Probabilistic Graphical Models

Lecture 10: Loopy Belief Propagation

Volkan Cevher, Matthias Seeger
Ecole Polytechnique Fédérale de Lausanne

28/10/2011

Outline

Loopy Belief Propagation

- Remember belief propagation? Exact inference for tree-structured graphical models in $O(n)$ (dynamic programming)
- Many graphs in practice have cycles. What to do then?

  Serious Statistician (Lauritzen): Convert graph to junction tree, run BP there (exact inference)

  Computer Scientist (Pearl, . . .):

Loopy Belief Propagation

- Remember belief propagation? Exact inference for tree-structured graphical models in $O(n)$ (dynamic programming)
- Many graphs in practice have cycles. What to do then?

Serious Statistician (Lauritzen):	Convert graph to junction tree, run BP there (exact inference)
Computer Scientist (Pearl, ...):	Run BP anyway, see what you get (approximate inference)

- Loopy belief propagation (LBP): Run BP iteratively, cross fingers that it will converge. Wacky, but enormously successful!

Loopy Belief Propagation

- Remember belief propagation? Exact inference for tree-structured graphical models in $O(n)$ (dynamic programming)
- Many graphs in practice have cycles. What to do then?

Serious Statistician (Lauritzen):	Convert graph to junction tree, run BP there (exact inference)
Computer Scientist (Pearl, ...):	Run BP anyway, see what you get (approximate inference)

- Loopy belief propagation (LBP): Run BP iteratively, cross fingers that it will converge. Wacky, but enormously successful!
- So it converges ...
  - When? Not always (why not?)
  - To what? Always to the same?
  - Corrections that are still feasible?

  None of these questions could seriously be approached before one thing became known: What is LBP doing at all?

Variational Inference

$$\log Z = \sup_{\boldsymbol{\mu} \in \mathcal{M}} \left\{ \boldsymbol{\theta}^T \boldsymbol{\mu} + \mathrm{H}[\boldsymbol{\mu}] \right\}$$

$$\mathcal{M} = \left\{ (\boldsymbol{\mu}_j) \,\middle|\, \boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x}) \right\}$$

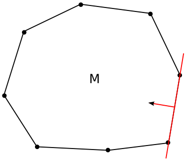
- $\mathcal{M}$ can be hard to fence in
  $\boldsymbol{\theta} \leftrightarrow \boldsymbol{\mu}$ can be hard to compute
  $\mathrm{H}[\boldsymbol{\mu}]$ can be hard to compute
- Variational mean field: Non-convex inner bound to $\mathcal{M}$

The Marginal Polytope

$$\mathcal{M} = \left\{ (\boldsymbol{\mu}_j) \,\middle|\, \boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x}) \right\}$$

Multinomial on graph. Minimal representation.

- $\mathcal{M}$ convex polytope: Described by finite number inequalities.
  Complexity of $\mathcal{M}$: Number of inequalities



- Complexity of $\mathcal{M}$ $\rightarrow$ complexity of exact inference

Local Consistency

$$\mathcal{M} = \left\{ (\boldsymbol{\mu}_j) \,\middle|\, \boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x}) \right\}$$

Multinomial MRF. Pairwise and single node potentials ($|C_j| \leq 2$)

- Cutting away to fence in $\mathcal{M}$. (L)BP steps are local:
  What can we do with local computations? F2

Local Consistency

$$\mathcal{M} = \left\{ (\boldsymbol{\mu}_j) \,\middle|\, \boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x}) \right\}$$

Multinomial MRF. Pairwise and single node potentials ($|C_j| \leq 2$)

- Cutting away to fence in $\mathcal{M}$. (L)BP steps are local:
  What can we do with local computations?

- Clique marginals are distributions

$$\mu_{ij}(\boldsymbol{x}_{ij}) \geq 0, \quad \mu_i(x_i) \geq 0, \quad \sum_{x_i} \mu_i(x_i) = 1$$

- Consistency with neighbours

$$\sum_{x_j} \mu_{ij}(\boldsymbol{x}_{ij}) = \mu_i(x_i)$$

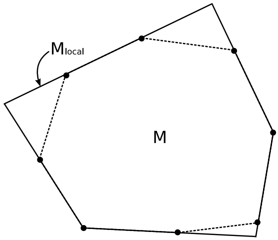
Local Consistency

$$\mathcal{M} = \left\{ (\boldsymbol{\mu}_j) \,\middle|\, \boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x}) \right\}$$

Multinomial MRF. Pairwise and single node potentials ($|C_j| \leq 2$)

- Cutting away to fence in $\mathcal{M}$. (L)BP steps are local: What can we do with local computations?
- Clique marginals are distributions

$$\mu_{ij}(\boldsymbol{x}_{ij}) \geq 0, \quad \mu_i(x_i) \geq 0, \quad \sum\nolimits_{x_i} \mu_i(x_i) = 1$$



- Consistency with neighbours

$$\sum\nolimits_{x_j} \mu_{ij}(\boldsymbol{x}_{ij}) = \mu_i(x_i)$$

- Local consistency polytope $\mathcal{M}_{\text{local}}$: Outer approximation, $\mathcal{M} \subset \mathcal{M}_{\text{local}}$

F2b

Variational Inference

$$\log Z = \sup_{\boldsymbol{\mu} \in \mathcal{M}} \left\{ \boldsymbol{\theta}^T \boldsymbol{\mu} + \mathrm{H}[\boldsymbol{\mu}] \right\}$$

$$\mathcal{M} = \left\{ (\boldsymbol{\mu}_j) \;\middle|\; \boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x}) \right\}$$

- $\mathcal{M}$ can be hard to fence in
  $\boldsymbol{\theta} \leftrightarrow \boldsymbol{\mu}$ can be hard to compute
  $\mathrm{H}[\boldsymbol{\mu}]$ can be hard to compute
- Local consistency polytope $\mathcal{M}_{\text{local}}$: Convex outer bound to $\mathcal{M}$. What about entropy term?

Variational Inference

$$\log Z = \sup_{\boldsymbol{\mu} \in \mathcal{M}} \left\{ \boldsymbol{\theta}^T \boldsymbol{\mu} + \mathrm{H}[\boldsymbol{\mu}] \right\}$$
$$\mathcal{M} = \left\{ (\boldsymbol{\mu}_j) \,\middle|\, \boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x}) \right\}$$

- $\mathcal{M}$ can be hard to fence in
  $\boldsymbol{\theta} \leftrightarrow \boldsymbol{\mu}$ can be hard to compute
  $\mathrm{H}[\boldsymbol{\mu}]$ can be hard to compute
- Local consistency polytope $\mathcal{M}_{\text{local}}$: Convex outer bound to $\mathcal{M}$.
  What about entropy term?
- Why should I care?
  This will show us what loopy belief propagation is doing!

Variational Inference on a Tree

$$\log Z = \sup_{\boldsymbol{\mu} \in \mathcal{M}} \left\{ \boldsymbol{\theta}^T \boldsymbol{\mu} + \mathrm{H}[\boldsymbol{\mu}] \right\}$$

$$\mathcal{M} = \left\{ (\boldsymbol{\mu}_j) \,\middle|\, \boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x}) \right\}$$

How does that look like for a tree graph $\mathcal{G}$? F3

- Multinomial MRF, overcomplete representation by indicators. Marginals $\mu_j(\boldsymbol{x}_{C_j})$ on cliques (factor nodes), $\mu_i(x_i)$ on variables

Variational Inference on a Tree

$$\log Z = \sup_{\boldsymbol{\mu}\in\mathcal{M}}\left\{\boldsymbol{\theta}^T\boldsymbol{\mu} + \mathrm{H}[\boldsymbol{\mu}]\right\}$$

$$\mathcal{M} = \left\{(\boldsymbol{\mu}_j)\,\middle|\,\boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x})\right\}$$

How does that look like for a tree graph $\mathcal{G}$?

- Multinomial MRF, overcomplete representation by indicators. Marginals $\mu_j(\boldsymbol{x}_{C_j})$ on cliques (factor nodes), $\mu_i(x_i)$ on variables
- Tree reparameterization. If the factor graph $\mathcal{G}$ is a tree: F3b

$$P(\boldsymbol{x}) = \frac{\prod_j \mu_j(\boldsymbol{x}_{C_j})}{\prod_i \mu_i(x_i)^{n_i-1}}, \quad n_i = |\{j \mid i \in C_j\}|$$

Variational Inference on a Tree

$$\log Z = \sup_{\boldsymbol{\mu} \in \mathcal{M}} \left\{ \boldsymbol{\theta}^T \boldsymbol{\mu} + \mathrm{H}[\boldsymbol{\mu}] \right\}$$

$$\mathcal{M} = \left\{ (\boldsymbol{\mu}_j) \,\middle|\, \boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x}) \right\}$$

$$P(\boldsymbol{x}) = \frac{\prod_j \mu_j(\boldsymbol{x}_{C_j})}{\prod_i \mu_i(x_i)^{n_i - 1}}, \quad n_i = |\{j \mid i \in C_j\}|$$

What does that mean?

- Marginal polytope $\mathcal{M}$ for tree graph: F4

Variational Inference on a Tree

$$\log Z = \sup_{\boldsymbol{\mu}\in\mathcal{M}}\left\{\boldsymbol{\theta}^T\boldsymbol{\mu} + \mathrm{H}[\boldsymbol{\mu}]\right\}$$

$$\mathcal{M} = \left\{(\boldsymbol{\mu}_j)\,\middle|\,\boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x})\right\}$$

$$P(\boldsymbol{x}) = \frac{\prod_j \mu_j(\boldsymbol{x}_{C_j})}{\prod_i \mu_i(x_i)^{n_i-1}}, \quad n_i = |\{j \mid i\in C_j\}|$$

What does that mean?

- Marginal polytope $\mathcal{M}$ for tree graph:
  Identical to local marginalization polytope: $\mathcal{M} = \mathcal{M}_{\text{local}}$
- Entropy term $\mathrm{H}[\boldsymbol{\mu}]$:

F4b

Variational Inference on a Tree

$$\log Z = \sup_{\boldsymbol{\mu}\in\mathcal{M}}\left\{\boldsymbol{\theta}^T\boldsymbol{\mu} + \mathrm{H}[\boldsymbol{\mu}]\right\}$$

$$\mathcal{M} = \left\{(\boldsymbol{\mu}_j) \,\middle|\, \boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x})\right\}$$

$$P(\boldsymbol{x}) = \frac{\prod_j \mu_j(\boldsymbol{x}_{C_j})}{\prod_i \mu_i(x_i)^{n_i-1}}, \quad n_i = |\{j \mid i \in C_j\}|$$

What does that mean?

- Marginal polytope $\mathcal{M}$ for tree graph:
  Identical to local marginalization polytope: $\mathcal{M} = \mathcal{M}_{\text{local}}$
- Entropy term $\mathrm{H}[\boldsymbol{\mu}]$:

$$\mathrm{H}[\boldsymbol{\mu}] = \sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})] - \sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]$$

Simple function of $\boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$

Variational Inference on a Tree

$$\log Z = \sup_{\boldsymbol{\mu} \in \mathcal{M}} \left\{ \boldsymbol{\theta}^T \boldsymbol{\mu} + \mathrm{H}[\boldsymbol{\mu}] \right\}$$

$$\mathcal{M} = \left\{ (\boldsymbol{\mu}_j) \,\middle|\, \boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x}) \right\}$$

$$P(\boldsymbol{x}) = \frac{\prod_j \mu_j(\boldsymbol{x}_{C_j})}{\prod_i \mu_i(x_i)^{n_i - 1}}, \quad n_i = |\{j \mid i \in C_j\}|$$

For tree graph:

1. $\mathcal{M} = \mathcal{M}_{\text{local}}$ [simple, just $O(n)$ inequalities]
2. $\mathrm{H}[\boldsymbol{\mu}] = \sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})] - \sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]$ [direct from $\boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$]
3. Exact inference

$\Rightarrow$ Wow, inference simple for a tree!

Variational Inference on a Tree

$$\log Z = \sup_{\boldsymbol{\mu}\in\mathcal{M}}\left\{\boldsymbol{\theta}^T\boldsymbol{\mu} + \mathrm{H}[\boldsymbol{\mu}]\right\}$$

$$\mathcal{M} = \left\{(\boldsymbol{\mu}_j)\,\middle|\,\boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x})\right\}$$

$$P(\boldsymbol{x}) = \frac{\prod_j \mu_j(\boldsymbol{x}_{C_j})}{\prod_i \mu_i(x_i)^{n_i-1}}, \quad n_i = |\{j \mid i \in C_j\}|$$

For tree graph:

1. $\mathcal{M} = \mathcal{M}_{\text{local}}$ [simple, just $O(n)$ inequalities]
2. $\mathrm{H}[\boldsymbol{\mu}] = \sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})] - \sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]$ [direct from $\boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$]
3. Exact inference

$\Rightarrow$ Wow, inference simple for a tree!
Knew that before, it's just BP. What's the point?

Variational Inference on a Tree

$$\log Z = \sup_{\boldsymbol{\mu} \in \mathcal{M}} \left\{ \boldsymbol{\theta}^T \boldsymbol{\mu} + \mathrm{H}[\boldsymbol{\mu}] \right\}$$

$$\mathcal{M} = \left\{ (\boldsymbol{\mu}_j) \,\middle|\, \boldsymbol{\mu}_j = \mathrm{E}_Q[\boldsymbol{f}_j(\boldsymbol{x}_{C_j})] \text{ for some } Q(\boldsymbol{x}) \right\}$$

$$P(\boldsymbol{x}) = \frac{\prod_j \mu_j(\boldsymbol{x}_{C_j})}{\prod_i \mu_i(x_i)^{n_i - 1}}, \quad n_i = |\{j \mid i \in C_j\}|$$

For any graph:

1. $\mathcal{M} \subset \mathcal{M}_{\text{local}}$ [simple, just $O(n)$ inequalities]
2. $\mathrm{H}[\boldsymbol{\mu}] \approx \sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})] - \sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]$ [direct from $\boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$]
3. Approximate inference (by loopy belief propagation)

The Bethe Approximation

Nobel Prize Winners affect Machine Learning F5

$$-\log Z \approx \underbrace{-\boldsymbol{\theta}^T\boldsymbol{\mu} - \sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})] + \sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]}_{\text{Bethe free energy}}$$

The Bethe Approximation

$$-\log Z \approx \underbrace{-\boldsymbol{\theta}^T \boldsymbol{\mu} - \sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})] + \sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]}_{\text{Bethe free energy}}$$

Variational Foundation of Loopy BP

Fixed point of loopy belief propagation [Yedidia, Freeman, Weiss, NIPS 13 (01)]

$\Rightarrow$ Saddle point of Bethe free energy, subj. to $\boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$

The Bethe Approximation

$$-\log Z \approx \underbrace{-\boldsymbol{\theta}^T \boldsymbol{\mu} - \sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})] + \sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]}_{\text{Bethe free energy}}$$

Variational Foundation of Loopy BP

Fixed point of loopy belief propagation [Yedidia, Freeman, Weiss, NIPS 13 (01)]
$\Rightarrow$ Saddle point of Bethe free energy, subj. to $\boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$

Make sure you understand: In general,

- $\boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$ not realizable marginals ($\boldsymbol{\mu} \notin \mathcal{M}$), not even at fixed point (call them pseudomarginals)

The Bethe Approximation

$$-\log Z \approx \underbrace{-\boldsymbol{\theta}^T\boldsymbol{\mu} - \sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})] + \sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]}_{\text{Bethe free energy}}$$

Variational Foundation of Loopy BP

Fixed point of loopy belief propagation [Yedidia, Freeman, Weiss, NIPS 13 (01)]

$\Rightarrow$ Saddle point of Bethe free energy, subj. to $\boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$

Make sure you understand: In general,

- $\boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$ not realizable marginals ($\boldsymbol{\mu} \notin \mathcal{M}$), not even at fixed point (call them pseudomarginals)
- $\sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})] - \sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)] \neq \mathrm{H}[\boldsymbol{\mu}]$, not an entropy (can be negative, even for $\boldsymbol{\mu} \in \mathcal{M}$)

F5b

The Bethe Approximation

$$-\log Z \approx \underbrace{-\boldsymbol{\theta}^T \boldsymbol{\mu} - \sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})] + \sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)]}_{\text{Bethe free energy}}$$

Variational Foundation of Loopy BP

Fixed point of loopy belief propagation [Yedidia, Freeman, Weiss, NIPS 13 (01)]
$\Rightarrow$ Saddle point of Bethe free energy, subj. to $\boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$

Make sure you understand: In general,

- $\boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$ not realizable marginals ($\boldsymbol{\mu} \notin \mathcal{M}$), not even at fixed point (call them pseudomarginals)
- $\sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})] - \sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)] \neq \mathrm{H}[\boldsymbol{\mu}]$, not an entropy (can be negative, even for $\boldsymbol{\mu} \in \mathcal{M}$)
- $\propto (\prod_j \mu_j(\boldsymbol{x}_{C_j}))/(\prod_i \mu_i(x_i)^{n_i-1})$ is distribution (of course), but not with marginals $\mu_j(\boldsymbol{x}_{C_j})$ [For tree: Remove not]

Messages are Lagrange Multipliers

$$\mathcal{F}_{\text{Bethe}} = -\boldsymbol{\theta}^T\boldsymbol{\mu} - \sum_j \mathrm{H}[\mu_j(\boldsymbol{x}_{C_j})] + \sum_i (n_i - 1)\mathrm{H}[\mu_i(x_i)], \quad \boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$$

- Local consistency: $\boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$ means that $\boldsymbol{\mu} \succeq \mathbf{0}$ and

$$\sum_{\boldsymbol{x}_{C_j \setminus i}} \mu_j(\boldsymbol{x}_{C_j}) = \mu_i(x_i),\ i \in C_j, \quad \sum_{x_i} \mu_i(x_i) = 1,\ i = 1, \dots, n$$

- Messages $\bigcirc \to \square$, $\square \to \bigcirc$ [recall exercises]

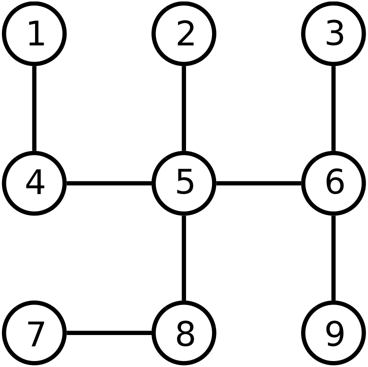
$$M_{i\to j}(x_i) \propto \prod_{j' \in \mathcal{N}_i \setminus j} M_{j' \to i}(x_i),$$

$$M_{j\to i}(x_i) \propto \sum_{\boldsymbol{x}_{C_j \setminus i}} \Phi_j(\boldsymbol{x}_{C_j}) \prod_{i' \in C_j \setminus i} M_{i' \to j}(x_{i'}),$$

$$\mu_i(x_i) \propto \prod_{j \in \mathcal{N}_i} M_{j\to i}(x_i), \quad \mu_j(\boldsymbol{x}_{C_j}) \propto \Phi_j(\boldsymbol{x}_{C_j}) \prod_{i \in C_j} M_{i \to j}(x_i)$$

- Proof [handout – study it!]:
  - Construct Lagrangian
  - Rewrite stationary equations. Log messages turn out to be Lagrange multipliers at fixed point

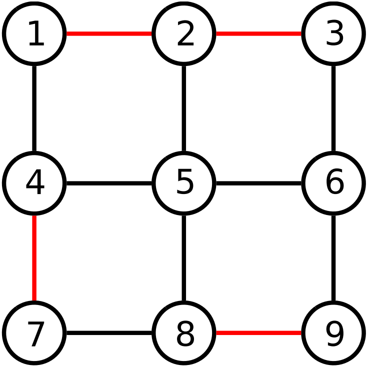
The Bethe Approximation



$$P(\boldsymbol{x}) = \frac{\mu_{14}\mu_{25}\mu_{36}\mu_{45}\mu_{56}\mu_{58}\mu_{69}\mu_{78}}{\mu_4\mu_5^3\mu_6^2\mu_8},$$

$$\boldsymbol{\mu} \in \mathcal{M}_{\text{local}} = \mathcal{M}$$

The Bethe Approximation



$$P(\boldsymbol{x}) \approx \frac{\mu_{14}\mu_{25}\mu_{36}\mu_{45}\mu_{56}\mu_{58}\mu_{69}\mu_{78}}{\mu_4\mu_5^3\mu_6^2\mu_8} \times \frac{\mu_{12}\mu_{23}\mu_{47}\mu_{89}}{\mu_1\mu_2^2\mu_3\mu_4\mu_7\mu_8\mu_9},$$

$$\boldsymbol{\mu} \in \mathcal{M}_{\text{local}} \supset \mathcal{M}$$

Remarks

- Stable fixed point of LBP
  $\Rightarrow$ Local minimum of $\mathcal{F}_{\text{Bethe}}$, subj. to $\boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$ [Heskes, NIPS 15 (03)]

Remarks

- Stable fixed point of LBP
  $\Rightarrow$ Local minimum of $\mathcal{F}_{\text{Bethe}}$, subj. to $\boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$ [Heskes, NIPS 15 (03)]
- Bethe neg-entropy not convex $\Rightarrow$ Bethe problem not convex F8
  - Double loop algorithms guarantee convergence, slower than LBP
  - Bethe neg-entropy convex in special cases (your sheet!)

Remarks

- Stable fixed point of LBP
  $\Rightarrow$ Local minimum of $\mathcal{F}_{\text{Bethe}}$, subj. to $\boldsymbol{\mu} \in \mathcal{M}_{\text{local}}$ [Heskes, NIPS 15 (03)]
- Bethe neg-entropy not convex $\Rightarrow$ Bethe problem not convex
  - Double loop algorithms guarantee convergence, slower than LBP
  - Bethe neg-entropy convex in special cases (your sheet!)
- Generalized belief propagation:
  Use $\mathcal{M}_{\text{local}}$ and Bethe (approximate) entropy on cluster graphs (aka. region graphs): cliques larger than C_j, smaller than in junction tree
  - Big improvement for regular 2D grids
  - Often worse in terms of convergence behaviour

Theory about LBP. Extensions

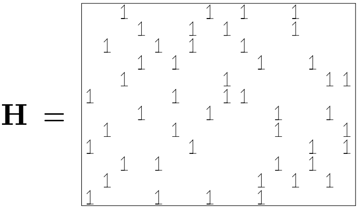
A certainly incomplete list:

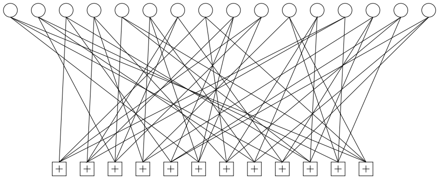
- Convergence analyses
  - Studying computation tree
  - Contraction arguments (unique fixed point!)
  - Convexity of Bethe neg-entropy
- Error bounds on marginals
  - Tree reparameterizations
  - Bound propagation
- Higher-order loop corrections
- Convexifications. Reweighted LBP

Loopy Belief Propagation in Practice

Coding / Information Theory F10

- LDPC codes and LBP decoding revolutionized this field (resurrection of Gallager codes)
- Used from deep space communication (Mars rovers) over satellite transmission to CD players / hard drives
- Theoretical guarantees by density propagation (with high probability over random LDPC graphs)





Courtesy MacKay: Information Theory . . . (2003)

Loopy Belief Propagation in Practice

Computer Vision: Markov Random Fields

F10b

- Denoising, super-resolution, restoration (early work by Besag)
- Depth / reconstruction from stereo, matching, correspondences
- Segmentation, matting, blending, stitching, impainting, . . .

Courtesy MSR

Why LBP for Gaussian Models?

- G-MRFs heavily used in spatial statistics (remote sensing, ...) and in low-level computer vision
- Isn't Gaussian inference tractable? $O(n^3)$ is $\mathrm{poly}(n)$.
  $\Rightarrow$ If $n = 10^7$, $O(n^3)$ is intractable

Why LBP for Gaussian Models?

- G-MRFs heavily used in spatial statistics (remote sensing, ...) and in low-level computer vision
- Isn't Gaussian inference tractable? $O(n^3)$ is $\mathrm{poly}(n)$.
  $\Rightarrow$ If $n = 10^7$, $O(n^3)$ is intractable
- Can't I just use conjugate gradients?
  For means you can (and you should), but not for variances.
  $\Rightarrow$ Convergent LBP is used to precondition CG

Why LBP for Gaussian Models?

- G-MRFs heavily used in spatial statistics (remote sensing, . . .) and in low-level computer vision
- Isn't Gaussian inference tractable? $O(n^3)$ is $\mathrm{poly}(n)$.
  $\Rightarrow$ If $n = 10^7$, $O(n^3)$ is intractable
- Can't I just use conjugate gradients?
  For means you can (and you should), but not for variances.
  $\Rightarrow$ Convergent LBP is used to precondition CG
- What if my signal is just not Gaussian (say: image)?
  Approximate inference for non-Gaussian continuous MRFs needs G-MRFs as major computational backbone

Why LBP for Gaussian Models?

- G-MRFs heavily used in spatial statistics (remote sensing, ...) and in low-level computer vision
- Isn't Gaussian inference tractable? $O(n^3)$ is $\mathrm{poly}(n)$.
  $\Rightarrow$ If $n = 10^7$, $O(n^3)$ is intractable
- Can't I just use conjugate gradients?
  For means you can (and you should), but not for variances.
  $\Rightarrow$ Convergent LBP is used to precondition CG
- What if my signal is just not Gaussian (say: image)?
  Approximate inference for non-Gaussian continuous MRFs needs G-MRFs as major computational backbone
- Advantage of continuous MRFs (based on G-MRFs) over discrete: Global covariances can be extracted. They are what (often) drives experimental design

LBP for Gaussian Models

LBP: Gaussian versus multinomial models

Similar

- Families closed under sum / product
- LBP, Bethe relaxation: Exactly same form

Different

- Inference "just" $O(n^3)$ for Gaussian models
- G-LBP can break down (negative variances) for valid G-MRF
- G-LBP: More known about convergence / correctness (where errors come from)

LBP for Gaussian Models

LBP: Gaussian versus multinomial models

Similar

- Families closed under sum / product
- LBP, Bethe relaxation: Exactly same form

Different

- Inference "just" $O(n^3)$ for Gaussian models
- G-LBP can break down (negative variances) for valid G-MRF
- G-LBP: More known about convergence / correctness (where errors come from)

Also: G-LBP seed of ideas to tackle general LBP

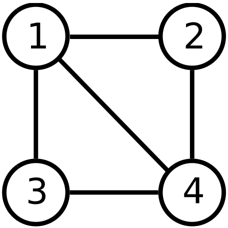
- Computation tree analysis [Weiss *et.al.*, NCOMP 01]
- Tree-based reparameterizations [Wainwright *et.al.*, NIPS 13 (01)]

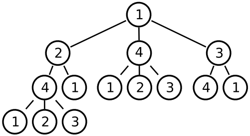
Results for G-MRF LBP

- Whenever LBP converges, the means are exact. Variances not correct in general (except trees) [Weiss *et.al.*, NCOMP 01]
- LBP converges for walk-summable models. [Malioutov *et.al.*, JMLR 06]

$$P(\boldsymbol{x}) \propto \prod_{(ij)\in\mathcal{E}} \underbrace{\Phi_{ij}(\boldsymbol{x}_{ij})}_{\text{Proper Gaussian}}$$

LBP variance estimates properly characterized





Wrap-Up

- Loopy belief propagation: Non-convex variational relaxation. Pretend graph is a tree
- Convergence / approximation error: “How wrong” is that conception?
- Bethe variational problem: Characterization of LBP. Leads to other relaxations [next lecture]
- Gaussian LBP: $O(n)$ approximate inference in Gaussian MRFs. Better characterized than discrete LBP