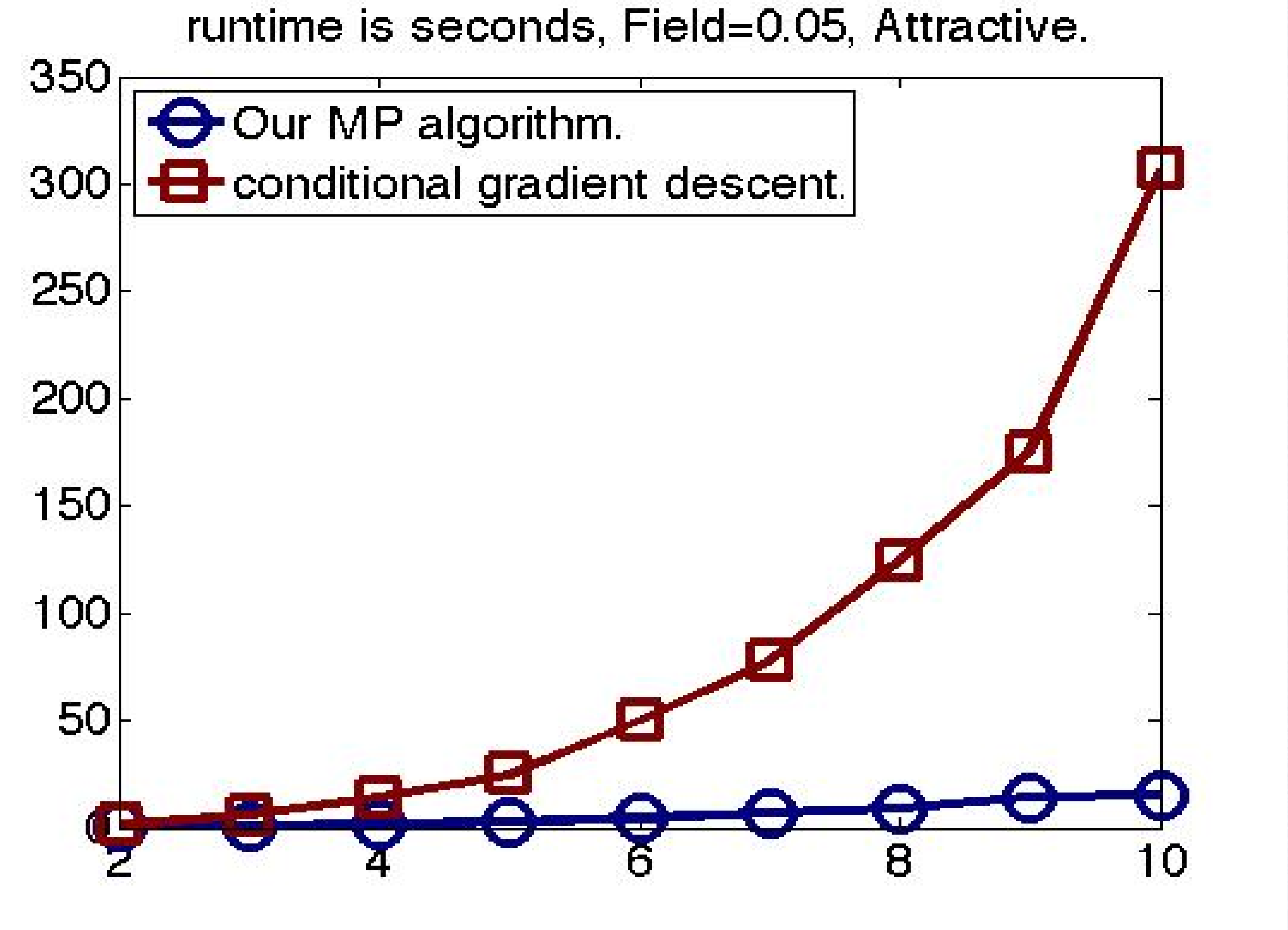
$$\Phi_i(x_i) = \exp(\theta_i x_i)$$

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

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

# Experiments – Efficiency

Seconds



Grid size

# Preliminary Experiments

**Accuracy: Compared to Matlab's linprog** $\theta^T b_\varepsilon - \theta^T b^*$

$$b^* = \operatorname*{arg\,min}_{b_\alpha \geq 0, \sum b_\alpha(x_\alpha)=1} \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha)$$

$$b_\varepsilon = \operatorname*{arg\,min}_{b_\alpha \geq 0, \sum b_\alpha(x_\alpha)=1} \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\tilde{H}$$

$$\tilde{H} = \sum_\alpha c_\alpha H(b_\alpha) + \sum_i c_i H(b_i) + \sum_{\alpha \in N(i)} c_{i\alpha}(H(b_\alpha) - H(b_i))$$

| | ε=1 | ε=0.5 | ε=0.1 | ε=0.01 |
|---|---|---|---|---|
| $c_\alpha=1, c_{i\alpha}=0, c_i=0$ | 2.39 | 0.89 | 0.08 | 0.0002 |
| $\tilde{H} = TRW$ | 1.26 | 0.42 | 0.02 | -9.6e-05 |
| Convex-Bethe | 1.45 | 0.49 | 0.01 | -5.6e-03 |

More accurate than linprog

# Preliminary Experiments

**Number of iterations**

$$b^* = \arg\min_{b_\alpha \geq 0, \sum b_\alpha(x_\alpha)=1} \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha)$$

$$b_\varepsilon = \arg\min_{b_\alpha \geq 0, \sum b_\alpha(x_\alpha)=1} \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\tilde{H}$$

$$\tilde{H} = \sum_\alpha c_\alpha H(b_\alpha) + \sum_i c_i H(b_i) + \sum_{\alpha \in N(i)} c_{i\alpha}(H(b_\alpha) - H(b_i))$$

| | ε=1 | ε=0.5 | ε=0.1 | ε=0.01 |
|---|---|---|---|---|
| $c_\alpha=1$, $c_{i\alpha}=0$, $c_i=0$ | 20 | 45 | 323 | 2025 |
| $\tilde{H} = TRW$ | 156 | 301 | 1378 | 1858 |
| Convex-Bethe | 624 | 1152 | 2141 | 2322 |

# Summary

**The Norm-Product algorithm:** $\arg\min \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha)\theta_\alpha(x_\alpha) - \varepsilon\tilde{H}$

- Message-passing algorithms with (1/ε)-norm. If $-\tilde{H}$ is convex: attains the optimum
- For ε→0 and convex $-\tilde{H}$ approaches the optimal solution of LP-relaxation
- For ε=1 it is a sum-product type. If $-\tilde{H}$ is convex it recovers the optimum. If $\tilde{H}$ = Bethe entropy → sum-product algorithm
- For ε=0 it is a max-product type. For some convex $-\tilde{H}$ there are convergence guarantees. If $\tilde{H}$ = Bethe entropy → max-product algorithm
- Region Norm-Product
- Non-convex Norm-Product